United States
Department of
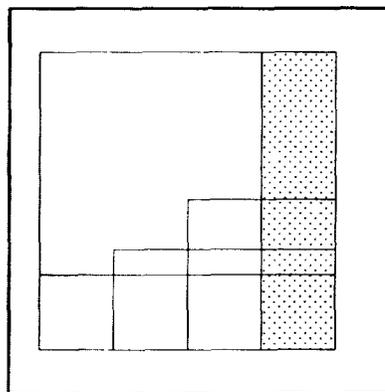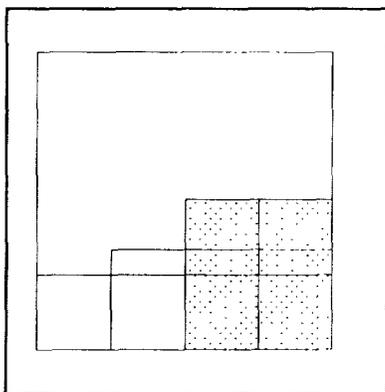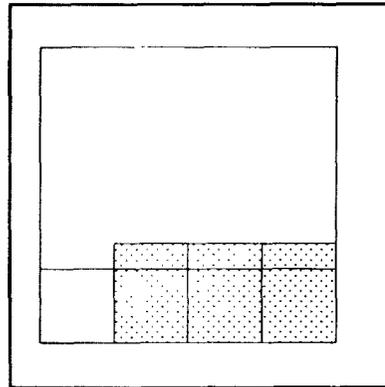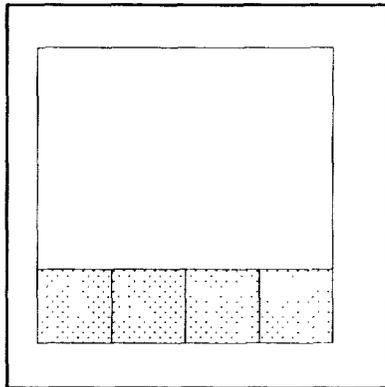Agriculture

National
Agricultural
Statistics
Service

Research Division

SRB Research Report
Number SRB-92-06

October 1992

ESTIMATION OF QAS LIST FRAME HOG TOTALS USING A REVISED ESTIMATOR -- AN EMPIRICAL INVESTIGATION

M. J. Fetter

ABSTRACT

The National Agricultural Statistics Service (NASS) implements the
Quarterly Agricultural Survey (QAS) to produce estimates for crops,
grain stocks, and livestock. It is a multiple frame survey which
uses an area frame and a list frame. This report is concerned with
QAS hog total estimates derived from the list frame. The
Reweighted and Adjusted Estimators are the two list frame hog total
estimators currently in use. The Reweighted Estimator uses none of
the partial information that is available on nonrespondents and
assumes that nonrespondents are a random sample of the population.
The Adjusted Estimator was developed by Crank (1979) to make the
best use of partial information that was available at that time.
Additional information is now available, making possible more
reasonable assumptions about nonrespondents. A Revised Estimator
is proposed in this report that makes use of this additional
information and also uses more effective weighting classes which
are based on hog control data.

List frame hog total estimates produced by the Revised Estimator
are somewhat higher than the estimates produced by the Adjusted
Estimator. Over fifteen consecutive QAS surveys (June 1988 -
December 1991), the average percentage increase relative to the
Adjusted Estimator's estimate ranged from a low of 0.64 percent in
Iowa to a high of 2.96 percent in Georgia.

It is recommended that the Revised Estimator be implemented for QAS
list hog total estimation as it is a statistically sound,
mathematically simple estimator and is based on reasonable
assumptions about the nonresponse mechanism.

TABLE OF CONTENTS

SUMMARY

The Quarterly Agriculture Survey (QAS) is an integrated multiple frame survey and is the instrument used to gather data for the estimation of the production of various crops and livestock items at state and national levels. The QAS list frame hog estimate is the focus of this report.

Cox (1992) outlines possible inadequacies of current list frame QAS hog estimators and outlines the development of an alternative estimator. This estimator differs from the NASS Adjusted Estimator in two ways. First, the alternative estimator (referred to as the Revised Estimator in this report) uses all partial data that are available for nonrespondents whereas the Adjusted Estimator uses only some of this information. Essentially, this results in the estimators being based on slightly different assumptions concerning the characteristics of the nonrespondents for which nothing is known about the presence or absence of hogs on the operation.

Second, the Revised Estimator uses poststratification to form weighting classes that are based strictly on hog control data whereas the Adjusted Estimator relies on the use of design strata as weighting classes. Because of its integrated nature, QAS design strata represent a compromise of optimality for each commodity. Some of these strata may be based on cropland, grain storage capacity, or other items not directly related to hogs. What the Revised Estimator attempts to do is form weighting classes that would be similar to the design strata that would be used for a hog-specific survey. This serves to group records together that have similar hog characteristics.

The Adjusted Estimator uses the design strata as weighting classes. Due to the large number of strata, some of which contain few hog operations, stratum hog total estimates are sometimes based on a very small number of complete respondents that have hogs (sometimes there are no complete respondents that are hog operators in a stratum). This can lead to excessive variance in the estimate. The Revised Estimator uses weighting classes that generally have at least 20 respondents. Weighting classes that fail to achieve a minimum number of respondents are collapsed with another similar weighting class to assure a statistically reasonable number of respondents.

List frame hog total estimates for 15 consecutive quarters were computed using QAS data for Georgia, Illinois, Indiana, Iowa, and North Carolina using both the Adjusted Estimator and the Revised Estimator. On the average, the Revised Estimator produced higher estimates than the Adjusted Estimator. The average relative increase ranged from a low of 0.64 percent in Iowa to a high of 2.96 percent in Georgia. The absolute difference between estimates produced by these two estimators was generally small in Iowa and

Indiana and comparatively large in Georgia and Illinois, with more unpredictable differences showing up in North Carolina. The percentage change varied from survey to survey and from state to state dependent on nonresponse counts and the effect of poststratification. Approximated variances produced CVs that closely matched current CVs of the Adjusted Estimator.

The Revised Estimator is derived from a particular set of assumptions made about nonrespondents, just as the Reweighted and Adjusted Estimators are derived from different sets of assumptions. The estimator that is based on the most reasonable set of assumptions is the estimator that should be employed. This paper argues that in light of the information presently being captured by the QAS, the most reasonable set of assumptions regarding nonrespondents is the set from which the Revised Estimator is derived. This paper recommends that the Revised Estimator be implemented for QAS list hog total estimation as it is a statistically sound, mathematically simple estimator and is based on reasonable assumptions.

ESTIMATION OF QAS LIST FRAME HOG TOTALS USING A REVISED ESTIMATOR-- AN EMPIRICAL INVESTIGATION

M.J. Fetter

INTRODUCTION

The National Agriculture Statistics Service (NASS) uses a multiple frame procedure for estimating hog totals in its Quarterly Agricultural Survey (QAS) Program. This procedure results in what is called a multiple frame estimate. The multiple frame estimate is composed of two independent components, the list frame expansion and the nonoverlap (NOL) expansion. The NOL expansion is derived from individuals sampled from the area frame that are not represented on the list frame. The list frame expansion and the NOL expansion are summed to produce the multiple frame estimate. All estimators discussed in this report are used only to produce the list frame estimate.

The need to account for survey nonresponse to reduce nonresponse bias is well known. The nonresponse rate for the hog section of the QAS typically ranges from 10 to 20 percent (Fuchs and Bass 1990, pp. 12-22) and the methods used to account for this nonresponse can have a significant effect on the overall estimate. Nonresponse occurs when a sampled unit from the population fails to provide all of the information that the survey was intended to capture (i.e., hog totals for each unit). It would be simple to deal with nonresponse if the set of survey respondents represented a random sample of the original sample. This would imply that nonrespondents would also represent a random sample of the original sample. However, in many survey situations, nonresponse is not random but is related to characteristics of the nonresponding units. Thus, the respondents cannot generally be considered as a random sample of the original sample.

The most widely used methods of accounting for nonresponse are imputation and weight adjustment. When estimating QAS list frame hog totals, NASS uses manual imputation for nonresponding operations from certain "prob-1" (self representing) strata and weight adjustment procedures for all other nonrespondents. NASS models nonresponse using two different list frame estimators -- the Reweighted Estimator and the Adjusted Estimator -- each of which assumes a different model for nonresponse.

The use of a particular estimator implies that a particular set of assumptions is valid concerning the characteristics of the nonrespondents. The research summarized in this report calls into question the validity of some of the assumptions on which the Reweighted and Adjusted estimators are based and outlines the

1

development of an alternative approach to modeling nonresponse for the hog portion of the QAS as outlined by Cox (1992). Readers not familiar with this approach to weight adjustment might find it helpful to refer to Cox's paper for a detailed description of the theory. The focus of this report is the application of the estimator rather than the theory behind the estimator.

One important aspect of modeling nonresponse is deciding how to form classes (cells) in which nonresponse adjustments are to be computed. Presently, NASS computes all nonresponse adjustments within list frame strata. Prior to December 1986, hogs were estimated through the use of a hog-specific survey; nonresponse adjustments computed within list frame strata were probably satisfactory. However, since then, crop acreage, grain storage, hogs, and other livestock surveys were integrated into one survey resulting in the present QAS survey design. This report points out the inefficiencies of using the design strata of the integrated QAS as nonresponse adjustment classes for hog total estimation and presents an alternative procedure for defining these classes.

CURRENT NONRESPONSE ADJUSTMENT PROCEDURES

The estimators currently in use for list frame hog total estimation model nonresponse differently. The models are different because they are based on different assumptions about the nonresponse mechanism. In developing an estimation procedure, it must first be decided what assumptions best describe this mechanism. Once a set of assumptions is decided upon, an estimator can then be developed.

Reweighted Estimator

The NASS Reweighted Estimator assumes that within a list frame stratum, the probability of being a complete respondent is the same for each sampled unit. A complete respondent is defined to be any unit for which the number of hogs associated with that unit is known. Thus, complete respondents include sampled units that are known to no longer be in business, agricultural operations that are known to raise no hogs, and agricultural operations that raise hogs and have known numbers of hogs. Complete respondents are assumed to represent a simple random sample of the original sample. Cox (1992) explains why this model for nonresponse is likely to be inappropriate under the integrated QAS design.

The expression for the Reweighted Estimator is given below:

$$\hat{Y}_{RW} = \sum_{h=1}^{H} \hat{Y}_h$$

where:

$$\hat{Y}_h - \sum_{i-1}^{n_h} \frac{N_h}{r_h + a_h + b_h} * x_{hi} \qquad (1)$$

and:

N_h = the number of classified list units in stratum h,

r_h = the number of hog operations in the stratum h sample that are complete respondents,

a_h = the number of known non-agricultural (out of business) units in the stratum h sample,

b_h = the number of known non-hog agricultural operations in the stratum h sample, and

x_{hi} = the number of hogs reported by respondent i in the stratum h sample.

Adjusted Estimator

When considering the form of the Adjusted Estimator and later, the alternative estimator (referred to as the _Revised Estimator_ in the sequel), it will be helpful to think of the estimation procedure as consisting of a sequence of three specific steps. For each sampled unit, three determinations need to be made. These are:

1) the sampled unit's status as an agricultural operation (ag-status)

2) the sampled unit's status as a hog operation (hog-status)

3) the sampled unit's status as a hog-total respondent (hog-total status).

The Adjusted Estimator (developed by Crank, 1979) can be expressed in the following form:

$$\hat{Y} - \sum_{h-1}^{H} \hat{Y}_h$$

where:

$\hat{Y}$ = the estimated population total number of hogs in a particular state for the list frame and

3

$\hat{Y}_h$ = the estimated number of hogs in list stratum h.

The estimator $\hat{Y}_h$ has the form:

$$\hat{Y}_h = N_h \, \hat{P}_h \, \overline{X_h} \qquad\qquad (2)$$

where:

N_h = the number of classified units in list stratum h,

$\hat{P}_h$ = the estimated proportion of units that currently have hogs in list stratum h, and

$\overline{X_h}$ = the estimated average number of hogs **per hog operation** in list stratum h.

Finally, the stratum h proportion of hog operations is estimated as:

$$\hat{P}_h = \frac{r_h + k_h}{r_h + k_h + a_h + b_h} \qquad\qquad (3)$$

where:

r_h = the number of hog operations in list stratum h that are complete respondents,

k_h = the number of known hog operations in list stratum h that are hog-total nonrespondents,

a_h = the number of known non-ag (out of business) units sampled in stratum h, and

b_h = the number of known non-hog ag units sampled in stratum h.

Two assumptions concerning the characteristics of nonresponse are implied by the Adjusted Estimator.

<u>Assumption 1.</u> The probability that hog-status will be determined is the same for all sampled units in a particular stratum. This implies that the hog-status respondents represent a simple random

sample of the original stratum sample.

Assumption 2. Within a stratum, amongst all units which have been determined to be hog operations, the probability that hog totals will be given is the same for each unit. This implies that within a stratum, hog operations that are complete respondents represent a simple random sample of all sampled records which have been determined to be hog operations.

Possible Weaknesses of the Adjusted Estimator

One problem with the Adjusted Estimator is that the design strata are not very effective weighting classes. A weighting class is defined as a group of units within which nonresponse weight adjustments are computed and applied. It is desirable for classes to be composed of units having similar target characteristics so that nonrespondents will be reasonably represented by respondents. Unfortunately, design strata do not have this property because the QAS is a multipurpose survey and the strata are designed to accommodate a variety of commodities. Stratification is not optimal for any one commodity. Survey responses regarding hogs might be quite variable within certain strata causing the resultant estimate to have a needlessly large variance.

Another factor that can contribute to needlessly large variances is the small number of responding hog operators in some strata. Due to a fairly large number of design strata in many states, the hog operation respondents are spread over a relatively large number of weighting classes. This results in many strata having very few hog operation respondents. This can cause excessive variance in the estimate. A popular rule of thumb is for weighting classes to contain at least 20 respondents at each nonresponse adjustment stage.

Another problem with the Adjusted Estimator is the possibility of a downward bias in the estimates it produces. As will be explained below, the validity of Assumption 1 is doubtful. If the assumption is indeed false, the model for nonresponse implied by the Adjusted Estimator is incorrect, causing a bias in the estimate. A better nonresponse model is needed to reduce this bias.

The Validity of Assumption 1 and the Model for Nonresponse

When the Adjusted Estimator was developed by Crank in 1979, partial information regarding ag-status was not being captured for any units with unknown hog-status. As stated earlier, Assumption 1 asserts that the probability of hog-status being determined is the same for all sampled units in the stratum. In light of the partial information concerning ag-status that is currently available, the original sample can be separated into two groups-- those units for which ag-status has been determined and those units for which ag-status has not been determined. Clearly the probability of hog-

5

status being determined is not the same for these two groups as hog-status determination implies ag-status determination. It is thus desirable to augment the nonresponse model so that the probability of determining ag-status is the same for all sampled units, while the probability of determining hog-status is the same for all sampled units which are **known to be ag-operations**.

It can be shown that under this augmented nonresponse model, the Adjusted Estimator is biased downward. In ignoring partial information concerning ag-status, the Adjusted Estimator estimates the probability of a randomly selected unit with unknown hog-status being a hog operation as though its ag-status was unknown. However, all hog-status nonrespondents can be divided into two groups-- those which are known to be ag-operations and those for which ag-status is unknown. It is reasonable to assume that units that are known to be ag-operations are more likely to be hog operations than units with unknown ag-status. Thus, under the augmented model, the Adjusted Estimator's estimate of the likelihood of a randomly selected unit with unknown hog-status being a hog operation is generally too low-- resulting in a downward bias.

ALTERNATIVE NONRESPONSE ADJUSTMENT PROCEDURES-- THE REVISED ESTIMATOR

The Revised Estimator as proposed by Cox (1992) implies the augmented nonresponse model described above. It treats units with unknown ag-status as less likely to be hog operations than units that are known to be ag-operations.

A complete respondent will now be defined as any unit for which its status for all three characteristics has been determined (ag-status, hog-status, and hog-total status). Any unit for which any of the three status determinations cannot be made is a nonrespondent.

Due to the inadequacies of the design strata as weighting classes, the Revised Estimator uses poststratification based on hog control data to form more homogeneous weighting classes. In addition, when implementing the Revised Estimator, checks are made and action taken to obtain a sufficient number of respondents within a weighting class at each adjustment stage for a more precise estimate. A detailed mathematical description of the Revised Estimator and the assumptions implied by the use of this estimator are presented in Cox (1992). The form of the estimator and the underlying assumptions are summarized here for completeness.

Description

Each unit's _Revised Expansion Factor (Weight)_ is composed of the product of the following six components.

1) _The DAF (Data Adjustment Factor)_. Some units on the list frame might be linked to other units on the list frame in the same stratum or a stratum with higher or lower priority. The DAF is used to adjust selection probabilities to account for multiple chances of selection in the highest priority stratum. It is also used to "zero out" units that are linked to a unit(s) in a higher priority stratum. The DAF is applied to the survey response and does not figure into the nonresponse adjustment.[1]

2) _The Original Sampling Weight_. This is simply the inverse of each unit's probability of selection. The sampling weight for unit i in stratum h is expressed as:

$$W_{sam}(hi) = \frac{N(h)}{n(h)}$$

where:

$N(h)$ = total classified units on the frame for stratum h
   and

$n(h)$ = the number of units sampled from stratum h.

3) _The Poststratification Adjustment Factor_. As stated earlier, the QAS is an integrated survey and the design strata represent a compromise between several commodities for which estimates are desired. To gain more homogeneity with respect to hogs within weighting classes, each sampled unit is poststratified into a poststratum based on its hog control value. These poststrata serve as the weighting classes for the first nonresponse adjustment. The poststrata are defined using the QAS hog strata definitions for each state. Within these poststrata, sampling weights can be summed to yield an estimate of the total number of classified units on the frame that would fall into a particular poststratum.

---

[1] This method of dealing with duplication is valid only under the assumption that all duplication is detected after a completed interview. This may not actually be a valid assumption for the QAS but current procedures make it impossible to determine at what point in the survey process the duplication was detected. See Cox (1992) for further details.

In order to reduce inefficiencies due to a possible disproportionate representation of the population by the sample in regards to its distribution across the weighting classes, a poststratification adjustment is applied to the sampling weights. These adjusted sampling weights, when summed within weighting classes will produce the list frame total for that weighting class. In this investigation, typical values for this adjustment factor lie between 0.80 and 1.10.

The poststratification adjustment factor is:

$$A_{ps}(c) = \frac{N(c)}{\sum_{hi\ \epsilon\ S(c)} W_{sam}(hi)}$$

where:

$N(c)$ = the total number of classified units on the frame for poststratum c (weighting class c) and

$S(c)$ = the set of all sampled units falling in poststratum c (weighting class c).

The poststratification adjusted weight is:

$$W_{ps}(chi) = W_{sam}(hi)\ A_{ps}(c) \quad .$$

Sarndal et al. (1991) discuss the advantage of the poststratification adjustment even in the absence of nonresponse.

4) <u>The Ag-Status Nonresponse Adjustment Factor</u>. For each unit, the first determination that must be made is:

The unit's status as an agricultural operation. (Does the unit represent an agricultural operation, is it out of business, or is its ag-status unknown?)

<u>Assumption 1:</u>

Within a particular ag-status weighting class (weighting class c; c= 1,...,C) the sampled units for which agriculture status is known represent a random sample of the total sampled units comprising weighting class c. Ag-status determination is assumed to be purely random within weighting class c.

The ag-status nonresponse adjustment factor is:

$$A_{AG-ST}(c) = \frac{\sum_{i \, \epsilon \, S(c)} W_{ps}(chi)}{\sum_{i \, \epsilon \, S_{AG+nonAG}(c)} W_{ps}(chi)}$$

where:

$S(c)=$ The set of all sampled units in weighting class c and

$S_{AG+nonAG}(c)=$ the set of all sampled units in weighting class c with ag-status determined (this would be all non-ag units and all known ag-operation units).

The ag-status adjusted weight is expressed as:

$$W_{AG-ST}(chi) = W_{ps}(chi) \; A_{AG-ST}(c) \quad .$$

Note:
All unknown ag-status units (code 921= 12) are given an ag-status adjusted weight of zero. For all units determined to be non-agricultural units (921 code= 9), hog-status and hog-total status determinations can be made immediately (by definition they are non-hog operations and have zero hogs). They can thus be considered as complete respondents. Because non-ag units will have identical characteristics concerning hog-status and hog totals, they are now placed in a separate weighting class. Within this weighting class, subsequent nonresponse adjustments are equal to one. For each unit in this weighting class, the hog-status adjusted weight will equal the ag-status adjusted weight.

5) <u>The Hog-Status Nonresponse Adjustment Factor</u>. For all units determined to be ag-operations, the second, determination that must be made is:

> **The unit's status as a hog operation.** (Does the unit represent a hog operation or is its hog-status unknown?)

<u>Assumption 2:</u>

Within a particular hog-status weighting class (weighting class d; d= 1,....,D) the sampled known ag-operations for which hog-status is known represent a random sample of the total sampled units comprising weighting class d. Hog-status determination is assumed to be purely random within weighting class d.

The hog-status nonresponse adjustment factor is expressed as:

$$A_{HOG-ST}(d) = \frac{\sum_{i \, \epsilon \, S_{AG}(d)} \, W_{AG-ST}(chi)}{\sum_{i \, \epsilon \, S_{HOG-AG+nonHOG-AG}(d)} \, W_{AG-ST}(chi)}$$

where:

$S_{AG}(d) =$ The set of all sampled units in weighting class d (these are identified agricultural operations only), and

$S_{HOG-AG+nonHOG-AG}(d) =$ the set of all sampled units in weighting class d known to be hog operations or an agricultural operation known not to raise hogs.

The hog-status adjusted weight is expressed as:

$$W_{HOG-ST}(cdhi) = W_{AG-ST}(chi) \, A_{HOG-ST}(d) \quad .$$

Note:
All ag-operation units with unknown hog-status (code 499= 2 and code 921 not equal 12), are given a hog-status adjusted weight of zero. For all ag-operation units that are now determined to be non-hog agricultural operations (code 499= 3 and code 921 ≠ 9 and ≠ 12), the number of hogs on these operations is immediately known (zero hogs) and are placed in the same weighting class for the hog-total status nonresponse adjustment as the units that were identified to be non-ag units. Like the non-ag units described earlier, identified non-hog ag-operations can be considered as complete respondents. For both non-ag units and non-hog ag-operations, the hog total nonresponse adjustment factor is equal to one and the hog-total adjusted weight is equal to the hog-status adjusted weight.

6) <u>The Hog-Total Status Nonresponse Adjustment Factor</u>.
For all units determined to be hog-ag operations, the third determination that must be made is:

> **The unit's hog total information.** (Is the number of hogs known for the unit?)

<u>Assumption 3</u>:

Within a particular hog-total status weighting class (weighting class e; e= 1,.....,E) the sampled known hog operations that indicate the number of hogs on their operation represent a random sample of the total sampled units comprising weighting class e. Hog-total status response is assumed to be purely random within weighting class e.

10

The hog-total status nonresponse adjustment factor is:

$$A_{totHOG-AG}(e) = \frac{\sum_{i \in S_{HOG-AG}(e)} W_{HOG-ST}(cdhi)}{\sum_{i \in S_{totHOG-AG}(e)} W_{HOG-ST}(cdhi)}$$

where:

$S_{HOG-AG}(e)=$ The set of all sampled units in weighting class e (these are identified hog operations only), and

$S_{totHOG-AG}(e)=$ the set of all units in weighting class e that are complete respondents.

The hog-total status adjusted weight is expressed as:

$$W_{totHOG-AG}(cdehi) = W_{HOG-ST}(cdhi) \; A_{totHOG-AG}(e) \quad .$$

Note:
All known hog operation units that are hog-total status nonrespondents (code 499= 1) are given a hog total adjusted weight of zero. For all hog operation units for which hog numbers are known (code 499= 0) and all other units which are complete respondents, the Revised Expansion Factor is:

$$W_{Complete-Resp}(cdehi) = W_{totHog-AG}(cdehi) \; DAF(cdehi) \quad .$$

Note:
The form of the Revised Estimator can easily be changed to accommodate different assumptions about nonrespondents. For example, if all interviewing is done face to face, one might be able to reasonably assume that any unit that is an ag-status nonrespondent is in business (an agricultural operation).

Weighting Classes - Poststratification Process

The first step in computing the Revised Estimate is to poststratify all sampled units into the initial weighting classes. The objective of poststratification is to group together units with similar hog characteristics.

The QAS has primarily three types of strata: cropland acreage, grain storage capacity, and hog strata. Within each type of stratum, there are size categories(i.e. small, medium, and large). Units are placed in a stratum using a priority ordering scheme that looks at the magnitude of the control data value for each of the three items.

11

**Table 1: Stratification Priority Scheme
for Indiana**

Priority	Stratum	Boundaries
1	98	HOGS 8000+
2	97	CAPACITY 500K+
3	95	CROPLAND 5000+
4	82	HOGS 2000-7999
5	80	HOGS 1000-1999
6	73	HOGS 400-999
7	72	CAPACITY 100K-499999
8	71	CROPLAND 1500-4999
9	70	HOGS 200-399
10	69	HOGS 100-199
11	67	CAPACITY 25K-99999
12	65	CROPLAND 600-1499
13	64	HOGS 1-99
14	63	CAPACITY 1-24999
15	62	CROPLAND 150-599
16	61	CROPLAND 1-149

Table 1 shows the 1991-1992 QAS stratification priority scheme for Indiana. To determine in which stratum a unit should be placed, a series of checks must be made. Suppose a unit has a cropland value of 1000 acres. This unit qualifies for stratum 65 based strictly on its cropland control value. Next it is determined what stratum the unit is qualified for based strictly on its capacity control value. Suppose this same record has a capacity control value of 250,000. This unit qualifies for stratum 72 based strictly on its capacity control value. Finally, it is determined which stratum the unit qualifies for based on its hog control value. Suppose this unit has a hog control value of 100. Thus the unit qualifies for stratum 69 based strictly on its hog control value. Since a unit can only be placed in one stratum, it is placed in the highest priority stratum for which it qualifies. For this example unit, stratum 72 is the highest priority stratum for which it qualifies.

To show one problem with using design strata as weighting classes for producing hog estimates, consider the following situation. Two units on Indiana's list are placed in stratum 71 based on their cropland control data. One of these units has a hog control data value of 350. The other has a hog control data value of 0. These two units may have similar characteristics with respect to acreage but are likely to have very dissimilar survey responses with respect to hogs.

12

The poststratification procedure looks only at the hog control value for each unit and places the unit in a weighting class based on this value. This insures to the greatest extent possible that units with similar hog characteristics will be placed in the same weighting class. Of course if the hog control value is not correlated with the response regarding hog totals, poststratification is of little use.

Table 2: Poststrata for Indiana

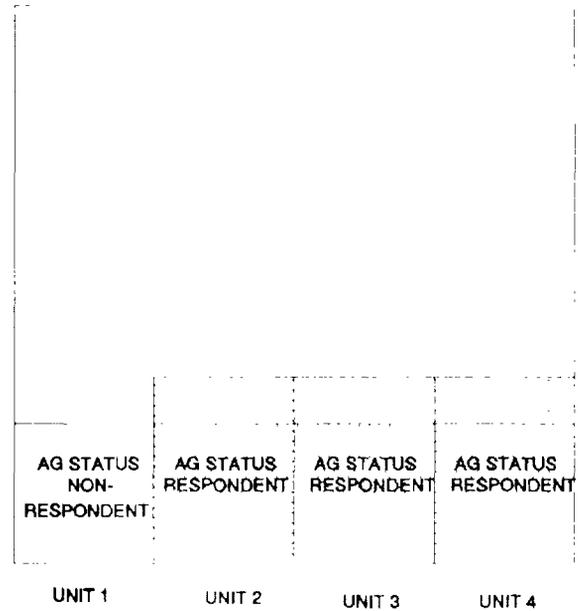| Poststratum	Boundaries
1	Missing Hog Control Data
2	HOGS 0
3	HOGS 1-99
4	HOGS 100-199
5	HOGS 200-399
6	HOGS 400-999
7	HOGS 1000-1999
8	HOGS 2000-7999
9	HOGS 8000+

The poststrata are defined for each state using that state's design hog strata intervals. Two additional poststrata are formed. One poststratum for units with missing hog control data and one poststratum for units with hog control values of zero. Table 2 shows the poststrata for Indiana. These poststrata are treated as weighting classes and they will now be referred to as the AG-STATUS weighting classes.

The ag-status nonresponse adjustment is done within the AG-STATUS weighting classes. A check is made to make sure that the number of units for which ag-status is determined (ag-status respondents) is sufficiently large. An ag-status respondent is defined to be any unit for which ag-status has been determined. This includes all units determined to be non-agricultural and those units determined to be agricultural operations. (Usually, if the number of respondents at any adjustment stage for a weighting class is less than 20, adjacent weighting classes are collapsed together.)

Iowa was the only state in this study for which it was necessary to do any collapsing at the ag-status adjustment stage. The AG-STATUS weighting class composed of units with missing control data was collapsed with the AG-STATUS weighting class composed of units with hog control data equal to zero. This collapse was necessary because of the low number of units with missing hog control data appearing in Iowa samples.

13
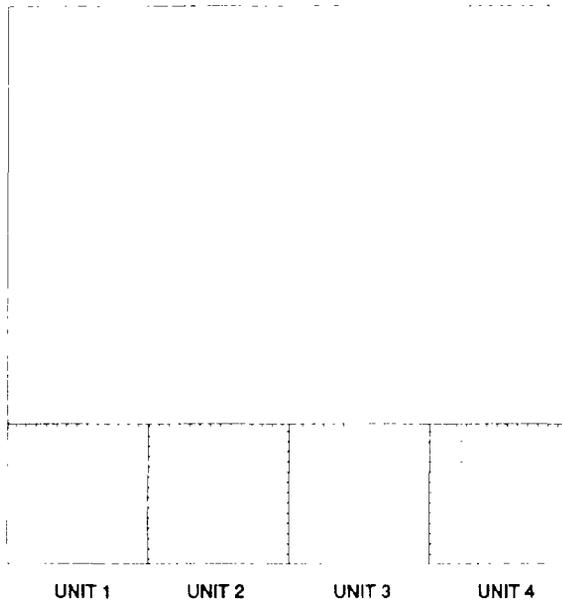
Figure 1: Poststratified
Sampling Weights
[W_{ps}(chi)]

| AG STATUS NON-RESPONDENT | AG STATUS RESPONDENT | AG STATUS RESPONDENT | AG STATUS RESPONDENT |
| UNIT 1 | UNIT 2 | UNIT 3 | UNIT 4 |

Figure 2: Weights After Ag-Status
Nonresponse Adjustment
[W_{AG-ST}(chi)]

When the ag-status nonresponse adjustment is calculated, each unit
for which ag-status is unknown is given a final analysis weight of
zero. Within each AG-STATUS weighting class, the weight of each
ag-status nonrespondent is set to zero and effectively distributed
over all the ag-status respondents. For an illustration of this
concept, the reader is referred to a simplified description of the
nonresponse adjustment process shown in Figures 1-4. This
illustration is not meant to describe the weight adjustment
procedure in all its complexity but to demonstrate how the
adjustment procedure preserves sampling weight totals in the
presence of nonresponse. In Figure 1 there are four units. Assume
that these four units are from the same design stratum and thus
have equal poststratification adjusted sampling weights (the weight
of each record is represented by the shaded area). (A word of
caution here. All records from the same design stratum that are in
the same AG-STATUS weighting class will have the same
poststratification adjusted sampling weight. This does not mean
that all records in a particular AG-STATUS weighting class have
equal poststratification adjusted sampling weights!)

As can be seen in Figure 2, unit 1 is an ag-status nonrespondent.
Thus unit 1 has an ag-status adjusted weight of zero. The weight
for unit 1 is effectively distributed equally over the three
responding units. At this point, all ag-status nonrespondents

14

(i.e., unit 1) could be dropped from the data base since they no longer have any weight associated with them.

Hog-status nonresponse adjustments are computed within weighting classes that will be referred to as the HOG-STATUS weighting classes. A hog-status respondent is defined to be any unit for which hog-status has been determined. This includes all units known to be non-ag units, all agricultural operations that have indicated that they have no hogs, and all agricultural operations that have indicated that they do raise hogs.

All sampled units that are identified to be non-ag units are placed in a separate HOG-STATUS weighting class. This is needed because members of a particular HOG-STATUS weighting class should have similar characteristics concerning hog-status. Hog-status is known for all non-ag units by definition (they produce no hogs). Thus, there is no nonresponse regarding hog-status for this weighting class. For these units, the hog-status nonresponse adjustment factor is equal to one.

All other HOG-STATUS weighting classes are composed entirely of ag-operations. For these HOG-STATUS weighting classes, a count of the number of hog-status respondents is made. If the number of respondents for a particular HOG-STATUS weighting class is too small, then that class is collapsed with an adjacent class. In the research conducted for this report, no collapsing was ever needed at this stage.

Like before, the weight of all units that are hog-status nonrespondents in a particular HOG-STATUS weighting class is effectively distributed over all the respondents within that class. (Figure 3 shows that unit 2 is a hog-status nonrespondent. The ag-status adjusted weight of unit 2 is now divided equally between unit 3 and unit 4. The weight of unit 2 is zero.)

15
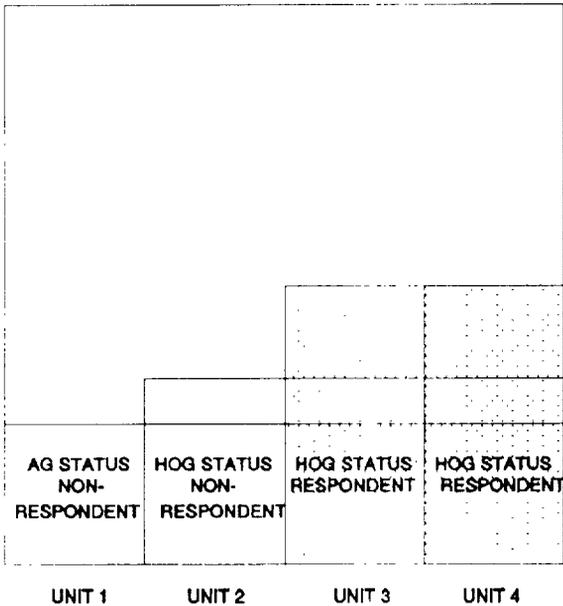
Figure 3: Weights After Hog-Status Nonresponse Adjustment $[W_{HOG-ST}(cdhi)]$
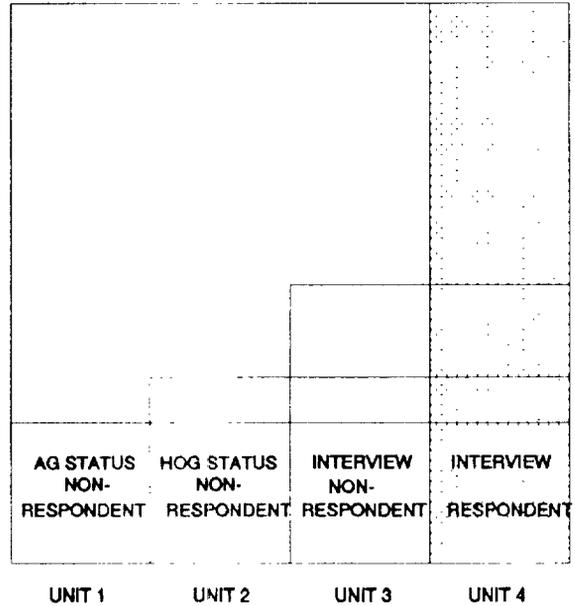
Figure 4: Weights After Hog-Total Nonresponse Adjustment $[W_{totHOG-AG}(cdehi)]$

After the necessary adjustments have been made at the hog-status stage, all hog-status nonrespondents (i.e., unit 2) can be dropped from the data base.

All hog-total status nonresponse adjustments are computed within weighting classes that will be referred to as the HOG-TOTAL STATUS weighting classes. A hog-total status respondent is defined as any unit for which the number of hogs operated by that unit is known. Hog-total status respondents include all units known to be non-agricultural, all units known to be non-hog agricultural operations, and all hog-producing agricultural operations for which the number of hogs on hand is known. All sampled units that are known to be non-hog agricultural operations can be placed in the same HOG-TOTAL STATUS weighting class as the identified non-ag units. Note that these units all have identical hog total responses (they all have zero hogs) and there is no nonresponse concerning hog totals. Thus the hog-total status nonresponse adjustment is equal to one for these units. All other HOG-TOTAL STATUS weighting classes are composed entirely of units that are known to be hog operations.

Like before, a check is made to determine if a sufficient number of respondents exist in each class and collapsing of adjacent classes is performed if necessary. When adjustments for hog-total status

nonresponse are computed, the weight for nonrespondents is effectively distributed over the respondents. For the example, unit 4 is the only complete respondent, the hog-status adjusted weight of unit 3 is added to the hog-status adjusted weight of unit 4. The hog-total status adjusted weight of unit 3 is now zero. Note that the sum of the weights is preserved throughout the nonresponse adjustment process. Hog-total status nonrespondents (i.e., unit 3) can now be dropped from the data base. At this point all nonresponse adjustments are complete. The result of the nonresponse adjustment procedure when multiplied by the DAF is the final weight for each unit. The number of hogs for each unit is expanded by the unit's final weight. Summing expanded totals over all units in the sample gives the Revised Estimator's estimate for total hogs.

In this study, some collapsing had to be done at the hog-total status adjustment stage. For all states except Iowa, (which was collapsed earlier), the HOG-TOTAL STATUS weighting class composed of known hog operations with missing hog control data had to be collapsed with the HOG-TOTAL STATUS weighting class composed of the known hog operations with hog control data equal to zero. This should not be surprising because few hog operations would be expected to be contained in these two weighting classes.

Summary of Factors Affecting the Difference Between Estimates Obtained by the Adjusted and Revised Estimators

The Revised Estimator is a complex function involving many different factors. These include (within list strata):

    1) the number of unknown ag-status units (code 921=12),
    2) the number of unknown hog-status units (code 499=2),
    3) the number of nonrespondent hog units (code 499=1),
    4) the number of non-ag units (code 921=9), and
    5) the number of ag-units with zero hogs.

In addition, the poststratification process influences the value of the estimate. The interaction of all these factors will determine the impact of the Revised Estimator's estimate as compared to the estimate produced by the Adjusted Estimator. The impact of the Revised Estimator will vary from state to state and from survey to survey depending on these counts, the effect of poststratification, and their interaction.

The reader is referred to the Appendix for a further discussion of the relationship between the Adjusted and Revised Estimators and their associated nonresponse models.

ANALYSIS

QAS survey data for June 1988 through December 1991 for Georgia,

Illinois, Indiana, Iowa, and North Carolina were used to compute estimates using the Revised Estimator.

These are five of the major hog producing states. Together, they account for approximately 50 percent of the total U.S hog production (USDA, September 1991). The validity of the Revised Estimator for these large hog states is an important factor in its evaluation. The September and December 1991 hog expansions were lower than expected by the Livestock Branch, so the performance of the Revised Estimator for these surveys was of special interest.

Analysis Level

Each unit on the list frame represents one tract as opposed to one operation. It is possible for a single tract to be composed of multiple operations **that are linked** to one operator or operation name. These operations **that are** linked together are called subtracts. Subtracts occur when new partners are detected in the interviewing process that yield additional operations other than the one originally selected for sampling. Another situation that will cause subtracts to occur is when one operation goes out of business and another newly formed operation is substituted for the original operation. QAS data are collected at the subtract level rather than the tract level. This permits the analysis to be done at either the subtract or the tract level. It was decided that all weighting for this report would be done at the tract level. This decision reflects the present NASS analysis level for hog estimation for the list frame. All subtracts **were** combined together to form one tract level unit **before** any poststratification or nonresponse adjustments were implemented. Because sampling is done at the tract level, all poststratification adjustments must be performed at the tract level. However, nonresponse adjustments can be performed at the subtract level if desired. Subtract level adjustments are likely to be preferable to tract level adjustments in some survey situations.

Unit Level Weights

The implementation of the Revised Estimator results in each unit in the sample having a specific weight attached to it. This weight can be thought of as the unit's expansion factor. The weight for all nonresponding units is zero. All complete respondents will have positive weights. In general, weights can differ among units in the same design stratum. The Adjusted Estimation procedure results in all responding units within a design stratum having the same weight. This includes non-ag units, non-hog ag-operations and hog operations alike.

Variance Estimation Procedure

The Revised Estimator models nonresponse as if the hog operations

18

that are hog-total respondents are the result of a poststratified four-phase sample design-- sample selection and poststratification (phase I), ag-status determination (phase II), hog-status determination (phase III), and hog total determination (phase IV). Due to the complexity of this nonresponse model, an exact design based expression for the variance is difficult to produce. When the variance of the adjusted weights is fairly small within the design strata and the respondent counts at each adjustment stage are reasonably high, an approximation for the variance can be obtained by using the poststratified nonresponse-adjusted weights under the assumption of a stratified simple random sample design. The data analysis software package SUDAAN was used to calculate an estimate of the variance in this manner.

Define:

Y = the total number of hogs in a particular state, and

$z_{hi} = (N_h/n_h)\ x_{hi}$

$x_{hi} = k_{hi}\ y_{hi}$

where:

N_h = the number of classified units in stratum h,

n_h = the number of sampled units with nonzero weights in stratum h,

k_{hi} = the product of the adjustments for the ith unit in stratum h,

y_{hi} = the number of hogs corresponding to complete respondent i in stratum h.

Also define:

$\hat{Y} = \Sigma_h \Sigma_i z_{hi}$

= Revised Estimator's estimate of total at the state level.

The form of SUDAAN's estimate of the variance is:

$$var(\hat{Y}) = \sum_h^H (1 - \frac{n_h}{N_h})\ n_h\ s^2_h \quad .$$

where:

$$S^2{}_h = \frac{\sum_i^{n_h}(z_{hi}-\overline{z_h})^2}{(n_h-1)} \qquad \overline{z_h} = \frac{\sum_i^{n_h}z_{hi}}{n_h}$$

As can be seen in Table 3, estimated CVs computed for the Revised Estimator tend to run about 0.1 percent above the estimated CVs for the Adjusted Estimator. The values shown here are typical for other surveys. The SUDAAN variance approximation does not fully take into account the effect of poststratification on the weighting class adjustment, nor the effect of using status information but is assumed here to be a reasonable approximation of the true variance. It is noted that comparing CVs produced by different estimators for evaluation purposes can be misleading if one or more of the estimators are significantly biased.

Table 3: Estimated CVs for March 1990 and June 1991

State	March 90			June 91		
	Rewght	Adjst	Rev	Rewght	Adjst	Rev
Georgia	4.1	4.0	4.3	3.1	3.2	3.3
Illinois	4.5	5.0	5.0	4.0	4.0	4.1
Indiana	2.9	2.9	3.0	3.3	3.2	3.3
Iowa	2.8	2.8	2.9	2.5	2.4	2.6
N. Carolina	3.3	3.5	3.5	4.0	4.1	4.2

RESULTS

The primary criterium for evaluating the Revised Estimator should be the set of assumptions from which it is derived. If these assumptions appear reasonable, the Revised Estimator should be implemented regardless of how it may shift the indications as compared with the Adjusted Estimator. The assumptions from which the Revised Estimator is derived are paraphrased below.

Assumption 1: Within a particular AG-STATUS weighting class, ag-status respondents represent a random sample of all sampled records comprising that weighting class.

Assumption 2: Within a particular HOG-STATUS weighting class, hog-status respondents represent a random sample of all sampled records comprising that weighting class.

Assumption 3: Within a particular HOG-TOTAL STATUS weighting class, hog-total status respondents represent a random sample of all sampled records comprising that weighting class.

These assumptions of purely random (non)response are probably never valid in the real world. However, any model for nonresponse will fall short of describing all the complexities of a nonresponse mechanism. It should be noted that the assumptions stated above concern relatively small groups of units having similar characteristics. If one could form weighting classes so that the units within a weighting class were identical with respect to target characteristics (i.e., the number of hogs), the reweighted estimators would be unbiased regardless of whether nonresponse was truly random. Thus, the more similar the units are within a weighting class, the more robust the model for nonresponse will be against departures from the random nonresponse assumption.

The above assumptions are actually relatively conservative compared to those implied by the Adjusted Estimator and Reweighted Estimator. For example, the Reweighted Estimator makes the assumption that within a stratum, **complete respondents** are a simple random sample from the entire original stratum sample. The Adjusted Estimator makes the more conservative assumption that within a stratum, **hog-status respondents** are a random sample of the entire original stratum sample. (Recall that for the Reweighted and Adjusted Estimators, the strata are the weighting classes.)

The Revised Estimator is more discreet than the Reweighted and Adjusted Estimators in its assumptions concerning nonrespondents. It does this by first making the assumption that within a particular AG-STATUS weighting class, **ag-status respondents** represent a random sample of the entire sample that comprises that particular weighting class. Then, within the HOG-STATUS weighting classes, hog-status is addressed separately.

Time Series Charts

To examine the effect the Revised Estimator has on the level of state list frame indications produced by the current estimators, estimates produced by the Revised Estimator were plotted against estimates produced by the Reweighted and Adjusted Estimators for 15 consecutive QAS surveys beginning with June 1988 and ending with December 1991 for the states of Georgia, Illinois, Indiana, Iowa, and North Carolina.

Bar charts illustrating the differences between estimates produced by the Adjusted and the Revised Estimator relative to the Adjusted Estimator's estimate are also provided.

21

GEORGIA: As can be seen in Figure 5, the Revised Estimator's indications for Georgia track well with the Reweighted and Adjusted Estimators' indications and are consistently higher than either one. The absolute difference between the indications given by the Revised and Adjusted Estimators appears to be increasing in more recent quarters. Figure 6 shows the relative differences between the estimate produced by the Revised and Adjusted Estimators for each quarter, expressed as a percent of the Adjusted Estimator's estimates for that quarter.
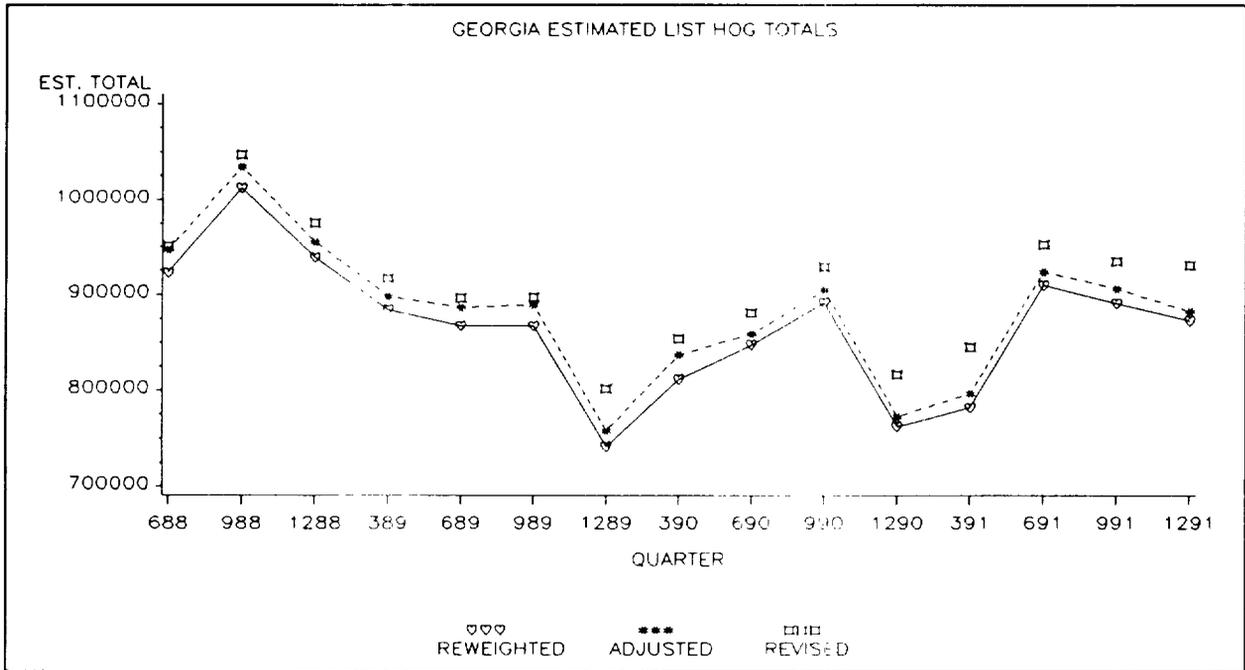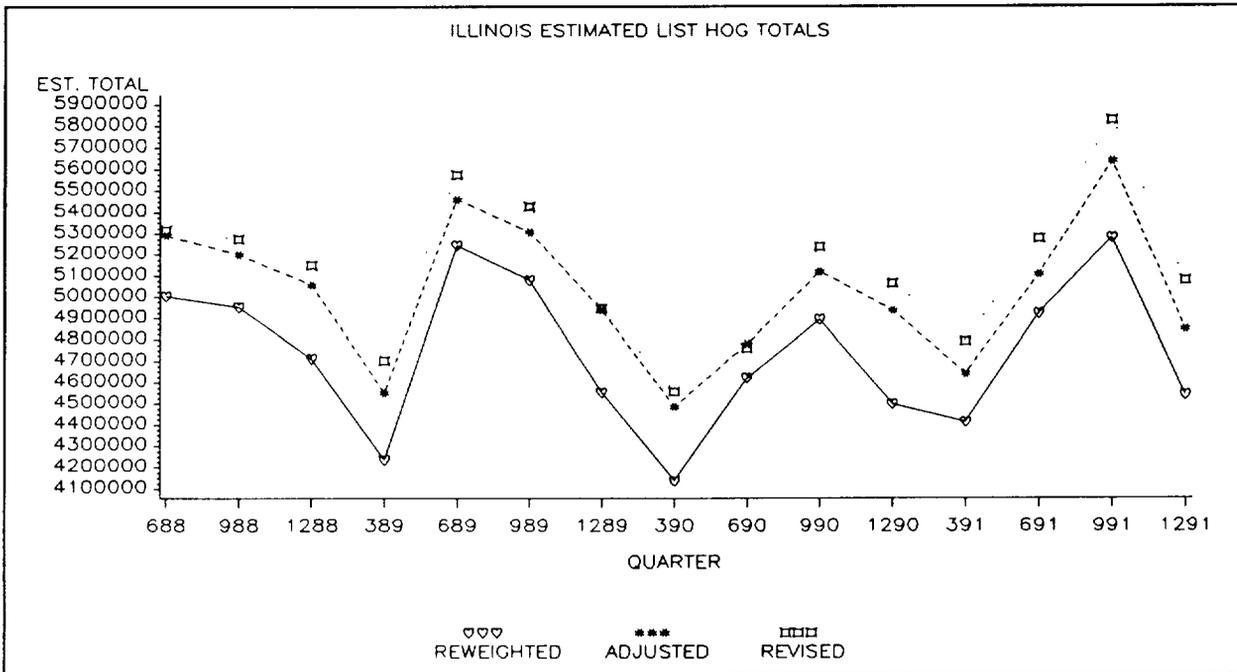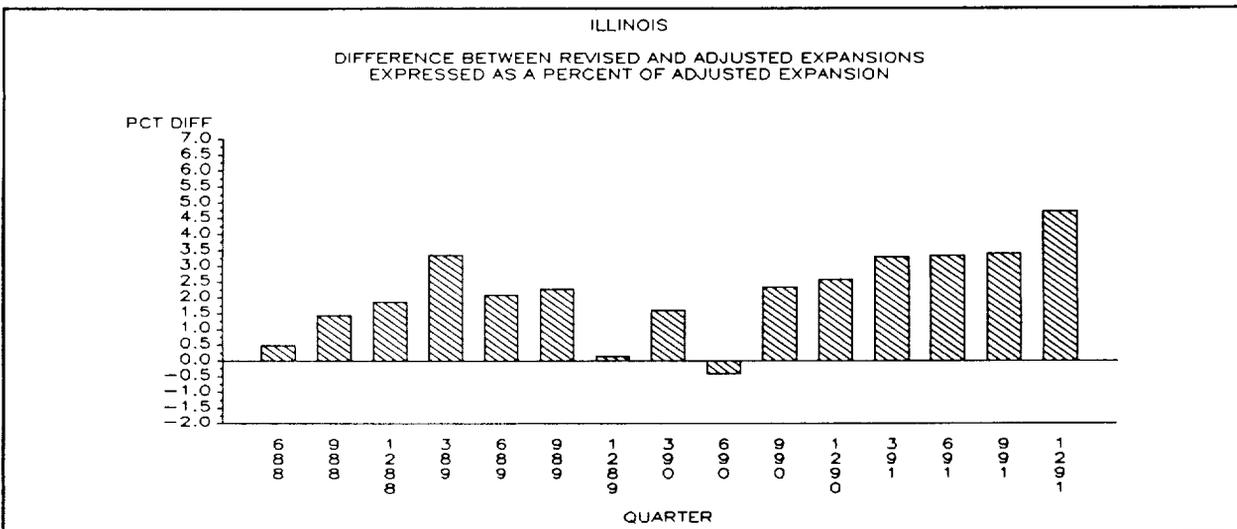

FIGURE 5
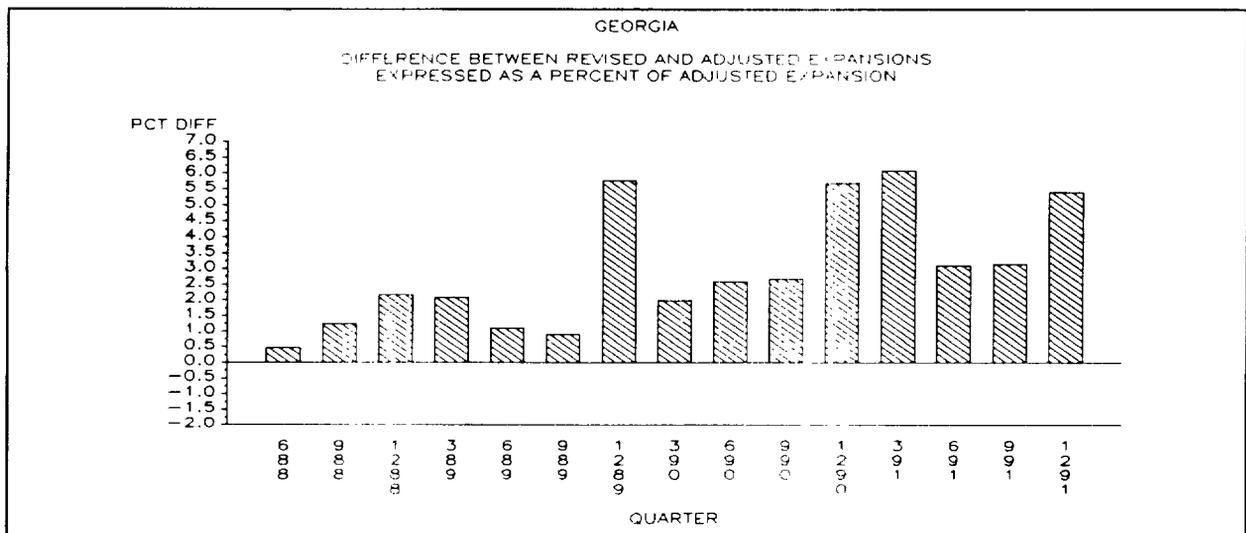
FIGURE 6

ILLINOIS: For Illinois, the Revised Estimator produces estimates that track fairly well with the current estimators. Its estimates are generally higher than those produced by the Adjusted Estimator, but not consistently so. In June 1990, the Revised Estimator gives a slightly lower estimate than the Adjusted Estimator's estimate. This can be seen in Figures 7 and 8.
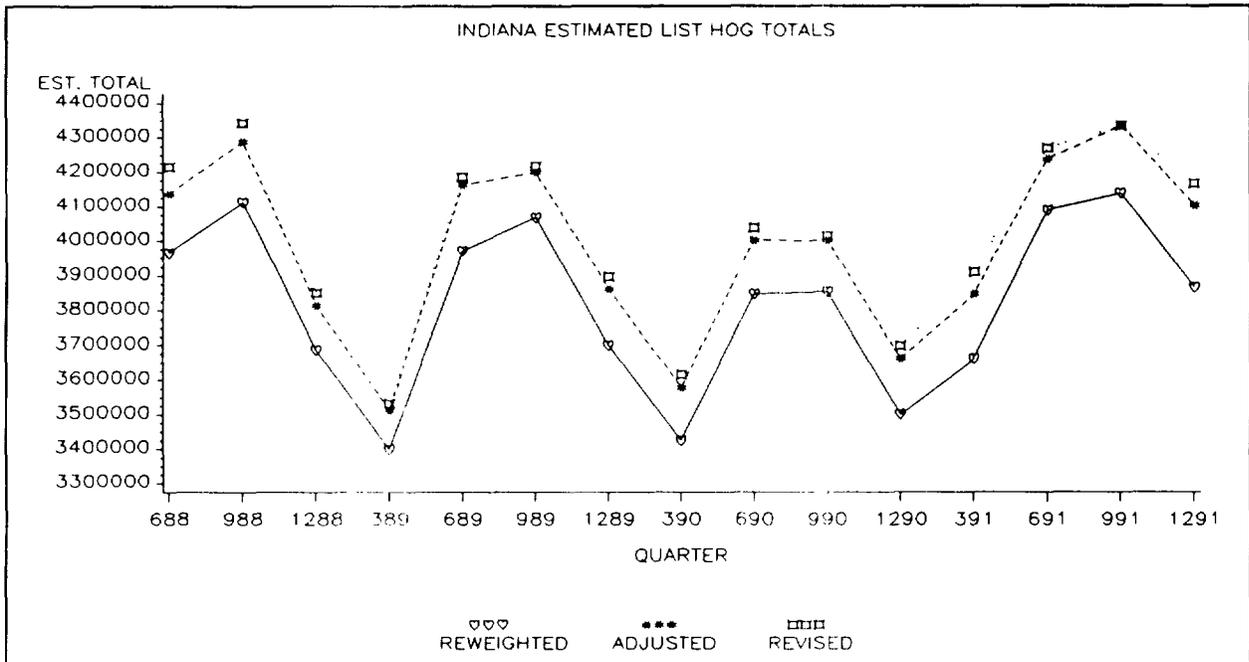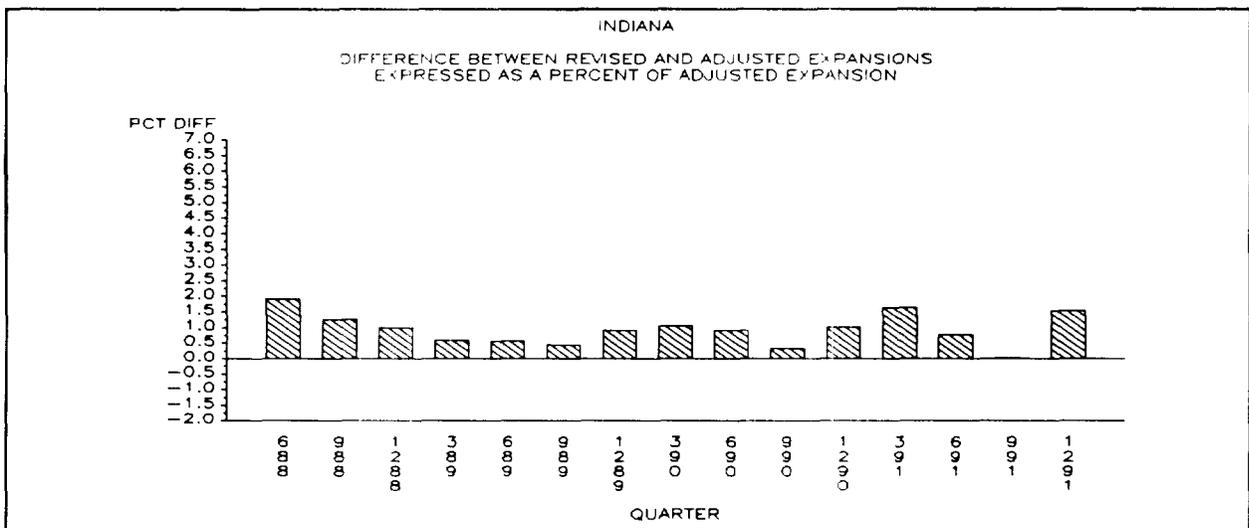

Figure 7

Figure 8

INDIANA: The Revised Estimator produced estimates that were consistently higher for Indiana than the Adjusted Estimator's estimates but the absolute and relative differences between the two are fairly small. This can be seen in Figures 9 and 10. The Revised Estimator tracks well with the other two estimators.
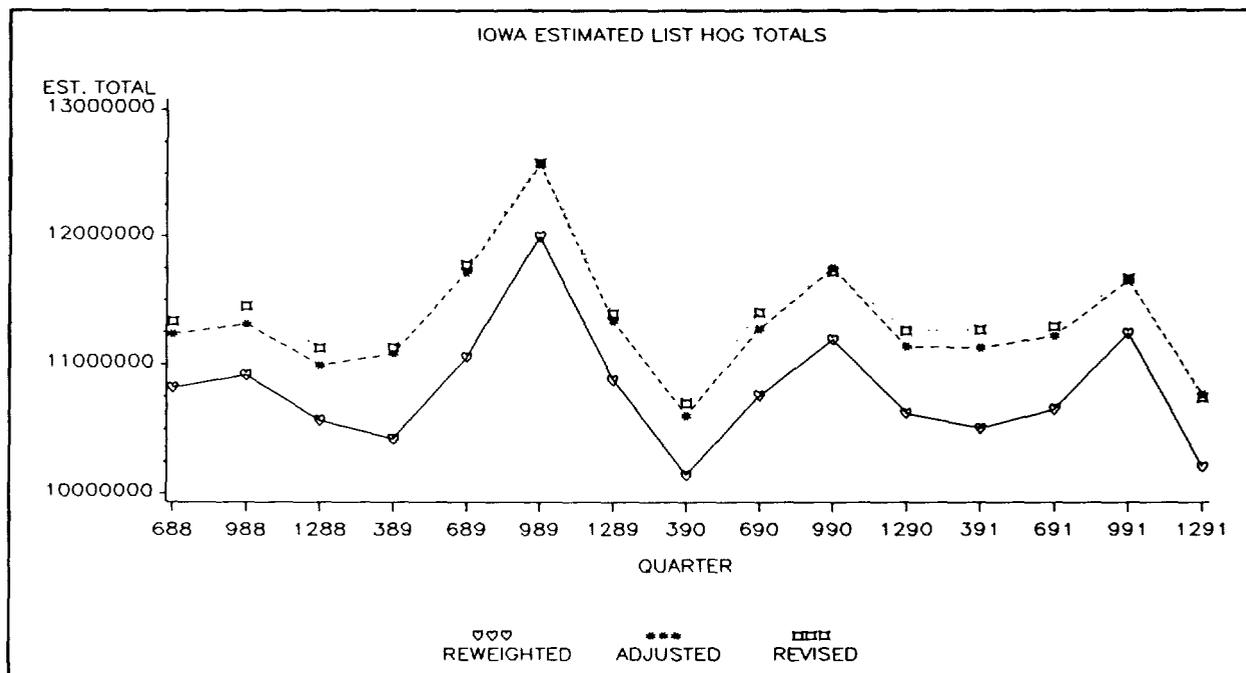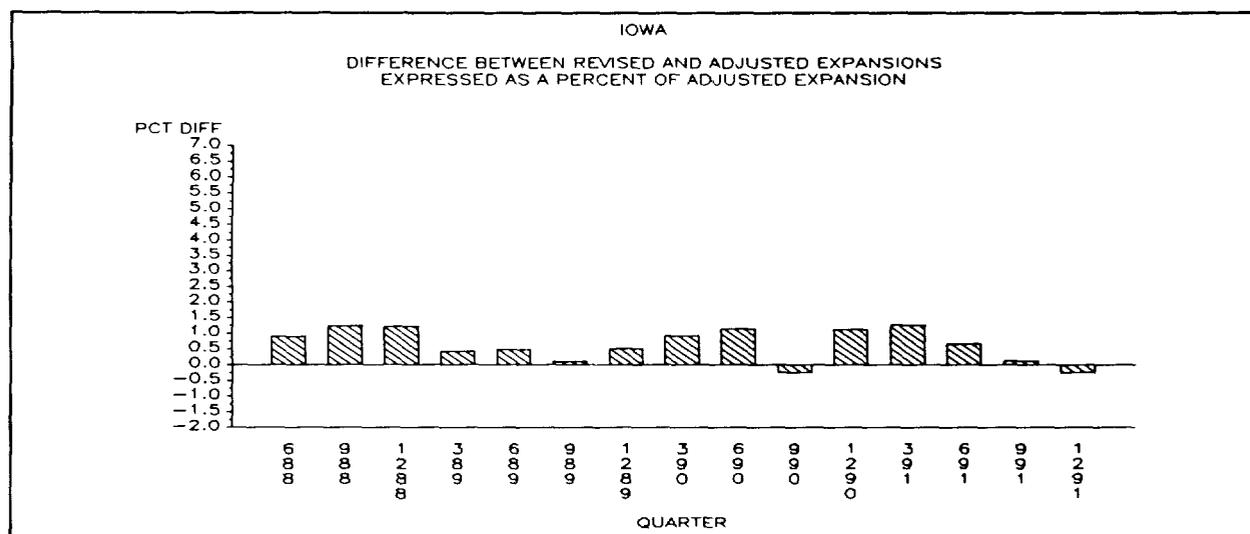

Figure 9

Figure 10

24

IOWA: For Iowa, the Revised Estimator gives generally higher estimates than the Adjusted Estimator but not consistently so. For two quarters, the Revised Estimator gives slightly lower estimates. The Revised Estimator tracks well with the other estimators with relatively small relative and absolute differences. This can be seen in Figures 11 and 12.

Figure 11

Figure 12

NORTH CAROLINA: For North Carolina, the behavior of the Revised Estimator in relation to the Adjusted Estimator is slightly erratic. For many of the earlier quarters, the Revised Estimator gives slightly lower estimates than the Adjusted Estimator. As shown in Figure 13, the Revised Estimator is tracking fairly well with the other estimators, but not as well as it did in the other four states. Figure 14 shows that the relative difference between these two estimators tends to be increasing significantly from June 1990 on. The Revised Estimator's estimates are consistently higher than the Adjusted Estimator's estimates since September 1990.
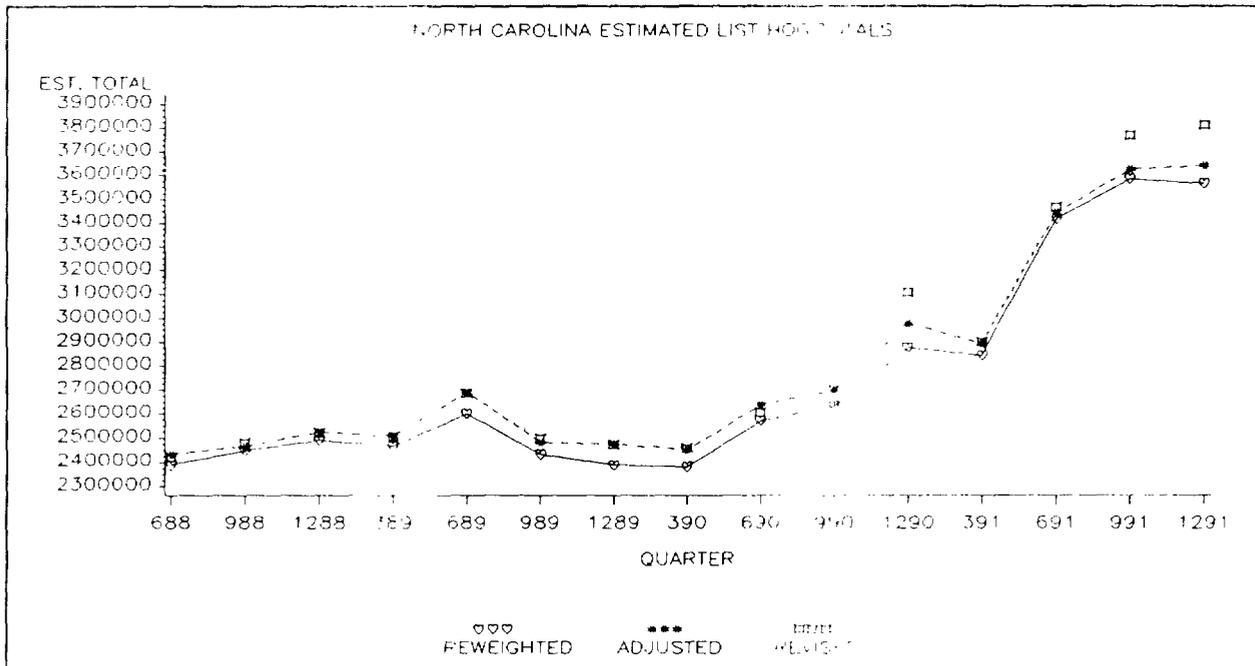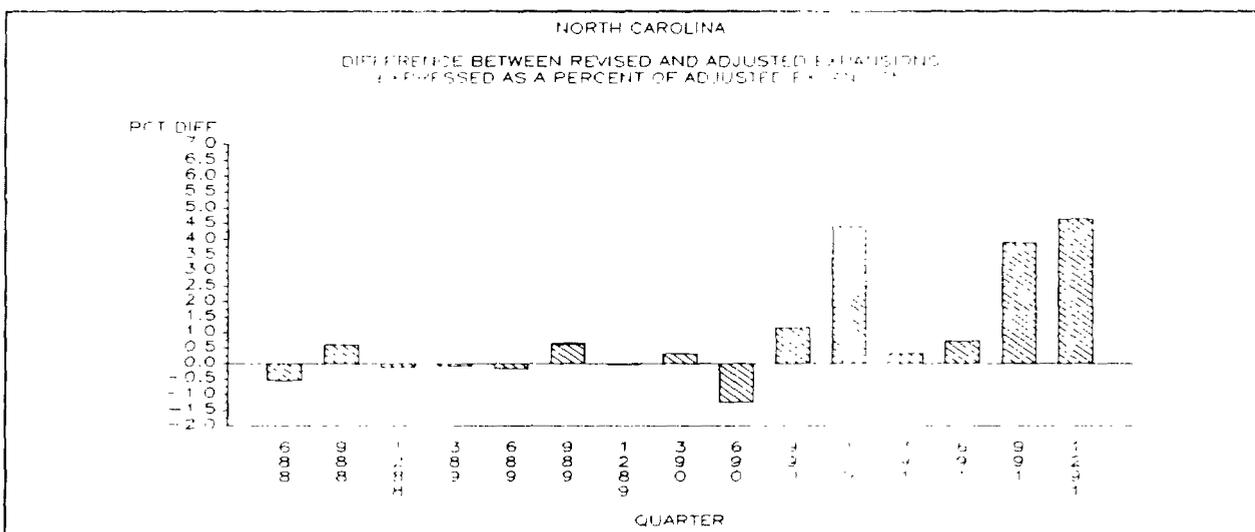

Figure 13

Figure 14

26

In Table 4, some summary statistics for the relative differences between the estimates produced by the Revised and Adjusted Estimators are presented.

Table 4

SUMMARY OF RELATIVE DIFFERENCES BETWEEN ESTIMATES PRODUCED BY THE REVISED ESTIMATOR AND THE ADJUSTED ESTIMATOR OVER 15 CONSECUTIVE SURVEYS [(REVISED - ADJUSTED)/ ADJUSTED]* 100

STATE	MINIMUM DIFFERENCE	MAXIMUM DIFFERENCE	MEAN DIFFERENCE
GEORGIA	+ 0.47 %	+ 6.09 %	+ 2.96 %
ILLINOIS	- 0.39 %	+ 4.72 %	+ 2.17 %
INDIANA	+ 0.01 %	+ 1.91 %	+ 0.93 %
IOWA	- 0.25 %	+ 1.25 %	+ 0.64 %
NORTH CAROLINA	- 1.23 %	+ 4.65 %	+ 0.97 %

CONCLUSIONS

The Revised Estimator for the list frame produces estimates of total hogs that are approximately 1 to 3 percent higher on average than those produced by the Adjusted Estimator. The percentage change varies from survey to survey and state to state dependent on response rates and the effect of poststratification.

The Adjusted Estimator was based on the most reasonable assumptions that could be made given the amount of partial information available at the time the estimator was developed. Additional information is now available making possible more reasonable assumptions, fully using all of the information currently available about nonrespondents.

The CVs calculated using a SUDAAN approximation of the variance closely match the CVs of the Adjusted Estimator. Even though there is no evidence that the variance of the Revised Estimator is lower, the variance estimates for the Revised Estimator should be more stable than the variance estimates for the Adjusted Estimator due to a larger number of respondents in each weighting class at each nonresponse adjustment stage.

To implement the Revised Estimator, a number of issues surface that need to be addressed. One of these is the need to know frame counts for poststratification adjustments. Obtaining the frame counts for old surveys is not a trivial matter but if provisions

are made to obtain these counts during the annual classifying process, they should be reasonably easy to obtain.

Another more difficult problem might be the need to collapse weighting classes to obtain sufficient number of respondents. Although the collapsing procedure used in this study was simple and consistent for all states and all quarters, this might fail to be true in general, particularly for states that are not large hog producers. Weighting classes in states producing few hogs would likely require extensive collapsing at the interview stage to obtain enough hog operation respondents. It might be necessary to collapse entire states together in certain cases.

Other Issues -- Nonsampling Errors

The QAS is a large and fairly complex survey. Regardless of the estimator used, the indications obtained by it are not only subject to sampling errors but are subject to nonsampling errors as well. The quality of the partial information about nonrespondents is of primary concern when trying to adjust for nonresponse. Any estimator that relies on this partial information in correcting for nonresponse bias is adversely effected by inaccurate partial information. The degree of harm is related to the extent of error in the partial data. Nonrespondents recorded as being agricultural operations when they are in fact out of business will cause an upward bias in the estimate (Kott, 1990). Nonresponding operations recorded as having hogs when they actually have none will also cause an upward bias in the estimate. Reinterview studies have provided evidence that these types of errors do occur in the QAS (Fetter in an unpublished memorandum to Bill Iwig, National Agricultural Statistics Service, 1991).

Quality of hog control data is another area of concern. Any estimator that relies on control data will be susceptible to inefficiencies caused by control data that is poorly correlated with responses. Continued effort to keep control data current and accurate is an essential part of producing a high quality estimate. Control data/response correlations at the state level for the states used in this study for the June 1990 QAS ranged from a low of .796 for Iowa to a high of .989 for North Carolina (Scott, 1990). These correlation values are typical.

RECOMMENDATIONS AND FURTHER RESEARCH

The states investigated in this report were all large hog producing states. Additional research is planned to determine what problems might be encountered when applying this estimation process to the nation as a whole. In addition, research should be conducted to evaluate the actual effectiveness of the new weighting classes and ways in which they can be improved. Research is presently being

28

conducted to determine if weighting classes with less than 20 respondents are feasible in some situations.

Currently, research on methods for dealing with outliers on the list and area frames is being planned. The Revised Estimator produces weights that vary within the design strata. It thus lends itself to alternative outlier treatments such as weight truncation and smoothing.

It is recommended that research should also be conducted to develop a more appropriate expression of the variance of the Revised Estimator that addresses the effect of the multi-phase nonresponse weighting and poststratification adjustment cells.

When changes are made in the survey design, the possible effect of these changes on the estimation procedure needs to be considered and appropriate adjustments made if necessary. As a result of capturing partial information concerning ag-status, it becomes necessary to modify some of the current assumptions being made about nonrespondents. The Revised Estimator is based on a set of assumptions that are reasonable and based on all the information that is currently available. Barring unforseen implementation problems, the use of the Revised Estimator for QAS list frame hog total estimation is recommended as an improvement over current estimators. The Revised Estimator is a statistically sound and mathematically simple estimator that properly accounts for nonresponse bias in the estimation of list frame hog totals.

REFERENCES

[1] Agricultural Statistics Board (1991), "Hogs and Pigs" (September 1991 Report), National Agricultural Statistics Service.

[2] Bosecker, R. (1989), "Integrated Agricultural Surveys," National Agricultural Statistics Service.

[3] Cox, B. G. (Publication Pending), "Weighting the Hog List Frame Portion on the QAS," National Agricultural Statistics Service.

[4] Cox, B. G. (Publication Pending), "Weighting Survey Data for Analysis," National Agricultural Statistics Service.

[5] Crank, K. N. (1979), "The Use of Partial Information to Adjust for Nonresponse," Agriculture, Economics, Statistics, and Cooperatives Service.

[6] Fuchs, D. R., and Bass, R. T. (1990), "1989 Agricultural Surveys Survey Administration Analysis," National Agricultural Statistics Service.

[7] Kott, P. S. (1990), "Nonresponse Adjustments in NASS Agricultural Surveys," National Agricultural Statistics Service.

[8] Garribay, R. and Huffman, J. E. Jr. (1991), "1990 Agricultural Survey Administration Analysis," National Agricultural Statistics Service.

[9] Scott, C., Patterson, R., and Geuder, J. (1992), "1991 List Frame Evaluation," National Agricultural Statistics Service.

[10] Shah, B. (1991), "Statistical Methods and Mathematical Algorithms; Technical Appendix to the Documentation for the Data Analysis Software Package, SUDAAN," Research Triangle Institute.

30

APPENDIX

The Relationship between the Nonresponse Models Implied by the Adjusted and Revised Estimators

There are two basic differences between the Revised Estimator and the Adjusted Estimator -- the weighting classes in which nonresponse adjustments are made, and the way in which nonresponse is modeled. There is a relationship between these two models. To see what this relationship is, it is necessary to isolate the nonresponse models from the confounding effect of the different weighting classes used by the estimators.

Though not optimal, it is conceivable to employ the Revised Estimator's model for nonresponse within design strata. If the Revised Estimator is implemented using the design strata as weighting classes with no weighting class collapsing, the stratum estimates produced by the Revised Estimator can be expressed in terms of the stratum estimates produced by the Adjusted Estimator. This relationship is expressed below (at the stratum level).

$$\hat{T}(h)_R = \hat{T}(h)_A * \frac{n_{HA-R}(h) + n_{N-HA}(h) + n_{HA-NR}(h) + n_{NON-AG}(h)}{n_{HA-R}(h) + n_{N-HA}(h) + n_{HA-NR}(h) + n_{NON-AG}(h) + n_{A-DK}(h)}$$

$$* \frac{n_{HA-R}(h) + n_{N-HA}(h) + n_{HA-NR}(h) + n_{A-DK}(h)}{n_{HA-R}(h) + n_{N-HA}(h) + n_{HA-NR}(h)} .$$

where:

$\hat{T}(h)_R$ - The Revised Estimator's estimate of the total hogs in stratum h,

$\hat{T}(h)_A$ - the Adjusted Estimator's estimate of the total hogs in stratum h,

$n_{HA-R}(h)$ = the number of responding hog operations in stratum h sample,

$n_{N-HA}(h)$ = the number of identified non-hog ag operations in stratum h sample,

$n_{NON-AG}(h)$ = the number of identified non-ag units in stratum h sample,

$n_{HA-NR}(h)$ = the number of nonresponding units in stratum h sample that are identified as being hog operations and finally,

$n_{A-DK}(h)$ = the number of units in stratum h sample that have been identified as being ag operations but with undetermined hog status.

A close inspection of the expression reveals that the two estimates are equal under either one or both of two conditions. Note also that if neither condition holds, $\hat{T}(h)_R > \hat{T}(h)_A$. This indicates that the Adjusted Estimator is biased downward under the nonresponse model implied by the Revised Estimator.

Note: The reader is reminded that sampled units that have been identified as being non-ag units are both ag-status respondents and hog-status respondents.

Condition 1:

If $n_{NON-AG}(h)$ = 0, there are no sampled units with known ag-status that are non-ag units in stratum h. The Revised Estimator assumes that the ag-status respondents represent a random sample of the original stratum sample. The Revised Estimator then assumes that the hog-status respondents that are also ag-operations (note that this **excludes** non-ag units) represent a random sample of the ag-status respondents that are ag-operations. Generally this is **not** equivalent to the Adjusted Estimator's Assumption 1 which states that the hog-status respondents (including identified non-ag units) represent a random sample of the original sample. However-- when Condition 1 holds, all ag-status respondents are ag-operations and thus have equal probability of providing hog-status under the assumptions made by the Revised Estimator. This immediately implies that the hog-status respondents represent a random sample of the original sample. Thus, the Revised Estimator's assumption concerning hog-status respondents becomes equivalent to the Adjusted Estimator's Assumption 1 when Condition 1 holds and the two estimators will produce identical stratum estimates.

Condition 2:

If $n_{A-DK}(h)$ = 0, there are no sampled units identified to be ag-operations that have undetermined hog-status in stratum h. When Condition 2 holds, all ag-status respondents are certain to be hog-status respondents (this includes identified non-ag units). Thus, when Condition 2 holds, ag-status respondents (which are assumed to be a random sample from the original sample by the Revised Estimator) are identically the hog-status respondents. This immediately implies that under Condition 2, the hog-status respondents are a random sample of the original sample. This is exactly equivalent to the Adjusted Estimator's Assumption 1 and the two estimators will produce identical stratum estimates.

32