

BUREAU OF THE CENSUS
STATISTICAL RESEARCH DIVISION REPORT SERIES
SRD Research Report Number: Census/SRD/RR-86/04

AN EVALUATION OF EDIT AND IMPUTATION
PROCEDURES USED IN THE 1982 ECONOMIC CENSUSES
IN BUSINESS DIVISION

by

Brian Greenberg
Thomas Petkunas
Statistical Research Division
Bureau of the Census

This series contains research reports, written by or in cooperation with staff members of the Statistical Research Division, whose content may be of interest to the general statistical research community. The views reflected in these reports are not necessarily those of the Census Bureau nor do they necessarily represent Census Bureau statistical policy or practice. Inquiries may be addressed to the author(s) or the SRD Report Series Coordinator, Statistical Research Division, Bureau of the Census, Washington, D.C. 20233.

Recommended by: Paul Biemer
Report completed: February 6, 1986
Report issued: February 6, 1986
Revised: May 1, 1986

1. EXECUTIVE SUMMARY

The focus of this study has been an analytic evaluation of the overall edit and imputation system used by Business Division to process the Retail, Wholesale and Service segments of the 1982 Economic Censuses. The data employed in this study were not used for a statistical analysis of procedures and edit system actions, but rather these data served as a vehicle to track the flow of the system and were used as a concrete focus during our investigations. Two primary features in every edit and imputation system are subject-matter expertise and operational (often mathematical) procedures through which this expertise is exercised. Computer code is designed to integrate subject-based rules and operational procedures within a coherent framework and to direct the flow of records between the edit and imputation system and the ambient data processing environment and within different segments of the system. In addition, features should be built into the computer code for an edit and imputation system that allow one to monitor its workings and aid in evaluating its performance.

The subject-matter expertise embedded within the Business Division edit and imputation system is of a very high caliber, and the subject-matter based rules for editing data and for imputation were carefully and well conceived. A lack in the overall system, however, is clear documentation and explicit description of rules and procedures. A clear narrative description of rules and procedures would facilitate evaluating, enhancing, and updating the system when needed, and would emphasize the distinction between (1) system rules and (2) code to implement the rules. The procedures built into the system to monitor and reflect its performance could also be improved and extended, in particular, the flagging routines already in the system to track the flow of records should be made more extensive and thorough. Based on an examination of individual establishment records and summary statistics, it is clear that the vast majority of changes made to data records were certainly warranted, and the imputation strategy is sound. Perhaps one of the most notable findings of this evaluation project is that the overall performance of Business Division's edit and imputation system appears to be extremely good.

An important factor that must be considered in the design of the overall system is the interaction between (1) automated routines and (2) individual clerk/analyst review for referral cases. An optimal strategy for an edit and imputation system will include

automated procedures to resolve the majority of cases and individual review for establishment records requiring special handling. This is the design built into the Business Division edit and imputation system and it is, by and large, effective. We did observe in our analysis, however, that the rate of cycling between automated procedures and clerk/analyst review for some kind-of-business classifications examined seemed high, and for some records the interplay between automated routines and clerk/analyst review did not flow smoothly. One remedy may lie in the use of on-line, interactive edit review procedures. Investigations into the development, optimal methods of utilization, and feasibility of such procedures have already begun in the economic area. We recommend that research into such procedures continue and accelerate over the next several years, with the objective of implementing them for the 1992 Economic Censuses.

2. DETAILED FINDINGS

2.1 Introduction

The objective of this project was to provide a broad analysis and evaluation of edit and imputation procedures for data collected in the 1982 Economic Censuses by Business Division. The study plan was to conduct an analytic evaluation of the methodologies and procedures incorporated into the automated routines within the Business Division edit and imputation system and to examine the design of the overall system. Among features considered were automated routines, clerk/analyst review processing, and the interplay between these two. In addition, it was our intention to compare data reported by respondents with the allocations and adjustments derived within the overall edit and imputation processing system. We focused on the edit and imputation system used by Business Division to process the Retail, Wholesale, and Service Censuses. We did not look at every item on the response form, but concentrated on the following basic items: Sales/Receipts, Annual Payroll, First Quarter Payroll, Number of Employees, Operating Expenses, Beginning Of Year Inventories, and End of Year Inventories.

Our first task was to acquire an understanding of the workings of the automated edit and imputation routines. To this end, we prepared a narrative description of the major operations within these routines, using as our source of information decision logic tables prepared by Business Division. In gaining this understanding, we benefited greatly from discussions with staff in Business Division who designed this system. In order to observe the workings of the edit and imputation system on respondent data, we acquired from Business Division all records from a selected set of kind-of-businesses (KB's). After the KB's to be used in this study were identified, all records from these KB's were sent to us after each pass through the edit and imputation processing cycle.

The KB's used in this study were selected by staff in Business Division to exhibit a cross-section of establishment types. That is, we wanted some KB's with many multi-units and some with few, some with primarily large establishments and some with primarily small, and we wanted to include each of these Censuses, or trade areas. Although fourteen KB's were originally selected for this study, we completed our analysis on only six due to costs incurred for computer processing. The six KB's studied in this report are:

Wholesale

509311 --- Iron and Steel Scrap Processors (using power processing equipment)
518100 --- Distribution of Beer and Ale

Service

751210 --- Passenger Car Rental Services (without drivers)
783300 --- Drive-in Motion Picture Theaters

Retail

531110 --- Conventional Department Stores
596110 --- Department Store Merchandise -- Mail Order

(Note: The kind-of-business category is a subclassification within the SIC identified by the first four digits of the classification code. The last two zero digits of 518100 and 783300 indicate that these both represent an entire SIC. Throughout this report we will refer to the KB's under analysis with the understanding that 518100 and 783300 are entire SIC's.) Two primary features in every edit and imputation system are subject-matter expertise and operational (often mathematical) procedures through which this expertise must operate. Computer code is then designed to integrate subject-based rules and operational procedures within a coherent framework and to direct the flow of records between the edit and imputation system and the ambient data processing environment and within different segments of the system. In addition, features should be built into an edit and imputation system that allow one to monitor its workings and aid in evaluating its performance.

The subject-matter expertise incorporated within the Business Division edit and imputation system is of a very high caliber, and the subject-matter based rules for editing data and for imputation were carefully and well conceived. A lack in the overall system, however, is clear documentation and explicit description of rules and procedures. A clear narrative description of rules and procedures would facilitate evaluating, enhancing, and updating the system when needed, and would emphasize the distinction between (1) system rules and (2) code to implement the rules. The procedures built into the system to monitor and reflect its performance could also be improved and

extended, in particular, the flagging routines already in the system to track the flow of records should be made more extensive and thorough. Based on an examination of individual establishment records and summary statistics, it is clear that the vast majority of changes made to data records were certainly warranted and the imputation strategy is sound. Perhaps one of the most notable findings of this evaluation project is that the overall performance of Business Division's edit and imputation system appears to be extremely good.

An important factor that must be considered in the design of the overall system is the interaction between (1) automated routines and (2) individual clerk/analyst review for referral cases. An optimal strategy for an edit and imputation system will include automated procedures to resolve the majority of cases and individual review for establishment records requiring special handling. This is the design built into the Business Division edit and imputation system and it is, by and large, effective. We did observe in our analysis, however, that the rate of cycling between automated procedures and clerk/analyst review for some KB's examined seemed high, and for some records the interplay between automated routines and clerk/analyst review did not flow smoothly. One remedy may lie in the use of on-line, interactive edit review procedures. Investigations into the development, optimal methods of utilization, and feasibility of such procedures have already begun in the economic area. We recommend that research into such procedures continue and accelerate over the next several years, with the objective of implementing them for the 1992 Economic Censuses.

In the body of this report we treat each of the items referred to above in more detail, discuss in further depth each of the recommendations, and exhibit the analysis on which these recommendations are based. In Section 2.2, we present an overview of the edit and imputation system used by Business Division to process the 1982 Economic Censuses for Wholesale, Retail, and Services. We prepared a narrative description of this processing system, which provides more details than are in the body of the text. Copies of this narrative are available from the authors for those interested in examining in more detail the workings of the Business Division edit and imputation system. In Section 2.3 we discuss rates of imputation for missing values and changes to reported data. The discussion and analysis are supported by the tables in Appendixes I, II, and III. The final segment of this report, Section 2.4, discusses the flow of records between the automated routines of this system and clerk/analyst review. In Appendix IV we present a summary

of cycle rates reflecting the flow of records. Throughout this report, we make various recommendations focusing on the specific topic under review.

2.2 Description of Business Division Edit and Imputation Processing System

2.2.1 Introduction

In this chapter we provide an overview of the edit and imputation system used in Business Division to process establishment records for the 1982 Economic Censuses. In subsequent chapters we will discuss some of the features of the system in more depth. In many respects this chapter forms the background for the rest of the report.

We first describe the data found on a typical establishment record which we employed in our analysis. We then describe the automated portion of the edit and imputation system and discuss the interplay between the automated portion of the system and the clerk/analyst review process. Lastly, we make recommendations regarding this system as a whole. A more detailed description of the workings of the automated portion of the edit and imputation system prepared for this report is available from the authors.

2.2.2 Establishment Data Records

Each establishment data record has a large number of fields describing the establishment, its composition, and its performance over the past several years. For the purposes of this study, the data types of primary interest are:

- (a) 1982 reported data,
- (b) 1982 Administrative data,
- (c) 1981 Administrative data,
- (d) 1977 Economic Censuses data,
- (e) 1982 Tabulation data,
- (f) Imputation flags (often referred to as "basic data flags").

For each of the four basic data items; Sales/Receipts (SLS), Annual Payroll (APR), First Quarter Payroll (QPR), and Number of Employees (EMP); each establishment record typically has six fields corresponding to (a)-(f) above. Of course, one or more of these values may be missing. In addition, establishments in the Wholesale sector also have data

fields for Operating Expenses (EXP), End of Year Inventories (INV), and Beginning of Year Inventories (BIN). The fields EXP and INV do not have data types (b) and (c), and BIN does not have data types (b), (c), or (d).

The data type "reported data" is in fact keyed reported data. That is, the value in an establishment data field has been entered or "keyed" onto the data record by a data entry machine operator from the actual response form. To the extent that there are keying errors, the keyed report data on the record will not correspond to the actual report data on the response form^{1/}. If a keying error was made that caused the keyed value to be inconsistent with other field values the edit system should detect it. The tabulation data fields contain the values for each establishment record after adjustments to data fields on the keyed reported data have been made and missing fields have been allocated. After all processing of an establishment record is completed, it is the tabulation data that one treats as the data representing the establishment under consideration, and the tabulation data is used to derive publication statistics released by the Census Bureau. The tabulation data will agree with the keyed reported data record unless data were imputed for blank fields or fields were altered due to edit failures.

For the purposes of this study, a typical establishment record may look like any one of the three following records:

		SLS	APR	QPR	EMP
1.	(a)	1000	83	19	5
	(b)	--	81	20	5
	(c)	800	64	13	6
	(d)	400	30	8	3
	(e)	1000	83	19	5
	(f)	R	R	R	R ,
2.	(a)	33500	83	19	5
	(b)	940	--	20	5
	(c)	800	64	13	5
	(d)	400	30	8	6
	(e)	940	83	19	5
	(f)	AR	R	R	R ,

^{1/} Data entry quality assurance consists of (1) rudimentary field edits (e.g., range checks) performed by the data entry system and (2) sample verification (rekey and compare). The keying errors on a record are those that escaped detection at the time of data entry.

3.	(a)	33500	83	--	5
	(b)	--	81	--	5
	(c)	--	64	13	6
	(d)	--	30	1	3
	(e)	1200	83	19	5
	(f)	IR	R	CP	R .

The dashes above are used to indicate a missing field. The values for SLS, APR, and QPR are in multiples of \$1,000. The value for EMP is the number of employees for the pay period including March 12, 1982. For Wholesale, there are three more columns for EXP, INV, and BIN, each recorded in multiples of \$1000. The labeled rows for each record above correspond to the data types listed at the beginning of this section. For example, row (a) represents keyed reported data, row (b) represents 1982 Administrative Data, etc.

In the first establishment record, all fields were reported and the tabulation values agree with the reported. The "R" in the field for imputation flag denotes that each field was reported. In the second record, SLS was changed from 33500 to 940 and the 1982 Administrative data were used. This is denoted by the flag "A" to the left of the flag "R" (flags are read from right to left); that is, the reported value was changed with new value determined by 1982 Administrative data. In the third record, QPR was not reported and its 1982 Administrative data counterpart was not available. The "P" flag indicates that it was imputed using 1981 Administrative data. The "C" flag to the left of the "P" indicates that this value was later changed by a clerk/analyst to "19". The reported value of SLS on this record is unreasonably large and it was determined to be inconsistent with other fields on the record because of edit failures, and it was to be changed. There are no values for SLS on other data types that can be used to derive an imputation, and the "T" flag indicates that the value 1,200 was determined by the Imputation and Tolerance (I&T) parameters. An establishment record can pass through the automated system more than once. Each time a record passes through the automated system, its cycle number should be incremented by one, and this cycle number also appears on each record.

We have described the fields from the establishment record that were of primary interest in this study. Several other fields on each record (such as "Months in Business") were employed in the analysis of the basic items. They will be introduced only as needed.

2.2.3 Automated Edit and Imputation Programs

As each establishment record is received by Business Division, it is sent through the automated complex edit and imputation programs. (The complex edit and imputation programs are preceded by a number of preliminary routines which make a variety of adjustments to the larger data record. These preliminary programs were not considered in this study, but it is worth noting there is some preparation of the record by Business Division prior to complex edit processing.) The automated edit and imputation system takes one of four actions. First, the system may determine that all fields are reported and consistent, create tabulation data fields equal to the (keyed) reported data, and pass the record out of the edit and imputation system. Second, it may impute for missing fields and alter selected edit-failing fields; determine that the revised record is acceptable; enter the revised record in the tabulation data fields; and, as with the first set of actions, send the record out of the edit and imputation system. The third possibility is that the system may impute for missing fields and/or alter other fields and then direct the record for clerk/analyst review. In this case, the tentatively revised field values are entered into the tabulation data fields and this revised record remains the tabulation data values pending clerk/analyst action to further revise or approve currently residing field values. Under the final possibility, the edit system cannot find a consistent/acceptable set of data values to impute. No revision to the data record is made in such instances and the record is referred to a clerk or analyst for review and correction or allocation.

Records in the third and fourth category (referral cases) are sent to a clerk/analyst for review and analysis. The reviewer may take no action and allow the values currently on the tabulation record (often generated by the automated system) to remain, or he/she may enter a new value in one or more fields of the tabulation data. The establishment data record is then sent through the automated system once again, and any of the four actions described above are again possible. The record being evaluated by the automated system at this point is the tabulation data type created at the previous stage of processing and reviewed (and possibly changed) by a clerk/analyst. This cycle of automated processing and review may be repeated several times. At each stage the cycle number should be incremented and the imputation flag character string should be updated to reflect the most recent changes in each data field. This process does terminate (a clerk/analyst can force it to terminate) and the tabulation data fields will contain the final revised values for each field.

2.2.4 Editing Criteria

The basic complex editing criteria consist of a family of ratio edit consistency checks. That is, the ratios of pairs of fields must lie between two predetermined bounds (entered into the system as parameters). A typical edit is of the form:

$$L_{IJ} \leq F(I)/F(J) \leq U_{IJ}$$

where $F(I)$ and $F(J)$ represent values for fields I and J respectively and L_{IJ} and U_{IJ} are respectively the lower and upper bounds for the ratio of these fields. These bounds are initially determined using 1977 Census data and adjusted for inflation as well as other factors. If all basic edits pass or there is at most one edit failure, the data are considered consistent and it only remains to impute for missing values. If two or more of these edits fail, ratio edits are evaluated comparing keyed reported data to administrative data. Based on the pattern of failed edits for a data record, selected fields may be targeted for change.

2.2.5 Basic Imputation Methodology

Fields targeted for change by the editing routines and missing fields (i.e., left blank by respondents) are imputed within the automated system. If a field is to be imputed, an imputation value is derived by the automated system based upon:

- (a) Other fields on the respondent record (e.g., direct substitution of sums for corresponding detail, etc.),
- (b) 1982 Administrative data (direct substitution),
- (c) 1981 Administrative data (ratio imputation),
- (d) 1977 Census data (ratio imputation with inflation factor),
- (e) Imputation and Tolerance (I&T) parameters.

We will briefly describe the procedures for deriving an imputation. A more detailed description of methods is included in the narrative prepared for this report.

(a) Other fields on respondent record

In a number of instances a suitable source for a missing value can be found in other reported fields. For example, the respondent may omit a total value yet report the

business and INF is an inflation factor (which is field dependent) to reflect changes in the ratio of field J to field I over the years 1977 to 1982. This inflation factor is determined using data from the current surveys. If any of the components for this imputation is not available or if the derived imputation is out of tolerance, then the derived value is not selected as the imputation. In either case, the Imputation and Tolerance parameters are used to derive an imputation.

(e) Imputation and Tolerance (I&T) Parameters

To impute for field J based on companion field I using the I&T parameters, the imputation is (basically):

$$F(J) = F(I)(AVE)(ADJ),$$

where F(J) and F(I) are as above, AVE is the average value of the ratio of values in field J to values in field I, and ADJ is as above. The imputation derived in this fashion usually passes all tolerance tests. If not, the record is targeted for analyst review.

It is important to note that we have described only the broad outline of imputation strategy. In addition to the flow of options described, there are a number of special purpose adjustment procedures employed on a field-by-field, KB-by-KB basis. These are all included under the general heading of special purpose complex edit changes to keyed reported data. They are assigned a "K" in the imputation flag field.

Later in this report we will discuss imputation strategy in greater depth, suggest alterations to the basic flow of imputation options, and provide summary descriptive statistics relating to this process.

2.2.6 Clerk/Analyst Review

As was discussed above, during processing through the automated portion of the edit and imputation system, selected records are targeted for clerk/analyst review. The basic criteria for selecting a record for clerical/analyst review are: (1) large change to a reported value, (2) change or imputation for a large establishment, or (3) failure of the I&T parameter imputation routine to provide an acceptable imputation. The clerk/analyst will examine and evaluate the actions taken by the automated system and, if needed, will further revise the field values in the tabulation record. He/she may

reinstate a reported value altered by the automated system, approve the action of the automated system, or enter into the tabulation field a value obtained or derived independently.

The actual establishment response form is present during the review process and may be used to: (a) determine if data were entered correctly from the form, (b) check for explanatory notes on the form, or (c) derive a valid response value based on other related variables (for example, details summing to a total). A respondent may be called on the telephone to verify, clarify, or provide responses. An analyst may derive an imputation using auxiliary data sources or other related information on the response form. The changes made by the clerk/analyst are then keyed onto the establishment record and the record is processed again through the automated portion of the system.

We have been using the phrase "clerk/analyst" in this report to denote a clerk or an analyst. By and large, the clerk performs more structured tasks (such as verifying response forms) following explicit directions catalogued in an instruction booklet which is designed to include many of the more typical error and resolution situations. The analyst is often a subject-matter specialist who is familiar with the survey subject-matter. He/she has more discretionary power to make changes on an establishment record and usually has more information and auxiliary sources to bring to bear in the review process. (An analyst will be the person making telephone calls to respondents when needed.) The clerk and analyst have been treated jointly in this report primarily because the "C" imputation flag does not distinguish between them.

2.2.7 Discussion and Recommendations

The design built into the Business Division edit and imputation system is well conceived. The blend and interplay of automated routines and individual review is an effective strategy both in the use of resources and treatment of establishment data records. No automated data editing and imputation system can handle every possible data record nor every possible response combination representing establishments with highly atypical attributes during the past operating year. On the other hand, one would not want every establishment record to be individually reviewed, if for no other reason than the number of respondents and attendant costs. An optimal design of a broad edit and imputation system has automated routines that will successfully treat the majority of cases. The automated portion of the system will have some mechanism built into it to recognize

cases which it cannot treat well. In addition, some edit actions, such as large changes or changes to large establishments, also warrant individual review. To the extent that the Business Division's automated edit and imputation routines do successfully resolve the large majority of fields on establishment records and the referral criteria are based on large changes to reported data, changes to large establishments, and unsuccessful derived imputations, it does conform to these criteria. Although the Business Division edit system is well conceived and designed, the recommendations below address some of the needs of this system and suggest methods for improvement.

One noticeable shortcoming in this system is the absence of a clear and comprehensive narrative description. For this project we prepared a narrative description of the main flow and rules built into the automated procedures. This description proved useful as a reference throughout our work. This narrative was developed working from a set of decision logic tables prepared by subject-matter specialists for use by computer programmers in developing code. Having only a set of decision logic tables (or flow chart) to serve as (1) rules, (2) procedures to effect to rules, and (3) computer code to implement the procedures, tends to merge all three and blur the distinction between them. Specifications for computer code and rules that determine a methodology are quite distinct items, and they should not be blended into one. By having a clear statement of basic rules, these rules can be analyzed, enhanced or updated, and evaluated in their own right. In addition, such a document would provide a far more accessible explanation of the system's intent (in contrast to operational procedures). A set of rules distinct from computer specifications can be used in monitoring the performance of an edit and imputation program so that actual program performance can be compared with the intent of the rules. We suggest that future edit and imputation programs have a clear narrative description of the rules to be employed and specifications detailing these rules.

One weak spot in the implementation of the edit and imputation system is the assignment of imputation flags. During our analysis we frequently observed that imputation flags were not correctly assigned to accurately reflect actions of the system. This made the actions of the system more difficult to monitor and hence evaluate. To the best we were able to determine, the vast majority of cases had flags accurately assigned. For records taking a more unusual route through the system, however, the assignment of flags was less accurate. In an analysis project, or in routine monitoring of an edit and imputation

system, it is often these records one wishes to track, and the faulty assignments made it difficult.

An additional shortcoming in the flag assignment rules is that the flag "C", representing clerk/analyst change, does not discriminate sufficiently from among the several types of clerk/analyst actions. One cannot tell from this flag whether the change was made by a clerk or an analyst or, for example, whether the revised value was obtained by observing a keying error in data entry or was based on an analyst derived number. If a clerk or analyst made a correction by dividing reported data by 1000 when the respondent reported in units rather than in thousands, observed a keying error, or noticed a value entered in the wrong place on the response form, the revised number should have virtually the status of a reported value. Other clerk/analyst derived values will have more the status of an imputation. These distinctions can prove valuable in attempting to evaluate clerk/analyst performance and more accurately reflect imputation and change rates. In this study we were able to say little of clerk/analyst behavior (other than frequencies) because the "C" flag only indicates that a value was entered during the review process. It provides no indication of how or why. In addition, when a clerk/analyst reviews and approves the action of the automated routines on a referral record, no flag is currently added to the imputation flag string. Also, knowing that a reported item or a change or imputation made by the automated system or item was reviewed and approved by a clerk/analyst would be valuable for evaluation and monitoring. We recommend that the single "C" flag be replaced by at least several other flags to enable monitoring of the clerk/analyst activities. For example, there should be flags for: (a) keyed data divided by 1000, (b) detected and corrected keying error, (c) value supplied by respondent on telephone or through correspondence, (d) direct inference from other values on response form, (e) reviewed current tabulation value and found acceptable.

The recommendations made above focus on enhancements that will improve capabilities to monitor the system and describe its workings. The actual performance of the system, to the best we were able to determine, appears quite good for the data items examined in this report. In general, an edit and imputation system is a blend of subject-matter expertise and computer programs to exercise this expertise. It is evident that the subject-based expertise incorporated into this system is of a high caliber. This expertise would be more effective if documented better. In subsequent sections we will make additional recommendations, as appropriate, primarily relating to the flow of the overall system.

2.3 Analysis of Imputation Rates and Changes to Reported Data

2.3.1 Introduction

When an establishment data record is entered into the automated edit and imputation routines, values for missing data are allocated and inconsistent values are detected and corrected. In this chapter, we discuss the rates of imputed fields, reported fields, and altered fields on the tabulation data file. In addition, for tabulation field values which were derived by the edit and imputation system, we investigate the source of these values. This analysis must be addressed on a field-by-field, KB-by-KB basis since a missing or erroneous response can occur for any field independently from any other, and these rates differ among KB's and between fields.

2.3.2 Sources of Data on the Tabulation File

In Appendix I we present a family of tables showing imputation and adjustment rates by field by KB. In the table below, we present the first column of Table 5 from Appendix I and explain how to read this table. This column presents the reported, imputed, and change rates for field SLS for KB 751210.

Reported Data Retained	2,241 (64.4 %)
	\$2,892,711 (83.0 %)
Reported Data Changed	80 (2.3 %)
	\$101,140 (2.9 %)
Imputed From Blank	1,158 (33.3 %)
(Nonresponse or nonmailout)	\$489,784 (14.1 %)
Totals	3,479 (100 %)
	\$3,483,635 (100 %)

The number 2,241 on the first line indicates that there are 2,241 cases for which SLS was reported and not changed. This number represents 64.4% of the total number of cases;

3479 on the second from last line. The number 2,892,711 represents a total of 2,892,711 thousands of dollars on the tabulation file reported by respondents included on the line above. That is, 83.0% of the total dollar amount of SLS for KB 751210 (3,483,635 thousands of dollars) was reported by respondents and not changed.

The remaining rows above are read similarly, and thus, each of the columns in Tables 1-6. That is, rows three and four indicate that 80 cases, accounting for 101,140 thousands of dollars had values on the tabulation data fields different from the keyed reported data. These values were respectively 2.3% and 2.9% of the corresponding totals on the last two lines. In each of the columns except EMP, the amount represents thousands of dollars; for EMP, the amount represents the number of employees for the pay period including March 12, 1982.

For most tables the percentage of amounts reported and retained exceeds the percentage of cases having reported retained data. (The reverse relationship exists for the row representing imputed from blank.) This is the case because the Census Bureau receives a higher percentage of responses from large establishments than from small establishments. One reason for this is that for each KB a number of small cases are not mailed at all, and their values are imputed directly from administrative data records. Even if we only consider mailout cases, however, the relationship between percentage of reported amounts compared with percentage of reported cases still prevails, in part, because extensive follow-up efforts are made to collect information from larger establishments.

For KB 531110, department stores, one can see from Table 3 of Appendix I that the reported data retained rate is very high, both for amounts and number of cases.

This is true, in part, because all establishments in this KB are mailed a questionnaire. Further, by definition, department stores must be fairly large. This, moreover, accounts for the quite small percentages in the imputed category. The reported data changed category is rather large for this KB, and we note this in passing here and will discuss it more in Section 2.3.3.

It is clear after examining Tables 1-6 that rates in each category vary from field to field but, more notably, from KB to KB. These rates, especially the imputed rate, are a factor of reporting behavior by establishments in each KB, percentage of establishments not mailed a questionnaire, and level of follow-up efforts. Rates for reported data changed

uniformly low (with the exception of KB 531110). These rates are not solely a function of the edit and imputation system but rather reflect the accuracy of reporting by respondents. In the next two sections we (1) discuss the source of data values that appear on each field when the reported data is not available or is not retained and (2) present a profile of changes to reported data.

2.3.3 Sources of Imputes

The next question studied was: what is the source of field values for the final tabulation data fields for values other than reported data retained? That is, we felt it would be of interest to know how often 1982 Administrative data, 1981 Administrative data, 1977 Census data, or the I&T parameters served as the basis of an imputation. We also sought to determine what percentage of the data was determined by a clerk/analyst. (As noted earlier, it is not possible to know the source of the data values assigned by a clerk/analyst due the lack of information conveyed by the "C" flag.) Such an analysis must be approached on a field-by-field, KB-by-KB basis as was the analysis in Section 2.3.2, above. The information on sources of imputes is contained in Appendix II, Tables 7-12, and we extract a portion of Table 7 below to discuss and interpret.

The table below shows the sources of data for the field APR for SIC 596110.

NUMBER OF FIELDS TAB VALUE OF FIELDS		3467 467602		
SOURCE OF FINAL TAB VALUE:				
	NUMBER OF CASES	AGGREGATE VALUE	% OF TAB	% OF IMPUTES
REPORTED DATA	2719(78.4%)	443605(94.9%)	
1982 ADMIN. DATA	522(15.1%)	17928(3.8%)	(74.7%)
1981 ADMIN. DATA	27(.8%)	376(.1%)	(1.6%)
1977 CENSUS DATA	6(.2%)	37(.0%)	(.2%)
I&T PARAMETERS	137(4.0%)	2385(.5%)	(9.9%)
77 DATA AVAIL.	49	875		
77 DATA NOT AVAIL.	88	1510		
CLERK/ANALYST CHANGE	33(1.0%)	1926(.4%)	(8.0%)
COMPLEX EDIT CHANGE	12(.3%)	19(.0%)	(.1%)
BLANK SET TO ZERO	8(.2%)	0(.0%)	(.0%)
CHANGE UNSPECIFIED	3(.1%)	1326(.3%)	(5.5%)
IMPUTE TOTALS =	748(21.6%)	23997(5.1%)	(100.0%)

The first line of the table, Reported Data, corresponds to reported data retained. Thus there were 2,719 cases of reported data retained, 78.4% of the total number of cases (3,467), accounting for 443,605 thousands of dollars, 94.9% of the total amount in this field (467,602 thousands of dollars). This line corresponds to the first two rows in Table 4, Column 2. Data were assigned to this category whenever the tabulation value was equal to the value of the keyed reported data. There were a number of imputation flag combinations that could accompany this assignment. First, the sole flag could be an "R". Second, the flag could be a "D" which indicates that the appropriate field was left blank but a value was directly allocated based on other reported values. For example, a missing total might be allocated based on the sum of reported detail items. Finally, the flag string might look like: "CIR". This would mean that a value was reported, changed by the automated edit system using the I&T parameters, and then the reinstated (as originally reported) in the tabulation field by a clerk/analyst. Note that the flags are read from right to left.

The next four lines indicate imputed values based on, respectively; 1982 Administrative data, 1981 Administrative, 1977 Census data or the I&T parameters. The first four columns of numbers are interpreted as in the case for reported data. The last column represents the percentage of imputes from each source. That is, the 1982 Administrative data were responsible for 74.7% of the total imputed value, the 1981 Administrative data based imputation accounted for 1.6%, and so on. Under the line for the I&T parameter imputes, we break out the number of cases for which 1977 Census data were available or were not available. We do this to try to determine if the I&T parameters were used primarily when 1977 Census data were not available or whether the I&T parameter based imputation was used in favor of that derived from the 1977 Census data.

Looking at each of the subtables of Tables 7-12, one observes that the imputations based on the I&T parameters were used more frequently than those based on 1977 Census data. This was surprising for two reasons. It would seem reasonable that an imputation based on auxiliary information (in this case 1977 Census data) about an establishment itself would be superior to one based on an estimate of a population parameter such as the I&T parameters. Secondly, the 1977 Census data imputation option is considered before the I&T parameter option, and is used as the imputation if within tolerance. After first observing this phenomenon, we kept count of the times 1977 Census data were

available when the I&T parameters were the source of imputation to see if the I&T parameters were used only in the absence of 1977 Census data. We observed that 1977 Census data frequently were available for the cases imputed by the I&T parameters. As noted earlier, if the 1977 Census data were neither respondent provided nor obtained from 1977 Administrative data, the 1977 Census data were not used as a basis for imputations. This rule most likely accounts for the use of I&T parameters even though 1977 Census data values were available. Several secondary reasons may be that: (1) due to establishment changes over time, the 1977 Census data values may not be a good predictor of the corresponding 1982 data, (2) the method of deriving a 1982 estimate from 1977 data is not optimal, or (3) the workings of the edit system do not always implement the rule of giving first preference to 1977 Census data when appropriate. We recommend that further investigations be undertaken to determine why the imputations based on I&T parameters are employed more frequently those based on 1977 Census data when the 1977 data are available.

One can see clearly from Tables 7-12 that the use of 1982 Administrative data to impute for SLS was quite limited in comparison to APR, QPR, and EMP. This is the case because there are no 1982 Administrative data for multi-unit establishments for SLS and also because the administrative data for SLS is frequently missing for single unit establishments. There are no administrative data for EXP, INV, and BIN and hence administrative data were never a source for their imputation.

The next line describes the imputations due to clerk/analyst change. The line following, "complex edit change," covers a variety of KB specific and field specific subject-based corrections to reported data. For example, in KB 531110 department stores should omit from their SLS total sales by departments leased out to other concerns which are not owned by the same company that operates the department store. If this was not done and is detected by the edit system, an appropriate amount will be subtracted from SLS. These types of adjustments are highly field by KB specialized and are covered under the complex edit change category, and are denoted by a "K" flag. The category, "change unspecified," consists of changes to reported data which we could not trace. Typically the keyed reported data and the tabulation field value are different, yet the only flag is an "R". This represents an error in flag assignment. In the case of KB 531110, many of the entries in the change unspecified category should have been in the complex edit change category for the reasons cited above.

Summing the appropriate columns for all imputation sources equals the last line, "impute totals." This line, in turn, corresponds exactly to the sum of "reported data changed" and "imputed from blank" in Table 4, Column 2. Thus, in Tables 1-6 one can see how much data was imputed; using Tables 7-12 one can see how it was imputed.

2.3.4 Magnitude of Changes to Keyed Reported Data

In Section 2.2.3 we discussed the rate of changes to keyed reported data, and we indicated the proportion of the final tabulated data derived by changing keyed reported amounts. In this section, we analyze the magnitude of changes to keyed reported data (as opposed to imputations for missing fields). As part of this discussion we will consider sources of error in the keyed reported data, clerk/analyst role in making changes, and overall effect of changed data on the final tabulation file. As in the previous sections, this analysis must proceed on a field-by-field, KB-by-KB basis since the reason for erroneous field values and the methods of resolution differs greatly by field and KB. Looking only at magnitudes of change in the absence of rationale for change can prove to be quite misleading. That is, a large change to a keyed reported data value is neither good nor bad in itself, but rather is good or bad depending on the appropriateness of the change.

One of the most common errors in reported data occurs when respondents report amounts in single dollars rather than thousands of dollars as per instructions. Such a reporting error only occurs in fields requiring a dollar amount response. This includes all fields under consideration in this report, with the exception of EMP. This sort of error is easy to detect and correct, and this corrected value, in many respects, has the status of a reported value rather than an imputed one. Another source of an invalid data field on an establishment record is due to an error from keying the response form. If the error was made in one of the last digits of a response field, the keyed value would probably pass edits and be treated as a consistent response. If, however, the error occurred in a leading digit or added a digit or two, the keyed value would likely be quite noticeably out of bounds. In each of the circumstances above, the revised value will be quite different from the keyed reported value; and the magnitude of change will be great. Yet the change is quite warranted, and not making this change would be poor processing.

For other cases, the changes made by the automated edit system to keyed reported data were fairly small. Such changes will often fit in the category "complex edit changes" as

discussed earlier, for example, subtracting leased departments from department store sales. Other small changes occur when the automated edit and imputation system detects edit failing responses which are also different from the corresponding administrative data. In these cases the imputed values often are derived from the 1982 Administrative data records, and at times a relatively small change will ensure consistency.

As noted earlier, for large changes to keyed reported data, the establishment record was referred for clerk/analyst review. As part of this review process the actual respondent form was examined to determine if the edit system change was appropriate. If not, the clerk/analyst entered a corrected value into the tabulation field and a "C" was assigned as the imputation flag. Accordingly, large changes to reported data were closely examined by a clerk/analyst and deemed suitable or corrected.

To the extent that large changes to reported data were carefully scrutinized, the next question to address is: what percentage of total change to reported data was contributed by the large change cases? The answer, which will be discussed more below, is that for (virtually) each field, the major portion of total change was contributed by very few cases, and these cases were reviewed during the referral process. This tendency is far more pronounced in fields representing dollar amounts because of the changes due to correcting for reports in units rather than thousands.

In order to formalize the observation that relatively few records accounted for the vast majority of total change we proceeded as follows. We formed a "change file" for each field by KB consisting of records for which the keyed reported value was different from the tabulation field value. Let x_i be the keyed reported value for case i and let y_i be the tabulation field value for case i , for all $i=1, \dots, N$, where N is the number of cases in this file. Furthermore, let

$$d_i = |x_i - y_i|,$$

and assume the cases are ordered such that

$$\text{if } i \geq j \quad \text{then } d_i \leq d_j$$

for all $i, j=1, \dots, N$. That is, we order the cases by the absolute value of the difference

between keyed reported and tabulation values. Let

$$D = \sum_{i=1}^N d_i$$

be the total amount of change, so that

$$d_i/D$$

represents the proportion of the total change contributed by the i^{th} case. We then form

$$q_i = \left(\sum_{j=1}^i d_j / D \right) 100\%$$

and

$$p_i = (i/N) 100\%$$

for $i=1, \dots, N$. That is, q_i represents the percentage of total change contributed by cases for which the change was equal to or greater than the change for case i . The value p_i represents the percentage of cases on the "change file" for which the change was greater than or equal to the change for case i .

Then for each change file (field by KB) we listed all cases, ordered by descending absolute difference, and printed out the corresponding q_i and p_i for each case. By a cursory examination of these listings, it was clear that for all fields (with the possible exception of EMP) and all KB's (except 531110), approximately 5% of the cases contributed over 90% of the total change. Many of these large change cases were due to reporting in units rather than thousands; many seemed to be keying errors; and almost all were reviewed by a clerk/analyst. To display this phenomenon, we prepared the graphs in Appendix III, Figures 1-6. On the horizontal axis of each graph we have the cumulative percentage of cases in the change file, and on the vertical axis we have the cumulative percentage of change; that is, we plot the values q_i against p_i .

In almost every graph we find a very steep initial rise followed by an almost horizontal tapering off. One can easily see that a small percentage of cases contributed to a large cumulative change, and beyond a certain point the remaining cumulative change is small. This phenomenon was not always present for EMP. Since one does not report in

thousands in that field, the corresponding correction was not necessary. We did examine the records of establishments contributing to a large change in EMP and noticed what appeared to be keying or transcription errors. We systematically examined other records having large changes and found that, based on other fields on these records and the corresponding administrative data, the changes were warranted.

For virtually all large change records, the "C" flag was present to indicate a clerk/analyst change. Here again, the lack of information in this flag prevented us from determining the basis of the corrected value. It was easy to tell when the keyed data were divided by one thousand, and we could frequently infer when there was a keying error, but beyond that, it was impossible to tell. If a clerk/analyst reviews a record and accepts the derived imputation or change made by the automated system, no review flag is set. Since there is no "C" flag one cannot tell when a clerk/analyst reviewed a record if no change was made. One can reasonably infer that many of the change cases were reviewed, but no record of this is available.

As can be seen from Tables 1-6 of Appendix I, the proportion of cases and amounts of data reported and changed is uniformly small. If one subtracts from those proportions corrections due to keying errors, dividing by 1000, callbacks, and similar types of adjustments and treats them as virtually reported data, these rates diminish even further. That is not to say one need not be concerned with the impacts of changes to reported data. When viewing subnational areas, some of these rates might be fairly high. Moreover, any attempt to do meaningful analysis on a longitudinal establishment file will be hampered by imprecise imputations or unwarranted changes.

2.3.5 Source of Parameters

In the discussions of the automated edit and imputation routines, reference was made to several families of parameters. Although the focus of this study is an analytic review of the overall edit and imputation system, rather than an evaluation of the parameters employed within the system, a few remarks about them may be of interest. Parameters are used in the automated routines both as upper and lower bounds for ratio edits and to derive an imputation based on the I&T parameter routines, (referred to as "AVE" in Section 2.2.5).

The upper and lower bounds of ratios used to edit two fields, say, I and J, are obtained from 1977 Census data. One considers the ratio,

$$F_m(I)/F_m(J) = r(m)$$

where $F_m(I)$ and $F_m(J)$ are the values of fields I and J for the m^{th} case, and these cases are selected from final tabulation records on the 1977 Census data file. The values L_{IJ} and U_{IJ} are approximately the lower and upper 2% - 3% cutoff values, respectively, for the distribution of the set of values $\{r(m)\}$. For an arbitrary record, k, if

$$L_{IJ} > F_k(I)/F_k(J) \text{ or } F_k(I)/F_k(J) > U_{IJ},$$

then the edit between fields I and J is said to fail. Based on the pattern of edit failures, one of these fields may be targeted for change.

Even as an automated review system begins to receive and edit current data, there is a need to have reasonable edit parameters. It is common practice to derive these bounds from an auxiliary data source. In the scenario under review, the editing bounds are derived from 1977 Census data and applied to 1982 Census data. Using bounds determined by 1982 Census data found acceptable (after screening by 1977 Census data bounds) may be preferable. However, since the major portion of changes to data were due to large changes to relatively few cases, and since these outliers would likely fail both 1982 Census data and 1977 Census data derived bounds, they would usually be detected in either case. Thus, the overall impact on data would probably change little using either set of bounds.

The parameters used in the I&T parameter imputation routines evolved in two stages during the processing of 1982 Economic Censuses data. As was noted earlier, to impute for field J, given field I, one lets

$$F(J) = F(I) (AVE)(ADJ)$$

where $F(J)$ is the imputed value for field J and $F(I)$ is considered a valid value for field I.

At the onset of processing, AVE is set equal to

$$\sum_{k=1}^N G_k(J) / \sum_{k=1}^N G_k(I)$$

where $G_k(J)$, $G_k(I)$ are values of fields J and I for the k^{th} case, for $k=1, \dots, N$, and these N cases are taken from final tabulation records from the 1977 Census data file. The value, AVE, is said to be a "cold-deck" parameter. At a certain stage of processing, the value of AVE is recomputed to be

$$\sum_{k=1}^M F_k(J) / \sum_{k=1}^M F_k(I)$$

where $F_k(J)$ and $F_k(I)$ are values of fields J and I for the k^{th} case, for $k=1, \dots, M$. These M cases are taken from final tabulation records from the 1982 Census data file. The new value of AVE is said to be "a warm-deck" parameter.

Data for the 1982 Economic Censuses were processed through the edit and imputation system on a flow-basis. That is, as soon as records start arriving, they were processed, (prior to receiving all returns). The value of AVE was initially set to the cold-deck value and later was updated to the warm-deck value. We were informed by staff from Business Division that the final I&T parameters were comprised primarily of warm-deck values for AVE. The cold-deck values for AVE were retained in cases where the number of 1982 observations were too small to provide a reliable AVE at the time adjustments were to be made to these parameters. Although the imputation flags should have distinguished between the two I&T parameter values, only the "T" flag was ever set. Thus, we were not able to distinguish the use of one from the other in this study.

2.4 Processing Flow

2.4.1 Introduction

As discussed earlier, every edit and imputation system must integrate subject-based rules with operational procedures to implement these rules. The subject-based rules, operational (often mathematical) procedures, and techniques to integrate them must then be converted to computer code and, hence, to a coherent computer program. In addition procedures must be designed to direct the flow of records between the edit and imputation system and the ambient data processing environment and within the different segments of the overall edit and imputation system. One aspect of this flow is the

cycling of referral records from the automated routines to a clerk/analyst for review and back again.

2.4.2 Cycle Counts

Within the automated edit and imputation routines, selected records were targeted as referral cases and were directed for clerk/analyst review. The criteria are: (1) large change to reported data, (2) imputations for large establishments, and (3) unsuccessful imputation of a value that will pass tolerance checks. The clerk/analyst will review referral cases, make adjustments if needed, and send establishment records back through the automated edit and imputation routines.

The automated routines may accept the clerk/analyst changes and send the record to the tabulation record file, or they may further adjust the tabulation data. In the latter case, the system may send the revised record directly to the tabulation file or it may, once again, direct the record for clerk/analyst review. Each time a record passes through the automated routines, its cycle number should be increased by one, and each time an adjustment is made to a field on the tabulation record, a new flag character should be added to the imputation flag string. Thus, for a particular establishment record the cycle number may be, say, six, and for an individual field, the imputation flag string may look like: CHCHCR.

In Appendix IV, Table 13, we provide final cycle rates for records in each of the KB's examined in this report. These rates vary greatly by KB. For some KB's, most records were resolved successfully within one or two cycles; for others, more cycles frequently were needed to resolve a problem. It should be noted that the decision to send a record to clerk/analyst review may be made based on fields other than the basic items treated in this report (the so-called trailer items). Nevertheless, a number of records have long imputation flag strings on basic items, indicating that, to some degree, they also were responsible for the cycling. As individual clerk/analyst review of referral cases is a time consuming (and hence, costly) aspect of the edit and imputation process, successful resolution of as many referral records as possible in as few cycles as feasible should be a major goal.

For some fields in establishment records, we observed an "S" flag, which indicates that additional administrative data were received for that field. Thus, an imputation flag

string might look like: ASPR. That is, the data field was reported; changed with an imputation based on 1981 Administrative data; and finally reimputed based on 1982 Administrative data following the receipt of updated 1982 Administrative data. Cycling based on the receipt of additional auxiliary information is highly desirable and enhances the quality of imputation.

However, when records cycle with no new auxiliary information and the automated system seems to be vying with the clerk/analyst for the final say, as in CHCHCR, one should seek methods to make the flow more effective and, hence, more time and cost efficient. The cycle of automated routines followed by clerk/analyst review and back again involves (1) processing records at headquarters, (2) sending referral listings to Jeffersonville, (3) hand corrections to referral documents, (4) keying of corrections, (5) sending corrections back to headquarters, and (6) a subsequent cycle of processing at headquarters. There are ample opportunities for delays and new errors (for example, in keying), and these multiply as the number of cycles grows. It is not clear whether cycling above the third cycle in the absence of additional auxiliary information represents more care in record processing or indicates a weakness in the interplay between automated routines and clerk/analyst review.

2.4.3 Recommendations

In order to make the cycle of automated routines followed by clerk/analyst review more efficient, active consideration is currently being given to a form of on-line editing. The proposed processing envisions a clerk/analyst performing the review of the automated edit and imputation actions using a computer terminal (or microcomputer) and keying corrections onto a data file. By conducting the review in this manner, it is hoped to streamline the processing, avoid additional keying errors, and eliminate the need for paper referral documents. Although such a step will contribute to making this process somewhat more efficient, the same general flow of records (sans paper) will still prevail. It is not clear how an on-line entry of clerk/analyst corrections will reduce the number of cycles needed to resolve referral records or treat the underlying causes of higher cycles.

We recommend that increased consideration be given to investigating a genuine on-line, interactive treatment of edit referral cases during clerk/analyst review. That is, we propose a system that will evaluate a clerk/analyst adjustment interactively, on-line. A

value entered onto the record by a clerk/analyst should be evaluated at the moment of data entry for edit failures and consistency. If a clerk/analyst should desire to over-ride the usual edits, he/she could do so by assigning a multiplier. However, keying errors or inadvertent inconsistencies can be detected and corrected immediately, thus obviating the need for additional cycles through the automated edit and imputation routines.

3. RELATION OF THIS STUDY TO SIMILAR STUDIES

The four studies listed below were undertaken as part of the 1982 Economic Census Evaluation Task Force. They each analyze some aspect of the impact of Census Bureau processing of individual establishment records from the 1982 Economic Censuses. The study by Dyke has the broadest focus and examines reported and processed data after each stage of the processing cycle. The study by Ernst analyzes the edit and imputation procedures employed for the 1982 Census of Construction Industries. The Ramos, Waite, Cole report examines the imputation methodology used in the 1982 Census of Manufactures. In the Corby study, estimates for totals of selected response items were compared using (1) reported values, (2) reinterview values, and (3) values obtained through Census Bureau processing.

Corby, Carol (1985). "Content Evaluation of the 1982 Economic Censuses -- Petroleum Distributors," Bureau of the Census, 1982 Economic Censuses Task Force report.

Dyke, T. Christopher (1985). "1982 Economic Censuses Processing Study," Bureau of the Census, 1982 Economic Censuses Evaluation Task Force report.

Ernst, Lawrence R. (1985). "Large Observation Study for the 1982 Census of Construction Industries," Bureau of the Census, 1982 Economic Census Evaluation Task Force report.

Ramos, Magda, Waite, Preston J., and Cole, Stacy, J. (1985). "Evaluation of the Imputation of Small Manufacturing Companies in the 1982 Census of Manufactures," Bureau of the Census, 1982 Economic Census Evaluation Task Force report.

Acknowledgement

The authors would like to thank Edward Walker, Business Division, for his assistance throughout this project and for his valuable comments and suggestions on an earlier version of this report. They also thank Carl Konschnik, Business Division, and Robert T. O'Reagan and T. Christopher Dyke, Statistical Research Division, Bureau of the Census, for their comments and suggestions on the earlier version of this report.

APPENDIX I

In this appendix we include Tables 1-6, which provide rates of reported data retained, changed data, and data imputed from blanks appearing on the tabulation file. These tables are discussed in the text, Section, 2.3.2, and they are presented by field within KB.

SIC 509311

	SLS	APR	QPR	EMP
REPORTED DATA RETAINED	1121 (52.3%) 2039480 (57.8%)	697 (32.5%) 266657 (67.0%)	691 (32.2%) 71514 (67.8%)	698 (32.6%) 16379 (61.9%)
CHANGED FROM REPORTED	91 (4.2%) 156987 (4.4%)	40 (1.9%) 10510 (2.6%)	27 (1.3%) 1648 (1.6%)	39 (1.8%) 611 (2.3%)
IMPUTED FROM BLANK	931 (43.4%) 1334191 (37.8%)	1406 (65.6%) 120892 (30.4%)	1425 (66.5%) 32260 (30.6%)	1406 (65.6%) 9455 (35.8%)
TOTALS	2143 (100.0%) 3530658 (100.0%)	2143 (100.0%) 398059 (100.0%)	2143 (100.0%) 105422 (100.0%)	2143 (100.0%) 26445 (100.0%)

KEY: FIRST NUMBER REPRESENTS NUMBER OF RECORDS
SECOND NUMBER REPRESENTS TAB VALUE OF RECORDS

Table 1

SIC 509311

	EXP	INV	BIN
REPORTED DATA RETAINED	899 (42.0%) 566855 (59.0%)	791 (36.9%) 263953 (60.2%)	788 (36.8%) 307153 (64.2%)
CHANGED FROM REPORTED	233 (10.9%) 83054 (8.6%)	110 (5.1%) 17401 (4.0%)	109 (5.1%) 22721 (4.8%)
IMPUTED FROM BLANK	1011 (47.2%) 310865 (32.4%)	1242 (58.0%) 157440 (35.9%)	1246 (58.1%) 148347 (31.0%)
TOTALS	2143 (100.0%) 960774 (100.0%)	2143 (100.0%) 438794 (100.0%)	2143 (100.0%) 478221 (100.0%)

KEY: FIRST NUMBER REPRESENTS NUMBER OF RECORDS
SECOND NUMBER REPRESENTS TAB VALUE OF RECORDS

Table 1 (continued)

SIC 518100

	SLS	APR	QPR	EMP
REPORTED DATA RETAINED	3651 (77.5%) 17447497 (83.8%)	2608 (55.4%) 1504992 (81.0%)	2547 (54.1%) 349800 (80.7%)	2610 (55.4%) 68436 (78.2%)
CHANGED FROM REPORTED	158 (3.4%) 779475 (3.7%)	82 (1.7%) 54223 (2.9%)	71 (1.5%) 8011 (1.8%)	89 (1.9%) 2259 (2.6%)
IMPUTED FROM BLANK	900 (19.1%) 2599592 (12.5%)	2019 (42.9%) 298947 (16.1%)	2091 (44.4%) 75735 (17.5%)	2010 (42.7%) 16860 (19.3%)
TOTALS	4709 (100.0%) 20826564 (100.0%)	4709 (100.0%) 1858162 (100.0%)	4709 (100.0%) 433546 (100.0%)	4709 (100.0%) 87555 (100.0%)

KEY: FIRST NUMBER REPRESENTS NUMBER OF RECORDS
SECOND NUMBER REPRESENTS TAB VALUE OF RECORDS

Table 2

SIC 518100

	EXP	INV	BIN
REPORTED DATA RETAINED	3259 (69.2%) 3004974 (77.3%)	3248 (69.0%) 872715 (78.2%)	3198 (67.9%) 834337 (73.7%)
CHANGED FROM REPORTED	283 (6.0%) 191402 (4.9%)	257 (5.5%) 50869 (4.6%)	217 (4.6%) 54188 (4.8%)
IMPUTED FROM BLANK	1167 (24.8%) 689906 (17.8%)	1204 (25.6%) 191960 (17.2%)	1294 (27.5%) 242881 (21.5%)
TOTALS	4709 (100.0%) 3886282 (100.0%)	4709 (100.0%) 1115544 (100.0%)	4709 (100.0%) 1131406 (100.0%)

KEY: FIRST NUMBER REPRESENTS NUMBER OF RECORDS
SECOND NUMBER REPRESENTS TAB VALUE OF RECORDS

SIC 531110

	SLS	APR	QPR	EMP
REPORTED DATA RETAINED	2260 (92.2%) 28736661 (90.5%)	2131 (86.9%) 3686569 (85.7%)	2046 (83.4%) 857238 (84.2%)	2099 (85.6%) 445807 (84.4%)
CHANGED FROM REPORTED	119 (4.9%) 2319328 (7.3%)	247 (10.1%) 518451 (12.1%)	250 (10.2%) 122086 (12.0%)	260 (10.6%) 70505 (13.3%)
IMPUTED FROM BLANK	73 (3.0%) 708419 (2.2%)	74 (3.0%) 94437 (2.2%)	156 (6.4%) 38881 (3.8%)	93 (3.8%) 12006 (2.3%)
TOTALS	2452 (100.0%) 31764408 (100.0%)	2452 (100.0%) 4299457 (100.0%)	2452 (100.0%) 1018205 (100.0%)	2452 (100.0%) 528318 (100.0%)

KEY: FIRST NUMBER REPRESENTS NUMBER OF RECORDS
SECOND NUMBER REPRESENTS TAB VALUE OF RECORDS

SIC 596110

	SLS	APR	QPR	EMP
REPORTED DATA RETAINED	2660 (76.7%) 3826994 (89.3%)	2719 (78.4%) 443605 (94.9%)	2581 (74.4%) 107156 (94.9%)	2741 (79.1%) 33922 (92.5%)
CHANGED FROM REPORTED	131 (3.8%) 235197 (5.5%)	76 (2.2%) 3146 (.7%)	77 (2.2%) 748 (.7%)	38 (1.1%) 410 (1.0%)
IMPUTED FROM BLANK	676 (19.5%) 221515 (5.2%)	672 (19.4%) 20851 (4.5%)	809 (23.3%) 5001 (4.4%)	688 (19.8%) 2738 (6.5%)
TOTALS	3467 (100.0%) 4283706 (100.0%)	3467 (100.0%) 467602 (100.0%)	3467 (100.0%) 112905 (100.0%)	3467 (100.0%) 42070 (100.0%)

KEY: FIRST NUMBER REPRESENTS NUMBER OF RECORDS
SECOND NUMBER REPRESENTS TAB VALUE OF RECORDS

SIC 751210

	SLS	APR	QPR	EMP
REPORTED DATA RETAINED	2241 (64.4%) 2892711 (83.0%)	2208 (63.5%) 427363 (82.4%)	2129 (61.2%) 99034 (81.4%)	2209 (63.5%) 31602 (79.7%)
CHANGED FROM REPORTED	80 (2.3%) 101140 (2.9%)	101 (2.9%) 7245 (1.4%)	87 (2.5%) 1552 (1.3%)	83 (2.4%) 833 (2.1%)
IMPUTED FROM BLANK	1158 (33.3%) 489784 (14.1%)	1170 (33.6%) 84025 (16.2%)	1261 (36.3%) 21145 (17.4%)	1187 (34.1%) 7233 (18.2%)
TOTALS	3479 (100.0%) 3483635 (100.0%)	3479 (100.0%) 518633 (100.0%)	3477 (100.0%) 121731 (100.0%)	3479 (100.0%) 39668 (100.0%)

KEY: FIRST NUMBER REPRESENTS NUMBER OF RECORDS
SECOND NUMBER REPRESENTS TAB VALUE OF RECORDS

Table 5

SIC 783300

	SLS	APR	QPR	EMP
REPORTED DATA RETAINED	1761 (77.2%) 293745 (83.6%)	1761 (77.2%) 57639 (82.8%)	1551 (68.1%) 8603 (72.1%)	1609 (70.5%) 8530 (75.8%)
CHANGED FROM REPORTED	49 (2.1%) 4669 (1.3%)	68 (3.0%) 971 (1.4%)	160 (7.0%) 692 (5.8%)	134 (5.9%) 460 (4.1%)
IMPUTED FROM BLANK	472 (20.7%) 53082 (15.1%)	453 (19.9%) 11037 (15.8%)	568 (24.9%) 2636 (22.1%)	539 (23.6%) 2268 (20.1%)
TOTALS	2282 (100.0%) 351496 (100.0%)	2282 (100.0%) 69647 (100.0%)	2279 (100.0%) 11931 (100.0%)	2282 (100.0%) 11258 (100.0%)

KEY: FIRST NUMBER REPRESENTS NUMBER OF RECORDS
SECOND NUMBER REPRESENTS TAB VALUE OF RECORDS

Table 6

APPENDIX II

In this appendix we include Tables 7-12 which indicate the sources of imputed data values and the rate of use of each of the imputation options discussed in Section 2.3.3 of the text. These tables are interpreted and discussed in detail in Section 2.3.3, and they are presented here by field within KB.

SIC 509311

FIELD SLS

NUMBER OF FIELDS 2143
TAB VALUE OF FIELDS 3530658

SOURCE OF FINAL TAB VALUE:

	NUMBER OF CASES	AGGREGATE VALUE	% OF TAB	% OF IMPUTES
REPORTED DATA	1121(52.3%)	2039480(57.8%)	
1982 ADMIN. DATA	347(16.2%)	193827(5.5%)	(13.0%)
1981 ADMIN. DATA	5(.2%)	4665(.1%)	(.3%)
1977 CENSUS DATA	24(1.1%)	88319(2.5%)	(5.9%)
I&T PARAMETERS	532(24.8%)	787073(22.3%)	(52.8%)
77 DATA AVAIL.	208	497182		
77 DATA NOT AVAIL.	324	289891		
CLERK/ANALYST CHANGE	48(2.2%)	182661(5.2%)	(12.2%)
COMPLEX EDIT CHANGE	0(.0%)	0(.0%)	(.0%)
BLANK SET TO ZERO	0(.0%)	0(.0%)	(.0%)
CHANGE UNSPECIFIED	66(3.1%)	234633(6.6%)	(15.7%)
IMPUTE TOTALS =	1022(47.7%)	1491178(42.2%)	(100.0%)

FIELD APR

NUMBER OF FIELDS 2143
TAB VALUE OF FIELDS 398059

SOURCE OF FINAL TAB VALUE:

	NUMBER OF CASES	AGGREGATE VALUE	% OF TAB	% OF IMPUTES
REPORTED DATA	697(32.5%)	266657(67.0%)	
1982 ADMIN. DATA	1058(49.4%)	93185(23.4%)	(70.9%)
1981 ADMIN. DATA	90(4.2%)	2619(.7%)	(2.0%)
1977 CENSUS DATA	6(.3%)	143(.0%)	(.1%)
I&T PARAMETERS	192(9.0%)	9488(2.4%)	(7.2%)
77 DATA AVAIL.	50	4942		
77 DATA NOT AVAIL.	142	4546		
CLERK/ANALYST CHANGE	27(1.3%)	7508(1.9%)	(5.7%)
COMPLEX EDIT CHANGE	7(.3%)	7(.0%)	(.0%)
BLANK SET TO ZERO	0(.0%)	0(.0%)	(.0%)
CHANGE UNSPECIFIED	66(3.1%)	18452(4.6%)	(14.0%)
IMPUTE TOTALS =	1446(67.5%)	131402(33.0%)	(100.0%)

FIELD QPR

NUMBER OF FIELDS 2143
TAB VALUE OF FIELDS 105422

SOURCE OF FINAL TAB VALUE:

	NUMBER OF CASES	AGGREGATE VALUE	% OF TAB	% OF IMPUTES
REPORTED DATA	691(32.2%)	71514(67.8%)	
1982 ADMIN. DATA	1173(54.7%)	26551(25.2%)	(78.3%)
1981 ADMIN. DATA	4(.2%)	131(.1%)	(.4%)
1977 CENSUS DATA	0(.0%)	0(.0%)	(.0%)
I&T PARAMETERS	16(.7%)	284(.3%)	(.8%)
77 DATA AVAIL.	2	33		
77 DATA NOT AVAIL.	14	251		
CLERK/ANALYST CHANGE	21(1.0%)	2180(2.1%)	(6.4%)
COMPLEX EDIT CHANGE	0(.0%)	0(.0%)	(.0%)
BLANK SET TO ZERO	172(8.0%)	0(.0%)	(.0%)
CHANGE UNSPECIFIED	66(3.1%)	4762(4.5%)	(14.0%)
IMPUTE TOTALS =	1452(67.8%)	33908(32.2%)	(100.0%)

FIELD EMP

NUMBER OF FIELDS 2143
TAB VALUE OF FIELDS 26445

SOURCE OF FINAL TAB VALUE:

	NUMBER OF CASES	AGGREGATE VALUE	% OF TAB	% OF IMPUTES
REPORTED DATA	698(32.6%)	16379(61.9%)	
1982 ADMIN. DATA	1124(52.4%)	7860(29.7%)	(78.1%)
1981 ADMIN. DATA	14(.7%)	66(.2%)	(.7%)
1977 CENSUS DATA	2(.1%)	11(.0%)	(.1%)
I&T PARAMETERS	35(1.6%)	200(.8%)	(2.0%)
77 DATA AVAIL.	8	47		
77 DATA NOT AVAIL.	27	153		
CLERK/ANALYST CHANGE	20(.9%)	431(1.6%)	(4.3%)
COMPLEX EDIT CHANGE	0(.0%)	0(.0%)	(.0%)
BLANK SET TO ZERO	173(8.1%)	0(.0%)	(.0%)
CHANGE UNSPECIFIED	77(3.6%)	1498(5.7%)	(14.9%)
IMPUTE TOTALS =	1445(67.4%)	10066(38.1%)	(100.0%)

SIC 509311

FIELD EXP					FIELD INV				
NUMBER OF FIELDS		2143			NUMBER OF FIELDS		2143		
TAB VALUE OF FIELDS		960774			TAB VALUE OF FIELDS		438794		
SOURCE OF FINAL TAB VALUE:					SOURCE OF FINAL TAB VALUE:				
REPORTED DATA	NUMBER OF CASES	AGGREGATE VALUE	% OF TAB	% OF IMPUTES	REPORTED DATA	NUMBER OF CASES	AGGREGATE VALUE	% OF TAB	% OF IMPUTES
	899(42.0%)	566855(59.0%)			791(36.9%)	263953(60.2%)	
1982 ADMIN. DATA	0(.0%)	0(.0%)	(.0%)		1982 ADMIN. DATA	0(.0%)	0(.0%)	(.0%)	
1981 ADMIN. DATA	0(.0%)	0(.0%)	(.0%)		1981 ADMIN. DATA	0(.0%)	0(.0%)	(.0%)	
1977 CENSUS DATA	142(6.6%)	56348(5.9%)	(14.3%)	1977 CENSUS DATA	0(.0%)	0(.0%)	(.0%)	
I&T PARAMETERS	979(45.7%)	240727(25.1%)	(61.1%)	I&T PARAMETERS	819(38.2%)	143574(32.7%)	(82.1%)
77 DATA AVAIL.	312	118771			77 DATA AVAIL.	309	90429		
77 DATA NOT AVAIL.	667	121956			77 DATA NOT AVAIL.	510	53145		
CLERK/ANALYST CHANGE	63(2.9%)	53463(5.6%)	(13.6%)	CLERK/ANALYST CHANGE	14(.7%)	13643(3.1%)	(7.8%)
COMPLEX EDIT CHANGE	0(.0%)	0(.0%)	(.0%)		COMPLEX EDIT CHANGE	45(2.1%)	4707(1.1%)	(2.7%)
BLANK SET TO ZERO	0(.0%)	0(.0%)	(.0%)		BLANK SET TO ZERO	402(18.8%)	0(.0%)	(.0%)	
CHANGE UNSPECIFIED	60(2.8%)	43381(4.5%)	(11.0%)	CHANGE UNSPECIFIED	72(3.4%)	12917(2.9%)	(7.4%)
IMPUTE TOTALS =	1244(58.0%)	393919(41.0%)	(100.0%)	IMPUTE TOTALS =	1352(63.1%)	174841(39.8%)	(100.0%)

FIELD BIN				
NUMBER OF FIELDS		2143		
TAB VALUE OF FIELDS		478221		
SOURCE OF FINAL TAB VALUE:				
REPORTED DATA	NUMBER OF CASES	AGGREGATE VALUE	% OF TAB	% OF IMPUTES
	788(36.8%)	307153(64.2%)	
1982 ADMIN. DATA	0(.0%)	0(.0%)	(.0%)	
1981 ADMIN. DATA	0(.0%)	0(.0%)	(.0%)	
1977 CENSUS DATA	0(.0%)	0(.0%)	(.0%)	
I&T PARAMETERS	905(42.2%)	131089(27.4%)	(76.6%)
77 DATA AVAIL.	0	0		
77 DATA NOT AVAIL.	905	131089		
CLERK/ANALYST CHANGE	16(.7%)	23923(5.0%)	(14.0%)
COMPLEX EDIT CHANGE	49(2.3%)	1311(.3%)	(.8%)
BLANK SET TO ZERO	314(14.7%)	0(.0%)	(.0%)	
CHANGE UNSPECIFIED	71(3.3%)	14745(3.1%)	(8.6%)
IMPUTE TOTALS =	1355(63.2%)	171068(35.8%)	(100.0%)

Table 7 (continued)

SIC 518100

FIELD SLS				
NUMBER OF FIELDS	4709			
TAB VALUE OF FIELDS	20826564			
SOURCE OF FINAL TAB VALUE:				
	NUMBER OF CASES	AGGREGATE VALUE	% OF TAB	% OF IMPUTES
REPORTED DATA	3651(77.5%)	17447497(83.8%)	
1982 ADMIN. DATA	362(7.7%)	953306(4.6%)	(28.2%)
1981 ADMIN. DATA	3(.1%)	6210(.0%)	(.2%)
1977 CENSUS DATA	45(1.0%)	217642(1.0%)	(6.4%)
I&T PARAMETERS	551(11.7%)	1458710(7.0%)	(43.2%)
77 DATA AVAIL.	332	1062906		
77 DATA NOT AVAIL.	219	395804		
CLERK/ANALYST CHANGE	67(1.4%)	555791(2.7%)	(16.4%)
COMPLEX EDIT CHANGE	0(.0%)	0(.0%)	(.0%)
BLANK SET TO ZERO	0(.0%)	0(.0%)	(.0%)
CHANGE UNSPECIFIED	30(.6%)	187408(.9%)	(5.5%)
IMPUTE TOTALS =	1058(22.5%)	3379067(16.2%)	(100.0%)

FIELD APR				
NUMBER OF FIELDS	4709			
TAB VALUE OF FIELDS	1858162			
SOURCE OF FINAL TAB VALUE:				
	NUMBER OF CASES	AGGREGATE VALUE	% OF TAB	% OF IMPUTES
REPORTED DATA	2608(55.4%)	1504992(81.0%)	
1982 ADMIN. DATA	1640(34.8%)	283064(15.2%)	(80.1%)
1981 ADMIN. DATA	183(3.9%)	14260(.8%)	(4.0%)
1977 CENSUS DATA	6(.1%)	242(.0%)	(.1%)
I&T PARAMETERS	171(3.6%)	14615(.8%)	(4.1%)
77 DATA AVAIL.	63	5673		
77 DATA NOT AVAIL.	108	8942		
CLERK/ANALYST CHANGE	78(1.7%)	27242(1.5%)	(7.7%)
COMPLEX EDIT CHANGE	2(.0%)	2(.0%)	(.0%)
BLANK SET TO ZERO	0(.0%)	0(.0%)	(.0%)
CHANGE UNSPECIFIED	21(.4%)	13745(.7%)	(3.9%)
IMPUTE TOTALS =	2101(44.6%)	353170(19.0%)	(100.0%)

FIELD QPR				
NUMBER OF FIELDS	4709			
TAB VALUE OF FIELDS	433546			
SOURCE OF FINAL TAB VALUE:				
	NUMBER OF CASES	AGGREGATE VALUE	% OF TAB	% OF IMPUTES
REPORTED DATA	2547(54.1%)	349800(80.7%)	
1982 ADMIN. DATA	1910(40.6%)	73246(16.9%)	(87.5%)
1981 ADMIN. DATA	8(.2%)	537(.1%)	(.6%)
1977 CENSUS DATA	4(.1%)	426(.1%)	(.5%)
I&T PARAMETERS	22(.5%)	1585(.4%)	(1.9%)
77 DATA AVAIL.	6	259		
77 DATA NOT AVAIL.	16	1326		
CLERK/ANALYST CHANGE	62(1.3%)	4922(1.1%)	(5.9%)
COMPLEX EDIT CHANGE	0(.0%)	0(.0%)	(.0%)
BLANK SET TO ZERO	136(2.9%)	0(.0%)	(.0%)
CHANGE UNSPECIFIED	20(.4%)	3030(.7%)	(3.6%)
IMPUTE TOTALS =	2162(45.9%)	83746(19.3%)	(100.0%)

FIELD EMP				
NUMBER OF FIELDS	4709			
TAB VALUE OF FIELDS	87555			
SOURCE OF FINAL TAB VALUE:				
	NUMBER OF CASES	AGGREGATE VALUE	% OF TAB	% OF IMPUTES
REPORTED DATA	2610(55.4%)	68436(78.2%)	
1982 ADMIN. DATA	1828(38.8%)	16590(18.9%)	(86.8%)
1981 ADMIN. DATA	8(.2%)	119(.1%)	(.6%)
1977 CENSUS DATA	3(.1%)	25(.0%)	(.1%)
I&T PARAMETERS	21(.4%)	196(.2%)	(1.0%)
77 DATA AVAIL.	9	93		
77 DATA NOT AVAIL.	12	103		
CLERK/ANALYST CHANGE	57(1.2%)	1082(1.2%)	(5.7%)
COMPLEX EDIT CHANGE	0(.0%)	0(.0%)	(.0%)
BLANK SET TO ZERO	137(2.9%)	0(.0%)	(.0%)
CHANGE UNSPECIFIED	45(1.0%)	1107(1.3%)	(5.8%)
IMPUTE TOTALS =	2099(44.6%)	19119(21.8%)	(100.0%)

SIC 518100

FIELD EXP				
NUMBER OF FIELDS	4709			
TAB VALUE OF FIELDS	3886282			
SOURCE OF FINAL TAB VALUE:				
REPORTED DATA	NUMBER OF CASES	AGGREGATE VALUE	% OF TAB	% OF IMPUTES
	3259(69.2%)	3004974(77.3%)	
1982 ADMIN. DATA	0(.0%)	0(.0%)	(.0%)	
1981 ADMIN. DATA	0(.0%)	0(.0%)	(.0%)	
1977 CENSUS DATA	354(7.5%)	270814(7.0%)	(30.7%)	
I&T PARAMETERS	1032(21.9%)	533292(13.7%)	(60.5%)	
77 DATA AVAIL.	470	300954		
77 DATA NOT AVAIL.	562	232338		
CLERK/ANALYST CHANGE	64(1.4%)	77202(2.0%)	(8.8%)	
COMPLEX EDIT CHANGE	0(.0%)	0(.0%)	(.0%)	
BLANK SET TO ZERO	0(.0%)	0(.0%)	(.0%)	
CHANGE UNSPECIFIED	0(.0%)	0(.0%)	(.0%)	
IMPUTE TOTALS =	1450(30.8%)	881308(22.7%)	(100.0%)	

FIELD INV				
NUMBER OF FIELDS	4709			
TAB VALUE OF FIELDS	1115544			
SOURCE OF FINAL TAB VALUE:				
REPORTED DATA	NUMBER OF CASES	AGGREGATE VALUE	% OF TAB	% OF IMPUTES
	3248(69.0%)	872715(78.2%)		
1982 ADMIN. DATA	0(.0%)	0(.0%)	(.0%)	
1981 ADMIN. DATA	0(.0%)	0(.0%)	(.0%)	
1977 CENSUS DATA	0(.0%)	0(.0%)	(.0%)	
I&T PARAMETERS	950(20.2%)	187591(16.8%)	(77.3%)	
77 DATA AVAIL.	517	132158		
77 DATA NOT AVAIL.	433	55433		
CLERK/ANALYST CHANGE	80(1.7%)	46447(4.2%)	(19.1%)	
COMPLEX EDIT CHANGE	116(2.5%)	6928(.6%)	(2.9%)	
BLANK SET TO ZERO	291(6.2%)	0(.0%)	(.0%)	
CHANGE UNSPECIFIED	24(.5%)	1863(.2%)	(.8%)	
IMPUTE TOTALS =	1461(31.0%)	242829(21.8%)	(100.0%)	

FIELD BIN				
NUMBER OF FIELDS	4709			
TAB VALUE OF FIELDS	1131406			
SOURCE OF FINAL TAB VALUE:				
REPORTED DATA	NUMBER OF CASES	AGGREGATE VALUE	% OF TAB	% OF IMPUTES
	3198(67.9%)	834337(73.7%)		
1982 ADMIN. DATA	0(.0%)	0(.0%)	(.0%)	
1981 ADMIN. DATA	0(.0%)	0(.0%)	(.0%)	
1977 CENSUS DATA	0(.0%)	0(.0%)	(.0%)	
I&T PARAMETERS	1041(22.1%)	236545(20.9%)	(79.6%)	
77 DATA AVAIL.	0	0		
77 DATA NOT AVAIL.	1041	236545		
CLERK/ANALYST CHANGE	108(2.3%)	52464(4.6%)	(17.7%)	
COMPLEX EDIT CHANGE	87(1.8%)	4651(.4%)	(1.6%)	
BLANK SET TO ZERO	245(5.2%)	0(.0%)	(.0%)	
CHANGE UNSPECIFIED	30(.6%)	3409(.3%)	(1.1%)	
IMPUTE TOTALS =	1511(32.1%)	297069(26.3%)	(100.0%)	

SIC 531110

FIELD SLS

NUMBER OF FIELDS 2452
TAB VALUE OF FIELDS 31764408

SOURCE OF FINAL TAB VALUE:

	NUMBER OF CASES	AGGREGATE VALUE	% OF TAB	% OF IMPUTES
REPORTED DATA	2260 (92.2%)	28736661	90.5%	
1982 ADMIN. DATA	12 (.5%)	75507	.2%	(2.5%)
1981 ADMIN. DATA	0 (.0%)	0	.0%	(.0%)
1977 CENSUS DATA	2 (.1%)	8179	.0%	(.3%)
I&T PARAMETERS	61 (2.5%)	561460	1.8%	(18.5%)
77 DATA AVAIL.	36	392531		
77 DATA NOT AVAIL.	25	168929		
CLERK/ANALYST CHANGE	16 (.7%)	240381	.8%	(7.9%)
COMPLEX EDIT CHANGE	0 (.0%)	0	.0%	(.0%)
BLANK SET TO ZERO	0 (.0%)	0	.0%	(.0%)
CHANGE UNSPECIFIED	101 (4.1%)	2142220	6.7%	(70.8%)
IMPUTE TOTALS =	192 (7.8%)	3027747	9.5%	(100.0%)

FIELD APR

NUMBER OF FIELDS 2452
TAB VALUE OF FIELDS 4299457

SOURCE OF FINAL TAB VALUE:

	NUMBER OF CASES	AGGREGATE VALUE	% OF TAB	% OF IMPUTES
REPORTED DATA	2131 (86.9%)	3686569	85.7%	
1982 ADMIN. DATA	74 (3.0%)	89026	2.1%	(14.5%)
1981 ADMIN. DATA	1 (.0%)	546	.0%	(.1%)
1977 CENSUS DATA	0 (.0%)	0	.0%	(.0%)
I&T PARAMETERS	5 (.2%)	2500	.1%	(.4%)
77 DATA AVAIL.	4	940		
77 DATA NOT AVAIL.	1	1560		
CLERK/ANALYST CHANGE	21 (.9%)	58623	1.4%	(9.6%)
COMPLEX EDIT CHANGE	0 (.0%)	0	.0%	(.0%)
BLANK SET TO ZERO	0 (.0%)	0	.0%	(.0%)
CHANGE UNSPECIFIED	220 (9.0%)	462193	10.8%	(75.4%)
IMPUTE TOTALS =	321 (13.1%)	612888	14.3%	(100.0%)

FIELD QPR

NUMBER OF FIELDS 2452
TAB VALUE OF FIELDS 1018205

SOURCE OF FINAL TAB VALUE:

	NUMBER OF CASES	AGGREGATE VALUE	% OF TAB	% OF IMPUTES
REPORTED DATA	2046 (83.4%)	857238	84.2%	
1982 ADMIN. DATA	148 (6.0%)	42116	4.1%	(26.2%)
1981 ADMIN. DATA	0 (.0%)	0	.0%	(.0%)
1977 CENSUS DATA	4 (.2%)	205	.0%	(.1%)
I&T PARAMETERS	1 (.0%)	211	.0%	(.1%)
77 DATA AVAIL.	0	0		
77 DATA NOT AVAIL.	1	211		
CLERK/ANALYST CHANGE	22 (.9%)	14139	1.4%	(8.8%)
COMPLEX EDIT CHANGE	0 (.0%)	0	.0%	(.0%)
BLANK SET TO ZERO	16 (.7%)	0	.0%	(.0%)
CHANGE UNSPECIFIED	215 (8.8%)	104296	10.2%	(64.8%)
IMPUTE TOTALS =	406 (16.6%)	160967	15.8%	(100.0%)

FIELD EMP

NUMBER OF FIELDS 2452
TAB VALUE OF FIELDS 528318

SOURCE OF FINAL TAB VALUE:

	NUMBER OF CASES	AGGREGATE VALUE	% OF TAB	% OF IMPUTES
REPORTED DATA	2099 (85.6%)	445807	84.4%	
1982 ADMIN. DATA	77 (3.1%)	11815	2.2%	(14.3%)
1981 ADMIN. DATA	1 (.0%)	90	.0%	(.1%)
1977 CENSUS DATA	4 (.2%)	982	.2%	(1.2%)
I&T PARAMETERS	17 (.7%)	2298	.4%	(2.8%)
77 DATA AVAIL.	10	500		
77 DATA NOT AVAIL.	7	1798		
CLERK/ANALYST CHANGE	22 (.9%)	6176	1.2%	(7.5%)
COMPLEX EDIT CHANGE	0 (.0%)	0	.0%	(.0%)
BLANK SET TO ZERO	16 (.7%)	0	.0%	(.0%)
CHANGE UNSPECIFIED	216 (8.8%)	61150	11.6%	(74.1%)
IMPUTE TOTALS =	353 (14.4%)	82511	15.6%	(100.0%)

0BRKPT PRINT\$

SIC 596110

FIELD SLS				
NUMBER OF FIELDS	3467			
TAB VALUE OF FIELDS	4283706			
SOURCE OF FINAL TAB VALUE:				
	NUMBER OF CASES	AGGREGATE VALUE	% OF TAB	% OF IMPUTES
REPORTED DATA	2660(76.7%)	3826994(89.3%)	
1982 ADMIN. DATA	234(6.7%)	42787(1.0%)	(9.4%)
1981 ADMIN. DATA	5(.1%)	309(.0%)	(.1%)
1977 CENSUS DATA	23(.7%)	11958(.3%)	(2.6%)
I&T PARAMETERS	464(13.4%)	167067(3.9%)	(36.6%)
77 DATA AVAIL.	158	66058		
77 DATA NOT AVAIL.	306	101009		
CLERK/ANALYST CHANGE	29(.8%)	150223(3.5%)	(32.9%)
COMPLEX EDIT CHANGE	0(.0%)	0(.0%)	(.0%)
BLANK SET TO ZERO	0(.0%)	0(.0%)	(.0%)
CHANGE UNSPECIFIED	52(1.5%)	84368(2.0%)	(18.5%)
IMPUTE TOTALS =	807(23.3%)	456712(10.7%)	(100.0%)

FIELD APR				
NUMBER OF FIELDS	3467			
TAB VALUE OF FIELDS	467602			
SOURCE OF FINAL TAB VALUE:				
	NUMBER OF CASES	AGGREGATE VALUE	% OF TAB	% OF IMPUTES
REPORTED DATA	2719(78.4%)	443605(94.9%)	
1982 ADMIN. DATA	522(15.1%)	17928(3.8%)	(74.7%)
1981 ADMIN. DATA	27(.8%)	376(.1%)	(1.6%)
1977 CENSUS DATA	6(.2%)	37(.0%)	(.2%)
I&T PARAMETERS	137(4.0%)	2385(.5%)	(9.9%)
77 DATA AVAIL.	49	875		
77 DATA NOT AVAIL.	88	1510		
CLERK/ANALYST CHANGE	33(1.0%)	1926(.4%)	(8.0%)
COMPLEX EDIT CHANGE	12(.3%)	19(.0%)	(.1%)
BLANK SET TO ZERO	8(.2%)	0(.0%)	(.0%)
CHANGE UNSPECIFIED	3(.1%)	1326(.3%)	(5.5%)
IMPUTE TOTALS =	748(21.6%)	23997(5.1%)	(100.0%)

FIELD QPR				
NUMBER OF FIELDS	3467			
TAB VALUE OF FIELDS	112905			
SOURCE OF FINAL TAB VALUE:				
	NUMBER OF CASES	AGGREGATE VALUE	% OF TAB	% OF IMPUTES
REPORTED DATA	2581(74.4%)	107156(94.9%)	
1982 ADMIN. DATA	650(18.7%)	4901(4.3%)	(85.2%)
1981 ADMIN. DATA	9(.3%)	30(.0%)	(.5%)
1977 CENSUS DATA	2(.1%)	5(.0%)	(.1%)
I&T PARAMETERS	25(.7%)	183(.2%)	(3.2%)
77 DATA AVAIL.	2	4		
77 DATA NOT AVAIL.	23	179		
CLERK/ANALYST CHANGE	28(.8%)	299(.3%)	(5.2%)
COMPLEX EDIT CHANGE	7(.2%)	7(.0%)	(.1%)
BLANK SET TO ZERO	163(4.7%)	0(.0%)	(.0%)
CHANGE UNSPECIFIED	2(.1%)	324(.3%)	(5.6%)
IMPUTE TOTALS =	886(25.6%)	5749(5.1%)	(100.0%)

FIELD EMP				
NUMBER OF FIELDS	3467			
TAB VALUE OF FIELDS	42070			
SOURCE OF FINAL TAB VALUE:				
	NUMBER OF CASES	AGGREGATE VALUE	% OF TAB	% OF IMPUTES
REPORTED DATA	2741(79.1%)	38922(92.5%)	
1982 ADMIN. DATA	520(15.0%)	2601(6.2%)	(82.6%)
1981 ADMIN. DATA	1(.0%)	2(.0%)	(.1%)
1977 CENSUS DATA	2(.1%)	45(.1%)	(1.4%)
I&T PARAMETERS	17(.5%)	209(.5%)	(6.6%)
77 DATA AVAIL.	3	61		
77 DATA NOT AVAIL.	14	148		
CLERK/ANALYST CHANGE	21(.6%)	136(.3%)	(4.3%)
COMPLEX EDIT CHANGE	0(.0%)	0(.0%)	(.0%)
BLANK SET TO ZERO	162(4.7%)	0(.0%)	(.0%)
CHANGE UNSPECIFIED	3(.1%)	155(.4%)	(4.9%)
IMPUTE TOTALS =	726(20.9%)	3148(7.5%)	(100.0%)

SIC 751210

FIELD SLS

NUMBER OF FIELDS 3479
TAB VALUE OF FIELDS 3483635

SOURCE OF FINAL TAB VALUE:

	NUMBER OF CASES	AGGREGATE VALUE	% OF TAB	% OF IMPUTES
REPORTED DATA	2241(64.4%)	2892711(83.0%)	
1982 ADMIN. DATA	291(8.4%)	95072(2.7%)	(16.1%)	
1981 ADMIN. DATA	2(.1%)	74(.0%)	(.0%)	
1977 CENSUS DATA	31(.9%)	48592(1.4%)	(8.2%)	
I&T PARAMETERS	884(25.4%)	401847(11.5%)	(68.0%)	
77 DATA AVAIL.	158	104624		
77 DATA NOT AVAIL.	726	297223		
CLERK/ANALYST CHANGE	20(.6%)	42248(1.2%)	(7.1%)	
COMPLEX EDIT CHANGE	0(.0%)	0(.0%)	(.0%)	
BLANK SET TO ZERO	0(.0%)	0(.0%)	(.0%)	
CHANGE UNSPECIFIED	10(.3%)	3091(.1%)	(.5%)	
IMPUTE TOTALS =	1238(35.6%)	590924(17.0%)	(100.0%)	

FIELD APR

NUMBER OF FIELDS 3479
TAB VALUE OF FIELDS 518633

SOURCE OF FINAL TAB VALUE:

	NUMBER OF CASES	AGGREGATE VALUE	% OF TAB	% OF IMPUTES
REPORTED DATA	2208(63.5%)	427363(82.4%)		
1982 ADMIN. DATA	1030(29.6%)	78887(15.2%)	(86.4%)	
1981 ADMIN. DATA	14(.4%)	441(.1%)	(.5%)	
1977 CENSUS DATA	2(.1%)	23(.0%)	(.0%)	
I&T PARAMETERS	133(3.8%)	5387(1.0%)	(5.9%)	
77 DATA AVAIL.	32	1993		
77 DATA NOT AVAIL.	101	3394		
CLERK/ANALYST CHANGE	28(.8%)	6468(1.2%)	(7.1%)	
COMPLEX EDIT CHANGE	64(1.8%)	64(.0%)	(.1%)	
BLANK SET TO ZERO	0(.0%)	0(.0%)	(.0%)	
CHANGE UNSPECIFIED	0(.0%)	0(.0%)	(.0%)	
IMPUTE TOTALS =	1271(36.5%)	91270(17.6%)	(100.0%)	

FIELD QPR

NUMBER OF FIELDS 3477
TAB VALUE OF FIELDS 121731

SOURCE OF FINAL TAB VALUE:

	NUMBER OF CASES	AGGREGATE VALUE	% OF TAB	% OF IMPUTES
REPORTED DATA	2129(61.2%)	99034(81.4%)		
1982 ADMIN. DATA	1084(31.2%)	21003(17.3%)	(92.5%)	
1981 ADMIN. DATA	2(.1%)	77(.1%)	(.3%)	
1977 CENSUS DATA	3(.1%)	14(.0%)	(.1%)	
I&T PARAMETERS	53(1.5%)	359(.3%)	(1.6%)	
77 DATA AVAIL.	6	35		
77 DATA NOT AVAIL.	47	324		
CLERK/ANALYST CHANGE	21(.6%)	1241(1.0%)	(5.5%)	
COMPLEX EDIT CHANGE	3(.1%)	3(.0%)	(.0%)	
BLANK SET TO ZERO	182(5.2%)	0(.0%)	(.0%)	
CHANGE UNSPECIFIED	0(.0%)	0(.0%)	(.0%)	
IMPUTE TOTALS =	1348(38.8%)	22697(18.6%)	(100.0%)	

FIELD EMP

NUMBER OF FIELDS 3479
TAB VALUE OF FIELDS 39668

SOURCE OF FINAL TAB VALUE:

	NUMBER OF CASES	AGGREGATE VALUE	% OF TAB	% OF IMPUTES
REPORTED DATA	2209(63.5%)	31602(79.7%)		
1982 ADMIN. DATA	1024(29.4%)	7366(18.6%)	(91.3%)	
1981 ADMIN. DATA	5(.1%)	12(.0%)	(.1%)	
1977 CENSUS DATA	10(.3%)	118(.3%)	(1.5%)	
I&T PARAMETERS	34(1.0%)	252(.6%)	(3.1%)	
77 DATA AVAIL.	6	54		
77 DATA NOT AVAIL.	29	200		
CLERK/ANALYST CHANGE	20(.6%)	318(.8%)	(3.9%)	
COMPLEX EDIT CHANGE	0(.0%)	0(.0%)	(.0%)	
BLANK SET TO ZERO	177(5.1%)	0(.0%)	(.0%)	
CHANGE UNSPECIFIED	0(.0%)	0(.0%)	(.0%)	
IMPUTE TOTALS =	1270(36.5%)	8066(20.3%)	(100.0%)	

DBRKPT PRINT\$

SIC 783300

FIELD SLS

NUMBER OF FIELDS 2282
TAB VALUE OF FIELDS 351496

SOURCE OF FINAL TAB VALUE:

	NUMBER OF CASES	AGGREGATE VALUE	% OF TAB	% OF IMPUTES
REPORTED DATA	1761(77.2%)	293745(83.6%)		
1982 ADMIN. DATA	134(5.9%)	16303(4.6%)	(28.2%)	
1981 ADMIN. DATA	0(.0%)	0(.0%)	(.0%)	
1977 CENSUS DATA	23(1.0%)	2090(.6%)	(3.6%)	
I&T PARAMETERS	340(14.9%)	33260(9.5%)	(57.6%)	
77 DATA AVAIL.	125	12070		
77 DATA NOT AVAIL.	215	21190		
CLERK/ANALYST CHANGE	22(1.0%)	5813(1.7%)	(10.1%)	
COMPLEX EDIT CHANGE	0(.0%)	0(.0%)	(.0%)	
BLANK SET TO ZERO	0(.0%)	0(.0%)	(.0%)	
CHANGE UNSPECIFIED	2(.1%)	285(.1%)	(.5%)	
IMPUTE TOTALS =	521(22.8%)	57751(16.4%)	(100.0%)	

FIELD APR

NUMBER OF FIELDS 2282
TAB VALUE OF FIELDS 69647

SOURCE OF FINAL TAB VALUE:

	NUMBER OF CASES	AGGREGATE VALUE	% OF TAB	% OF IMPUTES
REPORTED DATA	1761(77.2%)	57639(82.8%)		
1982 ADMIN. DATA	404(17.7%)	9377(13.5%)	(78.1%)	
1981 ADMIN. DATA	8(.4%)	92(.1%)	(.8%)	
1977 CENSUS DATA	9(.4%)	358(.5%)	(3.0%)	
I&T PARAMETERS	65(2.8%)	726(1.0%)	(6.0%)	
77 DATA AVAIL.	27	298		
77 DATA NOT AVAIL.	38	428		
CLERK/ANALYST CHANGE	24(1.1%)	1444(2.1%)	(12.0%)	
COMPLEX EDIT CHANGE	11(.5%)	11(.0%)	(.1%)	
BLANK SET TO ZERO	0(.0%)	0(.0%)	(.0%)	
CHANGE UNSPECIFIED	0(.0%)	0(.0%)	(.0%)	
IMPUTE TOTALS =	521(22.8%)	12008(17.2%)	(100.0%)	

FIELD QPR

NUMBER OF FIELDS 2279
TAB VALUE OF FIELDS 11931

SOURCE OF FINAL TAB VALUE:

	NUMBER OF CASES	AGGREGATE VALUE	% OF TAB	% OF IMPUTES
REPORTED DATA	1551(68.1%)	8603(72.1%)		
1982 ADMIN. DATA	475(20.8%)	2667(22.4%)	(80.1%)	
1981 ADMIN. DATA	6(.3%)	46(.4%)	(1.4%)	
1977 CENSUS DATA	17(.7%)	65(.5%)	(2.0%)	
I&T PARAMETERS	66(2.9%)	310(2.6%)	(9.3%)	
77 DATA AVAIL.	6	34		
77 DATA NOT AVAIL.	60	276		
CLERK/ANALYST CHANGE	28(1.2%)	236(2.0%)	(7.1%)	
COMPLEX EDIT CHANGE	4(.2%)	4(.0%)	(.1%)	
BLANK SET TO ZERO	132(5.8%)	0(.0%)	(.0%)	
CHANGE UNSPECIFIED	0(.0%)	0(.0%)	(.0%)	
IMPUTE TOTALS =	728(31.9%)	3328(27.9%)	(100.0%)	

FIELD EMP

NUMBER OF FIELDS 2282
TAB VALUE OF FIELDS 11258

SOURCE OF FINAL TAB VALUE:

	NUMBER OF CASES	AGGREGATE VALUE	% OF TAB	% OF IMPUTES
REPORTED DATA	1609(70.5%)	8530(75.8%)		
1982 ADMIN. DATA	428(18.8%)	2155(19.1%)	(79.0%)	
1981 ADMIN. DATA	1(.0%)	1(.0%)	(.0%)	
1977 CENSUS DATA	26(1.1%)	156(1.4%)	(5.7%)	
I&T PARAMETERS	45(2.0%)	239(2.1%)	(8.8%)	
77 DATA AVAIL.	13	103		
77 DATA NOT AVAIL.	32	136		
CLERK/ANALYST CHANGE	32(1.4%)	177(1.6%)	(6.5%)	
COMPLEX EDIT CHANGE	0(.0%)	0(.0%)	(.0%)	
BLANK SET TO ZERO	141(6.2%)	0(.0%)	(.0%)	
CHANGE UNSPECIFIED	0(.0%)	0(.0%)	(.0%)	
IMPUTE TOTALS =	673(29.5%)	2728(24.2%)	(100.0%)	

DBRKPT PRINT#

APPENDIX III

In this appendix we include Figures 1-6 which exhibit the effects of changes to reported data. The horizontal axis of each graph is the cumulative percentage of cases and the vertical axis is the cumulative percentage of total change. These graphs show, in essence, that a small percentage of cases account for a large percentage of total change. These graphs and attendant phenomenon are explained and discussed in Section 2.3.4 of the text, and they are presented here by field within KB.

WHOLESALE SIC 509311

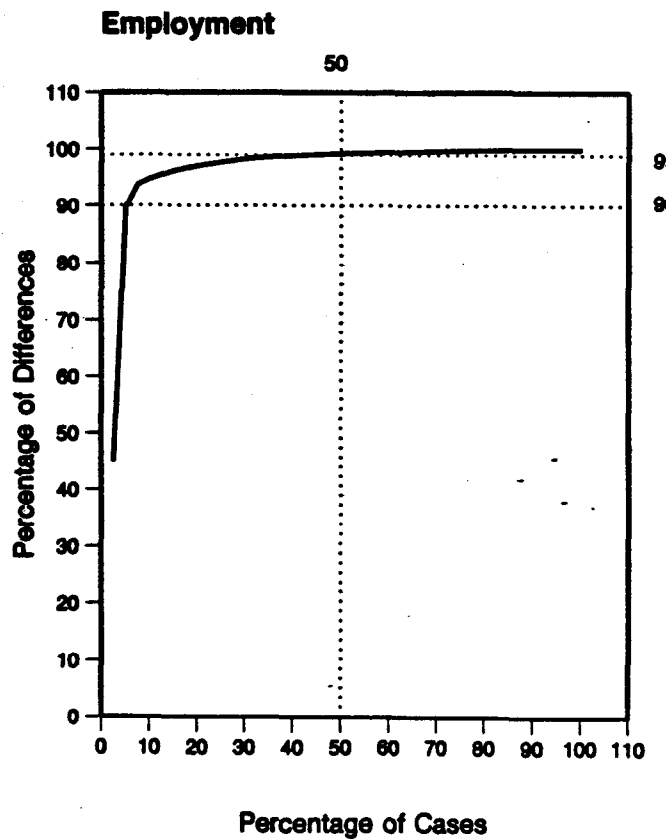
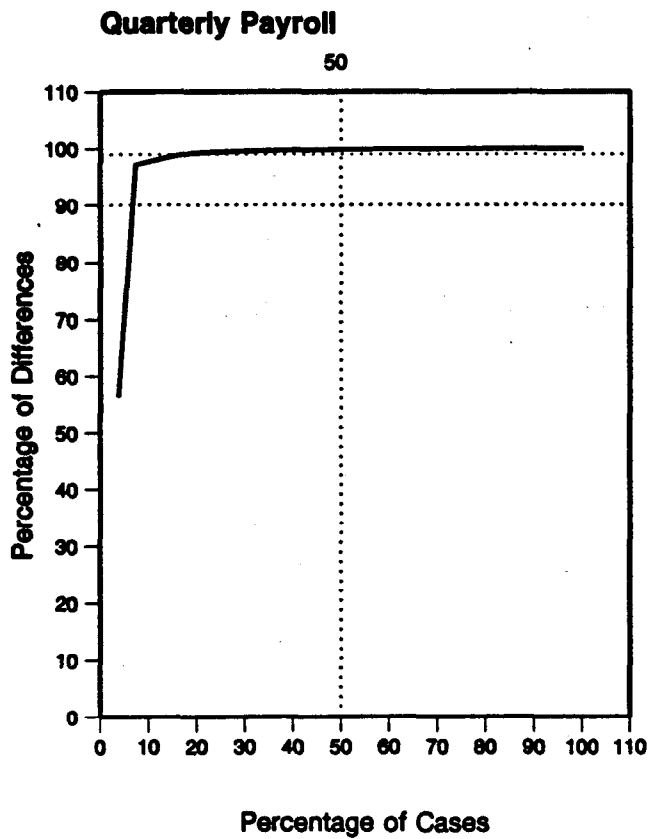
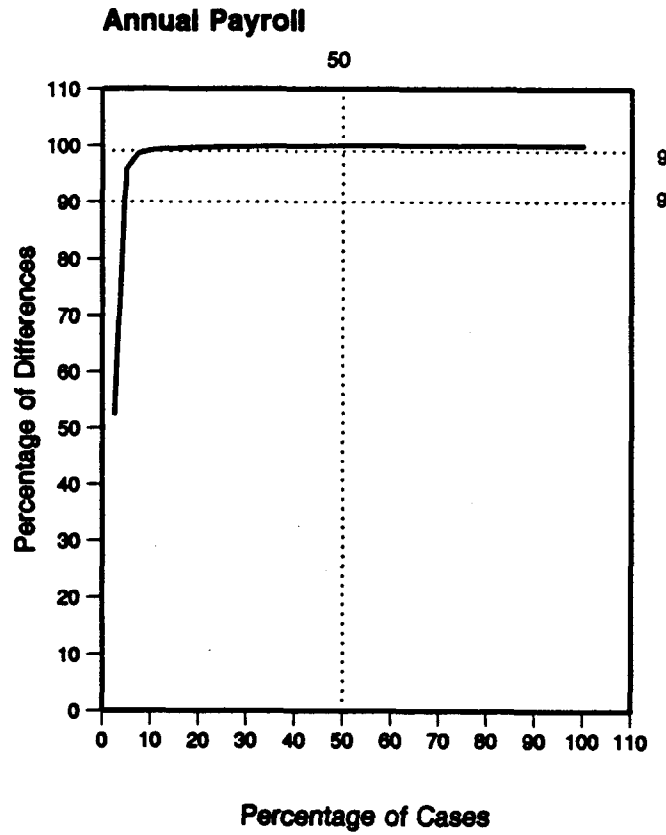
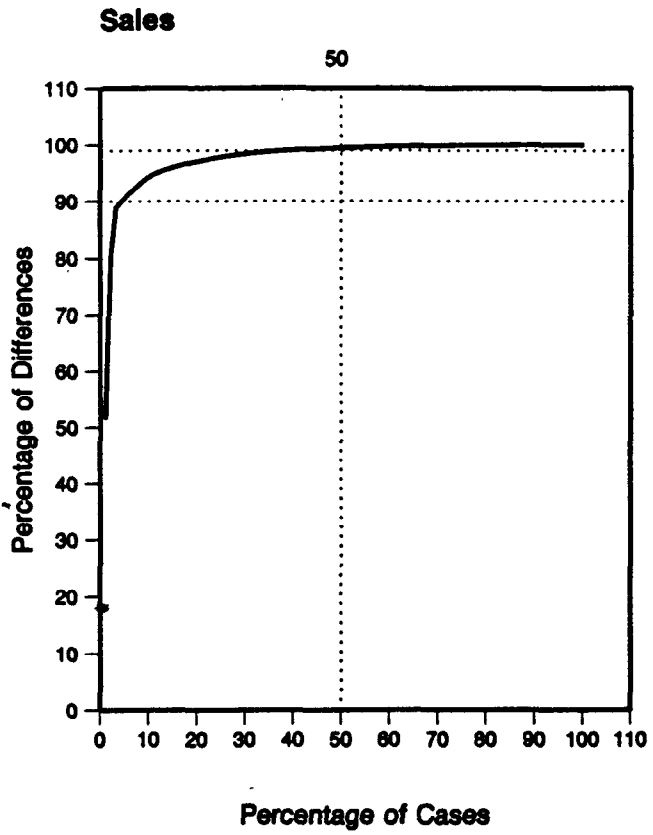


Figure 1

WHOLESALE SIC 509311

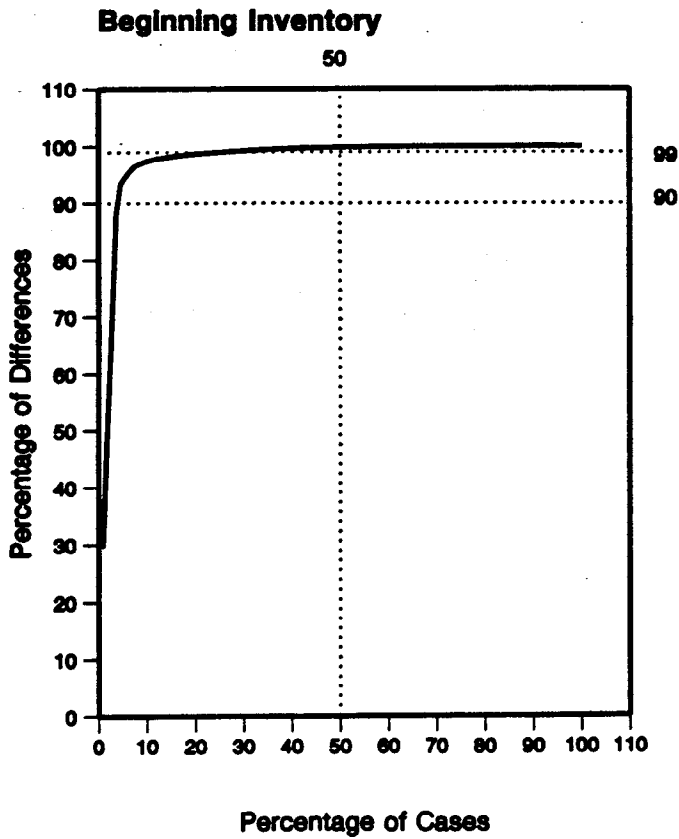
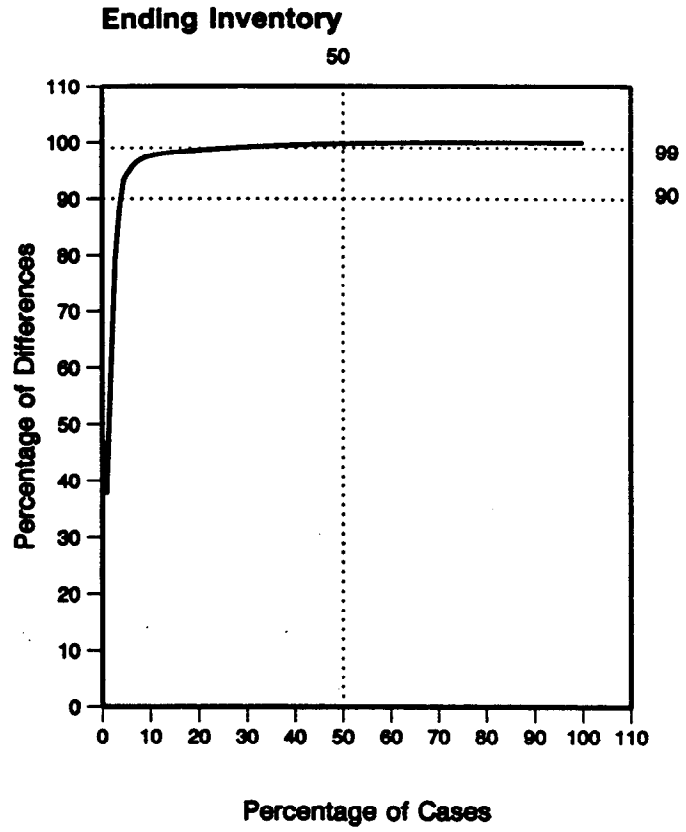
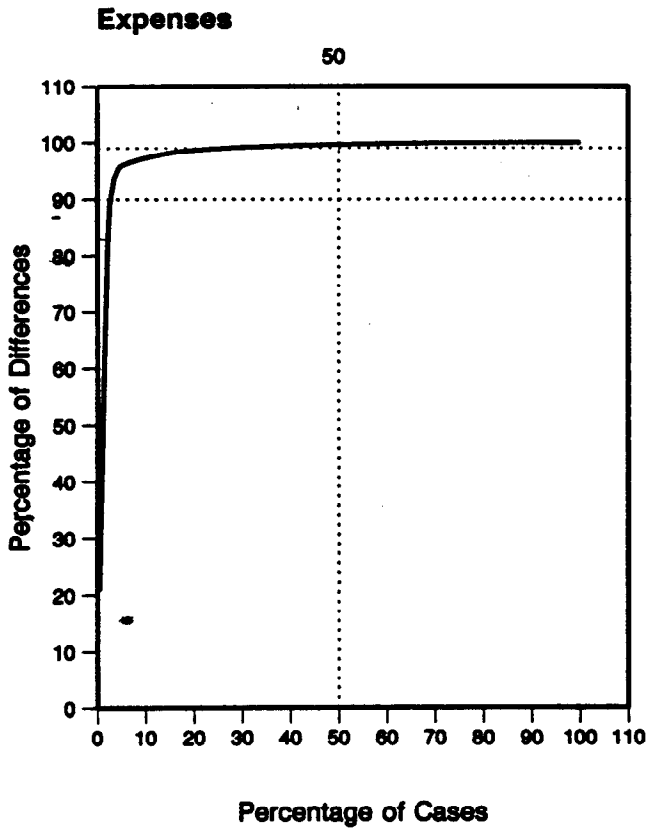


Figure 1 (continued)

WHOLESALE SIC 518100

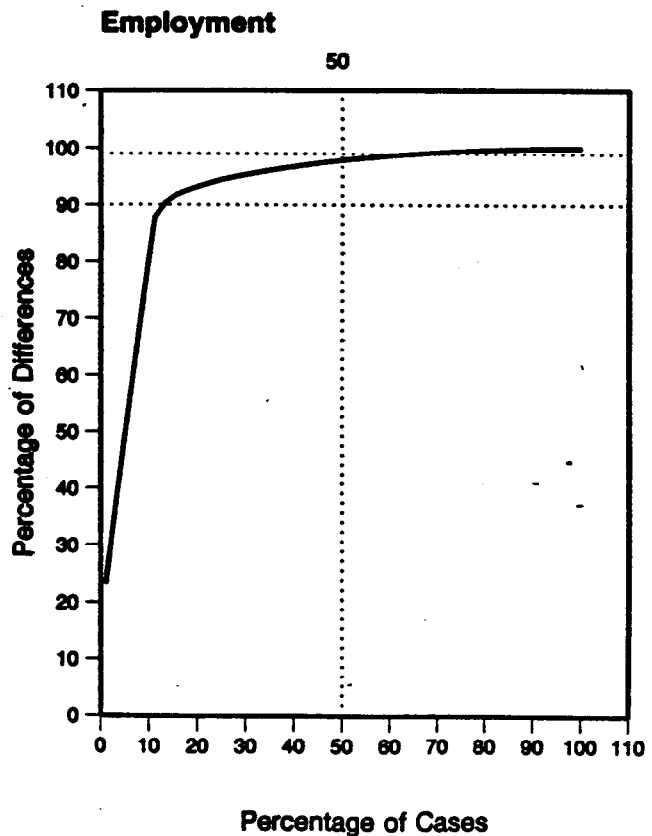
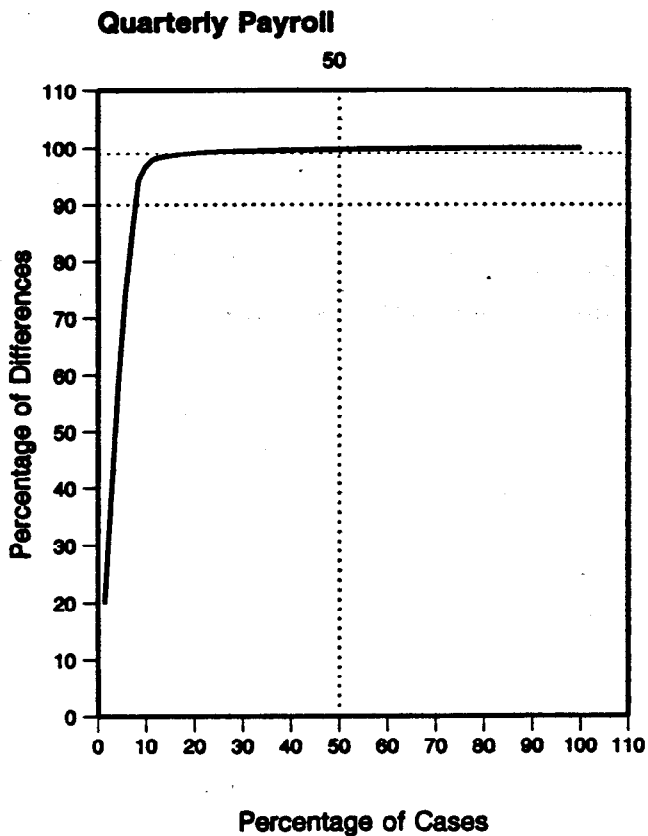
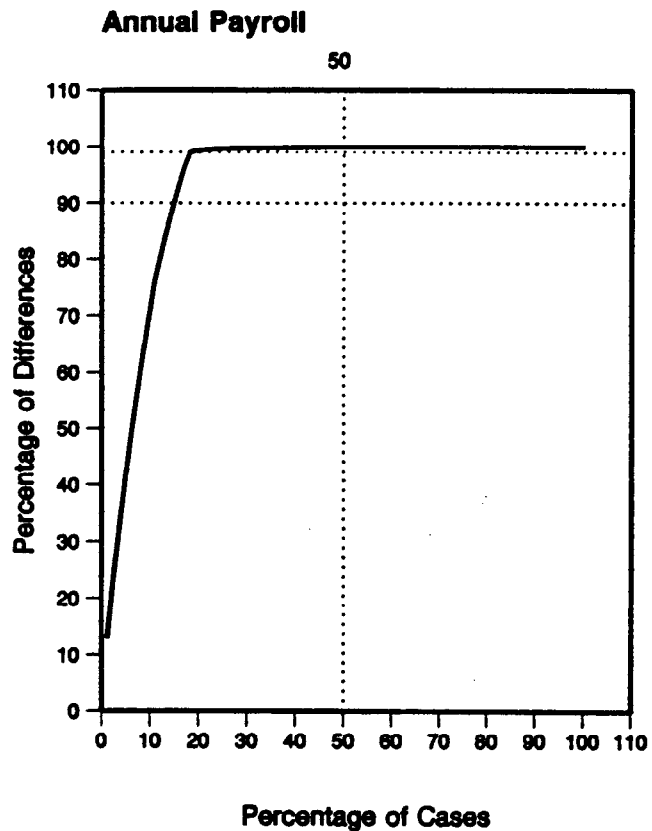
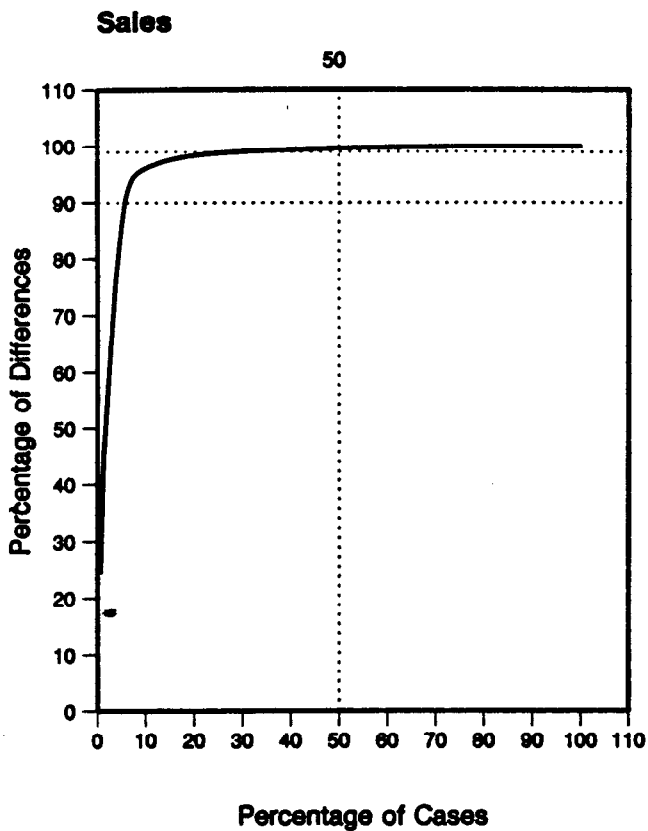


Figure 2

WHOLESALE SIC 518100

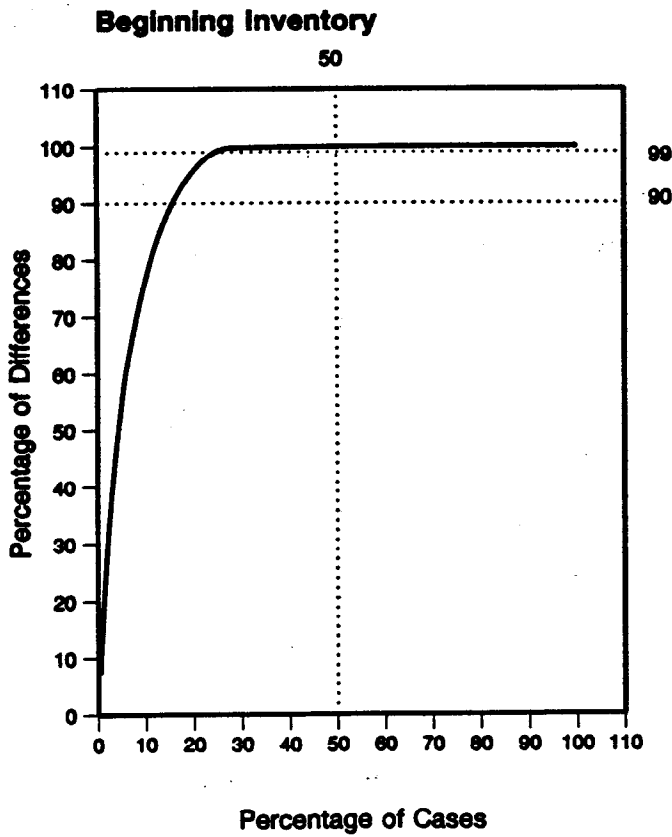
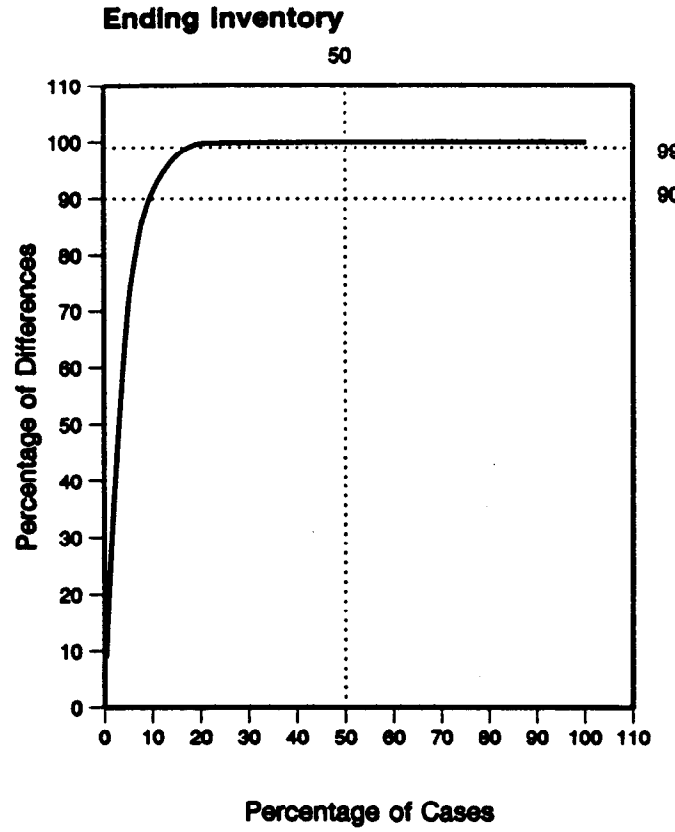
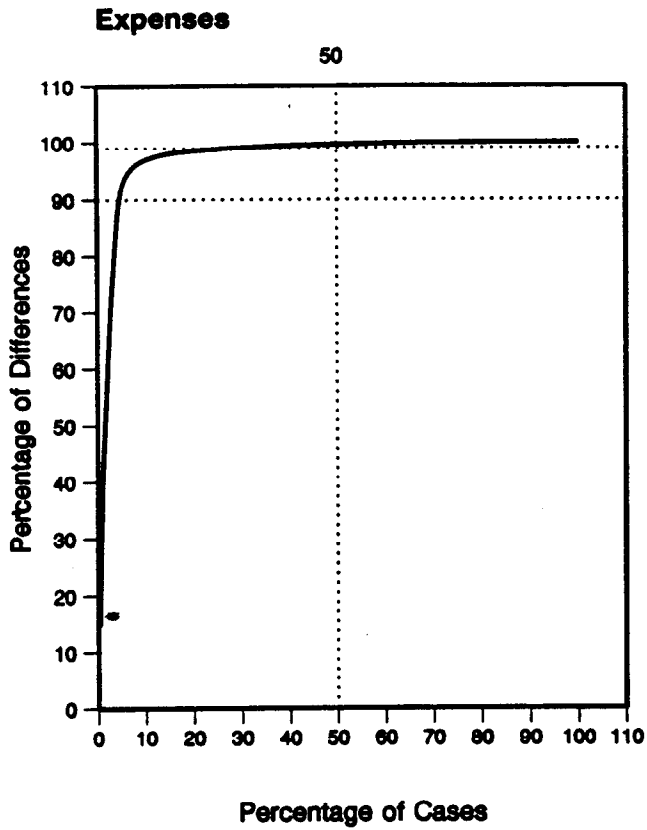


Figure 2 (continued)

RETAIL SIC 531110

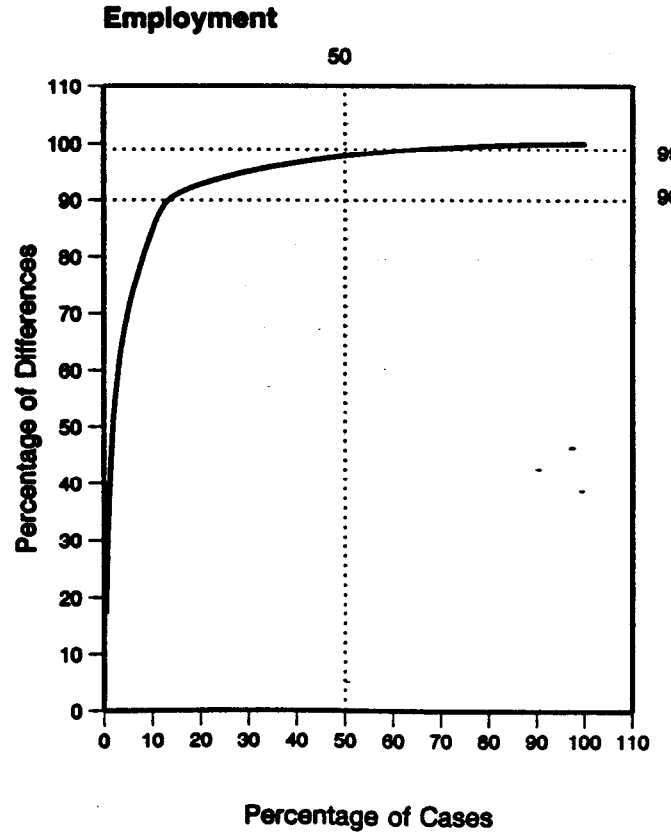
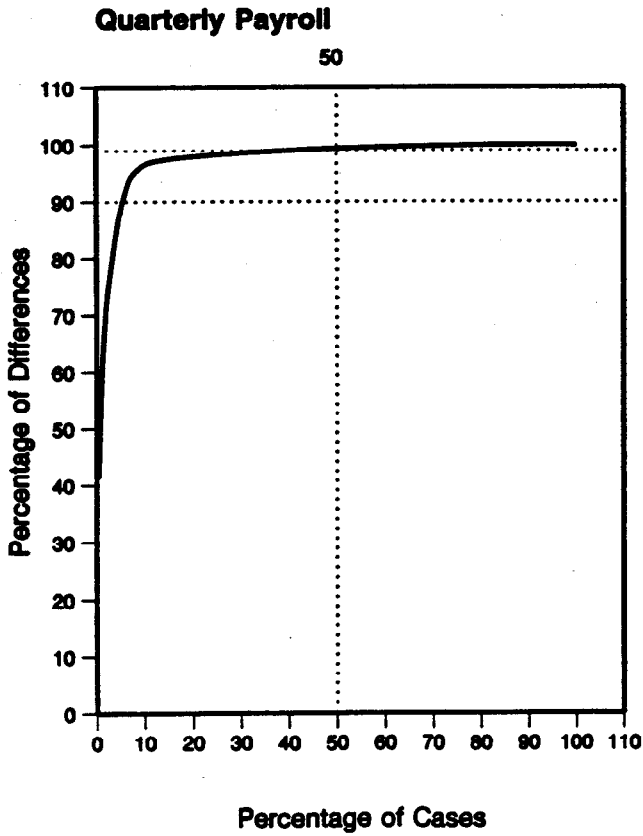
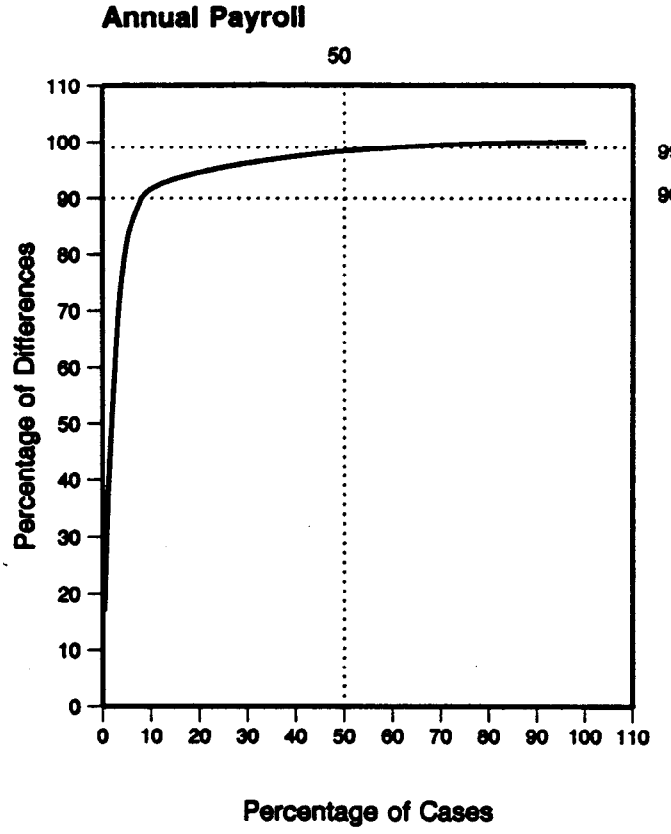
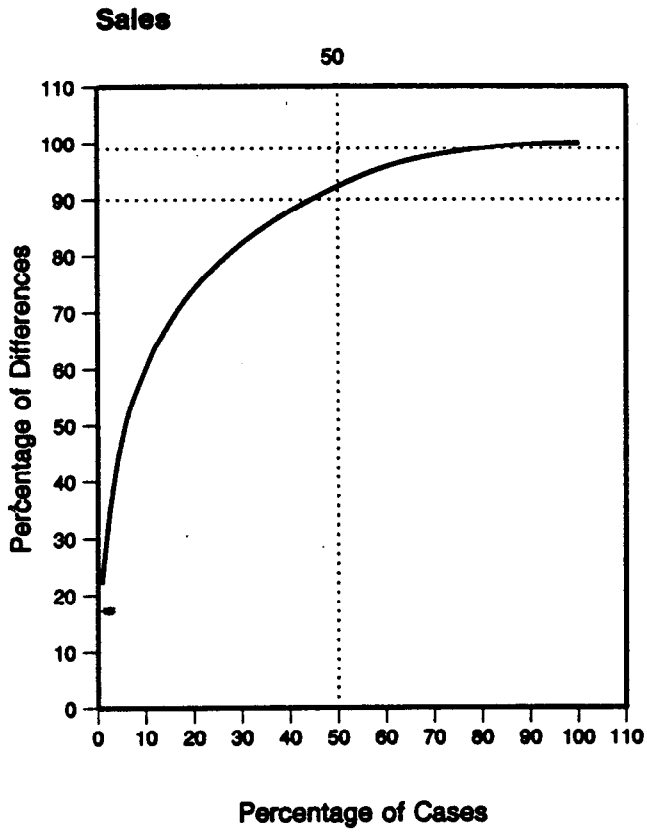


Figure 3

RETAIL SIC 596110

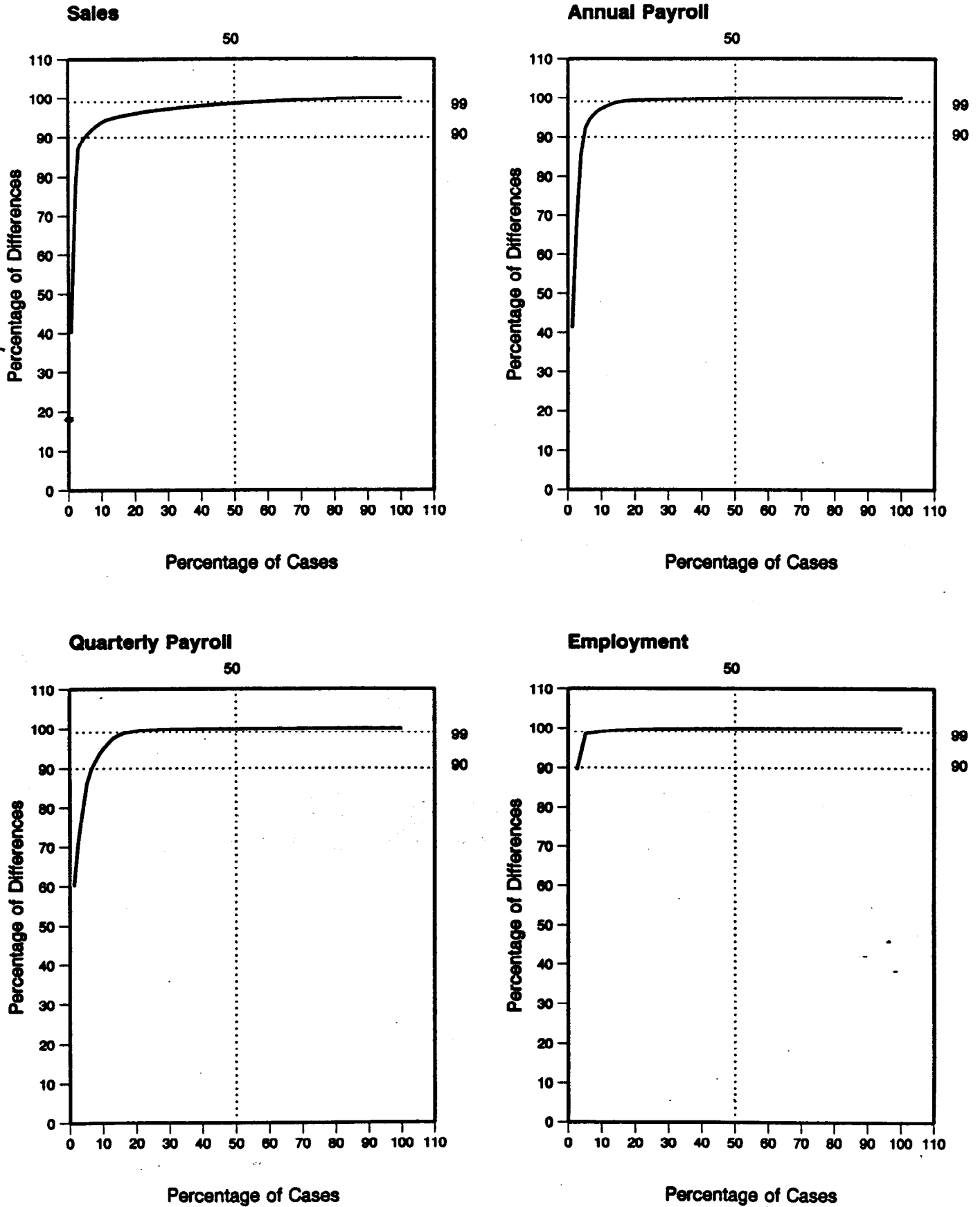


Figure 4

SERVICE SIC 751210

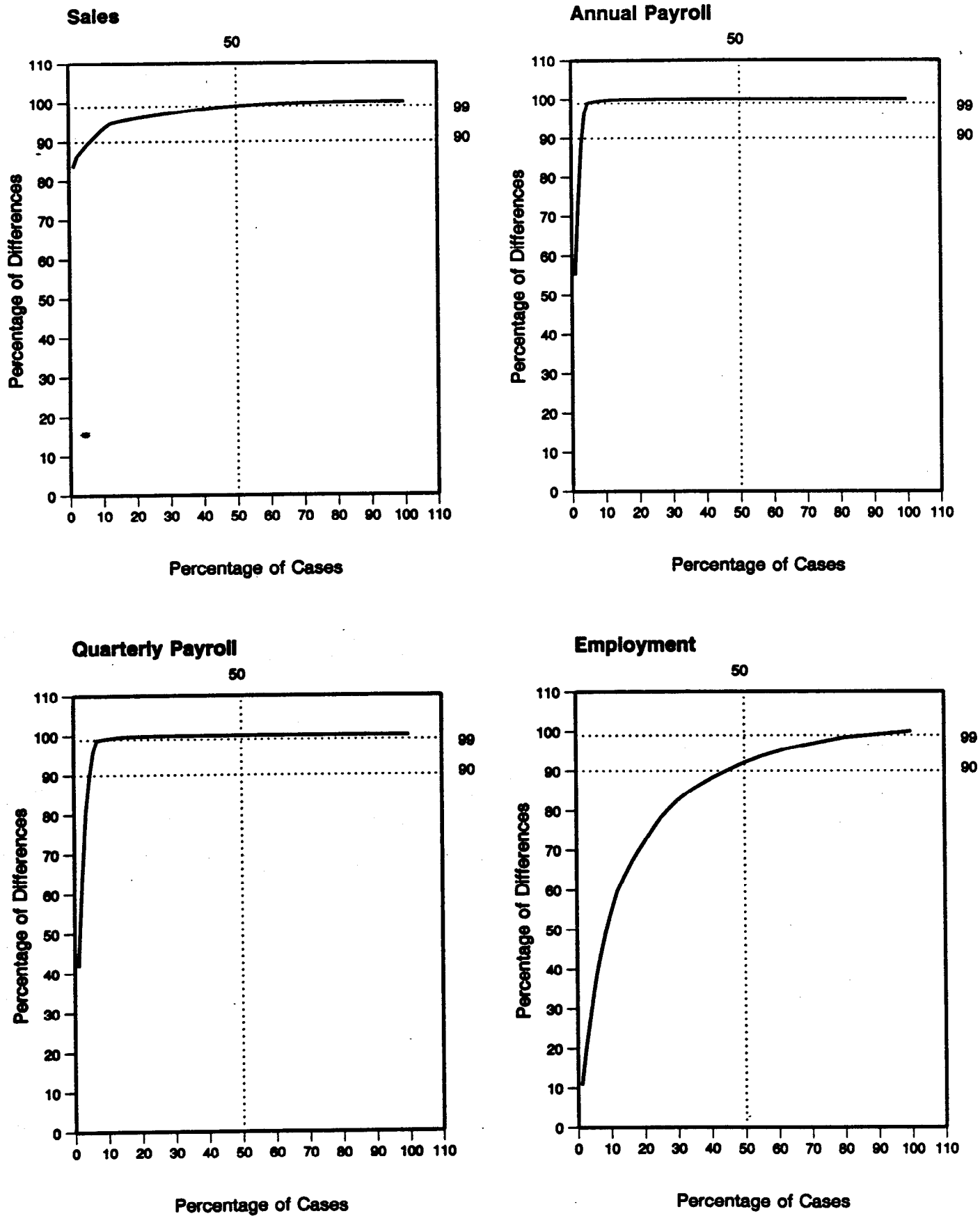


Figure 5

SERVICE SIC 783300

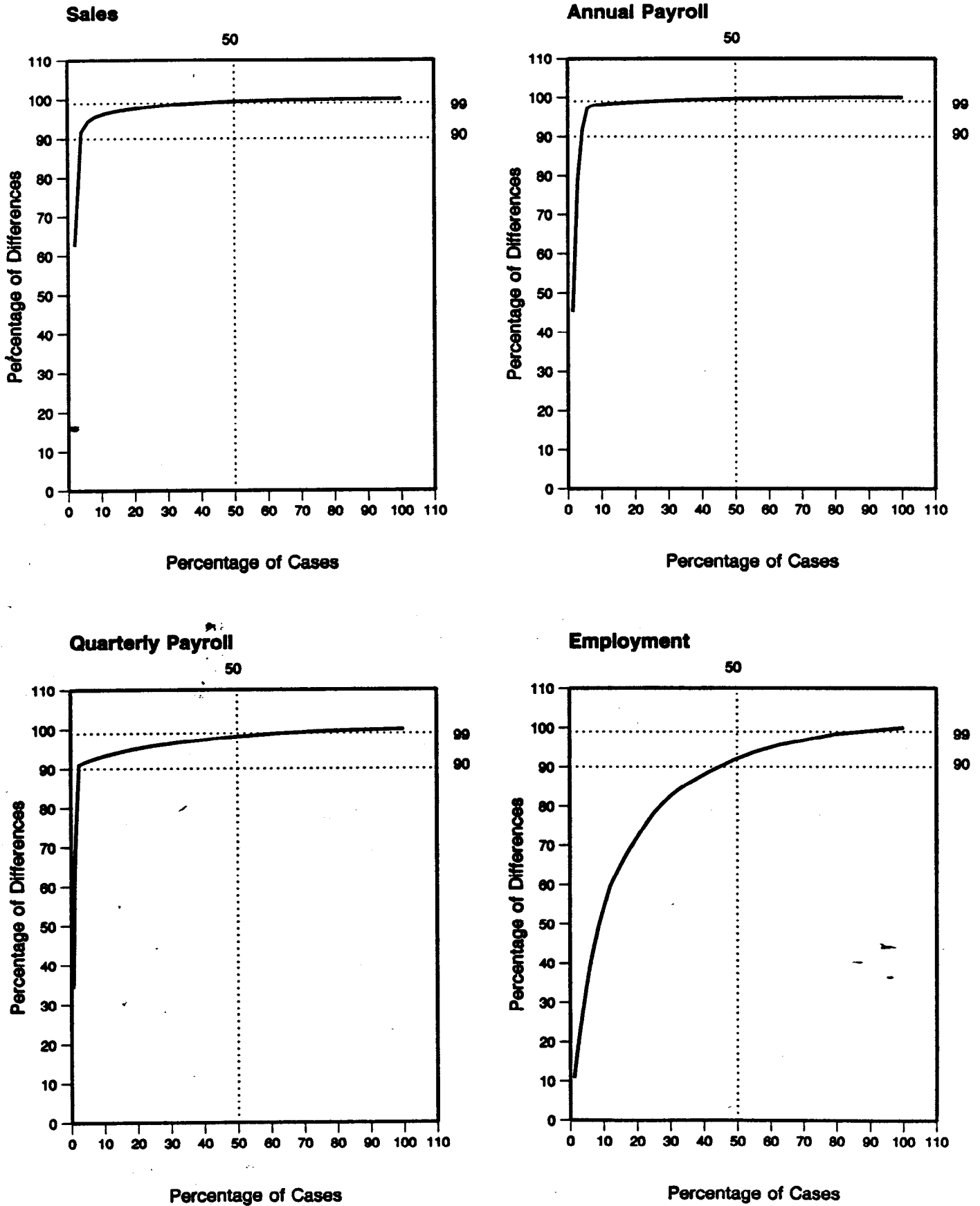


Figure 6

APPENDIX IV

In this appendix we include Table 13 which presents the rates for final cycle numbers of establishment records for each KB. These rates and the cycling process within the edit and imputation system are discussed in Section 3.4.2 of the text. We note that KB's 509311, 518100, and 531110 have relatively high rates for cycles above the first. We have been told by Business Division staff that these rates are atypically high, and these KB's were selected for this study, in part, because of difficulties in resolving many establishment records. The remaining three KB's all have relatively low rates for cycles above the first.

COUNTS OF FINAL CYCLE NUMBERS

SIC 509311		Number of Records: 2143							
Cycle Number	1	2	3	4	5	6	7	8	
Total Records	670	444	640	299	90	0	0	0	
Percent of Total	31.26%	20.72%	29.86%	13.95%	4.20%	.0%	.0%	.0%	

SIC 518100		Number of Records: 4709							
Cycle Number	1	2	3	4	5	6	7	8	
Total Records	2815	1614	228	41	4	2	1	4	
Percent of Total	59.78%	34.27%	4.84%	.87%	.08%	.04%	.02%	.08%	

SIC 531110		Number of Records: 2452							
Cycle Number	1	2	3	4	5	6	7	8	
Total Records	735	1118	462	97	32	7	1	0	
Percent of Total	29.98%	45.60%	18.84%	3.96%	1.31%	.29%	.04%	.0%	

TABLE 13

SIC 596110

Number of Records: 3467

Cycle Number	1	2	3	4	5	6	7	8
Total Records	2995	427	36	8	1	0	0	0
Percent of Total	86.39%	12.32%	1.04%	.23%	.03%	.0%	.0%	.0%

SIC 751210

Number of Records: 3479

Cycle Number	1	2	3	4	5	6	7	8
Total Records	3107	323	46	2	0	1	0	0
Percent of Total	89.31%	9.28%	1.32%	.06%	.0%	.03%	.0%	.0%

SIC 783300

Number of Records: 2282

Cycle Number	1	2	3	4	5	6	7	8
Total Records	1475	260	42	4	1	0	0	0
Percent of Total	86.55%	11.39%	1.84%	.18%	.04%	.0%	.0%	.0%

TABLE 13 (continued)