

United States
Department of
Agriculture

Statistical
Reporting
Service

**Statistical
Research
Division**

June 1983

AGES830516

Alternate Sampling Frame Construction

For the Prices Paid by Farmers Surveys --An Overview

Carol House
Nancy Carter
Doug Bond

ALTERNATIVE SAMPLING FRAME CONSTRUCTION FOR THE PRICES PAID BY FARMERS SURVEYS--AN OVERVIEW. By Carol House, Nancy Carter, and Doug Bond; Statistical Research Division; Statistical Reporting Service; U.S. Department of Agriculture; Washington, D.C. 20250; June 1983; SRS Staff Report AGES830516.

ABSTRACT

Inadequacies in the existing sampling frames for the Prices Paid Surveys led to an experiment in building lists of business establishments. Farm operators in the Farm Production Expenditure Survey were asked for the addresses of places where they purchased various commodities and for their expenditures at each place. These new frames allowed the sampling of establishments with probabilities proportional to sales to farmers. The study demonstrated the feasibility of building lists in this way, showed that the old frames were incomplete, and gave indications of price differences between the operational and research procedures. This report summarizes the results from the study, leaving technical discussions to a more detailed report, and recommends using a modification of the research procedures in future frame construction and maintenance.

Keywords: sampling frame, incompleteness, price statistics, PPS sampling, rare items

This paper was prepared for limited distribution to the research community outside the U.S. Department of Agriculture. The views expressed herein are not necessarily those of SRS or USDA.

CONTENTS

	<u>Page</u>
SUMMARY	1
INTRODUCTION	3
BACKGROUND	3
DESCRIPTION OF PROCEDURES	6
FEASIBILITY OF DATA COLLECTION PROCEDURES	8
MEASUREMENT OF FRAME INCOMPLETENESS	9
COMPARISON OF PRICE ESTIMATES	10
ESTIMATING RARE ITEMS	13
ALTERNATIVE PROCEDURES	15
CONCLUSIONS AND RECOMMENDATIONS	16
BIBLIOGRAPHY	18

SUMMARY

This study evaluates an alternative method of constructing sampling frames for the Prices Paid Surveys. The method involves asking farm operators on the annual Farm Production Expenditure Survey (FPES) for names and addresses of establishments where they made purchases during the past year and for estimates of their total expenditures at each establishment. The study examined the feasibility of these new procedures, measured the incompleteness of the existing frames and tested for differences in estimates produced by the research and operational procedures. The experiment was generally successful and the authors recommend using a modification of these procedures for future sampling frame construction.

The current operational frames for the Prices Paid Surveys suffer from a number of shortcomings. These include incompleteness, inclusion of firms that do not sell to farmers, high cost of frame construction, and lack of information on size of firm to allow estimation of a price per unit sold to farmers. New list building procedures were evaluated in Minnesota, Mississippi, and Montana during 1980 and in Louisiana, Minnesota, Montana and Pennsylvania during 1981. These procedures succeeded in building lists in a cost efficient manner that gives every firm selling to farmers a positive probability of being surveyed while excluding firms where farmers do not do business.

The analysis shows that the current operational frames for Prices Paid are on average only 50 percent complete. A comparison of price estimates from the operational and research procedures indicates that price differences may exist. This result highlights the central issue: the incompleteness of the operational frames has the potential to lead to serious biases in price estimates in the future.

The results of a simulation study show that the expenditure data from the FPES is not a satisfactory measure of size of firm because its use could lead to large biases in the estimates. The recommendation is to modify the research procedures to estimate an average price per firm, rather than an average price per unit sold as originally designed.

The study demonstrates that the FPES provides a productive way to build and maintain frames for the Prices Paid Surveys with a number of advantages over the operational procedures. The authors recommend using the method in future frame construction but suggest several modifications to the initial research procedures which may reduce cost or improve the reliability of the estimates. Besides the specific modifications discussed in the report there are any number of alternative designs which would follow the general methods described in this study. The authors recommend that such alternatives be evaluated by agency personnel in light of future needs of the Prices Paid program, and that an appropriate modified design be developed and tested.

This report summarizes the results from a more complete technical report on the same topic. It is available upon request from the Statistical Research Division.

Alternate Sampling Frame Construction

For the Prices Paid by Farmers Surveys --An Overview

Carol House
Nancy Carter
Doug Bond

INTRODUCTION

The study evaluates alternative frame construction and estimation procedures for the Prices Paid Surveys. The goals of the study are to examine the feasibility of collecting names and addresses using the new procedures, to measure the incompleteness of the existing Prices Paid lists, and to test for differences between the estimates of price from the research and operational procedures. This paper summarizes the results from a more complete report on the same topic (House and others), giving a basic overview of the study and its results without the technical detail. Readers interested in more information are referred to this technical report.

This paper first gives a background of the Prices Paid Surveys and the problems associated with the current methodology. The new research procedures and the project to evaluate them are described, and the feasibility of the data collection procedures is evaluated. The paper then presents the results from the analysis to measure the incompleteness of the current Prices Paid sample frames, to measure the difference in price between the operational and research procedures, and to examine the problems of estimating rare items. From the results of this analysis, the authors suggest modifications to the procedures tested and then present their conclusions and recommendations.

BACKGROUND

In 1910 the U.S. Department of Agriculture started collecting information on prices paid by farmers for some 86 items used in farm production. This basic program has continued to the present time. Prices are obtained from the sellers to estimate the average price at the time of sale of a wide variety of items farmers actually buy. The Government uses these estimates and the indexes derived from them to make policy decisions about farm programs and to compute parity prices and various income series used as national economic indicators.

The conceptual universe of inquiry for Prices Paid is all purchase transactions used for farm production. Theoretically, the total dollars spent on these transactions would be divided by the number of items purchased to produce an average price per unit sold. To obtain estimates of these prices, the Statistical Reporting Service (SRS) uses a series of surveys called the Prices Paid Surveys based on six major categories of expenditures: 1) feed, 2) fertilizer and pesticides, 3) fuel and motor supplies, 4) farm machinery, 5) autos and trucks, and 6) general farm supplies. For each commodity group, specific production items such as "chick starter feed, per ton bagged," or "12 1/2 gauge

galvanized barbed wire," or "2-wheel drive tractors" are chosen to be representative of related but unsurveyed items. Firms thought to handle these items are contacted regularly throughout the year to provide current prices.

Before 1980 the Prices Paid Surveys were nonprobability surveys. Each state maintained a small, geographically dispersed list of firms for each commodity group. Questionnaires were mailed to all firms on the list, and an average price per reporting firm was calculated. Response was generally low and chronic nonrespondents were eventually removed from the lists. New firm names were added when the total number of firms on the list dropped below a predetermined limit.

This survey methodology created a number of problems. First, with a nonprobability survey the precision of the estimates could not be measured. Second, the sampling frame covered only a fraction of the population of establishments selling to farmers, so that potentially large biases may have affected the estimates. Third, there was a large nonresponse rate on the surveys, further increasing the possibility of biased estimates. Finally, the lack of sample rotation created undue respondent burden and aggravated the nonresponse problem.

Renewed interest in agricultural prices, as evidenced by the emergence of such terms as "target prices," "deficiency payments," and "100% parity" created the impetus to improve this survey methodology. The Agency purchased lists of business establishments from American Business Lists (ABL) to supply a universe frame in each state for each commodity group.^{1/} The current Prices Paid Survey procedures based on these universe frames were implemented in 1980. The procedures called for the selection of a cluster sample of counties in each state and a simple random sample of firms within each selected county for each Prices Paid Survey.

Although these new procedures based the Prices Paid programs on probability surveys, a number of problems were still apparent. First, the universe frame supplied by ABL was believed to be incomplete. Second, units from outside the target population were included in the

^{1/}These lists were constructed primarily from advertisements and listings in yellow pages of telephone directories but were later supplemented by the efforts of SRS state office personnel.

frame. Third, no information was available on the size of the establishment for proportionate sampling or weighting. Finally, the cost of building and maintaining the lists was high. A discussion of these problems follows.

Incompleteness

The target population consists of establishments selling any one of a specified list of farm production items to farmers. Many of these establishments do not advertise in the yellow pages. Of those that do advertise, it is not always clear from the advertisement whether they sell a specific production item. For example, operators of a large drug store may advertise in the yellow pages but probably not in a way that tells if they sell spark plugs. The severity of the incompleteness problem is dependent on the specific item for which a price is to be estimated. Thus a list of firms selling farm machinery should be more complete than a list of firms selling fencing materials or motor oil. These latter items are sold in different types of establishments in almost every town.

Inclusion of Units
Outside the
Population

The target population for the Prices Paid Surveys excludes establishments not frequented by farmers. Therefore such firms should be excluded from the frame. Individual stores do not know how many of their customers are farmers and therefore cannot report this accurately in a screening interview. Although establishments are not expected to differentiate in their prices between farmers and the general public, farmers may buy at different types of stores than the general public and hence pay different prices. Lists of firms obtained from the yellow pages which are dominated by metropolitan area firms will include firms where farmers do not do business. Thus, the population actually surveyed will not be the target population.

Proportionate
Sampling or
Weighting

The purpose of the Prices Paid Surveys is to estimate, for particular production items, the average price per item (or unit) sold to farmers. In order to estimate "price per unit sold" from a sample of firms, the firms should be sampled proportional to sales of the item to farmers, or the prices reported by those firms should be weighted to reflect the size of firm in terms of the quantity of units sold. Otherwise, one estimates an average "price per firm" which does not reflect the quantity of units sold at different prices. Yellow page advertisements do not provide any information to calculate firm size. Therefore the operational procedures are designed to estimate an average price per firm rather than an average price per unit sold.

Cost

Building extensive lists of firms is expensive. Once such a list is built, additional resources must be expended to compensate for the inadequacies of the list building procedures. Also the maintenance effort must begin as soon as the frame is constructed. The Bureau of the Census estimates that each year over a third of all establishments in a monthly Current Business Survey undergo a change such as coming into existence, going out of business, merging, or splitting (Wolter and others). The Prices Paid frame would have to be virtually

DESCRIPTION
OF PROCEDURES

reconstructed each year to maintain a comparable level of completeness.

A straightforward method of identifying the establishments where farmers make purchases is to ask the farmers directly. This inquiry was conducted as part of the Farm Production Expenditure Survey (FPES) in several states during February of 1980 and 1981. The FPES is used to estimate annual expenditures for farm production and to determine the relative importance of various groups of expenditures. The sample is based on a list and area frame to provide efficient, complete coverage of the population of farm operations.

The study covered a variety of geographic areas and commodities over a number of months. During 1980, the study was conducted in Minnesota, Mississippi, and Montana. In the second year it was held in Louisiana, Minnesota, Montana, and Pennsylvania. The commodity groups examined were: farm supplies, fuel and motor supplies, farm machinery, feed, and fertilizer and pesticides.

During the FPES, questions were added at the end of each section where expenditures were recorded for a given commodity group. Respondents were asked to give the names and addresses of all firms where the reported purchases were made. They were then asked, for each group, to report the percent of total expenditures associated with each firm. This procedure constructed a separate list frame of establishments for each of the Prices Paid Surveys.

The research method of list construction has several apparent advantages over the traditional method. First, establishments on an FPES-produced list would be known to have sales to farmers. Second, every firm selling the commodity to farmers would have a positive probability of appearing on the frame. This is because the combined area and list frames used to define the sample for the FPES survey ensure 100 percent coverage of all farm operators, and thus of all firms selling to those operators. The statistical aspects of this are presented in the technical version of the report (House and others). Third, one could easily draw a self-weighting sample of firms with probabilities of selection proportional to the estimated expenditures by farmers. The calculation of these probabilities would require no additional size information from the establishments themselves. Fourth, the cost of constructing the frame would be minimal, the additional time required to collect the names and addresses being the principal component of cost. The second component would be staff hours to complete any incomplete addresses for firms selected in the sample. Finally, since the FPES survey is conducted each year, a frame such as this could be rebuilt annually, if necessary, to insure that the sample of establishments reflects current farmer buying patterns.

After FPES data was collected, state office personnel checked the names and addresses against the operational lists for Prices Paid and coded each firm as overlap or nonoverlap with the operational lists. This overlap check was done by commodity group. For example, a firm listed on an FPES questionnaire under feed would be checked and coded against just the Prices Paid list for feed, regardless of whether it appeared on a Prices Paid list for another commodity group.

The data collected during the FPES and the coding done by the state office personnel provided the information needed to evaluate the feasibility of collecting names and addresses and to measure the incompleteness of the operational lists. In order to measure any price differences between the research and operational procedures, research samples for two commodity groups in each state each year were selected from the FPES produced lists. Firms in these research samples were surveyed at the same time as the regular Prices Paid surveys for the remainder of that year. Price data from these research samples were edited by research personnel following operational editing procedures and summarized using the statistical software package SAS. Table I shows the selection of research samples by state and year.

Table I--Research samples by state and year.

State	1980	1981
Louisiana	--	Fuel and Motor Supplies
	--	Farm Machinery
Minnesota	Feed	Farm Machinery
	Farm Machinery	Farm Supplies
Mississippi	Feed	--
	Farm Machinery	--
Montana	Feed	Feed
	Farm Machinery	Fertilizer and Pesticides
Pennsylvania	--	Fuel and Motor Supplies
	--	Fertilizer and Pesticides

FEASIBILITY OF DATA COLLECTION PROCEDURES

The authors anticipated two problems involving respondent recall errors in collecting names and addresses. First, the operator might completely fail to mention a firm in which he or she had done business. For example, a farm operator might not recall every store where he purchased nails. There was no method available to count the number of establishments omitted in this way, but the supposition is that omitted firms are those with which the farmer had relatively small transactions. Second, the respondent might have trouble giving complete firm names and addresses. For example, a firm by the name of "A&A Feed" might be known locally as "Jones Feed" (the owner's name) or respondents might not recall the street address of a firm. Also, an establishment might be located in a crossroads community sharing a postal designation with a nearby town. The respondent might provide the local community name instead of the postal designation.

Despite the anticipated problems, the experience collecting names and addresses from the FPES was encouraging and demonstrated the feasibility of the method. In written evaluations after the first year of the study, enumerators reported that farm operators generally did not have difficulty listing the places where they made purchases. These evaluations tended to support the supposition that respondents could remember firms where they made their major purchases.

Enumerators reported that respondents sometimes had problems reporting street addresses and percentages of expenditures. If records were available at the time of the interview, the respondent generally had receipts with letterheads containing complete addresses. In many instances when the respondent did not supply a street address, enumerators supplied it using information obtained from telephone directories, post offices and other respondents. Some Montana enumerators reported that street addresses were not used nor needed by firms in many towns. Enumerators reported more problems with collecting percentages than with any other part of the data collection. They reported respondents would generally give their best guess, but they were often unsure.

In the office, most of the names on the FPES-produced list were declared a match or a nonmatch with a name on the operational list without further investigation. When more information was required, it was generally available in a telephone directory. State office personnel reported a rough estimate of 20 staff hours per state to do the overlap check on all commodities. This overlap check was done specifically for the research project to allow the measurement of the incompleteness of the operational lists. It would not be necessary if the FPES lists were used operationally unless they were used in a multiple frame design.

The entire FPES lists were not purged of duplication in the research study because this was not necessary for probability proportional to size (PPS) sampling based on expenditure data. Duplication was only

removed from the sample. However, removal of duplication from the entire FPES lists would probably be necessary if simple random sampling was used instead of PPS sampling. The amount of time needed to unduplicate the FPES lists would probably be comparable to the time needed to perform the overlap/nonoverlap checks.

MEASUREMENT
OF FRAME
INCOMPLETENESS

The collection of names and addresses on the FPES survey created new frames for the Prices Paid Surveys. An analysis was made using the expenditure data associated with each firm and the overlap checking between the FPES-produced frames and the operational frames. This analysis supported the conjecture that a substantial percentage of total farm-related expenditures for feed, fertilizer and pesticides, new farm machinery and farm supplies was made in establishments not identifiable through yellow page advertisements.

Incompleteness of an operational frame should be measured by the relative number of units purchased by farmers in firms not on the frame. This is distinct from the more traditional measurement of the "number of firms" not on the operational frame, but properly reflects the population of interest. Although the actual number of units sold by each firm is not known, the FPES provides expenditure data that allows estimates to be made of relative total expenditures by farmers in firms not on an operational frame. These can be used as a measure of incompleteness.

Estimates of the incompleteness of operational frames were obtained in the following way. For each state, year and commodity group, farmers' expenditures were calculated and expanded for overlap firms and nonoverlap firms. (Overlap firms are those that appear on both the FPES-produced frames and the operational frames.) Estimated expenditures by farmers in nonoverlap firms were calculated as a percent of the total expenditures by farmers (nonoverlap and overlap firms), and used to estimate the percent incompleteness of the operational frames. The commodity groups examined were feed, fertilizer and pesticides, farm supplies and farm machinery. Fuel was not a part of this analysis because there was no statewide operational frame for fuel. For descriptive purposes only, averages across states and commodity groups were generated.

The results in Table 2 show that the operational frames on average are estimated not to cover establishments where 46 percent of farm production expenditures are made. Over the four commodity groups, the average of all state estimates of incompleteness ranged from a low of 38 percent in Louisiana to a high of 52 percent in Minnesota. Even with farm machinery dealers (where yellow page produced lists were expected to be the most complete), state lists ranged from as high as 34 percent incomplete to as low as 13 percent. Farm supply lists were the most incomplete with approximately 75 percent of total expenditures in firms not on the operational lists. This incompleteness

is a serious problem for the operational frame since it creates the potential for substantial bias in the estimates of price.

COMPARISON
OF PRICE
ESTIMATES

The measurement of incompleteness of the operational frames showed that, on average, they were about 50 percent complete. Because of the incompleteness, this study attempted to measure any price differences between the operational and research procedures. The comparisons showed statistically significant, though small, differences between the estimates. Due to limitations in the test results, these comparisons should be viewed as rough indications of the relationship between the research and operational estimates rather than definitive results.

Paired comparisons were made for each state, month, and price item whenever operational and research data were both available. In all there were 841 comparisons. Although several statistics were generated to compare the two procedures, only the more relevant comparisons will be presented here. Further details are available in the technical report.

Table 2--Estimated percent incompleteness of the operational sampling frames.

State	Feed		Fertilizer and Pesticides		Farm Supplies		Farm Machinery		State Average
	1979	1980	1979	1980	1979	1980	1979	1980	
Minnesota	60	27	54	59	81	72	28	34	52
Mississippi	40	--	48	--	78	--	29	--	49
Montana	28	55	38	52	56	78	13	23	43
Louisiana	--	7	--	47	--	74	--	23	38
Pennsylvania	--	26	--	58	--	86	--	20	48
Average	43	29	47	54	72	77	23	25	46

Table 3--Comparison of 90 percent confidence intervals.

Data Class	Number of Comparisons	Percent of Comparisons With Intersecting 90% Confidence Intervals
--By Commodity Group--		
Farm Supplies	39	95
Fuel and Motor Supplies	82	80
Machinery	91	84
Feed	566	78
Fertilizer and Pesticides	63	76
--By State--		
Louisiana	48	64
Minnesota	236	93
Mississippi	137	79
Montana	319	70
Pennsylvania	101	81
Total	841	79

To compare the estimates of price, 90 percent confidence intervals are calculated for the research and operational estimates and the percentage of paired intervals which intersect are computed. If the two procedures are in fact measuring the same price one would expect the confidence intervals to intersect at least 90 percent of the time.

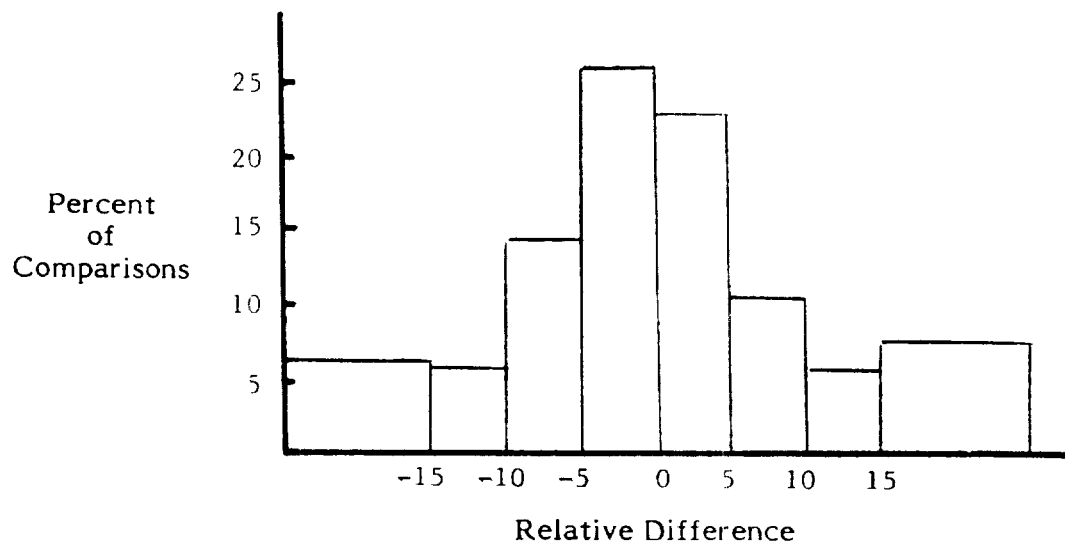
In the comparison of confidence intervals over all estimates, they intersect only 79 percent of the time. (See Table 3.) Thus, the number of disjoint confidence intervals is about twice what one would expect from random chance. At the state level, all states except Minnesota show a significant difference between the estimates of price. The intervals intersect 93 percent of the time in Minnesota. In other states the percentages range from 64 to 81 percent. When the comparison is made by commodity groups, the results are similar. The only commodity group not showing a statistical difference in price is farm supplies, where 95 percent of the confidence intervals intersect. The percents of intersection for other commodity groups range from 76 to 84 percent.

These comparisons give a rough indication that price differences exist between the operational and research procedures. There are, however,

limitations to the inferences that should be made from these tests. First, the validity of the price estimates and confidence intervals is questionable due to the small number of responses on which these estimates are based. Second, the possibility of assumption violations in probability proportional to size sampling is increased by the low response rate, leading to the introduction of possible biases into the estimates. Both issues are examined in the next section and in more detail in the technical report. Even if one accepts the validity of the tests and rejects the hypothesis that the estimates are the same, one is left with two possibilities: 1) there is a difference in price between firms frequented by farmers and firms not frequented by farmers, or 2) the tests reflect the difference in the "price per firm" estimator used in the operational program versus the "price per unit" estimator used in the research procedures. The results are likely a combination of the two.

Even though most commodity and state comparisons show significant price differences, the relative sizes of these differences are small. Figure 1 shows this graphically. The horizontal axis gives the difference in price (research minus operational) as a percent of the operational estimate of price, and the vertical axis gives the percent of all comparisons falling into each category. Approximately 50 percent of all comparisons have a relative difference less than five percent, although the research price estimates tend to be slightly less than the operational.

Figure 1--Relative price differences between research and operational procedures. 2/



2/ Graphed as the percent of all comparisons that show relative differences in the specified ranges.

$$\text{Relative Difference} = \frac{\text{Research Price} - \text{Operational Price}}{\text{Operational Price}} \times 100\%$$

ESTIMATING RARE ITEMS

In conclusion, the comparisons presented in this section, although subject to the limitations discussed, indicate that significant, but small, differences exist between the research and operational price estimates for most state and commodity groups. These findings provide evidence of potential deficiencies in estimating price from the current operational sampling frame, given its estimated level of incompleteness.

In this survey most firms returning a questionnaire reported a price for only a small portion of the surveyed items. This probably indicated that firms did not carry the omitted items. Since the analysis shows that over 50 percent of the operational estimates and almost 40 percent of the research estimates are based on less than 10 responses, many of the surveyed items are considered "rare items." Estimation of "rare items" can cause serious problems with the estimation procedures.

Table 4 gives a summary of the number of observations going into each estimate from the research and operational samples. It shows that 30 percent of the operational estimates and 21 percent of the research estimates were made with 5 or fewer observations. The frequency with which operational estimates are based on 10 or fewer observations is 14 percentage points higher than the frequency of the research estimates. A possible explanation for this is that firms known to be frequented by farmers are more likely to carry a wide range of products used by farmers. Hence, they provide greater item response than firms identified through other methods.

Estimating rare items creates a number of problems. First, estimates based on a small number of observations are generally more variable than those based on a greater number. Second, the estimators of price are ratio estimators. These estimators have a bias associated with them which is negligible for large sample sizes but could be substantial when dealing with sample sizes as small as those being used for Prices Paid. Third, the existence of so many rare items has indicated that the estimates of firm size based on farmer-reported expenditures of general commodity groups such as feed are probably too gross to provide appropriate weights when estimating prices of specific items such as hog concentrate.

A simulation study was conducted to see if poor estimates of firm size from the FPES affect the estimation of price in the research procedures. The results from this study indicate that unbiased estimation of relative size of firm leads to estimates of price that are not statistically different from what they would be if one knew the actual firm size. This is true even though the standard error of the estimate of size is very large. However, if the estimate of relative size of firm is biased by as little as 10 percent, then statistically different estimates of price may result. The large number of rare items on the surveys indicates that the estimates of size may very well be biased by

Table 4--Relative frequency distribution of the number of observations used in computing price estimates. 3/

Data Class	Percent of "Research/Operational" Estimates Made With the Indicated Number of Observations							Total
	1-5	6-10	11-15	16-20	21-25	26-30	> 30	
	--By Commodity Group--							
Farm Supplies	3/18	5/15	18/21	26/15	18/8	8/15	22/8	100%/100%
Fuel	13/17	13/21	23/13	21/13	5/7	5/17	20/12	100%/100%
Farm Machinery	18/33	22/34	20/10	13/10	14/11	11/2	2/0	100%/100%
Feed	26/34	21/24	20/15	12/9	6/8	9/5	6/5	100%/100%
Fertilizer	2/10	1/8	6/24	11/14	16/16	11/11	53/17	100%/100%
	--By State--							
Louisiana	21/27	13/35	27/17	19/13	8/4	10/4	2/10	100%/100%
Minnesota	16/19	22/23	16/15	14/11	13/12	8/10	11/10	100%/100%
Mississippi	27/39	16/28	22/15	13/9	7/8	12/0	3/1	100%/100%
Montana	28/39	21/24	21/16	14/9	5/6	8/4	3/2	100%/100%
Pennsylvania	5/15	7/10	19/14	19/11	10/12	11/19	29/19	100%/100%
Total	21/30	18/23	20/15	15/10	8/9	9/7	9/6	100%/100%

3/ Cell entries are of the form "percent of row total for research samples/percent of row total for operational samples."

10 percent or more. Therefore the authors conclude that the research procedures should be modified in the future to estimate an average price per firm (as is currently done in the operational procedures) to avoid biased estimates of price per unit sold.

The following is a brief description of the simulation study. The authors chose a data set of reported prices from the research study, assumed that the estimates of firm size (or weights) from the FPES were correct, and estimated average price. Next, a set of alternative weights were generated from a distribution having the same mean as the original weights, but with positive deviations from that mean. The average price was then calculated using these new weights. Fifty sets of alternative weights were generated, forming 50 estimates of average price and a confidence interval around their mean. This confidence interval was used to determine whether the estimate of price using the original weights was statistically different from the prices using the generated weights.

The simulation study was repeated using two different sets of price data: 1) September 1980 prices for large self-propelled combines in Minnesota, and 2) September 1980 prices for hog feed in Minnesota. Truncated normal and chi-square distributions with standard deviations ranging from 0.3 to 1.0 were used for the alternative weights. The means of the distribution were first set to that of the original weights and later adjusted by ten percent. The results were consistent for all simulations. If the simulated weights were distributed with a mean equal to the original weights, then the original estimate of average price was not statistically different from the simulated averages. However, if the mean was different so was the estimate of price.

In conclusion, the large levels of item rarity in both the research and operational samples cause the validity of many of the estimates of price and their standard errors to be questioned. Any inferences from estimates or tests discussed in this report should be made with these limitations in mind.

ALTERNATIVE PROCEDURES

The authors suggest modifications in the initial research procedures to reduce cost and improve the reliability of the estimates. In one possible alternative design the FPES is used as a tool to gather names and addresses for a more traditional sample frame. This list is unduplicated and a simple random sample is selected statewide. The major changes of this modification are the avoidance of probability proportional to size sampling and the avoidance of collecting names every year. The FPES is then used as needed, perhaps every two to four years, to update this frame.

The statistical advantages of this alternative are similar to those in the initial research design. The changes will eliminate the problems associated with the estimation of size of firm and provide greater

stability to the price estimates than is possible when the frame is completely changed each year. However, the estimate will be of an average price per firm rather than an average price per unit. Although the frame still will be reflective of farmers' buying patterns, changes will be incorporated into the frame only during the years when updates are done. Thus, a firm will have a positive probability of being selected on a price survey only if it sold merchandise to farmers during an update year.

There are trade-offs between the two research alternatives in the costs of building and maintaining the sampling frame. Under the modified design, updating would not occur yearly, thus reducing costs and respondent burden. However, the names and addresses would have to be unduplicated whenever building or updating occurred. Also, the entire frame would have to be coded and keypunched. Under the initial research design it is only necessary to code the sampled firms. Hence, these modifications require additional staff resources.

CONCLUSIONS AND RECOMMENDATIONS

The incompleteness of the operational sample frames ranged from a low of 24 percent for farm machinery to a high of 75 percent for farm supplies. This incompleteness is a serious problem for the operational frame since it creates the potential for price differences between the operational and research frames.

The tests showed significant differences in estimated prices using research and operational procedures for most states and commodity groups. However, the relative differences were generally less than 5 percent, and the number of responses used to calculate the estimates was generally so small that the results of tests in this report were only reliable enough to say that there are indications that price differences exist. Limitations in the testing, however, should not distract from the central issue that the operational frames are seriously incomplete. This incompleteness has the potential to lead to serious biases in price estimates in the future as the operational frames, newly constructed at the time of this study, get older. Even if testing showed no significant differences, the incompleteness of the operational frame is a serious handicap because of the potential it creates for errors in estimates.

Thirty percent of all operational estimates examined were made with five or fewer observations. Over half were made with ten or less. The research estimates were somewhat better (39 percent with 10 or fewer observations) but still not acceptable. Insuring an adequate number of responses to estimate properly all important items surveyed should be a primary consideration in any discussion of changes of the frame and/or survey design for the prices paid program. A simulation study measured the sensitivity of the research procedure estimates of price to the estimation of relative firm size. The results showed that firm size must be estimated unbiasedly in order to avoid significant effects on

the estimate of price. This sensitivity to bias causes the estimator of price per unit to appear less useful than an unbiased estimator of price per firm.

The modified sample design which involves simple random sampling of firms generated from the FPES seems to be a reasonable alternative. Although it does not allow for estimating price per unit sold, the estimate of price per firm is based on firms where farmers are known to do business. It eliminates the most serious statistical defect that was discovered in the original research design. The cost for frame building and maintenance would probably be less under the modified plan than under the original design.

The authors conclude that the FPES is a productive tool to build and maintain a frame of establishments for the Prices Paid Surveys. Besides the specific alternative design discussed in this paper, there are a number of other possible ways of using the FPES to build and maintain these frames. The authors recommend that alternatives which follow the same general methods described in this study be evaluated by agency personnel in light of future needs of the Prices Paid program, and that an appropriate modified design be developed and tested.

BIBLIOGRAPHY

- [1.] Cochran, William G. Sampling Techniques, 3rd Edition. New York: John Wiley & Sons, Inc. 1977.
- [2.] House, Carol C. Evaluating New Reporter Sources For a USDA Price Survey. U.S. Department of Agriculture, Economics, Statistics and Cooperatives Service, Statistical Research Division, 1980.
- [3.] House, Carol, Carter, Nancy and Bond, Douglas. Alternative Frame Construction For the Prices Paid By Farmers Survey--Technical Report. Unpublished Staff Report. U.S. Department of Agriculture, Statistical Reporting Service, Statistical Research Division, 1983.
- [4.] Kish, Leslie. Survey Sampling. New York: John Wiley & Sons, Inc., 1965.
- [5.] SAS Institute Inc., SAS User's Guide, 1979 Edition. Raleigh, NC: SAS Institute Inc., 1979.
- [6.] Wolter, Kirk M., and others. Sample Selection and Estimation Aspects of the Census Bureau's Monthly Business Surveys. U.S. Bureau of the Census.