33

APPLICATION OF THE MULTIPLE FRAME DESIGN IN AN ECONOMIC DISTRIBUTIONAL EFFECTS STUDY

by
Douglas G. Kleweno

Statistical Research Division
Economics and Statistics Service
U.S. Department of Agriculture
Washington, D.C. 20250

January 1981
ESS No. AGESS 801222.3

Application of the Multiple Frame Design in an Economic Distributional Effects Study. By Douglas G. Kleweno; Statistical Research Division; Economics and Statistics Services; U.S. Department of Agriculture; Washington D.C. 20250; December, 1980. ESS Staff Report No. AGESS801222.3

ABSTRACT

A multiple frame survey design was used to study establishment and household characteristics and to make inferences about the population of interest. The methodology required for frame construction, sampling, field enumeration, and estimation is illustrated. The nine-county Kentucky study was the first of its type attempted by the Statistics unit of the Ecomonics and Statistics Service, U.S. Department of Agriculture for data collection. Procedures were substantially more difficult than our traditional work. The study was carried out effectively and demonstrates our capability for this type of project. This document was developed because of the unusual nature of the survey; however, it would be beneficial to document all special surveys that the agency conducts. Suggested procedural enhancements are given for future applications and adaptation.

Keywords: List Frame, Area Frame, Multiple Frame, NOL Determination, Household, Establishment, Survey phases, Segment, Nonresponse

Acknowledgements

The author wishes to extend special thanks to Raymond Bosecker and Robert Tortora for their valuable assistance. Thanks must also go to the Sampling Frame and Survey Research Branch for helpful comments and suggestions.

CONTENTS

SUMMARY

An economic distributional effects study of household and establishment subpopulations in a rural area of Kentucky was conducted in 1979 using a multiple frame design. The study relied on two independent frames - a list frame of establishments (private and public inclusive) and an area frame of households and establishments. Both frame sources were stratified. The list was stratified by firm function (nine standard industrial classes) and substratified into three sizes: 1-19 employees, 20-99 employees, and 100 or more employees. The area frame was stratified into three geographical areas - urban, suburban, and rural. Sampled list establishments provided a subsample of their employees for enumeration. The list firms and their sampled workers formed the overlap domain of the population. Sampled households and firms in the area frame not represented on the list formed the nonoverlap domain of the population. This determination was necessary under the assumptions of multiple frame sampling and occurred during the screening process. Data collection was confined in the area frame to sampling units in the nonoverlap (NOL) domain because of time and monetary constraints.

Three questionnaires were used to gather data - a household version, a private establishment questionnaire, and a government version. Field work extended over a three month period with data collection divided into three stages. (At this date questionnaires have been edited and summarization has started).

All subpopulation estimates of means and totals were constructed at the strata level. Domain estimates of totals were made for each subpopulation group. The independent domain estimates were then combined into composite total estimates for each subpopulation. A combined ratio estimator was used to compute household mean estimates.

SECTION 1 - GENERAL DISCUSSION OF THE SURVEY

1.1 Introduction

A major study of the distributional effects of recent economic development on the rural population was conducted by the Economics and Statistics Service (ESS), U.S. Department of Agriculture. The Economic Development Division initiated the project while the Statistics unit was contacted for the survey design and data collection. Important components of the study were to access the impact of economic development: (1) on the supply of labor, especially for women; (2) on the demand for labor; and (3) on the concomitant relationships with rural residents' participation in USDA food programs. Other areas of interest were federal, state, and local income maintenance, manpower, and economic development programs.

1.2 Survey Purpose

The survey dealt with obtaining an understanding of the effects of recent rapid growth of population and employment for a particular rural area and the role that the Federal and other government units play in this process. The primary purposes of the survey were to: (1) measure employment and income growth in various population subgroups in relation to industrial structure and growth; and (2) examine the idea that economic development programs increase employment opportunities and improve the well-being of disadvantaged groups.

Information collected by this study will be used to help assess whether the monies expended by Federal economic development programs in rural areas do generate "favorable" employment and income changes. Also, the data will be used to determine how changes in numbers and types of employment opportunities over a five year period have affected various groups of people, such as, women entering the labor force. The information collected is important in both the public and private sectors to evaluate not only the impact of economic development but to utilize the information for economic development policy planning.

1.3 Survey Design

The distributional effects study was based on the target population in a nine-county area of southeastern Kentucky. The nine counties involved in the study area were Clay, Clinton, Knox, Laurel, McCreary, Pulaski, Russell, Wayne and Whitley. This area was selected from several nonmetropolitan labor market areas where rapid population and employment growth occurred in the 1970's. The study site had an economic base similar to a large number of nonmetropolitan areas which underwent rapid growth in recent years. Community leaders and plant managers also indicated a high degree of willingness to cooperate with ESS in data collection activities.

There were two major subpopulation groups--establishments and households. An establishment was defined as "an economic unit generally at a single physical location where business is conducted, industrial operations are performed, or services are provided." A household was defined as "all persons, not necessarily related by blood or marriage, occupying a single housing unit." Within households, data was collected for all members 16 years old and older.

The establishments were divided into private and public employers. Households were split into two areas of interest -- those with employed members and those with no employed members. Based on the most recent data available there were 2,987 private establishments and 498 public establishments in the study area. In 1977, the Census population estimate of this area was 208,900 persons. They resided in about 65,300 households. For the same period, total employment was estimated at 79,000 persons.

A multiple frame survey design consisting of a list frame and an area frame was used to study the population. The list frame was composed of private and public establishments. It was used to sample employers and subsample employees for establishment and worker traits. The area frame for households and firms was used to measure incompleteness of the list frame and to collect data on household characteristics of individuals unemployed, self-employed, and out of the labor force. Data collection was split into three phases to ease manpower needs and simplify logistics of training and supervision.

The multiple frame design was chosen over a sample design based only on a list or area frame for two main reasons: (1) Inferences were desired for the whole population, but a complete list of establishments or employed people could not be achieved because of limited frame construction materials. (2) Information was desired for certain rare items but area cluster sampling along would not provide the desired precision. The design of the survey and questionnaires would also permit cross referencing or linking of data between the household and firm for part of the sample. Because employees were subsampled from the employers payroll, household characteristics could be linked to firm characteristics. This feature allowed integration of labor demand and supply analysis and permitted analysis of economic development among population groups.

The most important variables estimated are listed below. The target precision for major categories was $\pm$ 10 percent of the true value with 95 percent confidence. Not all variables, however, could be estimated with the precision desired. Certain items were too rare to be estimated with this confidence given the sample size.

Important Variables Estimated

A. Establishments

1. Ownership
2. Total employment
3. Employment by major SIC (Standard Industrial Classification) groupings
4. Employment by occupational groupings
5. Wage structure
6. Employment policies
7. Participation in various Federal, state or local funded programs
8. The establishment is new, relocated, or long time in the area
9. Employment in the area has expanded, remained constant, or decreased

B. Employee

1. Employment status (employed, unemployed)
2. Jobs or occupations held over time
3. Demographics on individuals--age, sex, education, marital status, etc.
4. Earnings and/or wage rates earned by individuals
5. Training received by household members 16+ years of age

C. Household

1. Composition
2. Income from all sources (change in economic well-being)
3. Participation over time in various Federal, state, and local programs
4. Child care arrangements
5. Residential history
6. Sources of income, and changes in these sources over time
7. Employment status of the household (employed, unemployed, out of labor force)
8. Labor force participation
9. Occupational mobility

SECTION 2 - LIST FRAME

2.1 Introduction

The list frame was a name listing of establishments both private and public located in the nine-county area. Military units were excluded because of their insignificance as an area employer and a priori knowledge of being noncooperative. The private list was constructed mostly from a Kentucky State Employment Security list and an Economic Information System list. Both lists provided information on the type of business; and the System's list gave data on firm size. The public list was developed from a 1970 Census list for local government agencies and from contact with local authorities for state and Federal entities. Telephone directories were a supplemental source of information.

In retrospect additional work was needed in building the universe list of firms. Personnel developing the list should have been assigned their responsibilities much earlier. More time was needed to classify firms by type and size, to determine if a business operated at multiple locations, to compare differences among list sources, to remove duplication, and alphabetize the firms by primary name for ease in overlap determination. All of these tasks were an important part of list building; however, time prevented a better list from being constructed.

2.2 General List Description

The universe list of establishments was stratified by standard industrial class (SIC) and size of firm. The primary objective of the stratification was to allow comparisons by industry type while reducing estimate variation. Nine strata were used to constraint that a minimum of ten firms were within a category. The nine SIC strata were substratified into three sizes for sampling. The SIC codes were mining (10-14), construction (15-17), manufacturing (20-39), transportation (40-49), wholesale, (50-51) retail (52-59), finance (60-67), services (07-09 and 70-89), and government (91-97). A firm was further classified into one of three size groups: (1) 1-19 employees, (2) 20-99 employees, or (3) 100 or more employees. Therefore, each firm had a two digit stratum code - a SIC code to identify its function and a second number to identify its size.

2.3 Characteristics of Establishment Universe List

Private Establishments were placed on the universe list using the following guidelines:

1. Establishments with the same name and address but different industry (SIC) codes were considered one firm on the universe list using the major SIC code.
2. Establishments with the same name and SIC codes but different addresses were listed separately on the universe list. (The only exception was banks where the main office was listed. All branch locations were assumed part of the main unit.)
3. Establishments with the same name but different addresses and SIC codes were listed separately on the universe list.
4. Establishments located outside of the nine-county area were excluded.
5. Establishments with the same address but different names and SIC codes were listed separately on the universe list.

Public Establishments were placed on the universe list using the following procedures:

1. All military units were excluded.
2. Government units were combined based on ability to report data.
   a. The primary name (listed first) was the reporting unit for all secondary units.
   b. Secondary units (agencies) may not be in the same county as the primary unit.
   c. Primary and secondary units were all located in the nine-county area (Exception: There were two state district forestry offices located outside the area but were reporting for only counties in the study).

The universe list of establishments was in alpha order by primary name. A computer listing of the universe (in reporter format) was supplemented with a typed government list. The supplemental government list was used to show all secondary units (agencies) associated with a primary name. Most city and county offices could be identified (for classification) only by using the secondary name source.

2.4 Public Establishments

The government employer list provided a means to study public employment traits in the nine-county Kentucky area. This name list was stratum nine of the list frame. The list by most standards was "complete"; however, the area frame was still used to test this assumption. The combined government population count was 217 sampling units. This differed from the 498 units mentioned in section 1.3 because several agency units were combined under one primary reporter.

Care was exercised during the list development to insure that each sampling unit could be identified. Some government units were listed separately while others were grouped for ease in data collection activities. As illustrated below (from the typed government listing) most of the county and Federal government offices were separate sampling units. Selected units were circled. Each of the selected agencies were interviewed in phase II with a government questionnaire. The primary one was listed first if government units were combined, as was frequented the situation for cities, school districts, and State governments. The primary name was the reporting unit where information was available. All governmental units listed below the primary unit were considered part of the primary unit. For example, the city clerk of Manchester was a sampled unit. The city clerk's office included the fire department, police department, sewer plant, and water works.

List of Federal, State, and Local Government Units or Agencies in Nine-County Area of Kentucky

Clay County

Size of Unit or Agency		Primary Sampling Unit and/or Subunits	Address of Primary Unit	Telephone Number
		Local Government		
Under 20	1.	City of Manchester, City Clerk Fire Department Police Department Sewer Plant Water Works	City Hall 108 Richmond Road	548-2923
		County Government		
Under 20	2.	Clerk's Office	Manchester	598-2544
Under 20	3.	Extension Office	Manchester	598-2789
20-99	4.	Cumberland Valley District Health Office	Manchester	598-5564
Under 20	5.	Jail	115 Court	598-2133
Under 20	6.	Judge's Office	Court House	598-2071
Under 20	7.	Probation and Parole Office	Court House	598-5195

During the interview if the primary reporter for several government units could not report for all the units, the questionnaire was completed for as many units as possible. The additional information necessary to complete the questionnaire was obtained from other respondents.

Initial list development proved inadequate for government agencies during pretest activities. The government units were structured originally so the sampling unit was the smallest definable group. This structure was ineffective because many of the reporters did not have access to the information requested and the employees often crossed the bounds of several units in their work.

After considerable effort the government portion of the list was rebuilt. This involved combining units based on contact with city, county, state, and federal officials. A new sample was then selected for enumeration. A few problems still surfaced in locating someone who could report for the sampled units.

2.5 Private Establishments

Private employer characteristics were compiled from sampling the first eight strata of the establishment universe list. Private firms located in the study site were coded into three size groups 1-19, 20-99, and 100 plus employees. Ag service firms in this study were included in the service stratum.

The universe list format had the firms listed in alphabetic order across all counties. A firm could be listed several times if it had operations in more than one lication and/or had more than a single SIC code. For example, a mining company may not only extract the ore but clean and sell it wholesale at the same or different locations. The name, address, SIC code, and telephone number would therefore be essential to identify properly the sampling unit.

The universe was known to be somewhat "incomplete", particularly for firms in the 1-19 employee size group. This was expected with new firms locating in the area and the limited time available for list development. The area frame was used to measure this incompleteness.

2.6 The List Sample

The universe list of establishments as shown in Table 1 had 3,641 firms with 1-19 workers, 251 firms with 20-99 workers, and 60 firms with at least 100 workers. A systematic sample of firms was drawn for the two smaller employer groups within each SIC code.

Each firm size group was sampled at a different rate. Firms with 1-19 employees were sampled at a rate of 1 in 10. Firms with 20-99 employees were sampled at one in four. Large firms with 100 or more employees were completely enumerated. The sampling rate was largely determined by budget constraints.

The actual sample sizes shown in Table 1 for firms were: 1-19 employees = 335 firms; 20-99 employees = 63 employers; and 100 or more employees = 60 employers. This gave a total of 458 interviewed establishments. All operations were screened for a firm interview during phase II of data collection. A few nonoverlap (NOL) area firms were interviewed in phase III. A NOL firm was defined as an establishment not on the universe list.

One problem encountered in using the list was the selection of the sample by and before all firms had been keyed into machine media using the reporter format. It required many subsequent hours to make certain that every firm in the subpopulation when the sample was drawn) was also on the final universe list with proper identification. The sample of firms was drawn early to accommodate pretest activities using acutal selected firms.

Table 1--Nine-County Kentucky Establishments

STRATUM/(SIC)	POPULATION Substratum				SAMPLE Substratum			
	0	1	2	TOTAL	0	1	2	TOTAL
1: Mining (10-14)	161	20	3	184	16	5	3	24
2: Constr. (15-17)	408	-	-	408	41	-	-	41
3: Mfg. (20-39)	136	49	34	210	14	13	34	61
4: Transp. (40-49)	212	14	1	227	22	4	1	27
5: Wholesale (50-51)	208	33	1	242	21	8	1	30
6: Retail (52-59)	1154	75	-	1229	115	19	-	134
7: Finance (60-67)	173	22	-	195	17	5	-	22
8: Services (07-09)(70-89)	706	12	2	720	71	3	2	76
9: Gov't (91-97)	172	26	19	217	18	6	19	43
TOTAL	3330	251	60	3641	335	63	60	458

SECTION 3 - AREA FRAME

3.1 Introduction

The area frame was constructed by the Statistics unit from an agricultural land-use frame developed in 1976 for Kentucky. The original frame was modified to accommodate the household as the sampling unit of interest. This process required less resources than would have been necessary to modify the Census SMSA's or develop a completely new frame. The adjustment to the area frame involved combining land-use strata into three homogeneous strata based on population density. The major advantage of using the area frame was its completeness (all sampling units are contained in the frame). This insured that all units in the study area were represented by the sample.

3.2 Purpose

The area frame served several purposes: (1) to collect data on household characteristics of individuals unemployed, self-employed, or out of the labor force; (2) to estimate for incompleteness of the list frame of establishments; and (3) to estimate the incompleteness of the list frame for employee characteristics.

3.3 Design

The area sample was a two-stage stratified cluster design. The first stage of sampling was the segment while the secondary stage was the household or establishment. The households and establishments were identified and subsampled during the first of the three data collection phases. Area households were screened and interviewed during the third phase.

For the study, land area within the nine-county Kentucky site was stratified into three classes to improve the precision of the population estimate. Subsampling within each stratum eased field staffing problems for data collection. The strata were:

2 - Urban (densely populated areas with generally over 20 dwellings per square mile, includes business, industrial and recreation areas)

1 - Suburban (ag-urban, residential-commercial, and resort areas with over 10 residences per square mile generally)

4 - Rural (all land not included in the urban or suburban classification) NO large cluster of houses were in this stratum.

A segment was defined as a sampling unit selected for enumeration. Each segment was a continuous area of land varying in size. In stratum two a segment was a city block. In stratum one a segment covered an area about one-eighth of a square mile while stratum four segments were approximately one square mile in size. Segment size was determined by the expected number of dwellings, ease of modifying the 1976 Kentucky land-use frame, budget constraints, and time required to enumerate each segment.

Every segment was outlined on aerial photography to indicate the area to be enumerated. Photo dates ranged from 1968-1973 and were supplemented with more recent county highway and city maps. Large scale city maps were essential for work in the urban stratum because of the small land area. Only minor problems were encountered with the use of this material. A few of the problems were: boundaries were sometimes difficult to find because recent built up areas could not be detected until field staff arrived at the segment, some photography was out of date, and more households in the urban stratum were expected than screening identified.

The area frame, as shown in Table 2, had 318 segments for enumeration. Selection was made from 9011 sampling units identified within count units. The sample is composed of three replicates. The urban stratum accounted for 69 segments with every sixth household selected as a subsample. The suburban stratum had 183 segments with every fourth household interviewed. The rural stratum had 66 segments for enumeration with every second household visited. The household sampling rate was set in each stratum to provide an over all sample of one percent of the household population. This was consistent with the original guidelines set down by the Economics unit to have 2000 useable household interviews.

Table 2--Nine County Kentucky Area Segments

STRATUM	AREA FRAME		
Code - Descript.	Total Sample Units	Selected (SU) Segments	HH Sampling Rate
2 - Urban	1160	69	.167 = 1/6
1 - Suburban	4568	183	.250 = 1/4
4 - Rural	3283	66	.500 = 1/2
TOTAL	9011	318	.01 (overall)

SECTION 4 - DATA COLLECTION

4.1 Introduction

The Economic and Special Survey Section, Statistics unit was responsible for coordinating and supervising questionnaire design, survey interviews, organization and administration of field work, training, and supervision of field personnel. Crop Reporting Service supported the study with hiring, training, and supervision of field personnel. Instructive manuals for all stages of data collection were prepared. Pretesting for the survey began October 15, 1979 with interviewing completed by January 29, 1980 .

4.2 Questionnaire Design

Three questionnaires were designed for this study. A household questionnaire (Appendix D-7) was developed to obtain demographic employment history for 1974 and 1979, participation in job training programs, residence history, and household income data. This questionnaire required about one half to an hour for completion by a respondent depending on whether the person was currently employed. More than a single interview could be completed in a household. As expected, the most difficult information to obtain concerned income and data relating to 1974 activities.

There were two questionnaires used for obtaining information about firms. A government questionnaire (Appendix D-5) and an establishment questionnaire (Appendix D-6) were used for public and private employers respectively. The questionnaires asked for similar information in most instances. Firm or agency characteristics were asked relating to size, type of industry, employment characteristics, payroll, hours worked, benefits, gross sales (private firms only), and sources of capital or financing. A minimal amount of 1974 data was asked; however, it still provided the respondent with the most difficulty. Many times the information was not readily accessible or at least not available with any degree of accuracy. Interview time varied from less than 30 minutes for very small firms to over two hours for businesses with more than 100 employees.

A test version of each questionnaire was initially developed by the group efforts of the Data Collection Branch of Statistics and the Economics unit. These questionnaires were pretested and then revised before being used in the main survey.

4.3 Pretest

The survey questionnaires and survey design were pretested in the nine-county study site. Twenty-five list frame establishments and six area frame segments (24 households) were contacted. A minimum of two firm interviews were completed in each list frame stratum. Two segments were selected from each of the area stratum and four household interviews were completed for each segment. The firms were contacted from the establishment sample so the data could be used without another visit (except for the public stratum where a new sample was drawn). The area segments were not in the final sample.

The pretest was necessary for information to assess: (1) respondent burden — time required to complete each questionnaire; (2) whether questions elicit the information needed; (3) extent that information can be collected for 1974; (4) variances of data items collected; (5) whether employers would allow Statistics to select a subsample of their employees, and (6) other biases and nonsampling errors.

Considerable refinement was necessary to the questionnaires. Rewording and deletion of questions occurred. The flow of the questions was improved and the length of the questionnaire was reduced by asking for less historical information.

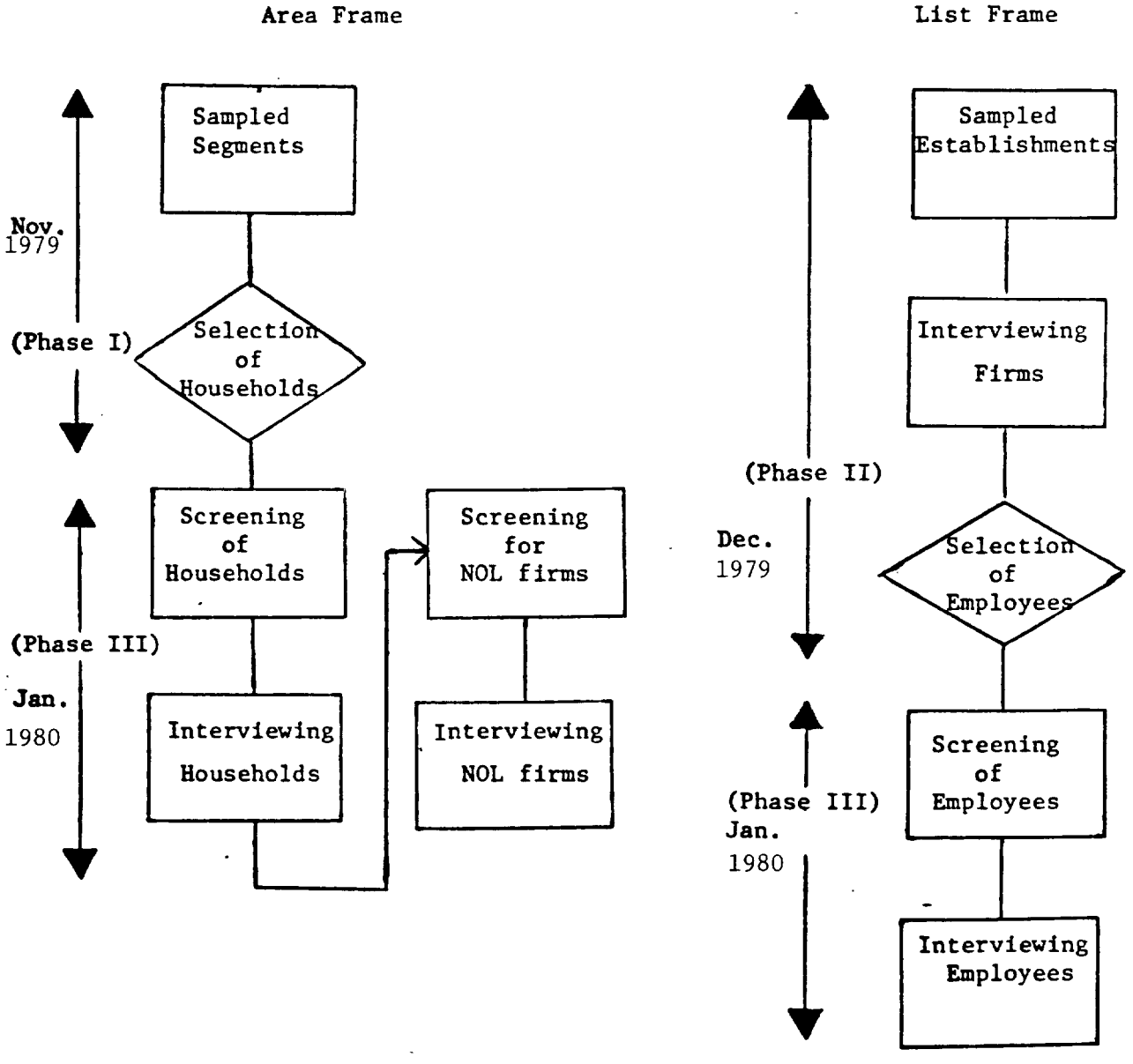
A primary concern of the survey design was the employer's willingness to allow sampling of their workers. Presurvey results indicated establishments were mostly supportive. An alternate survey plan was prepared in case the employers had not been cooperative with providing an employee list for sampling. The design stated briefly: (1) employee and household characteristics come exclusively from an area frame sample of households; (2) the area provided employer with employers; and (3) a second independent area frame would be constructed to assure complete coverage of employers. Because respondents contacted were generally cooperative, this plan was not used. No major design changes were necessary.

4.4 Flow of Field Work

The survey design permitted data collection to be completed in three stages. This was an important feature since time and manpower resources were limited. Coordination, training, and execution of the plan were also more manageable. Training schools were conducted before each phase of the field work. Phase I and II field activities were completed before starting the third phase.

The first phase of field work involved identifying households and establishments in the area frame. Phase II of enumeration required interviewing sampled establishments and selecting employees. Field activities in Phase III required screening area households and firms plus interviewing list households (employees) and NOL firms. This is illustrated in the diagram below.

FLOW OF FIELD WORK DIAGRAM



4.5 Phase I

This initial stage of data collection required field workers to locate and screen area frame segments. Enumerators with experience in reading aerial photography and maps were used to minimize training requirements and yet obtain quality results. All households and establishments in area segments were screened using the instructions given in section 4.5.1. Office personnel selected the sample of households following the steps outlined in Section 4.5.2.

4.5.1 Area Segment Screening

Field enumerators were asked to use the following instructions in screening all segments.

1. With your county highway map and aerial photo drive to the segment. At the segment, orient your aerial photo with ground features. Determine your location on the photo using roads, buildings, etc. as a guide. If possible, drive around the segment to identify ground features inside the map boundaries.

2. Starting at the initial point of entry into the segment, identify all dwelling units. Draw the household and firm locations on the map. List on the household and firm identification (ID) sheet (Appendix D-1) each unit (firm or household) so that the first unit listed and drawn on the map would correspond to the first unit observed. Each unit in the segment, whether a household or firm, would be accounted for in a clockwise manner with either a number or letter.

3. All area firms must be identified and listed for enumeration in Phase II.

4. In suburban and urban segments it will probably be necessary to sketch the segment and identify households on the sketch. Use street addresses with house numbers to identify units. Contact with a household unit should be a last resort. Be sure that the sketch and photograph agree on locations of households or buildings.

5. The segment map should identify each firm and household to correspond with the ID sheet.
   a. Firms and unoccupied houses are identified with an alphabet letter.
   b. Occupied households are identified with a numeric code.
6. Check to be sure all firms and unoccupied households are identified on the map and the identification sheet. Accurate addresses are needed for nonoverlap (NOL) determination of firms during Phase II of field enumeration.
7. Names and addresses may be obtained from mail boxes, but care must be taken to associate the correct mail box with each residence. Contact with a household should be held to a minimum and only used as a last resort to identify properly a dwelling.
8. Return completed segment with materials to supervisor or Kentucky field office.

4.5.2 Selection of Area Frame Households

All segments were returned to the office from Phase I data collection with dwelling units identified. Information was recorded on the Identification Worksheet. All occupied households had a unique numeric code and were then recorded in the appropriate column of the worksheet. All establishments and vacant houses had an alphabetic code recorded on the worksheet. These codes corresponded to those on the segment map (see completed example in Appendix B). Any problems were corrected before selecting the household subsample.

Before Phase III field work began a sample of households was selected from each segment. Each household sampled was then contacted for completion of the household questionnaire in Phase III. Firms identified and listed on the ID sheet were classified as to whether the firm was on the establishment universe list. This NOL (not on list) determination was completed by the office, based on information provided on the ID sheet. Firms identified as NOL were screened for completing an establishment questionnaire in Phase III. Any problem with making this determination in the office required a personal visit during Phase III of field work.

Steps to select household subsample used by the Kentucky office:

1. Gather Appropriate Materials:

   Aerial Photo with Segment
   Household and Firm Identification Sheet
   Random Number Table for Households

2. Use Random Number Table constructed in the office as recommended by Kish (2).

RANDOM NUMBER SHEET FOR HOUSEHOLDS
ECONOMIC HOUSEHOLD STUDY IN KENTUCKY

ENUMERATOR: _____
SUPERVISOR REVIEW: _____

URBAN STRATUM(2) S.R.=1/6						SUBURBAN STRATUM(1) S.R.=1/4						RURAL STRATUM(4) S.R.=1/2			
												SEG_NO__	N7	_SEG__NO	N8
	4		6		1	10410	(1)		1		3		2		2
	2		3		1		4		2		1		2		2
	3		1		5		3		3		2		1		1
	4		3		6		2		1		2		1		2
	3		6		4		4		2		2		1		1
	1		6		5		4		4		1		2		2
	6		4		6		4		1		2		1		1
	1		2		5		4		4		1		1		2
	4		6		2		3		4		1		1		1
	1		3		3		4		3		4		2		1
	4		5		5		3		4		3		1		1
	5		6		5		4		1		3		2		2
	2		4		6		3		3		4		1		2
	1		3		3		4		1		4		2		2
	6		2		6		1		1		4		1		2
	6		3		1		1		2		4		1		1
	1		4		1		2		1		2		1		2
	2		1		6		1		2		3		1		2
	4		3		3		4		2		4		1		2
	3		6		2		4		1		3		1		1
	5		4		3		3		3		1		2		1
	3		3		3		3		1		2		1		2
	6		2		5		3		2		3		2		1
	3		3		1		4		2		1		2		2
	5		1		5		3		2		4		1		1
	3		1		6		1		1		3		1		1
	4		2		3		3		1		4		1		2
	1		3		6		2		4		4		1		2
	4		6		5		1		3		4		1		2
	3		5		6		4		1		3		1		2
	4		1		3		2		2		3		2		1
	2		4		1		2		2		4		2		2
	6		4		6		2		4		3		2		1

A. For each segment use the appropriate columns of the table.

B. Start with the first unused number in the upper left corner from the appropriate stratum column.

   1. Record the segment number beside the selected random number.
   2. Circle the selected random number.
   3. Record the random number on the Identification Sheet.
   4. Note: The table of random numbers is only used to determine the starting point in each segment.

C. The first household selected for enumeration is the household in the occupied HH column with the same number as the selected starting point (In the example below this household is number 2).

AREA FRAME
HOUSEHOLD AND FIRM IDENTIFICATION SHEET

County: _____

Stratum: Suburban

Segment: 10410

H H Sampling Rate: 1/4

Random Starting Point: 2

1/		Occupied Households Unit No.	Firms & Vacant Houses	Firm Classification OL	Firm Classification NOL
Name: John Downe Address: 1431 Cub Circle, Toy, YS		1		☐	☐
Name: Rye Floor Coverings Address: 7450 Bear Road, Toy, YS		-	A	☐	☐
Name: Burger Bear Address: 1440 Cub Circle, Toy, YS		-	B	☐	☐
Name: Roy's Food Mart Address: 7452 Bear Rd., Toy, YS		-	C	☐	☐
Name: Toyland Address: 7453 Bear Road, Toy, YS		-	D	☐	☐
Name: Joe's Texaco Address: 7455 Bear Rd., Toy, YS		-	E	☐	☐
Name: Timothy Allen Address: 1460 Cub Circle, Toy, YS		(2)		☐	☐
Name: Rodney Duncan Address: 1465 Cub Circle, Toy, YS		3		☐	☐
Name: Sue Todd Address: 1533 Cub Circle, Toy, YS		4		☐	☐
Name: Address:		5		☐	☐
Name: Address:		(6)		☐	☐
Name: Address:		7		☐	☐
Name: Address:		8		☐	☐
Name: Address:		9		☐	☐
Name: Bond Construction Co. Address:		-	F	☐	☐
Name: Vacant House Address:		-	G	☐	☐
Name: Address:		(10)		☐	☐
Name: Address:		11		☐	☐
Name: Address:		12		☐	☐
Name: Address:		13		☐	☐
Name: Address:		(14)		☐	☐

Enumerator Check: (Total Occupied HH In Segment) X (Sampling Rate) = (Total HH In Segment)

(14) X (1/4) = 3.5 ⇒ 4 HH Samples*

*(If the random start had been three instead of two then 3 HH would have been in the sample.)

1/ Note: All households and firms are fictitious units.

D. Circle the selected household on the aerial map and also on the Identification Sheet.

E. The second and succeeding households are selected from the occupied HH column using the sampling rate given at the top of the ID Sheet.

1. If the sampling rate is 1/4 then, every fourth household following the first selected household is chosen (using occupied HH column).

2. In our example:
sample rate = 1/4 (every 4th household)
random start = 2 (household No. 2 is the first sample)
total HH in segment = 14
sampled HH in segment = 2, 6, 10, 14

F. Circle all selected households on the aerial photo ID Sheet.

G. Quality Check:

$$\left[\begin{array}{c}\text{Total Occupied House Holds} \\ \text{In Segment}\end{array}\right] \times \left[\begin{array}{c}\text{Sampling} \\ \text{Rate}\end{array}\right] = \left[\begin{array}{c}\text{Total Households in} \\ \text{Sample}\end{array}\right]$$

For example: (14) x (1/4) = 3.5 ≐ 4 households sampled.
(If the random start had been 3 instead of 2 then 3 households would have been in the sample.)

3. A completed example is provided in Appendix B. The households were then ready to be screened to determine if they qualified for an interview (Phase III).

4. All succeeding segments in their appropriate stratum used the next available number from the random number table of households for the first sampled household in the segment. (The order of using the table was from top to bottom and left to right.)

4.6 Phase II

Data collection for this phase required about 35 enumerators and supervisors. A two day training school was held December 5 and 6 prior to starting field work. Interviewing activities began immediately and were completed by January 5, 1980. The firms were interviewed the week of Christmas. Several call backs occurred because of the holiday season in retail and service industries.

4.6.1 Interviewing Firms

Establishments selected from the universe list of employers were interviewed subject to the condition they would provide a list for selection of a systematic sample of their employees. This listing was in a variety of forms and was not taken from the premises. Field supervisors reviewed all enumerators work for proper employee selection and completion of the establishment questionnaires.

A presurvey letter was sent to each firm in the sample. This letter (Appendix A) presented the purpose and general information to prepare the establishment for data collection activities.

If a firm refused to cooperate initially by not providing a list of employees to sample or by refusing to be interviewed, nonresponse procedures were followed (outlined in section 4.7).

4.6.2 Selection of Employees

Assuming the respondent cooperated, the interview with the firm representative was conducted. Upon completion of the interview, steps were outlined with the employer to draw the sample of employees.

A list of employees working full- or part-time at the time of the visit was secured first. Preferably, this was a listing by Social Security Number of employees living in the nine-county area. This insured that no employee would screen out during the interview because of the location of their residence. If the employer was not willing or able to identify employees living outside the study site area then the Social Security listing was used as it existed.

Example of Firm Register of Employees 1/

SSN	No. of Employees	Name	Address (Street-City-Co.-State-Zip)	Telephone
371-46-XXXX	1		1/	
371-82-XXXX				
499-21-XXXX				

1/ All information withheld until sample selected. Name-address was only necessary for selected sample (employees).

The concept used was that only a Social Security Number was needed to select the employee sample. This removed some of the concern about confidentiality. After the sample of Social Security Numbers was selected, the name and address associated with each of these employees was secured.

Example of Firm Register of Employees

SSN	No. of Employee on firm register	Name	Address	Telephone
371-46- XXXX	1	Bascon, Jerome E.	Box 437, Cary, Ferry, Y.S. 40218	301-344- XXXX
503-66- XXXX	5	King, Donald F.	470 N. 42nd St. Hunt, Gray, Y.S. 41268	301-344- XXXX
791-40- XXXX	9	Martin, Tim	8 Oak Circle, Brook, Madison Y.S.	301-344- XXXX

A small firm did not always have a listing of Social Security Numbers available. Instead, an alphabetic listing or other name listing was used to select the sample. If the names were on cards, each card was a possible sample unit. Large firms generally had employees identified by Social Security Number. If the information was on a computer file then a listing of Social Security Numbers was printed for the selection process. Ideally, employees living outside the nine counties were taken off the file before the numbers were printed. However, in most cases the employer could not or was not willing to make this distinction.

Enumerators were instructed to make the process of constructing and sampling the employee list as workable for the employer as possible. This was very important to insure a sample of workers to study employee characteristics.

To summarize, a variety of employee list formats given to enumerators are listed below in order of preference from most desirable to least desirable. Note the complete name-address was not needed until after employees were selected. Even the least desirable list was still useable.

1. Employee list of Social Security Numbers with individuals living outside the nine-county area excluded.
2. Employee list of Social Security Numbers for all workers.
3. Employee list of last (or full) names in alphabetic order excluding all individuals living outside the nine-county area.

4. Employee list of last (or full) names in alphabetic order for all workers.

5. Employee list (or cardfile) of last names in no order excluding all individuals living outside the nine-county area.

6. Employee list (or cardfile) of last names in no order for all workers.

The next step involved the selection of the employee sample for each firm. This process was very similar to the one used to draw households for enumeration. Field workers followed the below procedure:

Steps

1. A listing of the firm's employees had been prepared for sampling purposes.

2. Use the random number sheet (illustrated below) for employees to select the starting point to determine the first employee in the sample.

Random Number Sheet for Establishments
Economic Household Study in Kentucky

Enumerator: _____

Supervisor Review: _____

Business 1-19 Employees S.R.=1/4			Business 20-99 Employees S.R.=1/10		
_____ 1	_____ 1	_____ 1	_____ 1	_____ 1	_____ 1
_____ 3	_____ 4	_____ 4	_____ 3	_____ 1	_____ 2
_____ 3	_____ 4	_____ 1	_____ 2	_____ 5	_____ 8
_____ 2	_____ 1	_____ 2	_____ 6	_____ 6	_____ 2
_____ 1	_____ 3	_____ 1	_____ 7	_____ 6	_____ 4

a. Refer to the firm questionnaire or the employee sample list to find the prerecorded size group (1-19 , 20-99, or 100+ employees). In the example, the firm size is 1-19.

   NOTE: Always use the preentered firm size even if the firm is found to be a different size when interviewed.

b. The starting point in the table is the first unused number in the upper left corner of the appropriate size group column. Circle and record the establishment ID number beside it.

c. Record the random start number on the firm's employee sample list sheet (Appendix D-2).

3. From the firm's employee list, identify the employee that corresponds to the starting point. For example, if the random start point value was three, then the employee selected would be the third person from the top of the listing. In our example, the start point was one so the first employee (Jerome E. Bason) is selected as part of the sample.

4. Record this employee's name and address on the employee sample list (Appendix D-2) as the first employee to be interviewed.

Enumerator: _____

Employee Sample List

Business ID No: 8041
Establishment Size: 1-19 employees
Employee Sampling Rate: 1/4
Random Start No.: 1

Name	Address (#, Street, City, Co., State, Zip Code)
Bascon, Jerome E.	Box 437, Cary, Ferry Co., Y.S. 40218 Phone: 301-344-XXXX

---

5. The second and succeeding employees are selected using the sampling rate given at the top of the employee sample list sheet.

   a. The random number table is used only to select the first employee for each firm.

   b. If the sampling rate is 1/4 then every fourth employee following the first selected employee is chosen.

   In our example:
   sampling rate = 1/4
   random start = 1
   total employees in firm = 15
   sampled employees in firm = 1, 5, 9, 13

6. Record each selected employee and their address on the employee sample list.

7. The second firm and all those following in the size group will use the next unused random number as a starting point for selecting employees. Proceed in the table from top to bottom and left to right.

8. Enumerator check:
   (Total firm employees) x (Sampling rate) = (Total employees in sample)

   For example: (15) x (1/4) = 3.7 ≐ 4 employees sampled

   (If the random start had been 4 then only three employees would be selected instead of four.)

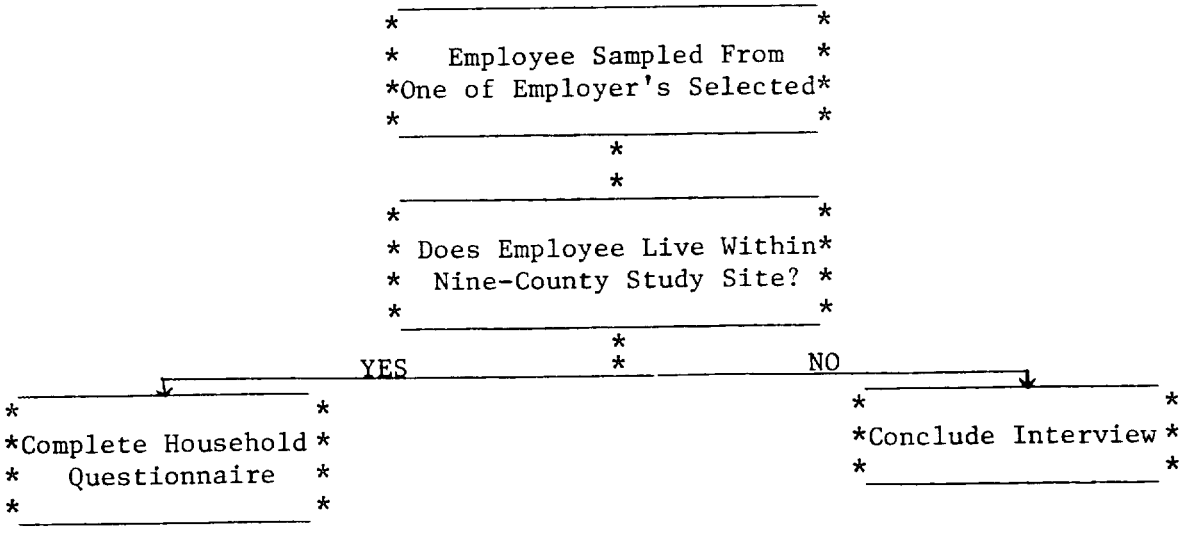
4.7 Phase III

Phase III data collection required screening and interviewing of households and area frame establishments. There were three groups of respondents involved in this procedure: (1) Households selected from subsampling firms' employee listings in Phase II (2) households selected from area frame segments during Phase I; and (3) firms identified in the area frame as not being on the list and firms not classified by the Kentucky office because information was lacking. Once screening was completed and any overlap determination made, using the list dominant concept (3), qualified respondents were interviewed.

4.7.1 Screening List Frame Employees (Households)

Employees (private and public) selected in Phase II were contacted for potential completion of the household questionnaire. The screening process which determined who qualified for an interview is described in Figure I. The major point was that household questionnaires were only completed for employees living in the study site area. This question was asked on the face page of the Phase III household questionnaire so NO screening form was used.

Figure I. Decision Diagram for List Frame Sampled Employees



Persons within sampled households were interviewed in Phase III based on responses to the screening form for area households (Appendix D-2). The area frame household screening form was completed before administering a household questionnaire.

Completion of the area household screening form required answering two questions. The first question identified households with working occupants. Households with no employed members completed a household questionnaire. The second question asked for a list of all employers by name and address if the household had employed occupants. These employers were then matched with the universe list of establishments. The enumerator was provided an alphabetic listing of firms (private and public) for this match. The respondent was asked to help identify the employer(s) on the list. The purpose of this step was explained to the respondent as the method to determine whether the household qualified for an interview.

Each firm's classification (overlap/nonoverlap) was recorded on the screening form. If any business listed on the area household screening form matched a firm on the universe list then the household was not interviewed. A household interview was completed only if no matches existed between recorded firms and the universe list of firms.

If a person was only self-employed and was not listed on the universe list, a household questionnaire was completed. If the household had just one person working and the employer was not on the universe list, a household interview was completed. If the household had no employed individuals, a household interview was completed.

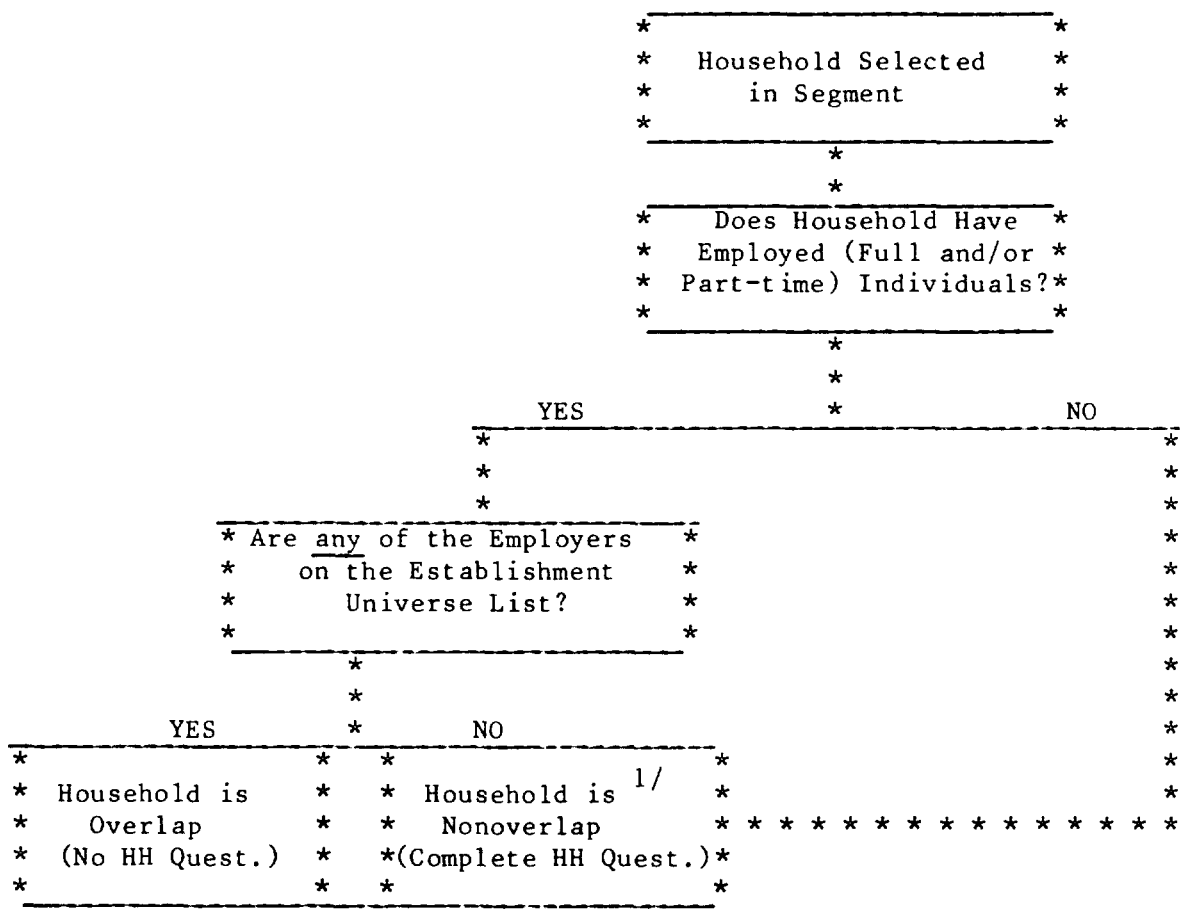
The flow chart in Figure II gives a view of the screening procedure for area frame households selected in Phase I of enumeration.

To summarize, the area frame household qualified for an interview based on the following conditions:

1. No member living in the household (minimum 16 years old) was employed.
2. The working member(s) in the household was (were) employed by a firm(s) not on the establishment universe list.
3. The working member(s) in the household was (were) self-employed and not on the establishment universe list.

A list frame household (from Phase II) was always interviewed so no screening form was necessary.

Figure II: Decision diagram for overlap concept applied to area frame households



1/ This instruction includes self-employed (nonoverlap) individuals.

The assumption was made that the respondent could provide the necessary information to classify household members accurately and identify corresponding employers. An individual's status related to the time of the interview. If possible all information was obtained in one visit. If the interview was concluded for any reason comments were provided.

4.7.3 Screening and Overlap Determination of Area Frame Firms

Firms located in sampled segment were identified during Phase I field activities and were recorded by and enumerator on the Identification Sheet. All area firms were reviewed and most were classified by the State office before Phase III work began. There were situations, however, when insufficient name and address information prevented classification. Only area firms which could not be classified (for lack of information) and firms classified as non-overlap (not on list) were contacted in Phase III.

Area firms visited in Phase III were all screened with an area frame establishment screening form (Appendix D-3). The forms had the identification fields coded and were placed in the segment kit envelopes. Each of the area firms were classified during the screening process by personal interview. All firms classified as nonoverlap (NOL) were interviewed with an establishment questionnaire. This screening and overlap determination was completed in the presence of the respondent. The procedure held regardless of segment stratum.

A sample of employees was not taken from any NOL area firms.

During this screening process it was essential that the firm's name-address-telephone be complete and accurately recorded. This information was then used to match firms with the universe list. Question three on the screening form was probably the most difficult to interpret. It read as follows:

3. Is this establishment currently operated as a unit with any other establishment in this nine-county area? (Units for which records on employees, sales, etc., cannot be separated.)

   ☐ YES - Continue.

   ☐ NO - Complete establishment questionnaire.

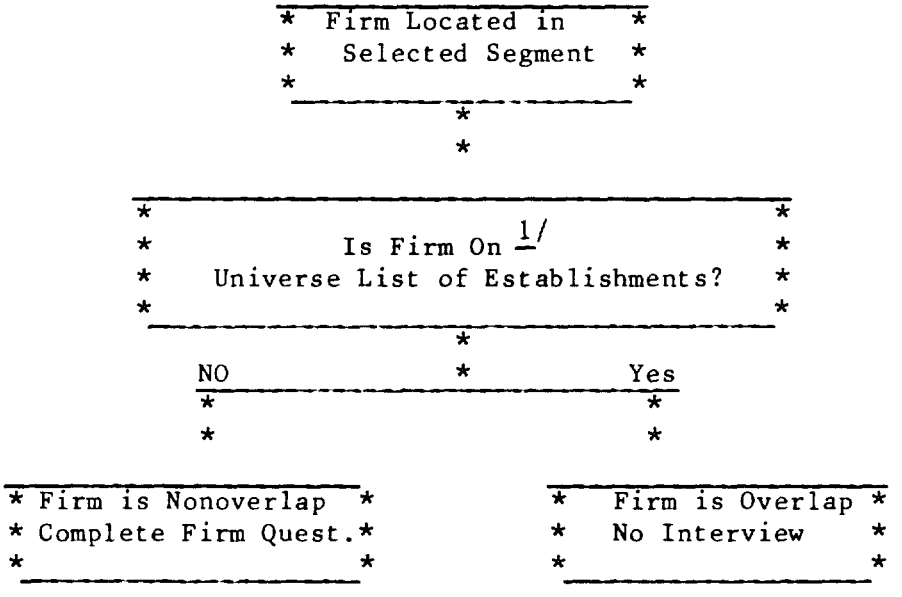
The rule was that if a firm operates at more than one location and each unit was kept separate for recordkeeping purposes, then each unit within the nine-county area was treated as a unique operation. This was necessary to be consistent with the development of the universe list of establishments.

The next step required comparison of the screened firm with the universe list of establishments. The alphabetic listing of firms (private and public) was again used for this name and address match. The universe list was in alphabetic order by primary name. The computer listing of the universe list was in reporter format. If there was a question as to the governmental units included in the primary name; the enumerator was instructed to review the secondary name field.

The universe list was used to match every name and address with the firm screened. Any question about the correct classification of a firm was reviewed by the supervisor before any further action was taken. This occasionally required the enumerator to dismiss themselves from the premises with the understanding that a return visit could be necessary. If an accurate OL determination could be made at the firm site, the second visit was not needed. If the respondent inquired where the firm list was obtained, it was explained.

A flow diagram of the classification procedure is given in Figure III.

Figure III: Decision Diagram for Area Frame Firm Overlap Determination



1/ Does firm in segment match firm name on universe list of business where the name in the area segment can be matched with the firm on the list?

Some general guidelines used for matching firms between the area and the list are given below.

General Guidelines for NOL Determination

Situation	Decision
Area frame firm and List frame firm had:	
Matching names and matching addresses	Overlap
Matching names but different addresses	Potential OL 1/
Different names but matching addresses	Potential OL 1/
Different names and different addresses	Nonoverlap

1/ Additional screening was necessary to verify if firm names or addresses were the same between frames to classify firms correctly.

Described below are examples of some frequent situations which arise in OL determination. These were reviewed with enumerators. Each case stated below was presented with an explanation of how to classify properly the firm.

Case	Problem	* Area Firm	* Firm Name on List	Firm Classified
1	Operation name change	Ted Bordon Gravel Oak Grove Road	Gordon Gravel Co. 114 Oak Grove	NOL
2	Different address for firm	Jones Bros. 423 Lonely Road	Jones Bros. P.O. Box 472	OL
3	Different spelling of name, same address	Tanner Shoe Co. 742 Main	Taner Shoe Co. 742 Main	OL
4	Company with two divisions of operation at same address	Stewart Supply Inc. 459 3rd Ave.	Stewart Advertising 459 3rd Ave. Eagle Stewart 459 3rd Ave.	OL
5	Company with two divisions of operation at different addresses	Ruebens Mining Co. Old Mine Road	Ruebens Mining Co. Railroad Station	NOL

* Note: *The firms used are fictitious units and appear only for the purpose of discussion.*

Discussion

CASE 1: This area firm would be classified NOL given the information presented. The firm names are different and addresses do not match. There is a "major" name change such that the two firms would not be identified as the same unit; therefore, this firm is nonoverlap.

CASE 2: This area firm appears to be overlap. The only difference in identity which needs verification is the address. The area firm address looks like a location whereas the list address is only a mail drop. This would be clarified during the screening process with the respondent.

CASE 3: This area firm is classified overlap even though the two names have a slightly different spelling. The same address is the decisive factor; however, the enumerator should still check to insure the companies match.

CASE 4: The area firm Steward Supply Inc. has two divisions located at the same address- Eagle Stewart and Stewart Advertising. Although Stewart Supply Inc. is not on the list, both divisions are present at the correct address so the classification is overlap.

CASE 5: The Reubens Mining Company has two divisions of operation at different locations. The mine is located on Old Mine Road and a wholesale office at Railroad Station. The area firm is nonoverlap because the division at Old Mine Road is not on the universe list.

Once the firm in the area segment was classified as overlap or nonoverlap, this information was recorded on the area frame ID sheet. A NOL firm had a check mark placed in the far right column NOL box. The necessary information for identification was then transferred to the firm questionnaire before starting the interview.

4.7.4 Interviewing List Households & NOL Area Households and Firms

This part of data collection was the most labor intensive requiring about seventy people. Two training schools, each two days long, were conducted the first few days of January 1980. Data collection was terminated by January 30.

People trained in Phase I and II to work with photography and interviewing of establishments handled all firm contacts. New workers were used for household interviewing so only the household questionnaire was covered in the training sessions.

At the school it was necessary to train a select group of experienced enumerators for making the NOL determination for area households and the few remaining area firms to be classified. Comprehension of this subject was faster than expected. Supervisors again reviewed all field work.

A major problem surfaced when it was necessary to contact household members using the name and address provided by the employer after selecting employees. Sufficient information was not available in some cases to locate the correct respondent's home. In other situations two people with the same name were confused when the street address or telephone number was not provided. The wrong household was thus interviewed.

Potential solutions for future surveys using this design are as follows: 1) Request the employer to provide street addresses and/or telephone numbers for all selected employees. 2) Use a screening question to insure the respondent actually works for the appropriate employer. 3) Over sample employees to insure the desired household interviews needed for summary. This would compensate for inaccessible samples and the number of units screening out because they lived outside the study site boundaries. Using this survey as a guide a 20 percent adjustment on the desired sample size would be resonable.

4.8 Nonresponse Follow-up Procedures

Discussion of this topic relates to Phase II and Phase III of data collection. List firms (100+) providing an interview in Phase II but refusing to give an employees list could not be replaced since all firms in this substratum were enumerated. To compensate for this problem the area frame was allowed to account for their employees in Phase III. A firm with 100 or more employees refusing to be interviewed or classified as inaccessible was considered a nonresponse. Expansion factors were adjusted for nonresponse firms.

A list firm in size group 1-19 or 20-99 employees was replaced if the firm refused to cooperate. The term refuse was defined as not willing to be interviewed and provide a list of employees for sampling. Enumerators were instructed to call the State supervisor for a replacement firm in the same stratum-substratum as the firm lost.

The state office used the following procedure to select firm replacements:

Steps

1. Use random number sheet for nonresponse (provided by D.C.) (Appendix E)
   a. Select a random number within interval (1-19: 1 in 9; and 20-99: 1 in 3) to determine an alternate firm.
   b. Take the first unused random number in the appropriate column.
2. Use the establishment universe list (from which sample was initially selected) to indentify firm replacement.
   a. Within sampling interval of nonresponse firm, number the firms from 1-3 or 1-9 consecutively. Start with the firm following the original sample.
   b. MATCH the random number selected to the firm and MARK to indicate it as the new replacement.
   c. Continue this process if the new firm refuses until all firms within their sampling interval have been chosen once. No further substitution is possible at this point. For firms in sizes 1-19 and 20-99 a maximum of nine and three substitutes are possible respectively.
   d. Provide the enumerator with the new alternate firms and make a special listing of the replacements for processing and recordkeeping.
      (1) Substitute questionnaires should be checked in with the original sample.
      (2) Only the completed interview should be coded and edited. Staple other questionnaires to the back.

During Phase II a nonresponse, whether it be a household or a NOL firm, had no replacement.

SECTION 5 - EDITING AND SUMMARIZATION

5.1 Introduction

Questionnaires were hand edited and then machine edited. Several open-ended questions required conversion to a machine code for summarization. The hand edit was done in the Kentucky office using both Statistics and Economics staffs. The machine edit was a general system of checking for relationships, consistency, and completeness of critical items. Errors in a report were flagged for review and correction before being passed to a data tape. The Statistics unit completed their responsibility with the study after a clean data tape was prepared for the household and establishment data.

The Economic Development Division was responsible for the summarization and analysis of the data. The data guidelines for summarization were outlined by the Statistics Survey Research Section.

Several steps were required to summarize the information. Some output cells involved one stratum while others included all strata of the list and area frames (overlap and nonoverlap domains). For example, a government output table used only stratum nine (public employers) employment practices by occupation for both domains. This contrasts with a household output table where total employment by occupation group was summarized across all strata for both domains. The flow of firm data is outlined in Figure IV. Estimates for domain totals were made separately for the establishment list and area frames and then were added to determine an overall (composite) value.

Household data originating in the two frames is illustrated in Figure V. This diagram shows one of the more complex situations since both the area and list frames were included in estimating households or household members' (subunit) attributes. Individual estimates were initially made for each domain and then the estimates were additive. The corresponding variances were also summed to a total for the household variables estimated.

To compute an estimate for individual items it was necessary first to expand each report based on the inverse of the sampling rate and any adjustment for nonresponse. Expansion factors and the estimation process are discussed in Section 6.

Figure IV: Flow of Establishment Data

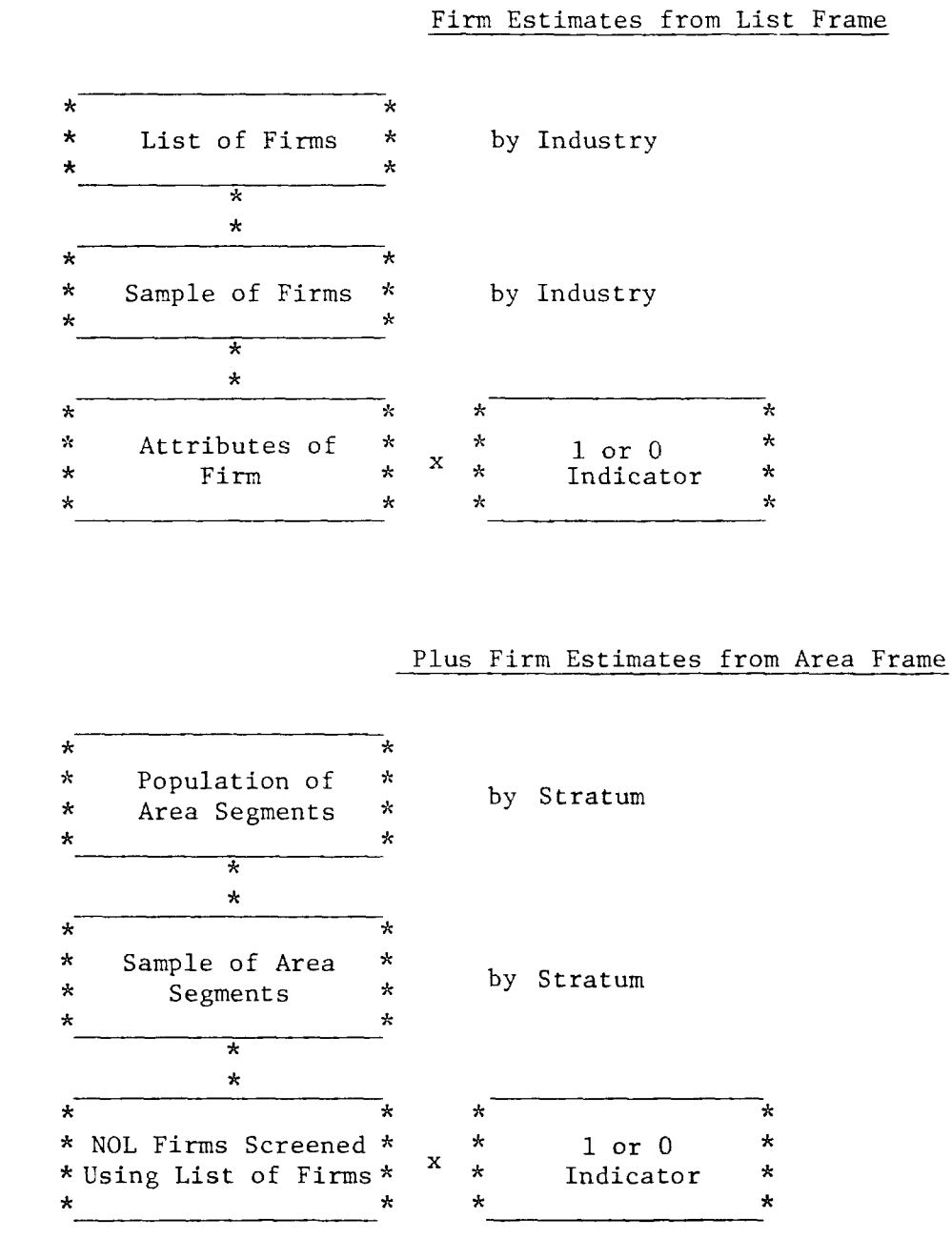
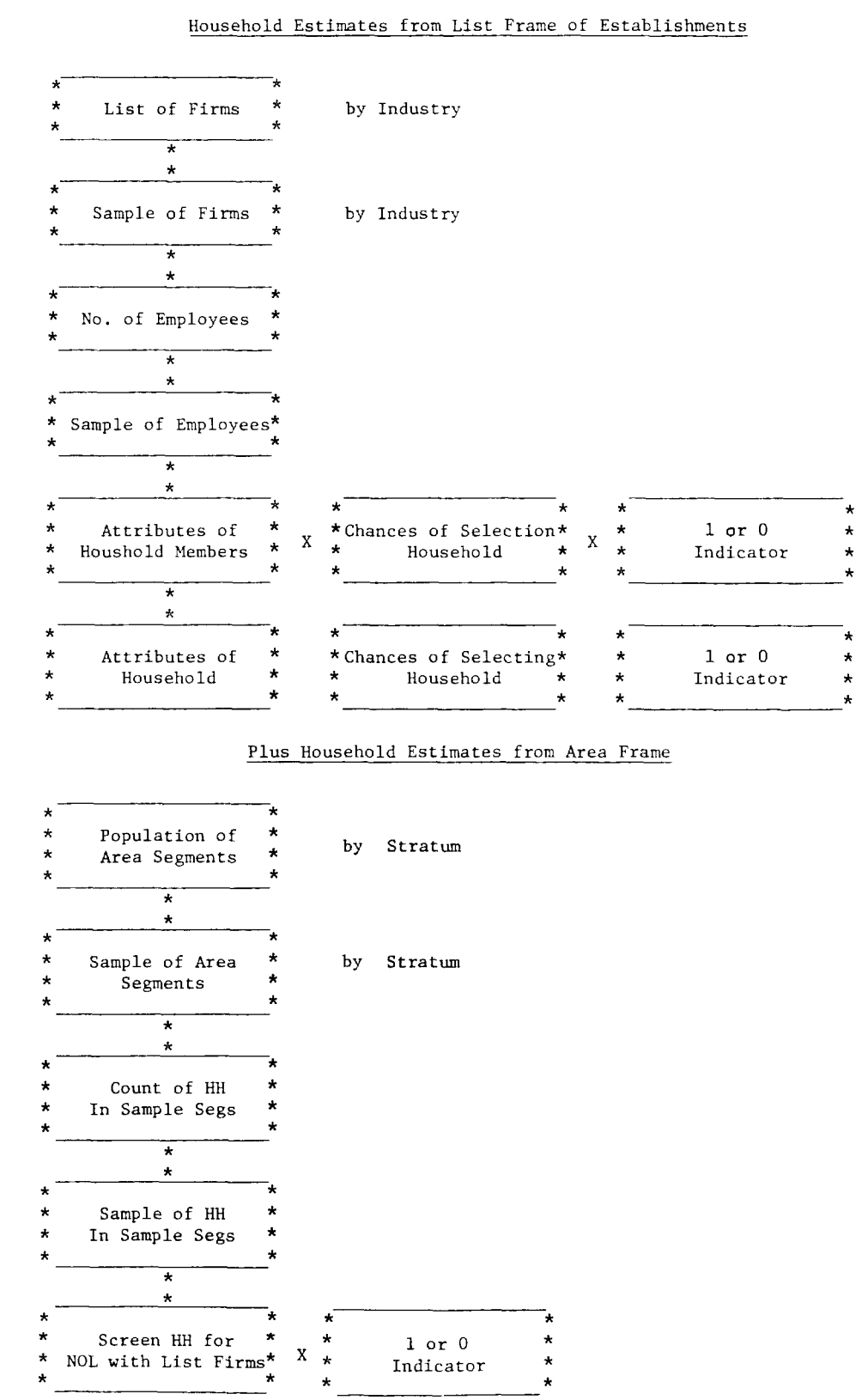


Figure V: Flow of Household Data



SECTION 6 - ESTIMATION

6.1 Introduction

Analysis of the survey data was developed around estimators for totals and a mean for households (employees) and establishments. To present the estimators, the format will be as follows: (1) provide the basic notation for the development of the estimators; (2) Estimate employee (household) domain totals; (3) estimate establishment domain totals; (4) combine domain estimates to arrive at subpopulation composite totals; and (5) estimate means for employee (household) and establishment subpopulations.

Notation used to develop the estimators for the domains is given in Table 3. Each estimate $\hat{X}_{(g)h}$ was identified by the subpopulation and frame source (g) and stratum (h) for the variable (X) of interest. For example, the estimator for variable X in stratum nine (government) of the establishment subpopulation would be shown as $\hat{X}_{(2)9}$. If an employee trait was cross classified with the employer for all strata then $\hat{X}_{(4)h}$ was summed for h = 1, 2, ..., 9. Household questionnaires received the employers' reporter numbers so employees could be identified with the firm for this link. Only data groups were linked to avoid disclosure of individual reports.

Table 3-- Domain Estimators by Stratum h = 1, 2, 4 for Area Frame; 1, ..., 9 for List Frame

Notation $\hat{X}_{(g)h}$; g=1,2, 3 ,4	Subpopulation Group	Domain	Frame Source
$\hat{X}_{(1)h}$	Establishment	Nonoverlap	area
$\hat{X}_{(2)h}$	Establishment	Overlap	list
$\hat{X}_{(3)h}$	Household (Employee)	Nonoverlap	area
$\hat{X}_{(4)h}$	Household (Employee)	Overlap	list

All estimators were constructed at the stratum level. Estimates of variance were formulated for all variables except those cross classified. The finite population correction was ignored since the sampling fraction was very small.

6.2 Estimating Employee (Household) Domain Totals

The general formula to estimate employee or household traits for the non-overlap area frame and the overlap list frame is given below by $\hat{X}_{h}^{(3)}$ and $\hat{X}_{h}^{(4)}$ respectively. The values of f_{oh}, P_{hij}, and the indicator variable depended on the conditions characterized by each data source. For example to estimate $\hat{X}_{h}^{(3)}$ where the household area frame was the frame source: the expansion factor $f_{oh} = 100$, $P_{hij} = 1.0$, and the indicator variable would be one if the attribute was present in the sampling unit and the household was classified nonoverlap.

Under the assumption of equal firm sizes within size groups the general unbiased employee (household) estimator for domain totals, referenced from Cochran's 3rd edition (1), for stratum h was

$$\hat{X}_{h}^{(g)} = \sum_{i}^{n_h} \sum_{j}^{m_{hi}} \left\{ f_{oh} \cdot P_{hij} \cdot \begin{bmatrix} 1 \text{ if sampling unit (su) has attribute} \\ 0 \text{ if otherwise} \end{bmatrix} \cdot X_{hij} \right\} \quad (1)$$

where,

$$f_{oh} = \begin{cases} 40 = \text{Expansion factor for employees from preselect firms in list frame (with 100+ employees)} \\ 40 = \text{Expansion factor for employees from selected firms in list frame (1-19 \& 20-99 employees)} \\ 100 = \text{Expansion factor for employees from selected segments in area frame} \end{cases}$$

P_{hij} = $0 < P_{hij} \leq 1$, P_{hij} is probability of selecting household where

$$P_{hij} = 1 \div \begin{bmatrix} \text{Number of establishments on Universe List who employ a member of the household.} \end{bmatrix}$$

1 or 0 = indicator variable for presence of attribute. For the NOL employee estimator $X_{h}^{(3)}$ the condition that the household be NOL is also required,

X_{hij} = variable to be estimated in domain,

g = 3 or 4 for employee (household) estimators,

h = stratum level: (industry h = 1, 2, ..., 9 or area h = 1, 2, 4)

i = primary sampling units (PSU) level: (firm and segment),

j = subunit (SU) level: (household and employee),

m_{hi} = sampled number of subunits (in the PSU of h^{th} stratum), and

n_h = sample size of PSU's within stratum h

To estimate domain totals for employees of all industries it was necessary to sum across all strata for each frame source. A code box was provided for the industry code on the area household questionnaire. This was necessary for computing domain totals for NOL households since there was no direct identity with the industry.

Because most output tables required only a total number of employees with a specific attribute, employees having the attribute caused X_{hij} to equal one when multiplied times the indicator variable. Otherwise it was zero. The appropriate expansion factor and household probability, as stated in the example, was then multiplied by X_{hij} producing either a positive value or zero if the attribute did not exist.

The variance estimate for employees' domain totals ($_{(g)}\hat{X}_h$) in stratum h where g = 3 or 4 was

$$v(_{(g)}\hat{X}_h) = \left\{ \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} (f_{oh}\, \hat{X}_{hi} - {}_{(g)}\hat{\bar{X}}_h)^2 \right\} \qquad (2)$$

where, $\hat{X}_{hi} = \sum_{j=1}^{m_{hi}} \left[(X_{hij})(P_{hij}) \right]$ = the sample total in the h^{th} stratum and the i^{th} PSU (firm or segment) adjusted for proration factor (P_{hij}),

$_{(g)}\hat{\bar{X}}_h = \frac{_{(g)}\hat{X}_h}{n_h}$ = the overall sample mean per stratum; $_{(g)}\hat{X}_h$ computed in (1),

f_{oh} = expansion factor from (1),

n_h = sample size within stratum h (firm or segment), and

P_{hij} = probability of selecting household from (1)

The variance estimate was computed based on only the first stage of sampling. This was possible since the subunits are self weighted.

6.3 Estimating Establishment Domain Totals

The estimator presented below for list frame establishments was confined to the strata level. Estimates were summed across all strata for an overall domain total. The selected PSU's (a total of n_h units) were drawn with probability $1/f_{oh}$. If a firm(i) could not report for multiple sampling units or was duplicated on the universe list the data was adjusted with the proration factor factor (P_{hi}). The factor was 1.000 for each firm unless an adjustment was needed.

The general unbiased list for domain totals within stratum h was

$$(2)\ \hat{X}_h = \sum_{i=1}^{n_h} \left\{ f_{oh} \cdot P_{hi} \cdot \begin{cases} 1 \text{ if PSU has attribute} \\ 0 \text{ otherwise} \end{cases} \cdot X_{hi} \right. \qquad (3)$$

where,

n_h = sample of firms in h^{th} stratum,

X_{hi} = variable to be estimated in domain,

P_{hi} = Proration Factor for i^{th} firm duplicated on list (from completion code 6 on questionnaire),

f_{oh} = $\begin{cases} 10 \text{ Exp. factor for } \underline{\text{list}}\ \underline{\text{firms}} \text{ in substratum } \underline{\text{zero}} \text{ (1-19 employees)} \\ 4 \text{ Exp. factor for } \underline{\text{list}}\ \underline{\text{firms}} \text{ in substratum } \underline{\text{one}} \text{ (20-99 employees)} \\ 1 \text{ Exp. factor for } \underline{\text{list}}\ \underline{\text{firms}} \text{ in substratum } \underline{\text{two}} \text{ (100+ employees)} \end{cases}$

1 or 0 = indicator variable for attribute presence,

h = stratum level: (industry h = 1, 2, ...9), and

i = primary sampling unit (PSU) level: (firm).

The variable X_{hi} was nonzero after multiplication with the indicator variable only if the PSU had the attribute in the domain; otherwise, X_{hi} was zero.

The general formula to estimate for area frame nonoverlap firms (NOL) was very similar to the computation in (3). The only differences were the addition of a subscript (j) to sum across secondary units, and the factor P_{hi} was a constant equal to one. The notation was adapted to add to the stratum (h), across primary sampling units called segments (i) and across secondary units which were NOL area firms (j). So modifying formula (3) and noting the change in the domain qualifications gave the

general area NOL firm estimator for domain totals within stratum h:

$$(1)\hat{X}_h = \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} \left\{ f_{oh} \cdot \left[\begin{matrix} 1 \text{ if sampling unit has attribute} \\ 0 \text{ otherwise} \end{matrix} \right] \cdot X_{hij} \right\} \quad (4)$$

where, n_h = sample of segments in h^{th} stratum (h = 1, 2, or 4),

m_{hi} = NOL firms in stratum h and segment i,

X_{hij} = variable to be estimated in domain,

f_{oh} = $\begin{cases} 1160/69 \text{ Exp factor for area NOL firms in stratum h = 2 (urban)} \\ 4568/183 \text{ Exp factor for area NOL firms in stratum h = 1 (suburban)} \\ 3283/66 \text{ Exp factor for area NOL firms in stratum h = 4 (rural)}, \end{cases}$

1 or 0 = indicator variable for attribute presence; the firm must also be classified nonoverlap,

h = stratum level for area frame 1, 2, or 4,

i = primary sampling unit: segment, and

j = secondary sampling unit: NOL firms.

If the domain total estimator $_{(1)}\hat{X}_h$ was used then the PSU had the attribute and the firm had to be nonoverlap for X_{hij} to be nonzero after multiplied by the indicator variable.

The variance estimate for list firm domain totals ($_{(2)}\hat{X}_h$) in stratum h was

$$v(_{(2)}\hat{X}_h) = \left\{ \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left[\left(f_{oh} \right) \left(P_{hi} \right) \hat{X}_{hi} - {}_{(2)}\bar{\hat{X}}_h \right]^2 \right\} \quad (5)$$

and,

n_h = sample of firms in h^{th} stratum

$\hat{X}_{hi}$ = variable to be estimated for in domain

$_{(2)}\bar{\hat{X}}_h = \left(\frac{_{(2)}\hat{X}_h}{n_h} \right)$ = firm estimate from domain mean in stratum h; $_{(2)}\hat{X}_h$ computed in (3)

f_{oh} = expansion factor from (3)

P_{hi} = proration factor from (3)

Because secondary units were self weighted, the variance formula was simplified to a single stage computation. To compute the variance for the domain total, it was necessary to add across strata.

The variance estimate for area NOL firm domain totals ($_{(1)}\hat{X}_h$) in stratum h was

$$v(_{(1)}\hat{X}_h) = \left\{ \frac{n_h}{n_h - 1} \left[\sum_{i=1}^{n_h} (f_{oh}\, \hat{X}_{hi} - {}_{(1)}\hat{\bar{X}}_h)^2 \right] \right\} \qquad (6)$$

where, f_{oh} = sampling rate from (4)

$\hat{X}_{hi} = \sum_{j}^{m_{hi}} X_{hij}$ = sample total in h^{th} stratum and i^{th} PSU (segment), and

$_{(1)}\hat{\bar{X}}_h = \left(\frac{{}_{(1)}\hat{X}_h}{n_h} \right)$ = overall sample mean per stratum; $_{(1)}\hat{X}_h$ computed in (4)

6.4 Combining Domain Total Estimates

To consider an attribute for the entire population it was necessary to sum the estimated domain totals across strata and then combine all appropriate frame sources. The first step was to sum the estimated domain totals in (1) and (3) across strata. The general notation to do this computation for any frame source (g) was $_{(g)}\hat{X} = \sum_{h} {}_{(g)}\hat{X}_h$.

The second step involved combining estimates and corresponding variances for various frame sources. This technique is presented in Table 4. The variances were additive without a covariance term since the frame sources were independent.

Table 4--Combining Domain Total Estimates

Composite Total Estimate	Composite Variance Estimates	Subpopulation Estimate Type
$_{T_1}\hat{X} = {}_{(1)}\hat{X} + {}_{(2)}\hat{X}$	$v(_{T_1}\hat{X}) = v(_{(1)}\hat{X}) + v(_{(2)}\hat{X})$	Composite Firm Estimator
$_{T_2}\hat{X} = {}_{(3)}\hat{X} + {}_{(4)}\hat{X}$	$v(_{T_2}\hat{X}) = v(_{(3)}\hat{X}) + v(_{(4)}\hat{X})$	Composite Employee Estimate

To estimate, for example, total employed people in the nine-county area the composite estimate $_{T_2}\hat{X}$ was used. The first component ($_{(3)}\hat{X}$) was the NOL domain estimate from the household area frame; and the second domain estimate ($_{(4)}\hat{X}$) was the total employees subsampled from the list frame. The composite estimate varied depending on the subpopulation of interest.

6.5 Estimating Household (Employee) Means

A combined ratio estimator (1) was used to estimate means for employee (household) characteristics. This estimator allowed the denominator to be a random variable since the total number of household units from the subpopulation was not known. So the estimator is really the ratio of two variables, both of which varied from unit to unit. Examples using this procedure were average income of employees in the retail industry and average hours worked per week by part-time retail employees.

To use the combined ratio estimator ($\hat{R}$) variable Y was introduced. The Y variable was estimated (using $_{T_2}\hat{Y}$) for domain totals of the appropriate subpopulation exactly as the X variable was estimated ($_{T_2}\hat{X}$). No new formulas were needed. It was essential, however, that domain totals for variables X and Y be summed across all strata and frame sources before computing $\hat{R} = (_{T_2}\hat{Y}/_{T_2}\hat{X})$. For the above example $_{T_2}\hat{X}$ was the estimated total employees in the retail industry and $_{T_2}\hat{Y}$ the total income for those employed in the retail industry. The ratio estimator then became the average income of retail employees.

The combined ratio estimator ($\hat{R}$) for estimating household (employee) averages was

$$\hat{R} = \frac{_{T_2}\hat{Y}}{_{T_2}\hat{X}} \tag{7}$$

The estimated variance for employee averages using the combined ratio estimator was

$$v(\hat{R}) = \frac{1}{_{T_2}\hat{X}^2}\left[v(_{T_2}\hat{Y}) - 2\,\hat{R}\,\mathrm{Cov}(_{T_2}\hat{X},\ _{T_2}\hat{Y}) + \hat{R}^2\,v(_{T_2}\hat{X})\right] \tag{8}$$

where,

$_{T_2}\hat{X}$ and $_{T_2}\hat{Y}$ = composite total estimates for variables X and Y (see Table 4),

$\hat{R}$ = combined ratio estimator computed in (7),

$v(_{T_2}\hat{Y})$ = composite variance estimate for Y (see Table 4),

$v(_{T_2}\hat{X})$ = composite variance estimate for X (see Table 4) and

$$Cov(_{T_2}\hat{X},\ _{T_2}\hat{Y}) = \sum_g \left[Cov(_{(g)}\hat{X},\ _{(g)}\hat{Y})\right] \text{ is the sample covariance} \qquad (9)$$

between $_{T_2}X$ and $_{T_2}Y$ where,

$$Cov(_{(g)}\hat{X},\ _{(g)}\hat{Y}) = \sum_h \left\{ \frac{n_h}{n_h - 1} (f_{oh})^2 \left[\sum_i^{n_h} X_{hi}\, Y_{hi} - \frac{\sum_i^{n_h} X_{hi} \sum_i^{n_h} Y_{hi}}{n_h} \right] \right\} \qquad (10)$$

and, g = 1 and 2 subpopulation and frame sources for each estimator,

h = 1, ..., ℓ strata

$X_{hi} = \sum_j \left(X_{hij}\ P_{hij}\right)$, where P_{hij} = probability of selecting jth household, in (1)

$Y_{hi} = \sum_j \left(Y_{hij}\ P_{hij}\right)$, where P_{hij} = probability of selecting jth household, and in (1)

f_{oh} = expansion factor for X_{hi} and Y_{hi} in (1).

6.5 Estimating Establishment Means

The procedure for estimating firm means was much simplier than for employees since the denominator of the estimator (N') was not a random variable.

The overall firm mean ($_{T_1}\hat{\bar{X}}$) was computed using the composite total estimates corresponding variances from strata and frame sources of interest (Table 4). The composite firm total estimates was $_{T_1}\hat{X}$ with variance $v(_{T_1}\hat{X})$. A typical characteristic of interest was average establishment employment. The generalized formula for the overall firm mean estimate was

$$_{T_1}\hat{\bar{X}} = {}_{T_1}\hat{X}/N' \qquad (11)$$

where, $\hat{X}_{T_1}$ = composite total firm estimator (see Table 4),

$$N' = \sum_g {}_{(g)}N \qquad \text{where,}$$

g = 1,2; area and list frame establishments, and

${}_{(g)}N$ = total firms in frame source of interest.

The corresponding variance estimate for firm means was

$$v(\hat{\bar{X}}_{T_1}) = \frac{1}{(N')^2} v(\hat{X}_{T_1}) \tag{12}$$

$(N')^2 = (\sum_g {}_{(g)}N)^2$ = total firms in subpopulation.

g = source establishment frame of interest, and

$v(\hat{X}_{T_1})$ = composite firm variance estimate (see Table 4).

6.6 Adjusting for Nonresponse

To summarize the data it was necessary to deal with the problem of nonresponse. Some rules and guidelines are given as a preface to the actual adjustment of the expansion factors.

1. Nonresponse is defined to be a refusal, inaccessible, or nonusable questionnaire (completion code 2 or 3 on the questionnaire.) Screenout is defined as a sampling unit outside of the study area, a vacant household, or a firm out of business.
2. All nonresponse adjustments are at the stratum or substratum level.
3. A partially completed questionnaire cannot be summarized.
4. A household questionnaire is only complete if at least one member completed the interview.
5. A stratum must have at least one completed report to be summarized.

6. The number of firm nonresponses should be minimal in substrata 0 and 1 because of the replacement procedure to select alternate samples for refusals.

7. Questionnaires with completion Code 4 (screen out) are summarized as zero reports.

The general formula to adjust the expansion factor for nonresponse was

$$f_{ah} = \left\{\left(\frac{G}{G-B}\right) \cdot f_{oh}\right\} \qquad (13)$$

where, f_{ah} = adjusted expansion factor in stratum h

G = total number of units sampled in stratum/substratum

B = total number of nonresponse units in stratum/substratum

f_{oh} = original expansion factor.

Adjustments in the original expansion factor (f_{oh}) only occurred for a stratum or substratum if there was an instance of nonresponse. Determination of the expansion factor to be adjusted was dependent on the estimate computed. Table 5 below gives a summary of the original expansion factors for the unit level of interest.

Table 5--Expansion Factors by Data Source

Data Source	Original Exp. Factor (f_{oh})
1. List Firms: substratum 0	10
substratum 1	4
substratum 2	1
2. Area (NOL) Firms: stratum 2	1160/69
stratum 1	4568/183
stratum 4	3283/66
3. List Household*	40
4. Area (NOL) Household*	100

* A household was considered complete if one questionnaire was completed.

Guidelines were also given for handling missing items in a useable questionnaire. Certain items were imputed during a hand and machine edit using relationships from reported information. This could not be done in many cases, because information was not available. If a missing cell could not be completed during the hand edit a negative one was entered. All respondents in the same stratum/substratum with a valid entry for the item of concern were identified. An average of the good entries (zero or positive value) was calculated with the new value replacing the minus one. This process assumed the missing value was similar to the mean of those reporting in that class. The averaging process allowed the entire questionnaire to keep the same expansion factor rather than each item having a different rate. Firm questionnaires were all handled in this manner; however, household questionnaires frequently required the more burdensome task of adjusting the expansion factor for missing items or sections of the questionnaire. Households in the list and area strata were not similar enough to support imputing averages from complete reports for missing items. Consideration was initially given to a system of individual substitutions from a neighboring questionnaire for imputing the missing cell value. This technique, however, according to Kish (2) could lead to serious selection beases.

6.8 Confidence Limits

Assuming that the normal approximation applied, confidence limits for attribute X and ratio of attributes R were obtained. A sample size of at least 30 was generally considered sufficient to assume normality. The confidence limits for $_{T_k}\hat{X}$ and $\hat{R}$ were

$$X: \quad {}_{T_k}\hat{X} \pm Z \sqrt{v({}_{T_k}\hat{X})} \tag{14}$$

$$R: \quad \hat{R} \pm Z \sqrt{v(\hat{R})} \tag{15}$$

where, Z = the normal deviate corresponding to the chosen probability

= 1.96 for a 95 percent statement of confidence

= 1.65 for a 90 percent statement of confidence

Expected confidence limits for certain key variables were estimated from coefficients of variation (C.V.) computations using 1970 census data. This information was compiled in an earlier report and is shown in Table 6 to provide some measure of assessing the value of certain estimates. Doubling the C.V. value gave an approximate 95 percent confidence limit subject to the following assumptions:

1. Assumed a simple random sample of 2,000 households from a complete list.
2. 1970 Census population data in Kentucky was used to initially determine the standard errors.
3. The primary sampling unit was assumed to be a census household.
4. Standard errors and the corresponding C.V. assume <u>each</u> category an independent event.
5. Certain categories were noted as rare items.

Table 6--Sample C.V.'s and Confidence Limits for Certain Household Variables Using 1979 Census Data

Variable	Population Value	Population C.V. (20% Sample)	4 Percent Sample (n=2000) C.V.	95% Confidence Level
		%	%	
Population Age 16+ Total	118,925	0.34	0.76	(117117, 120733)
Employed	44,447	0.63	1.41	(43194, 45700)
Unemployed	3,150	3.49	7.81	(2658, 3642)
Out of Labor Force	71,328	0.50	1.13	(69716, 72940)
Population Aged 16-24	24,757	1.03	2.31	(23613, 25901)
Aged 25-44	36,812	.75	1.67	(35582, 38042)
Aged 45-65	36,124	.75	1.69	(34903, 37345)
Aged 65+	21,232	1.13	2.52	(20162, 22302)
Ave. Household Income	4,846	2.71	6.06	(4259, 5433)
Ave. Age of Residents	28.6	1.29	2.89	(26.9, 30.3)
Ave. Years Education of Res.	8.2	1.09	2.43	(7.8, 8.6)
Number Blacks	2,317	3.73	8.34	(1931, 2703)
Number Households Public Assistance	6,188	2.51	5.60	(5495, 6881)
Number Households Poverty	14,362	1.38	3.08	(13477, 15247)
Percentage Classific. of				
White Collar Workers	.09	2.81	6.29	(.08, .10)
Blue Collar Workers	.11	2.60	5.81	(.10, .12)
Service Workers 1/	.03	5.13	11.48	(.02, .04)
Farm Workers 1/	.02	7.70	17.22	(.01, .03)
Private Household Workers 1/	.004	15.13	33.82	(.001, .007)
Percentage Industry Classific. of				
Agriculture 1/	.03	5.13	16.23	(.02, .04)
Manufacturing	.05	4.30	13.60	(.04, .06)
Non-manufacturing	.13	2.33	7.37	(.11, .15)
Wholesale and Retail Trade	.05	3.74	11.83	(.04, .06)
Household Income Distribution				
0-4999	14,143	1.25	3.96	(13023, 15263)
5,000-9,999	6,196	2.28	7.20	(5304, 7088)
10,000-14,999	2,270	4.01	12.68	(1694, 2846)
15,000-24,999 1/	792	6.94	21.96	(444, 1140)
25,000+ 1/	227	12.33	39.00	(50, 404)

Reference for table true values is U.S. Census of Population: 1970, Kentucky PC(1)-C19 App-46, 47.

1/ Rare items

SECTION 7 - COST SUMMARY

7.1 Introduction

The cost estimates provide only a rough approximation of data collection costs of nearly $200,000 associated with the survey. This did not include administrative and overhead costs, out of pocket expenses, training of field staff, frame construction (list and area), and machine editing of the data which would bring the total close to the alloted $350,000 for the project.

The following assumptions were made in projecting data collection costs (excludes training and overhead):

1) An enumerator staff day costs $170.
2) An enumerator could average 2.5 completed interviews per day.
3) An enumerator could screen out 7.5 households (HH) per day.
4) An enumerator could list and identify households (HH) for 3 area frame segments per day.
5) An enumerator could interview 2 firms per day.
6) 70 percent of area households would overlap with firm list.
7) .2 additional staff days were needed to screen HH area segments for NOL firms.

Job	Number	STAFF days	Cost
Identify and List HH	318 Segments	106	$ 18,020
Screen out HH	550 Households	73	12,410
Interview Area HH	250 Households	100	17,000
Interivew Firms	458 Firms	229	38,930
Interview HH from List Firms	1475 Households	590	100,300
Screen for NOL Firms	318 Segments	64	10,880
	TOTAL	1162	$197,540

REFERENCES

(1) Cochran, William G. Sampling Techniques, third edition, New York: John Wiley & Sons, 1977.

(2) Kish, Leslie. Survey Sampling, New York: John Wiley & Sons, 1965.

(3) Vogel, Frederic A., Bosecker, Raymond R., and Rockwell, Dwight A., "Multiple Frame Livestock Surveys - An Evaluation of Alternative Methods of Overlap Determination," Statistical Research Division, ESS, USDA, June 1976.