Recent Developments in Model-based Estimation Techniques in Agricultural Statistics

Wendy J. Barboza*
USDA, National Agricultural Statistics Service, Washington, DC, USA –
wendy.barboza@nass.usda.gov

Abstract

In June 2010, the National Agricultural Statistics Service (NASS) began an internal review of its survey and census programs to evaluate how well the agency adhered to the Office of Management and Budget's (OMB's) standards and guidelines related to disseminating estimates. The standard of interest states: "Agencies must use accepted theory and methods when deriving direct survey-based estimates, as well as model-based estimates and projections that use survey data. Error estimates must be calculated and disseminated to support assessment of the appropriateness of the uses of the estimates or projections. Agencies must plan and implement evaluations to assess the quality of the estimates and projections." The review found that many programs published direct survey-based estimates without including error estimates, so NASS changed these programs to adhere to the standard. For some programs, the review found that NASS utilized multiple data sources such as survey estimates and administrative data to produce crop and livestock estimates. To meet OMB's standard for these programs, NASS began to develop model-based estimates and forecasts with accompanying error measurements. This paper will provide an overview of three models, which are being incorporated into the operational environment. The model-based estimation approach utilized for these three projects depended on the data sources available as well as the nature of the data. This paper will discuss a hierarchical Bayesian model for crop yield, a time series model for hogs, and a bivariate area-level model for cash rental rates.

Keywords: NASS; model; multiple data sources; uncertainty.

1. Introduction

The National Agricultural Statistics Service (NASS) is a statistical agency located under the United States Department of Agriculture (USDA). NASS's mission is to provide timely, accurate, and useful statistics in service to U.S. agriculture. In order to successfully accomplish the agency's mission, NASS conducts hundreds of surveys every year and publishes numerous reports covering virtually every aspect of U.S. agriculture. Some examples of areas covered in NASS's reports are production and supplies of food and fiber, prices paid and received by farmers, farm labor and wages, farm income and finances, chemical use, and demographics of U.S. producers. A wide variety of topics are covered within these different areas. The subject matter ranges from traditional crops, such as corn and wheat, to specialty commodities, such as mushrooms and flowers; from agricultural prices to land in farms; from once-a-week publication of cheddar cheese prices to detailed census of agriculture reports every five years.

2. Office of Management and Budget

Within the United States, the Federal statistical system is decentralized. The Office of Management and Budget (OMB) provides oversight of the different Federal statistical agencies. OMB reports directly to the President and helps a wide range of executive departments and agencies across the Federal Government to implement the commitments and priorities of the President. Within the statistics arena, OMB is charged with developing and overseeing the implementation of Government-wide principles, policies, standards, and guidelines concerning the development, presentation, and dissemination of statistical information.

The Federal statistical system comprises over 100 programs that engage in statistical activities. These programs are decentralized among more than 70 agencies or organizational units. However, OMB has issued certain statistical directives that only apply to the "core" Federal statistical agencies and statistical units. As of December 2014, the core Federal statistical agencies/units are the Bureau of Economic Analysis (Department of Commerce); Bureau of Justice Statistics (Department of Justice); Bureau of Labor Statistics (Department of Labor); Bureau of Transportation Statistics (Department of Transportation); Census Bureau (Department of Commerce); Economic Research Service (Department of Agriculture); Energy Information Administration (Department of Energy); National Agricultural Statistics Service (Department of Agriculture); National Center for Education Statistics (Department of Education); National Center for Health Statistics (Department of Health and Human Services); National Center for Science and Engineering Statistics (National Science Foundation); Office of Research, Evaluation, and Statistics (Social Security Administration); Statistics of Income Division (Department of the Treasury); Microeconomic Surveys Unit, Federal Reserve Board; Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration (Department of Health and Human Services); and National Animal Health Monitoring System, Animal and Plant Health Inspection Service (Department of Agriculture). These Federal statistical agencies/units focus on different subject matter areas; NASS concentrates on publishing statistics on U.S. agriculture.

In September 2009, OMB published a document entitled "Standards and Guidelines for Statistical Surveys" [Office of Management and Budget (2009)]. This document provides 20 standards; these standards define the professional principles and practices that the core Federal statistical agencies/units are required to adhere to and the level of quality and effort expected in all statistical activities. Each standard has accompanying guidelines that present recommended best practices to fulfill the goals of the standards. Taken together, the standards and guidelines provide a means to ensure consistency among and within statistical activities conducted by the core statistical agencies/units within the Federal Government.

3. NASS Internal Review

In June 2010, NASS began an internal review of its survey and census programs to evaluate how well the agency adhered to OMB's standards and guidelines related to disseminating estimates. The standard of interest states "Agencies must use accepted theory and methods when deriving direct survey-based estimates, as well as model-based estimates and projections that use survey data. Error estimates must be calculated and disseminated to support assessment of the appropriateness of the uses of the estimates or projections. Agencies must plan and implement evaluations to assess the quality of the estimates and projections." The review found

that many programs published direct survey-based estimates without including error estimates, so NASS changed these programs to adhere to the standard. For some programs, the review found that NASS utilized multiple data sources such as survey estimates and *administrative data* to produce crop and livestock estimates. To meet OMB's standard for these programs, NASS commenced research into various methodologies to develop model-based estimates and forecasts with accompanying error measurements.

The term *administrative data* refers to data and reports obtained from agribusiness firms, regulatory agencies, and other government offices. These data are referred to as administrative data because they are obtained by other agencies and organizations to administer their programs and/or conduct their business. NASS is different from most other statistical agencies within the Federal Government because the agency has access to extensive administrative data. NASS obtains most of these data from other USDA agencies, but some are from other government and non-government entities. NASS uses administrative data to measure the performance of the agency's surveys and, in some cases, to assist in establishing the published estimates.

4. Hierarchical Bayesian Model for Corn and Soybean Yield

During the crop season, NASS publishes forecasts and estimates of crop yield on a monthly basis. The first forecast for corn and soybeans is published in August. The forecasts are updated monthly until November and the final yield estimates are published at the beginning of January. NASS conducts several surveys to obtain data for forecasting and estimating crop yields: the Agricultural Yield Survey (AYS); the Objective Yield Survey (OYS); and the Quarterly Crops Acreage, Production, and Stocks (APS) Survey. The Crop Progress and Conditions Survey is also utilized, but in a different manner.

The U.S. is divided into 50 administrative areas called states. A state is further divided into administrative areas called counties; the number of counties within a state varies from three to 254. Some surveys are only administered in certain states for various reasons. The AYS is a farmer-interview survey conducted monthly from May through November. Each month from August through November, enumerators ask questions on corn yield in 41 states and soybean yield in 31 states. The sample is the same each month and is selected from the NASS list frame, but excludes the largest producers to reduce respondent burden. The primary mode of data collection is by telephone. The OYS is a field-measurement survey conducted monthly from May through December. Each month from August through December, enumerators collect data on corn in 10 states and soybeans in 11 states. Enumerators measure crop characteristics in the sampled plot, such as number of plants, number of fruit, and fruit measurements. The sample is the same each month and is selected from the NASS area frame. The only mode of data collection is face-to-face. The Quarterly Crops APS Survey is a farmer-interview survey conducted quarterly in March, June, September, and December. In December, enumerators ask questions on corn and soybean yields in 48 states. The sample is selected from both the NASS list and area frames and the sample size is nearly three times larger than AYS. The primary mode of data collection is by telephone. The Crop Progress and Conditions Survey is an interview survey conducted monthly from early April until late November. Enumerators ask two types of questions, crop progress and crop condition, in 48 states. Crop progress questions ask respondents to estimate the county's percent of a particular crop that is at or beyond a specified stage of development, while crop condition questions ask respondents to estimate the county's percent of a particular crop that is in each of five condition categories ranging from

very poor to excellent. The sample is the same each month and consists largely of extension agents and Farm Service Agency staff (who are also employed by USDA). The most common mode of data collection is through a secure internet site.

The Crop Agricultural Statistics Board (ASB), whose members are commodity experts, uses the survey results along with weather data to determine yield forecasts and estimates. The Crop ASB compares the current data for a particular month to historical results for the same month and synthesizes the information to obtain a forecast/estimate for the geographic area represented by the OYS states. This process is not easily repeatable and does not result in an associated measure of uncertainty.

To meet OMB standards, NASS developed a model that mimics the Crop ASB's decision-making process. For the geographic area represented by the OYS states, a hierarchical Bayesian model is used to combine both current and historical corn and soybean yields from the AYS, OYS, and Quarterly Crops APS Survey (when available) for a particular month to obtain a composite forecast. The model also incorporates three covariates, although they differ somewhat for each crop. The first covariate is the state's percentage of the crop rated good or excellent for a particular week; this information comes from the Crop Progress and Conditions Survey. The particular week is different for corn and soybeans, week 30 for corn and week 34 for soybeans. The second covariate is the state's average monthly temperature in July for corn and in August for soybeans, although July is used for soybeans until August is available. The third covariate is the state's average monthly precipitation in July for corn and in August for soybeans, although again July precipitation is used for soybeans until August is available. Finally, the model includes a linear trend, which accounts for the fact that corn and soybean yields are increasing over time. In addition to the forecasts/estimates, the model provides standard errors. More detailed information about the model can be found in Nandram, Berg, and Barboza (2013) and Adrian (2012).

5. Time Series Model for Hogs

NASS publishes the *Quarterly Hogs and Pig*s report with hog inventory and farrowing information. This information is published in March, June, September, and December. Published estimates are subject to revision for a certain amount of time. Revisions are based primarily on data that become available following the initial estimates, such as marketing or slaughter data. When this occurs, the current survey estimates are evaluated in light of any recommended revisions.

To publish hog estimates, NASS performs a farmer-interview survey on a quarterly basis to obtain detailed inventory of breeding and market hogs and the future supply of market hogs based on pig crop and farrowing intentions. Hog owners, including contractors, are the target population. The sampling universe is all hog operations with the capacity to farrow, raise, or finish hogs. The sample is selected from both the NASS list and area frames. The December survey includes the entire U.S., whereas the March, June and September surveys contain operations in the 30 largest hog-producing states. Enumerators collect information by mail, telephone, and face-to-face, and some hog producers fill out the survey questionnaire electronically via a secure internet site. Information collected includes data on sows and gilts for breeding, sows and gilts expected to farrow, boars and young males used for breeding, weight groups of market hogs and pigs (under 50 pounds, 50 to 119 pounds, 120 to 179 pounds,

and 180 pounds and over), and sows and gilts that previously farrowed and the resulting pig crop.

The Hog ASB, whose members consist of experts in this commodity, uses the survey results in combination with other information such as slaughter, imports, and exports to determine the published hog estimates for the entire U.S. The Hog ASB also uses time trend analysis, balance sheets, and considers numerous relationships (e.g., the time between birth and slaughter for a pig is approximately six months). As with the Crop ASB, this process is not easily repeatable and does not result in an associated measure of uncertainty.

Again, to meet OMB standards, NASS developed a model to reflect the Hog ASB process. Using time series methodology, a state-space model with an Extended Kalman Filter is used to obtain estimates and standard errors of total hog inventory, pig crop, sows farrowed, breeding herd, and weight groups of market hogs at the U.S. level. The current model incorporates all measurements of hog inventory (including the Hog ASB estimates and the survey estimates) and administrative data such as slaughter, imports, and exports. In addition, the model uses constraints to enforce the many different relationships that must hold such as the relationship between hogs slaughtered and pigs farrowed two quarters prior. After the U.S. estimates are calculated, a restricted least squares technique is then used to calculate the state-level estimates. More detailed information about the model can be found in Busselberg (2013).

6. Bivariate Area-Level Model for County-level Cash Rental Rates

Prior to 2014, NASS published cash rental rates for every county with 20,000 or more acres of cropland and pastureland on an annual basis in October. When applicable, the county-level cash rental rates were published by three land types: irrigated cropland, non-irrigated cropland, and pastureland. In 2014, the agency announced county-level cash rental rates would be published on an as-needed basis. For this paper, discussion will focus on the annual estimation program (i.e., prior to 2014).

NASS conducts a farmer-interview survey called the Cash Rents Survey. Enumerators collect data on acreage rented from others and value of rent paid for the three land types listed above. Value of rent paid can be reported in rent per acre or total dollars paid. The sample is selected from the NASS list frame and consists of approximately 240,000 farms across the U.S. The data collection process begins in mid-February and concludes at the end of July. The survey questionnaire is mailed twice before a non-response follow-up is conducted by telephone. The response rate for mailed questionnaires is only about 35 percent, so the primary mode of data collection is by telephone.

Statisticians in the NASS Regional Field Offices have the primary responsibility for establishing the published county-level cash rental rate estimates. The only source of current county-level information is the survey estimates from the Cash Rents Survey. Another source of current information is the state-level cash rental rates for irrigated cropland, non-irrigated cropland, and pastureland. Prior to conducting the Cash Rents Survey, state-level estimates are published based on a different survey, the details of which are omitted here. In addition to these two pieces of current information, the previous year's survey estimates and previous year's published estimates are readily available. The published estimate may differ from the survey estimate if the survey estimate is deemed to be unreliable. This phenomena occurs for several

reasons. First, only a few respondents may report positive data (i.e., farmer paid rent for land in cash). Second, an extremely large cash rental rate may be reported within the county (i.e., an outlier). Finally, a county-level cash rental rate may need to be adjusted so that the county-level estimates are in agreement with the previously published state-level estimates. As with the Crop and Hog ASBs, this process is not easily repeatable and does not result in an associated measure of uncertainty.

Again, to meet OMB standards, NASS developed a model that uses experts' thought processes in addition to covariates that were not previously utilized. A bivariate area-level model is used; the two separate univariate area-level models incorporate the previous year's cash rent per acre as a random variable. One univariate model utilizes the average of the current and previous cash rent per acre and the other univariate model utilizes the difference of the current and previous cash rent per acre. One county-level covariate is utilized by the models; this is referred to as an index covariate because it is comprised of the following nine components: the National Commodity Crop Productivity Index for corn, wheat, cotton, and a maximum; a yield index for irrigated, non-irrigated, total, and hay acres; and the total value of production from the previous agricultural census. The county-level modeled estimates are then benchmarked to the published state-level estimates to ensure the county estimates agree with them. More detailed information about the model can be found in Berg, Cecere, and Ghosh (2014).

7. Conclusion

The purpose of this paper is to give a non-technical overview of recent model-based estimation strategies developed by NASS. After a NASS internal review of OMB's standard related to disseminating estimates, NASS developed several models depending on the data sources available as well as the nature of the data. The two largest gains from these endeavors are that the process is now repeatable and the estimates from these models are accompanied by an error measurement.

8. References

Office of Management and Budget. (2009). Standards and Guidelines for Statistical Surveys. https://www.whitehouse.gov/omb/inforeg_statpolicy.

Nandram, B., Berg, E., & Barboza, W. (2013). A Hierarchical Bayesian Model for Forecasting State-Level Corn Yields. Environmental and Ecological Statistics.

Adrian, D. (2012). A Model-Based Approach to Forecasting Corn and Soybean Yields. Fourth International Conference on Establishment Surveys Proceedings.

Busselberg, S. (2013). The Use of Signal Filtering of Hog Inventory Estimation. Federal Committee on Statistical Methodology Research Conference Proceedings.

Berg, E., Cecere, W., Ghosh, M. (2014). Small Area Estimation for County-Level Farmland Cash Rental Rates. Journal of Survey Statistics and Methodology.