Journal of Agricultural and Resource Economics: Preprint ISSN: 1068-5502 (Print); 2327-8285 (Online)

doi: 10.22004/ag.econ.358998

Does Complexity Pay? Forecasting Corn and Soybean Yields Using Crop Condition Ratings

Jiarui Li, Scott H. Irwin, and Todd Hubbs

We compare the accuracy of crop condition models from existing literature in forecasting U.S. corn and soybean yields. The data for the study consists of weekly state and national crop condition ratings from the USDA over 1986 through 2022. A battery of statistical tests is applied to perform out-of-sample forecasts over 2000 through 2022. While there are differences in the accuracy of the models, test results are uniform in suggesting that no model has statistically significant superior forecast accuracy. A key finding is that relatively simple models perform just as well as more complex and computationally demanding models.

Key words: accuracy, condition, corn, forecast, ratings, soybeans, USDA, yield

Introduction

Accurate forecasts of crop yield are highly valuable for several reasons. From a market perspective, yield forecasts are an essential component of supply, demand, and price forecasting. From a policy perspective, yield forecasts are important to governments around the world to assess drought impacts and food insecurity. In addition, these forecasts are crucial for farmers and agribusiness firms in developing marketing and risk management plans.

Given the importance of crop yield forecasts, it is no surprise that there is a very large literature on the relationship between weather, technology, and crop yields dating back to the early 1900s (e.g., Tannura, Irwin, and Good, 2008). Broadly speaking, this literature shows that summer precipitation and air temperature directly influence yield potential, along with other factors including soil quality, planting date, disease, insects, and technological improvements from seed genetics, fertilizers, and grower management techniques.

A popular approach among market analysts in both the private and public sectors is to forecast U.S. crop yields based on U.S. Department of Agriculture (USDA) condition ratings. The ratings are released weekly during the growing season and reflect the subjective judgment of thousands of observers about crop yield prospects. Importantly, the ratings are reported as the percentage of a crop rated in five mutually exclusive and exhaustive categories: very poor, poor, fair, good, and excellent. Many analysts use the sum of good and excellent condition ratings to build a simple condition index and relate this to trend-adjusted crop yields. Several representative articles applying this approach to forecasting U.S average corn and soybean yields can be found at the *farmdoc daily* website (Irwin and Good, 2017a, b; Irwin and Hubbs, 2018a, b, c, d).

Jiarui Li is a former PhD student in the Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign. Scott H. Irwin (corresponding author, sirwin@illinois.edu) is the Laurence J. Norton Chair of Agricultural Marketing, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign. Todd Hubbs is a former Cross Commodity Analyst in the Economic Research Service of the U.S. Department of Agriculture.
The authors thank Hongxia Jiao for her assistance with collecting the data for this study. We also than Lance Honig of NASS/USDA for answering numerous questions about condition ratings.

Review coordinated by Liang Lu.

Despite the widespread use of crop condition ratings to forecast crop yields in private and public organizations, there are only a few studies in the academic literature that investigate condition-based forecasts. The general idea behind these studies is to transform the ordinal condition ratings to a numeric condition index and then construct a time-series model between yields and the condition index. For example, Kruse and Smith (1994) developed a weighting system that estimates a changing yield weight for each crop condition class in the growing season for corn and soybean. By multiplying each crop condition ratings by the corresponding yield weight, they computed an average in-sample yield estimate at the state-level. Fackler and Norwood (1999) built a similar state-level yield forecasting model for corn, cotton, soybeans, and spring wheat with an estimated yield weight that is unchanging throughout the growing season. They showed that for each condition class, the product of estimated yield weight and condition rating reflects its average yields. Bain and Fortenbery (2017) used fixed weights to construct a condition index in a yield forecasting model for wheat. Their condition index is based on a straightforward system where for the lowest very poor condition is assigned a weight of zero, and as the condition increases by one category level, the corresponding weight increases by 0.25 until it reaches the highest excellent condition with a weight of one.

Begueria and Maneta (2020) developed a sophisticated two-stage yield forecasting model based on crop condition ratings for corn, cotton, soybeans, and winter wheat at the state level. They argued that spatial and temporal differences in crop condition information should be directly modeled before making yield forecasts. Hence, the authors developed a cumulative link mixed model to transform raw condition data to a continuous and almost normal-distributed crop condition index. After removing space and time effects, they argued that maximum information can be extracted from crop condition ratings, which offers a better possibility of providing unbiased and accurate yield forecasts. Begueria and Maneta (2020) provided evidence that their modeling approach achieves large improvements in accuracy over simpler condition-based forecasts, such as Jorgensen and Diersen (2014) and Irwin and Good (2017a, b).

The improvements in forecast accuracy reported by Begueria and Maneta (2020) are interesting for three reasons. First, the finding that a complex model beats simpler models in terms of forecast accuracy runs counter to a large body of literature on the forecasting of various variables, including GDP growth, inflation rates, unemployment levels, stock prices and market trends. Armstrong (2001, p. 693) summarizes the evidence as "…showing that while some complexity may improve accuracy, seldom does one need highly complex methods. In some studies, complexity harmed accuracy." The results reported by Begueria and Maneta (2020) may represent an important exception to this general result. Second, the forecast results in Begueria and Maneta are based on a cross-validation procedure that leaves out one observation at a time and forecasts the "missing observation" regardless of its ordering in time. This procedure is only applied to the second stage of the estimation, which is quite different from the recursive out-of-sample procedures that are standard in the time-series forecasting literature. Third, Begueria and Maneta (2020) did not compute forecast error statistics for simpler models using the same data set as in their study, but, rather, relied on forecast statistics reported in the original articles.

The purpose of this study is to evaluate the forecast accuracy of crop condition models for U.S. average corn and soybean yields. Specifically, we compare the forecast accuracy of the simpler models designed by Irwin and Good (2017a, b) and Bain and Fortenbery (2017) to the model developed by Begueria and Maneta (2020), which provides a representative set of models ranging from relatively simple to highly complex specifications. The data for the study consists of weekly state and national crop condition ratings from 1986 through 2022 for corn and soybeans. To evaluate the predictability of the yield forecasting models, we use data from 2000 through 2022 as the out-of-sample period. A battery of statistical tests is applied to the out-of-sample crop yield forecasts. The test results contradict Begueria and Maneta's (2020) finding that their model outperforms simpler models, which include the Kruse and Smith (1994) regression model for the period 1986-1993; the Jorgensen and Diersen (2014) regression model for the period 1986-2012; and the Irwin and Good (2017a) regression model for the period 1986-2016.

Data

From roughly late April through the end of November each growing season, USDA weekly Crop Progress reports provide progress and condition ratings for corn and soybean in 18 major producing states. The reports are published on the first business day of the week after 4:00 pm Eastern time. Estimates in the report are based on non-probability subjective surveys conducted by nearly 4,000 local crop observers, who are drawn from the ranks of extension agents, USDA Farm Service Agency (FSA) staff, elevator managers, and other agricultural professionals. Each local observer follows the standard definitions and guidelines provided by the USDA to conduct assessments of crops in their local area. Data are reported on the progress of producer activities (e.g., planting and harvesting), various phenological stages of development (e.g., emergence, flowering), and crop condition ratings. It is important to emphasize that weekly observations are entirely subjective and the result of visual field observations, direct conversations with farmers, and expert local knowledge. For this reason, the data collection process for USDA Crop Progress reports can be described as a system of "people as crop sensors." Finally, state-level estimates are based on weighting of local observer estimates, usually at the county level, and national-level estimates are based on weighting of each state's planted acreage estimate from the previous year (Irwin and Good, 2017a).[1]

The data released in the weekly Crop Progress report are followed closely by grain market participants. For example, Lehecka (2014) notes that these reports are among the most requested publications distributed by the USDA between monthly Crop Production and World Agricultural Supply and Demand Estimates (WASDE) reports. Using event study methods, Lehecka shows the strongest corn and soybean futures market reactions are found in July and August, when weather conditions are most critical for crop development. He also finds that market reaction to the release of the weekly Crop Progress report has increased over time.

Lehecka's work shows that Crop Progress reports have substantial informational value to participants in the grain futures markets. As discussed above, this is especially true during the heart of the summer growing season for corn and soybean. It is during these months that crop condition ratings take center stage. The ratings are reported in five exhaustive categories as follows:[2]

Very Poor – Extreme degree of loss to yield potential, complete or near crop failure. Pastures provide very little or no feed considering the time of year. Supplemental feeding is required to maintain livestock condition.

Poor – Heavy degree of loss to yield potential which can be caused by excess soil moisture, drought, disease, etc. Pastures are providing only marginal feed for the current time of year. Some supplemental feeding is required to maintain livestock condition.

Fair – Less than normal crop condition. Yield loss is a possibility, but the extent is unknown. Pastures are providing generally adequate feed but still less than normal for the time of year.

Good – Yield prospects are normal. Moisture levels are adequate and disease, insect damage, and weed pressures are minor. Pastures are providing adequate feed supplies for the current time of year.

Excellent – Yield prospects are above normal. Crops are experiencing little or no stress. Disease, insect damage, and weed pressures are insignificant. Pastures are supplying feed in excess of what is normally expected at the current time of year.

The ratings for a given crop in each condition category are expressed as a percentage, reflecting the proportion of the crop rated in a particular category. Since the categories are exhaustive, the percentages in the five categories sum to 100.

---

[1] See the discussion at the USDA/NASS website here: https://www.nass.usda.gov/Surveys/Guide_to_NASS_Surveys/Crop_Progress_and_Condition/index.php.
[2] The definitions are found on this page at the NASS website: https://www.nass.usda.gov/Surveys/Guide_to_NASS_Surveys/Crop_Progress_and_Condition/index.php.

We collected all weekly condition ratings for corn and soybeans at the state and national level starting in 1986, when the program was established, through 2022. For each year, the coverage of weeks in the growing season is not the same because ratings do not begin until a substantial part of the crop has emerged and do not end until most of the crop is mature. Since dates for emergence and maturity vary from year-to-year, the beginning and ending dates for condition ratings also vary. To obtain a consistent evaluation period for all competing models, we use weeks 23 through 39 for corn and weeks 25 through 39 for soybeans to evaluate the yield forecasts. The ranges roughly correspond to early June to late September for corn and late June to late September for soybean. Corn and soybean ratings are available for all years during the sample period for these weeks and for all but a few of the 18-states included for each crop in the Crop Progress report.

The Begueria and Maneta (2020) model provides weekly yield forecasts at the state-level for the 18 major-producing states included in the Crop Progress report for corn and soybean due to the design of their framework. We are interested in yield forecasts at the national level because this is a key determinant of market prices rather than yield in any individual state. To compare all competing models at the national level, we developed a straightforward method of converting a set of state-level forecasts to one national level forecast. Specifically, we use the ratio of weighted-average yields of 18 states to the national yields. Once the state-level yield forecasts are available, forecasts of national yields can be easily calculated using the estimated ratio. For these 18 states, each individual state has different productivity for corn and soybean. We use the proportion of individual state's harvested acres out of the total harvested acres of 18 states to estimate the yield weight for each state. Each year for each state, we use the previous five-year moving-average yield weight as a forecast for current year's yield weight. For the ratio of weighted sum of state-level yields to the final estimates of national yields, we apply a similar previous five-year moving-average procedures to acquire a forecast for the current year's state-to-national yield ratio.

Since a five-year moving-average procedure is applied to harvested acres, and the first year we use the crop condition ratings for yield forecasts is 1986, we collected harvested acres for each state from 1981 through 2022. The harvested acres data are obtained from the NASS Quick Stats website and are originally published in the Acreage report released each year at the end of June.[3] The Acreage report includes revised harvested acres for the previous year and forecasted harvested area for the current year. The timing of the Acreage report roughly lines up with the beginning of the forecast window each year for the present study.

Yield Forecasting Models

The yield forecasting models used in this study provide early yield projections when weekly condition ratings are available for corn and soybean. Figure 1 uses corn to illustrate a typical forecast cycle. Each year of our sample, the first yield prediction starts in week 23 (the week of June 3). The yield forecasts for week 23 are obtained using crop condition ratings published in this week. Importantly, all the forecast models are estimated recursively using samples that end before a given forecast week. The out-of-sample period is 2000 through 2022 and forecasts for corn are made for week 23 through week 39 (the week of Sep. 23) in each year and for soybean for week 25 (the week of June 17) through week 39. To evaluate the performance of yield forecasting models, we compare the weekly forecasts with final yield estimates published in the USDA's Crop Production Annual Summary report that is released in January after the growing season.

The design of the Irwin and Good (2017a) model makes it applicable for both state-level and national-level yield forecasts. At the national level, the Irwin and Good National model (IG National model, hereinafter) is specified as follows:

[3] https://www.nass.usda.gov/Quick_Stats/

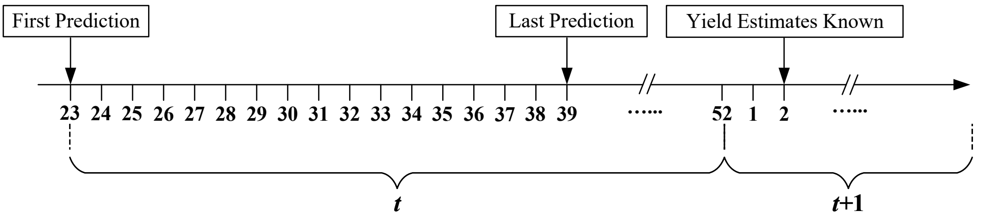


Figure 1. Yield Forecasting Cycle for Corn
Notes: we use corn as an example to illustrate the forecasting cycle. For soybeans, the first prediction is in week 25 and the last prediction is in week 39.

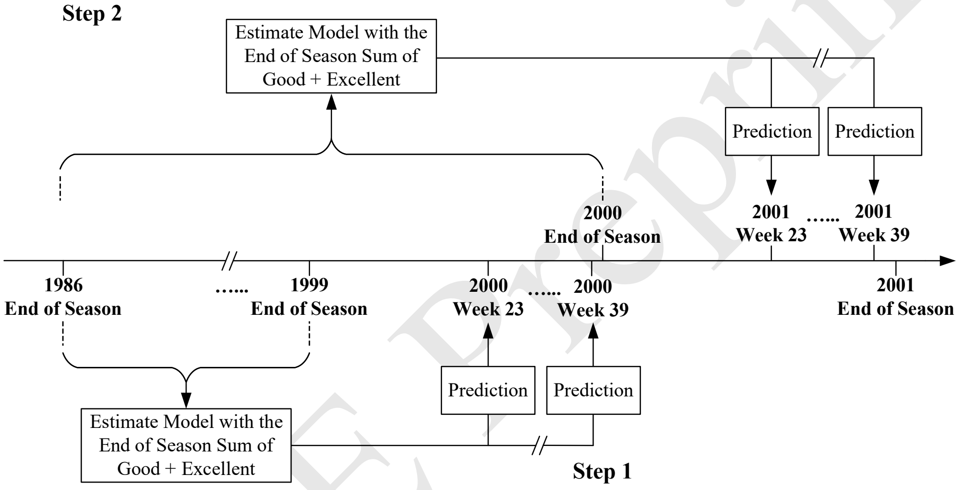


Figure 3. Recursive Out-of-Sample Yield Forecasts with Irwin and Good State Model and Irwin and Good National Model
Notes: We use corn as an example to illustrate the forecasting cycle. For soybeans, the first prediction is in week 25 and the last prediction is in week 39.

(1) $$Yield_t = \beta_0 + \beta_1 year\_index_t + \beta_2 SUM_t + \epsilon_t$$

where $Yield_t$ is national final yield estimates in year t; $year\_index_t$ is the time index in year t; SUM_t is the sum of excellent and good ratings at the end of the season in year t. With corn as an example, Figure 3 illustrates how to provide recursively out-of-sample yield forecasts with the model. In particular, yield forecasts for week 23 in 2000 are obtained with the following steps. First, first ordinary least squares (OLS) regression is used to estimate the IG National model from 1986 to 1999 with the time index, the percentage of corn rated in good and excellent condition at the end of the year, and the national final yield estimate. Second, the sum of ratings in week 23 and the year index for 2000 are entered in the regression model estimated over 1986 through 1999 to obtain a corn yield forecast for week 23 in 2000.

State-level yield forecasts follow the same procedure as at the national level. Instead of using national yield estimates, we use state-level final yield estimates to estimate the Irwin and Good State model (IG State model, hereinafter) and generate weekly yield forecasts for each state. State-level forecasts are aggregated to the national level using the procedure described earlier.

Irwin and Good (2017b) point out that a disadvantage of their approach is that it does not consider bias in the early week condition ratings during the growing season. Irwin and Good show that early condition ratings for corn and soybean, on average, are over-estimated. Early in the growing season, crops often are in a "green" state and retain full yield potential. However, in a few years adverse weather conditions (like the drought in 2012) and the development of plant disease (like the outbreaks of tar spot, a fungal disease damages corn leaf, since 2015) cause crop yield prospects deteriorate. As a result, ratings later in the growing season, on average, tend to be lower than early ratings. This is sometimes referred to as the "browning" of crop condition ratings. To measure the size of this bias, we follow definition of bias proposed by Irwin and Hubbs (2018a, c):

$$bias_t = final\ week\ rating_t - early\ week\ rating_t, \tag{2}$$

where $final\ week\ rating_t$ is the current year's sum of good and excellent ratings at the end of growing season and $early\ week\ rating_t$ is the sum of good and excellent ratings of each early week in year t. We expect the bias to be negative, hence, to adjust the bias in the early weeks, we need to add the bias to the early weeks' ratings as:

$$adj\_early\_rating_t = early\ week\ rating_t + bias_t. \tag{3}$$

For both corn and soybean, the data show that bias is minimal after week 30. Therefore, the bias adjustment is applied only to week 23 through week 30 for corn and week 25 through week 30 for soybean.

We apply moving-average procedures to estimate the size of bias. With ten-year and five-year moving-average approaches, we first calculate the weekly rating difference between the final week and each of the early weeks over the previous ten or five years. Then, we add the estimated bias to the reported ratings for the current forecast year. For some weeks, we do not have consecutive observations in all years. In these scenarios, we use all the available data we have from the previous ten or five years. These two augmented approaches are considered labeled the IG National with Bias Adjustment model.[4]

The Bain and Fortenbery (2017) fixed weight model (BF model, hereinafter) assigns fixed weights to each condition category to transform the ordinal condition ratings to a numerical crop condition index (CCI).[5] Below is the definition:

$$\begin{aligned} CCIndex = 1.0 \cdot Excellent + 0.75 \cdot Good + 0.50 \cdot Fair + \\ .25 \cdot Poor + 0.00 \cdot Very\ Poor. \end{aligned} \tag{4}$$

The ratings for each condition category are in percentages, therefore fixed weights CCI is bounded between 0 and 1. The BF model is specified as follows:

$$Yield_i = \alpha_0 + \alpha_1 \cdot Trend_i + \beta_1 \cdot CCIndex_i + e_i, \tag{5}$$

where $Yield_i$ is the final yields in year *i*, $Trend_i$ is the time index for year *i*, $CCIndex_i$ is the end of season $CCIndex$ value for year *i*. For example, the yield forecasts for week 23 in 2000 for corn are estimated with the following steps. First, crop conditions are transformed at the end of growing season to the fixed weight $CCIndex$ from 1986 through 1999. Second, model (5) is via OLS using the final yield estimates as the response variable and year index and fixed weight $CCIndex$ as explanatory variables. Third, once we obtain the crop condition ratings for week 23 in 2000, we transform them to the fixed weight $CCIndex$ and enter them in the model estimated over 1986 through 1999.

[4] Model comparisons with bias adjustment based on a five-year moving average are similar to those based on a ten-year moving average and are omitted to save space.
[5] Jorgensen and Dierson (2014) use the CCI developed by Bain and Fortenbery.

The Begueria and Maneta (2020) model (BM hereafter) is the most technically sophisticated model considered in this forecast evaluation. They argue that spatial and temporal differences in crop condition information should be directly modeled before making yield forecasts. Hence, a cumulative link mixed model (CLMM) is used to transform raw condition data to a continuous and almost normal-distributed crop condition index (CCI). After removing space and time effects, they argue that maximum information can be extracted from crop condition ratings, which offers a better possibility of providing unbiased and accurate yield forecasts.

In the first stage of the BM modeling approach a CLMM is estimated using a probit link function to connect ordinal response with numeric factors. The CLMM is specified as:

$$probit\big(P(Y_i \leq j|s,y,w)\big) = \theta_j + \beta_y y + \beta_w w + v_s + v_{y,s} y + v_{w,s} w + \epsilon_{si}, \tag{6}$$

where $probit\big(P(Y_i \leq j|s,y,w)\big)$ is the probability that the ith report's condition ratings are no greater than category j, and $j \epsilon [1,4]$ since there are five condition categories; s, y and w are state year and week in report i, respectively; and θ_j is a threshold parameter which remains constant and determines the range of the response variable in a certain category j. There are two fixed effects in the model: a long-term (year) effect and a temporal (week) effect. Three random effect components are included: state, the interaction between state and year, and the interaction between state and week. The error term ϵ_{si} is the unbiased CCI that is specific for each state and is free of any long-term or temporal time effects.

In the second stage of the BM modeling process, a mixed model is specified where the fixed effects are the long-term (year) and CCI effects and the random effect is conditional on year and CCI interactions. This model provides weekly yield forecasts for each state and is specified as:

$$\mu_i(s) = \beta_0 + \beta_y y_i + \beta_c CCI_i + v(s) + v_y(s) y_i + v_c(s) CCI_i + \epsilon_i, \tag{7}$$

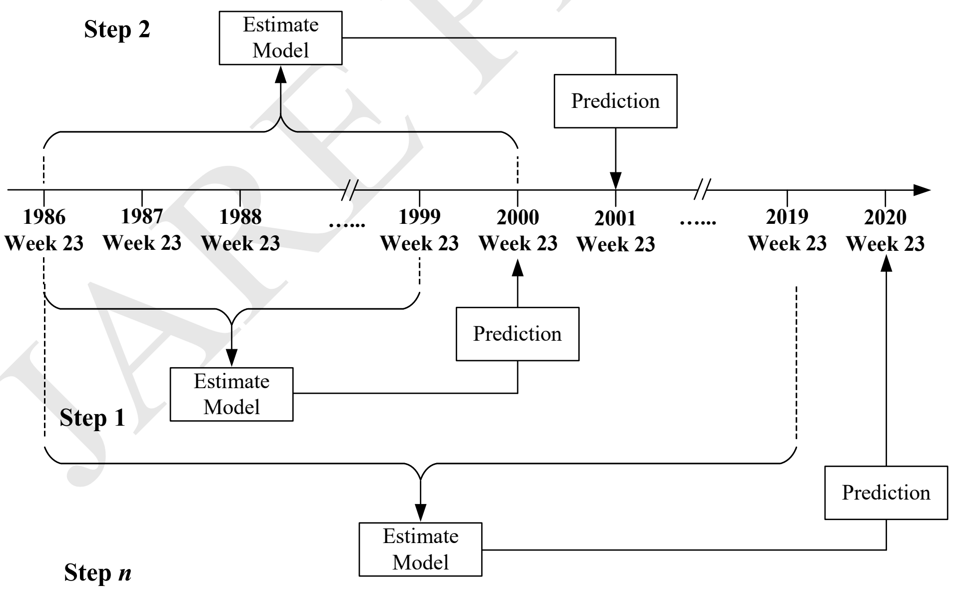


Figure 2. Recursive Out-of-Sample Yield Forecasts with Begueria and Maneta Model (2020)

Notes: We use corn as an example to illustrate the forecasting cycle. For soybeans, the first prediction is in week 25 and the last prediction is in week 39.

where $\mu_i(s)$ is the expected yield at state s and time i, y_i is the transformed year index at time i, CCI_i is the crop condition index at time i, β_0 is the global intercept, β_y is the long-term year effects and β_c is the CCI effect (they are both fixed effects and have the same effects on all the states). The BM model treats the state as a random component, which implies that different states have different temporal effects and CCI effects.

Figure 2

uses corn as an example to illustrate how the BM model recursively provide out-of-sample weekly yield forecasts. Yield forecasts for week 23 in 2000 are estimated with the following steps. First, the CLMM model is estimated using crop condition ratings from the first published Crop Progress report in 1986 to the most recent report published in week 23 of 2000. With the updated model, we can transform and update the ordinal crop condition ratings for all the weeks until week 23 in 2000. Second, we can estimate the mixed model using the updated CCI and other variables in week 23 from 1986 to 1999. Third, the updated CCI and year index for week 23 in 2000 are entered in the mixed model to obtain a yield projection for week 23 in 2000. Following these steps, as we move forward in the growing season, we generate weekly updates of yield forecasts. Fourth, national yield forecasts are generated from the state-level forecasts using the procedures described earlier.

Forecast Evaluation

We conduct two sets of model comparisons in our study. First, we compare all five yield forecasting models to a naïve trend yield model to evaluate the value of crop condition ratings as a yield indicator. Second, we set the BM model as a benchmark to compare it with the other four yield forecasting models. The comparisons are conducted at both the state and national levels.

Weekly forecast errors $e^i_{w,t}$ for model i are defined as the percentage difference between the USDA final yield and a model's yield forecast:

(8) $$e^i_{w,t} = 100 \cdot \frac{\left(y_t - \widehat{y^i_{w,t}}\right)}{y_t},$$

where y_t is the final USDA yield estimates and $\widehat{y^i_{w,t}}$ is the predicted yield in year t for week w produced by model i. We use the root mean squared percentage error (RMSPE) to measure each model's predictive accuracy. RMSPE is defined as

(9) $$RMSPE^i_{w,t} = \sqrt{\frac{1}{n}\sum\left(e^i_{w,t}\right)^2},$$

where n is the number of observations in each week over the out-of-sample period.

Naïve Trend Yield Model

One of the key factors that determines crop yields is technology improvement over time. Crops tend to increase in yield year-by-year, which is known as the "trend yield." A naïve trend yield model serves as the base model that we use to compare with five yield forecasting models since it only accounts for the variation in time. The Naïve trend yield model is specified as below:

(10) $$Yield_t = \beta_0 + \beta_{1,t} year\_index_t + \epsilon_t,$$

where $Yield_t$ is the national final yield estimates in year t, $year\_index_t$ is the corresponding year index running from 1 to 35 for the year from 1986 to 2022.

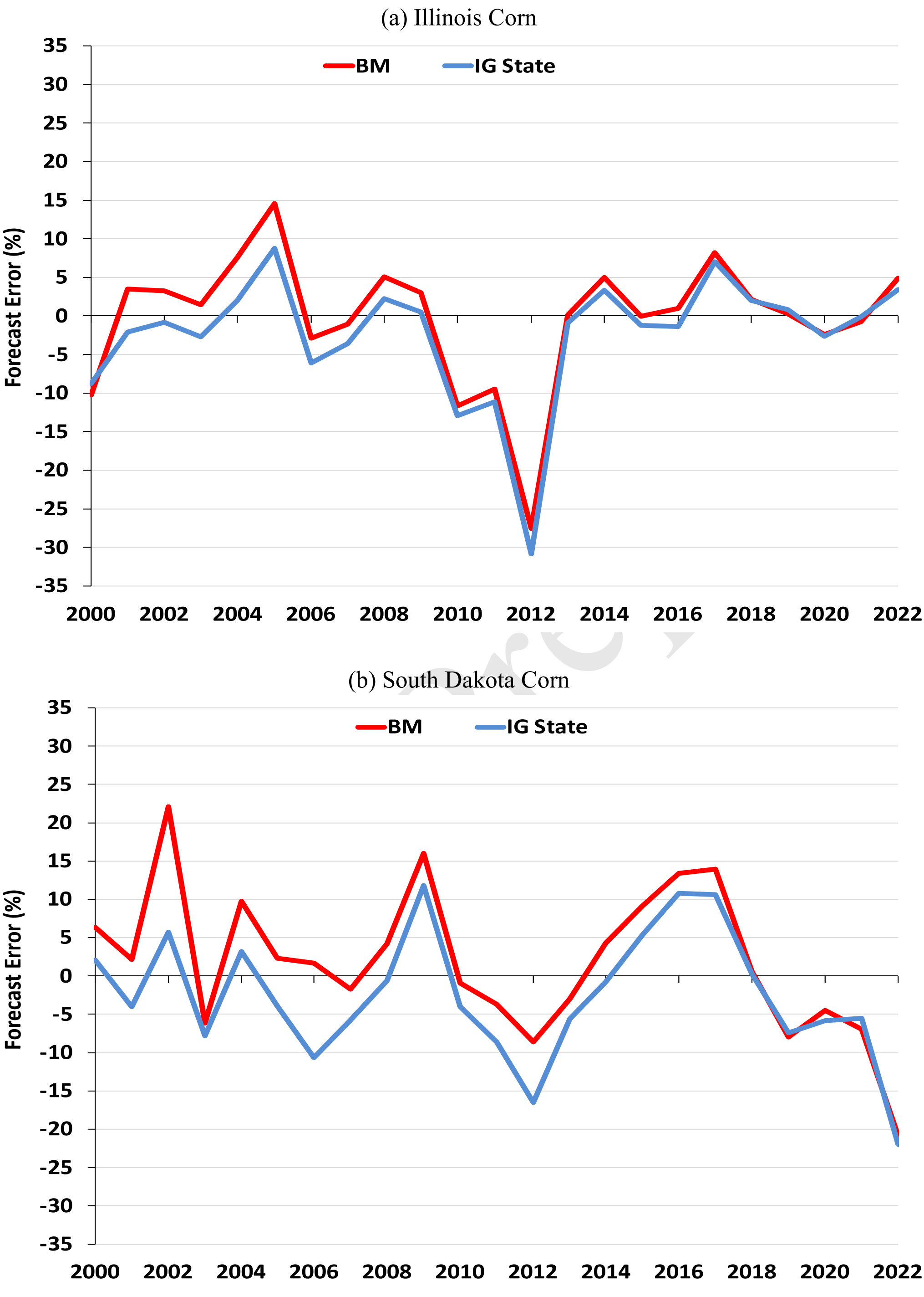


Figure 4. Forecast Error of the BM Model and IG State Model for Week 29 for Illinois and South Dakota in Corn, 2000 – 2022

Notes: BM model is proposed by Begueria and Maneta (2020), IG State model is proposed by Irwin and Good (2017a).

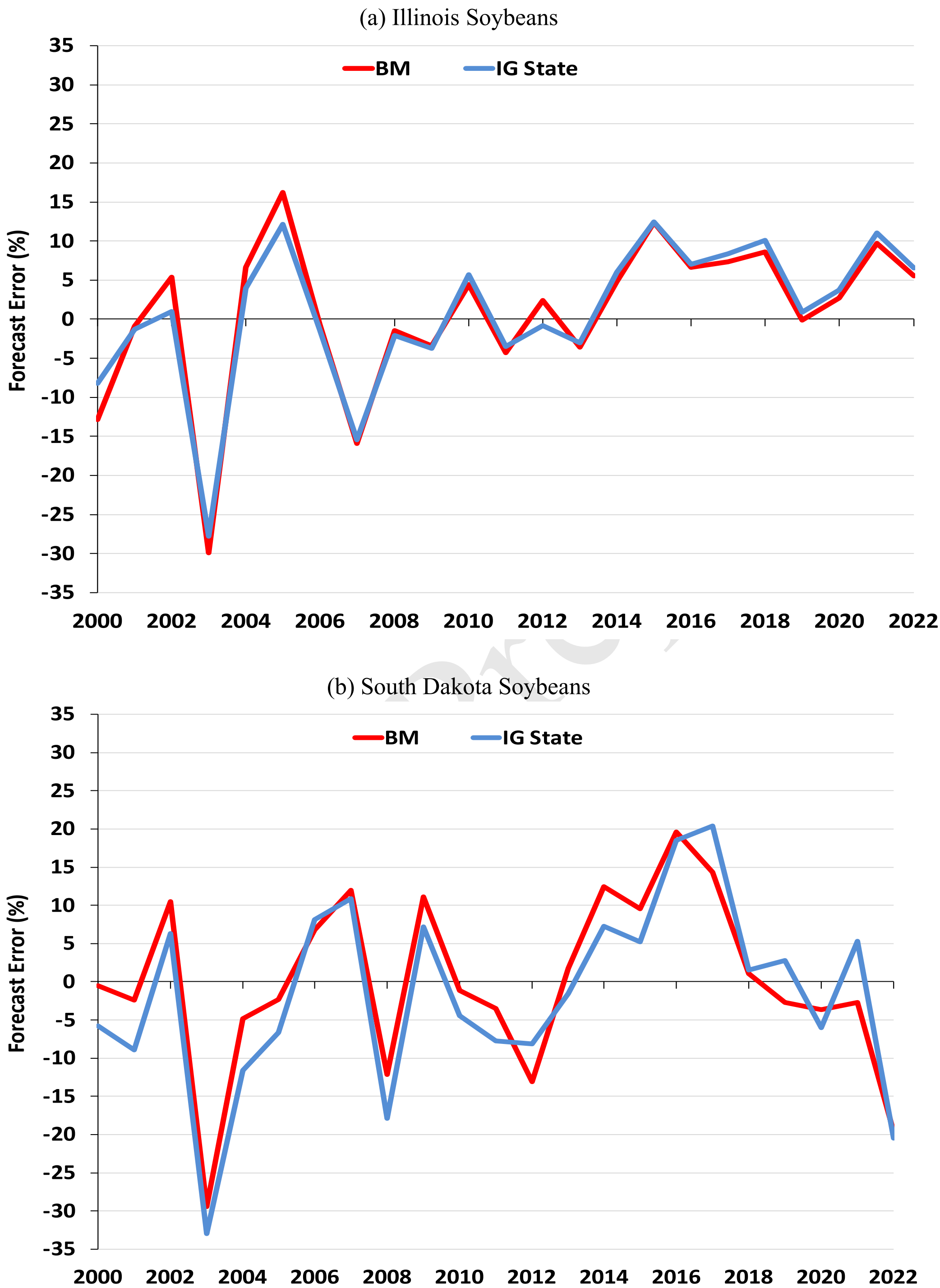
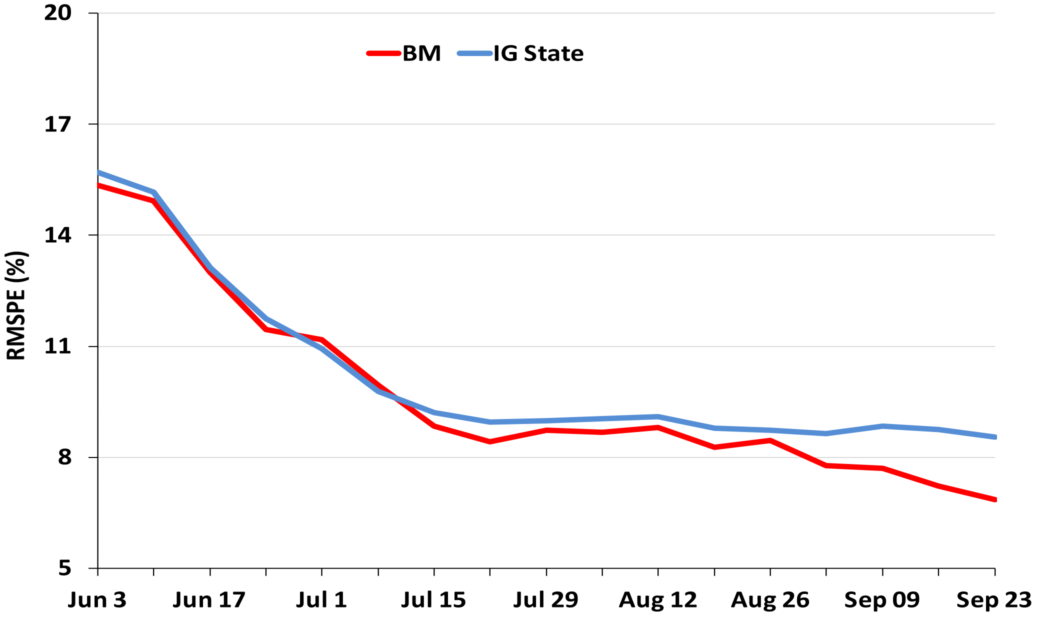


Figure 5. Forecast Error of the BM Model and IG State Model for Week 29 for Illinois and South Dakota in Soybeans, 2000 – 2022

Notes: BM model is proposed by Begueria and Maneta (2020), IG State model is proposed by Irwin and Good (2017a).

(a) Illinois Corn



(b) South Dakota Corn

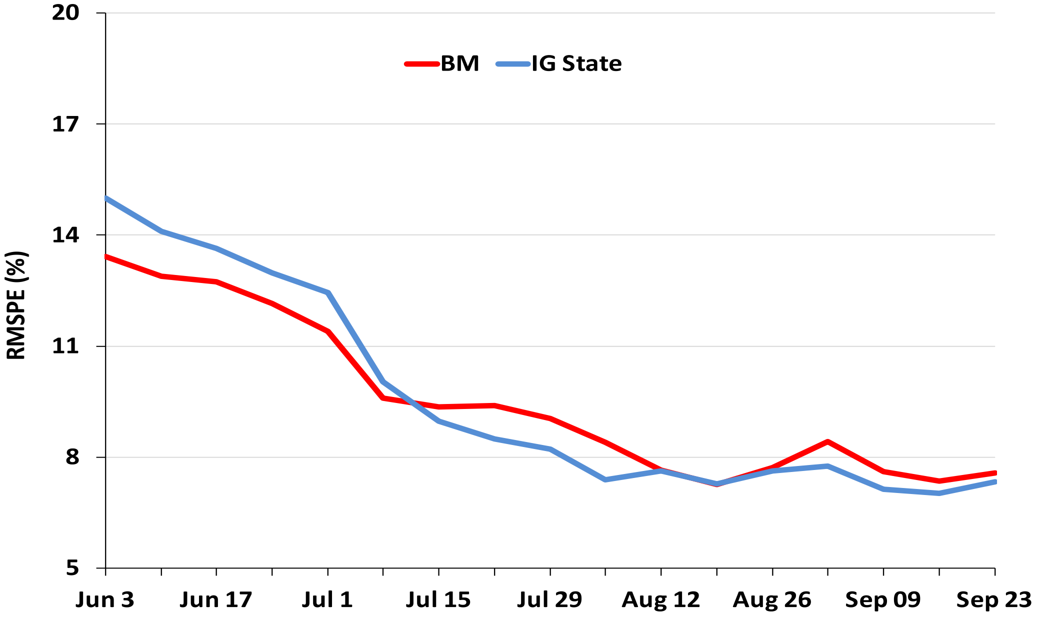
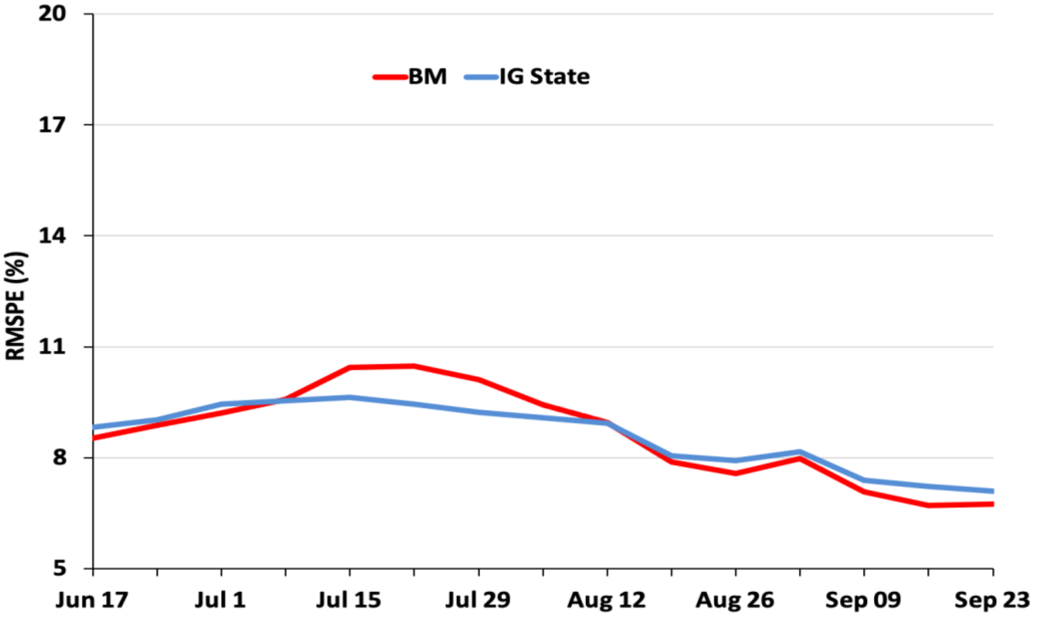


Figure 6. Weekly RMSPE of the BM Model and IG State Model for Illinois and South Dakota in Corn, 2000 – 2022

Notes: BM model is proposed by Begueria and Maneta (2020), IG State model is proposed by Irwin and Good (2017a).

(a) Illinois Soybeans



(b) South Dakota Soybeans

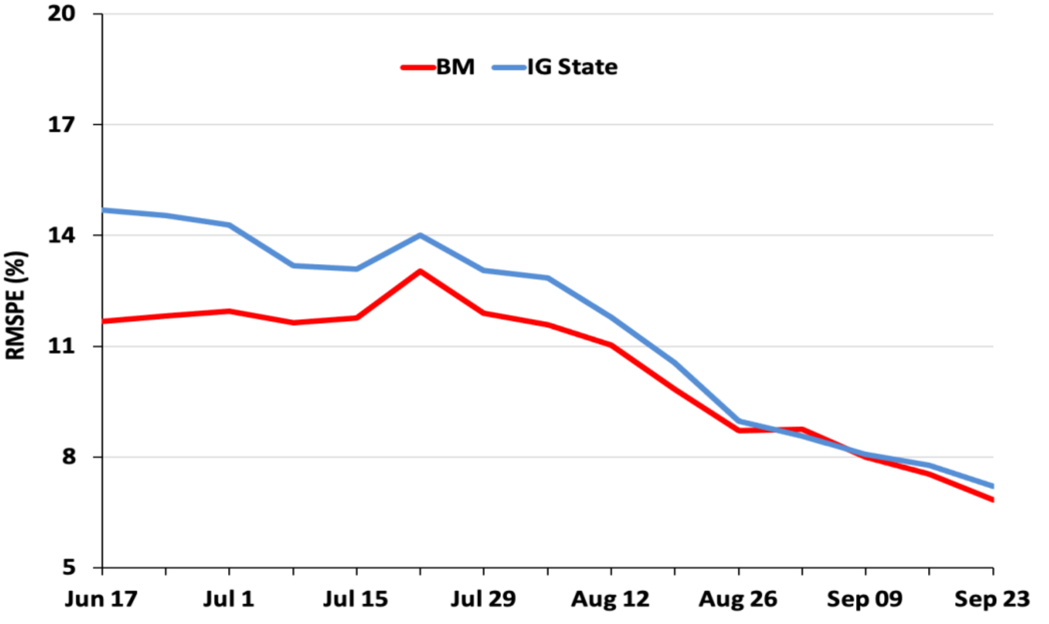


Figure 7. Weekly RMSPE of the BM Model and IG State Model for Illinois and South Dakota in Soybeans, 2000 – 2022
Notes: BM model is proposed by Begueria and Maneta (2020), IG State model is proposed by Irwin and Good (2017a).

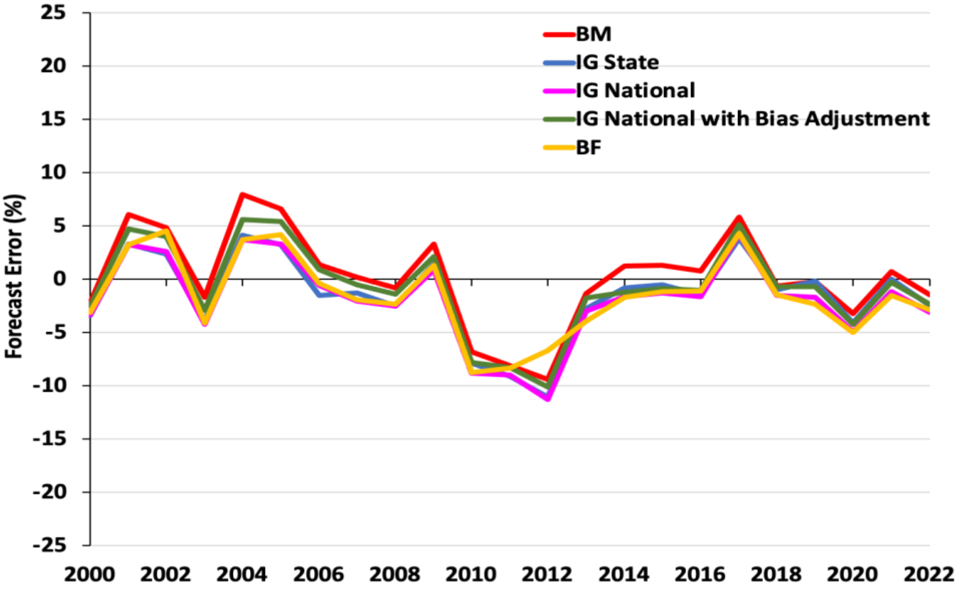


Figure 8: Forecast Error (%) of Five Yield Forecasting Models for Week 29 at the National Level in Corn, 2000 – 2022
Notes: BM model is proposed by Begueria and Maneta (2020), IG State model, IG National model, IG National with Bias Adjustment Model are proposed by Irwin and Good (2017a), and BF model is proposed by Bain and Fortenbery (2017).

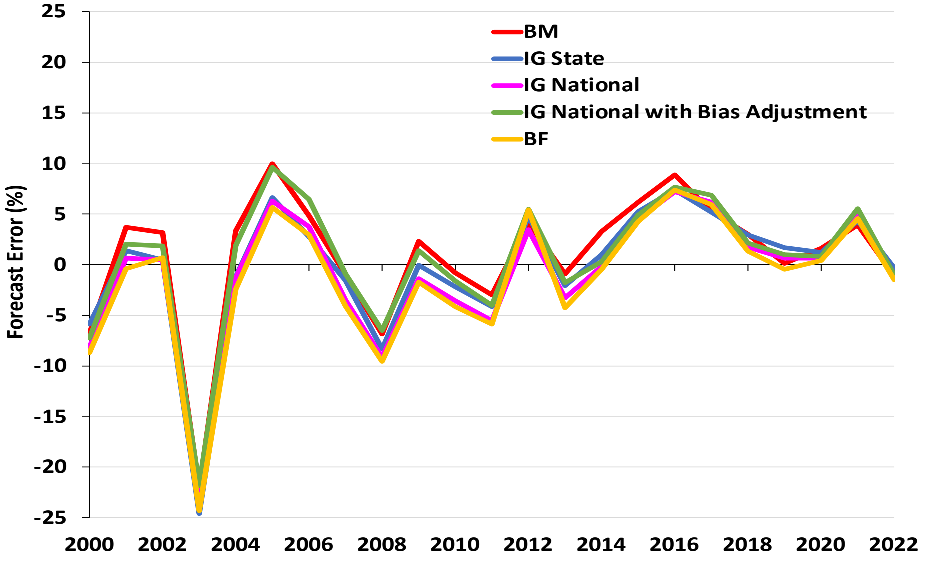


Figure 9. Forecast Error (%) of Five Yield Forecasting Models for Week 29 at the National Level in Soybeans, 2000 – 2022
Notes: BM model is proposed by Begueria and Maneta (2020), IG State model, IG National model, IG National with Bias Adjustment Model are proposed by Irwin and Good (2017a), and BF model is proposed by Bain and Fortenbery (2017).

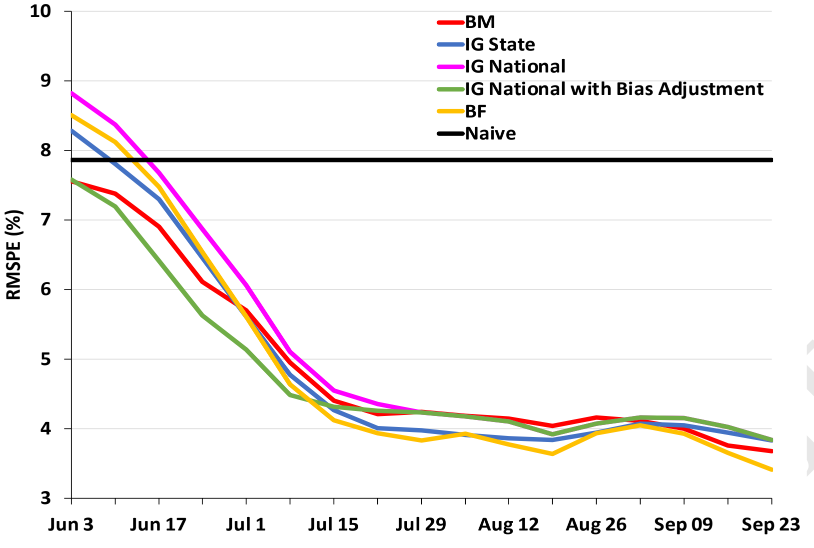


Figure 10. RMSPE of Five Corn Yield Forecasting Models at the National Level, 2000 – 2022

Notes: BM model is proposed by Begueria and Maneta (2020), IG State model, IG National model, IG National with Bias Adjustment Model are proposed by Irwin and Good (2017a), and BF model is proposed by Bain and Fortenbery (2017). We also include naïve trend yield model to illustrate the value of crop condition ratings as a yield indicator. The RMSPE of the IG National and IG National with bias adjustment models is the same starting on July 29th because the bias adjustment is set to zero starting this week.

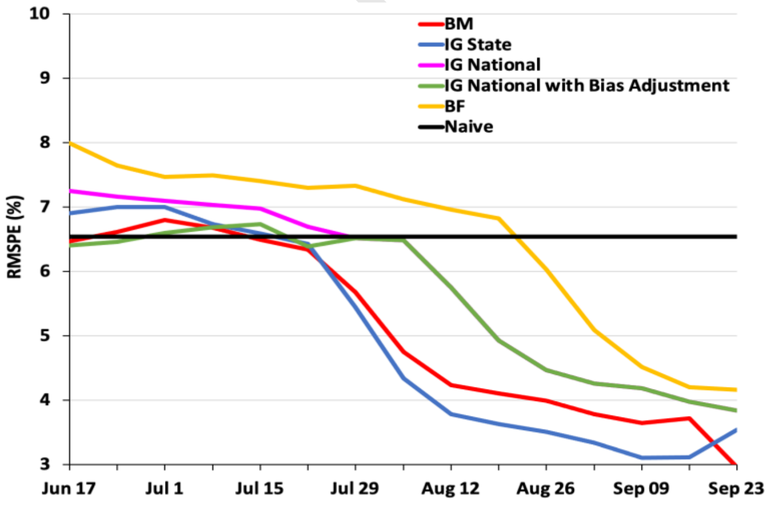


Figure 11. RMSPE of Five Soybeans Yield Forecasting Models at the National Level, 2000 – 2022

Notes: BM model is proposed by Begueria and Maneta (2020), IG State model, IG National model, IG National with Bias Adjustment Model are proposed by Irwin and Good (2017a), and BF model is proposed by Bain and Fortenbery (2017). We also include naïve trend yield model to illustrate the value of crop condition ratings as a yield indicator. The RMSPE of the IG National and IG National with bias adjustment models is the same starting on July 29th because the bias adjustment is set to zero starting this week.

The yield forecasts provided by naïve trend yield model also follow the recursive out-of-sample forecasting approach. For example, when we are in year 2000, we use yields and time indices from 1986 to 1999 to train the model. In 2000, we make yield predictions using the updated year index of 15 for all weeks during the growing season for corn and soybeans.

Comparisons at State Level

We begin the forecast comparisons for two representative states that have significant geographic differences in production characteristics—Illinois and South Dakota. The state-level forecasts are generated by the BM and IG State models. Figures 4 and 5 present the percentage errors of the forecast yields of these two models for a mid-season week (week 29) over the out-of-sample period for corn and soybeans, respectively. Figures 4a and 5a show that in Illinois the errors for the two models are quite similar and there is no clear pattern of which model performs best over time. Figures 4b and 5b show more variability for South Dakota, but there is no clear pattern of which model performs best over time. For 2012, when crop production was sharply impacted by a historic drought, we observe that the BM model provided more accurate yield forecasts than the IG State model in the mid-growing season for Illinois, whereas for South Dakota, the IG State model was more accurate.

Figures 6 and 7 show the RMSPE for each week during the growing season over the entire out-of-sample period (2000-2022) for corn and soybeans, respectively, in Illinois and South Dakota. Figure 6a indicates that the BM model has better performance from mid-July until the end of growing season for corn in Illinois. Figure 7a shows the IG State model outperforms the BM model in Illinois from mid-July to mid-August for soybean. For South Dakota, Figure 6b shows that the BM model takes the lead from early-June to early-July for corn, then the IG State model provides more accurate yield forecasts from early-July until the end of growing season for corn. Figure 7b suggests that the BM model has better forecasting performance from early-June to late-August for soybeans, then the Irwin and Good model takes the lead through the end of growing season.

Comparisons at National Level

All yield forecasting models in this study provide national-level yield forecasts for each week during the growing season over the out-of-sample period. Table 1 presents the RMSPE of the five forecasting models for each week for corn and soybeans. The RMSPE of all five models for corn are bounded with a maximum level of 8.8% (IG National model) to a minimum of 3.4% (BF model). The average RMSPE for corn is about 5% throughout the growing season. For soybean, the patterns are similar, with RMSPE are in the range of 3.6% to 8.0%, and the overall average RMSPE across the entire forecasting cycle about 6%. We begin by focusing on the forecast errors for mid-growing season from 2000 through 2022. Figures 8 and 9 present the forecast errors through the out-of-sample period for the five yield forecasting models for week 29, approximately the middle of the growing season. The variability of the forecast errors is similar to the mid-season errors at the state-level shown earlier in Figures 4 and 5. In general, it appears that the errors for the five forecasting models are highly correlated through time.

We also compare yield forecasts provided by the naïve tend yield model to the five forecasting models in Figures 10 and 11. Figure 10 shows the not too surprising result in corn that the individual models substantially outperform the naïve trend model except for the first few weeks of the growing season. It is interesting to note that both the BM and IG National with Bias Adjustment outperform the naïve trend for every week. Overall, these results indicate that crop condition ratings provide useful information to project corn yield early in the growing season.

Table 1. RMSPE of Weekly Yield Forecasting Models for Corn and Soybean at the National Level, 2000 – 2022

Date	BM Model	IG State Model	IG National Model	IG National with Bias Adjustment Model	BF Model
Panel A: Corn					
June 03	7.6	8.3	8.8	7.6	8.5
June 10	7.4	7.8	8.4	7.2	8.1
June 17	6.9	7.3	7.7	6.4	7.5
June 24	6.1	6.5	6.9	5.6	6.5
July 01	5.7	5.6	6.1	5.1	5.6
July 08	5.0	4.8	5.1	4.5	4.6
July 15	4.4	4.3	4.5	4.3	4.1
July 22	4.2	4.0	4.4	4.3	3.9
July 29	4.2	4.0	4.2	4.2	3.8
August 05	4.2	3.9	4.2	4.2	3.9
August 12	4.1	3.9	4.1	4.1	3.8
August 19	4.0	3.8	3.9	3.9	3.6
August 26	4.2	3.9	4.1	4.1	3.9
September 02	4.1	4.1	4.2	4.2	4.0
September 09	4.0	4.1	4.2	4.2	3.9
September 16	3.8	3.9	4.0	4.0	3.7
September 23	3.7	3.8	3.8	3.8	3.4
Panel B: Soybean					
June 17	6.5	6.9	7.2	6.4	8.0
June 24	6.6	7.0	7.2	6.5	7.6
July 01	6.8	7.0	7.1	6.6	7.5
July 08	6.7	6.7	7.0	6.7	7.5
July 15	6.5	6.6	7.0	6.7	7.4
July 22	6.3	6.4	6.7	6.4	7.3
July 29	5.7	5.4	6.5	6.5	7.3
August 05	4.8	4.3	6.5	6.5	7.1
August 12	4.2	3.8	5.8	5.8	7.0
August 19	4.1	3.6	4.9	4.9	6.8
August 26	4.0	3.5	4.5	4.5	6.0
September 02	3.8	3.3	4.3	4.3	5.1
September 09	3.6	3.1	4.2	4.2	4.5
September 16	3.7	3.1	4.0	4.0	4.2
September 23	3.0	3.5	3.8	3.8	4.2

Notes: For each week, there are 22 observations in the out-of-sample period from 2000 – 2022. The RMSPE measures the average forecast errors over the out-of-sample period, and it is measured in percentage. BM model is proposed by Begueria and Maneta (2020), IG State model, IG National model, IG National with Bias Adjustment Model are proposed by Irwin and Good (2017a), and BF model is proposed by Bain and Fortenbery (2017).

Figure 11 reveals a very different picture for soybeans. None of the individual models consistently beats the naïve trend model until early August. The BF model does not do so in soybeans until mid-July. These results likely reflect the fact that the critical growing period for soybeans occurs later in the summer than for corn. Recent research (Irwin, 2023b) shows that August weather is most important for determining soybean yields, whereas July weather is most important for corn (Irwin, 2023a).

Single-Horizon Forecast Tests

For each week we conduct a pairwise comparison between the benchmark BM model and the other four models. We apply the modified Diebold-Mariano (MDM) test for each week to test if the BM model provides more accurate yield forecasts than an alternative model at a given week during the growing season. The MDM test was developed by Harvey, Leybourne, and Newbold (1997), and has been shown to work well in small samples. Furthermore, as the forecasting horizon increases the test is over-sized and remains stable. For each week, there are 21 observations as the out-of-sample period covers 2000 through 2022.

The null hypothesis is that two models have the same predictive accuracy. The MDM test determines if the difference in RMSPE between the BM model and other models is significant. If we assume the loss function to be quadratic, we have:

(11) $$d_{w,t} = \left(e_{w,t}^2\right)^2 - \left(e_{w,t}^1\right)^2$$

(12) $$E\left(d_{w,t}\right) = 0,$$

where $e_{w,t}^1$ represents the yield errors from BM model, and $e_{w,t}^2$ represents the yield errors from one of its competing models.

For the h-step ahead yield forecasts, the MDM statistic is defined as:

(13) $$MDM = \left[\frac{n+1-2h+n^{-1}h(h-1)}{n}\right]^{\frac{1}{2}} \cdot \bar{d}_w \cdot \left[V(\bar{d}_w)\right]^{-\frac{1}{2}}$$

(14) $$V\left(\bar{d}_w\right) = \left[n^{-1}\left(\gamma_0 + 2\sum_{s=1}^{h-1}\gamma_s\right)\right],$$

where $\bar{d}_w$ is the sample mean of $d_{w,t}$, w is the forecast week and $w = 1,2,3,\dots,17$ for corn and $w = 1,2,3,\dots,15$ for soybeans, $\gamma_0 = n^{-1}\sum_{t=1}^{n}(d_{w,t} - \bar{d}_w)^2$ as the variance of $d_{w,t}$, $\gamma_s = n^{-1}\sum_{t=s+1}^{n}\left(d_{w,t} - \bar{d}_w\right)\left(d_{w,t-s} - \bar{d}_w\right)$, $s = 1,2,3,\dots,h-1$, as the sth auto-covariance of $d_{w,t}$. The weekly forecasts are one-step ahead forecasts (by year), and therefore, $h = 1$. Hence, the MDM statistic for each forecast week is:

(15) $$MDM_w = \left[(n-1)\right]^{\frac{1}{2}} \cdot \bar{d}_w \cdot \left[n^{-1}\left(\sum_{t=1}^{n}\left(d_{t,w} - \bar{d}_w\right)^2\right)\right]^{-\frac{1}{2}}$$

The MDM test statistics for corn and soybeans are shown in Tables 3 and 4, respectively. The null hypothesis is that each week throughout the out-of-sample forecasting period, the forecasting performance of BM model and one of its competing models is the same. Test statistics show that for corn, out of 68 cases of pair-wise yield forecast comparisons for week 23 to week 39, all test statistics are insignificant. These results suggest that we fail to reject the null hypothesis that BM model does not have better forecasting performance than other models. For soybeans, out of 60 cases of pair-wise yield forecast comparisons covering forecast weeks 25 to week 39, there is again no significant case. These results suggest that the benchmark BM model does not significantly outperform its competitors for each week throughout the growing season for both corn and soybeans.

Best Model Confidence Set Tests

Each week, all five yield forecasting models produce weekly yield forecasts for corn and soybeans. In the previous section, we applied the MDM test to conduct a pairwise yield performance test between the BM model and other models. To extend the pairwise comparisons, the Model Confidence Set (MCS) test allows model selection across all yield forecasting models (Hansen, Lunde, and Nason, 2011). For a given significance level α, the MCS test selects the model with best forecasting accuracy from a set of models.

Table 2. Modified Diebold Mariano (MDM) Test Statistics for Corn Yield Forecasting Models at the National Level, 2000 - 2022

Date	BM vs IG State	BM vs IG National	BM vs IG National with Bias Adjustment	BM vs BF
June 03	0.387	0.817	-0.409	0.579
	(0.702)	(0.423)	(0.686)	(0.568)
June 10	0.074	0.561	-0.831	0.341
	(0.942)	(0.580)	(0.415)	(0.737)
June 17	0.008	0.379	-1.360	0.173
	(0.993)	(0.709)	(0.188)	(0.864)
June 24	0.027	0.424	-1.391	0.131
	(0.978)	(0.676)	(0.178)	(0.897)
July 01	-0.457	0.059	-1.449	-0.427
	(0.652)	(0.954)	(0.162)	(0.674)
July 08	-0.576	-0.090	-1.396	-0.692
	(0.570)	(0.929)	(0.177)	(0.496)
July 15	-0.621	0.025	-0.640	-0.761
	(0.541)	(0.981)	(0.529)	(0.455)
July 22	-0.703	0.024	-0.103	-0.793
	(0.489)	(0.981)	(0.919)	(0.436)
July 29	-1.003	-0.278	-0.278	-1.053
	(0.327)	(0.784)	(0.784)	(0.304)
August 05	-1.103	-0.300	-0.300	-0.660
	(0.282)	(0.767)	(0.767)	(0.516)
August 12	-1.559	-0.406	-0.406	-0.826
	(0.133)	(0.689)	(0.689)	(0.418)
August 19	-1.375	-0.840	-0.840	-1.008
	(0.183)	(0.410)	(0.410)	(0.324)
August 26	-1.140	-0.729	-0.729	-0.391
	(0.267)	(0.474)	(0.474)	(0.700)
September 02	-0.138	0.363	0.363	0.093
	(0.891)	(0.720)	(0.720)	(0.927)
September 09	0.140	0.480	0.480	0.026
	(0.890)	(0.636)	(0.636)	(0.980)
September 16	0.482	0.789	0.789	-0.087
	(0.634)	(0.438)	(0.438)	(0.932)
September 23	0.229	0.239	0.239	-0.847
	(0.821)	(0.813)	(0.813)	(0.406)

Notes: This table presents the t-statistics and p-values (in parenthesis) for the MDM test. *, **, *** is the significance level at 10%, 5%, 1% respectively. The null hypothesis is that for each week, each of the four competing forecasting models have the same predictability as the BM model. BM model is proposed by Begueria and Maneta (2020), IG State model, IG National model, IG National with Bias Adjustment Model are proposed by Irwin and Good (2017a), and BF model is proposed by Bain and Fortenbery (2017).

BM model is proposed by Begueria and Maneta (2020), IG State model, IG National model, IG National with Bias Adjustment Model are proposed by Irwin and Good (2017a), and BF model is proposed by Bain and Fortenbery (2017). Colino et al. (2012) showed that an equal-weighted composite provides more accurate forecasts than individual outlook programs for hog prices. Following their approach, we build the Equal Weighted Model that produces composite forecasts which are the arithmetic average of the five individual yield forecasts. We include the composite forecasts in the set of yield forecasting models and apply the MCS test to test whether composite forecasts outperform individual forecasts.

Table 3. Modified Diebold Mariano (MDM) Test Statistics for Soybean Yield Forecasting Models at the National Level, 2000 – 2022

Date	BM vs IG State	BM vs IG National	BM vs IG National with Bias Adjustment	BM vs BF
June 17	0.24	0.445	-0.123	0.605
	(0.812)	(0.661)	(0.903)	(0.551)
June 24	0.094	0.413	-0.12	0.631
	(0.926)	(0.684)	(0.906)	(0.535)
July 01	0.006	0.242	-0.073	0.451
	(0.995)	(0.811)	(0.942)	(0.656)
July 08	0.047	0.136	-0.189	0.396
	(0.963)	(0.893)	(0.852)	(0.696)
July 15	-0.002	0.113	-0.347	0.691
	(0.998)	(0.911)	(0.732)	(0.497)
July 22	-0.322	-0.371	-2.081	0.452
	(0.751)	(0.714)	(0.049)	(0.656)
July 29	-0.214	-0.34	-0.34	0.495
	(0.832)	(0.737)	(0.737)	(0.625)
August 05	-0.23	-0.064	-0.064	0.559
	(0.82)	(0.95)	(0.95)	(0.582)
August 12	-1.462	-0.11	-0.11	0.564
	(0.158)	(0.914)	(0.914)	(0.578)
August 19	-1.677	0.473	0.473	1.202
	(0.108)	(0.641)	(0.641)	(0.242)
August 26	-1.006	0.948	0.948	1.511
	(0.325)	(0.354)	(0.354)	(0.145)
September 02	-0.694	1.24	1.24	0.881
	(0.495)	(0.228)	(0.228)	(0.388)
September 09	-0.65	1.251	1.251	1.224
	(0.522)	(0.224)	(0.224)	(0.234)
September 16	-0.601	1.266	1.266	0.559
	(0.554)	(0.219)	(0.219)	(0.582)
September 23	-0.814	1.219	1.219	0.811
	(0.424)	(0.236)	(0.236)	(0.426)

Notes: This table presents the t-statistics and p-values (in parenthesis) for the MDM test. *, **, *** is the significance level at 10%, 5%, 1% respectively. The null hypothesis is that for each week, each of the four competing forecasting models have the same predictability as the BM model.

The MCS test is built on an iterative procedure at each step, where it eliminates the worst performing model from the set of six models ($\mathcal{M}_0$) until the last model survives from the tests in all previous five steps. To select which model should be eliminated, the following *t*-statistic proposed by Hansen, Lunde, and Nason (2011) are used:

(16) $$t_{i\cdot} = \frac{\overline{d_{i\cdot}}}{\sqrt{\widehat{var}(\overline{d_{i\cdot}})}}, \text{ for } i, j \in \mathcal{M}_0$$

where $\overline{d_{i\cdot}} \equiv m^{-1}\sum_{j \in \mathcal{M}_0} \overline{d_{ij}}$, $\overline{d_{ij}} = n^{-1}\sum_{t=1}^{n} d_{ij,t}$, $d_{ij,t} = L_{i,t} - L_{j,t}$, $L(\cdot)$ is the squared error function. Corresponding *p*-values are collected from a bootstrap of the test statistics. The best model selected by MCS has *p*-value equals to 1. When more than one model has a *p*-value equal to 1, we use the equivalence test, $T_D \equiv \sum_{i \in \mathcal{M}_0} (t_{i\cdot})^2$, to test if the last model outperforms its competitors.

Our study reports the last model selected by the MCS test based on *p*-values produced by 2,000 bootstrap replicates for each week. We first show MCS test results for the set of models only consisting of the five individual yield forecasting models; next we show the MCS test results for the set of models adding the Equal Weight Model to the five individual yield forecasting models. The significance level for MCS test is 10% in order to be conservative in determining the best model. We also report the *p*-values for the equivalence test. When the estimated *p*-value is greater than 0.1, the MCS test indicates that the forecast accuracy of the best selected model is not statistically superior to the competing models in the set.

Weekly MCS test results for corn and soybeans are reported in Table 4 and 5. In Table 4, we report the best model from the set of five individual yield forecasting models. For both corn and soybeans, the best models early in the growing season are the IG National with Bias Adjustment model and the IG State model. By the end of growing season, the BM model and IG State model provide the most accurate yield predictions. In Table 5, we report the MCS test for the set of models that include the Equal Weighted composite model and five individual yield forecasting models. For both crops throughout the growing stages, the best model selected is the Equal Weighted model, except for corn in the week of August 19 (when the selected best model is BF model) and for soybean after the week of August 12 (when IG State model ranked the best). Equivalence test p-values are also reported in Tables 4 and 5. They fail to reject the null hypothesis of that the set of models have equal predictive ability. These findings indicate that there are models with more accurate corn and soybean yield forecasts, however, the differences are not statistically significant.

Multi-Horizon Forecast Tests

One limitation of the MDM test is that it only provides comparisons for two competing models at each horizon w. It is very common to find that at some horizons the first model outperforms the second, and at some other horizons the situation reverses. For two competing models that cover multi-horizons, it is helpful to also perform an omnibus test based on all forecasting horizons. Quaedvlieg (2021) introduced a multi-horizon superior predictive ability (SPA) test that enables the comparison of forecasts of different models jointly, combining the models' predictability across all horizons. The author proposed two tests, the first one is the uniform SPA test that tests if a model has superior forecasting performance at each individual horizon; the second is the average SPA test that determines if a model has superior forecasting performance considering the entire forecasting path. For our study, we follow the Quaedvlieg average SPA (aSPA) test as it is the less restrictive of the two tests.

We denote USDA final yields as y_t, and the weekly yield forecasts produced by model i as $\widehat{y_t^i}$. In a multi-horizon test framework, $\widehat{y_t^i}$ is a 17-dimension vector, $\widehat{y_t^i} = [\widehat{y_{1,t}^i}, \widehat{y_{2,t}^i}, \ldots \widehat{y_{h,t}^i}, \ldots, \widehat{y_{17,t}^i}]$, where h indicates the week of a yield forecast; i represents different choice of forecasting models; t is the year when the fixed-event forecasts are made. We define the loss function as $L_t^i = L(y_t, \widehat{y_t^i})$, and it projects the final yield estimates onto a 17-dimension space. The loss function is defined in a quadratic form, that is the square of the percentage difference between the final yield estimates and each week's yield forecasts provided by model i. Here we use notation "1" to stand for the benchmark BM model, and "2" for its competing model. Then we define the loss differential for the two competing yield forecasts as $d_t = L_t^2 - L_t^1$. D is the loss differential matrix and its dimensions are 21×17. $D = [d_1^T, \ldots, d_t^T, \ldots, d_{21}^T]^T$, where $d_t = [d_t^1, \ldots, d_t^h, \ldots, d_t^{17}]$. Each entry of the matrix D is denoted as d_t^h, and D is specified as:

(17)
$$D = \begin{bmatrix} d_1^1 & \ldots & d_1^h & \ldots & d_1^{17} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ d_t^1 & \ldots & d_t^h & \ldots & d_t^{17} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ d_{21}^1 & \ldots & d_{21}^h & \ldots & d_{21}^{17} \end{bmatrix}_{21\times 17}$$

We use the mean loss differential, $\mu^{aSPA} = \sum_{h=1}^{17} w_h \mu_h$, to compare overall predictability. μ^{aSPA} can be taken as the weighted sum of each week's average differentials, where w_h is the

Table 4. Best Individual Model Selected by the MCS Test for Corn and Soybeans at the National Level, 2000 – 2022

Panel A: Corn			Panel B: Soybean		
Date	Best Model with MCS Test	MCS p-values	Date	Best Model with MCS Test	MCS p-values
June 03	IG National with Bias Adjustment	0.175			
June 10	IG National with Bias Adjustment	0.261			
June 17	IG National with Bias Adjustment	0.284	June 17	IG National with Bias Adjustment	0.509
June 24	IG National with Bias Adjustment	0.298	June 24	IG State	0.433
July 01	IG National with Bias Adjustment	0.499	July 01	IG State	0.65
July 08	IG National with Bias Adjustment	0.638	July 08	IG National with Bias Adjustment	0.61
July 15	BF	0.664	July 15	IG National with Bias Adjustment	0.361
July 22	IG State	0.26	July 22	IG National	0.369
July 29	BF	0.59	July 29	IG National	0.321
August 05	IG State	0.585	August 05	IG State	0.316
August 12	IG State	0.221	August 12	IG State	0.273
August 19	BF	0.752	August 19	IG State	0.161
August 26	IG State	0.236	August 26	IG State	0.193
September 02	IG State	0.917	September 02	IG State	0.198
September 09	BM	0.883	September 09	IG State	0.756
September 16	BM	0.696	September 16	IG State	0.103
September 23	BF	0.758	September 23	IG State	0.518

Notes: MCS p-values are all greater than the 10% significance level, suggesting the selected best performing model fails to significantly outperform other individual yield forecasting models. The best model selected by MCS test is based on a significance level of 10%, with p-values are produced with 2000 bootstrap replicates for the test statistics.

Table 5. The Best Individual and Composite Model Selected by the MCS Test for Corn and Soybeans at the National Level, 2000 – 2022

Panel A: Corn			Panel B: Soybean		
Date	Best Model with MCS Test	MCS p-values	Date	Best Model with MCS Test	MCS p-values
June 03	Equal Weighted model	0.112			
June 10	Equal Weighted model	0.209			
June 17	Equal Weighted model	0.2	June 17	Equal Weighted model	0.436
June 24	Equal Weighted model	0.22	June 24	Equal Weighted model	0.373
July 01	Equal Weighted model	0.421	July 01	Equal Weighted model	0.596
July 08	Equal Weighted model	0.58	July 08	Equal Weighted model	0.551
July 15	Equal Weighted model	0.694	July 15	Equal Weighted model	0.331
July 22	Equal Weighted model	0.174	July 22	Equal Weighted model	0.301
July 29	Equal Weighted model	0.603	July 29	Equal Weighted model	0.253
August 05	Equal Weighted model	0.611	August 05	Equal Weighted model	0.273
August 12	Equal Weighted model	0.266	August 12	IG State	0.24
August 19	BF	0.813	August 19	IG State	0.141
August 26	Equal Weighted model	0.2	August 26	IG State	0.179
September 02	Equal Weighted model	0.908	September 02	Equal Weighted model	0.158
September 09	Equal Weighted model	0.877	September 09	IG State	0.759
September 16	Equal Weighted model	0.764	September 16	IG State	0.116
September 23	Equal Weighted model	0.796	September 23	IG State	0.133

Notes: The Equal Weighted model produced yield forecast composites of five yield forecasting models. MCS p-values are all greater than 10% significance level, suggesting the Equal Weighted model fails to outperform individual yield forecasting models. The best model selected by MCS test is based on the 10% significance level, with p-values produced via 2000 bootstrap replicates for the test statistics.

Table 6. Multi-Horizon Average Superior Predictive Ability (aSPA) Test between Yield Forecasting Models for Corn and Soybeans at the National Level with Fixed Weights, 2000 – 2022

Crop	BM vs IG State	BM vs IG National	BM vs IG National with Bias Adjustment	BM vs BF
Corn	-0.194	0.278	-0.724	-0.224
	(0.526)	(0.449)	(0.617)	(0.545)
Soybeans	-0.471	0.205	-0.206	0.558
	(0.697)	(0.491)	(0.581)	(0.446)

Notes: This table presents the t-statistics and p-values (in parenthesis) for the multi-horizon aSPA test. *, **, *** is the significance level at 10%, 5%, 1% respectively. The null hypothesis is that considering all horizons, on average, the competing yield forecasting model has better performance than BM model. BM model is proposed by Begueria and Maneta (2020), IG State model, IG National model, IG National with Bias Adjustment Model are proposed by Irwin and Good (2017a), and BF model is proposed by Bain and Fortenbery (2017).

weights for each forecast week; $\mu_h = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} d_t^h$ is the mean of each week's loss differentials and we use $\bar{d}_h = \frac{1}{21}\sum_{t=1}^{21} \widehat{d_t^h}$ to estimate μ_h. The null hypothesis of the aSPA test is $\mu^{aSPA} \leq 0$, which implies that, on average, the benchmark BM model fails to provide better performance than competitors across all forecast horizons. The studentized statistic for aSPA test is:

$$t_{aSPA} = \frac{\sqrt{T}\ \sum_{h=1}^{17} w_h \cdot \bar{d}_h}{\hat{\varsigma}}, \tag{18}$$

where $\hat{\varsigma} = \sqrt{w'\widehat{\Omega}w}$; $w = [w_1, \dots ,\ w_h, \dots, w_{17}]^T$ is the 17-dimentional weight vector. Ω is the variance-covariance matrix of matrix D. We denote $D = [D_t^1, \dots, D_t^h, \dots, D_t^{17}]$, where $D_t^h = [d_1^h, \dots, d_{21}^h]^T$. The variance-covariance matrix Ω of matrix D is defined as:

$$\Omega = \begin{bmatrix} var(D_t^1) & \dots & cov(D_t^1,\ D_t^{17}) \\ \vdots & \vdots & \vdots \\ cov(D_t^1,\ D_t^{17}) & \dots & var(D_t^{17}) \end{bmatrix}_{17\times 17}, \tag{19}$$

where $var(D_t^1) = \frac{1}{T}(D_t^1)^T(D_t^1) = \frac{1}{T}\sum_{t=1}^{21}(d_t^1)^2$; $cov(D_t^1,\ D_t^{17}) = \frac{1}{T}(D_t^1)^T(D_t^{17}) = \frac{1}{T}\sum_{t=1}^{21} d_t^1 \cdot d_t^{17}$. Since each week's differentials are highly correlated, we use the Newey-West HAC estimator to find its estimator, $\widehat{\Omega}$. The choice of weights is flexible. We follow the examples proposed by Quaedvlieg (2021): first, we select an equal weight where $w^h = \frac{1}{17}$ for each week; second, we use "efficient" weights to minimize ς as the yield forecasts during the growing season are based on accumulated information. We assign small weights to early forecasts where variance is high, and we assign large weights to near end-of-season forecasts where variance is low. Therefore, the inverse-variance weights are defined as $w^h = \frac{1}{\sigma_h^2(\sum_{i=1}^{17}\sigma_i^2)}$ and they satisfy the condition that the sum of weights is equal to 1. To obtain the critical values and p-values, we use a moving block bootstrap (MBB) technique to simulate the distribution. We focus on the significance level at 5%, and the significance level is the corresponding percentile of the bootstrap distribution.

The null hypothesis of the aSPA test is that the benchmark BM model is no more accurate than a competing model across all forecast horizons. Test results are summarized in Table 7. The

Table 7. Multi-Horizon Average Superior Predictive Ability (aSPA) Test between Yield Forecasting Models for Corn and Soybeans at the National Level with Varying Weights, 2000 – 2022

Crop	BM vs IG State	BM vs IG National	BM vs IG National with Bias Adjustment	BM vs BF
Corn	-0.816	-0.076	-0.398	-0.913
	(0.670)	(0.525)	(0.560)	(0.722)
Soybeans	-1.418	1.963	1.716	1.576
	(0.709)	(0.188)	(0.207)	(0.305)

Notes: This table presents the t-statistics and p-values (in parenthesis) for the multi-horizon aSPA test. *, **, *** is the significance level at 10%, 5%, 1% respectively. The null hypothesis is that considering all horizons, on average, the competing yield forecasting model has better performance than BM model. BM model is proposed by Begueria and Maneta (2020), IG State model, IG National model, IG National with Bias Adjustment Model are proposed by Irwin and Good (2017a), and BF model is proposed by Bain and Fortenbery (2017).

multi-horizon aSPA test *p*-values are all greater than 5%, suggesting the BM model fails to significantly outperform the other models considering all forecast horizons during the corn and soybean growing seasons. We also conduct the average SPA test with varying weights for each week of the growing season, and these test results are summarized in Table 8. Once again, the multi-horizon aSPA test *p*-values are all greater than 5%, suggesting the BM model fails to significantly outperform the other models considering all forecast horizons. Overall, the findings based on the aSPA tests are consistent with what we found with the single horizon MDM test. That is, the benchmark BM model fails to systematically outperforms competing models during the growing season for corn and soybeans. A plausible argument for this finding is that the BM model only controls for the time and spatial variations in the state-level crop condition ratings, so the transformed weekly CCI does not contain any more information relevant to forecasting corn and soybean yields than simple approaches that make similar adjustments.

Conclusions

Crop condition ratings provide unique information for predicting crop yields. The ratings are widely used by market analysts in the public and private sectors to forecast crop yields. In this study, we compare the accuracy of relatively simple single equation condition models (Irwin and Good, 2017a, b; Bain and Fortenbery, 2017) to a sophisticated and computationally demanding specification (Begueria and Maneta, 2020) in forecasting U.S. corn and soybean yields. The data for the study consists of weekly state and national crop condition ratings from the USDA over 1986 through 2022. To evaluate the predictability of all yield forecasting models, we use data from 2000 through 2022 as the out-of-sample period.

A battery of statistical tests is applied to the out-of-sample crop yield forecasts. The modified Diebold-Mariano test is used to conduct a weekly pair-wise comparison between models. Test results suggest that no model has statistically superior forecast accuracy. We also apply Model Confidence Set tests to select the best individual yield forecasting models. Moreover, we add composite forecasts as the arithmetic average of the five individual yield forecasts to the set of models. Test results for individual yield forecasting models suggest that early in the growing season the Irwin and Good model with bias adjustment is the best, and by the end of growing season the Bain and Fortenbery Begueria and Maneta models are selected as having the best yield forecasting performance. When we include equal-weighted composite forecasts, test results show composite forecasts provide the most accurate yield predictions. However, test statistics indicate all the best models fail to significantly outperform their competitors. Lastly, we apply the multi-horizon average Superior Predictive Ability (aSPA) test developed by Quaedvlieg (2021) to compare models across the entire growing season. Again, test results indicate no statistically significant difference in the accuracy of yield forecasts for simpler versus more complex models. In sum, the results of this study are consistent with the conventional wisdom in the forecasting literature that complex models generally do not outperform simpler models in terms of forecast accuracy. Complexity does not pay when forecasting corn and soybean yields based on crop condition ratings. The simple models widely used in the grain industry are at least as accurate as the most sophisticated model available in the literature.

[First submitted October 2024; accepted for publication May 2025.]

References

Armstrong, J.S. 2001. Principles of Forecasting: A Handbook for Researchers and Practitioners. New York: Springer.

Bain, R., and T.R. Fortenbery. 2017. "Impact of Crop Condition Reports on National and Local Wheat Markets." Journal of Agricultural and Applied Economics 49(1):97-119. DOI: https://doi.org/10.1017/aae.2016.31

Beguería, S., and M.P. Maneta. 2020. "Qualitative Crop Condition Survey Reveals Spatiotemporal Production Patterns and Allows Early Yield Prediction." Proceedings of the National Academy of Sciences 117(31):18317-18323. DOI: https://doi.org/10.1073/pnas.1917774117

Colino, E.V., S.H. Irwin, P. Garcia, and X. Etienne. 2012. "Composite and Outlook Forecast Accuracy." Journal of Agricultural and Resource Economics 37(2): 228-246.

Fackler, P.L., and B. Norwood. 1999. "Forecasting Crop Yields and Condition Indices." Paper presented at NCR-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management.

Hansen, P.R., A. Lunde, and J.M. Nason. 2011. "The Model Confidence Set." Econometrica 79(2): 453-497. DOI: https://doi.org/10.3982/ECTA5771

Harvey, D., S. Leybourne, and P. Newbold.1997. "Testing the Equality of Prediction Mean Squared Errors." International Journal of Forecasting 13(2):281-291. DOI: https://doi.org/10.1016/S0169-2070(96)00719-4

Irwin, S. 2023a. "The Relative Impact of Crop Weather Variables on the U.S. Average Yield of Corn." farmdoc daily (13):184, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, October 9.

Irwin, S. 2023b. "The Relative Impact of Crop Weather Variables on the U.S. Average Yield of Soybeans." farmdoc daily (13):194, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, October 23.

Irwin, S.H., and D.L. Good. 2017a. "When Should We Start Paying Attention to Crop Condition Ratings for Corn and Soybeans?" farmdoc daily (7):96, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, May 24.

Irwin, S.H., and D.L. Good. 2017b. "How Should We Use Within-Season Crop Condition Ratings for Corn and Soybeans?" farmdoc daily (7):101, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, June 1.

Irwin, S.H., and T. Hubbs. 2018a. "What to Make of High Early Season Crop Condition Ratings for Corn?" farmdoc daily (8):108, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, June 13.

Irwin, S.H., and T. Hubbs. 2018b."What to Make of High Early Season Crop Condition Ratings for Soybeans?" farmdoc daily (8):109, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, June 14.

Irwin, S.H., and T. Hubbs. 2018c. "Does the Bias in Early Season Crop Condition Ratings for Corn and Soybeans Vary with the Magnitude of the Ratings?" farmdoc daily (8):113, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, June 20.
Irwin, S.H., and T. Hubbs. 2018d. "Measuring the Accuracy of Forecasting Corn and Soybean Yield with Good and Excellent Crop Condition Ratings." farmdoc daily (8):118, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, June 27.
Jorgensen, N., and M. Diersen. 2014. "Forecasting Corn and Soybean Yields with Crop Conditions." Economics Commentator 547, South Dakota State University.
Kruse, J., and D. Smith. 1994."Yield Estimation Throughout the Growing Season." Center for Agricultural and Rural Development, Iowa State University.
Lehecka, G. V. 2014. "The Value of USDA Crop Progress and Condition Information: Reactions of Corn and Soybean Futures Markets." Journal of Agricultural and Resource Economics 39(1):88-105. DOI: 10.22004/ag.econ.168261
Tannura, M.A., Irwin, S.H., and Good, D.L. 2008. "Weather, Technology, and Corn and Soybean Yields in the U.S. Corn Belt." Marketing and Outlook Research Report 2008-01, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign.
Quaedvlieg, R. 2021. "Multi-Horizon Forecast Comparison." Journal of Business & Economic Statistics 39(1):40-53. DOI: https://doi.org/10.1080/07350015.2019.1620074