AN EVALUATION OF SAMPLE DESIGN CHANGES FOR THE 1997 COMMODITY FLOW SURVEY

Jock R. Black, William C. Davie Jr., and Jacklyn R. Jonas, U.S. Bureau of the Census
William C. Davie Jr., U. S. Bureau of the Census, SSSD Rm. 2754-3, Washington, DC 20233-6500

KEY WORDS: commodity flows, respondent-conducted sampling, sample design, respondent burden

1. Introduction

The U.S. Census Bureau and the Bureau of Transportation Statistics (BTS) conduct the Commodity Flow Survey (CFS) to measure important characteristics of shipments originating in the United States. Because the population of shipments is extremely large and variable, and shipping records are kept in a geographically wide-ranging universe of establishments, the sample design used for the CFS is necessarily complex. It is a three-stage design with the respondent conducting the last stage of sampling.

Based on experience gained from the 1993 CFS, we made several changes to the design and the questionnaire for the 1997 CFS. Research indicated that the focus of these changes should be on trying to improve timeliness, reduce respondent burden, and to lessen the influence of large and infrequent shipments (Black, 1997). The changes we made for the 1997 CFS, along with the reasons for the changes, are discussed in Section 3. In Section 4, we attempt to assess the overall effectiveness of these changes. Where possible, we make comparisons between the 1993 and 1997 surveys. Section 5 summarizes our findings and suggests additional areas for evaluation.

We begin by providing some general background on the CFS and an overview of the CFS design.

2. General Design of the CFS

The CFS is a mandatory survey conducted periodically as part of the Economic Census. The CFS covers shipments originating from a universe of approximately 800,000 employer establishments classified in mining, manufacturing, wholesale, or selected retail (catalog and mail order) industries, as defined by the 1987 Standard Industrial Classification (SIC) Manual (Ofc. of Mgmt. & Budget, 1987). Auxiliary establishments (e.g., warehouses) of multiple-establishment firms are also represented in the CFS; *auxiliary establishments* are establishments that are primarily involved in rendering support services for other establishments within the same multiple-establishment company.

The CFS is designed to produce origin-destination estimates of total value, tons, ton-miles shipped, and average miles traveled per shipment by commodity and mode of transportation for various levels of geography. Estimates of change between 1993 and 1997 are also produced.

Each establishment selected for the CFS is mailed a questionnaire for each of four reporting periods - one in each quarter of the survey year. For each of these periods, we ask the respondent to count the total number of shipments made by the establishment in the reporting week, find a selection rate based on that total, and, depending on the total number of shipments, report information for all shipments made in the reporting week or for a systematically selected sample of those shipments. We request the following information about each of the selected shipments: domestic destination (city, state, and ZIP code) or port of exit (for exports), commodity, value, weight, mode(s) of transportation, the date on which the shipment was made, and an indication of whether the shipment was an export, hazardous material, or containerized. For exports, we also ask for the mode of export and the foreign destination city and country. For hazardous materials, we request the 4-digit United Nations/North American code.

3. Sample Design Changes for the 1997 CFS

Based on experience gained from the 1993 CFS, we made several changes to the sample design for the 1997 CFS. Research indicated that the focus of these changes should be on trying to *improve timeliness*, *reduce respondent burden*, and *lessen the influence of large and infrequent shipments* (Black, 1997). The following paragraphs describe the CFS sample design, the changes made for the 1997 CFS, the reasons for the changes, and the potential benefits of each.

This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. We thank Julia Bienias, Patrick J. Cantwell, Ruth Detlefsen, David L. Kinyon, Dennis Schwanz, and Richard Sigman for their contributions and helpful comments.

3.1 Description of the 1993 and 1997 Sample Designs

The 1993 and 1997 surveys employed similar sample designs. Both were stratified three-stage designs in which the first-stage sampling units were establishments, the second-stage sampling units were one-week reporting periods (2-week periods were used for the 1993 survey) within the survey year, and the third-stage sampling units were shipments. The *primary strata* were defined by industry, geography, and type of operation (auxiliary or non-auxiliary). We based the industry stratification on groupings of related 3-digit SIC codes, using the SIC code as a proxy for commodity. The geographic stratification was based on the National Transportation Analysis Regions (NTARs). The NTARs, developed by the Department of Transportation as combinations of Bureau of Economic Analysis (BEA) Economic Areas, form a complete partition of the United States. We stratified by type of operation to account for differences in the shipping practices of auxiliary and non-auxiliary establishments. In the following sub-sections, we describe the specific changes made for the 1997 CFS.

3.2 Smaller Sample Size

Perhaps the most significant change between the 1993 and 1997 CFS samples was a decrease in the size of the sample from nearly 200,000 establishments in 1993 to just over 100,000 establishments in 1997. The reason for the decrease was to allow for intensive follow-up of problem reporters early in the survey. The 1993 sample was so large that the data were not completely keyed until mid-1994, decreasing the effectiveness of our follow-up efforts. Our thinking was that halving the sample should allow us to identify problems nearer to when they occur, resulting in higher quality, more accurate data.

3.3 Improved Sampling Efficiency

A large reduction in sample size could result in a large increase in the sampling variability of the estimates. To address this concern, we made two major changes to the CFS sample design to improve its efficiency. First, we changed the measure of size used for sampling. Instead of using an employment or payroll-based measure, as done in 1993, we used a measure of size that approximates an establishment's *annual total value of shipments*. Because this variable is better correlated with one of the primary variables of interest (value of shipments), we were able to more efficiently stratify the establishments represented on the sampling frame.

Second, we changed the stratification and sample selection procedures. We were able to do this because the primary sponsor (BTS) did not desire estimates for *every* commodity in *every* NTAR. This meant we could group establishments by combinations of 3-digit SICs and, in effect, not constrain estimates for commodities that were insignificant contributors to an NTAR's total value of shipments.

To accomplish this, we selected with certainty the establishments with the greatest measures of size from all establishments having the same 3-digit SIC (see Lavallee & Hidiroglou, 1988). Then we stratified the remaining establishments (noncertainty establishments) by SIC recode and NTAR, where an SIC recode was constructed from a group of related 3-digit SIC codes. Because there was less variability among auxiliary establishments across different 3-digit SIC industry groups, we re-grouped the auxiliary establishments so the primary strata were defined by trade area (mining, manufacturing, wholesale, and retail) and a flag used to differentiate between the three types of auxiliary establishments. This resulted in approximately *3,400 primary strata*.

To further increase the efficiency of the design, we stratified the noncertainty establishments within each primary stratum (sub-stratified) using the measure of size previously described. We then used Neyman allocation (Cochran, 1977) to determine the sample size required within each substratum to meet a coefficient of variation constraint on the primary stratum total measure of size. Within each substratum, a simple random sample of establishments was selected without replacement.

In contrast, within each of the nearly *18,000 primary strata* used for the 1993 CFS, we identified certainty establishments based on *employment size* and selected a sample of noncertainty establishments using a *probability-proportional-to-size (pps)* procedure, where the measure of size was an estimate of *annual payroll*. (See Smith et al. (1994) for additional details of the 1993 CFS design.)

3.4 Shorter Report Period

To lessen respondent burden, we shortened the reporting period from two consecutive weeks to one week. We hypothesized that a shorter reporting period would reduce the burden on the respondent because the task of constructing a sampling frame of shipping documents would be slightly easier - fewer shipments are made in one week and the information about these shipments should be more accessible. This, in turn, should increase the response rate and possibly the compliance with the sampling instructions.

3.5 Maximum First Stage Sampling Weight

Another change implemented with the 1997 CFS attempted to reduce the large effect that one establishment or a single shipment could have on the estimates, especially at finer publication levels. We selected the 1997 CFS sample so the maximum first stage sampling weight would be no greater than 100 (assuming 100% response). In the 1993 CFS, the maximum first stage sampling weight was nearly 3,200 and a total of 635 establishments (of nearly 200,000 establishments) were selected with a first stage sampling weight greater than 100.

3.6 Identification of Certainty Shipments

To further limit the effect a single shipment could have on the estimates, and to improve the representativeness of large and infrequent shipments in the sample, we implemented a procedure in which survey analysts could identify particular shipments as "certainty" shipments.

To aid the analysts in identifying these types of shipments, the 1997 CFS questionnaire asked, "In the last three months did this location have any individual shipments with a value more than $2,000,000?" If the respondent answered "Yes," the reported shipments were checked to see if they adequately represented large shipments. If they did, no action was taken. (Survey analysts performed this check by visually inspecting the reported shipments and by comparing the establishment's sampling measure of size to its annual value-of-shipments estimate computed from the reported shipments.) If they did not, our survey analysts contacted the respondent and asked for additional information about the large shipments made by the establishment. If data were obtained for *all* of these large shipments for the reporting week or for the entire quarter in which the reporting week fell, we identified these shipments as "certainty" shipments. That is, each of these shipments was given a third stage sampling weight of 1. We statistically adjusted the sampling weights of the remaining shipments in the establishment's reported weekly sample. If the respondent could not provide data for *all* of the large and infrequent shipments made by the establishment for the requested time period, no action was taken.

3.7 Additional Changes

In addition to these sample design changes, we made several changes to the questionnaire. We improved the sampling instructions by supplementing a written example of the sampling procedure with a diagram. We also illustrated how information about each sampled shipment should be transcribed onto the questionnaire by providing an example of a multiple-mode shipment as well as an export shipment of hazardous materials. These changes were designed to make the respondent's task of sampling shipments easier. Finally, we added a comprehensive, automated editing system that allowed our data collection staff and survey analysts to quickly identify and correct data for problem reporters.

4. Evaluating the Effectiveness of the Changes

In the remainder of this paper, we attempt to evaluate the effectiveness of the changes described in Section 3. Because of budget constraints and the additional burden that would have been incurred by respondents, we did not incorporate into the 1997 CFS any experiments designed to measure the effect of each of these changes. Instead, we used other measures, some statistical, some non-statistical, to assess the effectiveness of the changes we made.

4.1 Improve Timeliness

As previously noted, we wanted to improve the timeliness of the survey. Among the changes made to accomplish this goal were a 50% reduction in sample size and an automated editing system. One way to evaluate the effectiveness of these changes is to compare the publication release dates for 1993 and 1997. In Table 1, we present this comparison in terms of the number of months since the end of the survey year.

Table 1. Comparing Publication Release Dates for 1993 and 1997

	Number of Months Since Survey Year		
	1993	1997	Diff
Preliminary	**18**	**12**	**6**
Final			
National	**36**	**24**	**12**
Subnational			
Region	---	28	---
Division	---	28	---
State	**26**	**24**	**2**
MA/ROS	---	26	---
NTAR	39	---	---
Special Reports			
HAZMAT	---	24	---
Export	---	28	---
CD-ROM	**36**	**30**	**6**

The last column gives the difference between the two years. A positive difference indicates the 1997 CFS publication was released more quickly. The rows in boldface type identify publications that were produced

in both years. A dashed line means the publication wasn't produced for the survey year indicated; therefore, no comparison was made.

Table 1 clearly indicates that we improved the timeliness of the public release of the survey estimates. Additionally, we published many more estimates from the 1997 survey.

As previously mentioned, one of the major changes contributing to a timelier release of the 1997 CFS results was a large reduction in sample size. We were concerned that this change would result in less precise estimates. In Table 2, we present the estimated coefficients of variation (cv's) for the national value-of-shipments estimates for 1993 and 1997. We also present the average, median, and standard deviation of the cv's of the national value-of-shipments estimates for commodity and state-level estimates. Based on these comparisons, we notice only a small increase in the variability of the 1997 estimates. Therefore, we have improved the timeliness of the release of the survey results without sacrificing the quality of the estimates.

Table 2. Estimated Coefficients of Variation[1] for Value of Shipments Estimates: 1993 and 1997

	1993	1997
National	0.8	1.0
National by Commodity		
Average	5.7	6.1
Median	5.2	4.6
Std. Dev.	3.1	3.7
State		
Average	6.7	6.7
Median	5.7	5.3
Std. Dev.	4.1	3.9

[1]Expressed as percentages. Variance estimates for both years were computed using the method of random groups (Wolter, 1985).

4.2 Reduce Respondent Burden

The second goal we had for the 1997 CFS was to reduce respondent burden. Clearly, reducing the size of the sample reduced the overall burden on the business community. But we were also interested in reducing the burden imposed on respondents by asking them to construct sampling frames and select samples. We hypothesized that one way to reduce this burden would be to shorten the report period from 2 weeks to 1 week. Assuming an establishment makes similar numbers and kinds of shipments from one week to the next, shortening the report period would reduce the burden on the respondent simply by reducing the number of shipping documents that would have to be gathered to construct the sampling frame.

4.2.1 Analysis of Total Number of Shipments

To investigate this hypothesis, we analyzed establishments selected and tabulated (passed edits required for tabulation) for both the 1993 and 1997 CFS samples. For these establishments, we compared the difference in the total number of shipments made during the 2-week report period in 1993 to the total number of shipments made during the 1-week report period in 1997, as reported by the respondent. Table 3 shows the 25th, 50th, and 75th percentiles of the reported total number of shipments for each survey year. The third column gives the ratio of the 1997 total number of shipments for the percentile to the 1993 total number of shipments for the percentile. By examining the table, we can see that our hypothesis was correct - by shortening the report period from 2 weeks to 1 week, we have reduced, by almost half, the number of shipping documents the respondent must gather to construct the sampling frame.

Table 3. Comparison of the Total Number of Shipments Made: 1993 and 1997

Percentile (%)	1993	1997	Ratio
75	989	575	0.58
50	324	190	0.59
25	104	62	0.60
mean	3,173	866	
std. dev.	81,475	6,191	

n = 20,369 quarters

So, by shortening the report period, we have reduced respondent burden. However, did this change affect the "representativeness" of an establishment's shipment sample? To attempt to answer this question, we compared *each establishment's* value-of-shipments estimate to a proxy value-of-shipments total for the establishment derived from 1997 Economic Census data. This comparison is described in the next section.

4.2.2 Census Adjustment Ratio

Most of the establishments selected for the 1997 CFS also reported in the 1997 Economic Census. For each establishment reporting in the 1997 CFS, we computed an estimated total value of shipments (TVS) from the sample of shipments reported by the establishment. We compared this estimate with a proxy for TVS obtained from the census. Because respondents were not requested to report total value of shipments in the Economic Census, we derived the census TVS from a single data item or combination of data items collected

on the census questionnaire. (For example, for mining and manufacturing establishments, we used the sum of "value of products" and "value of resales" as the value-of-shipments proxy.) Therefore, we consider the census TVS a proxy for the CFS TVS. Similar comparisons were made for establishments responding to the 1993 CFS using data obtained from the 1992 Economic Census, adjusted to a 1993 level.

Because the census TVS is not based on a sample of shipments (it is a single data item or combination of data items reported for the entire reference year), it has no sampling variability associated with it. Therefore, we consider the census TVS a more reliable measure of an establishment's annual total value of shipments.

For both the 1993 and 1997 surveys, we used the ratio of the census TVS to the CFS TVS to adjust the CFS estimates. We refer to these ratios as census adjustment ratios (CARs). CARs which differ from 1 by a large amount could indicate problems in the sampling of shipments by the respondent (e.g., purposive sampling, respondent constructed an incomplete frame of shipments from which to sample, more than one person completed the form, etc.). (See Evans and Cantwell (1995) for a detailed description of the census adjustment methods used for the 1993 CFS.)

Table 4 contains a comparison of census adjustment ratios for establishments tabulated for both the 1993 and 1997 surveys. We limited the comparisons to non-auxiliary establishments because we could not identify a reliable census proxy for total value of shipments for auxiliary establishments. Based on this analysis, we don't appear to have affected the "representativeness" of the sample in terms of shipment value. However, this does not address the "representativeness" of other shipment characteristics such as shipment weight, mode of transportation, or commodity.

Table 4. Comparison of Census Adjustment Ratios for Establishments Tabulated in 1993 and 1997

Percentile (%)	1993	1997
75	1.33	1.41
50	0.95	1.00
25	0.69	0.73

n = 28,718 establishments

4.3 Reduce the Influence of Large Shipments

The third change made to the sample design for the 1997 CFS was to reduce the influence of large and infrequent shipments. In the following sections, we examine the expected and observed results of our procedure (described in Section 3.6) for reducing the influence of these shipments. We begin with a discussion of our simulation study.

4.3.1 Simulation Study

To conduct the simulation study, we used actual shipment data reported in the 1997 CFS. We combined all shipments from all quarters reported by one establishment and used these as the total number of shipments for a hypothetical reporting week. The establishment used for the simulation study provided information for a total of 76 shipments, three of which were identified as certainty shipments. We chose this particular establishment because it reported a small number of shipments in each of the four quarters, thus reducing the time it took to run the simulation program. Furthermore, analysts had identified three of the shipments as certainty shipments.

Using these 76 shipments as the total number of shipments for our reporting week, the sampling instructions direct us to select a sample of 38 shipments from these 76. For the simulation study, we randomly selected 1,000 samples of size 38 from the shipments in our hypothetical reporting week.

From each sample, we computed two value-of-shipments estimates. The first estimate took the certainty shipments into account. If at least one certainty shipment was selected, then the remaining certainty shipments were added to the sample and a value-of-shipments estimate was computed. If a certainty shipment was not selected, and *assuming we did not obtain any additional information about the large and infrequent shipments made by the establishment*, the shipments were all considered to be noncertainty shipments in the computation of the estimate. To compute the second estimate, we treated all shipments in each of the 1,000 samples as noncertainty shipments. So, no certainty shipments were added.

For each group of 1,000 estimates, we compared the empirical distribution of the estimates to the actual value for the week, which was $1,072,796. Taking certainty shipments into account, the estimates of the 1,000 samples had a mean of $974,737 and a standard deviation of $272,465. Treating all sampled shipments as noncertainty shipments, the estimates of the 1,000 samples had a mean of $1,077,027 and a standard deviation of $685,456. While the first method is slightly biased downward, the second method is unbiased. However, the first method produces far less variable estimates and a smaller estimated mean square error.

4.3.2 Empirical Results

After completion of the data collection, edit, and review processes, our survey analysts had identified approximately 5,100 certainty shipments across 1,267 establishments. Considering only the 1,164 non-auxiliary establishments having certainty shipments, we computed two separate value-of-shipments estimates for each establishment. (As previously noted, our analysis was limited to non-auxiliary establishments because we could not identify a suitable census proxy of an auxiliary establishment's total value of shipments.) We computed the first of these estimates by treating the certainty shipments as designed. This estimate is equivalent to the TVS estimate described in Section 4.2.2. The second total value-of-shipments estimate (TVS2) was computed as if we had no additional information about the large and infrequent shipments. Any shipments selected in the reported sample and later identified as certainty shipments were treated as noncertainty shipments (i.e., they received the same sampling weight as the other shipments selected in the sample for that reporting week). Furthermore, those shipments added as certainties after contacting the establishment were ignored because we would not have had a procedure to account for them.

The estimated standard deviation of the 1,164 TVS estimates was $1.5 million, while the estimated standard deviation for the TVS2 estimates was $3.4 million for TVS2 - nearly a 45% reduction.

To examine the bias introduced by this "certainty shipment" procedure, we computed two census adjustment ratios for each of the 1,164 establishments. We used TVS as the denominator of one of the ratios and TVS2 as the denominator of the other ratio. Call these ratios CAR and CAR2, respectively. For each establishment having one or more certainty shipments, we found about 76% of the time CAR was closer to 1.00 than CAR2. In addition, the medians for CAR and CAR2 were 0.56 and 0.33, respectively. Based on this analysis, our procedure for reducing the influence of large and infrequent shipments successfully reduced the variability of the value-of-shipments estimates and seems to have biased the estimates for establishments making large shipments in the proper direction.

5. Summary and Future Work

As mentioned in the introduction, the CFS is a complex survey in which the respondent conducts the final stage of sampling. Based on experience gained from the 1993 CFS, we made several changes to the sample design used for the 1997 CFS. These changes were made to improve timeliness, reduce respondent burden, and lessen the influence of large and infrequent shipments. From the results presented in this paper, it appears the changes worked well. However, many areas of evaluation remain.

Preliminary results from an evaluation of the automated editing system identified several edits that could be improved. The recommended improvements were made to these edits and tested using the 1997 CFS data. Those improvements that worked well will be implemented in future CFS surveys. We have produced numerous measures of response and plan to examine how these measures change by industry and relative size of the establishment. This will allow us to focus our data collection efforts on problem industries. We also plan to conduct an evaluation of the completeness and accuracy of the sampling frame. This would include an investigation of the accuracy of the sampling measure of size and the industry codes assigned to establishments represented on the sampling frame. Finally, we have started to examine the relationship between survey response quality and the occupational class of the respondent. For the 2002 CFS data collection, we hope to use the results of this investigation to target those respondents within the establishment who are able to provide the most accurate and reliable information.

References

Black, Jock. (1997) "Can Respondents Construct a Frame and Draw a Sample? Experiences from the 1993 Commodity Flow Survey." *American Statistical Association 1997 Proceedings of the Survey Research Methods Section*. Alexandria, VA: American Statistical Association, pp. 216-221.

Cochran, W.G. (1977) *Sampling Techniques*, 3rd edition. New York: Wiley, pp. 98-99.

Evans, B. Timothy & Patrick J Cantwell. (1995) "Adjusting the 1993 Commodity Flow Survey to the 1992 Economic Census." *American Statistical Association 1995 Proceedings of the Survey Research Methods Section*. Alexandria, VA: American Statistical Association, pp. 216-221.

Lavalee, P. & M.A. Hidiroglou. (1988) "On the Stratification of Skewed Populations." *Survey Methodology*, 14, pp. 33-43.

Office of Management and Budget. (1987) *Standard Industrial Classification Manual*. Available from National Technical Information Service, Springfield, VA (Order no. PB 87-100012).

Smith, Robert T., Evans, B. Timothy, & Fowler, John L. (1994) "The Commodity Flow Survey." *American Statistical Association 1994 Proceedings of the Business and Economic Statistics Section*. Alexandria, VA: American Statistical Association, pp. 89-94.

Wolter, Kirk M. (1985) *Introduction to Variance Estimation*. New York: Springer-Verlag.