



**United States
Department of
Agriculture**

**National
Agricultural
Statistics
Service**

**Estimates
Division**

**SMB Staff Report
Number SMB-91-04**

May 1991

A REVIEW OF THE NONOVERLAP COMPONENT OF THE JUNE 1990 MULTIPLE FRAME INDICATIONS

**J. DONALD ALLEN
JERRY THORSON**

A REVIEW OF THE NONOVERLAP COMPONENT OF THE JUNE 1990 MULTIPLE FRAME INDICATIONS by J. Donald Allen and Jerry Thorson, Estimates Division, National Agricultural Statistics Service, Washington, D.C. 20250, May 1991. NASS Staff No. SMB-4-91.

ABSTRACT

The National Agricultural Statistics Service uses a multiple frame design for its Quarterly Agricultural Surveys. The June quarter serves as a base with three "follow on" surveys conducted in March, September, and December. Data collected during these subsequent surveys often uncover problems in the June data file. These errors typically involve the duplication of data. This analysis of the June 1990 - March 1991 survey data shows that there was an overexpansion of the June data because of these inaccuracies.

Keywords: Multiple frame, nonoverlap, post survey, weight.

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

TABLE OF CONTENTS

INTRODUCTION 1

THE POST SURVEY PERIOD 1

FINDINGS 3

CONCLUSION 10

REFERENCES 11

INTRODUCTION

Multiple frame sampling implies the use of two or more sampling frames. Typically the approach is used when 1) the use of a single frame does not provide complete population coverage or 2) when a single frame is unable to adequately address rare or highly variable commodities without unduly large samples. The area frame used by the National Agricultural Statistics Service (NASS) is a complete frame but alone is unable to provide reliable indications for livestock totals nor for the acreages of minor crops, particularly at the state level. Thus, the decision was made in the 1970s to construct a list frame consisting of all known farming operations, with it being the primary source of data. To further sampling efficiency, control data are maintained on the list records so as to allow for stratification. The list is inherently incomplete and in a state of flux. Thus, to obtain complete population coverage, it is necessary to use the area frame to account for those farming operations not on the list frame.

One of the requirements of a multiple frame design is that there be a mechanism for removing duplication between frames as well as duplication within frames. NASS has such a mechanism in place with procedures for handling both situations outlined in the 1991 Agricultural Surveys Supervising and Editing Manual. Flowcharts are provided as a guide for statisticians to use in their decision making. These diagrams are effective tools for resolving duplication when all the pertinent information is correct (e.g., name, address, operating arrangement, etc.). Unfortunately, the two-week data collection period allowed for NASS's Quarterly Agricultural Surveys does not always provide adequate time for ensuring that all the needed information is accurate. In this latter case, the statistician might follow procedures in a completely proper manner and still have duplication in the end just because some minor name variation caused it to go undetected. Ultimately, duplication decisions should be a matter of understanding the concept: i.e., every farming arrangement should be covered by one and only one frame.

THE POST SURVEY PERIOD

For the Quarterly Agricultural Surveys, the June survey serves as the base. That is, the list frame is frozen and each subsequent quarter's sample is drawn prior to this survey. Ideally, all the duplication between frames is resolved during the June data collection and data editing period. Such is seldom the case. In subsequent quarters, some name and address errors are nearly always detected as are errors in operation descriptions. Most of the corrections result in a change in the duplication determination, and in essence, this means duplication was allowed into the prior quarter's summary. In an effort to minimize these changes,

NASS designed a computer program to aid statisticians in detecting "overlap" errors. After the June survey has been completed, each state office runs the Automated Overlap/Nonoverlap system which allows them to review all of their duplication coding. Any needed updates can then be posted to the June data file. Of course, the system is not foolproof since coding is based on the available names and addresses.

During this post survey period, states can also update other data fields based on late reports, refusal conversion, and the like. One of principal data fields that should be reviewed is the entire farm acreage for the nonoverlap (NOL) area tracts. This number is a key component of the weighted estimator. Basically, the weighted estimator uses entire farm data factored back to the segment/tract level; this adjustment allows the area expansion factor to be applied. The proration factor or weight is simply the ratio of tract acreage to total acreage. NASS uses the weighted indication as its primary estimator for the NOL component of the multiple frame expansion, with the one exception being crop acreages in June which rely on the NOL tract indication (Nealon, 1984).

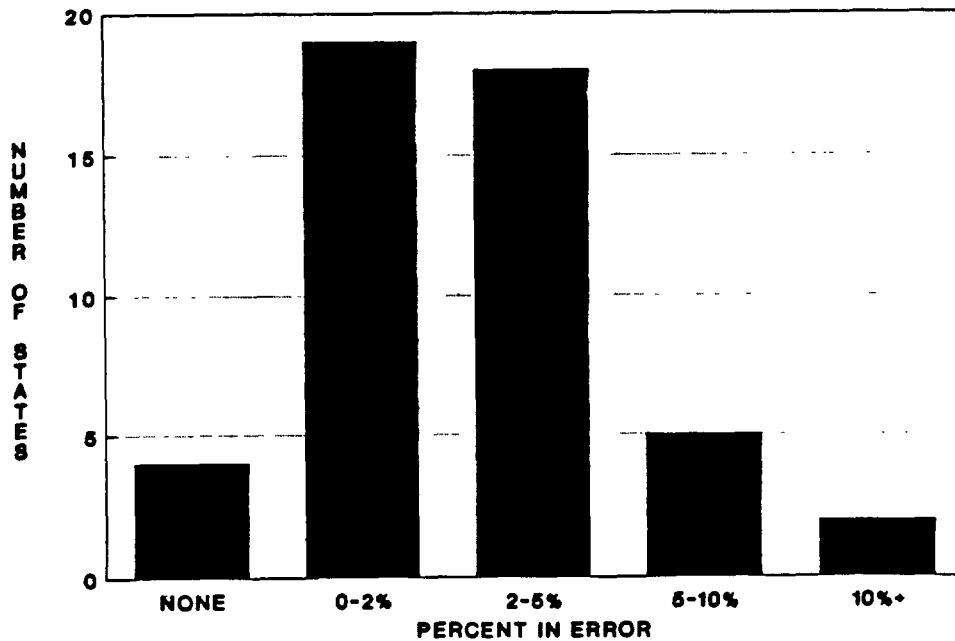
Errors in total land can have negative effects in subsequent quarters since the weights established in June are assumed to be true and, in a way, are carried forward as part of the expansion factor in the "follow on" surveys. There is no machine imputation used for the area frame operations (except for stocks) in June. Thus, by necessity, tract acreage and total land acreage must be estimated for all the refusals/inaccessibles found in the frame. In essence, during the post survey period, states should make an effort to "true up" any weights based on estimated data. One of the best tools for determining records that possibly have erroneous weights is data listings. Using this approach, the statistician generates computer printouts of all the NOL records with a weight approximately equal to one as well as listings of the acreages for all refusals/inaccessibles. These are typically the records that one might wish to review further for accuracy. In some cases, the statistician may wish to examine any reports for which the respondent was not the actual farm operator.

The post survey updates serve two functions. First, states are allowed to review a resummarization of their corrected June data; the reassessment can be used to make revisions in the crop and livestock estimates. Second, the updates can make area frame sampling for the "follow on" surveys more effective. Only the nonoverlap records are used in the subsequent surveys with the samples coming from a stratified population. If the OL/NOL status on any of these records is incorrect, the expansions for the "follow on" surveys will tend to be too high. Overlap records do not get sampled so errors made in that direction are not detectable.

FINDINGS

Unfortunately, the work done during the post survey period does not guarantee a completely accurate base for subsequent surveys. June mistakes continually surface during the ensuing contacts. This is shown by an analysis of the Quarterly Agricultural Survey data for June 1990 - March 1991. After the June 1990 post survey period, there were 21,519 total nonoverlap tracts available for sampling. Typically, there are five replications in each state's area sample with the oldest replication being replaced each year through a rotation scheme. Prior to the sampling for the "follow on" quarterly surveys, the two oldest replications in the June sample (i.e., forty percent of the segments) were set aside to be used for NASS's economic surveys. From the remaining sixty percent, a total of 10,407 distinct tracts were sampled for the September, December, and March quarters. Of these, three percent, or 331, were later found to have had an error in their June name (see 1991 Agricultural Surveys Supervising and Editing Manual). Figure 1 shows the percent of tracts coded in error in June with the number of states falling in each category. Eighty eight percent, or 291, of those changed to overlap; essentially, this means that the June expansions were inflated since these records were accounted for by both frames.

Figure 1: Percent of Tracts Coded as Being in Error (June 1990 - March 1991)



Additionally, during the subsequent surveys, states requested that total land (i.e., weights) be changed on 29 records. Of these, 26 occurred in December due to an overexpansion of grain stocks. It should be pointed out that a request of this nature is normally only made when a state is faced with a expansion that is totally unfeasible. The individual records that are contributing to any abnormalities are typically identified through the use of NASS's crop analysis package or through the newly developed "high/low" frequency prints. Again, overexpansion of the June data has occurred, but this time because a weight that was too high was used during the base survey.

To illustrate these points, the corrected overlap statuses and weights were posted to the June data file and the expansions recalculated. In Figure 2, the total change in the nonoverlap expansion for cropland is shown with the change in the hog nonoverlap expansion shown in Figure 3. Note that no adjustment was made for the 11,112 tracts not sampled for the "follow ons." Those not sampled include not only all the tracts from the forty percent of the segments that were set aside but also all those tracts not selected from the sixty percent. Figure 4 shows the downward percent change in the NOL expansion for cropland and storage capacity first due to weight changes and then due to OL/NOL changes.

Figure 2: June 1990 Cropland NOL Expansion

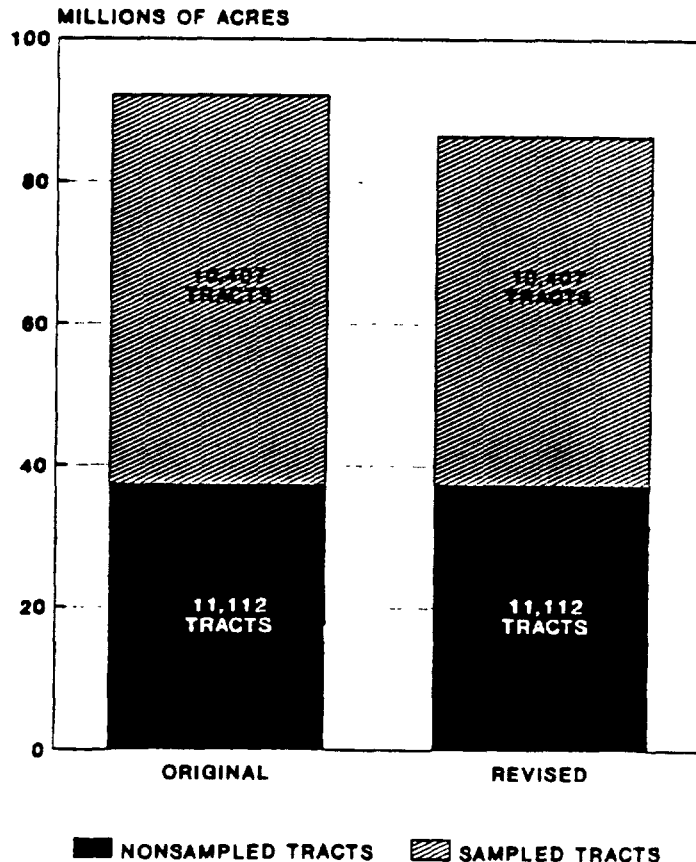


Figure 3: June 1990 Hog NOL Expansions

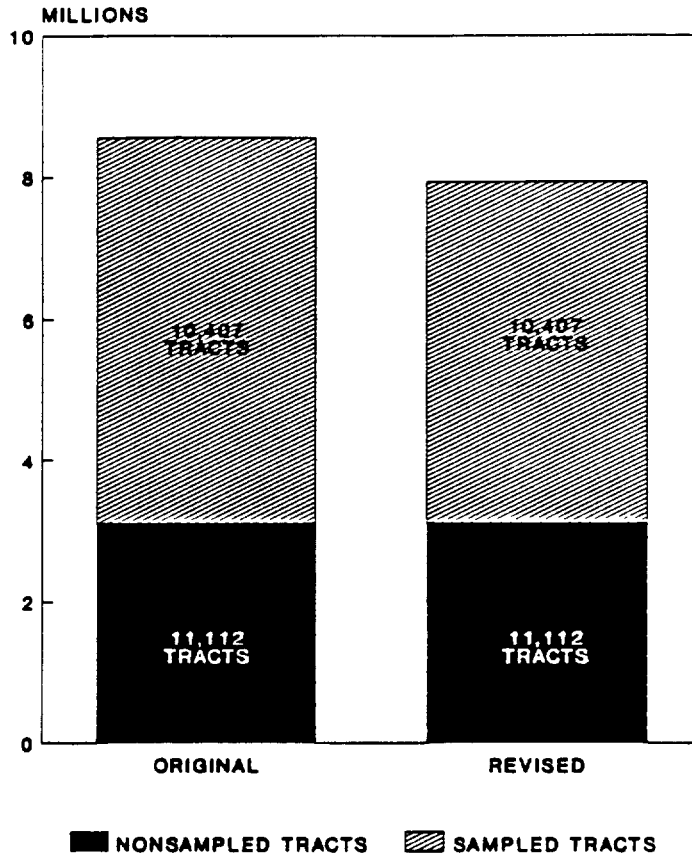
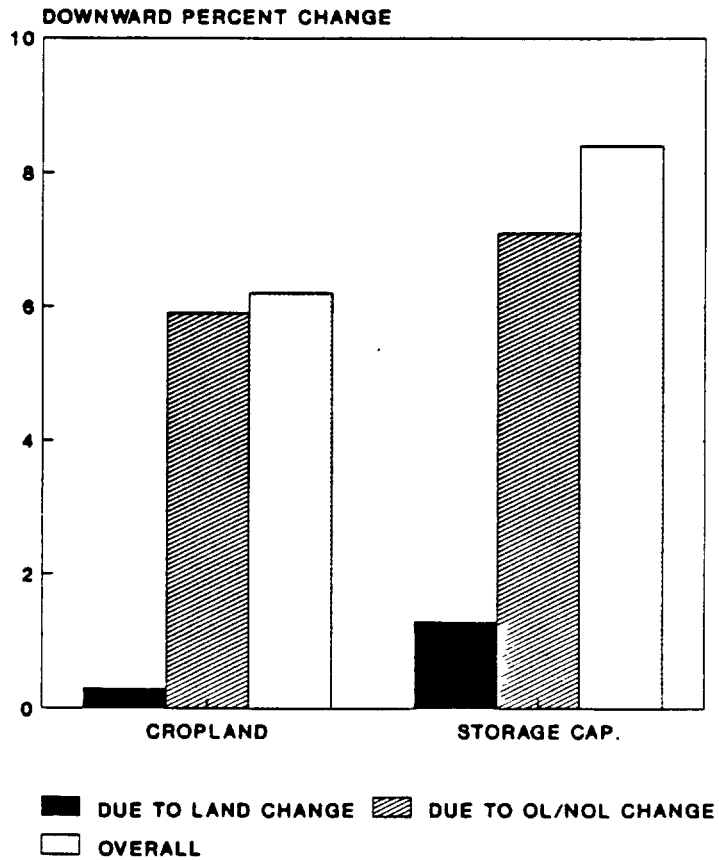


Figure 4: June 1990 NOL Expansions



Percentages for the NOL area expansions for other commodities are shown in Table 1. These figures can be viewed as minimums since uncorrected errors still exist in the 11,112 tracts not sampled.

Table 1: Differences in NOL Expansions Due to June Errors

<u>Commodity</u>	<u>Downward Percent Change</u>
Winter Wheat Acreage	2.4
Corn Acreage	3.6
Barley Acreage	3.2
Soybean Acreage	3.6
Storage Capacity	8.4
Cropland Acreage	6.2
Barley Stocks	0.4
Corn Stocks	11.4
Soybean Stocks	12.9
Total Hogs	7.4

Naturally, one might think that since the list frame is the primary source of data that errors in the area expansions would have little impact on the overall multiple frame expansion. In reality though, the area expansion is a significant piece of the indication with its contribution ranging from approximately 15 to 20 percent for major commodities (see Table 2).

Table 2: NOL Expansion as a Percent of the June 1990 Multiple Frame Expansion

<u>Commodity</u>	<u>Downward Percent Change</u>
Winter Wheat Acreage	18.6
Corn Acreage	16.1
Barley Acreage	14.7
Soybean Acreage	16.6
Storage Capacity	16.9
Cropland Acreage	21.3
Barley Stocks	16.6
Corn Stocks	13.9
Soybean Stocks	15.3
Total Hogs	15.9

Multiple frame expansions based on the original June data and the corrected file are shown for soybeans planted acreage and for hog inventories in Figures 5 and 6. These two charts also show the original NASS estimates (June board) and the current estimates (revisions). The soybean indication would have been about 400,000 acres less in June if the OL/NOL determination had been handled correctly on the 10,407 sampled tracts. If all the nonoverlap tracts had been revisited during the "follow ons," a much greater error may have been detected. The original acreage figure published by NASS was approximately 58 million acres; this figure was subsequently revised down by 250,000 acres. For hogs, the original June indication was found to be 600,000 head too high; the original NASS estimate was 54.4 million head with subsequent revisions lowering the estimate 500,000 head to 53.9 million. The question is "would these revisions have been necessary if the correct indications had been available in June?"

Figure 5: June 1990 Soybeans Pltd. Acres

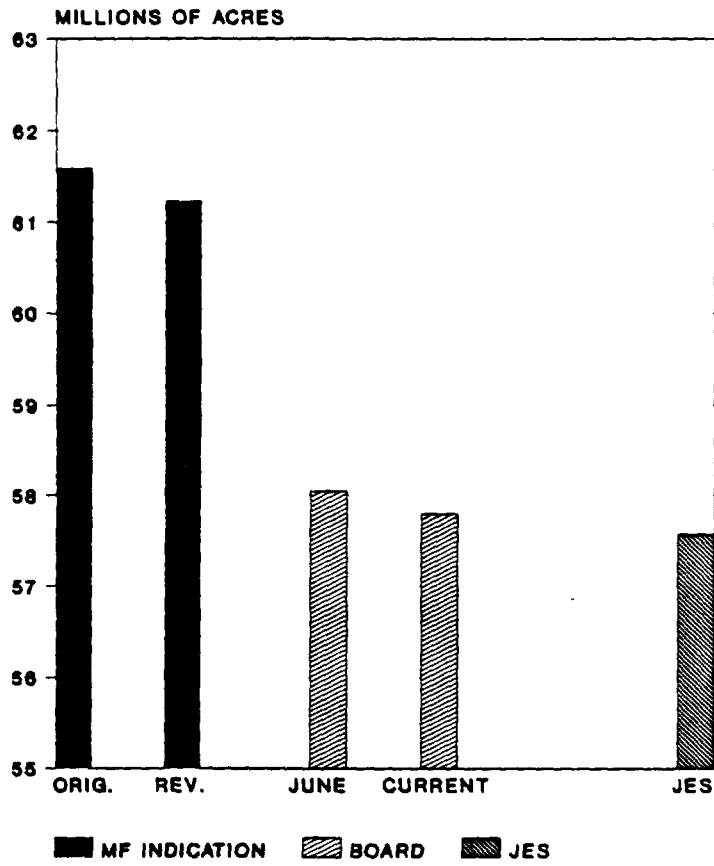
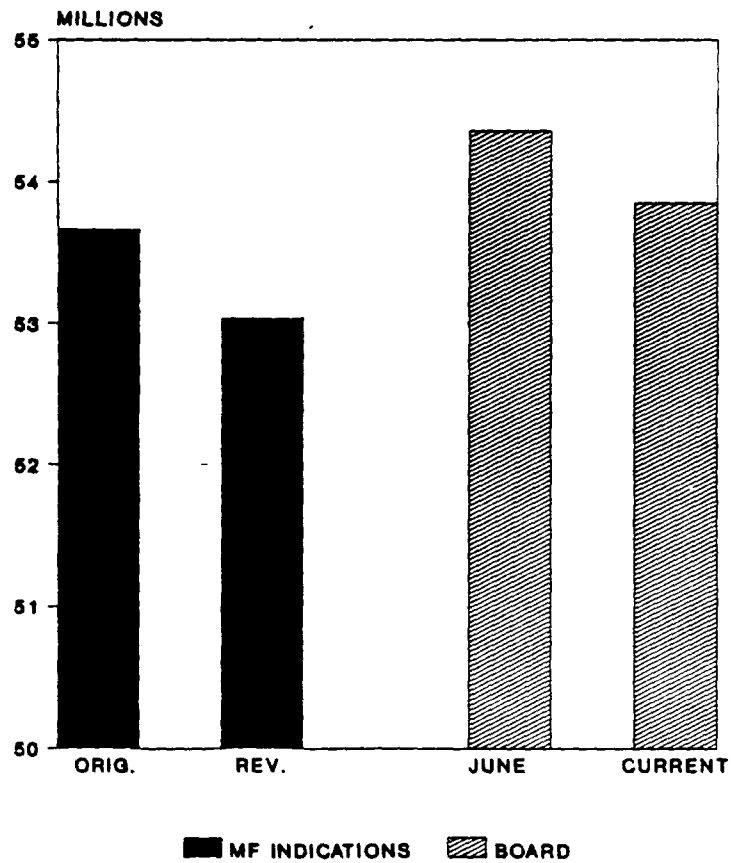
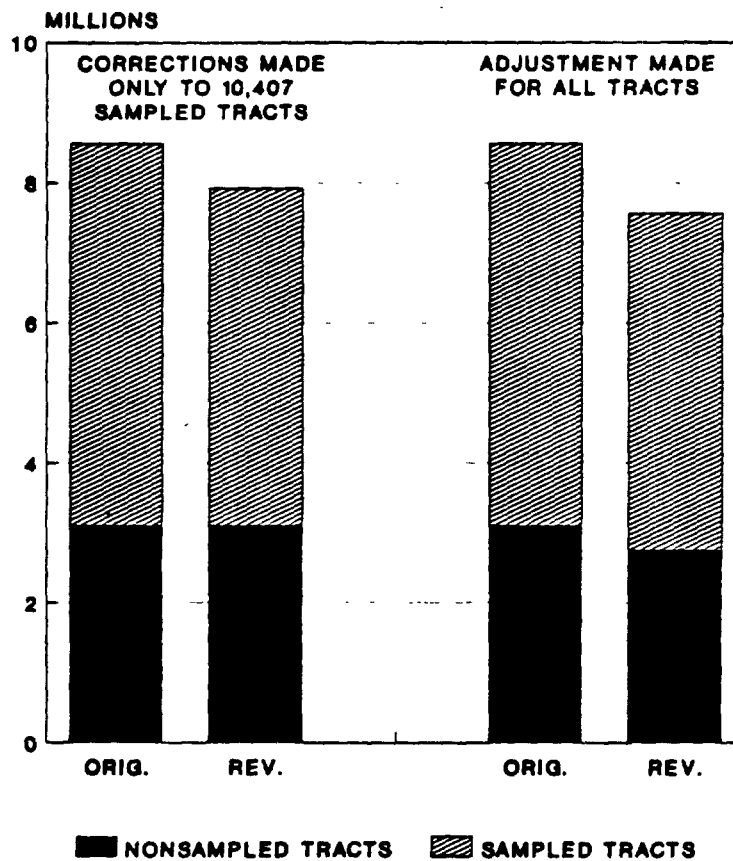


Figure 6: June 1990 Hog Inventory



To this point, only the sampled NOL tracts have been considered. What about the 11,112 nonsampled tracts? Nothing can really be said about them since they were not recontacted. However, if the assumption is made that what was found in the sampled tracts would also be found in the nonsampled tracts, then the analysis can be taken one step further. In Figure 7, the first two bars reflect the levels of the NOL expansion based on changes in only the sampled tracts while the second set of bars reflect the extrapolation of that information to the 11,112 nonsampled tracts. This was done by applying the percent change in the NOL for the sampled tracts to the total NOL expansion.

Figure 7: Hog NOL Expansions



PERCENT CHANGE IN NOL FOR THE SAMPLED TRACTS HAS BEEN APPLIED TO NONSAMPLED TRACTS IN THE SECOND SET OF BARS

The results for hogs and soybeans are shown in Table 3. In both cases, the overall change in the multiple frame indication based on corrected data exceeds the original survey indication by over one standard error.

Table 3: Relative Changes in the June 1990 Multiple Frame Expansion with Findings in Sampled Tracts Extrapolated to Nonsampled Tracts

<u>Commodity</u>	<u>June Standard Error</u> (thousands)	<u>Change in MF Expansion</u>	<u>Change as Factor of Std. Error</u>
Hogs	852	993	1.17
Soybeans	510	577	1.13

There are of course limitations to this approach. Namely, the nonsampled tracts are known to be somewhat different than those that were sampled. For instance, the forty percent of the segments not sampled are older segments which have been visited over several years which means that the names and addresses associated with those tracts are probably less prone to error. Additionally, twenty percent of the segments available for sampling are new and therefore are probably more prone to having name and address errors. The tracts contained in the available segments are also subsampled with the criteria varying by quarter; this too points to possible differences between selected and nonselected records. All tracts reporting hogs or intentions to have hogs as well as all unknowns are sampled in the September quarter with a subsampling occurring in the subsequent quarters. Soybean acreage is not sampled as such, but is most likely reflected by the cropland and stocks strata that are used in the "follow on" surveys. From this, it is not readily apparent whether the extrapolation is overstating or understating the problem in the NOL contribution. Nevertheless, the point is that there was a problem in the June 1991 NOL indication due to erroneous weights and the miscoding of June tracts. To reiterate, the indication for hogs was found to be about 600,000 too high *without the extrapolation*, with soybean acreage found to be 400,000 acres too high.

CONCLUSION

This analysis illustrates the impact that incorrect OL/NOL determination (and erroneous weights) can have on survey indications. This, in turn, exemplifies the need for a quality list frame as well as adequate procedures for OL resolution. This further translates into the need for enumerators and state office staff to understand multiple frame concepts as well as why correct data are so important. Problems in domain determination were addressed early on by Beller (1979) and, as illustrated here, still persist. Although this report only examined the nonoverlap portion of the Quarterly Agricultural Surveys, it is felt that the same problems exist for the economic surveys. NASS must find ways to reduce the nonsampling errors in its multiple frame approach or devise estimators that will lessen their impact. The reports issued by the Survey Quality Team may provide some guidance in this respect.

REFERENCES

Beller, N.D.(1979). Error profile for multiple-frame surveys. Washington, D.C.: U.S. Dept. of Agr., ESCS. (ESCS-63)

Nealon, J.P.(1984). Review of the multiple frame and area frame estimators. Washington, D.C.: U.S. Dept. of Agr., Stat. Rep. Serv. (SF&SRB Staff Report No. 80)

Pafford, B. et al.(1990). The concept of accuracy and how to approach reducing the total survey error of NASS survey statistics, with an example from the agricultural survey program. Washington, D.C.: U.S. Dept. of Agr., NASS. (NASS Survey Quality Team Report No. 90-1)

Survey Quality Team for the Agricultural Survey Program.(1990). Agricultural survey program: baseline quality report. Washington, D.C.: U.S. Dept. of Agr., NASS.

Vogel, F.A.(1986). Survey design and estimation for agricultural sample surveys. Washington, D.C.: U.S. Dept. of Agr., Stat. Rep. Serv.