

Evaluating U.S. Department of Agriculture's Long-Term Forecasts for U.S. Harvested Area

David Boussios, Sharon Raszap Skorbiansky, and Matthew MacLachlan

What Is the Issue?

The long-term baseline projections of the U.S. Department of Agriculture (USDA) are a departmental consensus that provides commodity projections for the trajectory of global agricultural markets 10 years into the future. The baseline model is used in policy and budgetary matters and by USDA to evaluate the effect of shocks on agricultural markets. For example, the baseline projections are a key component for forecasting USDA outlays in the President's budget. These 10-year projections, which span many commodities and countries, are based on a combination of modeling and expert opinion. The baseline projections are based on incomplete knowledge about the future, so it is to be expected that some divergence between the projections and actual values will occur. Several modeling techniques are available, and USDA engages in continuous efforts to improve the accuracy of projections. This report analyzes the historical baseline projections for U.S. corn, soybean, and wheat harvested area for the period from 1997 to 2017 and examines the potential for two types of modeling approaches to improve the performance of USDA's baseline projections. While these crops are only one piece of the entire set of baseline projections, the three crops analyzed in this report make up the lion's share of U.S. planted acreage and, as such, these estimates are widely reviewed by USDA stakeholders due to their importance for policy and the public. As of the release of this report, USDA has taken steps to introduce statistical forecast models to the baseline process. First, in 2020, USDA estimated time-series models alongside the current baseline process for corn, soybeans, and wheat as part of the 2020 baseline projections. Second, USDA is in the process of testing these models with additional structure and for additional commodities.

What Did the Study Find?

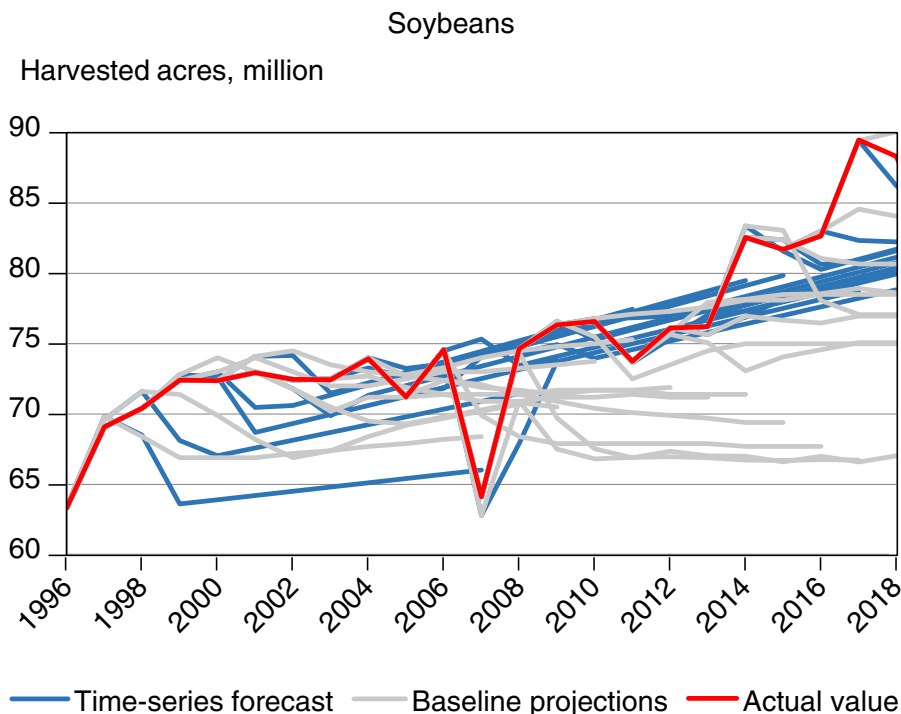
This study found that the baseline projections consistently overestimated the harvested area of wheat area harvested while simultaneously underestimating soybean area. For example, when predicting seven periods ahead, the baseline projections had a mean percentage error of 10.4 percent and -7.3 percent for wheat and soybeans, respectively. The mean percentage error is a measure of forecast bias, i.e. the tendency to over-forecast or under-forecast repeatedly. Harvested corn area was also statistically underestimated, but to a lesser degree than the other two crops studied, with a mean percentage error of -3.3 percent when projecting seven periods ahead.

Naïve and statistical forecasts were used to examine whether the performance of the baseline projections could be improved. A naïve forecast uses only the most recent observation and assumes it will continue at that level for the duration of the forecast horizon. The statistical or time-series econometric models used here use the historical relationship between outcomes over time to fore-

ERS is a primary source of economic research and analysis from the U.S. Department of Agriculture, providing timely information on economic and policy issues related to agriculture, food, the environment, and rural America.

cast 10 periods into the future (or 10 forecast horizons). The study found that generally, both naïve and statistical forecasts outperformed the baseline projections over longer horizons but were less accurate for projecting some less distant outcomes. Results varied by crop.

Economic forecasts result in smaller forecast errors relative to baseline



Note: The red line is the final realized outcome. Gray lines are the baseline's projected values. Blue lines are time-series projections from econometric results.

Source: USDA, Agricultural Baseline Projections, 1997–2017, and USDA, National Agricultural Statistics Service, QuickStats. Time-series projections calculated by USDA, Economic Research Service using data from USDA, Agricultural Baseline Projections, 1997–2017

Forecasts generated using time-series models did not improve performance relative to the current baseline for projections in the first 3 years but did for more distant ones. The statistical forecasts outperformed baseline projections for soybean and wheat areas in 9 and 7 of 10 forecast horizons considered, respectively. The forecasts outperformed the baseline projections in 3 out of 10 horizon lengths for corn. The statistical forecasts reduced the mean absolute percentage error, a measure of forecast accuracy, by 60 percent relative to baseline projections for soybean forecasts of horizon lengths of 10 years. The reduction in forecast error (the difference between the realized and predicted values) was equivalent to improving the performance of the forecast by 6 million acres in 2018 values. The results indicate that statistical forecasting methods are a useful addition to the baseline process, particularly in forecasting more distant outcomes for corn, soybeans, and wheat. USDA has begun to incorporate these models into the baseline process for corn, soybeans, and wheat and evaluate their performance for additional countries and commodities.

How Was the Study Conducted?

This study examined the accuracy of the USDA's long-term (10-year) projections of the harvested area of U.S. corn, soybeans, and wheat compared to the realized (actual) harvested area from 1997 to 2017. The study used the harvested area for corn, soybeans, and wheat as known at the time of each baseline projection. USDA researchers compared the relative performance of forecasts generated by three econometric models and one naïve specification to the baseline projections for these three crops using the same harvested acreage data as known at the time of each baseline projection. The econometric models were selected based on their suitability to reproducing the data generating process, as well as simplicity (parsimony). The quality of the estimates was statistically evaluated based on how frequently they over- or underestimated the actual values, as well as the size of the difference between the projected and actual values.