NCCC-134
APPLIED COMMODITY PRICE ANALYSIS, FORECASTING AND MARKET RISK MANAGEMENT

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

by

Jiarui Li, Scott H. Irwin, Todd Hubbs

Suggested citation format:
Li, J., S. H. Irwin, and T. Hubbs. 2023. "Does Complexity Pay? Forecasting Corn and Soybean Yields Using Crop Condition Ratings." Proceedings of the NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management. [http://www.farmdoc.illinois.edu/nccc134].

DOES COMPLEXITY PAY? FORECASTING CORN AND SOYBEAN YIELDS USING CROP CONDITION RATINGS

Jiarui Li*, Scott H. Irwin, Todd Hubbs

Jiarui Li was a PhD student at Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign; Scott H. Irwin is the Lawrence J. Norton Chair of Agricultural Marketing, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign; Todd Hubbs is a Cross Commodity Analyst at USDA. *Correspondence may be sent to: E-mail: jli180@illinois.edu.

1. Introduction

Accurate forecasts of crop yield are highly valuable from several perspectives. 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 (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 producer management techniques.

Kaufmann and Snell (1997) and others observe that the various methods of forecasting crop yields can be categorized into three groups. The first group consists of crop simulation models that directly assess the effects of weather and soil properties on plant physiology. While such models have a strong foundation in biological theory and experimental data, they are nonetheless highly complex and difficult to generalize to aggregate areas such as crop reporting districts or states (e.g., Walker 1989). The second group consists of multiple regression models that estimate the relationship of weather and technology to crop yields. Regression models are relatively simple to specify and estimate for aggregate areas, an advantage when forecasting, but aggregation of variables across time and space can harm accuracy (Shaw, 1964). The third group consists of models based on remote-sensing data collected from earth-orbiting satellites. By far the most popular approach is to convert remote-sensing data into a vegetative index (NDVI) and correlate this to yield (e.g., Sakamoto, Gitelson, and Arkebauer, 2013). While there has been a great deal of work along these lines, the advantage in forecasting crop yields has not been proven convincingly to date.

There is still another approach to crop yield forecasting that is widely used by market analysts in both the private and public sectors. The U.S. Department of Agriculture (USDA) publishes weekly condition ratings for important crops during the growing season. The condition ratings reflect the subjective judgment of nearly 4,000 observers about crop yield prospects and are reported as the percentage of a crop rated in five mutually exclusive and exhaustive categories: very poor, poor, fair, good, and excellent. A popular approach is to 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).

Despite the widespread use of crop condition ratings to forecast crop yields in the 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 its yield weights, they computed an average in-sample yield estimate at state-level. Fackler and Norwood (1999) built a similar state-level yield forecasting model for corn, cotton, soybean, and spring wheat with estimated yield weight that is unchanging throughout the growing season for each crop condition level. They showed that for each condition class, the product of estimated yield weight and condition ratings 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 0, and as the condition increase by one level, the corresponding weight will increase by 0.25 until it reaches the highest excellent condition with a weight of 1.

Most recently, Begueria and Maneta (2020) developed a sophisticated two-stage yield forecasting model based on crop condition ratings for corn, cotton, soybean, and winter wheat at the state level. They argue 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 argue 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 (BM) purport that their modeling approach achieves large improvements in accuracy over simpler condition-based forecasts, such as Irwin and Good (2017a,b).

The improvements in forecast accuracy reported by BM 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 forecasting. 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 provided by BM may represent an important exception to this general result. Second, the forecast results in Begueria and Maneta model (BM model, hereinafter) 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 as well. This approach is quite different from the recursive out-of-sample procedures that are standard in the time-series forecasting literature. Third, BM 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 conduct a forecast competition between BM model and simpler crop condition models in forecasting U.S. average corn and soybean yields. Specifically, we compare the forecast accuracy of BM model to Irwin and Good (2017a,b) and Bain and Fortenbery (2017) models. The data for the study consists of weekly state and national crop condition ratings from 1986 through 2022 for corn and soybean. To evaluate the predictability of all yield forecasting models, we use data from 2000 through 2022 as the out-of-sample period. We first recursively estimate all yield forecasting models and provide true out-of-sample yield forecasts. Next, we apply the modified Diebold-Mariano test to conduct a weekly pair-wise comparison between BM model and its competitors. Test results suggest that BM model does not have a systematic superior predictability than other more straightforward yield forecasting models. We also apply the Model Confidence Set tests to select the best individual yield forecasting models. Moreover, we add the composite forecasts as the arithmetic average of the five individual yield forecasts in the set of models. Test results for individual yield forecasting models suggest early in the growing season, Irwin and Good Bias Adjustment model is the best model, and by the end of growing season model BM model and Bain and Fortenbery model are the two models selected out with the best yield forecasting performance. When we include composite forecasts from Equal Weighted model, test results show composite forecasts provide the most accurate yield predictions. However, test statistics are not significant, indicating all best models fail to outperform their competitors. Last, we apply the multi-horizon average Super Predictive Ability (aSPA) test developed by Quaedvlieg (2021) to compare BM model with its four competitors across the entire growing season. Again, test results indicate that BM model fails to provide more accurate yield forecasts than the competing and simpler forecasting models.

2. Data

2.1 Crop condition ratings

From roughly late April until 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 (NASS, 2021; Irwin and Good, 2017a).

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 reactions have 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 (NASS, 2021):

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.

We collected all weekly condition ratings for corn and soybean 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 have a consistent evaluation period for all competing models, we use weeks 23 – 39 for corn and weeks 25 – 39 for soybean to evaluate the weekly yield forecasts. The ranges of these weeks 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.

2.2 Harvested acres

BM 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 modelling 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 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 five-year moving-average procedures are 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 1976 – 2022. The harvested acres data are obtained from NASS Quick Stats, and they are published in the Acreage reports released by NASS each year by the end of June. The Acreage report produces the 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.

2.3 Annual yield estimates

All weekly yield forecasts are compared to the final yield estimates published at the NASS Quick Stats website. As only one five-year moving-average procedure is applied to yields, we collected yield data from 1981 through 2022.

3. Yield Forecasting Models

3.1 Yield forecasting cycles

The goal of all yield forecasting models in this study is to 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 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 – week 39 in each year and for soybean for week 25 – 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.

3.2 Begueria and Maneta model

Begueria and Maneta (BM) model (2021) is the most technically sophisticated model considered in this forecast competition. They argue that spatial and temporal differences in crop condition information should be directly modeled before making yield forecasts. Hence, BM developed a cumulative link mixed model (CLMM) 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 formal terms, the first step of BM model is to estimate the CLMM using a probit link function to connect ordinal response with numeric factors. The CLMM is specified as:

$$probit\big(P(Y_i \le j|s,y,w)\big) = \theta_j + \beta_y y + \beta_w w + v_s + v_{y,s} y + v_{w,s} w + \epsilon_{si}, \tag{1}$$

where $probit\big(P(Y_i \le j|s,y,w)\big)$ 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 their modeling process, BM develop a mixed model, where the fixed effects are the long-term (year) effects and CCI effects and the random effect is conditional on each state including the intercept two slopes with the interactions from year and CCI. This model provides weekly yield forecasts for each state and is specified as:

$$\mu_i(s) = \beta_0 + \beta_y y_i + \beta_c CCI_i + v(s) + v_y(s) y_i + v_c(s) CCI_i + \epsilon_i \tag{2}$$

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 state as the random components, meaning for different states, they have different temporal effects and CCI effects.

Figure 2 uses corn as an example to illustrate how BM model recursively provide out-of-sample weekly yield forecasts. Yield forecasts for week 23 in 2000 were estimated with the following steps: i) 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 till week 23, 2000. Second, we can estimate the mixed model using the updated CCI and other variables in week 23 from 1986 to 1999. Third, updated CCI and year index of week 23, 2000 were entered in the mixed model and we can obtain a yield projection for week 23, 2000. Following these steps, as we move forward in the growing season, we can have weekly updates of yield forecasts.

We also present BM yield forecasts for corn and soybean at the national level. We calculate the weighted average yields of 18 states where the weight for each state is the proportion of harvested acres. To transform the weighted average yields of 18 states to the national yields, we apply the ratio of weighted average yields of 18 states to the national yields. Yield forecasts at national level are available from 2000 to 2022.

3.3 Irwin and Good model

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

$$Yield_t = \beta_0 + \beta_1 year\_index_t + \beta_2 SUM_t + \epsilon_t \tag{3}$$

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. Yields forecasts for 2000 week 23 are obtained with the following steps. First, we run IG National model with time index, the percentage of corn rated in "good" and "excellent" conditions at the end of years, and the national final yield estimates from 1986 to 1999. Second, the sum of ratings in week 23 and the year index for 2000 are entered in the model to get the yield forecasts for week 23, 2000.

State-level yield forecasts follow the same procedure of national level. Instead of using national yield estimates, they use state-level final yield estimates to build Irwin and Good State model (IG State model, hereinafter) and eventually receive weekly yield forecasts for each state. For individual state, we can also compare the predictability of BM model and IG State model to examine if there is a significant trade-off between model complexity and forecast accuracy.

Irwin and Good (2017b) pointed out that the disadvantage of this straightforward approach is that it does not consider the bias in the early weeks' condition ratings within the growing season. Their weekly analysis (Irwin and Good, 2017a) showed that in early weeks the correlations coefficients between the sum of "good" and "excellent" ratings and the yields are lower than that of final weeks. The reason behind this observation is that, on average, the early weeks' ratings for corn and soybean are over-estimated. Early in the growing season, crops usually are in a normal or a better than normal condition. However, for a few years (like drought in 2012), when adverse weather conditions occur, crop yields would deteriorate and become worse than normal. This would make crop ratings of "good" and "excellent" in the final week lower than that in early weeks, which makes the final week's ratings lower than the average ratings in early weeks. To measure the size of bias, we follow definition of bias proposed by Irwin and Hubbs (2018a,c) and specify the bias as:

$$bias_t = final\ week\ rating_t - early\ week\ rating_t, \tag{4}$$

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. Because on average final ratings are lower than that of early weeks' ratings, we expect the bias to be negative. To adjust the bias in the early weeks, we need to add the bias to the early weeks' ratings as:

$$adj\_early\_rating_t = early\ week\ rating_t + bias_t. \tag{5}$$

For both corn and soybean, we consider weeks before 31 as the early weeks that need bias correction. Therefore, these weeks are week 23 – week 30 for corn and week 25 – week 30 for soybean.

We apply the 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, as we have the average bias for each week of the early weeks, for the current year, we can add the bias to the reported ratings to have the adjusted ratings for a week that is in the range of early weeks. For some weeks, we do not have consecutive observations in all each year. In these scenarios, we use all the data we have from the previous ten or five years, but we might not have all the ten or five data points to calculate the average biases. These two approaches are considered as two augmented bias-adjusted Irwin and Good models (IG National with Bias Adjustment model, hereinafter)[1].

[1] Model comparisons between BM model and IG National with Bias Adjustment model with five-year moving average approach are available in Online Appendix.

3.4 Bain and Fortenbery model

Bain and Fortenbery (2017) fixed weight model (BF model, hereinafter) assigned fixed weights to each condition category to transform the ordinal condition ratings to a numerous crop condition index (CCI). Below is the definition of fixed weights CCI:

$$CCIndex = 100\% \cdot Excellent + 75\% \cdot Good + 50\% \cdot Fair$$

$$+25\% \cdot Poor + 0 \cdot Very\ Poor \quad (6)$$

The ratings for each condition category are in percentages, therefore fixed weights CCI is bounded between 0 and 1. Bain and Fortenbery built the weekly crop yield forecasting model by having the end of season $CCIndex$ in the framework, and the model is specified as:

$$Yield_i = \alpha_0 + \alpha_1 \cdot Trend_i + \beta_1 \cdot CCIndex_i + e_i \quad (7)$$

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, 2000 for corn are estimated with the following steps. First, we transform crop conditions of the end of growing season to the fixed weight $CCIndex$ from 1986 through 1999. Second, we run the model with annual final yield estimates as the response variable and year index and the fixed weight $CCIndex$ as explanatory variables. Third, once we obtain the crop condition ratings for week 23, 2000, we transform them to the fixed weight $CCIndex$ and enter them in the model with updated year index for 2000 to have the yield forecasts.

3.4. Model Comparison and Forecast Evaluation

There are two sets of comparisons conducted by our study. First, we compare all five yield forecasting models with naïve trend yield model to evaluate the value of crop condition index as a yield indicator. Second, we set BM model as a benchmark to compare it with other four simpler yield forecasting models. The comparisons are conducted both at state level and national level, by focusing on the forecast errors for week 29 of mid-July, over the out-of-sample period from 2000 through 2022, and the root mean squared percentage error (RMSPE) of each yield forecasting model over the out-of-sample period throughout the entire growing season.

We use the absolute value of the difference between final yield estimates and the yield forecasts for week 29 to measure the forecast error in week 29. The weekly forecast errors $e_{w,t}^{i}$ for model i are defined as the percentage difference between the USDA final yields and this model's yield forecasts:

$$e_{w,t}^{i} = 100 \cdot \frac{(y_t - \widehat{y_{w,t}^{i}})}{y_t} \tag{8}$$

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

$$RMSPE_{w,t}^{i} = \sqrt{\frac{1}{n}\sum\left(\frac{(y_t - \widehat{y_{w,t}^{i}})}{y_t}\right)^2} \tag{9}$$

where n is the number of observations for each week over the out-of-sample period. One advantage of RMSPE error is that it transforms the error to the positive percentage value, so it avoids offsetting positive and negative errors, and we only need to consider one direction of the error. The other advantage is that RMSPE makes the errors comparable for corn and soybean.

4.1 Naïve trend yield model

One of the key factors that determines crop yields is the technology development over the years. Crops tend to increase their yields year by year, which is known as the "trend yield" (Irwin, Good, and Tannura, 2009). Naïve trend yield model serves as the base model that we use to compare with five yield forecasting models. This is because naïve trend yield model only includes time to account for the variations in yields over time, and yet yield forecasting models include time and crop condition ratings to explain the development of yields. The comparisons provide clear evaluations of whether additional crop condition ratings contain valuable yield information as naïve model only includes year index to account for the variations in national yields. Naïve trend yield model is specified as below:

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

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.

The yield forecasts provided by naïve trend yield model also follow the recursively 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 can make yield predictions using the updated year index of 15 for all weeks during the growing season for corn and soybean.

4.2 Comparison at state level

Both BM model and IG State model provide the state-level yield forecasts. To compare which model systematically provide better yield forecasts, we compare: (i) the absolute value of yield forecast errors for mid-growing season, that is approximately week 29 for both corn and soybean over the out-of-sample period; (ii) weekly RMSPE over the out-of-sample period for each week during the growing season. We conduct the comparisons for two representative states given their geographic difference: Illinois and South Dakota.

Figure 4 and Figure 5 present the percentage difference between forecasted yields provided by BM model and IG State model for week 29 over the out-of-sample period for corn and soybean, respectively. Figure 4 (a) and Figure 5 (a) show that in Illinois, there is no clear pattern of which model outperforms over time. Figure 4 (b) and Figure 5 (b) show a similar pattern in South Dakota.

For year 2012 when crop productions are largely impacted by droughts, we observe that for Illinois, BM model provided more accurate yield forecasts than IG State model in the mid-growing season, whereas for South Dakota, it shows IG State model is more accurate.

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

4.3 Comparison at national level

All yield forecasting models been discussed in this study provide national-level yield forecasts for each week during the growing season over the out-of-sample period. First, we focus on the forecast errors for mid-growing season from 2000 – 2022. Figure 8 and Figure 9 presents the forecast error between national yield forecasts provided by five yield forecasting models and the final USDA yield estimates for corn and soybean, respectively. We also present the yield forecasts provided by naïve tend yield model. It suggests that from 2000 – 2022, all forecasting models provide more accurate yield forecasts than naïve trend yield model, which shows the value of crop condition ratings for yield forecasts. To compare the forecast errors of each forecasting model, we observe that there is no clear pattern to show which model has the superior forecasting performance.

Table 1 summarizes the RMSPE of five forecasting models for each week during the growing season for corn and soybean. RMSPE of all five models for corn are bounded with a maximum level of 8.8% from IG National model, indicating that for week 23, the yield forecasts provided by IG National model are within 8.8% of the final average yields. A minimum level of 3.4% from BF model for week 39, indicating the yield forecasts provided by Bain and Fortenbery model are within 3.4% of the final yields estimates. The average of RMSPE for corn is about 5% throughout the growing season. For soybean, the pattern is similar. RMSPE are in the range of (3.6%, 8.0%), and the overall average RMSPE across the whole forecasting path is about 6%.

Figure 10 shows for corn, all five forecasting models provide more accurate yield forecasts than naïve trend yield model since week 24, about early-June till the end of growing season. Figure 11 shows for soybean, all five forecasting models show the forecasting advantage since early-August. Both plots present the pattern of the yield forecasts provided by BM model and its four competitors: near the end of growing season, all models provide the most accurate yield forecasts; and later in the growing season, there is no forecasting improvements. This pattern indicates by the mid-August, yield forecasting models apply the crop condition ratings reach the limits as human observations cannot fully capture the true underlying information in the fields.

4.4 Single-horizon yield forecasts comparison

For each week we conduct a pairwise comparison between BM model and its four competitors. We apply the modified Diebold-Mariano (MDM) test for each week to test if BM model provides more accurate yield forecasts at single horizon throughout the growing season. The MDM test is developed by Harvey, Leybourne, and Newbold (1997) with the advantage that the MDM test works well for small samples; and with the increase in forecasting horizons, the over-sized MDM test results remain stable. The MDM test is applied on two models' out-of-sample forecast errors. For each week, there are 21 observations as the out-of-sample period covers years from 2000 – 2022.

The null hypothesis is that two models have the same predictive accuracy, and it lies upon the loss function between two models' errors. To be more specific, we test if the difference in RMSPE between BM model and other of its competitors is significant. Here we assume the loss function to be quadratic, and when we fail to reject the null hypothesis, we have:

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

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

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 statistic of MDM is defined as:

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

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

where $\bar{d}_w$ is the sample mean of $d_{w,t}$, w is the forecast week and $w = 1,2,3,\ldots,17$ for corn and $w = 1,2,3,\ldots,15$ for soybean, $\gamma_0 = n^{-1}\sum_{t=1}^{n}(d_{w,t} - \bar{d}_w)^2$ as the variance of $d_{w,t}$, $\gamma_s = n^{-1}\sum_{t=s+1}^{n}\left(d_{w,t} - \bar{d}_w\right)\left(d_{w,t-s} - \bar{d}_w\right)$, $s = 1,2,3,\ldots,h-1$, as the sth auto-covariance of $d_{w,t}$. As each week we make yield predictions for a year ahead, we have one-step ahead forecasts where $h = 1$. Therefore, the MDM statistics for each forecast week is:

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

The MDM test statistics for corn and soybean are summarized in Table 3 and Table 4. 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 have the same predictability. Test statistics show that for corn, out of 68 cases of pair-wise yield forecast comparisons, from 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 less-computation demanding models. For soybean, out of 60 cases of pair-wise yield forecast comparisons, covering forecast weeks from week 25 – week 39, there is no significant cases. These results suggest for soybeans, BM model does not outperform its competitors for each week throughout the growing season.

4.5 Best model selected by the Model Confidence Set (MCS) test

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

Colino et al. (2012) show that equal-weighted composites provide more accurate forecasts than individual outlook programs for hog prices. Following their approaches, we build the Equal Weighted Model that produce 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 we apply the MCS test to test whether composite forecasts outperform individual forecasts.

As MCS is built on the iterative procedures where each step, 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. Each step, to select which model should be eliminated, it is based on the t-statistics proposed by Hansen, Lunde, and Nason (2011):

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

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

Our study reports the last model selected by MCS test based on the p-values produced by 2,000 bootstrap replicates for each week. We first show MCS test results for the set of models only consists of five individual yield forecasting models; next we show the MCS test results for the set of models with Equal Weight Model and five individual yield forecasting models. The significance level for MCS test is 10%. We also report the p-values for the equivalence test. When p-values greater than 0.1, it suggests the best selected model fails to have superior predictability than its competitors.

Weekly MCS test results for corn and soybean are reported in Table 5 and Table 6. In Table 5, we report the best model that survives four steps of model selections for each week. The set of models consist of five individual yield forecasting models. For both corn and soybean, early in the growing season, the best model is IG National with Bias Adjustment, and by the end of growing season, BM model and BF model provide the most accurate yield predictions. In Table 6, we report the MCS test for the set of models including Equal Weighted model and five individual yield forecasting models. For both crops throughout the growing stages, the best model selected is Equal Weighted model that provide equal-weighted composite forecasts of five individual yield forecasts, except for corn in the week of August 19, the selected best model is BF model. For Table 5 and Table 6, each week, the p-value of each step's model elimination is greater than the significance level of 0.1, and there is more than one model has p-value equals to 1. The equivalence test p-values are also reported in Table 5 and Table 6. They fail to reject the null hypothesis of whether these models have equal predictive ability. These findings suggest the selected best models indeed provide more accurate yield forecasts, however, they to significantly outperform their competitors.

4.6 Weekly forecasting errors correlation

During the growing season, crop conditions ratings are published each week. In the estimation process, the data under preparation are compared with data reported in previous week and in surrounding counties. This procedure makes the weekly condition ratings correlated in a year. We want to test if such dependence is available in the forecasting error between the USDA final yields and the yield forecasts produced by one of our selected yield forecasting models. We conduct the correlation test for corn and soybean forecasting errors using BM model as we assume that for other models because the out-of-sample yield forecasts are also produced recursively, they should follow a similar pattern. We run multiple OLS models between the first week and the weeks ahead. Each OLS estimation, the first week is treated as the independent variable, and each week ahead is the dependent variable.

Correlation test results for corn and soybean are summarized in Table 2. We use Heteroskedasticity and Autocorrelation Corrected (HAC) standard errors to produce t-statistic and p-value. Test results show that the first week are correlated significantly to all weeks in the growing season at the 5% significant level. As all competing yield forecasting models have similar patterns of yield forecast errors over the growing season, we can expect this correlation embedded in these forecasting models as well.

4.7 Multi-horizon yield forecasts comparison

One limitation of the MDM test is that it only provides comparison test results 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, such situation reverses. Sometime, only using the single horizon forecast comparison test like the MDM test is likely to conclude contradictory results. For two competing models that cover multi-horizons, it is necessary to perform an omnibus test on all the forecasting horizons. When we argue which model has better forecasting performance, the omnibus test adds more conclusive evidence. Quaedvlieg (2021) introduced the multi-horizon superior predictive ability (SPA) tests that enable the comparison of forecasts of different models jointly, combining these 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 one is the average SPA test that tests if a model has superior forecasting performance considering the whole forecasting path. For our study, we follow Quaedvlieg (2021) average SPA test as we can see in Figure 10 and Figure 11 that there are some cross-over points between BM model and its competing models for corn and soybean, suggesting for some weeks, BM model has lower forecast errors while in some other weeks, its competing models achieve lower forecast errors. Therefore, it is more appropriate to apply the multi-horizon average Super Predictive Ability (aSPA) test (2021) pair-wisely between BM model and the other four forecasting models. This test extends the MDM test and compare two models' yield forecasts across the whole growing season.

Each year USDA final yields are denoted as y_t, and the weekly yield forecasts produced by model i is denoted as $\widehat{y_t^i}$. In multi-horizon test framework, $\widehat{y_t^i}$ is a 17-dimension vector, $\widehat{y_t^i} = [\widehat{y_{1,t}^i}, \widehat{y_{2,t}^i}, \dots \widehat{y_{h,t}^i}, \dots, \widehat{y_{17,t}^i}]$, where h indicate the week that produces the yield forecasts; i represents different choice of forecasting models; t is the year when the fixed-event forecasts happen. We define the loss function as $L_t^i = L\left(y_t, \widehat{y_t^i}\right)$, 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 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 differentials matrix and its dimensions are 21×17. $D = [d_1^T, \dots, d_t^T, \dots, d_{21}^T]^T$, where $d_t = [d_t^1, \dots, d_t^h, \dots, d_t^{17}]$. Each entry of the matrix D is denoted as d_t^h, and D is specified as:

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

We use the mean loss differentials, $\mu^{aSPA} = \sum_{h=1}^{17} w_h \mu_h$, to compare two models' overall predictability. μ^{aSPA} can be taken as the weighted sum of each week's average differentials, where w_h is the 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 $\mu^{aSPA} \leq 0$, meaning considering all horizons, on average, BM model fails to provide better performance than its competitors. 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 $var(D_t^1) = \frac{1}{T}(D_t^1)^T(D_t^1) = \frac{1}{T}\sum_{t=1}^{21}(d_t^1)^2$; $cov(D_t^1,\ D_t^{17}) = \frac{1}{T}(D_t^1)^T(D_t^{17}) = \frac{1}{T}\sum_{t=1}^{21} d_t^1 \cdot d_t^{17}$. For a given year, from our correlation test results, we found each week's differentials are highly correlated. Instead of directly estimating the full variance-covariance matrix Ω, we use the Newey-West HAC estimator to find its estimator, $\widehat{\Omega}$. The choices of weights are flexible. We follow the examples proposed by Quaedvlieg (2021): first, we select the 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 crop growing survey 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 the moving block bootstrap (MBB) technique to draw the distribution. We focus on the significance level at 5%, and the significance level is the corresponding percentile of the bootstrap distribution.

From the single-horizon MDM test, throughout the growing season, for both corn and soybean, the test results produce mixed evidence. To test if BM model has better forecasting performance throughout the whole growing season, we perform multi-horizon average SPA test. First, we assign equal weights to each horizon for the loss differentials. The null hypothesis of the average SPA test is that considering all horizons, on average, simple yield forecasting model has better performance than BM model. Test results are summarized in Table 7. The multi-horizon average SPA test p-values are all greater than 5%, suggesting BM model fails to have significantly better predictability than its four competitors. Second, we conduct the average SPA test with varying weights for each week of the growing season. Test results are summarized in Table 8. The findings with average SPA tests are consistent with what we found with the single horizon MDM test: BM model fails to systematically outperforms its competing models during the growing season for corn and soybean. A plausible argument for this finding is that BM model only controls for the time and spatial variations in the state-level crop condition ratings, so the transformed weekly CCI do not contain more determinate factors to account for the variations in yields than the other yield forecasting models which apply simple approaches to transform the ordinal condition ratings.

5. Conclusions

Crop production forecasts have been an important indicator for price changes in agricultural commodity markets. A small group of studies use the crop conditions data to build the yield forecasting models. Condition ratings are the products of human sensors, and they provide consistent subjective assessment about crops conditions that highly correlated with crops' yields.

This study examined the forecasting accuracy of a batch of yield forecasting models that directly transform the ordinal crop condition ratings to the numeric condition index along with a recently developed model introduced by Begueria and Maneta in 2020 that applies the cumulative link mixed model to transform the condition ratings to the continuous condition index. We conduct the out-of-sample yield forecasts recursively for corn and soybean from 2000 through 2022 for all models. We compared each model's yield forecasts with USDA final yield estimates and we used RMSPE to measure each model's forecasting accuracy. We found all models provide a pattern of the forecasting accuracy: in early weeks of the growing season, RMSPE are relatively higher than that in the final weeks. This pattern is reasonable as we move toward the end of season, crops are about to mature, so their conditions connect to yields more closely. The average RMSPE level for all models throughout the growing season is about 5% for corn and 6% for soybean. Our findings suggest this group of models that use crop conditions data provide accurate yield forecasts.

This study compared the forecasting performance of BM model with its four competing yield forecasting models that have already been widely applied by industry practitioners. One disadvantage of BM model is that it is more complex and computational demanding. Our study evaluates if the BM model provides significantly superior yield forecasts than its competitors so its disadvantage can be compensated. We use the modified Diebold Mariano test for the single-horizon pair-wise forecasts comparisons. Test results suggest for both corn and soybean, BM model fails to outperform its competitors. With Model Confidence Set test, we find among individual yield forecasting models, in early weeks IG with Bias Adjustment is usually the best model, and in final weeks, BM model and BF model are selected as the best models. With composite forecasts in the set of models, Equal Weighted model is selected as the best model. However, all best models fail to significantly outperform their competitors. Furthermore, we conduct the multi-horizon average Superior Predictive Ability test to test whether averaging out the forecasting performance over the growing season, BM model has superior predictability. Test results show that for both corn and soybean, BM model fails to provide significantly more accurate yield forecasts than its competing yield forecasting models.

6. Tables and Figures

Table 1: The RMSPE of weekly yield forecasting models for corn and soybean at national level over 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	6.3	8.0	7.2	6.4	8.0
June 24	6.3	7.9	7.2	6.5	7.6
July 01	6.5	7.9	7.1	6.6	7.5
July 08	6.6	8.0	7.0	6.7	7.5
July 15	6.8	7.9	7.0	6.7	7.4
July 22	6.7	7.8	6.7	6.4	7.3
July 29	6.5	7.7	6.5	6.5	7.3
August 05	6.3	7.4	6.5	6.5	7.1
August 12	5.7	6.5	5.8	5.8	7.0
August 19	4.8	5.5	4.9	4.9	6.8
August 26	4.2	5.1	4.5	4.5	6.0
September 02	4.1	5.0	4.3	4.3	5.1
September 09	4.0	5.1	4.2	4.2	4.5
September 16	3.8	5.0	4.0	4.0	4.2
September 23	3.6	4.9	3.8	3.8	4.2

Notes: For each week, there are 21 observations in the out-of-sample period from 2000 – 2021. 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).

Table 2: Correlation test results for BM model for corn and soybean over 2000 – 2022

Panel A: Corn, independent variable: forecasting error week 23 (June 03)				Panel B: Soybean, independent variable: error in week 25 (June 17)			
Dependent Week Date	Coefficient	t-Statistic	p-Value	Dependent Week Date	Coefficient	t-Statistic	p-Value
June 10	**0.970**	68.894	0.000				
June 17	**0.905**	49.597	0.000				
June 24	**0.798**	22.151	0.000	June 24	**1.00**	50.90	0.00
July 01	**0.715**	15.020	0.000	July 01	**1.00**	29.51	0.00
July 08	**0.588**	10.474	0.000	July 08	**1.00**	19.59	0.00
July 15	**0.455**	7.419	0.000	July 15	**0.97**	14.49	0.00
July 22	**0.396**	5.732	0.000	July 22	**0.90**	11.55	0.00
July 29	**0.396**	6.112	0.000	July 29	**0.89**	13.50	0.00
August 05	**0.389**	6.825	0.000	August 05	**0.85**	13.06	0.00
August 12	**0.408**	7.546	0.000	August 12	**0.78**	13.85	0.00
August 19	**0.387**	6.336	0.000	August 19	**0.63**	10.30	0.00
August 26	**0.374**	6.270	0.000	August 26	**0.55**	9.27	0.00
September 02	**0.374**	6.925	0.000	September 02	**0.53**	9.15	0.00
September 09	**0.357**	7.601	0.000	September 09	**0.49**	8.31	0.00
September 16	**0.328**	7.223	0.000	September 16	**0.48**	8.91	0.00
September 23	**0.322**	6.892	0.000	September 23	**0.46**	9.71	0.00

Notes: Correlation test results with the OLS regression where the independent variable is the forecast error in the first week; and the dependent variable is each one of the other weeks in the growing season. Regression function is specified as: $e_{1+i} = \alpha_i + \beta_i e_1 + \sigma_i, i = 1, \dots, 16$ for corn and $i = 1, \dots, 14$ for soybean. HAC estimation is used to correct for autocorrelation and heteroskedasticity. Coefficients in bold indicate they are significantly at the level that is no greater than 5%. BM model is proposed by Begueria and Maneta (2020).

Table 3: The Modified Diebold Mariano (MDM) test statistics between the BM model and other yield forecasting models for Corn

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. 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).

Table 4: The Modified Diebold Mariano (MDM) test statistics between the BM model and other yield forecasting models for soybean

Date	BM vs IG State	BM vs IG National	BM vs IG National with Bias Adjustment	BM vs BF
June 17	0.808	0.444	-0.124	0.604
	(0.428)	(0.662)	(0.902)	(0.552)
June 24	0.795	0.412	-0.122	0.630
	(0.435)	(0.685)	(0.904)	(0.535)
July 01	0.724	0.241	-0.075	0.451
	(0.477)	(0.812)	(0.941)	(0.657)
July 08	0.811	0.135	-0.190	0.396
	(0.426)	(0.894)	(0.851)	(0.696)
July 15	0.886	0.111	-0.351	0.690
	(0.385)	(0.912)	(0.729)	(0.497)
July 22	0.748	-0.371	-2.083	0.452
	(0.462)	(0.714)	(0.049)	(0.656)
July 29	0.801	-0.339	-0.339	0.495
	(0.432)	(0.738)	(0.738)	(0.625)
August 05	0.785	-0.065	-0.065	0.558
	(0.441)	(0.949)	(0.949)	(0.582)
August 12	0.680	-0.112	-0.112	0.563
	(0.504)	(0.912)	(0.912)	(0.579)
August 19	0.629	0.469	0.469	1.198
	(0.536)	(0.644)	(0.644)	(0.244)
August 26	0.630	0.942	0.942	1.505
	(0.535)	(0.356)	(0.356)	(0.147)
September 02	0.665	1.238	1.238	0.877
	(0.513)	(0.229)	(0.229)	(0.390)
September 09	0.715	1.251	1.251	1.224
	(0.482)	(0.224)	(0.224)	(0.234)
September 16	0.759	1.266	1.266	0.559
	(0.456)	(0.219)	(0.219)	(0.582)
September 23	0.718	1.221	1.221	0.813
	(0.480)	(0.235)	(0.235)	(0.425)

Notes: *, **, *** is the significant level at 10%, 5%, 1% respectively. This table presents the t-statistics and p-values (in parenthesis) for the MDM test. 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).

Table 5: Weekly best model selected by MCS test for corn and soybean from 2000 – 2022

Panel A: Corn Date	Best Model with MCS Test	MCS p-values	Panel B: Soybean 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	0.338
June 24	IG National with Bias Adjustment	0.298	June 24	IG National with Bias Adjustment	0.470
July 01	IG National with Bias Adjustment	0.499	July 01	IG National with Bias Adjustment	0.486
July 08	IG National with Bias Adjustment	0.638	July 08	IG National with Bias Adjustment	0.377
July 15	BF	0.664	July 15	IG National	0.441
July 22	IG State	0.260	July 22	IG National	0.474
July 29	BF	0.590	July 29	IG National	0.501
August 05	IG State	0.585	August 05	IG National	0.516
August 12	IG State	0.221	August 12	IG National	0.623
August 19	BF	0.752	August 19	BM	0.653
August 26	IG State	0.236	August 26	BM	0.646
September 02	IG State	0.917	September 02	BM	0.562
September 09	BM	0.883	September 09	BM	0.537
September 16	BM	0.696	September 16	BM	0.516
September 23	BF	0.758	September 23	BM	0.504

Notes: MCS p-values are all greater than the significance level of 0.1, 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 the significance level of 0.1, p-values are produced with 2000 bootstrap replicates for the test statistics.

Table 6: Weekly best model selected by MCS test for corn and soybean from 2000 – 2022

Panel A: Corn Date	Best Model with MCS Test	MCS p-values	Panel B: Soybean 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	0.200	June 17	Equal Weighted model	0.381
June 24	Equal Weighted model	0.220	June 24	Equal Weighted model	0.574
July 01	Equal Weighted model	0.421	July 01	Equal Weighted model	0.538
July 08	Equal Weighted model	0.580	July 08	Equal Weighted model	0.461
July 15	Equal Weighted model	0.694	July 15	Equal Weighted model	0.521
July 22	Equal Weighted model	0.174	July 22	Equal Weighted model	0.556
July 29	Equal Weighted model	0.603	July 29	IG National	0.503
August 05	Equal Weighted model	0.611	August 05	IG National	0.606
August 12	Equal Weighted model	0.266	August 12	Equal Weighted model	0.642
August 19	BF	0.813	August 19	Equal Weighted model	0.675
August 26	Equal Weighted model	0.200	August 26	Equal Weighted model	0.648
September 02	Equal Weighted model	0.908	September 02	Equal Weighted model	0.597
September 09	Equal Weighted model	0.877	September 09	Equal Weighted model	0.554
September 16	Equal Weighted model	0.764	September 16	Equal Weighted model	0.504
September 23	Equal Weighted model	0.796	September 23	Equal Weighted model	0.535

Notes: Equal Weighted model produce yield forecast composites of five yield forecasting models. MCS p-values are all greater than the significance level of 0.1, suggesting Equal Weighted model fails to significantly outperform individual yield forecasting models. The best model selected by MCS test is based on the significance level of 0.1, p-values are produced with 2000 bootstrap replicates for the test statistics.

Table 7: The multi-horizon average superior predictive ability (aSPA) test between BM model and other yield forecasting models for corn and soybean with fixed weights

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)
soybean	0.886	0.341	-0.0198	0.713
	(0.199)	(0.419)	(0.524)	(0.354)

Notes: This table presents the t-statistics and p-values (in parenthesis) for the multi-horizon aSPA test. 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).

Table 8: The multi-horizon average superior predictive ability (aSPA) test between BM model and other yield forecasting models for corn and soybean with varying weights

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)
soybean	0.843	1.565	1.088	1.262
	(0.151)	(0.270)	(0.310)	(0.332)

Notes: This table presents the t-statistics and p-values (in parenthesis) for the multi-horizon aSPA test. 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).

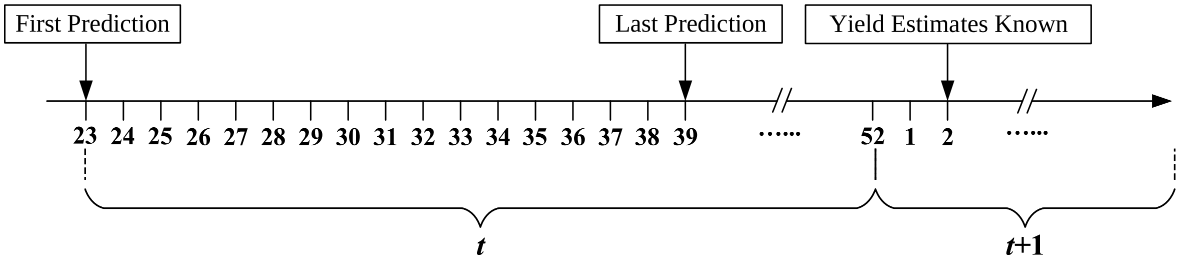


Figure 1: Yield forecasting cycle for corn

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

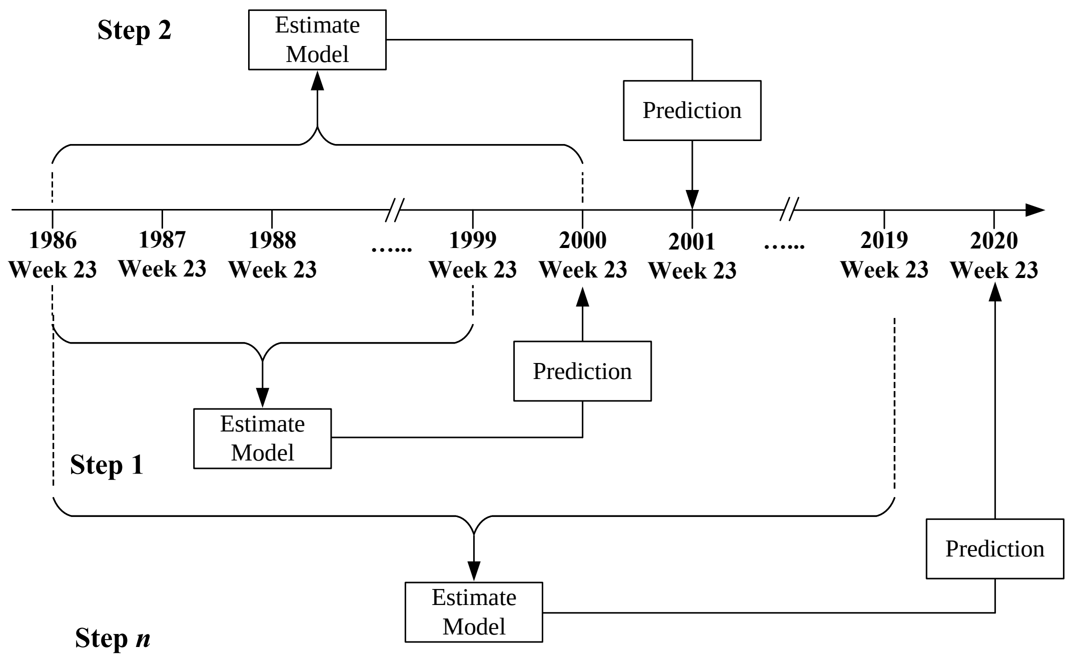


Figure 2: Recursive out-of-sample yield forecasts with Begueria and Maneta model (2017)

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

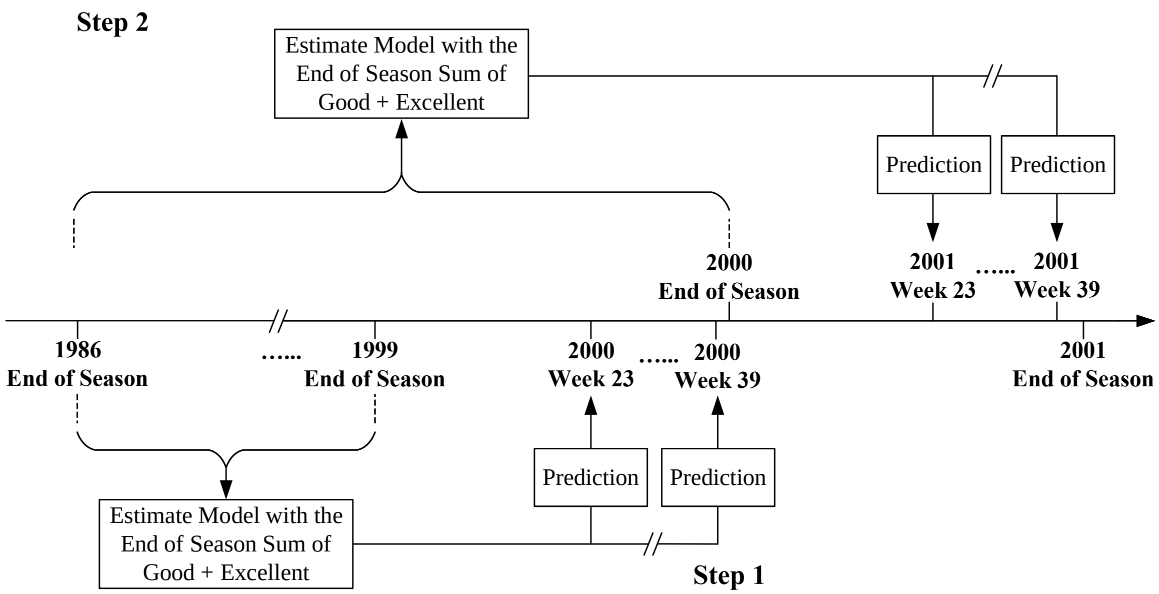
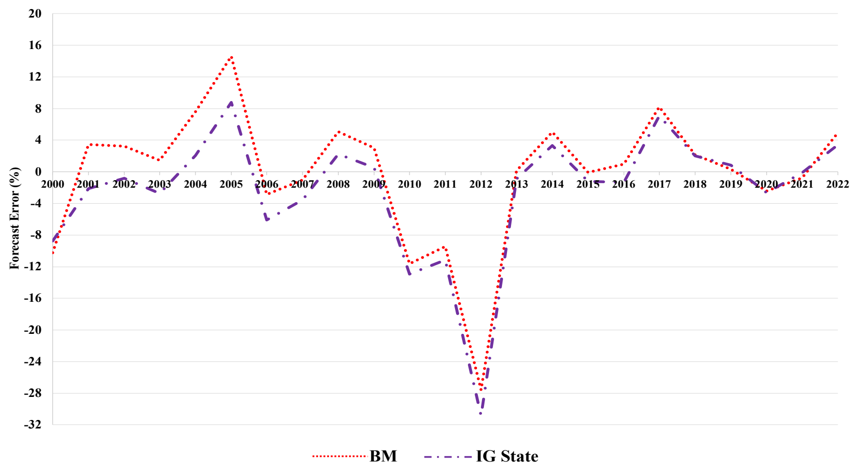
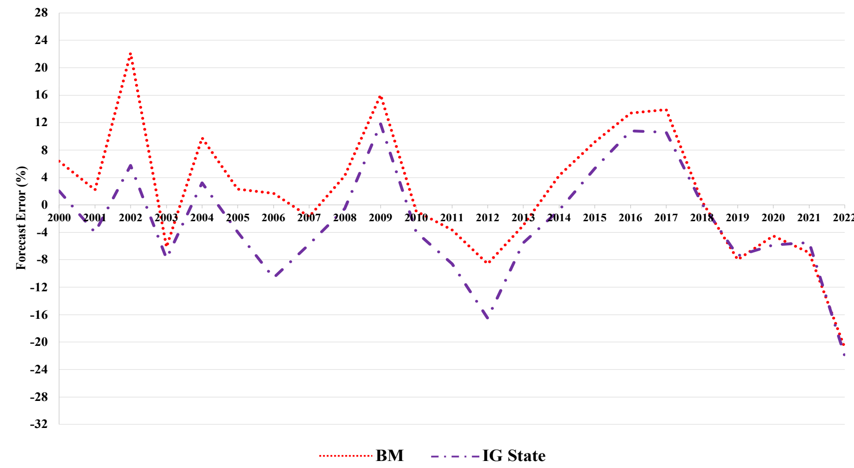


Figure 3: Recursive out-of-sample yield forecasts with IG State model and IG National model (2017a)

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



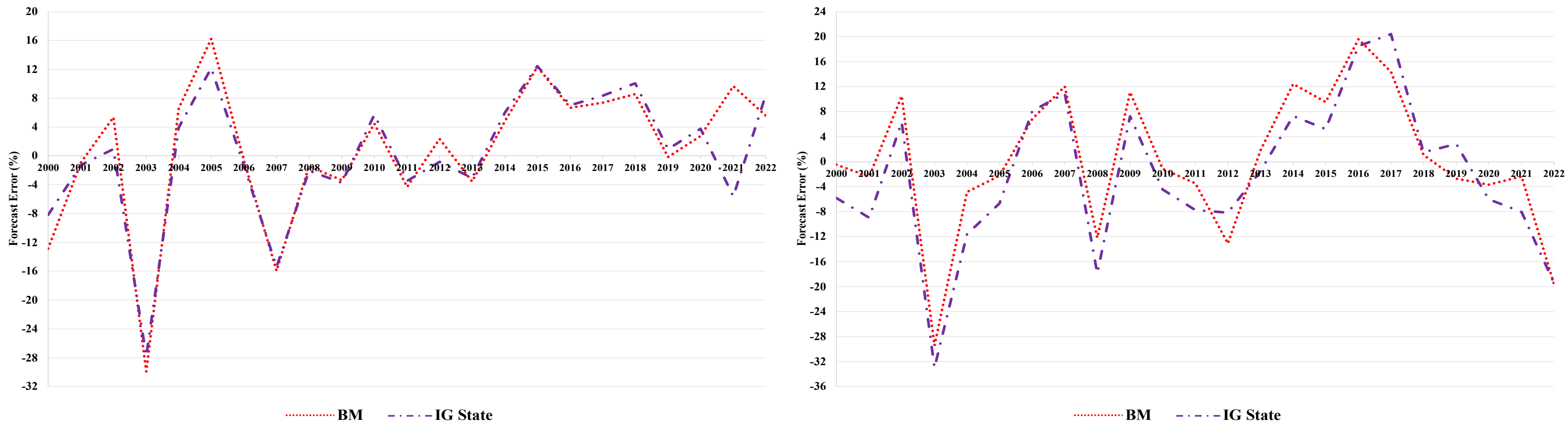
(a) Illinois Corn



(b) South Dakota Corn

Figure 4: The forecast error (%) of BM model and IG State model for week 29 for corn, for Illinois and South Dakota, 2000 – 2022

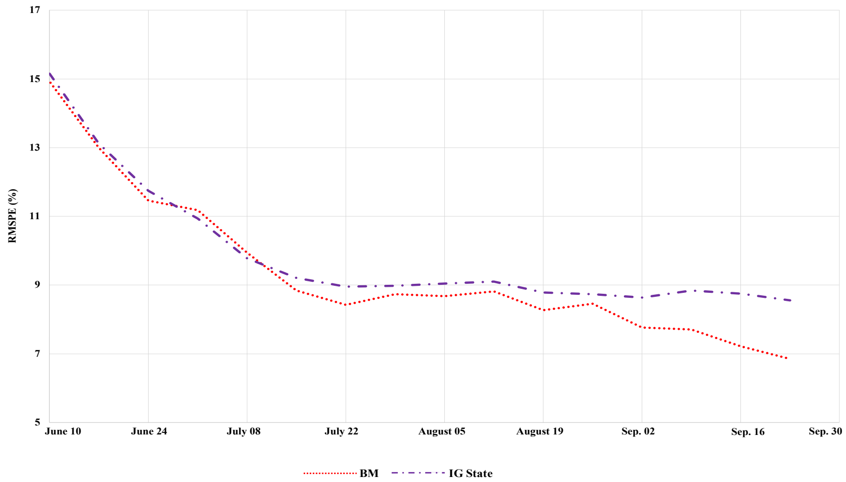
Notes: BM model is proposed by Begueria and Maneta (2020), IG State model is proposed by Irwin and Good (2017a).



(a) Illinois Soybean (b) South Dakota Soybean

Figure 5: The forecast error (%) of BM model and IG State model for week 29 for soybean, for Illinois and South Dakota, soybean, 2000 – 2022

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

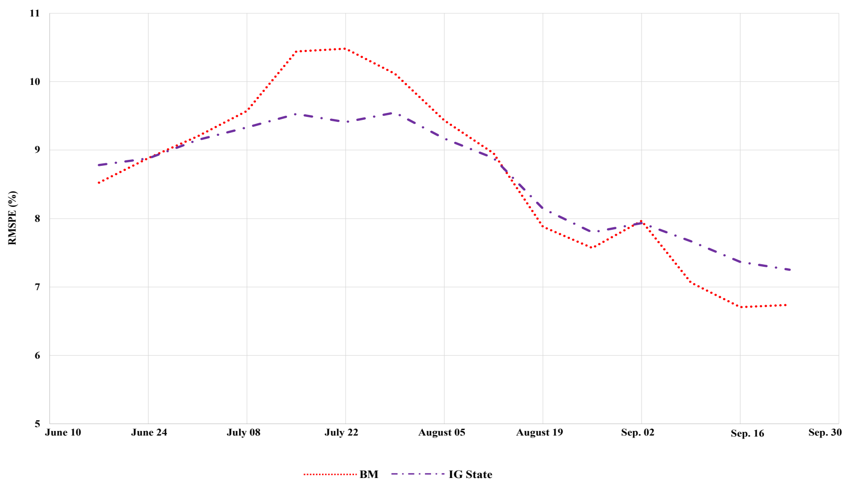


(a) Illinois Corn

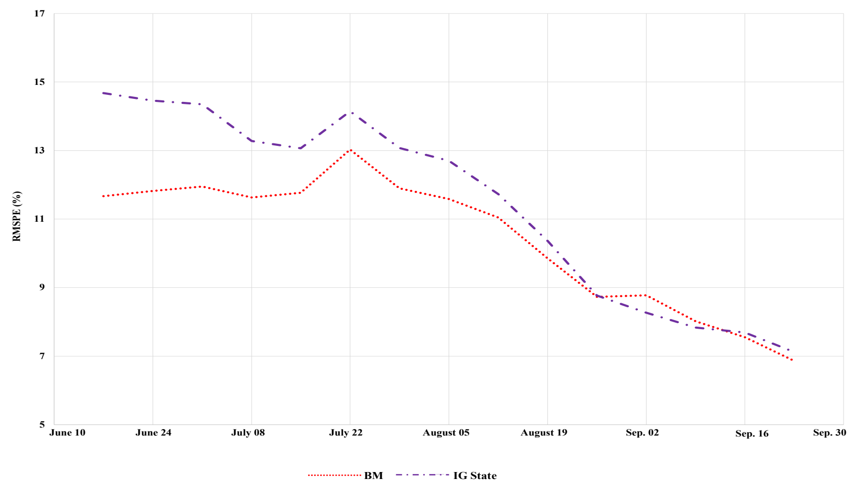
(b) South Dakota Corn

Figure 6: Weekly RMSPE of BM model and Irwin and Good model for Illinois and South Dakota for 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 Soybean



(b) South Dakota Soybean

Figure 7: Weekly RMSPE of BM model and Irwin and Good model for Illinois and South Dakota for soybean, 2000 – 2022

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

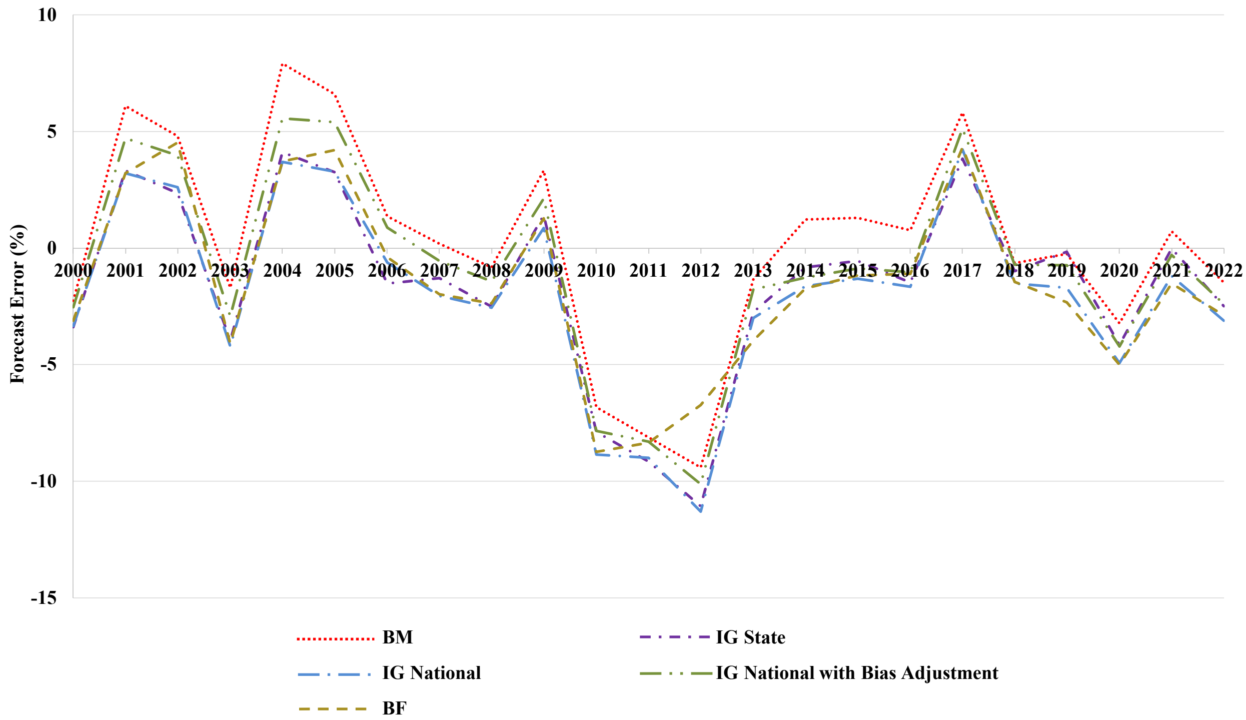


Figure 8: The forecast error (%) of five yield forecasting models for week 29 for corn, 2000 – 2022

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).

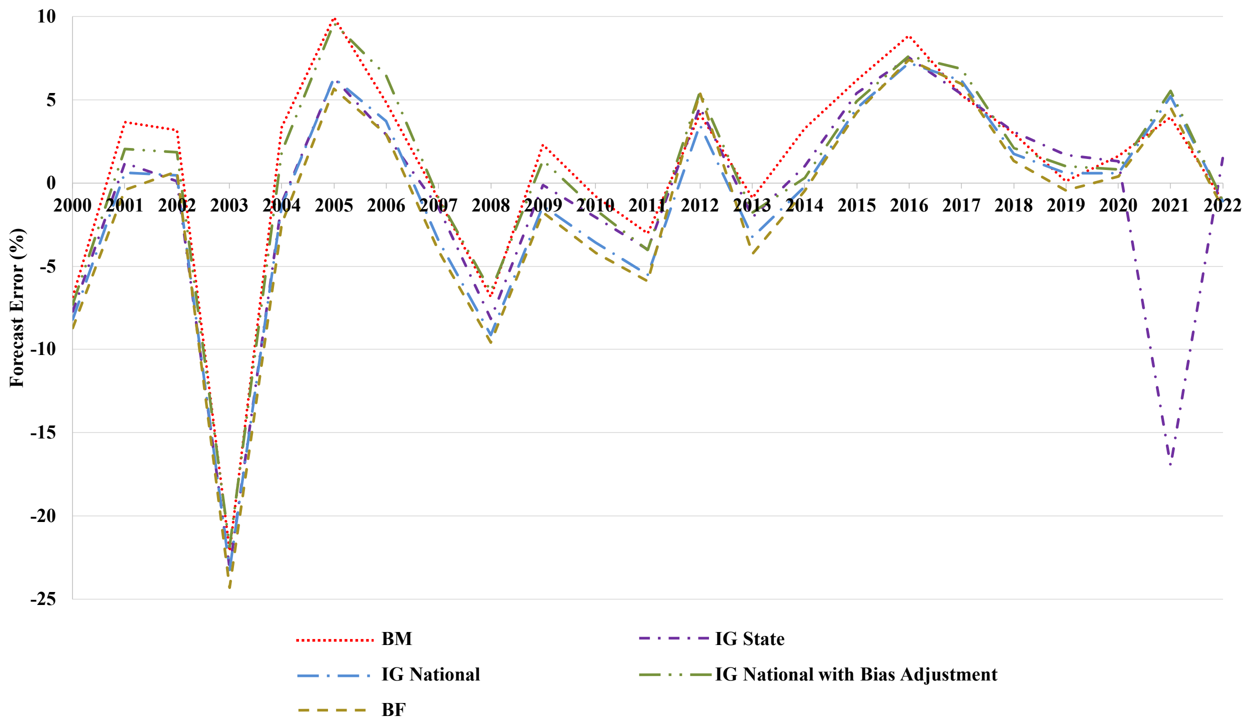


Figure 9: The forecast error (%) of five yield forecasting models for week 29 for soybean, 2000 – 2022

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).

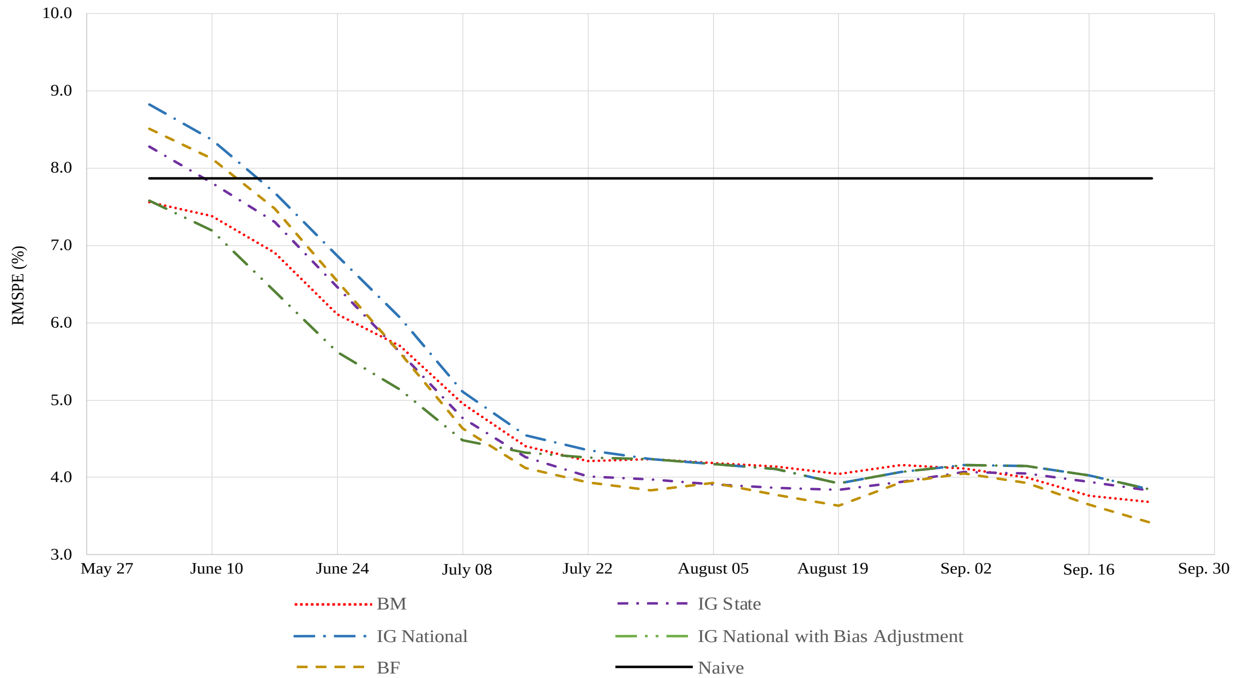


Figure 10: RMSPE of five yield forecasting models at national level from 2000 – 2022 for corn

Notes: we also include naïve trend yield model to illustrate the value of crop condition ratings as a yield indicator. 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).

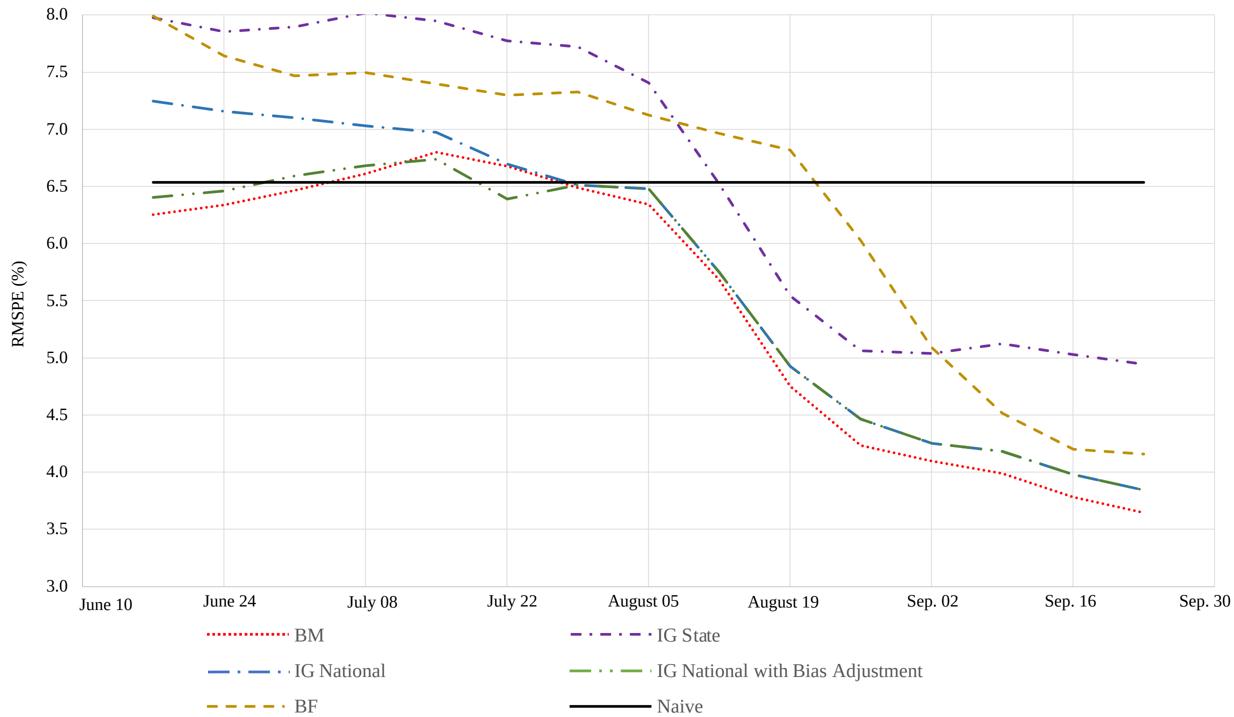


Figure 11: RMSPE of five yield forecasting models at national level from 2000 – 2022 for soybean

Notes: we also include naïve trend yield model to illustrate the value of crop condition ratings as a yield indicator. 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).

References

Armstrong, J.S. *Principles of Forecasting: A Handbook for Researchers and Practitioners.* New York: Springer Science+Business Media, Inc., 2001.

Bain, R., and T.R. Fortenbery. "Impact of Crop Condition Reports on National and Local Wheat Markets." *Journal of Agricultural and Applied Economics* 49,1(2017):97-119.

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

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

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

Funk, C., and M.E. Budde. "Phenologically-Tuned MODIS NDVI-Based Production Anomaly Estimates for Zimbabwe." *Remote Sensing of Environment*, 113,1(2009):115-125.

Harvey, D., S. Leybourne, and P. Newbold. "Testing the Equality of Prediction Mean Squared Errors." *International Journal of Forecasting* 13,2(1997):281-291.

Hansen, P.R., A. Lunde, and J.M. Nason. "The Model Confidence Set." *Econometrica* 79.2(2011): 453-497.

Hoogenboom, G. "Contribution of Agrometeorology to the Simulation of Crop Production and Its Applications." *Agricultural and forest meteorology* 103,1-2(2000):137-157.

Irwin, S.H., D.L. Good, and M. Tannura. "2009 Final Corn and Soybean Yield Forecasts." Marketing and Outlook Briefs, MOBR 09-06, Urbana-Champaign: Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, 2009.

Irwin, S.H., and D.L. Good. "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, 2017a.

Irwin, S.H., and D.L. Good. "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, 2017b.

Irwin, S.H., and T. Hubbs. "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, 2018a.

Irwin, S.H., and T. Hubbs. "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, 2018b.

Irwin, S.H., and T. Hubbs. "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, 2018c.

Irwin, S.H., and T. Hubbs. "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, 2018d.

Kauffman, R.K., and S. Snell. "A Biophysical Model of Corn Yield: Integrating Climatic and Social Determinants." *American Journal of Agricultural Economics* 79(1997):178-190.

Kruse, J., and D. Smith. "Yield Estimation Throughout the Growing Season." Center for Agricultural and Rural Development, Iowa State University, 1994.

Lehecka, G. V. "The Value of USDA Crop Progress and Condition Information: Reactions of Corn and Soybean Futures Markets." *Journal of Agricultural and Resource Economics* 39,1(2014):88-105.

Quaedvlieg, R. "Multi-Horizon Forecast Comparison." *Journal of Business & Economic Statistics* 39,1(2021):40-53.

Sakamoto, T., A.A. Gitelson, and T.J. Arkebauer. "MODIS-Based Corn Grain Yield Estimation Model Incorporating Crop Phenology Information." *Remote Sensing of Environment* 131(2013):215-231.

Shaw, L.H. "The Effect of Weather on Agricultural Output: A Look at Methodology." *Journal of Farm Economics* 46(1964):218-230.

U.S. Department of Agriculture, National Agricultural Statistics Service (USDA-NASS). *Surveys: Crop Progress and Conditions*. Retrieved December 13, 2021. Internet site: https://www.nass.usda.gov/Surveys/Guide_to_NASS_Surveys/Crop_Progress_and_Condition/index.php

Walker, G.K. "Model for Operational Forecasting of Western Canada Wheat Yield." *Agricultural and Forest Meteorology* 44(1989):339-351.