Frame Improvements for the 2007 Commodity Flow Survey[1]

Ruth E. Detlefsen, U.S. Census Bureau
Yukiko T. Ellis, U.S. Census Bureau

Abstract

The Commodity Flow Survey (CFS) is conducted approximately once every five years to provide data on the movement of goods in the United States. Information on the frame available for the CFS was inadequate for determining shipper status or developing accurate measures of shipping activity for auxiliaries (establishments of large companies that provide support to the rest of the company). For the 2007 survey, steps were taken to address this issue. This paper will describe the problem and its effect on the 2002 CFS, discuss the 2007 approach and its implications for the frame and subsequent sample design for 2007, and provide an initial evaluation of the 2007 approach.

Keywords: Frame, Commodity Flow

1. Commodity Flow Survey Description

The U.S. Census Bureau has partnered with the Bureau of Transportation Statistics (BTS), U.S. Department of Transportation to conduct the Commodity Flow Survey (CFS) for reference years 2007, 2002, 1997, and 1993. The first publicly available results from the 2007 CFS are expected in December 2008. This survey provides information about the movement of goods in the United States and is a major source of transportation statistics not available elsewhere. Information from the survey includes value of shipments, weight of shipments, commodities shipped, mode(s) of transportation used, origin and destination of shipments, ton-miles, and average miles per shipment. Data from the survey are used by public policy analysts and for transportation planning and decision-making to assess the demand for transportation facilities and services, energy use, safety risk and environmental concerns.

In-scope to the CFS are establishments classified in mining, manufacturing, wholesale, and selected retail and service industries. Also included are auxiliary establishments that ship in support of multi-establishment companies. (Prior to the 2007 survey auxiliaries were limited to those that ship in support of retail and other in-scope multi-establishment companies.) Foreign establishments, most retail and services establishments, and all establishments classified as farms, forestry, fisheries, governments, finance, utilities, construction, and transportation are out-of-scope.

The sample for each Commodity Flow Survey has been selected using a stratified three-stage design in which the first-stage sampling units were establishments, the second-stage sampling units were groups of four 1-week (2-week for 1993) reporting periods within the survey year, and the third-stage sampling units were shipments.

The first-stage sample has consisted of a stratified simple random sample without replacement (except that a probability-proportional-to-annual payroll design was used for 1993). The sample size has varied from approximately 50,000 to 200,000 establishments, depending on the survey year. Primary stratification has been by industry and geography with secondary stratification by size. The sample has been allocated by imposing coefficient of variation (CV) constraints on the value of shipments estimate for each primary strata. For the 2007 survey we also imposed a CV constraint at the national level.

For sample stratification, industry is used as a proxy for commodity. Geographic strata have varied in definition and number depending on the survey year, but have generally been based on the 50 states and the District of Columbia and selected metropolitan areas or transportation areas of interest.

[1] *This report is released to inform interested parties of research and to encourage discussion. The views expressed on statistical, methodological, technical, or operational issues are those of the authors and not necessarily those of the U.S. Census Bureau.*

For the size stratification we have estimated an annual value of shipments for each establishment. The methods used have varied some with each survey year.

The CFS collects data on shipments originating from establishments located in the 50 states and the District of Columbia. Each establishment selected into the sample was contacted once each quarter of the reference year. Respondents were instructed to select a sample of their establishment's shipments. For each selected shipment, they were requested to provide domestic destination or port of exit, commodity, value, weight, mode(s) of transportation, the date of the shipment, and an indication of whether the shipment was an export, or a hazardous material. Additional information was requested for exports and hazardous materials shipments.

2. Auxiliary Issue

Auxiliaries present special scoping and sizing challenges. An auxiliary is an establishment of a multi-establishment company that provides support to the rest of the company. Examples include: headquarters offices, regional offices, data processing facilities, training facilities, warehouses, and distribution centers.

For each CFS conducted we have had difficulty using information available on the frame to determine whether auxiliary establishments are doing any shipping for their company, and if so, how much. As a result, for past surveys many sampled auxiliaries turned out to be non-shippers and measures of shipping activity estimated for sampling and editing purposes were frequently substantially inaccurate.

Section 3 describes our efforts to overcome these difficulties for the 2002 CFS along with the outcome. Section 4 discusses studies conducted in preparation for the 2007 CFS. At the same time that these studies were underway, a decision was made to take another approach as described in section 5. An analysis of **t**he 2007 approach is presented in section 6.

3. Treatment of Auxiliaries in the 2002 CFS

3.1 Identifying Shippers

To cover auxiliaries that ship in the 2002 CFS, we included in the frame, certain establishments classified into Warehousing and Storage (NAICS 4931) and all Corporate, Subsidiary, and Regional Managing Offices (551114) that provided support to manufacturing, mining (except oil and gas), wholesale, and retail establishments of multi-establishment companies. Managing Offices were included since they may perform a warehousing/shipping function as well as their primary management function.

We excluded any auxiliary from the 2002 frame that:

- was a managing office classified as a nonshipper in the 1997 CFS,
- had no other establishments in its parent company that were on the sampling frame or no others classified in the Retail Trade Sector, or
- had an industry code from the 1992 Economic Census that was considered out-of-scope to CFS. The Standard Industrial Classification (SIC) system was used for the 1992 census. SICs assigned to auxiliaries reflected primary industry served.

We also considered excluding managing offices if they had not reported sales or end-of- year inventory in the economic census completed just prior to the CFS (1997 for the 2002 CFS), since that had been done for the 1997 CFS. However, an evaluation showed that some managing offices that reported shipping activity in the 1997 CFS reported no sales or end-of-year inventories in the 1992 census. Therefore, the 1997 census results were not used to exclude managing offices from the 2002 CFS sampling frame.

We identified 18,512 managing offices and 4,432 warehouses as in-scope to the 2002 CFS frame.

3.2 Measure of Size

We assigned a measure of size (MOS) to each auxiliary included on the 2002 CFS sampling frame. The MOS was a proxy for an establishment's annual value of shipments and was used in the sample design, edits, and industry adjustment of estimates.
For about 3 % of the auxiliaries, the MOSs were set to the value of shipments estimates from the 1997 CFS, adjusted to preliminary 1997 economic census results. For 97% of the auxiliaries the MOSs were

derived using estimated regression coefficients applied to the 2000 annual payroll. The coefficients were estimated from regression models that attempted to approximate value of shipments from administrative payroll data.

3.3 Sampling

The auxiliaries identified as being in-scope to CFS were stratified by industry and geography into 219 primary strata. From those we selected a sample of 1,479 warehouses and 2,314 managing offices.

3.4 Outcomes of Approach

Only 41% of the auxiliaries selected into the sample were tabulated as shippers – 67% of warehouses and 33% of managing offices.

For those that did have shipping activity, 90% of the warehouses and 65% of the managing offices provided useable responses. An examination of assigned measures of size (MOSs) indicated that we should consider other methods to improve the MOSs. Auxiliary establishments typically do not report sales or receipts in the economic census, our primary source of frame data. These establishments provide support for the rest of the company and so do not directly produce revenue. If data are reported in the census, these may represent more than just the shipping activity at the establishment, such as receipts for the entire company or they may include the payroll of company executives at a headquarters.

4. Auxiliary Studies Conducted In Preparation for the 2007 CFS

Two studies were conducted prior to the 2007 CFS. One study concerned better identification of shippers. The second concerned improving the measure of size.

4.1 Identifying Shippers

We examined information collected in the 2002 Economic Census to determine if we could use it to more accurately identify shipping status of auxiliaries. We looked at whether an auxiliary reported employment in Trucking (NAICS 484) or Warehousing (NAICS 4931), reported positive inventory, reported secondary activity in Trucking or Warehousing, or whether the auxiliary's name contained variants of words such as "Warehouse" or "Distribution Center." If none of these were the case, we speculated that the auxiliary was likely not a shipper. About 23,800 of the roughly 31,900 auxiliaries that responded to the 2002 census, gave no indication of being shippers under this speculation.

To test the speculation, we matched the auxiliaries included in the 2002 CFS sample to the census. About 72% matched. Of those that matched, using the census information would have resulted in 24% being incorrectly classified as nonshippers and 10% being incorrectly classified as shippers. Thus, 34% would have been misclassified.

4.2 Measure of Size (MOS)

Research was conducted to determine if additional methods, equivalent to those used for non-auxiliaries, could be incorporated into the MOS calculations for auxiliaries. The correlations between shipment value and the variables used in those calculations were weak even for methods for which sufficient data were available. Short of obtaining size information from an advance survey, there seemed to be no good way to improve the MOS calculation for auxiliaries.

5. Treatment of Auxiliaries for the 2007 CFS

A decision was made to conduct an advance survey to obtain information on shipping status and value of shipments for the auxiliaries. The advance survey was conducted in the spring of 2006. There were 4,845 establishments classified into warehousing and 35,152 establishments classified into managing offices. The number of managing offices included in the advance survey for the 2007 CFS was much higher than the number included in the sampling frame for the 2002 CFS. This was the case because for the advance survey we did not limit the auxiliaries to only those that support retail or in-scope multi-unit companies. Also, we did not match the auxiliaries to the 2002 CFS to eliminate any that had reported that they were non-shippers.

5.1 Identifying Shippers

A questionnaire was mailed to each auxiliary to determine if any of the following were performed by the auxiliary:

- Ship or deliver products to customers or clients
- Ship or deliver products to other locations of this company
- Provide for customer pick-up of products
- Maintain records of shipments or deliveries leaving this location

Responses were received for 87% of the auxiliaries. Auxiliaries that performed none of the four activities were deemed to be non-shippers and were excluded from the CFS sample frame. All others, including non-respondents, were assumed to be potential shippers. Table 1 below shows the final shipper status of the auxiliaries.

Table 1: Shipper Status of Advance Survey

	Warehouses	Managing Offices	Total
Shipper Status			
Total Shippers	**3,770**	**12,007**	**15,777**
Shipper/responded	2,942	7,774	10,716
Shipper/no response	828	4,233	5,061
Non-Shippers	**1,075**	**23,145**	**24,220**
Total	**4,845**	**35,152**	**39,997**

5.2 Measure of Size

We provided eight annual value of shipment ranges and asked that the appropriate one be marked for auxiliaries that ship.

Of the 10,716 auxiliaries with responses indicating that they were shippers, 9,586 responded to the annual value of shipments inquiry. For shipping auxiliaries that marked a range, the MOS used for sampling and editing was randomly assigned from a uniform distribution within the range selected.

For shippers not reporting annual value of shipments (including those who did not report at all), the MOS was based on their 2002 Economic Census responses or other administrative data contained on the Census Bureau's Business Register using the same method as for the 2002 CFS.

5.3 Sampling

In the fall of 2006, we constructed the frame and selected the 2007 CFS sample.

The warehouse and managing office auxiliaries that responded to the advance survey and were considered to be shippers were stratified together into 123 geographic strata.

Two national level strata were created for auxiliary establishments that did not respond to the advance survey or that were added to the sampling frame subsequent to the advance survey. One stratum included warehouses and the other included managing offices. We stratified these separately because warehouse auxiliaries are more likely to be shippers than are managing office auxiliaries.

This gave us 125 auxiliary strata from which we selected a sample of 4,713 auxiliaries.

Table 2 below summarizes the auxiliary frame and sample sizes for the 2002 and 2007 surveys.

Table 2: Auxiliary Frame and Sample Size for 2002 and 2007

	2007 CFS		2002 CFS	
Industry	**Frame**	**Sample**	**Frame**	**Sample**
4931	4,065	1,668	4,432	1,479
551114	10,813	3,045	18,512	2,314
Total	14,878	4,713	22,944	3,793

6. Analysis of the 2007 Approach

We conducted an analysis to determine whether the advance survey gave us a better way to identify auxiliaries that ship and improved the measures of size assigned to the auxiliaries. The analysis follows.

6.1 Shipper Status of Auxiliaries – 2007 vs 2002

We compared the shipper status of auxiliaries selected for the 2007 and 2002 surveys to assess the impact of the advance survey in distinguishing shippers from non-shippers. We also compared the shipper status for non-auxiliaries. Table 3 shows the results.

Table 3: Shipper Status of Units Selected into the CFS for 2007 and 2002

Industry	2007 Advance Result	2007 CFS Shipper	2002 CFS Shipper
4931	All selected	76.3%	66.9%
4931	Nonrespondent	47.0%	--
4931	Shipper	77.5%	--
551114	All selected	46.3%	33.2%
551114	Nonrespondent	8.1%	--
551114	Shipper	48.1%	--
Non-Auxs	--	63.8%	65.0%

We found that:

- About 76% of the warehouses (NAICS 4931) selected for the 2007 CFS that were also included in the advance survey responded as shippers to the 2007 CFS. This was about 10 percentage points more than those that responded as shippers to the 2002 CFS. It was about 12.5 percentage points higher than the non-auxiliaries that were not afforded the opportunity to respond to the shipper status questions on the advance survey.
- 47% of the warehouses that were non-respondents to the advance survey responded as shippers to the 2007 CFS.
- About 77% of the warehouses that responded as shippers to the advance survey, responded as shippers to the 2007 CFS.
- About 46% of the managing offices (NAICS 551114) selected for the 2007 CFS that were also included in the advance survey responded as shippers to the 2007 CFS. This was about 13 percentage points higher than those that responded as shippers to the 2002 CFS. On the other hand it was about 17.5 percentage points lower than the non-auxiliaries that were not afforded the opportunity to respond to the shipper status questions on the advance survey.
- Around 8% of the managing offices that were non-respondents to the advance survey responded as shippers to the 2007 CFS.
- About 48% of the managing offices that responded to the advance survey as shippers responded as shippers to the 2007 CFS.

6.2 Impact of Size on Final Shipper Status

By industry, we examined whether the size reported in the advance survey by an auxiliary was a factor in the discrepancies observed between the shipping status reported in the advance survey and the shipper designation assigned in the CFS.

Tables 4 and 5 provide some results for the warehouses and managing offices, respectively. These tables display results for auxiliaries that were canvassed in the advance survey and later selected in to the CFS sample. Each row shows the percent of total auxiliaries for a given advance survey measure of size category that 1) reported being shippers in the advance and in the CFS, 2) reported being shippers in the advance but not in the CFS, 3) did not report in the advance but reported as shippers in the CFS, and 4) did not report in the advance and did not respond to the CFS. The row percentages sum to 100%.

Table 4: Comparison of Shipper Status by Advance Survey MOS

Warehouses				
ADVANCE	Shipper		Nonresponse	
CFS	Shipper	NR/NS	Shipper	NR/NS
MOS				
<$5M	67.1%	32.9%	--	--
$5M-$20M	78.8%	21.2%	--	--
$20M-$50M	72.3%	27.7%	--	--
$50M-$200M	71.2%	28.8%	--	--
$200M-$500M	85.1%	14.9%	--	--
$500M-$1B	91.5%	8.5%	--	--
$1B-$5B	81.6%	18.4%	--	--
>$5B	53.9%	46.2%	--	--
No Response[2]	32.5%	9.7%	27.2%	30.7%
All	74.3%	21.6%	1.9%	2.2%

Table 5: 2007 Advance Survey Response to MOS by Shipper Status

Managing Offices				
ADVANCE	Shipper		Nonresponse	
CFS	Shipper	NR/NS	Shipper	NR/NS
MOS				
<$5M	33.3%	66.7%	--	--
$5M-$20M	56.3%	43.7%	--	--
$20M-$50M	59.9%	40.1%	--	--
$50M-$200M	50.7%	49.2%	--	--
$200M-$500M	43.6%	56.4%	--	--
$500M-$1B	50.0%	50.0%	--	--
$1B-$5B	38.7%	61.3%	--	--
>$5B	21.9%	78.1%	--	--
No Response[2]	9.7%	20.0%	5.7%	64.6%
All	45.9%	49.4%	0.4%	4.3%

We found that:

- There is no clear indication that size is a factor in the observed shipper status discrepancies except that there is some indication (as well as anecdotal evidence) that respondents reporting more than $5 billion in shipments, may have been reporting for their entire company rather than a specific establishment.
- The patterns differ between the warehouses and the managing offices. There are larger discrepancies between the shipper status from the advance and the CFS for the managing offices than for the warehouses.

6.3 Measure of Size Compared to Value of Shipments

For auxiliary units that reported in the CFS, we examined the ratio of the measure of size (MOS) to the appropriately weighted value of shipments reported in the CFS. These ratios should be close to 1. We compared ratios for cases with MOSs derived from their advance report 1) to those derived using the methodology described in section 5.2., 2) to those derived by using the 2002 methodology for all cases in the advance survey, and 3) to those derived for the 2002 CFS. Results are shown in Table 6.

We found that:

- For 2007, the MOSs derived from the advance survey appear to be better than those derived using other methods.

[2] Some advance survey respondents indicated that they were shippers but failed to mark a measure of size checkbox.

- For 2007, the MOSs derived from the advance survey appear to be better than the MOSs that would have resulted using the 2002 methodology for warehouses. Results are not clear for managing offices.
- For MOSs derived from the advance survey, the means for both warehouses and managing offices are not as extreme as for the MOSs from the 2002 CFS.
- For MOSs derived from the advance survey, the median for warehouses is closer to 1 than for the MOSs from the 2002 CFS.
- For MOSs derived from the advance survey, the median for managing offices is further from 1 than for the MOSs from the 2002 CFS.

Table 6: MOS to Value of Shipments Comparisons

	2007 CFS MOS Source			2002 CFS
	Advance	**All Others**	**2002 Method**	
Industry				
4931				
# of units	1,149	93	1,242	931
mean	9.81	48.84	9.49	12.65
3rd quartile	2.13	1.93	1.48	1.65
median	1.16	0.75	.60	0.63
1st quartile	0.62	0.22	.24	0.30
551114				
# of units	1,191	79	1,270	608
mean	139.94	212.33	146.27	380.31
3rd quartile	3.67	17.19	1.50	5.09
median	1.52	3.02	.40	1.17
1st quartile	0.75	0.50	.07	.43
All Other				
# of units	--	61,653	--	30,299
mean	--	244.36	--	60.28
3rd quartile	--	2.15	--	2.40
median	--	1.05	--	1.24
1st quartile	--	0.63	--	.76

7. Summary/Future Research

An advance survey was sent to auxiliary establishments to try to improve the shipper status and the measures of size for the shippers.

The advance survey enabled us to more accurately assign shipper status to both the warehouse and managing office auxiliaries on the 2007 CFS sampling frame. The shipper status was more accurate for the warehouses on the frame than for the non-auxiliaries. However, the accuracy of shipper status for managing offices on the frame was less than for non-auxiliaries.

For both the warehouses and the managing offices tabulated in the 2007 CFS, the measures of size derived from advance survey responses were more accurate than were those computed from administrative data. Measures of size derived from the advance survey for the 2007 warehouses seem to be better than those that would have resulted had we used the 2002 methodology and also than those derived for the 2002 CFS. Results are mixed for the managing offices.

There is still room to improve the accuracy of the shipper status and the measure of size. Some respondents misunderstood what was being asked on the advance survey questionnaire. Some auxiliaries thought they were responding for their entire company rather than just for their particular establishment. In addition, no allowance was made for respondents to indicate that their value of shipments was zero. Clarifying the instructions and adding a zero box could lead to further improvements in the accuracy of both the shipper status and the measure of size. Asking respondents for estimated value of shipments rather than asking them to mark the appropriate range might improve the MOS accuracy, assuming response to the question remained relatively high.

Other research that could lead to further improvements includes:

- Conduct follow-up research on cases that said they shipped in the advance survey, but said otherwise in the CFS.
- Conduct follow-up on MOS derived from the advance survey that differed considerably from the CFS response.
- Conduct follow-up research on advance survey cases that indicated that they did not ship to determine if that is the case.
- Use advance survey results to verify some assumptions about auxiliaries that ship. For example, for Commodity Flow Surveys prior to 2007 an assumption was made that service auxiliaries do not ship. Since we mailed all auxiliaries in the advance survey, we should be able to test this assumption. If the assumption holds, we can exclude these from surveys, saving mailing and processing costs.
- Consider using an advance survey to improve the shipper status of non-auxiliaries. This has the potential to increase the efficiency of the CFS sample as only 63.8% of sampled non-auxiliary establishments were determined to be shippers for the 2007 CFS and 65.0% for the 2002 CFS.

References

Chandhok, P., "Precanvass Survey for the 2007 Commodity Flow Survey," 2007 Proceedings of the American Statistical Association, Section on Government Statistics, Alexandria, VA: pp. 1415-1421.

Dahl, S. and Davie, Jr., W., "Designing the 2007 Commodity Flow Survey," 2007 International Conference on Establishment Surveys, CD-ROM, Alexandria, VA: American Statistical Association: pp. 158-168.

Ellis, Y., "2002 CFS Review of Sample Auxiliaries in NAICS 4931 and 5511: Action Item #1 from the Scope/Industry Coverage Joint Investigative Team Meeting on April 7, 2005," Internal Census Bureau document.

Ellis, Y., "Identifying Non-Shipping Auxiliaries in 2002 Commodity Flow Survey (CFS) Based on Economic Census Responses," June 20, 2008, Internal Census Bureau document.

Kinyon, D.L., "2002 CFS Extraction Specifications," Internal Census Bureau document.

Kinyon, D.L. and King, C.S., "Overview of the Establishment Sampling Frame, Computer Edits, and Imputation Methodology for the 2002 Commodity Flow Survey," 2005 Proceedings of the American Statistical Association, Section on Survey Research Methods, Alexandria, VA: pp. 3220-3227.

Smith, R. T., Evans, B. T., & Fowler, J. L., "The Commodity Flow Survey." 1994 Proceedings of the American Statistical Association, Section on Business and Economic Statistics Section. Alexandria, VA: American Statistical Association, pp. 89-94.

U.S. Census Bureau, Economic Census. www.census.gov.

U.S. Department of Transportation and U.S. Department of Commerce (December 2004). 2002 Commodity Flow Survey: United States: 2002, www.census.gov/svsd/www/cfsdat/2002cfs.html.

U.S. Office of Management and Budget (1998). North American Industry Classification System: United States, 1997. Lanham, MD: Bernan Press.

U.S. Office of Management and Budget (2002). North American Industry Classification System: United States, 2002. Lanham, MD: Bernan Press.

U.S. Office of Management and Budget (1987). Standard Industrial Classification Manual 1987. Springfield, VA: National Technical Information Service.

Washington, K. and Clark, K., "Combined Measure Of Size Write-up," October 15, 2005, Internal Census Bureau document.