

APPENDIX G

NATIONAL UNDERGROUND STORAGE TANK SURVEY NATIONAL SAMPLE OF FARMS

I. INTRODUCTION AND SUMMARY

The survey of underground motor fuel storage tanks is designed to provide national estimates of the number of underground motor fuel storage tanks at the end use point and the number and percent of these tanks which leak. The survey design defined three segments of the overall target universe of establishments with underground motor fuel storage tanks:

- o Fuel establishments (gas stations and establishments in other fuel-related or fuel-using industries) which by the nature of their business are likely to have such tanks;
- o Large establishments (20 or more employees) which by virtue of their size may have an underground motor fuel storage tank; and
- o Farms, of which over half have motor fuel storage capacity, but an unknown proportion store motor fuel underground.

The sample design for the survey is a two-stage cluster design. The first stage is survey locations, called Primary Sampling Units (PSUs) and consisting of counties or groups of counties. The contiguous United States was divided into six survey regions, based on rough similarity of soil type and condition, as defined in Figure G-1. Thirty-four PSUs were drawn, six from each region, except four PSUs were drawn from Region 5.

Figure G-1. Six regions for National Survey of Underground Storage Tanks

1 -- Northeast

Maine
New Hampshire
Vermont
Connecticut
Massachusetts
Rhode Island
New York
New Jersey
Pennsylvania
Maryland
Delaware
Virginia
West Virginia
Washington, D. C.

2 -- Southeast

Kentucky
Tennessee
Arkansas
Louisiana
Mississippi
Alabama
Georgia
North Carolina
South Carolina
Florida

3 -- Midwest

Wisconsin
Minnesota
Iowa
Missouri
Illinois
Indiana
Ohio
Michigan

4 -- Central

North Dakota
South Dakota
Nebraska
Kansas
Oklahoma
Texas

5 -- Mountain

Montana
Wyoming
Idaho
Nevada
Utah
Colorado
Arizona
New Mexico

6 -- Pacific

Washington
Oregon
California

Among the three survey segments, fuel establishments and large establishments are both concentrated in the same areas, where the population is. Drawing a sample of PSUs which is optimal for both of these segments is therefore no problem, because they occur together. Farms, however, tend to be found in the opposite places, those with sparse population. So optimizing the design for farms is in direct opposition to optimizing the design for fuel establishments and large establishments. Since the fuel establishments are the major focus of the survey, accounting for about 800 of the approximately 920 expected establishments with underground motor fuel storage tanks, the sample of PSUs was optimized for fuel establishments by being drawn in proportion to the number of fuel establishments in each PSU. As noted above, the resulting sample of PSUs is not optimal for studying farms.

The second stage of sampling is the sample of establishments within the selected PSUs. Three sample frames (master lists) were developed for the 34 sampled PSUs -- one for fuel establishments, one for large establishments, and one for farms. Samples were drawn from each list:

- o 1618 fuel establishments;
- o 600 large establishments; and
- o 600 farms.

These establishments were contacted to determine whether they were eligible for our survey; that is, whether they had

underground motor fuel storage tanks. The eligibility rates were (approximately):

- o 50 percent for fuel establishments;
- o 15 percent for large establishment; and
- o Less than 5 percent for farms.

This appendix discusses the national farm sample of 600 farms to be screened. Subsection II discusses the target universe of farms and describes the farm sampling frame on a national basis. The 1982 Census of Agriculture conducted by the Census Bureau is taken as the standard count of farms, and a list developed by the Agricultural Stabilization and Conservation Service (ASCS) of the U.S. Department of Agriculture (USDA) is the sample frame used. For the nation, overall, this frame offers good coverage of the farm universe. Subsection III reviews the survey design with reference to the farm sample and compares Census figures with ASCS figures for the selected PSUs. In this subsection, it is seen that the coverage of farms by the frame is weak in some parts of the country. Section IV concludes the appendix with a discussion of the ratio-adjustment weighting method proposed to minimize total sampling error in the farm estimates.

II. TARGET UNIVERSE OF FARMS AND SAMPLING FRAME

A. Two Farm Data Sources

Two sources of information on farms were used in designing and conducting this survey. One is the 1982 Census of Agriculture (the most recent) conducted by the Bureau of the

Census. This source is used as the most reliable source of national statistics about farms. The second is the "1983 Deficiency Master File" developed by the Agricultural Stabilization and Conservation Service (ASCS) of the U.S. Department of Agriculture (USDA), which is used as the list, or sampling frame, for farms.

The Census of Agriculture is a data collection and tabulation effort which is as inclusive as possible. The 1982 Census lists 2,240,976 farms in the U.S. A farm is defined by Census as "any place from which \$1000 or more of agricultural products were sold or normally would have been sold during the Census year." Tables provide breakdowns of these farms by size of farm, value of sales, type of crop, etc., both nationally, by state and by county. Some of these figures are reviewed later in this section.

What the Census of Agriculture does not provide is a list of farms or farm operators in specific places. Thus, for an actual sampling frame we used the USDA/ASCS 1983 Deficiency File. This is a list of farms developed by the USDA containing about 1,942,000 listings (87 percent as many as the Census total). The original impetus for the development of the file was to provide a mechanism for payment distribution for the PIK (Payment-in-Kind) program for 1983. In 1983, the PIK program was so popular that USDA believes that almost everyone engaged in growing PIK program crops (which include various cash grains and upland cotton) applied for it, and hence is listed on the Deficiency File. Because they saw a chance to have a near-Census of farms on a data file, USDA made a special effort to also include listings of farms not eligible for the PIK program. The basic data were gathered by the ASCS county agents.

The official USDA/ASCS statistics indicate that of 2,018,000 farms known to the ASCS, 1,942,000 (or 96 percent) are listed on the Deficiency File. The ASCS definition of a farm is all of the land farmed under one operation.

Only about 57 percent of the farms listed on the Deficiency File (1,116,000 farms) are farms that are eligible for the PIK program. The remaining 43 percent of farms on the list are not eligible for the PIK program. Some portion of the ineligible farms are ineligible because they were not growing PIK program crops, others because they did not choose to apply for the PIK program. Because of the 96 percent coverage of farms known to them, ASCS believes the Deficiency File is a very complete list of farms in the U.S.

In exploring the universe of farms and comparing the two data sources, we take the 1982 Census of Agriculture as the primary source of information on the nation's farms. Although the ASCS total is less than the Census total, it is probable that the ASCS list is not simply a subset of the farms counted by the Census, but a partially overlapping list. This is due to the fact that the two lists are constructed by different organizations for different purposes, are based on different information, and have different definitions as for including and counting specific cases. However, we can get a summary of the nation's farms from the Census and a rough idea of the ASCS coverage of those farms.

B. Summary of the Target Universe Based on the 1982 Census of Agriculture

The figures presented here are taken from Vol. 1, Part 51, U.S. Summary and State Totals of the 1982 Census of Agriculture.

The first table lists total numbers of farms by size and sales categories.

It seems likely that farms with small acreage or low sales volume would be less likely to have underground motor fuel storage tanks and would also be less likely to be included on the ASCS file than large farms. Table G-1 indicates that a number of farms are quite small, with 8 percent of farms reported having one to nine total acres. Also, many farms have quite low sales figures. Nearly one-quarter of farms reported on had less than \$2,500 in sales in 1982.

The Census also gives figures for storage of various fuels (although unfortunately for our survey, no question was asked as to whether the storage was underground). Table G-2 summarizes the storage capacity data for 1982.

This indicates that roughly half of all farms reported gasoline or gasohol storage, and about 40 percent reported diesel storage. The overlap of the two groups is not given but is presumably fairly high. However, the number of farms with substantial storage capacity is much less -- 2 percent reported 2,000 gallons or more diesel storage capacity, and 1 percent reported that much gas storage capacity. Taking 1,000 gallons or more as a cutoff, 7 percent of farms reported this much gasoline storage capacity and 8 percent reported this much diesel storage capacity.

In conclusion, based on the 1982 Census of Agriculture, there were about 2.2 million farms, of which 8 percent were smaller than 10 acres, one-quarter had less than \$2,500 in sales for the year, and perhaps 10 percent have 1,000 gallons or more fuel storage capacity. This last assumes a substantial overlap between storers of gasoline and diesel fuel. If there is little

Table G-1. Farms by acreage and sales
(1982 Census of Agriculture)

Total U.S. Farms		2,240,976
<u>By acreage</u>		
1 - 9		187,665
10 or more		2,053,311
10 - 49		449,252
50 - 499		1,238,162
500 - 1,999		301,320
2,000 or more		64,577
<u>By sales</u>		
Less than \$2,500		536,327
\$2,500 or more		1,702,973
\$2,500 - \$9,999		560,010
\$10,000 or more		1,142,963
\$10,000 - \$99,999	840,583	
\$100,000 - \$499,999	274,580	
\$500,000 or more	27,800	

(1,676 abnormal farms not reported by sales - institutional, research and experimental farms, and Indian reservations.)

Table G-2. Fuel storage capacity, 1982*
(1982 Census of Agriculture)

Farms reporting fuel expenses	Gasoline and Gasohol	Diesel fuel
Storage capacity reported, farms	1,123,463	924,863
1,000's gallons	583,853	648,605
Farms with storage capacity of:		
1 - 499 gallons	616,650	471,646
500 - 999 gallons	352,925	262,902
1,000 - 1,999 gallons	136,455	140,896
2,000 or more gallons	17,433	49,419
Storage capacity reported as "no", farms	451,895	150,210
Storage capacity not reported, farms	422,083	245,380

*Includes above-ground tanks and containers, as well as under-ground tanks.

overlap, as many as 15 percent of farms may have 1,000 gallons or more motor fuel storage capacity.

C. Comparison of Census and Sample Frame

The sampling frame, the ASCS 1983 Deficiency File, is primarily a data base of farms rather than a source of statistics. Hence, we do not have extensive national or state statistics on this file. Nationally, we can compare the number of farms from Census (2,240,976) and the ASCS file (1,942,437), showing that the sample frame file has 87 percent as many farms as the Census. (Note that these are not necessarily completely a subset of the Census farms, as mentioned above.)

We also can compare total cropland acreage between the two data sources. The Census shows 445,362,028 acres of total cropland on 2,010,609 farms with cropland, while ASCS shows 443,850,049 acres of total cropland on its 1,942,437 farms. The ASCS definition of cropland is "tillable soil" -- the land does not have to have been planted, only to be suitable for planting. The Census definition includes three categories:

- o Harvested cropland;
- o Cropland use only for pasture or grazing; and
- o Other cropland.

The two definitions appear to be quite similar.

The sample frame thus covers 99.7 percent of the total cropland reported in the Census and has 96.6 percent as many farms as those reporting cropland in the Census. It appears that farms with no cropland is an area of sparse coverage for the ASCS

list. The major categories of land in farms not included in total cropland are:

- o Pasture and rangeland other than cropland and woodland pastured (418,264,264 acres);
- o Woodland (87,088,255 acres); and
- o Land in house lots, ponds, roads, etc. (36,082,032 acres).

So farms with pasture, rangeland or woodland and no cropland are more likely to be in the Census but not the ASCS list. However, in the Census 90 percent of farms listed had cropland, so farms with none are relatively rare.

Other types of farms which may tend to under-represented by the ASCS list (based on discussions with Tom Meyer of ASCS) would be growers of fruits and vegetables. Most farms grow more than one crop, and so many fruit or vegetable farms may also have a PIK-eligible crop or may be listed as an ineligible farm on the ASCS file. According to Census data, 69,109 (3.1%) of farms reported vegetables harvested for sale and 123,663 (5.5%) reported land in orchards. On a national basis, these farms do not represent a major portion of the target universe, although on a regional basis their proportion varies. These figures are presented as a way of assessing the potential for undercoverage, but we have no direct way of determining the ASCS coverage of these types of farms.

III. SAMPLE DESIGN FOR UST SURVEY, FARM SEGMENT

In this subsection we again review the survey sample design, emphasizing the aspects relevant to the farm sample. The design was a two-stage cluster design. The contiguous U.S. was divided

into six survey regions, as presented in Figure G-1 shown earlier. The first stage of the sample was survey locations, known as Primary Sampling Units (PSUs). These PSUs consisted of counties or groups of counties and were chosen by region with probability proportional to number of fuel establishments. The second stage was the within-PSU selection of farms. Farms were selected from a sampling frame based on the ASCS list for the selected counties with within-PSU probabilities determined so that the overall probabilities of selection would be equal for all farms. We give more details in the following sections.

A. First Stage Sample of Survey Sites (PSUs)

The first stage in the two-stage sample design was of PSUs, which were counties or groups of counties. Within each region, six PSUs (four in the Mountain Region) were selected with probability proportional to their number of gas stations and fuel-related establishments. As discussed in Subsection I, this is the optimal design for studying fuel establishments -- the main focus of the survey.

Table G-3 shows some statistics on number of farms, by region, based on the 1982 Census of Agriculture. The first two columns give the total farms in each region and the corresponding expected sample size, by region, for an equal probability sample of 600 farms to be screened for underground motor fuel storage tanks. Regions 1, 5 and 6 have expected sample sizes of less than 100, with Regions 5 and 6 less than 50. Next, in column 3, we have used the inverse of the PSU probability of selection as a PSU weight and weighted the 1982 Census of Agriculture farm counts for the selected PSUs up to the regional level. By comparing these figures with column 1, we see that our sample of PSUs has considerable variance from the actual totals. As

Table G-3. Farm summary based on 1982 Census of Agriculture, all farms

Region ¹	Agriculture Census count	Expected farm sample ²	Weighted count, sampled PSU's	Expected farm sample ²
1-Northeast	222,099	60	123,714	36
2-South	548,926	147	283,226	82
3-Midwest	725,699	195	908,358	264
4-Central	464,680	125	494,029	144
5-Mountain	121,777	33	147,071	43
6-Pacific	152,630	41	104,164	30
Continental U.S. Total	2,235,811	601	2,060,562	599

¹Regions are defined in Figure G-1.

²These farms are to be screened for the presence of underground motor fuel storage tanks.

mentioned in Subsection II, this is due to the PSU sample selection being based on the number of fuel establishments, a measure inversely correlated with the number of farms.

Finally, column 4 gives the expected sample size based on the 1982 Agriculture Census counts for our PSUs. Regions 5 and 6 are still very low, and Regions 1 and 2 have a lower sample size than expected from the regional totals.

B. ASCS List for Selected PSUs

The actual sample was drawn from a sample frame based on the ASCS 1983 Deficiency File. This file was described in Subsection II above on a national basis. Here, we compare the ASCS file counts to the Census counts for our sampled PSUs and present some relevant Census figures on a regional basis. The actual sample frame used was a modification of the ASCS file, which we describe below, leading to the final sample sizes.

In Table G-4, the Census of Agriculture counts are compared with the ASCS file counts for the sampled PSUs on a region-by-region basis. The third column shows the percent coverage the ASCS file had. For the 76 counties in our 34 PSUs as a group, the ASCS file had 70 percent as many listings as there were farms counted in the Census of Agriculture. On a region-by-region basis there is quite a bit of variation in this coverage. The ASCS list has good to excellent coverage of Regions 2 through 4, which together contain 70 percent of all farms according to the Census; and fair to poor coverage of the rest of the country. For Region 3, the Midwestern region, ASCS actually has more listings -- 118 percent as many as the Census. For Region 2 (South) and 4 (Central), the ASCS had fairly good coverage -- 90 percent and 79 percent as many listings, respectively, as the

Table G-4. Raw farm count based on sampled PSUs (1982 Census of Agriculture and 1982 ASCS Deficiency File)

Region ¹	Raw counts, sampled PSU's		Percent Coverage ASCS File
	1982 Agriculture Census	1983 ASCS Deficiency File	
1-Northeast	3,743	1,573	42%
2-South	6,619	5,969	90%
3-Midwest	13,367	15,787	118%
4-Central	11,025	8,706	79%
5-Mountain	4,472	2,305	52%
6-Pacific	10,851	504	5%
Continental U.S. Total	50,077	34,844	70%

¹Regions are defined in Figure G-1.

Census. For Regions 5 (Mountain) and 1 (Northeast), the coverage was only about half -- 52 and 42 percent as many listings, respectively, in ASCS as the Census count. Finally, for Region 6 (Pacific), the coverage was very low -- the ASCS list had only 5 percent as many listings as the Census for this region.

Several attempts to understand these discrepancies have met with limited success. The two data sources rely on different bases to get their lists of farms and farm operators, employ different (and to a great extent not thoroughly documented) definitions of "a farm" and have different basic philosophies of the importance of complete coverage. We were able to determine that our ASCS list is a list with one record per farm, as defined by the County Agent, so that the comparison in Table G-4 is the relevant one.

We expected that vegetable, fruit or livestock farms would be at greater risk of under-representation on the ASCS list, so Table G-5 presents the counts of these types of farms by region, with the percent of all farms in the region, based on the 1982 Census. A farm may, of course, have crops in more than one category. For example, a cattle ranch with pastureland would likely also grow feed grain and be eligible for the PIK program. Farms with land in vegetables or orchards might also have PIK-eligible crops, or be on the ASCS File as ineligible. The most striking statistic in Table G-5 is that, while nationally 5.4 percent of farms have land in orchards, in Region 6 (Pacific), 33.7 percent of farms have land in orchards. It seems quite probable that this is a contributing factor to the severe discrepancy between the ASCS frame and the Census in that region. Region 1 (Northeast) has a higher range of farms with vegetables (7% versus 3.1%) than the national average but scarcely enough to account for listing less than half of all farms in that region.

Table G-5. Regional data from 1982 Census of Agriculture on farms with land in vegetables, orchards, and pastureland

Region*	Farms with land in vegetables		Farms with land in orchards		Farms with pastureland	
	Number	Percent	Number	Percent	Number	Percent
1-Northeast	15,458	7.0%	12,740	5.7%	151,287	68%
2-South	19,978	3.6%	28,063	5.1%	355,467	65%
3-Midwest	17,629	2.4%	11,784	1.6%	413,446	57%
4-Central	4,761	1.0%	12,524	2.7%	353,149	76%
5-Mountain	2,858	2.3%	5,271	4.3%	82,766	68%
6-Pacific	7,638	5.0%	51,456**	33.7%	71,679	47%
Continental U.S. Total	68,322	3.1%	121,838	5.4%	1,427,794	64%

*Regions are defined in Figure G-1.

** California has 39,801 farms with land in orchards, including 10,481 with grapes, 7,512 with citrus, 6,119 with avocados, 3,664 with plums and prunes, 2,904 with apples and 2,898 with peaches.

Washington has 6,946 such farms including 5,406 with apples, 2,235 with pears, 2,066 with cherries and 1,042 with grapes.

Oregon has 4,709 such farms including 2,053 with apples, 1,717 with cherries and 1,316 with pears.

The basic pattern in Table G-4 is good coverage to over-coverage in those parts of the country which contain the majority of all farms (Regions 2, 3, and 4 contain 1,739,305 farms, or 78 percent of the total, see Table G-3), and fair to poor coverage in the remainder of the country. This underlying distribution of farms, combined with the pattern of over- and under-coverage and the PSU selection probabilities, results in a fairly decent national estimate of number of farms based on weighted ASCS data, even though the regional estimates are poor. These weighted figures are shown in Table G-6, along with the expected sample size based on weighed ASCS file counts. Regions 1, 5, and 6 continue to lose sample cases due to list undercoverage of those regions.

D. Sampling Frame and Actual Farm Sample

In order to use the ASCS list as a sampling frame, two modifications were made. First, the list of farms was collapsed into a list of farmers by aggregating records with the same name and address. We would thus be able to increase the number of farms sampled without increasing the costs by sampling 600 operators and interviewing them regarding "any farm land you own or operate" in the specific counties they were sampled for. For those few who reported underground storage tanks, we then determined which distinct farms have such tanks and how many. The second frame modification was due to the use of a purchased list for the large establishment segment of the overall survey. Any large establishments with agricultural SICs were removed from the large establishment frame and matched against the ASCS list. If they did not already appear on it, they were added to the frame.

Table G-6. Weighted farm counts from ASCS 1983 File, expected and actual sample sizes

Region ¹	Weighted counts, sampled PSU's		Farm sample size expected from ASCS file ²
	1982 Agriculture Census	1983 ASCS Deficiency File	
1-Northeast	123,714	52,376	15
2-South	283,226	301,055	86
3-Midwest	908,358	1,105,519	314
4-Central	494,029	512,376	146
5-Mountain	147,071	132,621	38
6-Pacific	104,164	5,652	2
Continental U.S. Total	2,060,562	2,109,599	601

¹Regions are defined in Figure G-1.

²These farms are to be screened for the presence of underground motor fuel storage tanks.

From the final frame of farm operators thus established, a sample of 600 cases was drawn with within-PSU probabilities set so that the entire sample had equal probability. Table G-7 reviews the results of farm operators by region, column 1 shows the distribution of farm operators by region, column 2 gives the number of distinct farms this represents, and column 3 shows the farm estimate based on the unadjusted sample weights. Comparing these estimates back to the Census totals in Table G-3, we see that there is quite a bit of region to region variation, although the grand total is fairly close. This indicates that a ratio adjustment would improve the sampling error of estimation for this survey, which we describe in the next subsection.

IV. STATISTICAL ADJUSTMENT OF WEIGHTS TO MINIMIZE SAMPLING VARIANCE

In the previous subsection, it became apparent that the actual sample of farms based on the ASCS list does not accurately reflect the regional distribution of farms as measured by the 1982 Census of Agriculture. Further, in subsection II we found that the underground tank survey regions are very unequal in numbers of farms. In order that our final estimates of number and proportion of farms with underground tanks reflect regional variation and totals more closely, we propose a system of adjustments to the sample weights by region. Since some of the six survey regions have such small sample sizes, we also propose, for farm estimates only, consolidating the survey regions into three areas which have about the same number of farms and which will have over 100 sample cases each. The proposed consolidation is given in Table G-8, which shows the three consolidated regions, their Census totals, the unadjusted sample estimates, and the approximate adjustment factor to apply to the sample weights so that our final sample estimates (of numbers of farms)

Table G-7. Results of farm sample draw

Region ¹	Number of farmers (operators) sampled ²	Number of farms operated	Weighted number of farms using sample weight
1-Northeast	11	17	53,395
2-South	88	94	295,242
3-Midwest	324	354	1,111,868
4-Central	142	159	499,398
5-Mountain	33	33	103,649
6-Pacific	2	2	6,282
Continental U.S. Total	600	659	2,069,834

¹Regions are defined in Figure G-1.

²These farms are to be screened for the presence of underground motor fuel storage tanks.

Table G-8. Consolidated regions for farm estimates and ratio adjustment factors

Regions	Consolidated region	1982 Census of Agriculture	Weighted sample, selection weights	Ratio adjustment factor (rounded)
1&2 - Northeast and Southeast	East	771,025	348,637	2.21
3-Midwest	Midwest	725,699	1,111,868	0.65
4, 5&6-Central, Mountain and Pacific	West of the Mississippi	739,087	609,329	1.21

will equal the Census totals. The actual adjustment was made after the field work had been completed, so that the final number of actual farms contacted was used. After this adjustment, the ratio of largest to smallest weight was about 3.4 to 1, not an excessive gap.

In assessing the quality of the final estimates for farms, for these three consolidated regions and nationally, we have computed sample variances based on the final weights. There is a qualitative aspect to the accuracy as well, in which we acknowledge that coverage of the far West Coast especially is fairly low, and the estimates for the Western consolidated region may contain some bias if these three states are strongly different in terms of underground motor fuel storage from the rest of the west. However, since the West Coast accounts for only 20 percent of farms in Survey Regions 4, 5 and 6, it would have to be extremely different for the survey estimates of this consolidated region to be significantly affected.