

Non-response Bias Analysis

April, 2020

BACKGROUND

USDA NASS conducts the Crops Acreage, Production, and Stocks (APS) survey quarterly in March, June, September, and December. The survey of interest for this study is the December 2017 survey. The response rate for this survey was below 80 percent and the coverage was below 70 percent. A nonresponse bias analysis was completed to compare our survey expansion indications for select crop harvested acres to expansion indications from a more complete sample. Crop yields were also evaluated against other sources. The response rate from the more complete sample was 82.5% compared to the original response rate of 58.4%. This comparison was made to identify whether the survey indications were biased.

The Crops APS survey uses a Multivariate Probability Proportion to Size (MPPS) sample design for the items of interest in the survey. Imputation is used to make all records complete and account for nonresponse in the expansion indications. The imputation program imputes for missing data using data from farms of similar size, type, and location with complete survey data.

PROXY DATA

To create a more complete dataset, proxy data were used for nonresponses when possible. First, data was processed and edited for any late reports received by mail after the data was originally summarized for publication. Additionally, data from the 2017 Census of Agriculture (COA) was used for any additional nonresponse list frame records that matched to a complete COA report. Any remaining nonrespondents had their data imputed by the summary using the normal survey procedures discussed above. Note that the more complete dataset is not a perfect estimation of our population parameter, but it is the best obtainable comparison.

Since the COA and the December Crops APS questions differed in structure, we adjusted some of the Crops APS questions to reflect the COA structure. For example, the Crops APS survey asks for corn harvested for grain and corn harvested for seed separately whereas the COA asks for the two combined. Additionally, several of the crops analyzed ask irrigated and non-irrigated acres separately but only total acres are collected on the COA. To make the data comparable to COA, corn harvested for grain and corn harvested for seed were combined as were the irrigated and non-irrigated acres for the various crops.

The crops of interest that were summarized were corn harvested for grain or seed, corn harvested for silage, soybeans, upland cotton, sorghum harvested for grain or seed, and sorghum harvested for silage. Each crop has a set of states that are published annually. Corn had 48 states, soybeans had 31 states, upland cotton had 17 states, and sorghum had 14 states. There were a total of 172 list frame indications compared in this study.

RESULTS

After rerunning summaries with the more complete datasets, there was some bias observed in the list frame indications. In general, there were more indications that underestimated than overestimated but some commodities performed better than others. Overall, 62 indications were overestimated while 109 were underestimated, a 36% over and 63% under comparison (one indication remained unchanged). If the indications were unbiased, we would expect roughly the same amount of states overestimated as were underestimated. The proportion of indications underestimated is statistically different from 50 percent.

The breakouts for the individual crops are reflected in Table 1 below:

Table 1 - HARVESTED ACRES COMPARISON

Crop	Total States	Overestimated	Pct Over	Underestimated	Pct Under
Corn for Grain or Seed	48	14	29%	34	71%
Corn for Silage	48	16	33%	32	67%
Soybeans	31	14	45%	17	55%
Upland Cotton	17	6	65%	11	35%
Sorghum for Grain or Seed	14	7	50%	7	50%
Sorghum for Silage ¹	14	5	57%	8	36%

¹One indication remained unchanged.

The majority of the downward bias was in the two corn harvested acres indications. The U.S. re-summarized indications for corn harvested for grain or seed and corn harvested for silage were both significantly different than the original indications. The P values for both indications were well under 0.05. There were also several major corn states with P values less than 0.05.

ADDITIONAL YIELD ANALYSIS

The December 2017 Crops APS yield indications from the operational summary for Corn for Grain and Seed, Corn for Silage, Soybeans, Sorghum for Grain or Seed, and Sorghum for Silage were compared to yields from the 2017 COA. Yield indications are generated using a reweighted estimator (the weights of the usable reports are adjusted to account for non-respondents) and not the imputed estimator like the acreage expansions. The same comparison was made for production for Upland Cotton as this reflects how estimates are set. Production indications are generated using the imputed estimator. Overall, 105 out of the 172 yield indications that were compared were above the number from the COA and 67 were below. Or, by percent, 61% overestimated while 39% underestimated.

The breakouts for the individual crops are reflected in Table 2 below:

Table 2 - YIELD COMPARISON

Crop	Total States	Overestimated	Pct Over	Underestimated	Pct Under
Corn for Grain or Seed	48	35	73%	13	27%
Corn for Silage	48	32	67%	16	33%
Soybeans	31	23	74%	8	26%
Upland Cotton ¹	17	1	6%	16	94%
Sorghum for Grain or Seed	14	6	43%	8	57%
Sorghum for Silage	14	8	57%	6	43%

¹Comparison made for production, not yield.

The majority of the upward bias was in the two corn and the soybean yield indications. There was an obvious downward bias in the upland cotton production indications. However, cotton ginnings data is collected separately from a census of cotton gins which is available when upland cotton production estimates are set. Also for all commodities except sorghum for grain or seed, there were several states with P values less than 0.05.

Additional yield comparisons were made for two crops. The December 2018 Crops APS operational summary yield indications for corn for grain or seed and soybeans were compared to yields calculated from the 2018 Market Facilitation Program (MFP) data for states that had data available. This program was created to provide assistance to farmers with commodities negatively impacted by the recent foreign tariffs. Nearly all farmers reported their production data to USDA Farm Service Agency (FSA) to sign up for the program and receive payments. Of the 76 indications compared, 29 were overestimated and 47 were underestimated, a 38% over and 62% under comparison.

The breakouts for the individual crops are reflected in Table 3 below:

Table 3 - MFP YIELD

Crop	Total States	Overestimated	Pct Over	Underestimated	Pct Under
Corn for Grain or Seed	45	18	40%	27	60%
Soybeans	31	11	35%	20	65%

Based on the MFP yields, there was a downward bias in the corn and soybean indications from the December Crops APS survey. However, there is possible incentive for farmers to round up when reporting their production to increase their payment amount. This bias is in the opposite direction of what we saw in the comparisons to the COA data for the previous year. There were also a large number of states with P values less than 0.05 for both commodities.

DISCUSSION

While this study did show some bias in the acreage indications, FSA administrative acreage data is used to complement the survey data and set planted acreage estimates. FSA data coverage at the time estimates are set is generally more than 95%. The June Area frame survey provides additional planted acreage indications based on a complete sampling frame of land area and very little nonresponse since planted acreage can be observed. NASS is also currently exploring expanded use of remote sensing and machine learning techniques. A harvested to planted acres ratio from the December Crops APS survey is primarily used to set harvested acreage estimates. While a direct analysis of this ratio was not possible, ratios generally have less bias than totals.

A comparison of yields to two different sources indicated bias in the December Crops APS survey but in opposite directions. There was also a noticeable bias in the production indications for upland cotton. However, as previously mentioned, we have additional data from our cotton ginnings survey. This survey collects ginning data biweekly from all active cotton gins as identified by the USDA Agricultural Marketing Service (AMS).