CAN RESPONDENTS CONSTRUCT A FRAME AND DRAW A SAMPLE? EXPERIENCES FROM THE 1993 COMMODITY FLOW SURVEY

Jock Black, Services Division, Bureau of the Census, Washington, D.C. 20233

The responsibility for frame construction and sample selection usually belongs to the professional statistician. However, in a large multistage cluster design that requires sampling in all stages, constructing the frames and drawing the samples for all stages may be impractical or even impossible for the statistician. In the 1993 Commodity Flow Survey (CFS), the Census Bureau asked respondents to construct a frame of the shipments made by their establishments and to select a systematic sample from it. The decision to delegate these statistical activities to respondents was the result of the relationship between the goals of the survey and the characteristics of the survey universe. To accomplish the respondent sampling, we set up a uniform sampling plan for the respondents.

This paper discusses the circumstances that led to the unconventional survey design, explains the sample design we imposed on respondents, presents some statistics we used to gauge the respondents' accuracy, and makes recommendations for future applications.

1. Survey Design

1.1. Survey Goals

The CFS was funded for a variety of reasons related to the measurement, regulation, and appropriation of funds for the improvement of the transportation infrastructure and sector of the economy. Among the survey's goals were to estimate characteristics associated with the ultimate origin and destination of shipments; the distances traveled by shipments of goods; the commodities shipped; the modes of transportation used to transport shipments; and the volume of shipments measured by weight and value of shipments. The sponsor's ideal publication would be a four-way table (commodity by mode by origin by destination) quantifying the total value, tonnage, and ton-miles of shipments and the average miles traveled by shipments. Besides the unusually large number of estimates required to build this table, estimates of other characteristics of interest were planned, some of which were still being negotiated and refined when the design decisions had to be made.

These goals greatly influenced the survey design because they focus on characteristics of *shipments*, not establishments. This requires the survey to collect either aggregate shipment data or data about individual shipments that must be aggregated or both. In any case, the data targeted for collection was the cluster of shipments made by individual establishments.

1.2. Anticipated Problems with Aggregate Data

With this realization came the question of whether aggregate data could be collected. To collect aggregate data and still produce the variety of estimates needed by the sponsor would require either of two situations. The easiest for us and the most burdensome for the respondent would be to ask the respondent to report the value and weight of shipments for each cell in the four-way table of estimates we wanted. Not only would this be extremely burdensome but, we also believed, the respondents' ability to provide accurate responses for (occasionally) several hundred cells was unknown and unlikely. Unlikely because, for many potential respondents, there is no or little economic incentive to keep track of the information we were seeking. While it is a common business practice to record the value and commodity involved in individual transactions, and less common to record the weight and mode for individual transactions (because a single shipment may involve several transactions), it is *not common* for businesses to produce summary data of the type we wanted.

An alternative would be to ask the respondent to report the marginal totals for each dimension of the four-way total. That is, we would ask the respondent to report the grand total value and grand total weight of all shipments and the percentage of the total weight and value of shipments for (1) each mode the establishment used to ship its commodities, (2) each commodity it produced, (3) each location from which shipments originated, and (4) each ultimate destination of shipments. (Alternatively, we considered asking for the absolute total value and weight of shipments by each category but this would inevitably lead to the problem of the total value over all modes not equaling the total value over all commodities.) To produce the individual cell estimates we wanted, we would assume some relationship between the marginal percentages reported by the respondents and the cell totals needed for the survey.

For instance, we could estimate the percentage of shipments of corn that traveled by inland water for a particular establishment as the product of corn's percentage (of commodity shipments) and inland water's percentage (of the shipments transported by the various modes). This would be appropriate if the commodity shipped and the mode used to ship it were independent. Unfortunately, our intuition and consultation with transportation professionals suggested that complex and variable relationships exist between mode of transportation, commodity, and destination. These are based on not only those three variables but also external factors such as weather, season,

and transportation provider.

Beyond these problems, a 1986 evaluation of respondents' estimates given in a pretest for the 1987 Commodity Transportation Survey (CTS) revealed large differences between the respondents reported distributions (of mode and state of destination for all shipments made in 1985) and the actual distributions computed by Census enumerators from the respondents' records of shipments. The same evaluation concluded that respondents could not reliably estimate the total annual poundage of shipments.

1.3. Anticipated Problems in Collecting Shipment Data

Gathering individual shipment information was not seen as a panacea. The first-stage universe of the Commodity Flow Survey included about 800,000 establishments. Preliminary sample size research showed that we would need to include about 200,000 establishments to produce the desired estimates. With that many establishments, expecting Census employees to visit them all and draw samples was impossible. In addition, Census enumerators would require many weeks of training to be able to handle all of the special situations they might encounter.

One way to solve this problem would be to pay for more enumerators. To be able to complete the survey quickly and before the selected companies destroyed the shipment records, we would need several thousand temporary enumerators. The logistical problems of hiring and training many temporary employees, concerns about the quality of any data collected, and the establishments' potential adverse reactions to government employees leafing through internal company documents caused us quickly to reject this approach. Another alternative was to dramatically reduce the scope of the survey so that we could use as few as 5,000 establishments. However, it was believed this would have made the survey not worth doing.

This left us in the position of requiring respondents to provide information on their shipments.

1.4. Problems with a Shipment Census

One alternative that avoids the problems associated with collecting aggregate shipment data is to require the respondents to report all of their shipments. This approach has its problems.

Getting all shipments for a year is unrealistic. The data would be too burdensome both for the respondents or enumerators to provide and for the Census Bureau to process. So, we would have at most a sample of shipments.

In fact, the number of shipments we could expect for any period varied widely. Mail order department stores might have millions of shipments a year and some manufactures may have fewer than 100 shipments a year. Some way of handling this variability had to be developed. The issue was whether to collect all shipments for a time period that varied by establishment or a variable number of shipments reported for a constant period.

To collect a set number of shipments for each establishment a few things must be true. First, the respondents must be capable of reporting shipments as they occurred. This is necessary to avoid the respondents' getting information from files of already completed shipments. Some shippers order their shipments by customer, others order them by commodity, and others order them by the carrier of the shipment. Some shippers use multiple filing schemes. If we had only the first shipments from a file or files so ordered, our estimates would be unrepresentative and biased. Second the respondents must be able to measure the time required to produce those shipments. This would allow us to estimate a yearly volume.

As for the first requirement, research conducted for the Commodity Transportation Survey showed that most respondents preferred to use their files of completed shipments rather than sample as they are making shipments. Concentrating on one task at a time was easier for them. We rejected the second requirement because we believed that keeping the length of reporting period constant would be easier for the respondents to do and the results would be more reliable. This plan requires that we get a sample of shipments.

1.5. The Sampling Plan

Because of the problems with other methods of collecting the data required to meet the survey's objectives, the design team concluded that the best approach for the Commodity Flow Survey was for respondents to provide information for a sample of shipments that they would select. Necessarily, the respondents must also construct the frame of shipments.

Research conducted for the CTS suggested that for most establishments, a sample of 200 shipping documents would yield estimates of annual total value of shipments that had a coefficient of variation of about 30% or less. Empirical comparisons of several sampling schemes showed that selecting quarterly samples of 50 documents covering a two-week period gave the smallest median coefficient of variation (22%) for estimating the establishment's total value of shipments. The schemes varied the frequency of sampling (once a month, once every two months, once a quarter, twice a year, or once a year), the period covered (one week, two weeks, one month, one year) and the number of documents sampled (to give a total of 200 documents). A systematic sample of shipping documents would be selected to avoid biases caused by the order of the shipping documents in the files sampled.

The systematic sampling scheme was set up in the following manner. Respondents were asked the establishment's total number of shipments for the given two-week sampling period. Then they used a lookup table to translate the total number of shipments to a "take-every" number that, properly applied, would result in a sample of 20 to 50 shipments. The exact instructions are presented in figure 1. Note that respondents with 40 or fewer shipments for the two-week period were to report them all.

SAMPLE SELECTION INSTRUCTIONS

1. Enter your total number of shipments for the 2-week period. ——→ [ ]

NOTE — Remove any voided invoices, credit memoranda, etc. from the files, if possible, before estimating the total number of shipments.

2. Find the range in column (1) at right that includes the number entered in 1 above. Put an (X) in column (2) beside it.

3. If your total number of shipments is 40 or less, provide data for **every** shipment during the 2-week period in Item F. If the number of shipments is 41 or more, continue with steps 4 and 5 to select the shipments to report.

Number of shipments (1)	*Mark (X) one* (2)	"Take every" number (3)	Expected sample size (4)
0–40		*Select every shipment*	1–40
41—100		2	20—50
101—200		5	20—40
201—400		10	20—40
401—800		20	20—40
801—1600		40	20—40
1601 or more		*Call Census 1-800-528-3049*	

CONTINUE ON NEXT PAGE.

Figure 1

2. Evaluating the Response

The CFS sample comprised 197,163 establishments. We tabbed about 119,000 establishments (60.7% of the total sample). The establishments that were not tabulated included those that did not respond (18.8%), did not report usable data (1.5%), did not ship (7.9%), made no shipments during the reporting period (2.3%), had been misclassified in an inscope industry (3.1%), and those that did not have representative samples of shipments according to our analysis (5.7%).

2.1. The Census Adjustment Ratio

Most of the establishments in the Commodity Flow Survey were also reporting in an Economic Census for 1992. Based on the sample of shipments reported by the establishment we computed an estimated 1993 total value of shipments (TVS). We compared this estimate with either the TVS reported in the census or a reasonable proxy for the TVS from the census. In fact, we used the ratio of the census TVS to the CFS TVS to adjust the estimates.

Before seeing any data and based on the earlier study's median 22% cv, we believed this ratio would rarely be greater than three or less than one-third. Upon tabulating estimates for all usable CFS establishments with a Census TVS, we found that many factors were outside this range. Figure 2 gives the distribution of this ratio by the number of useable quarters of response.

Based on the wide variations in the 'census adjustment' ratio and particularly because of the large and small size of some adjustment factors, we excluded establishments whose Census TVS was more than 10 times or less than 1/10 of

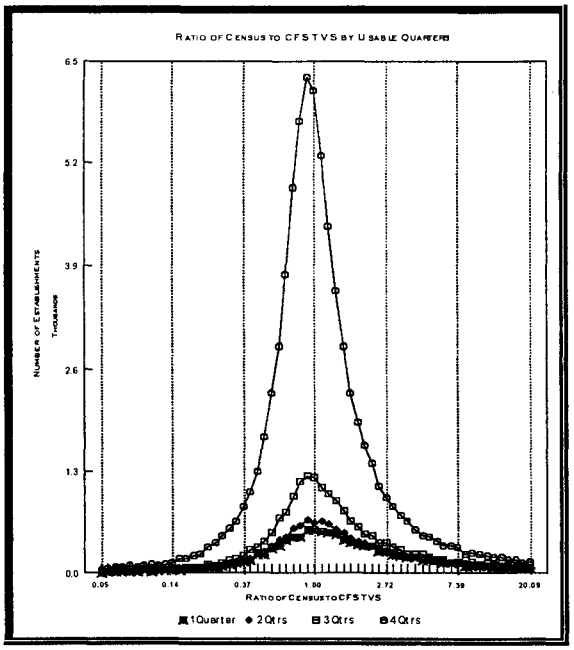

Figure 2

the CFS TVS. This eliminated about 8,900 establishments from the estimates. We felt that the inaccuracy of the CFS-

level estimate tainted the associated estimates of shipments' characteristics such as commodity and mode of transport.

Reviewing this graph, the number of establishments with four usable quarters of data is, at all levels, greater than the number of establishments with fewer than four usable quarters. Although seeing it from this graph is hard, the *relative* number of establishments with a census adjustment ratio between one-half and two increases with the number of usable quarters. Also, the number of establishments with a census adjustment ratio outside that range decreases with the number of usable quarters. In other words, the more usable quarters the better the estimate.

2.2. Response Problems

As we looked into the causes of these large (and small) adjustments, we noticed several problems that occurred repeatedly. Each of these problems contributes to the bias and nonsampling variance of the estimate and not to the sampling variance. Because of the extraordinary respondent effort required, it would not be a surprise to find that many respondents made mistakes filling out the form. Some of our findings follow.

2.2.1. Large and Infrequent Shipments

Some establishments did not report any commodities consistent with their SIC classification. This problem was particularly prevalent in SIC 37, Transportation Equipment Manufactures. One cause of this could be that the establishment's SIC classification is wrong. For example, an establishment might be a wholesale parts distributor rather than a manufacture. SIC misclassification is a common problem, but there is another important possibility. We believe that some respondents do not consider the delivery of a fully-assembled vehicle or group of vehicles to be a "shipment." It is also possible that some companies do not keep the records for huge shipments with the records for other less extraordinary shipments. If so, the respondent (who might be a secretary or warehouse manager) might not have known to include such shipments.

This exposed the biggest problem with using a uniform sampling rate with a small sample size. The variation of the sample estimates is extremely large for any establishment that had many highly variable shipments. If the establishment had extremely large but infrequent shipments as well as smaller, more frequent shipments, we faced two possibilities. A large, rare shipment is included in the sample and the estimates are extremely high or it is not and the estimates are extremely low.

2.2.2. Inconsistency of Sampling-Related Data Items

The instructions ask respondents to count the total number of shipments sent during the two-week period, find a selection rate based on that total, and select a systematic sample of shipments. The sample size should equal the total divided by the selection rate, truncated. Encouragingly, we found that 90% of the respondent's samples were within two shipments of the correct size. We also tabulated the agreement of the total number of shipments, selection rate, and the shipment sample size for all useable respondents. Only 22% of the respondents gave internally consistent responses to three sampling-related items each time they responded. Of those establishments that reported nonblank data for four quarters, only 13% had these three items consistent each quarter. Of course, this does not measure how many respondents constructed a complete and unduplicated frame, nor does it measure the number that failed to take a systematic sample of shipments, it serves as an indicator of response inconsistencies.

2.2.3. Errors in the Reported Total Number of Shipments

We used the ratio of the universe to the sample size to weight the shipments, so incorrect universe counts translate to biased estimates. We believe some respondents rounded off their universe size to the nearest 10 or 100. Figure 3 shows this distribution. To illustrate the bias toward estimating even multiples of ten, we have one line connecting the multiples of ten and a second line connecting the nonmultiples of ten. The line joining the multiples of ten is always above the line joining the nonmultiples of ten. Adjusting estimates to the census will remedy errors in level but not errors of misrepresenting the relative proportion of some characteristic (mode, commodity, etc.) that could occur by combining data from different quarters from the same establishment.

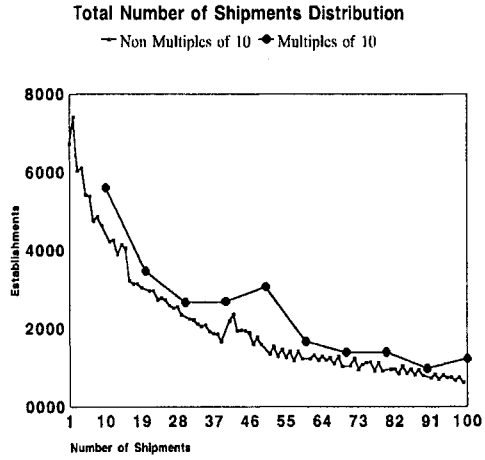

Figure 3

Another error that we occasionally saw was the inclusion of noncommodity shipments (such as payroll checks or contracts) in both the universe and the sample.

These shipments were found by inspection since for economy, we keyed the commodity code not the description and the respondents coded them as shipments of paper products. To have accurate weights we need the exact number of commodity shipments in the universe.

2.2.4. Differences in Actual and Expected Sample Sizes

The number of establishments that reported a sample of exactly 40 or 50 shipments was much greater than those that reported other numbers of shipments. Figure 4 contrasts the distribution of the actual shipment sample sizes with the distribution that we should have seen based on the reported distribution of the total number of shipments for the two-week period. Overall the agreement between the expected samples reported and the actual reported was close. About 90% of all establishments reported an actual number of shipments that was within two shipments of the expected number based on their reported total number of shipments. Nevertheless, we gained some insights from Figure 4.

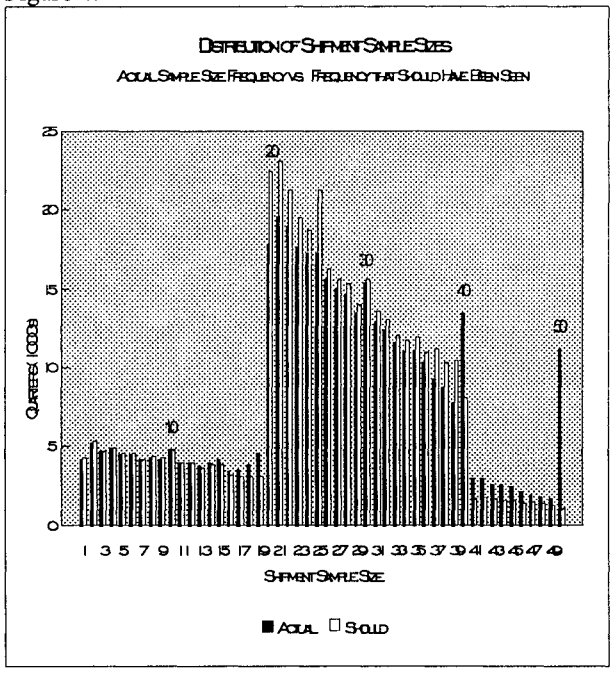

Figure 4

To report fewer than 20 shipments in the sample, the establishment had to have fewer than 20 shipments for the two-week period. Similarly, to report a sample of 41 to 50 shipments, the establishment had to have from 82 to 100 shipments for the reporting period. Any other total number of shipments for the reporting period should have resulted in a sample size of between 20 and 40 shipments. With that in mind, it is clear that the number of quarters that had 50 shipments in a sample was an outlier. We suspect this is because the questionnaire had 50 lines to report shipments and many respondents simply ignored the sampling instructions. The explanation of the spike at 40 is similar. The questionnaire was formatted to allow the reporting of 15 shipments on the second and third pages, 25 shipments on the fourth and fifth pages, and 10 shipments on the last pages. The instructions specified an 'expected sample size' of 20 - 40 shipments for most cases. Respondents eager to help, misinterpreted 'expected' and reported to the natural end of the fourth page -- a total of 40 shipments. The biggest problem with doing this is that we cannot tell from where the additional shipments were drawn. In addition, combining an incorrect sample size and an estimated universe size will adversely affect the estimation.

2.2.5. Patterns of Reported Shipments

Looking at the shipment identification numbers suggested that inaccurate shortcut sample methods were used. The most common shortcut was to enter the first 20 to 40 shipments and avoid sampling. Other respondents had patterns that showed interesting deviations from a strict systematic sample. For example, one respondent followed a pattern of reporting four shipments in a row having a numeric serial number followed by a fifth shipment that had an alphabetic serial number. The fifth shipment was always many times heavier and higher valued than any of the previous four shipments. He repeated this 4-1 pattern for the entire questionnaire. We believe the respondent tried to represent various files or locations by ensuring that he included some shipments from each in the sample. Unless he sampled the shipments in the correct proportions, this could do more harm than good.

2.2.6. Confusing Terms

The following table compares the median adjustment ratio for different levels of quarterly response and for different incidences of sampling. Our expectation is those estimates from establishments that report all of their shipments (40 or fewer) will have an adjustment ratio closer to one than establishments that take a sample of shipments.

What we see is exactly the opposite. Whatever the number of quarters reported, the average adjustment ratio for respondents that never sampled (the zero row) is further from 1 than the ratio for respondents that did sample. It is also interesting that the adjustment ratio decreases with increasing sampling (going down a column). We believe that the major cause of this result is that the respondents that we classified as reporting all of their shipments misunderstood the instructions as described below and those that report data suggesting they were drawing a sample understood the instructions.

The instructions ask for the "total number of shipments

for the 2-week period." Upon contacting some of these respondents we found that they thought we wanted the number of shipments for which they were transcribing data. For instance, they may have had, say, 600 shipments and reported data for 30 of them. They then entered "30" as the total number of shipments and checked the "select every shipment" take-every box. Here, we would never know that they had misinterpreted our instructions and so our evaluation of the number of samples reported for a particular establishment may be underestimated. We should note that the adjustment ratio could possibly inflate the report to an appropriate level.

Median Ratio of Census TVS to CFS TVS					
		Quarters Reported			
		1	2	3	4
Samples Reported	0	1.442	1.420	1.314	1.147
	1	1.035	1.338	1.183	1.051
	2		0.971	1.053	0.998
	3			0.936	0.978
	4				0.904

3. Changes for the 1997 CFS

The problems we found in the 1993 data made us reevaluate our methods. Broadly speaking, we found that while most respondents followed our instructions well, some had difficulty understanding and complying with our instructions. The net effect is an unknown amount of bias in the data and estimates. We therefore decided the focus of any survey improvements should be on reducing the bias and nonsampling error in the data.

For reasons described in the first section, we kept the 1993 sample design intact. Nevertheless, we made several changes in the survey. Perhaps the most significant was a 50% cutback in the sample size. The 1997 CFS sample comprises 102,000 establishments. This decrease was done so that we could target the data collection staff to followup problem reporters early in the survey. The 1993 sample was so large that the data were not completely keyed until mid-1994 and preemptive education of respondents was precluded. By halving the sample, we can identify problems nearer to when they occur. We can also contact many respondents to encourage them to respond for the entire year. By doing this, we hope to have not only a greater response rate, but also a higher quality, more accurate response.

A second change attempts to reduce problems created by large and infrequent shipments. On the new questionnaire we ask, "In the last three months did this location have any individual shipments with a value more than $2,000,000?" If the answer is "No," no action is needed. If the respondent answers "Yes," the reported shipments are checked to see if they adequately represent large shipments. If they do, we take no action. If they do not, we will call and ask the respondent to report all of its large shipments. This strategy effectively creates one certainty and one noncertainty stratum for the shipment universe and should reduce the variance of the estimates.

A question added for 1997 asks the weekly value of shipments. This will help us followup respondents whose sample-estimated weekly value of shipments differ substantially from their reported total. For now, the responses will be used only for editing and followup. However, it could potentially be used for weighting.

We have also shortened the reporting period from two weeks to one week. We hope this improves the response because we will require that fewer respondents sample. In about 38% of the reporting periods in 1993, respondents reported all of their shipments. If we can assume that the number of shipments is evenly distributed between the first and second week and that the same distribution of number of shipments holds for 1997, we can expect that 50% of the reporting periods in 1997 will not require sampling.

A final improvement, begun in 1997, is a concurrent evaluation of response to measure some attributes that will be a basis for targeting refinements for the next survey. Some characteristics we are measuring are the frequency of including noncommodity shipments (such as contracts), the job title of the respondent, the presence of multiple files or multiple locations of shipments, and unusual patterns of shipments that may suggest departures from our instructions. Our goal is to quantify the occurrence of problems or ambiguous situations so that we can concentrate our efforts on solving the most significant ones.

References

Dunn, Nancy H. *The 1987 Commodity Transportation Pretest I-- Analysis of the Best Time Frame for Future Establishment Sampling.* Unpublished Census Memorandum dated April 25, 1986.

Dunn, Nancy H. *The 1987 Commodity Transportation Pretest II-- Initial Analysis of the Summary and Shipment Form.* Unpublished Census Memorandum dated March 31, 1986.