

Strengthening methodological architecture with multiple frames and data sources

James M. Harris^{a,b,*} and Cynthia Z.F. Clark^{a,c}

^a*United States Department of Agriculture, National Agricultural Statistics Service, USA*

^b*Fairfax, Virginia, USA*

^c*Washington, DC, USA*

Abstract. The United States Department of Agriculture, National Agriculture Statistics Service (NASS) has long conducted statistical surveys and published official statistics about almost every aspect of agriculture. Survey frames include both a list of farm establishments and an area frame, which are often combined to produce multiple-frame estimates. The survey results have also been compared to administrative records on the disposition of commodities. The disposition of commodities, or other intermittent administrative records, which are often received sometime after the survey results are published, must align with published results. For example, hog and pig inventories are published quarterly, while complete slaughter administrative information that matches the inventories is available six to seven months later. NASS has utilized balance sheets and other analysis tools in an expert review process to align the published results to the administrative information. However, the tools lack the statistical rigor that is required by the U.S. Office of Management and Budget (OMB) statistical standards. They are also deficient in the area of statistical measures of error, and are somewhat dependent upon the analyst. This paper will describe efforts to combine the multiple-frame survey results and other data sources using rigorous, statistically defensible methodologies that strengthen the overall results and meet OMB guidelines.

Keywords: Administrative records, model-based estimation, balance sheet, expert analyst

1. Introduction

The United States Department of Agriculture (USDA), National Agricultural Statistics Service (NASS) has published information on agriculture production for more than 150 years. In fact, gathering and publishing information on agricultural production was a founding function of the department. In more recent times the information published on agriculture has expanded from a singular focus on the production of agricultural commodities to include a vast array of other topics. NASS publishes information on the economics of the farm economy, provides estimates regarding environmental factors, and collects demographical infor-

mation which is then published in the quinquennial census of agriculture. All of these topics are highlighted in a series of publications that are census and survey-based.

Almost all of the surveys and censuses that are conducted within NASS are based on two primary types of sampling frames – NASS's list sampling frame and its area sampling frame. Often these two frames are combined utilizing multiple-frame estimators. Furthermore, the designs are often complex, target multiple goals, and can result in more than one design-based estimator. These surveys are conducted in relatively short periods of time, and results are published shortly after the data collection window is closed and cleaning and analysis of the data has been completed. An example of the timing is the data collection for the national *Quarterly Hogs and Pigs* release, which begins on the first day of the month and the publication is released on the last Friday of that same month.

*Corresponding author: James M. Harris, United States Department of Agriculture, National Agricultural Statistics Service, USA. E-mail: mark.harris@nass.usda.gov.

One common component most agriculture estimates have, especially those within commodity production, is that the administrative information on the production is available sometime after the initial estimates are published. An example of this is with the hog and pig inventories. Before the animal can be consumed, it is inspected by either Federal or state officials. Thus, the administrative records on hog slaughter are collected and even published by NASS. Consequently, the information on inventories should align with the published slaughter. The disposition information on agriculture production is published for many commodities. Disposition reports for livestock provide the annual balance sheet and income estimates for cattle, hogs, and sheep. These disposition reports include estimates of beginning and ending inventories, births, deaths, in-shipments, marketings, and ending inventories. The reports also note total production and marketings in pounds, average prices, value of production, cash receipts, home consumption, and gross marketings by state and across the country.

Historically, NASS has combined the survey-based estimators, and utilized the administrator information to construct balance sheets. During the process the design-based estimators are compared in analytic processes by subject matter experts to adjust the estimators, with the goal of being consistent to the balance sheets. When actual administrative results are received, the official estimates of production can be revised to ensure coherence with the balance sheet.

Many analytical tools have been developed over time, but almost always the final administrative information on production is compared to the design-based estimators or the historic relationship between the survey results and official published numbers. This process has the advantage of providing plausible and consistent estimates of production, but it also lacks any measures of errors, and can be somewhat dependent on the set of analysts during the review process.

2. Challenge

The challenge presented to the Research and Development Division of NASS by the NASS Administrator was to develop model-based estimators for the commodity estimates with measures of error. The commodity estimates were expected to be coherent between components of the series, and also coherent at the multiple levels of publication. If this was not enough of a challenge, the revisions to match the administrative in-

formation should not be more than the current expert analysis process has produced. The challenge of producing model-based estimators is consistent with the U.S. Office of Management and Budget statistical standards.

3. Working towards a solution

There were enough statistical problems within the challenge, but there were also additional organizational challenges that the Research and Development Division faced. The Division had not attempted to model both survey data and administrative data together in the past. And, this particular challenge was just one of several arduous research topics the Division was undertaking. Not only did the Division lack staff members, but general training of the Division staff often existed at the master's level. These staff members had historically conducted projects investigating improvements to current survey estimators or processes. In fact, much of the staff had been working on operational issues to successfully complete the previous census of agriculture. That is, the research and development staff was actually not researching, but actively engaged in supporting the census of agriculture.

The division's challenge was not identifying research problems. The agency administrator as well as three outside reviews of NASS programs had provided more research opportunities than were possible for the Division to handle. So, the research leadership undertook the challenges to build research capacity by initiating activities in the following four areas: hiring new staff with appropriate skills to begin working on the research problems; initiating Cooperative Research Agreements on key topic areas identified with university academics and research organizations; initiating outside reviews with panels of experts with an eye on implementing the research; and working internally with the Information Technology Division to implement the research into the agency's production processes.

Hiring new staff members at the Ph.D. level was challenging and continues to be so. However, engaging and cultivating graduate students has been accomplished through several tactical maneuvers. Ultimately, we discovered that the key was to provide interesting work and ensure that those researchers had opportunities to collaborate with top academic and governmental partners. Making contact with graduate students and recruiting those students was a result of much personal

cultivation via telephone and online, and even reaching out to these talented individuals at statistical meetings and universities. Recruiting successes have been partially attributed to being present at statistical meetings and making the extra effort to go to universities and meet with the students to describe our statistical problems, our challenges, and our organization. The more graduate students know about your organization, the more they see themselves being researchers within the organization.

Since the methodology challenges were so great, we initiated multiple cooperative research agreements with outside research organizations and university faculty. USDA has a unique ability to enter into cooperative research with universities and other research organizations. Our general model has been to approach individuals or organizations with the unique expertise we believe will benefit the research, and then to work together to develop statements of work, or problem statements. Finally, we assemble an agreement and a team to begin the research. Usually the agreements are with academic institutions, and they naturally bring graduate students. We also put agency employees on the team. These agency employees are very receptive to joining teams and working with the top researchers. They gain valuable experience conducting research and developing their statistical skills. They provide a link between the organization and its complex survey program, which adds to the research team's productivity and the final implementation of the project.

On several occasions, we have assembled panels of technical experts to advise both the researchers and agency leaders on the research and the appropriate ways to advance the methodological architecture. These panel meetings are often one or two days in length. The agency provides background and research information on the topic in advance of the meeting and our researchers present their work on the topic during the meeting. These panels have been particularly helpful with focusing the work and advancing the research product towards implementation. They suggest where future efforts should be concentrated and make recommendations on possible solutions. These panels have been organized internally and externally. Both ways of organizing the panels have been positive, but the more significant the challenge, the more one should consider bringing an outside organization to assemble the panel. As an example, NASS worked with the Committee on National Statistics, National Academies of Science to conduct a review of the census weighting methodology. The panel of experts made several significant sug-

gestions to the researchers, and also provided the leadership team with the confidence to adopt the research proposal to change the weighting methodology for the 2012 Census of Agriculture.

An additional challenge for any production survey organization with a research group is to have the research adopted and integrated into the operational program. Researchers are often focused on the research problem. However, integrating the research into the operational program still requires tremendous efforts and diverse skills sets. We have made the implementation of the research product a goal of the academic researcher, and we have expected all products and computer programs to be documented such that the NASS research staff can take the finalized research and move it into the operation program. Often NASS researchers that were a part of the research team are included in the development or implementation teams. These teams follow information technology (IT) protocol for development and implementation, and are paired with IT specialists and project managers during the development/implementation stages. In addition to the researcher and IT specialist, business experts are often added to the teams for further integration. They deliver the product to the operational staff.

4. Summary

The authors would like to summarize by saying that strengthening methodological architecture with multiple frames and data sources has been challenging, and a dedicated effort will be required before we can say that it has been completely accomplished. However, early successes have shown that the framework we have convoked is working for the National Agricultural Statistics Service. Defining the problems in both a business description and a statistical problem statement requires a significant amount of up-front discussions. However, if we can continue to cooperatively exchange ideas with these discussions, as well as acquire research teams with both the business knowledge and researchers skills, and finally retain leadership teams who are willing to tackle large problems, we will ultimately see success. Hiring young statisticians with the basic skills and teaming them with academic experts, while also bringing other experts to review the work, enhances the research efforts. The expert review panel also provides the leadership with the confidence to move forward on implementing the new methodologies, and it provides the organization's staff with the confidence to embrace the methodology.

This paper has been a general overview of the challenges and processes that the National Agricultural Statistics Service has been undergoing to strengthen methodological architecture with multiple frames and data sources. The authors would like to take an opportunity to say that the advancements at NASS would not have happened without the support of our agency staff and leadership, and the many researchers who have cooperated on the projects. The list of academics are long, but we would like to highlight the efforts of Dr. Nell Sedransk, National Institute of Statistical Science; Drs. Linda Young and Malay Ghosh, University of Florida; Dr. Partha Lahiri, University of Maryland; and Drs. Jae-Kwang Kim and Sarah Nusser, Iowa State University.

References

- [1] Busselberg, Stephen. (2011) Bridging Livestock Survey Results to Published Estimates through State-Space Models: A Time Series Approach. Proceedings of the Government Statistics Section, Joint Statistical Proceedings 2011. Online: http://www.nass.usda.gov/Education_and_Outreach/Reports,_Presentations_and_Conferences/reports/conferences/JSM-2011/JSM-2011-Busselberg.pdf.
- [2] Davies, Carrie (2009). Area Frame Design for Agricultural Surveys, National Agricultural Statistics Service, U.S. Department of Agriculture, RDD Research Report-09-2009. Online: http://www.nass.usda.gov/Publications/Methodology_and_Data_Quality/Advanced_Topics/AREA%20FRAME%20DESIGN.pdf.
- [3] Office of Management and Budget, Standards and Guidelines for Statistical Surveys, Guideline 4.1.2, p 21 (2006) Online: http://www.whitehouse.gov/sites/default/files/omb/assets/omb/inforeg/statpolicy/standards_stat_surveys.pdf.
- [4] F.A. Vogel and G.A. Bange, Understand USDA Crop Forecast, U.S. Department of Agriculture, Miscellaneous Publication No. 1554. Online: http://www.nass.usda.gov/Education_and_Outreach/Understanding_Statistics/pub1554.pdf.