TABLE OF CONTENTS

Summary .......................................................... iii
Introduction .......................................................... 1

Methods ............................................................ 2
   Current procedure ................................................. 2
   Estimation program .............................................. 2
   Estimators .......................................................... 2
   Outlier detection procedures ..................................... 2
New procedures ....................................................... 5
   Standardized Outlier Component Control Charts .............. 6
   Robust Estimator .................................................. 7
December 1989 Hog Board ........................................... 8
Applications since December 1989 ................................. 9

Results ............................................................. 10
   Results for December 1989 ....................................... 10
   Results of September 1989 Simulation .......................... 15

Conclusions ........................................................ 19

Recommendations ................................................... 19

References .......................................................... 21

Appendix .............................................................. 22

ACKNOWLEDGEMENTS

The authors appreciate the initiation of this project by Rich Allen and George Hanuschak. Throughout the project, they provided managerial support and encouragement. It was clear they both strongly support investigating and establishing formal robust estimation procedures for all multiple frame estimates.

The authors also extend a special thanks to Scot Rumburg for his help and investigation of specific outliers observations.
SUMMARY

The National Agricultural Statistics Service (NASS) conducts quarterly Multiple Frame (MF) hog inventory surveys. The survey direct expansion estimator for hogs is separated into outlier and non-outlier components during summary calculations. This report documents a procedure for identifying when the outlier component is unusually large or small. A simple robust estimator that reduces the impact of unusual outlier components is also presented. The Agricultural Statistics Board used information from these procedures when setting the December 1989 hog estimates and revising September 1989 estimates.

To identify unusual outlier components, standardized outlier components (SDOC) are calculated and charted for several quarters. The SDOC gives reviewers a way to judge if the current quarter's outlier component is influential when compared with previous outlier components. The robust estimator substitutes the current quarter's outlier component with an average outlier component over several quarters. This discounts the current outlier component and spreads its influence over time. However, the SDOC and the robust estimator can be heavily influenced by unusual historic outlier components if not enough historic data are used in the calculations.

Review of December 1989 robust indications and the Board estimates indicate that the robust indications influenced Board actions for States with unusual SDOC's. Simulation of the standardized outlier totals and robust indications for September 1989 indicate that this information would have been helpful to the Board. SDOC control charts clearly show the presence of unusual outlier components for seven States in September. The December Board revised the total hog estimates for five of these seven States in the direction of the robust indication. These revisions probably would have been reduced or eliminated if the SDOC's and robust indications had been available in September.

Clearly these procedures provided useful information to the December 1989 Hog Board. Currently, only the robust indication is supplied for the 10 State, 16 State and U.S. total hog estimates. No State level robust indications are used and no standardized outlier charts are used at any level.

NASS research and operational statistical methodology units should put high priority on jointly developing and implementing robust multiple frame procedures. Based on this review and evaluation of the procedures used for the December 1989 Hog Board, we recommend the following:
1. Examine methods that reduce the impact of unusual historic outlier components on the SDOC and robust estimator. This should include reviewing the amount of historic data used for calculating the SDOC and robust indications.

2. Review present outlier cutoff limits and examine other criteria for determining these cutoff limits. Current cutoff limits do not provide meaningful outlier components for many States.

3. Research staff should study the statistical characteristics of individual outliers and other components to aid in the development of effective outlier detection and robust estimation procedures.

4. Modify the operational summary systems with several new experimental outlier cutoff limits and store individual outlier observations for robust estimation analysis and review.

Once 1 through 4 have been resolved then:

5. The Agency should make it an important goal to implement outlier detection and robust estimation techniques for SSO and National Board reviews of most MF crop and livestock estimates.
A Review and Evaluation of Unusual Outlier Component Detection and Robust Estimation Procedures used by the December 1989 Hog Board

Gary Keough and Charles R. Perry, Jr.

INTRODUCTION

Outliers are common in agricultural commodity survey data because populations are often highly skewed with a large number of small values and a few very large values. However, limited attention is given to outliers until they dominate the estimator. Thomas, Perry, and Viroonsri examined Empirical Bayes and right censored estimators for highly skewed populations in order to dampen outlier effects. This report documents procedures to identify when the outlier component in repeated survey data are unusually larger or smaller than expected, and a simple robust estimator to be used when this occurs. Dr. Charles Perry developed these procedures for the National Agricultural Statistics Service (NASS) December 1989 Hog Board. The Chairperson of the Agricultural Statistics Board, Rich Allen, felt outliers overly influenced the September 1989 hog data and a procedure for dampening their effect was needed. Since outlier observations frequently occur in successive NASS Agricultural Surveys within a survey year (June-March) due to sample design, numerous outliers were expected in the December 1989 survey.

For this report, an outlier will be broadly defined as an expanded value considered unusual or influential by some defined procedure. The outlier component is the sum of these expanded values.
Methods

Current procedures

Estimation program

NASS has estimated quarterly hog numbers in the 10 major hog producing States since 1975. These States are Georgia, Illinois, Indiana, Iowa, Kansas, Minnesota, Missouri, Nebraska, North Carolina, and Ohio. In 1988 NASS started making quarterly estimates for the 16 major hog producing States by adding Kentucky, Michigan, Pennsylvania, South Dakota, Tennessee, and Wisconsin. These 16 States produce about 90 percent of the U.S. total hogs. Only annual estimates as of December 1 are made for the remaining States.

Estimators

NASS uses a multiple frame (MF) approach for estimating major agricultural commodities from the Agricultural Survey Program (ASP). This approach utilizes a list frame and an area frame for providing survey indications. The list frame consists of a list of farming operations and associated control data. Control data are used to stratify the list frame to improve sampling efficiency. A major disadvantage of almost any list frame is that it seldom contains the entire population of interest. The NASS area frame is complete and allows any area of land in the continuous 48 States to be selected with a known probability. Consequently, the area frame is used to account for the list frame's incompleteness in the MF approach.

Each reporting unit in the area frame sample is matched against the list frame to determine whether it is on the list or not. Reporting units not on the list frame represent the nonoverlap (NOL) population. The MF estimator is the sum of the list and NOL components. The NOL component is the main contributor to the variance of the estimator. Outliers typically come from small operation list strata and the NOL domain. These have small sampling fractions and therefore have large expansion factors. For a more detailed discussion of NASS's multiple frame approach see Nealon\(^2\) or Thomas et al.\(^4\)

A second MF indication adjusts the MF direct expansion (DE) for nonresponse based on information about presence or absence of hogs for the nonrespondents\(^1\). Any outliers that affect the original MF DE would also affect the adjusted indication.

Outlier detection procedures

NASS uses two different procedures in the current hog summary and analysis system for detecting outliers. Each procedure is described below.
The oldest procedure, Procedure 1, uses fixed State level cutoff values to identify if an observation's expanded value is an outlier. Table 1 gives the cutoffs by State in 1,000's. This procedure is used on both the LIST and NOL components of the multiple frame data. The history of Procedure 1 is obscure. No historical documentation describing its development is available. The procedure has been a part of NASS's Enumerative Summary System since its implementation in 1977.

TABLE 1 -- Procedure 1 cutoff values in 1,000's of hogs

<table>
<thead>
<tr>
<th>16 major hog States</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Georgia</td>
<td>25</td>
<td>Kansas</td>
<td>20</td>
<td>N. Carolina</td>
</tr>
<tr>
<td>Illinois</td>
<td>50</td>
<td>Minnesota</td>
<td>40</td>
<td>Ohio</td>
</tr>
<tr>
<td>Indiana</td>
<td>40</td>
<td>Missouri</td>
<td>40</td>
<td>Kentucky</td>
</tr>
<tr>
<td>Iowa</td>
<td>80</td>
<td>Nebraska</td>
<td>40</td>
<td>Michigan</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Remaining 32 States</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Alabama</td>
<td>10</td>
<td>Idaho</td>
<td>5</td>
<td>N. Hampshire</td>
</tr>
<tr>
<td>Arizona</td>
<td>5</td>
<td>Louisiana</td>
<td>5</td>
<td>N. Jersey</td>
</tr>
<tr>
<td>Arkansas</td>
<td>10</td>
<td>Maine</td>
<td>3</td>
<td>N. Mexico</td>
</tr>
<tr>
<td>California</td>
<td>5</td>
<td>Maryland</td>
<td>5</td>
<td>N. York</td>
</tr>
<tr>
<td>Colorado</td>
<td>10</td>
<td>Massachusetts</td>
<td>5</td>
<td>N. Dakota</td>
</tr>
<tr>
<td>Connecticut</td>
<td>5</td>
<td>Mississippi</td>
<td>10</td>
<td>Oklahoma</td>
</tr>
<tr>
<td>Delaware</td>
<td>5</td>
<td>Montana</td>
<td>5</td>
<td>Oregon</td>
</tr>
<tr>
<td>Florida</td>
<td>10</td>
<td>Nevada</td>
<td>4</td>
<td>Rhode Island</td>
</tr>
</tbody>
</table>

It appears the State cutoffs are set at approximately one percent of State's historical hog inventory. Except for the addition of some smaller States, cutoff limits have the not changed since 1977.

The average percent of the MF DE due to the outlier component was calculated using 1975 to 1989 quarterly information for the 10 major hog States. Average percentages for the remainder of States were calculated using 1988 and 1989 information. Figure 1 shows the outlier component is typically 5 to 10 percent of the MF DE for the 10 major hog States and about 6 percent at the 48 State level. The average percent varies more among the remaining States. Some States have had outlier components averaging over 45 percent of the MF DE for these 2 years. Nine states have averaged over 25 percent.

Cutoff limits should be reviewed to insure that meaningful outlier components are provided. Outlier components that are 40 percent or more of the estimate are hardly useful or meaningful. However, it may not be possible to provide consistently meaningful cutoff limits for the very small states with just a few positive reports since any of these reports can have a large influence on the total DE. Thomas et al.4 discuss cutoff limits in terms of upper quantiles.
Figure 1. Outlier Component as a Percent of the Direct Expansion

State/Region

10 Major States' percents calculated from 1975-1989 quarterly data, others calculated from 1988-1989 quarterly data
The historical NASS database contains state level outlier components and standard deviations from all observations classified as outliers by this procedure, but neither the individual outliers nor the number of outliers have been archived. Quarterly state level outlier components and standard deviations are available for the 10 major hog producing States starting with the December 1975 quarter. All other States start with the March 1988 quarter.

The newer procedure, Procedure 2, uses historical stratum variances to detect influential list observations at the stratum level. Procedure 2 was implemented in 1982. This procedure compares each observation's value to the stratum mean. List values larger than 10 standard deviations from the stratum mean are classified as outliers. This procedure's outlier component was about 3.5 percent of the National estimate in June 1989 and about 4 percent in September 1989.

Two censored estimates that adjust for outliers are computed from Procedure 2. The first excludes outlier observations and treats them as inaccessible responses. The summary system adjusts the sample count of usable observations used to calculate the stratum expansion factors. The second also excludes outlier observations but treats them as positive inaccessible responses in the nonresponse adjustment procedure.

The historical NASS database does not contain individual observations or the summary statistics from this procedure. However, the list frame individual outliers are available for all quarterly surveys starting with the December 1988 quarter in special databases managed by Statistical Methods Branch. The censored estimates are also available starting with the June 1989 quarter. Several additional quarters of censored estimates from microfilm copies of the hog analysis package output could be recovered with a sufficient clerical effort.

It should be noted that Procedure 2 involves only the list frame and therefore gives no indication of outliers in the NOL component of the multiple frame estimates. Because the NOL accounts for between 2/3 to 3/4 of the total variance in the hog series and because of limited data, no further analyses using Procedure 2 data was justified.

**New Procedures**

The techniques described below were produced using Procedure 1 historical data. They are intended to supplement current indications used by the Agricultural Statistics Board in setting National, Regional and State hog estimates.

The techniques can be divided into two logical groups for ease of discussion;
1. Standardized outlier component control charts for the detecting unusual outlier components, and

2. A simple robust procedure which will dampen the effect of unusual outlier components.

These procedures are based on the assumption that an average outlier component is expected each survey. It is recognized that a sample is not perfect and the outlier component from a sample may not truly represent the population. The new procedures are designed to identify when unusually large or small outlier component occur and dampen their affect on the survey expansion.

**Standardized Outlier Component Control Charts**

Standardized outlier component (SDOC) control charts identify when the current outlier component is unusual compared to previous outlier components. The SDOC for quarter t, SDOC\(_t\), is calculated as:

\[
SDOC_t = (O_t - \overline{O_{h1}})/S_o
\]

where

\[
O_t = \text{the outlier component for quarter } t;
\]

\[
\overline{O_{h1}} = \left[\frac{1}{n_1}\sum_{k=1}^{n_1} O_k\right]
\]

is the mean of the historic outlier components;

\[
S_o = \left\{\left[\frac{1}{n_1-1}\sum_{k=1}^{n_1} (O_k - \overline{O_{h1}})^2\right]\right\}^{\frac{1}{2}}
\]

is the standard deviation of the historic outlier components; and

\[
n_1 = \text{the number of quarters from } t=1 \text{ to } t-1.
\]

The quarters 1 through \(n_1\) make up a sliding window. Each quarter, the previous \(n_1\) quarter's outlier components make up the historic base used to calculate SDOC's.

For general guidelines to interpret SDOC control charts, consider the standardized normal distribution, a normal distribution with mean zero and variance of one. Typically, any observation from a standardized normal distribution with a value greater than two or less than minus two is considered unusual. The SDOC values can be interpreted similarly. This is a generalization and does not imply we can assign probabilities to the event a SDOC is larger than a
given value. However, as the SDOC gets larger in absolute value, the more unusual the outlier component is compared to historic outlier components. We can consider SDOC's within plus or minus two as common. However, when SDOC's are beyond plus or minus two the outlier component is influencing the MF DE and more consideration should be given to the robust estimator.

Note the outlier component for the current quarter is not used to calculate the historic outlier component mean and standard deviation. Consequently, the current quarter SDOC is an independent comparison of the current outlier component against its historic base. The SDOC is not an indication of the size of the outlier component. Also, SDOC's, as defined here, are to be recalculated each quarter because each current quarter has a different historic base.

Robust Estimator

When an unusually large or small outlier component occurs, the MF will provide an unreliable indication of the population total. Recall that an underlying assumption of this robust approach is that it is possible for an unusual outlier component to occur on any survey but its value may not be representative of the true outlier component. A better estimate is an average which spreads unusual outlier levels over several surveys. Thomas et al. examined an empirical Bayes method to stabilize the current survey estimate. A right-censored estimator for the NOL component of the MF estimate was also examined. The robust estimator presented here is similar to the right-censored estimator but is applied to the NOL and list components of the sample. The robust estimator introduced to the December 1989 Hog Board smooths out the current quarter's outlier component's influence by replacing it with the average outlier component from several quarters. The main reason for using the procedure is it's simplicity. The robust estimator for quarter $t$, $R_t$, is calculated as:

$$R_t = X_t - O_t + \bar{O}_{h_2}; \quad (1)$$

where

- $X_t$ = indication for quarter of interest,
- $O_t$ = the outlier component for quarter of interest,
- $\bar{O}_{h_2} = \left( \frac{1}{n_2} \right) \sum_{k=1}^{n_2} O_k$
- is the mean outlier component, and

$n_2 = n_1$ plus the current quarter.
Notice the robust estimator uses all outlier components in calculating the mean outlier component. This procedure spreads an unusual outlier component's influence out over several quarters so that it does not dominate the current quarter's indication.

The number of quarters, \( n_2 \), is presently a subjective choice. Guidelines for choosing the number of quarters should be examined. If \( n_2 \) is too small, influential outlier components are not spread over enough quarters, possibly causing the robust estimator to be biased. Similarly, if \( n_2 \) is too large, any trend in the size of the outlier component may also cause the robust estimator to be biased.

**December 1989 Hog Board**

For the December 1989 Hog Board, the robust estimator was applied to the MF DE and the adjusted MF DE indications. SDOC control charts and time series charts showing the relationship of the MF DE, adjusted MF DE, robust DE, adjusted robust DE, MF DE minus outlier component, adjusted MF DE minus outlier component, and outlier component were provided. Time series charts and additional computer listings displayed warning messages when the SDOC was 2 or larger. Charts and listings were generated for the individual 48 multiple frame States, the 16 State Region, 32 State Region, and 48 State Region.

SDOC's and robust indications were calculated using outlier components from the March 1988 through December 1989 quarters. This corresponds to the start of MF estimates for 48 continuous States. This was done for convenience since data prior to March 1988 is not available for all States. The standardized outlier component for December 1989, \( SDOC_{DE9} \), was calculated as:

\[
SDOC_{DE9} = (O_{DE9} - \bar{O}_{h1})/S_o
\]

where

- \( O_{DE9} \) = outlier component for December 1989,
- \( \bar{O}_{h1} \) = mean outlier component calculated using outlier totals from March 1988 through September 1989,
- \( S_o \) = standard deviation calculated using outlier totals from March 1988 through September 1989.

The robust indications were calculated as:

\[
R_{DE9,i} = X_{DE9,i} - O_{DE9} + \bar{O}_{h2},
\]

where
\[ X_{DE_{i}} = \text{the MF DE when } i=1, \]
\[ \tilde{\sigma}_{h_{2}} = \text{the adjusted MF DE when } i=2, \]
\[ \tilde{\sigma}_{h_{2}} = \text{mean outlier component calculated using outlier components from March 1988 through December 1989.} \]

Note that March 1988 through September 1989 robust indications were estimated using the \( \tilde{\sigma}_{h_{2}} \). Thus, robust indications for these quarters are calculated using one or more future outlier components. This was done because of the limited amount of data available. In an operational setting, only previous outlier components would be used.

**Applications since December 1989**

Currently, only U.S., 10 State, and 16 State Regional charts showing the robust indications and Board estimates are provided to the Board. These charts show March 1988 to present quarter indications. Robust indications are calculated using the March 1988 to present outlier components. No State level time series charts are used and no SDOC control charts are used at any level by the Board.
RESULTS

Results will be presented in two sections. The first section documents materials provided to the December 1989 Hog Board. The second section presents an analysis of the September 1989 survey with SDOC's and robust indications.

Results for December 1989

Charts A1 through A11 in the Appendix are copies of 48 State region material provided the December 1989 Hog Board. Charts A1 through A8 show different combinations of the Multiple frame indications, Multiple frame indications minus outlier components, robust indications, first Board estimates, first Board revisions, and the outlier component. Chart A9 shows only the indications from Chart A3 on a different scale. Chart A11 is a SDOC control chart. Figures 2-4 are reproductions of the SDOC control charts for the 48 State region, 10 State region, and Georgia.

Figures 2 and 3 show that the December SDOC is about -0.5. This implies the December outlier components are not unusual compared to the previous seven quarters. Figure 4 shows Georgia's SDOC for December was about 3.5. This indicates the Georgia's robust indications should be followed more closely than the MF DE and adjusted MF DE. The large December SDOC is due to one NOL operation that expanded to about 714,000 head in December, or about 60 percent of the State's total.

Figure 2. SDOC Control Chart
U.S. level, December 1989

![Graph showing SDOC control chart for December 1989, indicating a SDOC between -0.5 and 3.5, with March 1988 to December 1989 time period.](image-url)
Figure 3. SDOC Control Chart, 10 State level, December 1989

Figure 4. SDOC Control Chart, Georgia, December 1989
Table 2 shows the SDOC's for the 10 State, 16 State, 32 State, 48 State Region, and the 16 individual States from the December 1989 Survey. Besides the unusually large SDOC for Georgia, unusually small outlier components for Minnesota and the 32 State Region are present. Although Table 2 alone would identify potential outlier problems, review of the control charts is recommended to identify any long term problems or trends. For instance, Figure 4 shows that Georgia's outlier component had been fairly constant until March 1989 then it substantially increased in June, September, and December. This information should help with post-survey analysis and revisions. It is also noteworthy that the Regional SDOC's show no apparent outlier problems, but unusual SDOC's could exist for individual states. In the December 1989 survey, Georgia and Minnesota help offset each other so that the 10 State and 16 State regional SDOC's are small. It is important to review each published State's control chart.

Recognizing the unusual outlier levels for Georgia, Minnesota, and the 32 State Region, Figures 5, 6, 7, show how the Board used the simple robust indications in these situations. For Georgia, the December MF DE and adjusted MF DE are biased upward due to outliers so the Board more closely followed the robust indications. For Minnesota and the 32 State Region the MF DE and adjusted MF DE are biased downward due to small outlier levels, so the Board more closely followed the higher levels indicated by the robust indications. In all three cases, apparently the robust indications were influential on the Board estimate. Table A1 (in the Appendix) shows the indications, first Board and revised estimates, and differences for all 16 States and the regions.

Additional review of Figure 5 illustrates the problem of not spreading influential outlier components over enough quarters. Note that the March 1988 through March 1989 robust DE and adjusted robust DE are substantially higher than the MF DE, adjusted MF DE, and Board values. This is because the large September and December outlier components are included in the average outlier component. Recall equation 1,

\[ R_t = X_t - O_t + \bar{O}_{h2}. \]

If the average outlier component (\( \bar{O}_{h2} \)) is substantially larger than the current outlier component (\( O_t \)), the robust indication will be substantially larger than the DE. The intent of the estimator is that this average will be a "better" indication of the true outlier component level. However, the average used in Figure 5 includes the large September and December outlier component values. Consequently, averages and robust estimates are probably biased upward for all quarters. This is most noticeable in March 1988 through March 1989 where the original outlier component was relatively small. This impact could possibly be reduced by increasing the number of quarters used to calculate \( \bar{O}_{h2} \). Although this is an artificial setting since future data are used to
calculate the average outlier component, it still shows large outlier components in previous quarters have a major impact on the average outlier component if it is calculated with too few quarters.

Figure 5. Total Hogs--Georgia
December 1989
Figure 6. Total Hogs—Minnesota
December 1989

Figure 7. Total Hogs—32 State
December 1989
Table 2 -- SDOC's for December 1989 Multiple Frame Survey by State

<table>
<thead>
<tr>
<th>State</th>
<th>Standardized Outlier Totals</th>
<th>Standardized Outlier Totals</th>
<th>Standardized Outlier Totals</th>
<th>Standardized Outlier Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>GA.</td>
<td>3.5</td>
<td>KANSAS -0.4</td>
<td>N.CAR. -1.2</td>
<td>PA. 0.3</td>
</tr>
<tr>
<td>ILL.</td>
<td>-1.1</td>
<td>MINN. -2.0</td>
<td>OHIO -1.4</td>
<td>S.DAK. 1.0</td>
</tr>
<tr>
<td>IND.</td>
<td>0.3</td>
<td>MO. -1.0</td>
<td>KY. 0.4</td>
<td>TENN. -1.6</td>
</tr>
<tr>
<td>IOWA.</td>
<td>-1.5</td>
<td>NEB. 1.2</td>
<td>MICH. 0.5</td>
<td>WISC. 1.6</td>
</tr>
<tr>
<td>10 HOG</td>
<td>-0.4</td>
<td>16 HOG -0.2</td>
<td>32 HOG -2.0</td>
<td>48 HOG -0.6</td>
</tr>
</tbody>
</table>

Results of September 1989 Simulation

During the September 1989 survey, it was recognized that outliers were greatly affecting the survey indications. Analysis was conducted to examine the possible impact of the outlier detection and robust estimation techniques if they had been available. To do this the SDOC's and robust indications were recalculated as if in an operational setting. That is, each charted robust indication was calculated using only previous survey data. The SDOC values for the original 10 major hog States were calculated using the previous seven quarters' (December 1987-June 1989) outlier components, while the previous six quarters (March 1988-June 1989) outlier components were used for the remaining six States. This simulation is only an example of many possible approaches. This approach was chosen to mirror the December 1989 procedures as much as possible.

Figure 8 is an example of a simulated September 1989 time series chart using Georgia hogs indications. Figure 9 is the SDOC control chart. These figures are similar to Charts A5 and A6, respectively.

Figure 8 shows the robust indications support the Board estimates up through June 1989. The large September outlier component causes the MF survey indications to increase over 200,000 hogs from the June indications while the robust indications decline slightly. Figure 9 shows Georgia's September SDOC is about 8, extremely unusual. Figures 8 and 9 clearly show the September 1989 outlier component in Georgia was unusually large and affected the multiple frame indications.
Figure 8. Total Hogs--Georgia
September 1989

Figure 9. SDOC Control Chart,
Georgia, September 1989
Comparing Figure 9 (Georgia's September SDOC control chart) with Figure 4 (Georgia's December SDOC control chart) illustrates the impact that an unusually large outlier component can have on SDOC values. Georgia's September outlier component was 598,574. The September chart (Figure 9) shows this value had a SDOC of about eight while the December chart (Figure 4) shows September with a SDOC of about two. This is because a different historic base is used each month to calculate the historic outlier component average ($\bar{O}_h$) and standard deviation ($S_o$). September's large outlier component is not used to calculate Figure 9's SDOC's while it is used to calculate Figure 4's SDOC's. This also shows that one previous quarter's large outlier component can have a major impact on the current quarter's SDOC if it is calculated with too few quarters of data. Table 3 shows the historic mean and standard deviations are very different for each chart. This demonstrates that the current quarter SDOC is simply a comparison of the current quarter outlier component against its historic base. It does not indicate the actual size of the outlier component.

### Table 3 -- Comparison of Georgia's SDOC values

<table>
<thead>
<tr>
<th>SDOC Chart</th>
<th>Historic Mean</th>
<th>Historic Standard Deviation</th>
<th>September Outlier Component</th>
<th>September SDOC Value 1/</th>
</tr>
</thead>
<tbody>
<tr>
<td>September</td>
<td>127,398</td>
<td>585,810</td>
<td>598,574</td>
<td>8</td>
</tr>
<tr>
<td>December</td>
<td>192,323</td>
<td>188,331</td>
<td>598,574</td>
<td>2</td>
</tr>
</tbody>
</table>

1/ SDOC = September Outlier Component - Historic Mean

Historic Standard Deviation

Table 4 shows six States, Georgia, Indiana, Minnesota, Nebraska, Ohio, and Wisconsin would have had SDOC's greater than two. The SDOC for N. Carolina would have been -2.7. Also the, 10 State Region, 16 State Region, and National levels would have had similarly high SDOC's. It is obvious there were several unusual outlier components in the September data.

### Table 4 -- Simulated standardized outlier components for September 1989

<table>
<thead>
<tr>
<th>State</th>
<th>SDOC</th>
<th>State</th>
<th>SDOC</th>
<th>State</th>
<th>SDOC</th>
<th>State</th>
<th>SDOC</th>
<th>State</th>
<th>SDOC</th>
</tr>
</thead>
<tbody>
<tr>
<td>GA.</td>
<td>8.0</td>
<td>KANSAS</td>
<td>-1.9</td>
<td>N.CAR.</td>
<td>-2.7</td>
<td>PA.</td>
<td>1.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ILL.</td>
<td>0.6</td>
<td>MINN.</td>
<td>3.5</td>
<td>OHIO</td>
<td>3.4</td>
<td>S.DAK.</td>
<td>-0.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IND.</td>
<td>3.0</td>
<td>MO.</td>
<td>-1.2</td>
<td>KY.</td>
<td>0.5</td>
<td>TENN.</td>
<td>-1.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IOWA.</td>
<td>-0.6</td>
<td>NEB.</td>
<td>3.4</td>
<td>MICH.</td>
<td>-0.3</td>
<td>WISC.</td>
<td>3.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 HOG</td>
<td>3.4</td>
<td>16 HOG</td>
<td>4.4</td>
<td>32 HOG</td>
<td>1.3</td>
<td>48 HOG</td>
<td>3.3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Recognizing the impact that the SDOC charts and robust indications had on the December Board estimates, it is probable that the September estimates would have been similarly affected. Estimates for States with large SDOC's (larger than two) would have followed the lower robust indications more closely than the operational MF indication. Estimates for States with small SDOC's (smaller than minus two) would have followed the larger robust indication more closely. The initial Board estimates for these targeted States and Regions were subsequently revised as indicated in Table 5. In general, these revisions were in the direction anticipated by the SDOC value and robust indication. Only North Carolina and Ohio were not revised. Quite likely, revisions would not have been necessary or at least reduced if these procedures had been available in September.

Table 5 -- September 1989 Hog Indications and Board Estimate

<table>
<thead>
<tr>
<th>State</th>
<th>Simulated SDOC</th>
<th>MF DE</th>
<th>Robust DE</th>
<th>First Board</th>
<th>Revised Board</th>
</tr>
</thead>
<tbody>
<tr>
<td>GA.</td>
<td>8.0</td>
<td>1482</td>
<td>1070</td>
<td>1300</td>
<td>1250</td>
</tr>
<tr>
<td>IND.</td>
<td>3.0</td>
<td>5158</td>
<td>4440</td>
<td>4650</td>
<td>4550</td>
</tr>
<tr>
<td>MINN.</td>
<td>3.5</td>
<td>5036</td>
<td>4802</td>
<td>5050</td>
<td>4950</td>
</tr>
<tr>
<td>NEB.</td>
<td>3.4</td>
<td>4424</td>
<td>4132</td>
<td>4450</td>
<td>4350</td>
</tr>
<tr>
<td>N.CAR.</td>
<td>-2.7</td>
<td>2627</td>
<td>2780</td>
<td>2700</td>
<td>2700</td>
</tr>
<tr>
<td>OHIO</td>
<td>3.4</td>
<td>2277</td>
<td>2125</td>
<td>2300</td>
<td>2300</td>
</tr>
<tr>
<td>WISC.</td>
<td>3.2</td>
<td>1407</td>
<td>1239</td>
<td>1350</td>
<td>1300</td>
</tr>
<tr>
<td>10 HOG</td>
<td>3.4</td>
<td>45927</td>
<td>44511</td>
<td>45800</td>
<td>45200</td>
</tr>
<tr>
<td>16 HOG</td>
<td>4.4</td>
<td>53199</td>
<td>51526</td>
<td>53045</td>
<td>52395</td>
</tr>
</tbody>
</table>

Table A2 shows the indications, first Board and revised estimates, and differences for all States and Regions.
CONCLUSIONS

Occasionally and unpredictably, outliers have a large impact on State, Regional, and National survey indications. Techniques are needed to identify when unusual outliers occur and to provide robust indications that dampen the effect of the outliers. After a thorough review of outlier detection and robust techniques, simple procedures were developed and implemented for the December 1989 Hog Board. SDOC control charts clearly identified that unusually large and small outlier components were present for some States and Regions in the December 1989 data. Review of the Official estimates indicate that the robust indications did influence the Board action for these States and Regions. Analysis also indicates that SDOC control charts would have identified unusual outlier components in the September 1989 data and that using the robust indications could have reduced or eliminated the revisions that were later necessary. However, unusual outlier components will have a large impact on the SDOC and robust estimator if not enough historic data are used. Also, current outlier cutoff limits are not meaningful for many States.

RECOMMENDATIONS

NASS research and operational statistical methodology units should put high priority on jointly developing and implementing more robust multiple frame procedures. After reviewing how the SDOC control charts and the robust estimator performed for the December 1989 Hog Board and in the September 1989 simulation, we recommend the following:

1. Examine methods that reduce the impact of unusual historic outlier components on the SDOC and robust estimator. This should include reviewing the amount of historic data used for calculating the SDOC and robust indications.

2. Review present outlier cutoff limits and examine other criteria for determining these cutoff limits. Current cutoff limits do not provide meaningful outlier components for many States.

3. Research staff should study the statistical characteristics of individual outliers and other components to aid in the development of effective outlier detection and robust estimation procedures.

4. Modify the operational summary systems with several new experimental outlier cutoff limits and store individual outlier observations for robust estimation analysis and review.

Once 1 through 4 have been resolved then:
5. The Agency should make it an important goal to implement outlier detection and robust estimation techniques for SSO and National Board reviews of most MF crop and livestock estimates.
REFERENCES


**APPENDIX**

Table A1 -- Multiple frame and robust indications, first and revised Board estimates, 16 major hog States, December 1989

<table>
<thead>
<tr>
<th>State</th>
<th>Multiple Frame Ind.</th>
<th>Robust Ind.</th>
<th>Board</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Adj. DE *diff</td>
<td>Adj. DE *diff</td>
<td>First *diff</td>
</tr>
<tr>
<td>GA.</td>
<td>1670 -470 1687 -487</td>
<td>1086 114</td>
<td>1103 97</td>
</tr>
<tr>
<td>ILL.</td>
<td>5307 393 5696 4</td>
<td>5460 240</td>
<td>5849 -149</td>
</tr>
<tr>
<td>IND.</td>
<td>4307 43 4489 -139</td>
<td>4196 154</td>
<td>4379 -29</td>
</tr>
<tr>
<td>IOWA.</td>
<td>13022 478 13481 19</td>
<td>13500 0</td>
<td>13959 -459</td>
</tr>
<tr>
<td>KANSAS</td>
<td>1367 83 1461 -11</td>
<td>1378 72</td>
<td>1472 -22</td>
</tr>
<tr>
<td>MINN.</td>
<td>4062 388 4369 81</td>
<td>4262 188</td>
<td>4569 -119</td>
</tr>
<tr>
<td>MO.</td>
<td>2539 161 2612 88</td>
<td>2598 102</td>
<td>2672 28</td>
</tr>
<tr>
<td>NEB.</td>
<td>3793 407 4131 69</td>
<td>3628 572</td>
<td>3966 234</td>
</tr>
<tr>
<td>N.CAR.</td>
<td>2564 6 2651 -81</td>
<td>2664 -94</td>
<td>2751 -181</td>
</tr>
<tr>
<td>OHIO</td>
<td>1807 273 1916 164</td>
<td>1906 174</td>
<td>2015 65</td>
</tr>
<tr>
<td>10 HOG</td>
<td>40437 1763 42492 -292</td>
<td>40680 1520</td>
<td>42735 -535</td>
</tr>
<tr>
<td>KY.</td>
<td>961 14 990 -15</td>
<td>940 35</td>
<td>968 7</td>
</tr>
<tr>
<td>MICH.</td>
<td>1247 13 1282 -22</td>
<td>1224 36</td>
<td>1258 2</td>
</tr>
<tr>
<td>PA.</td>
<td>940 35 955 20</td>
<td>926 49</td>
<td>941 34</td>
</tr>
<tr>
<td>S.DAK.</td>
<td>1754 -34 1862 -142</td>
<td>1709 11</td>
<td>1817 -97</td>
</tr>
<tr>
<td>TENN.</td>
<td>653 47 671 29</td>
<td>795 -95</td>
<td>813 -113</td>
</tr>
<tr>
<td>WISC.</td>
<td>1298 -48 1329 -79</td>
<td>1166 84</td>
<td>1197 53</td>
</tr>
<tr>
<td>16 HOG</td>
<td>47289 1791 49579 -499</td>
<td>47438 1642</td>
<td>49728 -648</td>
</tr>
<tr>
<td>32 HOG</td>
<td>4526 246 4573 199</td>
<td>4840 -68</td>
<td>4886 -114</td>
</tr>
<tr>
<td>US HOG</td>
<td>51816 2036 54152 -300</td>
<td>52278 1574</td>
<td>54615 -763</td>
</tr>
</tbody>
</table>

*diff is the Revised Board Estimate minus the previous indication.
Table A2 -- Multiple frame and simulated robust indications, first and revised Board estimates, 16 major hog States, September 1989

<table>
<thead>
<tr>
<th>State</th>
<th>Multiple Frame Ind.</th>
<th>Simulated Robust Ind.</th>
<th>Board</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Adj. *diff</td>
<td>Adj. *diff</td>
<td>Adj. *diff</td>
</tr>
<tr>
<td>DE</td>
<td>DE</td>
<td>First</td>
<td>Revised</td>
</tr>
<tr>
<td>GA.</td>
<td>1482 -232</td>
<td>1070 180</td>
<td>1300 -50</td>
</tr>
<tr>
<td>ILL.</td>
<td>5951 149</td>
<td>5879 221</td>
<td>6100 0</td>
</tr>
<tr>
<td>IND.</td>
<td>5158 -608</td>
<td>4440 110</td>
<td>4650 -100</td>
</tr>
<tr>
<td>IOWA.</td>
<td>14635 -35</td>
<td>14816 -216</td>
<td>14800 -200</td>
</tr>
<tr>
<td>KANSAS</td>
<td>1536 14</td>
<td>1579 -29</td>
<td>1600 -50</td>
</tr>
<tr>
<td>MINN.</td>
<td>5036 -86</td>
<td>4802 148</td>
<td>5050 -100</td>
</tr>
<tr>
<td>MO.</td>
<td>2801 49</td>
<td>2889 -39</td>
<td>2850 0</td>
</tr>
<tr>
<td>NEB.</td>
<td>4424 -74</td>
<td>4132 218</td>
<td>4450 -100</td>
</tr>
<tr>
<td>N.CAR.</td>
<td>2627 73</td>
<td>2780 -80</td>
<td>2700 0</td>
</tr>
<tr>
<td>OHIO</td>
<td>2277 23</td>
<td>2125 175</td>
<td>2300 0</td>
</tr>
<tr>
<td>10 HOG</td>
<td>45927 -727</td>
<td>44511 689</td>
<td>45800 -600</td>
</tr>
<tr>
<td>KY.</td>
<td>1028 -3</td>
<td>1017 8</td>
<td>1025 0</td>
</tr>
<tr>
<td>MICH.</td>
<td>1262 38</td>
<td>1277 23</td>
<td>1300 0</td>
</tr>
<tr>
<td>PA.</td>
<td>1038 -18</td>
<td>982 38</td>
<td>1020 0</td>
</tr>
<tr>
<td>S.DAK.</td>
<td>1756 -6</td>
<td>1767 -17</td>
<td>1750 0</td>
</tr>
<tr>
<td>TENN.</td>
<td>781 19</td>
<td>874 -74</td>
<td>800 0</td>
</tr>
<tr>
<td>WISC.</td>
<td>1407 -107</td>
<td>1239 61</td>
<td>1350 -50</td>
</tr>
<tr>
<td>16 HOG</td>
<td>53199 -804</td>
<td>51526 869</td>
<td>53045 -650</td>
</tr>
<tr>
<td>32 HOG</td>
<td>5280 -80</td>
<td>5400 -200</td>
<td>5078 122</td>
</tr>
<tr>
<td>US HOG</td>
<td>58479 -884</td>
<td>58445 -850</td>
<td>56687 908</td>
</tr>
</tbody>
</table>

*diff is the Revised Board Estimate minus the previous indication.
TOTAL HOGS -- U.S. LEVEL

LEVEL=1 GROUP=No apparent outliers SIGN=2 SIZE=0.6 FIPS=99 ST_REG=48 HOG

DATE

Dec87 Mar88 Jun88 Sep88 Dec88 Mar89 Jun89 Sep89 Dec89 Mar90

D: Direct Expansion.
A: Adjusted.
d: Direct Expansion Less Current Outliers.
a: Adjusted Less Current Outliers.
: First Board Estimate.
\: Sum of all Outliers.
TOTAL HOGS -- U.S. LEVEL

LEVEL=1 GROUP=No apparent outliers SIGN=2 SIZE=0.6 FRPS=99 ST_REC=48 HOG

D: Direct Expansion
A: Adjusted
a: Direct Expansion Less Current Outliers
A: Adjusted Less Current Outliers
- First Board Revision
* Sum of all Outliers
TOTAL HOGS -- U.S. LEVEL

LEVEL=1 GROUP=No apparent outliers SIGN=2 SIZE=0 6 FIPS=99 ST_REG=48 HOG

- D: Direct Expansion
- A: Adjusted
- a: Direct Expansion Less Current Outliers
  Plus Mean of historical Outliers
- o: Adjusted Less Current Outliers
  Plus Mean of historical Outliers
- *: First Board Estimate
- 0: Sum of all Outliers
TOTAL HOGS -- U.S. LEVEL

LEVEL=1 GROUP=No apparent outliers SIGN=2 SIZE=3 6 FIPS=99 ST_REG=48 HOG

DATE

0 10000000 20000000 30000000 40000000 50000000 60000000 70000000
Dec87 Mar88 Jun88 Sep88 Dec88 Mar89 Jun89 Sep89 Dec89 Mar90

D: Direct Expansion.
A: Adjusted.
d: Direct Expansion Less Current Outliers
   Plus Mean of historical Outliers.
O: Adjusted Less Current Outliers
   Plus Mean of historical Outliers
- First Board Revision
s: Sum of all Outliers
TOTAL HOGS -- U.S. LEVEL

DATE

- Direct Expansion.
A. Adjusted.
* First Board Estimate.
o. Sum of all Outliers.
Chart A6

TOTAL HOGS -- U.S. LEVEL

LEVEL=1 GROUP=No apparent outliers SIGN=2 SIZE=0 6 FIPS=99 ST_PC=48 HOG

D: Direct Expansion
A: Adjusted
= First Board Revision
> Sum of all Outliers
TOTAL HOGS — U.S. LEVEL

LEVEL=1 GROUP=No apparent outliers SIGN=2 SIZE=0.6 FIPS=99 ST_REG=48 HOG

---

d: Direct Expansion Less Current Outliers
Plus Mean of historical Outliers.
a: Adjusted Less Current Outliers
Plus Mean of historical Outliers.
-: First Board Estimate.
o: Sum of all Outliers.
TOTAL HOGS -- U.S. LEVEL

LEVEL = 1  GROUP = No apparent outliers  SIGN = 2  SIZE = 0 6  FIPS = 99  ST_REG = 48 HOG

Chart A8

DATE

Dec87  Mar88  Jun88  Sep88  Dec88  Mar89  Jun89  Sep89  Dec89  Mar90

0  10000000  20000000  30000000  40000000  50000000  60000000

Direct Expansion Less Current Outliers
Plus Mean of historical Outliers
Adjusted Less Current Outliers
Plus Mean of historical Outliers
First Board Revision
Sum of all Outliers
TOTAL HOGS — U.S. LEVEL

LEVEL = 1
GROUP = No apparent outliers
SIGN = 2
SIZE = 0.6
FIPS = 99
ST_REG = 48 HOG

Chart A9

D: Direct Expansion.
A: Adjusted.
a: Direct Expansion Less Current Outliers
   Plus Mean of historical Outliers.
a: Adjusted Less Current Outliers
   Plus Mean of historical Outliers.
- First Board Estimate.
TOTAL HOGS -- U.S. LEVEL

LEVEL = 1  GROUP = No apparent outliers  SIGN = 2  SIZE = 0 6  FIPS = 09  ST, REG = 48 HCG

Chart A10

Dec87  Mar88  Jun88  Sep88  Dec88  Mar89  Jun89  Sep89  Dec89  Mar90

50000000  51000000  52000000  53000000  54000000  55000000  56000000  57000000  58000000  59000000  60000000  61000000

D: Direct Expansion
A: Adjusted
D': Direct Expansion Less Current Outliers
Plus Mean of historical Outliers.
A': Adjusted Less Current Outliers
Plus Mean of historical Outliers.
- First Board Revision.
TOTAL HOGS -- U.S. LEVEL

LEVEL=1 GROUP=No apparent outliers SIGN=2 SIZE=0.6 FIPS=99 ST_REG=48 HOG

current outliers = historical outliers mean
historical standard error of outliers