Milk Production Nonresponse Bias Analysis

BACKGROUND

USDA NASS conducts the quarterly Milk Production surveys in January, April, July, and October for all states; the January survey serves as the base month for the survey year. The survey of interest for this nonresponse bias study is the January 2017 Milk Production since it was the most recent base survey available.

The response rate for this survey was below 80 percent, and the coverage rate was below 70 percent. As a result, a nonresponse bias analysis was conducted to compare our production survey milk cow survey estimates to "complete" survey estimates in order to identify whether the survey estimates were biased. The nonresponse adjustment for this survey uses a version of a reweighted estimator which adjusts records based on the presence of the commodity. The nonresponse adjustment is calculated within each stratum and applied to the sampling weight. The presence of the commodity for nonrespondents with greater than 500 milk cows on the sampling frame, less than two year old frame data, or reporting milk cows on another survey in the last year are assumed to be known positive. All other nonrespondents are assumed to have unknown presence.

The Milk Production Survey uses a stratified sample design created by using the number of milk cows from NASS's list sampling frame for each given record. The design across all 50 states is standardized into as many as nine sampling strata according to size.

PROXY DATA

To create a "complete" January 2017 Milk Production survey dataset, proxy data were used for all nonresponse records. The proxy data were implemented using a hierarchical structure, starting with the most recent quarterly Milk Production Survey data. First, proxy data were used from the October 2016 Milk Production survey. If there were no October reported data then July was used, if no report in July then April is used. If there were no previously reported data (PRD) available at this point, then sampling frame data were used. All frame data will be positive for the number of milk cows. As a result, every record had usable data for number of milk cows. The most common source of the sampling frame data was the 2012 Census of Agriculture, if they were not in the Milk Production or Cattle Survey since that time. The goal was to create a "complete" January 2017 Milk Production dataset without nonresponse records.

RESULTS

The January 2017 Milk Production dataset contained 11,891 records, with 4,081 nonresponse records. The PRD datasets added back 961 complete records, and the sampling frame data added back 3,120 records.

Survey estimates were recalculated with the "complete" datasets, and some differences were observed. The number of milk cows is higher in the "complete" dataset overall by about 0.3% at the US level. Thirty-two states had higher estimates while 18 states had lower estimates. If the survey estimates were unbiased, we would expect roughly the same amount of states above and below the "complete" estimates. The proportion overestimated is 65% which is statistically different from 50% with a p-value of 0.016. The CV for the operational estimates is 0.7%, so the difference at the US level is less than half of the standard error.

About 75% of the data used to create the "complete" dataset came from the sampling frame and is a positive number of milk cows. Using the sampling frame data provides somewhat of an upper bound on the potential bias, since it assumes all nonrespondents have positive milk cows.

DISCUSSION

The results indicate there may be a small downward bias in the adjusted estimate from the Milk Production Survey. The amount of bias is less certain since the study really only provides a maximum on the amount of bias. NASS is discussing an estimator that would better adjust for nonresponse, but has not begun any research on this. We will soon have data from the 2017 Census of Agriculture to further analyze nonresponse bias in this survey.

USDA NASS is currently placing special emphasis on improving response rates across all surveys. A Response Rate Research team has been active for over a year, and the team has about 14 sub teams working on different options to improve the response rates. In fact, a few of our surveys have shown an increase in response rates. For example, the December 2017 Crops Acreage, Production and Stocks Survey is up almost one percent over the two previous years.

Additionally, our phoning centers are expanding data collection periods, where possible. Our metadata team is collecting better contact information, including updating cell phones numbers and current home addresses. Our Regional Field Offices are contacting our chronically inaccessible operators with different strategies. Our marketing department is giving more promotional information to the operators on why they should respond.