

EVALUATION OF THE 1977 ECONOMIC CENSUSES

Barbara A. Bailar, U.S. Bureau of the Census
Shirley Kallek, U.S. Bureau of the Census

I. Introduction

The economic censuses comprise a body of individual censuses covering major sectors of the U.S. economy: manufacturing, mining, construction, and distributive trades and services. These are mandatory programs required by law and are conducted every 5 years for the years ending in "2" and "7."

Each individual census has a long history; the census of manufactures, for example, was started in 1810. The timing, periodicity, and content of each of the sectors varied considerably. Each was done independently and used different techniques and sources as the universe frame. Beginning in 1954, however, the various census programs were integrated into a comprehensive economic censuses program. Tax records became the source of the frame and operations common to all censuses, such as mail-out, check-in, clerical processing, and so forth, were implemented on a unified basis. This permitted economies of scale as well as the development of better techniques for the control and processing of the censuses.

Together, the economic censuses cover approximately 7 million establishments in the selected economic sectors, but only about 2.8 million of the larger establishments are sent a report form to fill out. Administrative record data from other Government agencies are used as proxy records for all but a sample of small single establishment firms. Nonemployer firms, for which data are obtained from administrative records, are included only in the censuses of retail trade and services. Each establishment is classified into a 4-digit standard industrial classification (SIC) industry and coded to a detailed geographic area.

The censuses are conducted on a mail-out, mail-back basis, and nonrespondents are followed up through the use of mail procedures. Telephone contacts are also used, primarily as a means to urge respondents to report, but are confined to the bigger companies. Relatively few report forms are completed through telephone followup procedures.

Each of the censuses has a set of common questions which are asked of all respondents classified in the sector. For example, in the census of manufactures there are 16 items covering such things as employment, payroll, inventories, and capital expenditures. In addition, there are specialized inquiries relating to the characteristics of the individual industries. These normally cover detailed information on the inputs and outputs of the specified industry.

The report forms are reviewed centrally for gross errors by a clerical unit prior to keying, then passed through a complex edit which examines internal relationships. Rejects are reviewed and corrected by analysts where necessary and the results are tabulated. Another extensive review is done at this stage; comparisons are made to

previous periods, and analytical ratios and outliers are examined.

The publication series for each census flows down two major paths: by industry and geographic area. Data are shown in the aggregate and cross-classified by various size classes, by legal form of organization, and by other economic variables. Since each establishment is identified separately and is classified by industry and geographic area, the potential for tabular presentations is great. Special care is taken to prevent disclosure of information for any individual respondent in the tables that are published.

The censuses provide a detailed picture of the economy at a point in time. Because the censuses are so detailed and so complete, it is possible to study changes in the structure of the economy over periods of time. They are used by the Government and the public as an analytical and marketing tool and are the basic foundation for many other statistical programs. They serve as the sampling frame for the Bureau's current programs, permit reconciliation of data reported in different surveys, and are a major input into the Standard Statistical Establishment List (SSEL). Important uses are made of the data by other Federal agencies in updating the national accounts, the producer price index, the index of industrial production and the like.

II. Role of Evaluation

Evaluation has three main purposes in censuses and surveys conducted by the Bureau of the Census. These are as follows:

- to insure that the data collected meet the needs for which the collection system was created;
- to provide relative magnitudes of error for the various potential sources of error so that data producers can focus on the areas where improvements will have the largest payoff;
- to help users in interpreting survey and census data by providing information on the limitations of the data.

Evaluation means a systematic investigation of the entire operation of the census including:

- the translation of objectives into clear operational terms;
- the use of sampling;
- the application of definitions;
- the mail-out and follow-up procedures;
- the editing routines;
- the effect of nonresponse and imputation;
- the interpretation of data.

After a review of the potential sources of error, the sources should be grouped into some structure, permitting potential sources of error to be identified as contributing to variance or bias terms in a mathematical model. Estimating the relative magnitudes of these error components then leads to reaching the goals of the evaluation process.

At the Bureau of the Census we have made use of a mean-square error model developed by Hansen, Hurwitz, and Bershad (5) as the basis for evaluation efforts in demographic censuses and surveys. This model is also applicable to economic censuses and surveys. In the demographic area, the model is explained as follows. Consider an area in which n units are assigned to each of k interviewers. This can be an area the size of a block, a tract, a city, or any other size. The size of the area will affect the value of k , the number of interviewers needed to conduct the census or survey. Then, the mean for a given characteristic for that area, based on the census or survey, is \bar{x} . The mean-square error of that mean can be expressed as:

$$MSE(\bar{x}) = \frac{\sigma_s^2}{kn} + \frac{\sigma_r^2}{kn} [1 + (n-1)\rho] + \frac{2(n-1)}{kn} \sigma_{rs} + B^2.$$

In this expression, σ_s^2 is the sampling variance. For items collected on a 100-percent basis, this term is zero. The term σ_r^2 represents the simple response or measurement variance, the basic variability in measuring the response from one time to the next. This component of variance is present in censuses as well as surveys. It is the expected value of the square of the response or measurement deviation, the difference between the recorded value for that unit on a particular trial of the procedure and the expected value for the unit over all trials. The measurement deviations are functions of the respondents, coders, keyers, editors, and analysts.

The term $\rho\sigma_r^2$ is the correlated component of response variance. It is the covariance among measurement deviations of units that are interviewed by the same person, coded by the same person, responded for by the same person, or in the assignment of some person who exerts a common influence over them. This term may be present in both censuses and surveys.

The term σ_{rs} is the covariance between measurement deviations and sampling deviations. It will be zero for items collected on a 100-percent basis.

Finally, B^2 is the square of the bias. The bias term is present in both censuses and surveys. It arises through respondents' misunderstanding the questions, through refusals, through misapplication of sampling procedures, and through other processes.

Though the model is equally appropriate to either demographic or economic censuses, the factors responsible for causing a large variance or bias vary. The simple response variance reflects the difference in response on one trial of the survey procedure from the average over all trials. In a demographic census, such differences can be caused by the respondent, the interviewer, or the interaction of the two. In an economic census, there is no interviewer who visits the establishment, though one may call on the telephone. Generally, the establishment sends in a report by mail. Thus, the difference in response may be caused by the respondent, the analyst, the edit clerk, or the interactions

among them. Just as one may get a different respondent in a demographic census on a repetition of the census, one may get a different respondent in an economic census. Different respondents often provide different answers.

A parameter that is very different in interpretation between demographic and economic censuses is that of the correlated component of response variance. In the decennial censuses we have measured the effect of interviewers, crew leaders, and coders. Interviewers and crew leaders are not part of the economic census activity. However, in the economic censuses, there are respondents, keyers, and analysts. A respondent will not contribute to the correlated component of response variance in single-unit companies or multi-unit companies in which each establishment has its own respondent. But there are multi-unit companies, some with hundreds of establishments, for which there is one company respondent. The tendency of these company respondents to overestimate or underestimate, their interpretations of various questions, their conversions of responses into codes, their methods of rounding data, their tendencies to estimate rather than look up book values--all of these enter into the variability that can be associated with a company respondent.

It is also the case in some large multi-unit companies that the headquarters unit will provide written instructions for each individual establishment to use in filling the questionnaire. To the extent that these instructions cause all the establishments to report in a way different from other establishments, this practice contributes to the correlated component of response variance.

What is the potential effect of such variability? It depends on both the magnitude of the correlation, ρ , and the number of establishments for which a company respondent reports. The distribution of companies by size is skewed. There are many small companies and fewer very large ones. Suppose a particular company respondent always estimated for a particular census item and the estimates were, in general, a little high. Suppose the value of ρ was .90. In the case of a multi-unit company with 11 establishments, the total variability of the statistics would be increased by a factor of 10. In the case of a multi-unit company with 200 establishments, the total variability would be increased by a factor of 180. Thus, these correlations can lead to large increases in variances. It is clear that this component of variance may be substantial given that a number of multi-unit companies have over 100 establishments.

There are several reasons this component of variance caused by company respondents is disturbing. First, it may be large and it is not measured in the usual sampling variance formulas. Second, it does not seem feasible to design an experiment to estimate this variance. In the decennial census, we estimate this variance by designing an interpenetrated study in which the work assignment of one interviewer is randomly divided among several interviewers. Since the interviewers are Bureau employees, we can redesign assignment areas without causing undue hardship. It would be difficult, if not

impossible, to do this with company respondents. We would have to ask the company to provide us with two or more respondents and then assign the establishment reports randomly among them. Most companies consider it enough of a chore to fill out the forms, much less participate in an experiment.

Another way a correlated component of response variance can be introduced into the data is through keyers. There are no keyers in the decennial census because the form is read by an optical sensing device, but this is not the case in the economic censuses. To the extent that all keyers do not carry out their keying in a uniform manner in the economic census processing, they contribute to this component of variance. This parameter could be estimated in an experiment if it was thought to be an important source of variability. However, its effect is probably not important since control features are aimed at keeping the keyers' error rate low. Trade associations may also influence the magnitude of the correlated component. They may send out recommendations to their constituents on how to report for certain items.

Finally, the analysts can introduce correlated errors into data. Analysts are referred individual establishment records after the original edit. These may be records for which large changes have been made in the original data, large imputations made, or for which extreme ratios of data items exist even after the edit. The analysts are responsible for determining the final disposition of the record. They may decide to call the respondents, they may decide to accept large changes, or they may decide to use the tabulations. They review the totals and make changes that seem reasonable to them. To the extent that an analysts has a "personal equation" of his or her own, there is a contribution to the correlated component.

The term σ_{rs} reflecting the covariance between measurement and sampling deviations is zero in data collected on a 100-percent basis. However, if some of the questions are asked only of a sample, this component may appear. It is difficult to measure this term individually and it is usually included in the estimation of the correlated component. Fellegi (3) attempted to estimate this parameter in the 1961 Canadian Census.

Finally, the bias term in the model may come from some of the same sources in economic as in decennial censuses, but there are additional sources of bias as well. One source of bias that is the same is the misunderstanding of a question by respondents. If a substantial number of respondents interpret a question in the same way, a positive or negative bias may result. Another source of bias is incomplete coverage of the universe. In demographic censuses, there are coverage problems caused by missing persons within households, missing households, and unrecognizable households. In economic censuses, there are coverage problems caused by not having a complete frame, excluding establishments from the sample incorrectly, and not finding establishments. Another common source of bias is nonresponse. In all censuses there are some units that refuse to report or are never available for obtaining a report. There are other units that

respond to some but not all questions. Item nonresponse can be a serious problem in any kind of census. Since item nonresponse usually leads to imputation, the imputation process itself can lead to bias. It may reduce the total bias of nonresponse, or it may exaggerate it.

Because of the desire to reduce respondent burden in the economic censuses, there is a reliance on administrative records that does not exist for population censuses. Smaller firms are not, in general, asked to fill out census forms. Instead data items are filled from administrative records for those firms. Administrative records are also used for imputing for nonrespondents. To the extent that the administrative records are not complete, up-to-date, or based on the same concepts, a bias will result.

An evaluation program has as its goal the estimation of some of these parameters in the model. It may not be feasible, useful or cost-effective to measure all of them. A choice must be made about which parameters to measure. Basically, the evaluation program for the 1977 economic censuses supported studies of biases. There were two studies to measure coverage bias; there was a study to measure the bias caused by the use of administrative records; there was a study of bias caused by the misapplication of sampling procedures; there was a study of bias caused by the coding of geographic area; and there was a study of the bias caused by respondents' interpretations of questions. A brief description of these studies follow. A more detailed report on some of them is in the session at this meeting called, "Studies in the Evaluation of the 1977 Economic Censuses."

III. Coverage Studies

There were two studies to measure the coverage of economic censuses. One was evaluation of the completeness of the administrative record frame. This evaluation concentrated on problems arising from Census processing and cases considered out-of-scope of the economic censuses for either activity or size. A sample was selected of the cases in the administrative record frame that were classified as out-of-scope of the censuses. The employer establishments were mailed a schedule which specifically asked about major activities so that a classification could take place. Non-employer cases were also mailed a classification card. The study of these two groups will give a measure of the amount of bias in the census statistics resulting from accepting the classification on the administrative record when, in fact, it may be incorrect. Paralleling this activity was a study of the coverage of activities that were originally classified as in-scope. A sample of these cases was selected and traced through the census processing operations.

A description of the results of the study on the out-of-scope employer cases is available in the paper, "Evaluation of the Coverage of the Administrative Records Frame for the 1977 Economic Censuses--Employer Segment" by Hanczaryk and Sullivan (4). The total misclassification of all records originally classified as out-of-scope was 3.1 percent for establishments, 0.4 percent for employment, and 0.3 percent for annual payroll. This is the amount left out of all

censuses. A report on the nonemployer cases and the coverage study of those originally classified as in-scope is not yet available.

The other coverage study was one to learn more about the coverage of individual proprietorships. Some persons who are self-employed may report their non-farm self-employment income as wages and salaries on their tax return rather than filing a Schedule C. Presumably this is more common with individuals whose self-employment income is a secondary source of income. Using the income supplement from the March 1978 Current Population Survey (CPS), which asked about 1977 income, any persons identified as having non-farm self-employment income in the CPS were sent a questionnaire requesting some information about the kind of business and their Employer Identification number if they had one. These cases were matched to the economic census returns. Any unmatched cases were potential missed units from the census universe. These cases are currently being matched to other files. A report on this study is not yet available.

These two studies will give an indication of the completeness of the census frame. Though it is anticipated that any missed units are probably smaller units in the sense of their contribution to total sales or value of product, it is important to know where there are gaps in the frame.

IV. Content Studies

A. Evaluation of reported cases information.

This study is a content analysis of the information reported by respondents to the

questions on employment, payroll, and total sales or value of product. The focus of this study is on the application of the definitions specified in the census instructions. The conceptual definitions incorporate a number of specifications which should be observed by respondents. For example, included in total value of shipments should be the value of interplant transfers and excluded should be the amount of freight charges. The instructions for many census items are long and do not appear on the census form, but in a separate instruction booklet.

A sample of about 4,000 establishments from those which reported were selected for the study. Table 1 shows the distribution of these cases among census areas. Interviewers contacted each of these establishments and asked a set of probing questions about what was included or excluded from the reported figures on employment, payroll, and total sales or value of product. For the single-unit and small multi-unit companies, field interviewers conducted the interviews. For the large multi-units, professional staff from Bureau headquarters conducted the interviews. Because of the expected large costs per case for the larger firms and because the headquarters staff was used, the samples were limited to the sizes shown in Table 1. There will be an over-all assessment at the national level only for the censuses shown. A preliminary report on this study is in the paper, "Content Evaluation of the Economic Censuses" by Corby, Farrell, Blum, and Clark (2).

TABLE 1. DISTRIBUTION OF NUMBER OF ESTABLISHMENTS IN CONTENT REINTERVIEW STUDY

<u>Census Area</u>	<u>Number of Establishments</u>		
	<u>Single-unit</u>	<u>Multi-unit</u>	<u>Total</u>
Retail Trade	1,156	191	1,347
Wholesale Trade	451	216	667
Selected Services	536	165	701
Construction	556	55	611
Manufacturing	360	315	675
Totals	3,059	942	4,001

The output of this study will be useful in many ways. First it provides a measure of respondent bias in the estimates. For items with which many respondents had difficulty, we may be able to improve the wording or the instructions. Second, the study will encourage particular problem areas to be given more attention. It may be true that one census area has little difficulty while another has great difficulty. This could lead to improved resource allocations. Finally, the study may shed some light on the existence of the correlated component of response variance caused by company

respondents or company instructions. If, in the multi-unit companies included in the study, there were some who interpreted the instructions one way, another set who interpreted them in a different way, and maybe some other groups who had still different interpretations, then that is a signal that the variance caused by company respondents exists. To the extent that the company respondents in different companies report for different numbers of establishments, their effect varies. We will not be able to quantify their effect, but we will have some evidence that the effect is real.

B. Evaluation of administrative record data.

The objective of this study was to evaluate the extent of differences between the census data as published, collected with the use of administrative records, and as it would have been if report forms were obtained from the entire census universe. Thus, the respondent reports are taken to be the operational standard. If it is found that administrative record data and respondent reports are measuring different concepts, we want to find improved ways of using administrative record data. This use of

administrative records would not be abandoned since they have great advantages in reducing respondent burden and census costs. The main items being compared are the Standard Industrial Classification (SIC), payroll, sales or receipts, and the number of employees.

The studies were conducted to provide information for the three areas in which administrative records are used--the below cutoff cases, the nonemployers, and the delinquents. The sample sizes for each type of case by census area are shown in Table 2.

TABLE 2. DISTRIBUTION OF NUMBER OF ESTABLISHMENTS IN STUDY OF USE OF ADMINISTRATIVE RECORDS

Census Area	Number of Establishments		
	Below Cutoff	Nonemployers	Delinquents
Retail Trade	57,600	3,989	1,144
Wholesale Trade	-	-	458
Services	45,700	6,059	1,069
Construction	-	2,610	1,486
Manufacturing and Minerals	3,450	-	-
Totals	106,750	12,658	4,157

The processing of the part of the study dealing with the below cutoff and nonemployer cases is complete and is described in the paper, "Evaluation of the Use of Administrative Record Data in the Economic Censuses" by King and Ricketts (7).

and the evaluation questionnaire was made in order to detect any apparent sampling problems on the part of the respondent or the interviewer. The total number of shipping documents, the "take every" sampling interval, the number of documents selected in the sample, the total given weight of shipments, and the reasons for differences between the original and evaluation questionnaires were recorded.

V. Studies of Procedures

A. Commodity Transportation Survey.

In the 1977 Commodity Transportation Survey (CTS), establishments were requested to sample their own files of shipping documents and transcribe information from each sampled document onto the 1977 survey questionnaires. This evaluation study of in-scope establishments in the 1977 CTS was conducted in order to (1) obtain a measurement of the effect sampling by respondents had on the CTS results, (2) categorize the types of sampling problems encountered by them, (3) determine biases in published 1977 CTS data, and (4) determine whether or not improvements should be made in future surveys of this kind. A sample of establishments that responded to the 1977 CTS (370 sampled establishments) were visited by Bureau interviewers to (1) attempt to obtain a subjective measure of respondents' understanding of an adherence to the sampling instructions and (2) select an independent sample of each establishments' shipping documents to determine any sampling problems and to measure the difference between the samples.

In the second phase of the evaluation study the estimated weight shipped from the two samples on an establishment-by-establishment basis was analyzed. For each establishment, an estimate of weight shipped was calculated for each commodity, mode of transportation, and month, as well as the total for each firm.

In the third phase, estimates of total weight shipped were made from the sample of establishments in the evaluation study, using the sample taken by the respondents in the 1977 survey and the sample used in the evaluation study. All errors that were correctable in the latter sample were corrected in order to make the "best" estimate. Gross keying errors detected during the edit processing of the 1977 CTS sample also were rectified. Estimates were made by mode of transportation, commodity, code, and month. The results of this study were reported in a paper, "Can U.S. Businesses Take a Sample of Their Own Records?", by Tupek and Perez (9).

B. Geographic coding.

The analysis of the data was divided into three phases. In the first, a comparison of the questionnaires completed in the 1977 CTS

The purpose of this study was to assess the accuracy of the geographic location coding of

establishments included in the Economic Censuses. A sample of cases was selected from establishments included in all economic census areas. There was both a clerical evaluation phase and a field evaluation phase. The clerical evaluation phase was designed to determine the adequacy of the computerized coding operation using all available reference materials except the files used in computer coding. The field evaluation phase was to determine the quality of the physical-location addresses used for geographic coding and the codes assigned based on those addresses. A report on this study is in the paper, "An Evaluation of the Geographic Coding in the 1977 Economic Censuses: An Overview", by Judge (8). Early results indicate that for those addresses for which geocodes could be determined clerically or by telephone, the correct geocodes were assigned to approximately 98 percent of the addresses. This is a gain over the 1972 experience.

C. Current survey reconciliation.

This evaluation activity was concerned with comparisons of data from the economic censuses with current survey estimates for the same data items. The paper, "Reconciliation of Economic Censuses Results and Current Survey Programs" by Helfand and Bernhardt (6) gives results of this reconciliation for manufacturing, retail trade, wholesale trade, and service industries. For the censuses of retail trade, wholesale trade, and service industries, reconciliation was limited to 3,500 companies representing the "certainty" cases in the current samples. Further, only very large differences were reconciled. Therefore, no probability statements can be made about comparability of estimates for the total U.S. from the two sources. However, the reconciliation was useful since it identified a number of problem areas and corrections were made to the data before publication both in the censuses and the current surveys.

Many of the firms reporting large differences were not covering the same establishments in the census and current surveys, or were not reporting as instructed. Some inconsistencies between the census and current surveys included the assigned "type-of-operation" classification in wholesale trade, reporting of sales for departments and concessions, treatment of franchise operations, and coverage of taxes. Reporting differences resulted from different people completing the questionnaires, from dissimilar instructions, and from timing differences. Requirements for early reporting in the monthly surveys often result in estimates, whereas the data for the annual surveys and censuses are usually based on accounting records.

In the manufacturing area, data from about 70 current surveys in the Current Industrial Reports (CIR) Series were compared with data from the Census of Manufactures. Over 750

product groups and approximately 30,000 reporting units were involved in the comparisons. Based on adjustments, or corrections made during this review, the current survey and census results were reconciled to pre-established tolerances. These tolerances ranged from 2 to 10 percent.

D. Comparison of new construction and census of construction industries reconciliation.

This study was also a reconciliation study between a current survey of construction activity and the census of construction industries. Details are given in the paper, "Comparison of the 1977 Census of Construction Industries and the Value of New Construction Put in Place Series" by Blum, Roff, Visnansky, and Fondelier (1). After adjustments for differences in concepts and coverage, the census value of construction put in place was 92 percent of the current survey value overall, but only 64 percent for single family houses. Some of these differences were caused by differences in coverage and classification by construction category.

VI. Conclusions

The studies undertaken give indications of the magnitude of the bias term in the mean-square-error model. All of the studies related to biases--of the frame, of the data-collection, of the geographic coding, or of the concepts. In addition, there are variances, not included in this set of evaluation studies. There are also processing errors leading to biases and variances that were not evaluated.

A careful evaluation of all phases of the census operations is necessary. In the 1982 evaluation, more attention will be focused on other areas in addition to studies of biases.

REFERENCES

- (1) Blum, A., Roff, G., Visnansky, A.W., and Fondelier, D., "Comparison of the 1977 Census of Construction Industries and the Value of New Construction Put in Place Series", paper presented at the 1980 American Statistical Association meeting in Houston.
- (2) Corby, C., Farrell, M., Blum, A., and Clark, D., "Content Evaluation of the Economic Censuses", paper presented at the 1980 American Statistical Association meeting in Houston.
- (3) Fellegi, I.P., "Response Variance and Its Estimation", Journal of the American Statistical Association, Vol. 59, 1964, pp 1016-1041.
- (4) Hanczaryk, P. and Sullivan, J., "Evaluation of Coverage of the Administrative Records Frame for the 1977 Economic Censuses--Employer Segment", paper presented at the 1980 American Statistical Association meeting in Houston.

- (5) Hansen, M.H., Hurwitz, W.N., and Bershad, M.A., "Measurement Errors in Censuses and Surveys", Bulletin of the International Statistical Institute, Vol. 38, Part 2, Tokyo, 1961, pp 359-374.
- (6) Helfand, S. and Bernhardt, M., "Reconciliation of Economic Censuses Results and Current Survey Programs", paper presented at the 1980 American Statistical Association meeting in Houston.
- (7) King, C.S. and Ricketts, E., "Evaluation of the Use of Administrative Record Data for Nonmail Cases in the Economic Censuses", paper presented at the 1980 American Statistical Association meeting in Houston.
- (8) Judge, J., "An Evaluation of the Geographic Coding in the 1977 Economic Censuses: An Overview", paper presented at the 1980 American Statistical Association meeting in Houston.
- (9) Tupek, A., and Perez, A., "Can U.S. Businesses Take A Sample of Their Own Records?", paper presented at the 1979 American Statistical Association meeting in Washington, D.C.