United States Department of Agriculture

National Agricultural Statistics Service

Research and Applications Division

SRB Research Report Number SRB-90-12

September 1990

RESPONSE BIAS IN THE QUARTERLY AGRICULTURAL SURVEY HOG DATA: DECEMBER 1987 AND MARCH 1988

Vic Tolomeo

RESPONSE BIAS IN THE QUARTERLY AGRICULTURAL SURVEY HOG DATA: DECEMBER 1987 AND MARCH 1988. by Vic Tolomeo and Gretchen McClung. Research and Applications Division, National Agricultural Statistics Service, U.S. Department of Agriculture, Washington, D.C. 20250. July, 1990. NASS Research Report No. SRB 90-12.

ABSTRACT

Reinterview projects were conducted for the December 1987 and March 1988 Agricultural Surveys. The purpose was to measure the quality of the survey data by the estimation of response bias for selected crops, grain stocks, and hog numbers. A subsample of the Agricultural Survey CATI sample was recontacted through face to face reinterviews. Differences between the original and reinterview responses were reconciled to determine a final value. This paper presents results of hog bias estimates using the final reconciled value as a proxy for the truth. Both multivariate and univariate test results indicated statistically significant levels of negative hog bias or underreporting of hog numbers. States involved in the December 1987 Survey were Indiana, Minnesota, and Ohio. States from the March 1988 survey were Iowa, Nebraska, and Pennsylvania.

ACKNOWLEDGEMENTS

Special thanks to the State Statistical Offices, headquarter's staff, enumerators and supervisory enumerators who participated in these projects.

TABLE OF CONTENTS

SUMMARY

Face to face reinterviews of a subsample of Computer Assisted Telephone Interviewing (CATI) respondents were conducted in three states each for the December 1987 and March 1988 Agricultural Surveys. The purpose was to evaluate the quality of the survey data by estimating response bias. Enumerators reasked specific crop and hog survey questions, compared the reinterview responses to the original telephone responses, and reconciled differences to determine a final value. Bias estimates were computed under the assumption that the final reconciled value was a valid proxy for the true value. This paper focuses on hog bias estimation and characteristics.

Both multivariate and univariate tests were used to determine if the hog bias was significantly different from zero. Multivariate test results were significant at the combined state level for both survey periods indicating negative biases or underreporting of hog inventory items. Univariate tests were significant for December 1987 only. Total hog bias was estimated at -11.7 percent (p-value=.01) in December 1987 and -4.5 percent (p-value=.08) in March 1988.

As in previous studies, larger biases were associated with respondents other than the operator, indicating that the magnitude of the bias is influenced by the type of respondent. Although nearly 90 percent of the original CATI respondents were the operator, changes in the mix of respondents could have a strong influence on the level of bias.

The reinterview procedures limited the time between the original interview and reinterview to 10 days. Within this interval, time did not appear to influence the magnitude of the bias.

The bias data was shown to consist mostly of zeros with relatively few large negative values. This presents problems in terms of point estimation and the level of precision. Negative biases were indicated in both survey periods but the precision of the estimates was low. The CV's of the total hog bias estimates were 40.5 percent in December 1987 and 57.6 percent in March 1988. Therefore, the precision of the estimates may be lower than desired for direct adjustment of survey indications.

The recommendation is that the coverage of the reinterview project be expanded. Efforts should be directed towards detecting and monitoring bias level changes over time. Analysis of the reasons for the biases can be used to develop improvements in the questionnaire design and survey training.

INTRODUCTION

The National Agricultural Statistics Service (NASS) has conducted several reinterview projects designed to measure the quality of data collected during the Agricultural Surveys (AS). The specific purpose of these projects was to measure response bias for selected crops, grain stocks, and hog inventory items. This paper presents results of bias estimates for hog inventory items from the December 1987 and March 1988 AS. The states involved in the December 1987 reinterview project were Indiana, Minnesota, and Ohio. Those in the March 1988 project were Iowa, Nebraska, and Pennsylvania.

Earlier papers reported negative biases or underreporting of grain stocks [4,5]. They also indicated that 48-75 percent of the estimated biases were associated with definitional type problems. Readers should refer to these reports for a more complete description of the reinterview sample design, specific reinterview procedures, training manuals, reinterview questionnaires, and reconciliation forms. Reference material related to response errors can also be found [5].

REINTERVIEW PROCEDURES

The reinterview projects were designed to measure response bias through the use of face to face reinterviews of operations originally interviewed using Computer Assisted Telephone Interviewing (CATI). The reinterview/reconciliation technique used is similar to that used by the U.S. Census Bureau, although the focus is on bias rather than response variance or consistency of response. The assumption is that the final reconciled value represents the true value, or at least a better measure of the true value.

The procedures used by NASS involve supervisory or experienced field interviewers for 1) face to face reinterviewing of a subsample of the Agricultural Survey CATI sample conducted within 10 days of the original CATI interview, 2) reconciling differences between the original CATI responses and the reinterview responses to determine the correct or 'true' values, and 3) determining reasons for differences.

Questionnaires used in the reinterview were similar to the AS questionnaires. However, not all questions asked on the original interview were reasked on the reinterview. The goal of the reinterview was to obtain the best possible information regarding a subsampled operation by contacting the most knowledgeable person. It was not to recontact the same individual originally interviewed by CATI.

STATISTICAL MEASURES

Bias and variance estimates were based on a stratified sample design. Estimates and tests of significance were computed for both the original

CATI data and the corresponding Survey Processing System (SPS) edited CATI data.

For the i^{th} observation in stratum h, bias was measured as

$$B_{hi} = X_{hi} - T_{hi} \qquad \text{stratum } h = 1,\ldots,L \text{ and unit } i = 1,\ldots,n_h$$

where X_{hi} = original CATI or SPS edited response
T_{hi} = final or reconciled value

A negative bias indicates underreporting of the survey item. Both univariate and multivariate test procedures were used in the analysis (See Appendix A).

THE SAMPLE

The reinterview subsample was drawn from the list portion of each state's AS sample completed on CATI. CATI completed samples eligible for reinterview included completed interviews, out-of-businesses, and interviews with item refusals or item don't knows. Questionnaire refusals were not eligible for reinterview. List strata not placed on CATI and area nonoverlap tracts were also not eligible for reinterview.

Reinterview sample sizes and response rates by state and survey period are shown in Table 1. The percent of each state's total Agricultural Survey list sample completed on CATI is also shown.

Table 1. State reinterview sample sizes and response rates.

	Response				% AS list completed
Survey/State	Completed	Refusals	Inaccessible	n	on CATI
December 1987					
Indiana	82%	3%	15%	369	57%
Minnesota	87%	6%	7%	402	31%
Ohio	89%	3%	8%	313	61%
Total	86%	4%	10%	1084	49%
March 1988					
Iowa	85%	4%	11%	357	48%
Nebraska	74%	5%	21%	483	49%
Pennsylvania	89%	3%	8%	320	71%
Total	82%	4%	14%	1160	52%

RESULTS

Response Bias

Bias estimates for breeding, market, and total hogs for the two survey periods are summarized in Table 2. Bias estimates are shown for the original CATI data and the corresponding SPS edited CATI data. Both multivariate and univariate test (H_o: Bias = 0) results are shown. Bias estimates for market hogs included the market hogs plus hogs no longer used for breeding. Statistically significant biases at α equal to .05 are indicated by an asterisk. Levels of significance close to .05 are indicated in parentheses.

Multivariate testing included breeding and market hogs but excluded total hogs. This was done since a multivariate test solution is difficult to obtain if one component approximates the sum of the other components. In this case, total hogs is the sum of the breeding, no longer used for breeding, and the market weight categories.

Multivariate tests were significant for both December 1987 and March 1988. Univariate tests were significant in December 1987 but not in March 1988.

Table 2. Combined state bias estimates for breeding, market, and total hogs by survey.

Survey	Original CATI - reconciled		SPS edited CATI - reconciled	
	Number	Percent	Number	Percent
December 1987				
Breeding	-111,201	-10.7 *	-111,201	-10.7 *
Market	-788,887	-11.4 *	-819,977	-11.9 *
Total[1/]	-946,181	-11.9 *	-930,804	-11.7 *
Multivariate test p-value < .05				
March 1988				
Breeding	-41,856	-3.7	-34,511	-3.0
Market	-296,157	-4.3 (.13)	-320,088	-4.7 (.10)
Total[1/]	-344,147	-4.3 (.10)	-360,541	-4.5 (.08)
Multivariate test p-value < .05				

* indicates significant bias at α = .05

[1/] Not the sum of breeding and market due to item refusals and item don't knows.

Table 3 shows bias estimates for total hogs by state and combined states by survey. With the exception of the original CATI data in Pennsylvania, a negative total hog bias was estimated for all states. A single observation accounted for the difference between the estimated positive original CATI data bias and the negative SPS edited CATI bias in Pennsylvania.

Table 3. Bias estimates for total hogs.

Survey/State	Original CATI - reconciled			SPS edited CATI - reconciled		
	Number	Percent		Number	Percent	
December 1987						
Indiana	-504,209	-10.8	(.08)	-457,582	-9.7	(.11)
Minnesota	-222,050	-8.6	*	-230,325	-9.0	*
Ohio	-219,922	-33.2		-242,897	-37.8	
Total	-946,181	-11.9	*	-930,804	-11.7	*
March 1988						
Iowa	-88,167	-1.9		-88,167	-1.9	
Nebraska	-267,564	-9.7	*	-267,564	-9.5	*
Pennsylvania	11,584	2.0		-4,810	-.9	
Total	-344,147	-4.3	(.10)	-360,541	-4.5	(.08)

* indicates significant bias at $\alpha = .05$

Reinterview respondents were asked weight breakdowns for market hogs in Minnesota only for the December 1987 survey and in Iowa and Nebraska for the March 1988 survey. A general question that asked for total market hogs was used in the remaining states. (See Appendix B). Table 4 summarizes the combined survey weight category bias estimates for Iowa, Minnesota, and Nebraska. Multivariate test results were significant at $\alpha = .10$ for the combined survey period level. Univariate tests were significant for the smallest and largest weight categories. Separate multivariate tests by survey period, Minnesota in December 1987 and Iowa/Nebraska in March 1988, were not statistically significant.

Table 4. Combined December 1987 and March 1988 bias estimates for market weight categories. (Iowa, Minnesota, and Nebraska).

Weight Category	n	Original CATI - reconciled Number	Original CATI - reconciled Percent	SPS edited CATI - reconciled Number	SPS edited CATI - reconciled Percent
0 - 60 lbs.	161	-218,929	-7.4 *	-233,149	-7.9 *
61 - 119 lbs.	158	50,793	2.2	59,280	2.5
120 - 179 lbs.	139	-94,886	-5.6	-95,896	-5.6
over 180 lbs.	139	-212,587	-13.8 (.06)	-220,231	-14.2 *

Multivariate p-value < .10

* indicates significant bias at $\alpha = .05$

From the above test results we can conclude that respondents tend to underreport their hog numbers. Both the original CATI response data and the corresponding SPS edited data underestimate hog numbers. In addition, larger biases were reported for the smallest and largest market weight categories.

The precision of the total hog bias estimates by survey period is shown for the SPS edited CATI data in Table 5. The percent bias variance formula is shown in Appendix C. The large coefficients of variation (CV) indicate that, although we are able to conclude that biases exist, the precision of the estimates are low.

Table 5. Precision of total hog bias estimates by survey (SPS edited CATI data).

Survey	Estimated Percent bias	Standard Error	CV (%)
December 1987	-11.7	4.7	40.5
March 1988	-4.5	2.6	57.6

Bias Characteristics

Table 6 shows the frequency of differences between the original and reinterview responses for total hogs. Differences, positive or negative, occurred in approximately 16 percent of the reinterviews. Although the frequencies of positive and negative differences were approximately the same, in terms of magnitude the average negative difference was nearly twice as large. The range or variability of negative differences was also greater.

Table 6. Combined December 1987 and March 1988 frequency of differences between the original and reinterview responses for total hogs.

Difference	Frequency	Percent	Average Difference	Range
CATI < reinterview	161	8.8	-101.0	1270
CATI > reinterview	131	7.1	59.8	692
CATI = reinterview	1541	84.1	--	--
Total	1833	100.0		

As shown in Table 7, the reinterview response was determined to be the correct or final response in 62 percent of the reinterviews with differences, compared to 20 percent for the CATI response. About 14 percent of the differences resulted in a third value being defined as the final value. The remaining categories included instances where the respondent indicated that both responses were estimates and equally likely to be correct or where a final value was not reconciled.

Table 7. December 1987 and March 1988 combined two-way frequency table of reinterview differences by correct response for total hogs. 1/

	Correct response				
Difference	CATI	Reinterview	Third	Either	Missing
CATI < Reinterview	23	110	24	4	0
CATI > Reinterview	36	70	16	7	2
Total	59 (20.2%)	180 (61.6%)	40 (13.7%)	11 (3.8%)	2 (.7%)

1/ Includes only observations with a difference between the original and reinterview response.

Reinterview procedures required enumerators to complete their reinterviews within 10 days of the original CATI interview. This was done to minimize recall problems associated with the first of the month reference date. The average time between the original CATI interview and the reinterview was 6.2 days in December 1987 and 6.4 days in March 1988.

Table 8 summarizes the relationship between the magnitude of the bias, positive or negative, and the time between interviews. The proportion of relatively large biases is approximately the same in the 3-5 and 6-8 day intervals and it is actually less in the 9-10 day interval. Therefore, it does not appear that time was a factor affecting the bias, at least within the 10 day required reinterview time.

Table 8. Combined December 1987 and March 1988 two-way frequency table of relative bias for total hogs by time (days) between original interview and the reinterview. 1/

Relative bias 2/ Absolute value	Number of days between Interviews 3 - 5	6 - 8	9 - 10	Total
less than 10%	34 (42%)	52 (43%)	13 (43%)	99
10% - 20%	16 (20%)	19 (16%)	9 (30%)	44
Greater than 20%	30 (38%)	49 (41%)	8 (27%)	87
Total	80 (100%)	120 (100%)	30 (100%)	230

Table 9. Combined December 1987 and March 1988 two-way frequency table of relative bias by respondent category for total hogs. 1/

Relative bias 2/ Absolute value	Respondent Combinations Operator-operator	Other-operator	Other
Less than 10%	85 (44%)	7 (41%)	7 (35%)
10% - 20%	40 (21%)	2 (12%)	2 (10%)
Greater than 20%	68 (35%)	8 (47%)	11 (55%)
Total	193 (100%)	17 (100%)	20 (100%)

1/ Includes only observations with a bias.
2/ Relative bias =

$$100 * \frac{\text{(CATI value - reconciled value)}}{\text{reconciled value.}}$$

Finally, Table 9 shows the relationship between respondent combinations and the magnitude of the bias. Combinations of the original respondent and the corresponding reinterview respondent were grouped into the three categories shown above. For example, the operator-operator category includes operator-same operator and partner-different partner. This category accounted for nearly 81 percent of the completed reinterviews and 84 percent of the reinterviews with reported hog differences. The data suggests that larger biases were associated with the other-operator and other combinations of respondents compared to the operator-operator combination. However, with so few observations in the other-operator and other categories, the results were not statistically significant. Similar results had been reported for grain stocks biases [4,5].

Proportion of Farm Population Studied

The CATI reinterview subsample represents a restricted portion of each state's population of farm operations since the nonoverlap area tracts and some list strata were not eligible for reinterview.

Table 10 shows the proportion of each state's operational list expansion for total hogs represented by the CATI list strata and eligible for reinterview. As shown in the table, CATI representation of list strata varies from 92% in Indiana to 40% in Iowa.

Table 10. Proportion of operational list expansion represented in the reinterview project by state and survey period for total hogs.

Survey/State	percent[1/]	State % U.S. Board Est.
December 1987		
Indiana	92.0	8.5
Minnesota	63.7	8.2
Ohio	79.4	3.9
Total	78.4	20.6
March 1988		
Iowa	40.0	25.7
Nebraska	72.9	7.4
Pennsylvania	83.3	1.6
Total	49.9	34.7

1/ (Reinterview strata Direct Exp. ÷ Total List Direct. Exp.) x 100

Assumption of Normality

A goal of survey sampling is to determine a sample size large enough so that the sampling distributions of the means are approximately normal regardless of the underlying distribution. A standard t or F test of the null hypothesis, in this study that the expected value of the bias is equal to 0, is valid if the assumption of near normality for the mean bias is correct.

This assumption is particularly likely to be invalid when a few large or extreme values are present. Cochran describes the situation for positively skewed data and the resulting impact on the estimates and confidence intervals [1]. He also presents a general rule under simple random sampling which can be used to determine the sample size needed to satisfy the normality assumption.

Kott has applied this general rule to stratified sample designs, where the focus is on the distribution of the stratified sample mean [3]. The coefficient of skewness, $G_{\bar{x}}$, is computed to measure the impact of

extreme values on the sample distribution. (See Appendix D). As a general rule, if $|G_{\bar{x}}|$ is less then .2 then the stratified mean can be treated as a normal random variable.

Table 11 shows values of $G_{\bar{x}}$ by survey and state for total hogs. Values of $|G_{\bar{x}}|$ greater than .2 were computed for both survey periods. As a result, the estimated confidence interval for the bias is shifted towards zero and slightly narrower than the true confidence interval for a given α level.

Outliers were detected by identifying the observations within strata which contributed most to the value of $G_{\bar{x}}$. Three outliers identified in the December 1987 data included one each from strata 69 and 72 in Indiana and one from stratum 65 in Ohio. In March 1988, one outlier was identified in stratum 67 from Nebraska. The state/strata contributions to the computed $G_{\bar{x}}$ at the 3-state level by survey period are shown in Appendix D, Tables 1 and 2.

Table 11 also shows the impact of the outliers on the value of $G_{\bar{x}}$ and the estimated bias. P-values are indicated in parentheses. Deleting the outliers improved the normality of the data resulting in values of $|G_{\bar{x}}|$ less than .2 and, although reduced, significant negative biases were still estimated.

Table 11. Computed values of $G_{\bar{x}}$ by state and all states combined for December 1987 and March 1988 total hogs.

	Before outliers deleted		After outliers deleted	
State	$G_{\bar{x}}$	Total hog % bias	$G_{\bar{x}}$	Total hog % bias
December 1987				
Indiana	-0.496	-9.7% (.11)	0.140	-2.0%
Minnesota	-0.546	-9.0% (.02)	-0.546	-9.0% (.02)
Ohio	-0.940	-37.8%	-0.434	-9.7% (.10)
Total	-0.425	-11.7% (.01)	0.039	-4.9% (.02)

	Before outliers deleted		After outliers deleted	
State	$G_{\bar{x}}$	Total hog % bias	$G_{\bar{x}}$	Total hog % bias
March 1988				
Iowa	-0.144	-1.9%	-0.144	-1.9%
Nebraska	-0.498	-9.5% (.05)	-0.228	-5.8% (.07)
Pennsylvania	0.027	-0.9%	0.028	-0.9%
Total	-0.301	-4.5% (.08)	-0.124	-3.2% (.15)

Frequency tables of differences between the CATI response and the final reconciled value illustrate the skewness of the bias distribution for total hogs. (See Appendix E).

DISCUSSION AND RECOMMENDATIONS

NASS has conducted four reinterview projects for the AS since December 1987. The purpose was to estimate biases, if they existed, and to identify reasons for their occurrence. Information derived from these studies could then be used to adjust the operational survey indications, to suggest improvements in the operational survey methods, and to monitor survey quality over time [5]. This paper has focused on hog bias estimation with respect to the reinterview population coverage, precision of the bias estimates, and the distribution of the observed biases. The analysis of reasons for differences will be presented in a future report combining the results of all four reinterview projects.

The reinterview project did not include the AS nonoverlap area sample, although area nonoverlap tract operations are now on CATI. The selection of list strata to place on CATI depends on each state's operational sample, CATI capacity, and particular data collection procedures. For the reinterview states, the proportion of the list sample completed by CATI, and therefore eligible for reinterview, was 49 percent in December 1987 and 52 percent in March 1988 (Table 1). Individual state values ranged from 31 percent to 71 percent.

A more descriptive measure of the reinterview coverage of the population of hog farm operators is the proportion of the AS list strata expansions represented in the reinterview sample (Table 10). The reinterview proportion of the total hog list direct expansion was 78 percent in December 1987 and 50 percent in March 1988. State values ranged from 40 percent to 92 percent. Iowa, with the lowest coverage at 40 percent, accounted for 26 percent of the U.S. estimate for total hogs. As a result, these studies are only directly applicable to the population of farm operators who are represented on CATI and eligible for reinterview.

The current reinterview procedures and sample were designed to test whether or not biases exist for major crop and grain stocks items. Sample sizes were sufficient to detect biases in the hog data; however, the precision of the estimates may be too low for direct adjustment of the operational survey value. The coefficient of variation (CV) of the total hog percent bias is 40.5 percent for December 1987 and 57.6 percent for March 1988. The 95% confidence intervals for the total hog percent bias were -21.0 percent to -2.4 percent in December 1987 and -13.3 percent to +4.4 percent in March 1988.

The large CV's are in part the result of the highly skewed nature of the bias data. Twenty-three percent of the reinterview sample reported hogs. Two-thirds of these showed a difference between the original and reinterview responses. However, over the entire reinterview sample a bias was observed in only 13 percent of the reinterview sample units. A positive or negative bias greater than 20 percent was observed in five percent of the reinterview sample units. Small strata sample sizes together with the skewness of the data resulted in the relatively

large variance estimates. As a result, sample sizes would need to be increased approximately 10 times if a 95% confidence interval of ± 2 percent is desired.

Such highly variable data makes interpretation of survey to survey changes in the bias estimates difficult since they may be the result of sample variability rather than any real changes in bias levels. Emphasis needs to be placed on detecting major shifts in the bias level over time.

The following are recommended:

1. The reinterview sample states and strata should be increased to improve population coverage and precision of the estimates. The projected increase in the number of CATI states would provide the additional states for study. New states should be selected based on their proportion of the national estimate for large bias items.

2. Include the nonoverlap area tract samples placed on CATI in the reinterview study. This would increase the coverage of the bias estimates, providing information about a portion of the population not currently under study. The nonoverlap domain accounts for 15 - 18 percent of the total hog multiple frame indication for the 10 Quarterly Hog States.

3. A data series of bias estimates should be developed to identify major changes in bias levels over time. The application of statistical process control procedures can help identify an actual shift in the bias based on previous reinterview estimates.

REFERENCES

1. Cochran, W. G. (1977). Sampling Techniques, Third Edition, John Wiley & Sons, New York.

2. Fahimi, M. , Kasala, S., Fecso, R. (1990). "Hypothesis Testing for Multivariate Stratified Sampling", draft paper, National Agricultural Statistics Service, U.S. Department of Agriculture.

3. Kott, P. S. (1990). "On the Asymptotic Normality of Stratified Means and their t-Statistics", unpublished paper , National Agricultural Statistics Service, U.S. Department of Agriculture.

4. McClung, G., Tolomeo, V., Pafford, B. (1990). "The Measurement of Response Bias in March 1988 Quarterly On-Farm Grain Stocks Data", SRB Research Report Number SRB-90-09, National Agricultural Statistics Service, U.S. Department of Agriculture.

5. Pafford, B. (1989). "USE OF REINTERVIEW TECHNIQUES FOR QUALITY ASSURANCE: The Measurement of Response Bias in the Collection of December 1987 Quarterly Grain Stocks Data Using CATI," SRB Research Report Number SRB-89-08, National Agricultural Statistics Service, U.S. Department of Agriculture.

APPENDIX A: Stratified Multivariate and Univariate Tests

Stratified Multivariate Test: [2]

H_o: $\mu = \mu_o$ *vs* H_A: $\mu \neq \mu_o$

If $T^2 > \chi^2_p(\alpha)$ *then reject* H_o

$$T^2 = (\bar{\boldsymbol{x}} - \mu_o)\, \boldsymbol{S}^{-1} (\bar{\boldsymbol{x}} - \mu_o)'$$

where

$\bar{\boldsymbol{x}}$, μ are (1 x p) row vectors

$$\bar{\boldsymbol{x}} = \sum_{h=1}^{L} W_h\, \bar{\boldsymbol{x}}_h \qquad W_h = \frac{N_h}{N}$$

$$\bar{\boldsymbol{x}}_h = \frac{1}{n_h}\, \mathbf{1}'_{n_h}\, \boldsymbol{x}_h$$

$$\boldsymbol{S} = \sum_{h=1}^{L} W_h^2 (1-f_h)\, \frac{1}{n_h}\, \boldsymbol{S}_h \qquad f_h = \frac{n_h}{N_h}$$

$$\boldsymbol{S}_h = \frac{1}{n_h - 1} (\boldsymbol{x}_h - \mathbf{1}_{n_h} \bar{\boldsymbol{x}}_h)' (\boldsymbol{x}_h - \mathbf{1}_{n_h} \bar{\boldsymbol{x}}_h)$$

$\boldsymbol{x}_h$ = an (n_h x p) matrix of observations on p variables for stratum h

$\mathbf{1}_{n_h}$ = an (n_h x 1) vector of 1's

Stratified Univariate Test:

$H_o: \mu = \mu_o \qquad vs \qquad H_A: \mu \neq \mu_o$

if $z > Z_{\frac{\alpha}{2}}$ *then reject* H_o

$$z = \frac{\hat{x}_{st} - \mu_o}{\sqrt{Var(\hat{x}_{st})}}$$

$$\hat{x}_{st} = \sum_{h=1}^{L} N_h \bar{x}_h \qquad \bar{x}_h = \frac{1}{n} \sum_{i=1}^{n_h} x_{hi}$$

$$Var(\hat{x}_{st}) = \sum_{h=1}^{L} N_h (N_h - n_h) \frac{s_h^2}{n_h} \qquad s_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (x_{hi} - \bar{x}_h)^2$$

SECTION 5 - HOGS AND PIGS

1. On March 1, were any **HOGS** or **PIGS**, regardless of ownership, on the total acres operated? 006

☐ YES ☐ NO → 1a. Were any **HOGS** or **PIGS** on the total acres operated, at any time, during the period of December 1, 1987 through February 29, 1988?

☐ YES - *GO TO Section 6.* ☐ NO - *GO TO Section 6.*

2. Of the **HOGS and PIGS for BREEDING** on hand March 1, how many were:

a. **Sows, gilts and young gilts** bred and to be bred ..... 301 ★

b. How many were **boars and young males** for breeding ..... 302 ★

c. How many were **sows and boars** no longer used for breeding ..... 303 ★

3. Of the **HOGS and PIGS FOR MARKET and HOME USE**, how many were in each of the following four weight groups? *(Exclude breeding hogs reported in Item 2.)*

a. Under 60 lbs. *(Include pigs not yet weaned)* ..... 311 ★

b. 60 - 119 lbs. ..... 312 ★

c. 120 - 179 lbs. ..... 313 ★

d. 180 lbs. and over *(Exclude hogs no longer used for breeding)* ..... 314 ★

4. **TOTAL NUMBER OF HOGS and PIGS** on hand March 1. *(Add ★ Items 2a through 3d)* 300

Hogs: Incomp.	1 - Has 2 - Unk 3 - No	499
Comp.	4	

SECTION 6 - PARTNER'S NAMES

1. Did you check **partners** in Section 1, Item 4, on Face Page?

☐ NO - *GO TO Section 8.*

☐ YES - *Continue.*

2. Please identify the other person(s) in this partnership in boxes below, then go to Section 8.
(make necessary corrections if names have already been entered)

Name _____ (First) (Middle) (Last)
Address _____ (Rt. or St.)
City _____ State _____
Zip Code _____ Phone _____
Did this person operate land individually in this State on June 1, 1987? ☐ YES ☐ NO

Name _____ (First) (Middle) (Last)
Address _____ (Rt or St.)
City _____ State _____
Zip Code _____ Phone _____
Did this person operate land individually in this State on June 1, 1987? ☐ YES ☐ NO

Continue On Next Page

SECTION 5 - HOGS AND PIGS

1. On March 1, were any **HOGS** or **PIGS**, regardless of ownership, on the total acres operated? 006

☐ YES ☐ NO → 1a. Were any **HOGS** or **PIGS** on the total acres operated, at any time, during the period of December 1, 1987 through February 29, 1988?

☐ YES - *GO TO Section 6.* ☐ NO - *GO TO Section 6.*

2. Of the **HOGS and PIGS** for **BREEDING** on hand March 1, how many were:

a. **Sows, gilts and young gilts** bred and to be bred ..... 301 ★

b. How many were **boars and young males** for breeding ..... 302 ★

c. How many were **sows and boars** no longer used for breeding ..... 303 ★

3. How many **HOGS and PIGS FOR MARKET AND HOME USE** were on hand March 1? *(Exclude breeding hogs reported in Item 2.)* ..... 007 ★

4. **TOTAL NUMBER OF HOGS** and **PIGS** on hand March 1. *(Add ★ Items 2a through 3)* 300

Hogs: Incomp.	1 - Has 2 - Unk 3 - No	499
Comp.	4	

SECTION 6 - PARTNER'S NAMES

1. Did you check **partners** in Section 1, Item 4, on Face Page?

☐ **NO** - *GO TO Section 8.*

☐ **YES** - *Continue.*

2. Please identify the other person(s) in this partnership in boxes below, then go to Section 8.
(make necessary corrections if names have already been entered)

Name _____ *(First) (Middle) (Last)*

Address _____ *(Rt. or St.)*

City _____ State _____

Zip Code _____ Phone _____

Did this person operate land individually in this State on June 1, 1987? ☐ YES ☐ NO

Name _____ *(First) (Middle) (Last)*

Address _____ *(Rt. or St.)*

City _____ State _____

Zip Code _____ Phone _____

Did this person operate land individually in this State on June 1, 1987? ☐ YES ☐ NO

Continue On Next Page

Appendix C: Variance Computation of the Percent Bias

The variance of the percent bias was calculated from the formula for a combined ratio estimate as described in Cochran [1].

$$V(\hat{R}) = \frac{\sum_{h=1}^{L}\left(\frac{N_h(N_h-n_h)}{n_h(n_h-1)}\sum_{i=1}^{n_h}(u_{hi}-\bar{u}_h)^2\right)}{(\sum_{h=1}^{L}N_h\bar{x}_h)^2}$$

where

$\hat{R}$ = Proportion of total bias to total hogs (percent bias)

$$\hat{R}=\frac{\hat{Y}}{\hat{X}}$$

$f_h = n_h/N_h$

$$\hat{Y}=\sum_{h=1}^{L}N_h\bar{y}_h \qquad \bar{y}_h=\frac{\sum_{i=1}^{n_h}y_{hi}}{n_h}$$

$$\hat{X}=\sum_{h=1}^{L}N_h\bar{x}_h \qquad \bar{x}_h=\frac{\sum_{i=1}^{n_h}x_{hi}}{n_h}$$

$u_{hi} = y_{hi} - \hat{R}x_{hi}$

y_{hi} = bias for i^{th} observation in stratum h

x_{hi} = total hogs for i^{th} observation in stratum h

APPENDIX D: Computation of $G_{\bar{x}}$ and Contributions by Strata

The coefficient of skewness is computed as:

$$G_{\bar{x}} = \frac{\sum_{h=1}^{L} W_h^3 m_{3h}}{\left(\sum_{h=1}^{L} \frac{W_h^2 s_h^2}{n_h} \right)^{3/2}}$$

where

$$\bar{x} = \sum_{h=1}^{L} W_h \left(\sum_{i=1}^{n_h} \frac{x_{hi}}{n_h} \right) \qquad W_h = \frac{N_h}{N}$$

$$s_h^2 = \frac{\sum_{i=1}^{n_h} (x_{hi} - \bar{x}_h)^2}{n_h - 1}$$

$$m_{3h} = n_h \sum_{i=1}^{n_h} \frac{(x_{hi} - \bar{x}_h)^3}{(n_h - 1)(n_h - 2)} \qquad \text{when} \quad n_h > 2$$

$$m_{3h} = (2/3)\left(x_{h1}^3 + x_{h2}^3 - \frac{(x_{h1} + x_{h2})^3}{4} \right) \qquad \text{when} \quad n_h = 2$$

x_{hi} = bias for the i^{th} observation in stratum h

Table 1. Contributions to $G_{\bar{x}}$ by Strata for December 1987 Total Hogs.

state	strata	n_h	mean	variance	W_h	$G_{\bar{x}}$
IN	60	57	-0.070	2.10	0.26870	-0.00000
IN	62	18	-1.056	22.41	0.09574	-0.00000
IN	63	84	-0.286	4.93	0.22834	-0.00000
IN	64	33	2.061	182.95	0.13094	0.00003
IN	65	10	-2.800	78.40	0.03465	-0.00000
IN	67	40	-7.600	2678.25	0.09052	-0.00089
IN	69	8	-110.875	82611.84	0.04515	-0.20970
IN	70	9	9.222	1342.94	0.03989	0.00023
IN	71	6	-1.000	6.00	0.01214	-0.00000
IN	72	7	-157.286	84740.57	0.03760	-0.05977
IN	80	6	66.333	11613.87	0.01194	0.00024
IN	82	5	71.800	20841.20	0.00440	0.00005
MN	60	54	-1.185	50.00	0.23688	-0.00002
MN	61	39	0.000	0.00	0.16049	0.00000
MN	62	20	0.000	0.00	0.10300	0.00000
MN	63	67	-0.104	3.70	0.17066	-0.00000
MN	64	5	0.000	0.00	0.01356	0.00000
MN	65	39	-0.667	219.86	0.10793	0.00001
MN	66	7	-12.143	1040.14	0.03724	-0.00017
MN	67	13	0.000	0.00	0.03554	0.00000
MN	68	40	-0.050	6.25	0.06008	-0.00000
MN	69	14	-28.357	4669.94	0.02642	-0.00015
MN	70	12	-81.417	42371.36	0.02180	-0.00370
MN	71	12	-2.500	34.09	0.01651	-0.00000
MN	73	13	0.000	0.00	0.00593	0.00000
MN	75	3	-33.333	3333.33	0.00221	-0.00000
MN	76	3	13.333	533.33	0.00173	0.00000
OH	60	34	0.000	0.00	0.29034	0.00000
OH	62	18	0.000	0.00	0.13280	0.00000
OH	64	179	-1.838	494.27	0.34854	-0.00023
OH	65	3	-128.000	49152.00	0.03196	-0.15042
OH	66	6	0.000	0.00	0.05403	0.00000
OH	68	10	-6.000	360.00	0.08685	-0.00031
OH	69	3	0.000	0.00	0.00880	0.00000
OH	70	7	0.000	0.00	0.01942	0.00000
OH	71	3	-0.667	324.33	0.02157	-0.00000
OH	74	2	0.000	0.00	0.00127	0.00000
OH	80	2	-70.500	9940.50	0.00442	-0.00008

Table 2. Contributions to $G_{\bar{x}}$ by Strata for March 1988 Total Hogs.

state	strata	n_h	mean	variance	W_h	$G_{\bar{x}}$
IA	60	43	-0.651	14.57	0.22854	-0.00004
IA	62	38	0.000	0.00	0.13178	0.00000
IA	64	16	-2.688	68.63	0.09177	-0.00010
IA	65	48	-3.375	333.01	0.18013	-0.00170
IA	66	25	8.280	2297.38	0.10621	0.01094
IA	68	26	3.115	443.71	0.03641	0.00005
IA	69	54	-10.454	9598.91	0.10325	-0.04147
IA	71	51	-1.245	4550.74	0.12192	0.00814
NE	61	13	0.000	0.00	0.15638	0.00000
NE	63	14	2.214	68.64	0.07392	0.00008
NE	65	36	-0.167	4.14	0.15333	-0.00000
NE	66	63	-0.881	16.41	0.17432	-0.00001
NE	67	36	-13.889	13913.59	0.12691	-0.23658
NE	69	38	6.289	1477.56	0.07263	0.00161
NE	70	18	0.167	339.32	0.07098	-0.00012
NE	71	10	-26.300	5029.12	0.03408	-0.00402
NE	72	11	-11.636	1379.25	0.03775	-0.00063
NE	73	92	-12.587	17130.18	0.07995	-0.01770
NE	77	7	0.000	0.00	0.00335	0.00000
NE	80	13	-151.538	77462.94	0.01640	-0.01947
PA	60	7	0.000	0.00	0.17741	0.00000
PA	62	7	0.000	0.00	0.07701	0.00000
PA	65	5	0.000	0.00	0.06952	0.00000
PA	66	33	-0.030	0.03	0.33551	-0.00000
PA	67	4	0.000	0.00	0.03779	0.00000
PA	68	69	-0.246	2.19	0.07931	-0.00000
PA	70	98	1.061	91.19	0.15915	0.00005
PA	71	7	1.929	3754.20	0.01924	0.00016
PA	72	6	-14.333	1232.67	0.01099	-0.00007
PA	73	17	0.000	0.00	0.00570	0.00000
PA	75	6	0.000	0.00	0.01259	0.00000
PA	76	2	0.000	0.00	0.00440	0.00000
PA	77	3	0.000	0.00	0.00233	0.00000
PA	78	3	0.000	0.00	0.00108	0.00000
PA	79	4	0.000	0.00	0.00041	0.00000
PA	81	8	-0.063	1239.32	0.00634	0.00000
PA	82	4	-141.750	75932.25	0.00121	-0.00008

APPENDIX E: Frequency Histograms of Total Hog Bias

Figure 1. Frequency distribution of unexpanded total hog biases for December 1987 (excludes observations with zero bias).

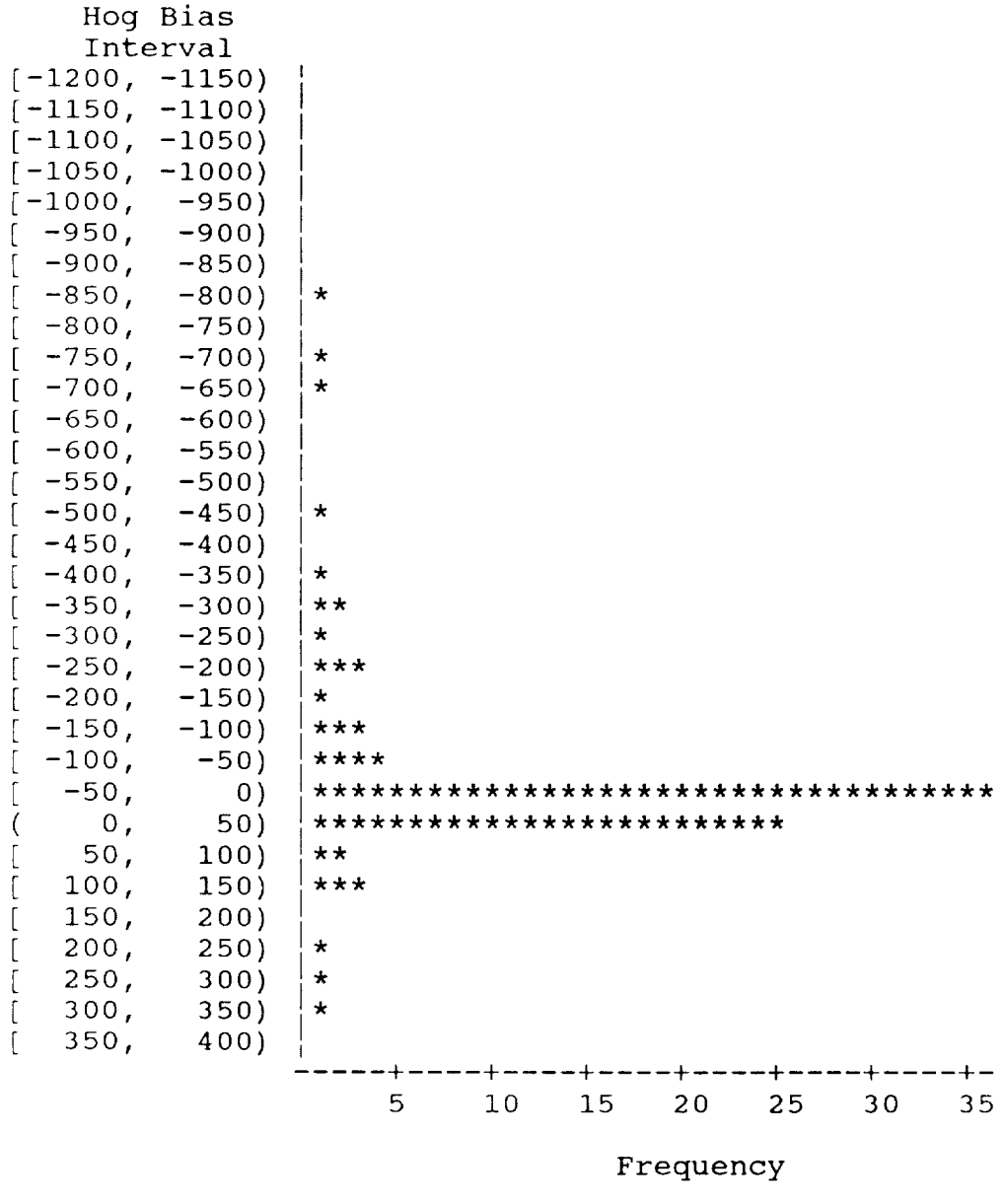

Figure 2. Frequency distribution of unexpanded total hog biases for March 1988 (excludes observations with zero bias).