Guest Editorial

Agricultural data integration: The 2019 International Conference on Agricultural Statistics in New Delhi

Linda Young
United States Department of Agriculture, Washington, DC, USA
E-mail: ljyoung999@googlemail.com

Surveys based on samples with known probabilities of inclusion have served as the gold standard for official statistics for decades. National Statistical Systems (NSSs) have used information from other data sources to inform survey estimates. In recent years, the availability of data that could be used to improve estimates, including administrative, weather, soil, and remotely-sensed data, has exploded. At the same time, list frame coverage and response rates have tended to decrease, leading to increasing concerns about the quality of the statistics produced. These and other factors, such as increasing costs, have led NSSs to include more non-survey data into their estimation processes. Integrating data from multiple sources in a principled manner so that the resulting estimates are of good quality and have valid measures of uncertainty is a major challenge that NSSs are addressing. During the 2019 International Conference on Agricultural Statistics (ICAS) in New Delhi, India, and the 2019 World Statistics Congress in Marrakesh, Morocco, agricultural data integration was a major topic. Four papers from those conferences are published in this journal and consider issues from data collection and integration to the analysis of these integrated data.

Two of these articles appeared in the last issue of the *SJIAOS* (Volume 37, Issue 1). Both articles focused on collecting census data. In "Agricultural census 2020 – how to reduce costs and burden? The European Statistical System Approach [1]," Co-authors Selenius, Wirtz, Florescu, and Lazar described technical and methodological features of the newly modified system for the European agricultural censuses. The integration of data from a variety of data sources, leading to a reduction in respondent burden, is one feature. In the second paper "Integrating agricultural statistical operations for optimal data collection [2]," Castano emphasized that censuses should focus on structural farm features while surveys are more useful for producing current statistics. He promoted integration of the census and survey programs as a cost-effective approach to producing official statistics, especially in countries with less developed NSSs.

In this issue, two papers consider other aspects of agricultural data integration. D'Alberto and Raggi, in "From collection to integration: Statistical matching of primary and secondary farm data [3]," D'Alberto and Raggi suggest using Micro Statistical Matching (MiSM) to integrate survey (primary) data with administrative and other data collected by others and acquired by an NSS (secondary data). They provide a case study in which the matching results from non-parametric MiSM are compared to predictive mean matching and Bayesian linear regression. Finally, in "Integration of statistical and administrative agricultural data from Namibia [4]," Berg, Im, Zhu, Lewis-Beck, and Li highlight the strengths and weaknesses of survey and administrative data. Using survey and administrative data from Namibia, they carefully describe

the model development and selection for producing the official statistics, which others may find useful when modeling integrated data.

These four papers illustrate the progress that NSSs are making in acquiring data from diverse sources, integrating the data, and producing official statistics. At the same time, the need for further research is evident.

Linda Young
April 2021

References

[1] Selenius, J., Wirtz, C., Florescu, D., & Lazar, A.C. (2021). Agricultural census 2020 – how to reduce costs and burden? The european statistical system approach. Statistical Journal of the IAOS, 37(1), 327–332. doi: 10.3233/sji-200656.

[2] Castano, J. (2021). Integrating agricultural statistical operations for optimal data collection. Statistical Journal of the IAOS, 37(1), 333–339. doi: 10.3233/sji-200631.

[3] D'Alberto, R., Raggi, M. (2021). From collection to integration: non-parametric statistical matching between primary and secondary farm data. Statistical Journal of the IAOS, 37(2). doi: 10.3233/sji-200644.

[4] Li, J., Lewis-Beck, C., Zhengyuan, Z., Im, J., Berg, E. (2021). Integration of statistical and administrative agricultural data from Namibia. Statistical Journal of the IAOS, 37(2). doi: 10.3233/sji-200634.