NCCC-134

APPLIED COMMODITY PRICE ANALYSIS, FORECASTING AND MARKET RISK MANAGEMENT

Basis Forecasting Performance of Composite Models: An Application to Corn and Soybean Markets

by

Kexin Ding and Berna Karali

Basis Forecasting Performance of Composite Models:
An Application to Corn and Soybean Markets

Kexin Ding and Berna Karali [1]

Paper presented at the NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management
Minneapolis, Minnesota, April 15-16, 2019.

[1] Kexin Ding and Berna Karali are a master student and associate professor in the Department of Agricultural and Applied Economics at the University of Georgia.

Basis Forecasting Performance of Composite Models:
An Application to Corn and Soybean Markets

Numerous studies have examined the performance of different models on basis forecasting while none of them has compared relative performance of composite models. In order to further improve basis forecasting accuracy, the crux of hedge management strategies, we investigate basis forecasting performance of selected composite models, as well as various individual models. Empirical results based on weekly futures and cash prices for major North Carolina corn and soybean markets indicate that composite models have more stable and better performance in forecasting basis compared to individual models' forecasts. The information-theoretic forecast combination method is found to be superior among the composite models considered.

Key words: basis forecasting, composite models, futures markets, information-theoretic forecast combinations.

Introduction

Basis, which is the difference between the local cash price of a commodity and the price of a futures contract of the same commodity at any given point in time, is a key to making informed risk management decisions (CBOT 2004). Theoretically, cash prices and futures prices tend to move together and converge to each other at maturity, making the concept of effective hedging possible. However, basis is not a constant value and it can be either positive or negative. Basis directly depends on market factors such as transportation costs, storage costs, handling costs and profit margins, as well as local supply and demand conditions (CBOT 2016). Basis risk is defined as the inherent risk a trader takes when hedging away cash price risk. In recent years, volatility in both corn and soybean basis has risen along with unprecedented price volatility experienced in both global and U.S. markets (Bekkerman and Pelletier 2009; Williams 2012; Taylor, Tonsor, and Dhuyvetter 2014). Figure 1 presents the changes in global prices of corn and soybean from 1980 to 2016, revealing the evident increase in price volatility since mid-2000. The same phenomenon can also be verified in figure 2, which presents the U.S. producer price index for corn and soybean.

Against this backdrop, the importance of improving basis forecasting for market participants who use futures or options to construct a hedging strategy is self-evident. Not only the prospective price they will receive or pay, but also the best time to accept a supplier's offer or a buyer's bid can be obtained from accurate basis forecasts (Tonsor, Dhuyvetter, and Mintert 2004; CBOT 2004). Given the importance of accurate basis forecasts, this study investigates whether basis forecasting can be further improved by utilizing composite forecasts.

Previous studies investigated basis forecasting using different approaches, varying from the most typical historical average approach (Hayenga et al. 1984; Dhuyvetter and Kastens 1998) and incorporation of market information into the historical average basis (Tonsor, Dhuyvetter, and Mintert 2004) to various complex time-series models such as threshold autoregressive model (Goodwin and Piggott 2001), smooth-transitioning autoregressive model (Sanders and Baker

2012), and generalized additive models (Onel and Karali 2014; Guney, Goodwin, and Riquelme 2019). Since alternative forecasting models include different information sets and rarely result in the same prediction of basis, combination of individual forecasts has then been investigated using the Bayesian model averaging approach (Payne, Karali, and Dorfman 2019).

Similar to individual forecasting models, different composite forecasting methods also deserve to be compared with each other since they are weighted averages of individual forecasts based on different theoretical criteria, and they might yield improved basis forecasts. Therefore, the objective of this study is to explore the relative performance of several individual forecast models, including both linear and nonlinear, as well as that of a number of composite models with different choices of weight for each individual model. In particular, a linear autoregressive (AR) model is introduced to compare with three nonlinear models: the threshold-autoregressive (TAR) model, smooth-transitioning autoregressive (STAR) model and generalized additive model (GAM). Then, three composite forecast models are constructed including commonly used equally weighted forecasts (EW), inverse mean-squared error method (IN), and the data-driven, information-theoretic forecast combination method (IT-AIC). For an application, corn and soybean markets are chosen due to their important role in U.S. agriculture. In 2018, for example, the cash receipts of corn and soybean account for nearly half of the total crop cash receipts which can be seen in figure 3, with $48.5 billion of cash receipts (24% share) for corn and $40.4 billion of cash receipts (20% share) for soybean. In particular, we use spot prices of corn and soybean in major North Carolina markets and nearby futures prices. Both corn and soybean futures contracts have long been the most active agricultural commodity futures traded at the Chicago Board of Trade (CBOT). Their average daily volume notional value makes up the majority of agricultural commodity futures.

Results of individual basis forecasts show nonlinear forecast models outperform the linear model. Especially, GAM performs best among individual forecast models in long horizons. In general, composite forecast models have more stable and improved basis forecasting performance than individual forecast models. Further, information-theoretic forecast combinations compare favorably with other models.

Literature Review

Basis Forecasting

New basis forecasting methods have been developed during the last two decades. Earlier studies tend to use naïve models, non-regression methods like variants of the simple moving averages, or simple time-series models. Heifner (1966) uses a least-squares regression method and show that much better predictions are obtained for basis changes than for cash price changes for corn in Michigan. Taylor and Tomek (1984) construct an econometric model based on the difference between Chicago and New York cash prices and the difference between Chicago cash and futures prices to forecast the corn basis and compare it to naïve forecasts. They show that their econometric model yields better goodness-of-fit measures than do the naïve forecasts. Hauser, Garcia, and Tumblin (1990) focus on the role of basis expectations in measuring hedging effectiveness. They evaluate the performance of various methods in forecasting soybean basis, including regression models and historical average methods. Their results suggest when much of the information is still forthcoming, the historical average method using previous year or past three years is the best predictor. Dhuyvetter and Kastens (1998) use various models to forecast basis for wheat, corn, and soybean for each week of the year, and show that basis forecasts based on simple historical averages compare favorably with more complex forecasting models. Tonsor, Dhuyvetter, and Minert (2004) evaluate live cattle and feeder cattle basis forecast accuracy using forecasts generated by a historical calendar-date technique and a time-to-futures-contract-expiration technique. Their results indicate that the time-to-expiration approach has little impact on forecast accuracy compared to a simple calendar approach. Sanders and Manfredo (2006) compare several time series methods for basis forecasts in the soybean futures complex, which are generated with an exponential smoothing technique, autoregressive moving average (ARMA) model, and vector autoregressive (VAR) model. They find that alternative naïve techniques, such as year-ago and no-change methods, outperform the 5-year average method, and the improvement gained by time-series modeling is relatively small.

More recent work applies various complex time-series models to basis forecasting. Sanders and Baker (2012), for instance, apply a STAR model using data from ten reporting locations in Ohio, and compare it to the standard AR model and to the commonly used 5-year moving average. They find that time-series models can provide better basis forecasts in the short run. Moreover, although there is statistical evidence in favor of the regime-changing models, they provide no real forecasting improvement over simpler autoregressive models. Onel and Karali (2014) use a semi-parametric GAM as well as simple autoregressive models in forecasting corn and soybean basis and find the GAM to yield better forecasts. Guney, Goodwin, and Riquelme (2019) apply semi-parametric, the vector GAM to basis linkages among North Carolina corn and soybean markets. The results of GAM are compared to standard vector TAR models and are found to reveal more statistical significance and substantially more nonlinearity in basis adjustments. Payne, Karali, and Dorfman (2019) construct a Bayesian averaging approach which combines cattle basis forecasts from different models and compare it to simple moving averages. Their results suggest that the average model typically performs favorably compared to regression models. However, except for very short-horizon forecasts, simple moving averages have a lower out-of-sample forecast errors than the regression models.

We extend this body of literature by applying the simple time-series model, the AR model, and various complex nonlinear time-series models mentioned above to basis forecasting. Further, our study focuses on comparing different composite models that assign different weights to aforementioned individual forecasting models. This can be thought as a natural next step following the Bayesian averaging approach used by Payne, Karali, and Dorfman (2019). The innovation of this paper is comparing the performance of different composite forecasts to see whether basis forecasting can be further improved.

Composite Forecasting

Mingled with accuracy gains, composite forecasting models have been widely applied in various fields. The most common combination approach is equally weighting different forecasts, which has been proven to perform better than regression weights. Einhorn and Hogarth (1975) compare the predictive ability between the equal weighting method and standard linear regression composite method and show that equal weighting is superior in certain situations and not greatly inferior in other cases. Dawes (1979) presents evidence that equal weighting is quite robust for making predictions compared with linear composite models. Aiolf, Capistrán, and Timmermann (2011) consider combinations of subjective survey forecasts and model-based forecasts from linear and nonlinear univariate specifications as well as multivariate factor-augmented models. They conclude that using equal weights leads to better forecast performance than using estimated combination weights in roughly two-thirds of all cases. Colino et al. (2012) investigate whether the accuracy of hog price forecasts can be improved using composite forecasts. Their findings favor the use of equally-weighted forecasts.

Another simple composite forecasting method that has been frequently used is the inverse mean-squared error method. Capistrán and Timmermann (2009) explore the consequences of frequent entry and exit of individual forecasters for several composite forecasting methods containing the inverse mean-squared error method. Although the results show the best predictor is the bias-adjusted mean approach, the inverse mean-squared error method outperforms many other methods. Aiolf, Capistrán, and Timmermann (2011) indicate even if forecast combinations, such as inverse mean-squared error method, do not always deliver the most precise forecasts they generally do not deliver poor performance.

More complicated composite forecasting methods are also applied. Hansen (2007) compares estimators obtained from information-theoretic model averaging based on several different criteria finding that the Mallows model average (MMA) estimator compares favorably compared to weights based on the Akaike information criterion (AIC) and Bayesian information criterion (BIC). Kapetanios, Labhard, and Price (2008) consider the use of information-theoretic model averaging in forecasting U.K. inflation with a large dataset, and find that it can be a powerful alternative to Bayesian averaging schemes.

In consideration of performance, we first choose the most common equal weighting method. Then the inverse mean-squared error method is considered. Finally, the information-theoretic model averaging based on AIC is considered.

Methodology

Forecast Models

In order to compare the relative performance of forecast models with different properties, four individual forecast methods consisting of both linear and nonlinear, parametric and nonparametric models are considered. Then, three composite forecast methods are employed assigning different weights to each individual methods.

First, consider a simple linear model, an AR(p) process, defined as:

$$y_t = c + \sum_{i=1}^{p} \varphi_i y_{t-i} + \varepsilon_t, \qquad (1)$$

where y_t denotes basis at time t, y_{t-i} denotes the i^{th} lagged value of y_t, c and φ are parameters to be estimated, and ε_t is the disturbance term.

The first nonlinear model considered is the self-exciting threshold autoregressive (SETAR) which is a special class of TAR models first proposed by Tong (1978) and further discussed by Gonzalo and Pitarakis (2002). Unlike the AR models, SETAR models allow the parameters to change according to the value of an exogenous threshold variable y_{t-d}:

$$y_t = \phi_0^{(j)} + \sum_{i=1}^{p} \phi_i^{(j)} y_{t-i} + \sigma^{(j)} \varepsilon_t,\ r_{j-1} \le y_{t-d} < r_j, \tag{2}$$

where y_t and y_{t-i} are defined as above, $\sigma^{(j)}$ denotes variance terms of j^{th} regime, j=1, 2, ..., k, and the thresholds for regimes are $-\infty = r_0 < \ldots < r_k = \infty$. The threshold variable y_{t-d} is the d^{th} lagged value of y_t, with the delay parameter d being a positive integer. Thus, the $k-1$ thresholds $(r_1, r_2, \cdots, r_{k-1})$ divide the domain of y_{t-d} into k different regimes. For each j, y_t follows a linear AR(p) model.

Although the SETAR model can capture many nonlinear features, the regime switch happens when the threshold variable y_{t-d} crosses a certain value which is discontinuous. By replacing the threshold value with a smooth transition function, the SETAR model can be generalized as a STAR model introduced by Chan and Tong (1986). The STAR model of order p is defined as follows:

$$y_t = \pi_{10} + \sum_{i=1}^{p} \pi_{1i} y_{t-i} + (\pi_{20} + \sum_{i=1}^{p} \pi_{2i} y_{t-i}) F(y_{t-d}, \gamma, c) + u_t, \tag{3}$$

where y_t, y_{t-i}, and y_{t-d} are defined as above, $F(y_{t-d}, \gamma, c)$ is the transition function bounded between 0 and 1, and $u_t \sim iid(0, \sigma^2)$. Thus, the series y_t switches between regimes smoothly as the dynamics of y_t may be determined by more than one regime. There are two different transition functions specified by Teräsvirta (1994), logistic smooth transition autoregressive (LSTAR) model and exponential smooth transition autoregressive (ESTAR) model, given respectively as:

$$F(y_{t-d}, \gamma, c) = (1 + exp[-\gamma(y_{t-d} - c)])^{-1}, \gamma > 0. \tag{4}$$

$$F(y_{t-d}, \gamma, c) = 1 - exp(-\gamma(y_{t-d} - c)^2), \gamma > 0. \tag{5}$$

The last nonlinear forecast model considered is the GAM proposed by Hastie and Tibshirani (1986) assuming that the mean of the dependent variable depends on an additive predictor through a link function. The model is:

$$y_t = \beta_0 + \sum_{i=1}^{p} f_i(y_{t-i}) + \varepsilon_t, \tag{6}$$

where y_t and y_{t-i} are defined as above, $f_i(\cdot)$ is an unspecified smooth nonparametric function whose estimate transforms the explanatory variable so as to maximize the fit to the dependent variable, subject to constraints about the smoothness of the link function. This nonparametric smoothing has the advantage of avoiding overfitting compared with purely parametric models, but arguably with some loss of interpretability (Guney, Goodwin, and Riquelme 2019). Among various link functions, there are two major types, one of which is locally weighted regression smoothers (LOESS) that fits multiple regressions in local neighborhood. The other one is cubic smoothing spline (SPLINE) whose estimated function within the space of all smooth functions is defined to be the minimizer of a penalized residual sum of squares.

All the models presented above are combined in three different composite models. The most common combination approach is weighting forecasts equally:

$$y_{t+h|t} = M^{-1} \sum_{m=1}^{M} \hat{y}_{t+h|t}^{m}, \tag{7}$$

where $y_{t+h|t}$ denotes the composite basis forecast for time $t+h$, $\hat{y}_{t+h|t}^{m}$ denotes the m^{th} individual model's basis forecast for time $t+h$, and M=4.

The second composite model we consider is the inverse mean-squared error method using weights that are inversely proportional to their historical mean squared error values (Timmermann 2006; Capistran and Timmermann 2009). This method reduces the effect of parameter estimation error on the composite forecast. The normalized weight for the m^{th} individual model is denoted as:

$$w_m = \frac{\frac{1}{MSE_m}}{\sum_{m=1}^{M} \frac{1}{MSE_m}}, \tag{8}$$

where MSE_m denotes the mean-squared error of the m^{th} individual model. Then, the composite forecast is defined as follows:

$$y_{t+h|t} = \sum_{m=1}^{M} w_m \hat{y}_{t+h|t}^{m}. \tag{9}$$

The last composite forecasting model considered is the information-theoretic forecast combinations based on AIC which is applied by Kapetanios, Labhard, and Price (2008). This method allows the amount of shrinkage enforced on each individual forecast driven by data. The normalized weight for the m^{th} individual model is given by:

$$w_m = \frac{exp(-1/2\psi_m)}{\sum_{m=1}^{M} exp(-1/2\psi_m)}, \tag{10}$$

where

$$\psi_m = AIC_m - min_j AIC_j. \tag{11}$$

AIC_m denotes the AIC of the m^{th} individual model, $min_j AIC_j$ denotes the AIC of the j^{th} individual model with the minimum AIC. Thus, the term $exp(-1/2\psi_m)$ can be interpreted as the odds for the m^{th} model to minimize the estimated information loss in M models, also known as being the best Kullback and Leibler (KL) distance model. Using these normalized weights, the composite forecast is defined as in equation (9).

Forecast Evaluation Methods

Three measures of accuracy are employed to evaluate the relative performance of all models presented above. Two measures use forecast errors:

$$e_{t+h}^{n} = A_{t+h} - F_{t+h|t}^{n}, \tag{12}$$

where e_{t+h}^{n} denotes the forecast error of the n^{th} model at time $t+h$, A_{t+h} is the actual basis at time $t+h$, and $F_{t+h|t}^{n}$ denotes the basis forecast of the n^{th} model for time $t+h$ as of t, for n=1, 2, …, 7. The other measure uses percent forecast errors defined as follows:

$$p_{t+h}^{n} = \frac{A_{t+h} - F_{t+h|t}^{n}}{A_{t+h}} * 100, \tag{13}$$

The first accuracy measure considered is root mean squared error (RMSE), with the n^{th} models RMSE given by:

$$RMSE^{n} = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (e_{t+h}^{n})^2}, \tag{14}$$

where T denotes out-of-sample size. The mean absolute error (MAE) of the n^{th} model is:

$$MAE^n = \frac{1}{T}\sum_{t=1}^{T} |e_{t+h}^n|, \tag{15}$$

Finally, the mean absolute percentage error (MAPE) of the n^{th} model is computed as follows:

$$MAPE^n = \frac{1}{T}\sum_{t=1}^{T} |p_{t+h}^n|. \tag{16}$$

In order to test if a given accuracy measure is statistically different across models, the Modified Diebold Mariano (MDM) test introduced by Harvey, Leybourne, and Newbold (1997) is considered. We test each other model with the best performed model. The test statistic is computed as:

$$MDM = \sqrt{\frac{T-1}{\frac{1}{T}\sum_{t=1}^{T}(d_t - d)^2}}\, d, \tag{17}$$

where d_t is the difference between respective loss functions at time t, such as squared error, absolute error, and absolute percentage error, d is the average difference, T is the total sample period. The null hypothesis is the forecast accuracy of two models is same. The alternative hypothesis is the forecast accuracy of two models is different.

Data

Our application is to weekly corn and soybean basis observations from major North Carolina markets. Weekly corn cash prices in Candor, Cofield, and Roaring River are from January 1988 to February 2017, and soybean prices in Elizabeth City and Fayetteville are from January 1980 to February 2017. Prices are quoted as cents per bushel. For the weeks with missing price data (less than 7% of total), interpolation is performed with a cubic spline method.

We use weekly Wednesday settlement prices of corn and soybean futures contracts. These contracts are traded at CBOT and have a size of 5,000 bushels, and their prices are quoted as cents per bushel. Nearby futures price series are constructed by rolling over the futures contracts at the end of the month preceding the delivery. Corn futures have maturity months of March, May, July, September, and December, while soybean futures have contract months of January, March, May, July, August, September, and November. Table 1 presents the specific futures contract months used in constructing the nearby price series. Finally, we compute the weekly basis in each location as the difference between the cash price and nearby futures price.

The descriptive statistics of both cash and futures prices as well as basis are shown in table 2. In corn markets, Candor and Roaring River have higher average basis than Cofield. In soybean markets, Elizabeth City has negative average basis while Fayetteville's is positive. The standard deviation of all basis series is similar indicating they have similar amount of variation. As figure 4 demonstrates both cash and futures prices fluctuated more in the last decade. Further, increased basis volatility since 2008 can be observed in figure 5. These observations are consistent with the unprecedented price volatility experienced in global and U.S. markets for corn and soybean (figures 1 and 2).

Prices of storable commodities, and therefore the basis, exhibit seasonal patterns. Specifically, price levels increase as approaching to harvest due to declining inventories, and start to decrease after the harvest as more commodity becomes available. To account for seasonality, we fit a sinusoidal polynomial function to basis data. The order of the polynomial in each location varies, with Candor and Roaring River having 2 terms each, Cofield having 4, Elizabeth City having 5, and Fayetteville having 3. Figure 6 shows the predicted basis obtained with these sinusoidal polynomial functions, verifying such seasonal patterns. In order to remove the intra-year seasonality caused by the harvest cycle, we use the residuals obtained from these regressions to represent deseasonalized basis. We then test deseasonalized basis series for unit root using the augmented Dickey-Fuller tests with one augmenting lag and Phillips-Perron tests with one lag. Results shown in table 3 indicate that both corn and soybean deseasonalized basis are stationary according to two tests.

Each model presented in the previous section are estimated using the stationary, deseasonalized basis series. To compare the relative performance of different forecasting models, the last 24 observations are reserved for out-of-sample accuracy evaluation. We apply the iterated strategy, in which, for a given model, multi-period forecasts are obtained by recursively using the one-step ahead forecast until reaching the horizon. Thus, in our application, we obtain forecasts for horizons 2, 6, 12, 18, and 24 (weeks) by recursively forecasting 1-week ahead basis using the in-sample parameter estimates. The iterated method has the advantage of estimating only one model, but leads to accumulated forecast errors over multiple horizons (Xiong, Bao, and Hu 2013). The accuracy evaluation constructed between actual basis values and restored basis predictors that are recovered from deseasonalized basis forecasts.

Results

Estimation Results of Individual Forecasting Models

Selecting the order by AIC, we estimate an AR model for corn and soybean markets using the Yule-Walker method. The estimation results are presented in table 4. While all corn markets have selected order of 5, soybean market has 4 lags in Elizabeth City and 5 lags in Fayetteville.

The estimation results for SETAR models are shown in table 5 after fine tuning the hyper-parameters, with embedding dimension setting as 5, delay as 1, forecasting steps as 1, and threshold delay being bounded between 1 and 4. All cities have a higher proportion of the first regime indicating that regime switch happens near the end of the sample period. The autoregressive parameters estimates in each regime are significantly different from the pure AR model. Further, estimations in the first regime significantly vary from parameters in the second regime of each location. Specifically, insignificant parameters exist in both regimes of Cofield, Roaring River, and Elizabeth City, as well as high regime of Fayetteville.

We set the maximum number of possible regimes to 5 in the STAR model for simplicity and tractability of the model. According to the linearity test results, all series are nonlinear (p-values are smaller than 0.01), which is consistent with the findings in Guney, Goodwin, and Riquelme (2019). Table 6 presents the STAR estimation results showing the parameters of transition functions, linear parameters of lagged basis, and nonlinear parameters of lagged basis.

Table 7 presents the estimation results of GAM with LOESS smoother. Results with SPLINE smoother are similar; therefore, we use the LOESS smoother as representative for the GAM forecasting. Both linear parameters of lagged basis and smooth parameters of lagged basis are significant in all markets. Figure 7 shows the nonlinear effects of the lagged value of basis on the current basis. For all three corn markets, we observe that when the last period's basis is large and negative, its effect on this period's basis is almost linear which is larger and positive. However, when the last period's basis is large and positive, its effect on this period's basis turns out to be nonlinear which is negative. Similar nonlinear effects are also found in soybean markets. However, when the last period's basis of Fayetteville is small and negative, its effect on this period's basis tends to be nonlinear.

Forecast Accuracy

Tables 8 through 13 present the performance accuracy measures (RMSE, MAE, MAPE) of four individual forecasting models and three composite forecasting models with 1-, 2-, 6-, 12-, 18-, and 24-week forecasting horizons for corn and soybean markets. For convenient horizontal comparison among different markets, the transformed MAPE index is computed by dividing the MAPE value of a given model by the MAPE of the AR model. If the transformed MAPE index is less than 1, the model outperforms the AR model. Similar to MAPE, the smaller the transformed MAPE index, the better the model forecast. Figures 8 through 13 present the transformed MAPE index with the same six forecasting horizons. Ranking results shown in these figures are consistent with the results based on MAPE presented in tables 10 and 13.

In short forecast horizons, 1- and 2-week horizons, the information-theoretic forecast combination has the best performance according to all three accuracy measures in soybean markets and in Roaring River corn market. One exception is the 2-week-ahead forecasts in Elizabeth City under the MAPE criterion, in which case EW performs best. For Candor, GAM and STAR outperfom composite forecast models for 1-week horizon forecasts while the IT-AIC has the best performance for 2-week horizon forecasts. For Cofield, several individual forecast models outperfom composite forecast models. Specifically, SETAR performs best for 1-week horizon forecasts, AR performs best for 2-week horizon forecasts under the RMSE criterion, and SETAR performs best for 2-week horizon forecasts under the MAE and MAPE criteria. We can see in figures 8 and 9 that while the IT-AIC has the best forecast in many cases, the relative performance of other models is unstable. Also, according to the MDM test based on MAPE criterion shown in table 14 and marked bold in tables 10 and 13, the 2-week horizon performance differences of STAR, GAM, EW, and IN with the best model are insignificant for Candor. All performance differences for 2-week horizon are insignificant for Cofield and Elizabeth City. SETAR and STAR have insignificant performance differences with the best model for Roaring River. For Fayettville, only GAM's forecasting performance statistically differs from the best model.

For longer forecast horizons, 6- and 12-week horizons, the IT-AIC has the best performance in the majority of cases, except for corn in Candor for 6-week horizon forecasts under the MAE and MAPE criteria, Roaring River for 6-week horizon forecasts, and soybean in Fayetteville under the MAPE criterion. Further, GAM frequently has the second best forecasts. Specifically, for 6-week horizon forecasts, GAM performs second best in Cofield under the MAE and MAPE criteria, and Elizabeth City under the MAE criterion. For 12-week horizon forecasts, GAM performs second best in Candor under all criteria, Cofield under the MAE and MAPE criteria, and Elizabeth City under the MAE criterion. It can be seen in figures 10 and 11 that the basis forecasting performance of different models varies widely. Also, according to the MDM test based on MAPE criterion shown in tables 15 and 16 and marked bold in tables 10 and 13, the 6-week horizon performance differences of GAM and IT-AIC with the best model are insignificant for Candor, while only GAM is insignificant in 12-week horizon. For Cofield, only SETAR has significant performance differences with the best model in 6-week horizon, while performance differences of GAM, EW, and IN with the best model are insignificant in 12-week horizon. AR and EW have significant performance differences with the best model for Roaring River in 6- and 12-week horizons. The 6-week horizon performance differences between STAR, GAM, EW, and IN with the best model are insignificant for Elizabeth City, while GAM, EW, and IN with the best model are insignificant in 12-week horizon. All 6- and 12-week horizons performance differences are insignificant for Fayettville.

In much longer forecast horizons, 18- and 24-week horizons, the IT-AIC has absolute advantage over all models as it performs best or second best in all cases, except for Fayettville for 18-week horizon forecasts under MAPE criterion in which EW performs best. Similar to 6- and 12-week horizons, GAM frequently has the best and second best forecasts. Specifically, GAM performs best in Cofield under the MAE and MAPE criteria, and Candor for 24-week horizon forecasts under the MAPE criteria. Meanwhile, GAM generally performs best among individual forecast models. Similar polarization trend can be seen in figures 12 and 13. Also, according to the MDM test based on MAPE criterion shown in tables 17 and 18 and also marked bold in tables 10 and 13, the performance differences between GAM with the best model are insignificant for Candor in both18- and 24-week horizons. For Cofield, the performance differences between IT-AIC with the best model are insignificant in both18- and 24-week horizons. The 18-week horizon performance differences between GAM, and IT-AIC with the best model are insignificant for Roaring River, while only IT-AIC is insignificant in 24-week horizon. The performance differences between GAM, EW, and IN with the best model are insignificant for Elizabeth City in both 18- and 24-week horizons. All 18- and 24-week horizons performance differences are insignificant for Fayettville.

Conclusion

Basis forecasting has long been an important research question as many successful hedge strategies based on futures or options are contingent on their ability to avoid basis risk by accurately forecasting basis. Various previous studies have investigated basis forecasting with different models. However, the relative performance of different composite models in forecasting crop basis has not been explored in the literature. Thus, we investigate whether basis forecasting can be further improved by focusing on basis forecasting using composite models. This study compares the relative performance of seven models, including four individual forecast models and three composite forecast models. Accuracy measures are based on out-of-sample forecasts.

Although the best individual forecast model in a given forecast horizon varies across different markets, nonlinear forecast models outperform the linear model. Especially, GAM performs best among individual forecast models in long horizons. This conclusion is consistent with Guney, Goodwin, and Riquelme (2019) which states the vector GAM is found to perform better in basis adjustments compared with threshold VAR models. In general, composite forecast models have more stable and accurate basis forecasting performance than individual forecast models in all horizons, except for the very short-run horizon in corn markets. This finding is consistent with previous studies stating combined forecasting substantially reduces forecast errors (Adams 1978; Colino et al. 2012; Payne, Karali, and Dorfman 2019). The relative performance of three composite forecast models is found to be consistent across different horizons. The information-theoretic forecast combination method generally outperforms the equally-weighted and inverse mean-squared error methods. This disparity in the best individual forecast model and the superiority of information-theoretic forecast combination method highlights the necessity of considering composite models when forecasting basis.

References

Akaike, H. 1978. "On the likelihood of a time series model." *The Statistician* 27: 217-235.

Adams, A.J. 1978 "Modeling and forecasting seasonal time series of product sales across territories: A comparative study." The University of Iowa.

Aiolf, M., C. Capistrán, and A. Timmermann. 2011. "Forecast Combinations." *The Oxford Handbook of Economic Forecasting*.

Bekkerman, A., and D. Pelletier 2009. "Basis Volatilities of Corn and Soybean in Spatially Separated Markets: The Effect of Ethanol Demand." Agricultural and Applied Economics Association.

Chan, W.S., and H. Tong. 1986. "On tests for non-linearity in time series analysis." *Forecast* 5: 217-228.

CBOT. 2004. "Understanding Basis." Chicago Board of Trade.

Capistrán, C., and A. Timmermann. 2009. "Forecast Combination with Entry and Exit of Experts." *Journal of Business and Economic Statistics* 27 (4): 428-440.

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): 228-46.

CBOT. 2016. "Learn about Basis: Grains." Chicago Board of Trade.

Dawes, R.M. 1979. "The robust beauty of improper linear models in decision making." *American Psychologist* 34(7): 571-582.

Dhuyvetter, K.C., and T.L., Kastens. 1998. "Forecasting Crop Basis: Practical Alternatives." In Proceedings of the 1998 NCR-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management.

Einhorn, H.J., and R.M. Hogarth. 1975. "Unit weighting schemes for decision making." *Organizational Behavior and Human Performance* 13(2): 171-192.

Goodwin, B.K., and N.E. Piggott. 2001. "Spatial market integration in the presence of threshold effects." *American Journal of Agricultural Economics* 83: 302-17.

Gonzalo, J. and J. Pitarakis. 2002 "Estimation and Model Selection Based Inference in Single and Multiple Threshold Models" *Journal of Econometrics* 110: 319-352.

Guney, S., B.K. Goodwin, and A. Riquelme. 2019. "Semi-Parametric Generalized Additive Vector Autoregressive Models of Spatial Basis Dynamics." *American Journal of Agricultural Economics* 101 (2): 541–562.

Heifner, R.G. 1966. "The Gains from Basing Grain Storage Decisions on Cash-Futures Spreads." *Journal of Farm Economics* 48 (5): 1490-1495.

Hayenga, M.L., D.D. DiPietre, J.M. Skadberg, and T.C. Schroeder. 1984. "Profitable Hedging Opportunities and Risk Premiums for Producers in Live Cattle and Live Hog Futures Markets." *The Journal of Futures Markets* 4(2): 141-154.

Hastie, T.J. and R.J. Tibshirani. 1986. "Generalized Additive Models." *Statistical Science* 1(3): 297-318.

Hauser, R.J., P. Garcia, and A.D. Tumblin. 1990. "Basis Expectations and Soybean Hedging Effectiveness." *North Central Journal of Agricultural Economics* 12: 127-136.

Harvey, D., S. Levbourne, and P. Newbold 1997. “Testing the equality of prediction mean squared errors.” *International Journal of forecasting* 13(2): 281-291.

Hansen, B.E. 2007. “Least Squares Model Averaging.” *Econometrica* 75 (4): 1175–1189.

Kapetanios, G., V. Labhard, and S. Price. 2008. “Forecasting Using Bayesian and Information-Theoretic Model Averaging: An Application to U.K. Inflation.” *Journal of Business & Economic Statistics* 26: 33-41.

Onel, G., and K. Berna. 2014. “Relative Performance of Semi-Parametric Nonlinear Models in Forecasting Basis.” Agricultural and Applied Economics Association.

Payne, N., B. Karali, and J.H. Dorfman. 2019. “Can Cattle Basis Forecasts Be Improved? A Bayesian Model Averaging Approach.” *Journal of Agricultural and Applied Economics.*

Sanders, D.R., and M.R. Manfredo. 2006. “Forecasting Basis Levels in the Soybean Complex: A Comparison of Time Series Methods.” *Journal of Agricultural and Applied Economics* 38: 513- 523.

Sanders, D.J., and T.G. Baker. 2012. “Forecasting Corn and Soybean Basis Using Regime-Switching Models.” Proceedings of the NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management. St. Louis, MO.

Tong, H., 1978. “On a threshold model.” In: Chen, C.H. (Ed.), Pattern Recognition and Signal Processing. Sijthoff and Noordhoff, Amsterdam.

Taylor, P.D., and W.G. Tomek. 1984. “Forecasting the Basis for com in Western New York.” *Journal of Northeast Agricultural Elconomic Council* 13: 97-102.

Teräsvirta, T. 1994. “Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models.” *Journal of the American Statistical Association* 89 (425): 208-218.

Tonsor, G.T., K.C. Dhuyvetter, and J.R. Mintert. 2004. “Improving Cattle Basis Forecasting.” *Journal of Agricultural and Resource Economics* 29(2): 228-241.

Timmermann, A. 2006. “Forecast Combinations.” In Graham Elliott, Clive W.J. Granger and Allan Timmermann (eds.), Handbook of Economic Forecasting. Elsevier, 135-196.

Taylor, M., G. Tonsor, and K. Dhuyvetter 2014. “Structural change in forward contracting costs for Kansas Wheat.” *Journal of Agricultural and Resource Economics* 39 (2): 217-229.

Wilson, W.W., and B. Dahl 2009. “Grain contracting strategies to induce delivery and performance in volatile markets.” *Journal of Agricultural and Applied Economics* 41 (2): 363-376.

Xiong, T., Y. Bao., and Z. Hu 2013. “Beyond One-Step-Ahead Forecasting: Evaluation of Alternative Multi-Step-Ahead Forecasting Models for Crude Oil Prices.” *Energy Economics* 40: 405-415.

Table 1. Contract Months Used in Nearby Futures Series

Month of cash price	Corn futures contract	Soybean futures contract
January	March	March
February	March	March
March	May	May
April	May	May
May	July	July
June	July	July
July	September	August
August	September	September
September	December	November
October	December	November
November	December	January (next year)
December	March (next year)	January (next year)

Table 2. Descriptive Statistics of Prices and Basis

Crop	Price series	Observation	Mean	Standard Deviation	Median	Minimum	Maximum	Skewness	Kurtosis
Corn	Cash price-Candor	1513	371.98	155.91	305.00	187.00	902.00	1.40	1.25
	Cash price-Cofield	1513	354.13	149.70	295.00	178.00	845.00	1.43	1.31
	Cash price-Roaring River	1513	372.99	157.59	305.00	194.00	905.00	1.38	1.12
	Futures price	1513	329.30	142.25	271.75	175.25	830.25	1.53	1.60
	Basis-Candor	1513	42.68	22.19	35.50	-84.47	129.43	0.50	1.77
	Basis-Cofield	1513	24.83	19.30	22.75	-34.25	154.75	0.83	2.80
	Basis-Roaring River	1513	43.69	24.40	39.25	-17.25	210.75	1.24	2.89
Soybean	Cash price-Elizabeth City	1924	750.22	288.53	634.00	379.00	1803.00	1.29	0.85
	Cash price-Fayetteville	1924	778.57	292.35	664.00	418.00	1820.00	1.27	0.74
	Futures price	1924	765.45	282.33	653.00	415.75	1763.25	1.26	0.73
	Basis-Elizabeth City	1924	-15.23	29.43	-15.00	-312.71	190.00	0.21	12.03
	Basis-Fayetteville	1924	13.11	28.12	7.00	-91.00	250.00	2.75	13.81

Table 3. Unit Root Tests of Basis

	Corn			Soybean	
	Candor	Cofield	Roaring River	Elizabeth City	Fayetteville
Augmented Dickey-Fuller Tests					
Zero mean	-3.00***	-4.20***	-2.33**	-12.85***	-10.86***
Single mean	-7.08***	-7.96***	-5.44***	-14.90***	-12.40***
Trend	-9.66***	-8.94***	-9.03***	-14.90***	-13.52***
Phillips-Perron Tests					
Zero mean	-3.56***	-4.88***	-2.82***	-13.89***	-12.51***
Single mean	-8.68***	-9.48***	-6.57***	-16.19***	-14.62***
Trend	-12.12***	-10.72***	-11.35***	-16.21***	-15.90***

Note: The asterisks *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Table 4. AR Estimation Results

Crop	City	Parameters estimate					
		Order selected	Lag				
			1	2	3	4	5
Corn	Candor	5	0.55***	0.08***	0.19***	-0.08***	0.20***
	Cofield	5	0.67***	0.05***	0.24***	-0.23***	0.18***
	Roaring River	5	0.53***	0.08***	0.20***	-0.02***	0.17***
Soybean	Elizabeth City	4	0.67***	0.06***	-0.09***	0.17***	
	Fayetteville	5	0.56***	0.11***	0.06***	0.10***	0.04***

Table 5. SETAR Estimation Results

Crop	City	Regimes		Parameters estimate					
				Intercept	Lag				
					1	2	3	4	5
Corn	Candor	1	71.50%	3.14**	0.60***	0.24***	0.17***	-0.26***	0.16***
		2	28.50%	8.36***	0.44***	-0.08*	0.20***	0.13**	0.19***
	Cofield	1	76.48%	1.94***	0.58***	0.27***	0.07	-0.24***	0.22***
		2	23.52%	11.09***	0.69***	-0.10*	0.39***	-0.27***	0.05
	Roaring River	1	67.45%	2.70**	0.68***	-0.01	0.07	0.00	0.18***
		2	32.55%	5.00***	0.41***	0.13***	0.25***	0.07	0.06
Soybean	Elizabeth City	1	82.32%	-4.22***	0.63***	0.14***	-0.02	0.03	0.00
		2	17.68%	-2.66*	0.73***	-0.03	-0.24***	0.44***	-0.03
	Fayetteville	1	83.11%	1.23**	0.45***	0.11**	0.17***	0.24***	-0.14***
		2	16.89%	8.12***	0.62***	0.08*	-0.02	0.03	0.07*

Note: The order of SETAR process is chosen as 5. The asterisks *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Table 6. STAR Estimation Results

Crop	City	Regime	Estimation						
			γ	c	Linear parameter			Nonlinear parameter	
					1	2	3	1	2
Corn	Candor	1	1.00	22.11	152.08	2.32	0.00		
		2	0.07	21.87	-151.72	-1.46	0.14	0.33	-5.91
		3	38.62	89.92	2798.51	-32.48	-1.12	39.86	84.32
		4	43.15	89.91	-2796.65	32.03	1.55	45.24	83.17
		5	38.77	89.91	140.09	-0.86	-0.65	38.77	89.91
	Cofield	1	1.00	7.76	43.43	2.10	2.35		
		2	46.70	-7.33	-42.05	-1.24	-2.26	88.44	-8.78
		3	24.14	52.38	258.28	-5.72	0.41	9.98	47.40
		4	48.44	52.19	513.79	-11.23	0.90	33.76	48.83
		5	49.25	50.41	-750.40	16.66	-1.39	49.25	50.41
	Roaring River	1	100.00	19.41	1.28	-0.12	1.05		
		2	100.00	19.43	-0.35	0.81	-0.77	103.37	19.81
		3	42.97	86.13	415.30	-4.78	-0.15	53.81	82.21
		4	40.50	85.85	-65.90	-0.41	1.19	40.64	89.10
		5	40.14	85.88	-315.18	5.05	-1.26	40.14	85.88
Soybean	Elizabeth City	1	26.38	-42.18	-42.31	0.34	0.02		
		2	26.47	-42.11	-895.77	-6.61	-1.24	3360.58	-136.93
		3	428.66	15.86	935.08	6.98	1.32	532.99	-105.96
		4	310.51	-106.01	104.47	-0.42	-0.52	562.49	58.59
		5	39.36	92.50	-187.08	0.74	0.44	39.36	92.50
	Fayetteville	1	18.77	36.77	1.94	0.58	0.27		
		2	18.87	36.72	221.18	-5.58	-0.24	20.67	36.92
		3	41.79	39.65	-91472.05	2330.83	-38.85	39.86	39.76
		4	35.69	40.02	32366.57	-434309.01	419660.58	40.64	40.07
		5	40.77	40.07	58894.85	431983.75	-419621.63	40.77	40.07

Table 7. GAM Estimation Results

Crop	City	Linear parameter		Smoothing parameter
		Intercept	Linear(lagged basis)	Loess(lagged basis)
Corn	Candor	5.26***	0.88***	0.18***
	Cofield	3.38***	0.86***	0.15***
	Roaring River	3.49***	0.92***	0.15***
Soybean	Elizabeth City	-3.81***	0.75***	0.62***
	Fayetteville	2.97***	0.78***	0.62***

Note: The asterisks *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Table 8. Forecasting Performance of Models in Corn Markets, RMSE

Location	Model	Horizon					
		1	2	6	12	18	24
Candor	AR	8.04	6.51	4.03	4.45	5.13	5.88
	SETAR	10.28	10.03	6.39	4.59	6.11	5.37
	STAR	1.92	3.95	6.05	8.01	7.48	7.75
	GAM	0.75*	4.45	2.59	2.26	5.08	4.41
	EW	4.87	3.48	2.32	3.17	4.35	4.36
	IN	4.43	3.15	1.94	2.29	4.11	3.90
	IT-AIC	3.73	2.70*	1.77*	1.61*	4.04*	3.58*
Cofield	AR	1.09	3.43*	4.94	4.54	3.87	4.78
	SETAR	0.33*	3.67	6.09	4.73	5.48	5.13
	STAR	2.55	3.48	3.05	5.76	10.08	10.72
	GAM	0.71	5.89	3.55	3.81	3.40	3.17
	EW	1.01	4.03	3.82	3.37	4.40	4.03
	IN	0.67	4.74	3.74	3.51	3.35	3.07
	IT-AIC	3.24	3.53	2.23*	2.93*	2.88*	2.66*
Roaring River	AR	12.31	14.06	15.48	20.38	21.87	22.97
	SETAR	4.90	3.47	2.57	6.32	6.13	6.06
	STAR	3.58	2.78	3.08	4.67	4.10	3.94
	GAM	2.16	5.45	3.68	5.05	4.27	3.77
	EW	3.29	4.88	4.29	7.84	7.58	7.55
	IN	1.66	3.02	2.45*	4.63	3.80	3.31
	IT-AIC	0.95*	2.46*	2.69	4.21*	3.52*	3.06*

Note: The asterisks * denotes best forecast. AR = autoregressive model; SETAR = self-exciting threshold autoregressive model; STAR = smooth-transitioning autoregressive model; GAM = generalized additive model; EW = equally-weighted forecast, IN = inverse mean-squared error forecast; IT-AIC = information-theoretic forecast combinations based on AIC.

Table 9. Forecasting Performance of Models in Corn Markets, MAE

Location	Model	Horizon					
		1	2	6	12	18	24
Candor	AR	8.04	6.27	3.17	3.98	4.59	5.28
	SETAR	10.28	10.03	5.07	3.07	4.01	3.43
	STAR	1.92	3.58	5.75	7.62	7.16	7.45
	GAM	0.75*	3.50	1.40	1.45	2.73	2.19
	EW	4.87	2.78	1.71	2.69	3.46	3.62
	IN	4.43	2.44	1.30*	1.82	2.79	2.82
	IT-AIC	3.73	2.28*	1.34	1.22*	2.41*	2.15*
Cofield	AR	1.09	2.91	4.48	4.08	3.34	4.08
	SETAR	0.33*	2.75*	5.33	3.75	4.70	4.38
	STAR	2.55	3.38	2.81	4.99	8.45	9.42
	GAM	0.71	4.50	2.30	2.40	2.18*	2.16*
	EW	1.01	3.31	3.20	2.73	3.74	3.28
	IN	0.67	3.67	3.01	2.71	2.73	2.53
	IT-AIC	3.24	3.52	1.91*	2.34*	2.43	2.26
Roaring River	AR	12.31	13.96	15.26	19.29	21.04	22.27
	SETAR	4.90	2.61	2.10*	4.07	4.56	4.87
	STAR	3.58	2.60	2.96	3.92	3.37	3.37
	GAM	2.16	4.77	2.93	3.48	2.69	2.36
	EW	3.29	4.68	3.78	6.15	6.43	6.68
	IN	1.66	2.80	2.14	3.22	2.33*	1.87*
	IT-AIC	0.95*	2.15*	2.17	3.13*	2.44	1.94

Note: The asterisks * denotes best forecast. AR = autoregressive model; SETAR = self-exciting threshold autoregressive model; STAR = smooth-transitioning autoregressive model; GAM = generalized additive model; EW = equally-weighted forecast, IN = inverse mean-squared error forecast; IT-AIC = information-theoretic forecast combinations based on AIC.

Table 10. Forecasting Performance of Models in Corn Markets, MAPE

Location	Model	Horizon					
		1	2	6	12	18	24
Candor	AR	13.52	10.22	5.07	5.93	6.87	7.60
	SETAR	17.29	16.16	8.05	4.79	6.34	5.32
	STAR	3.22	**5.65**	8.88	11.17	10.45	10.61
	GAM	**1.25***	**5.44**	**2.18**	**2.16**	**4.07**	**3.24***
	EW	8.19	**4.63**	2.74	3.98	5.22	5.28
	IN	7.44	**4.07**	**2.10***	2.72	4.26	4.16
	IT-AIC	6.26	**3.78***	**2.16**	**1.86***	**3.72***	**3.24***
Cofield	AR	3.71	**8.61**	12.87	10.67	8.62	9.61
	SETAR	**1.11***	**7.95***	15.25	10.23	12.17	10.90
	STAR	8.64	**10.34**	**8.25**	13.06	21.12	22.14
	GAM	2.40	**13.06**	**6.64**	**6.31**	**5.60***	**5.24***
	EW	3.41	**9.71**	**9.24**	**7.51**	9.70	8.27
	IN	2.28	**10.67**	**8.64**	**7.24**	7.09	6.30
	IT-AIC	10.99	**10.91**	**5.74***	**6.26***	**6.30**	**5.62**
Roaring River	AR	15.45	16.91	18.71	22.26	23.47	24.50
	SETAR	6.14	**3.26**	**2.61**	**4.45**	4.91	5.22
	STAR	4.50	**3.20**	**3.67**	**4.47**	3.78	3.73
	GAM	2.70	5.70	**3.56**	**3.92**	**3.00**	2.61
	EW	4.13	5.64	4.58	6.86	7.03	7.24
	IN	2.09	3.36	**2.60***	**3.55**	**2.56***	**2.06***
	IT-AIC	**1.20***	**2.57***	**2.68**	**3.53***	**2.72**	**2.16**

Note: The asterisks * denotes best forecast. AR = autoregressive model; SETAR = self-exciting threshold autoregressive model; STAR = smooth-transitioning autoregressive model; GAM = generalized additive model; EW = equally-weighted forecast, IN = inverse mean-squared error forecast; IT-AIC = information-theoretic forecast combinations based on AIC.

Table 11. Forecasting Performance of Models in Soybean Markets, RMSE

Location	Model	Horizon					
		1	2	6	12	18	24
Elizabeth City	AR	58.99	42.70	28.20	21.83	19.05	17.62
	SETAR	60.76	43.80	28.80	22.57	19.13	17.19
	STAR	59.09	42.36	27.04	20.94	18.35	17.00
	GAM	56.14	46.55	27.01	21.49	17.67	15.39
	EW	58.75	41.54	26.10	20.24	17.26	15.55
	IN	58.59	41.46	25.86	20.06	17.06	15.32
	IT-AIC	6.76*	11.70*	9.33*	9.42*	8.31*	7.60*
Fayettville	AR	41.00	29.24	19.75	16.80	14.32	13.03
	SETAR	41.81	30.04	19.85	16.91	14.33	12.63
	STAR	40.85	29.23	19.80	16.80	14.27	12.90
	GAM	43.86	37.47	26.80	21.91	18.38	16.08
	EW	41.88	29.67	20.36	17.04	14.37	12.77
	IN	41.69	29.48	20.08	16.88	14.25	12.70
	IT-AIC	3.20*	2.34*	8.15*	10.49*	9.07*	8.27*

Note: The asterisks * denotes best forecast. AR = autoregressive model; SETAR = self-exciting threshold autoregressive model; STAR = smooth-transitioning autoregressive model; GAM = generalized additive model; EW = equally-weighted forecast, IN = inverse mean-squared error forecast; IT-AIC = information-theoretic forecast combinations based on AIC.

Table 12. Forecasting Performance of Models in Soybean Markets, MAE

Location	Model	Horizon					
		1	2	6	12	18	24
Elizabeth City	AR	58.99	35.96	23.12	17.42	15.19	14.39
	SETAR	60.76	36.40	23.32	17.84	14.45	13.01
	STAR	59.09	34.46	20.83	16.11	14.29	13.63
	GAM	56.14	45.26	17.11	14.54	10.65	8.60
	EW	58.75	29.43	18.14	14.71	12.37	11.30
	IN	58.59	30.30	18.02	14.59	12.18	11.02
	IT-AIC	6.76*	10.93*	8.46*	8.37*	7.27*	6.61*
Fayettville	AR	41.00	23.16	14.29	12.38	9.91	9.44
	SETAR	41.81	24.66	14.02	12.63	10.23	8.80
	STAR	40.85	23.60	14.19	12.39	9.88	9.28
	GAM	43.86	36.80	22.98	17.96	13.49	11.28
	EW	41.88	22.26	14.71	12.78	9.87	8.97
	IN	41.69	20.88	13.90	12.34	9.67	8.84
	IT-AIC	3.20*	2.02*	7.09*	8.70*	7.27*	6.32*

Note: The asterisks * denotes best forecast. AR = autoregressive model; SETAR = self-exciting threshold autoregressive model; STAR = smooth-transitioning autoregressive model; GAM = generalized additive model; EW = equally-weighted forecast, IN = inverse mean-squared error forecast; IT-AIC = information-theoretic forecast combinations based on AIC.

Table 13. Forecasting Performance of Models in Soybean Markets, MAPE

Location	Model	Horizon					
		1	2	6	12	18	24
Elizabeth City	AR	122.36	**204.99**	220.97	151.01	111.76	94.96
	SETAR	126.05	**196.78**	214.52	151.91	109.11	89.82
	STAR	122.59	**170.56**	**180.92**	129.29	97.20	83.72
	GAM	116.46	**440.20**	**175.92**	**114.92**	**79.58**	**61.98**
	EW	121.87	**62.15***	**133.55**	**103.00**	**76.59**	**64.88**
	IN	121.50	**83.03**	**135.02**	**102.81**	**76.08**	**64.02**
	IT-AIC	**14.02***	**174.87**	**132.94***	**92.08***	**66.64***	**54.26***
Fayettville	AR	58.86	**38.46**	**479.17**	**539.30**	**430.39**	**514.41**
	SETAR	60.02	42.75	**314.56***	**607.46**	**560.46**	**460.42**
	STAR	58.65	40.07	**428.31**	**544.80**	**440.12**	**496.35**
	GAM	62.97	81.87	**898.75**	**689.69**	**526.81**	**597.74**
	EW	60.13	**34.53**	**523.90**	**531.59***	**419.56***	**466.40**
	IN	59.85	**30.05**	**485.31**	**540.52**	**435.81**	**471.56**
	IT-AIC	**4.60***	**3.72***	**562.51**	**645.61**	**568.67**	**444.01***

Note: The asterisks * denotes best forecast. AR = autoregressive model; SETAR = self-exciting threshold autoregressive model; STAR = smooth-transitioning autoregressive model; GAM = generalized additive model; EW = equally-weighted forecast, IN = inverse mean-squared error forecast; IT-AIC = information-theoretic forecast combinations based on AIC.

Table 14. MDM Test with the Best Model, 2-week Horizon

Crop	Location	Model						
		AR	SETAR	STAR	GAM	EW	IN	IT-AIC
Corn	Candor	-7.88***	-9.11***	-0.38	-0.25	-0.79	-0.33	/
	Cofield	-0.34	/	-0.46	-1.34	-1.30	-1.76	-0.43
	Roaring River	-158.72***	-0.16	-0.24	-8.93***	-21.87***	-8.05***	/
Soybeans	Elizabeth City	-1.00	-1.03	-1.01	-0.99	/	-0.98	-0.51
	Fayetteville	-1.78	-2.38*	-2.05*	-3.95***	-1.25	-0.91	/

Note: Loss function is based on absolute percentage error. The asterisks *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively. AR = autoregressive model; SETAR = self-exciting threshold autoregressive model; STAR = smooth-transitioning autoregressive model; GAM = generalized additive model; EW = equally-weighted forecast, IN = inverse mean-squared error forecast.

Table 15. MDM Test with the Best Model, 6-week Horizon

Crop	Location	Model						
		AR	SETAR	STAR	GAM	EW	IN	IT-AIC
Corn	Candor	-2.60**	-2.53**	-3.01**	-0.04	-1.60*	/	-0.14
	Cofield	-1.92**	-1.95**	-1.05	-0.32	-1.18	-1.01	/
	Roaring River	-9.66***	0.00	-1.10	-1.21	-1.62*	/	-0.11
Soybeans	Elizabeth City	-3.18***	-2.66**	-1.36	-0.50	-0.01	-0.03	
	Fayetteville	-1.02	/	-1.01	-1.24	-1.37	-1.40	-1.05

Note: Loss function is based on absolute percentage error. The asterisks *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively. AR = autoregressive model; SETAR = self-exciting threshold autoregressive model; STAR = smooth-transitioning autoregressive model; GAM = generalized additive model; EW = equally-weighted forecast, IN = inverse mean-squared error forecast.

Table 16. MDM Test with the Best Model, 12-week Horizon

Crop	Location	Model						
		AR	SETAR	STAR	GAM	EW	IN	IT-AIC
Corn	Candor	-5.31***	-2.05**	-7.10***	-0.33	-3.35***	-2.22**	/
	Cofield	-1.96**	-1.39*	-2.69***	-0.03	-0.74	-0.66	/
	Roaring River	-13.26***	-0.86	-0.98	-0.56	-2.97***	-0.06	/
Soybeans	Elizabeth City	-3.36***	-3.22***	-2.02**	-0.53	-0.33	-0.37	/
	Fayetteville	-0.11	-0.35	-0.13	-0.79	/	-0.23	-0.69

Note: Loss function is based on absolute percentage error. The asterisks *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively. AR = autoregressive model; SETAR = self-exciting threshold autoregressive model; STAR = smooth-transitioning autoregressive model; GAM = generalized additive model; EW = equally-weighted forecast, IN = inverse mean-squared error forecast.

Table 17. MDM Test with the Best Model, 18-week Horizon

Crop	Location	Model						
		AR	SETAR	STAR	GAM	EW	IN	IT-AIC
Corn	Candor	-3.44***	-1.94**	-4.29***	-0.38	-2.63***	-1.76**	/
	Cofield	-1.72*	-2.78***	-4.02***	/	-2.26***	-1.48*	-0.61
	Roaring River	-18.99***	-3.18***	-2.19**	-0.88	-6.79***	/	-0.48
Soybeans	Elizabeth City	-3.60***	-3.11***	-2.47**	-0.46	-0.46	-0.49	/
	Fayetteville	-0.23	-0.91	-0.31	-0.78	/	-0.61	-1.27

Note: Loss function is based on absolute percentage error. The asterisks *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively. AR = autoregressive model; SETAR = self-exciting threshold autoregressive model; STAR = smooth-transitioning autoregressive model; GAM = generalized additive model; EW = equally-weighted forecast, IN = inverse mean-squared error forecast.

Table 18. MDM Test with the Best Model, 24-week Horizon

Crop	Location	Model						
		AR	SETAR	STAR	GAM	EW	IN	IT-AIC
Corn	Candor	-5.14***	-2.01**	-6.10***	0.01	-4.13***	-3.11***	/
	Cofield	-2.97***	-2.99***	-5.64***	/	-1.98**	-1.32*	-0.43
	Roaring River	-22.48***	-5.03***	-3.65***	-1.48**	-9.33***	/	-0.39
Soybeans	Elizabeth City	-4.30***	-3.40***	-3.20***	-0.36	-0.66	-0.68	/
	Fayetteville	-0.58	-0.20	-0.49	-0.63	-0.18	-0.25	/

Note: Loss function is based on absolute percentage error. The asterisks *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively. AR = autoregressive model; SETAR = self-exciting threshold autoregressive model; STAR = smooth-transitioning autoregressive model; GAM = generalized additive model; EW = equally-weighted forecast, IN = inverse mean-squared error forecast.

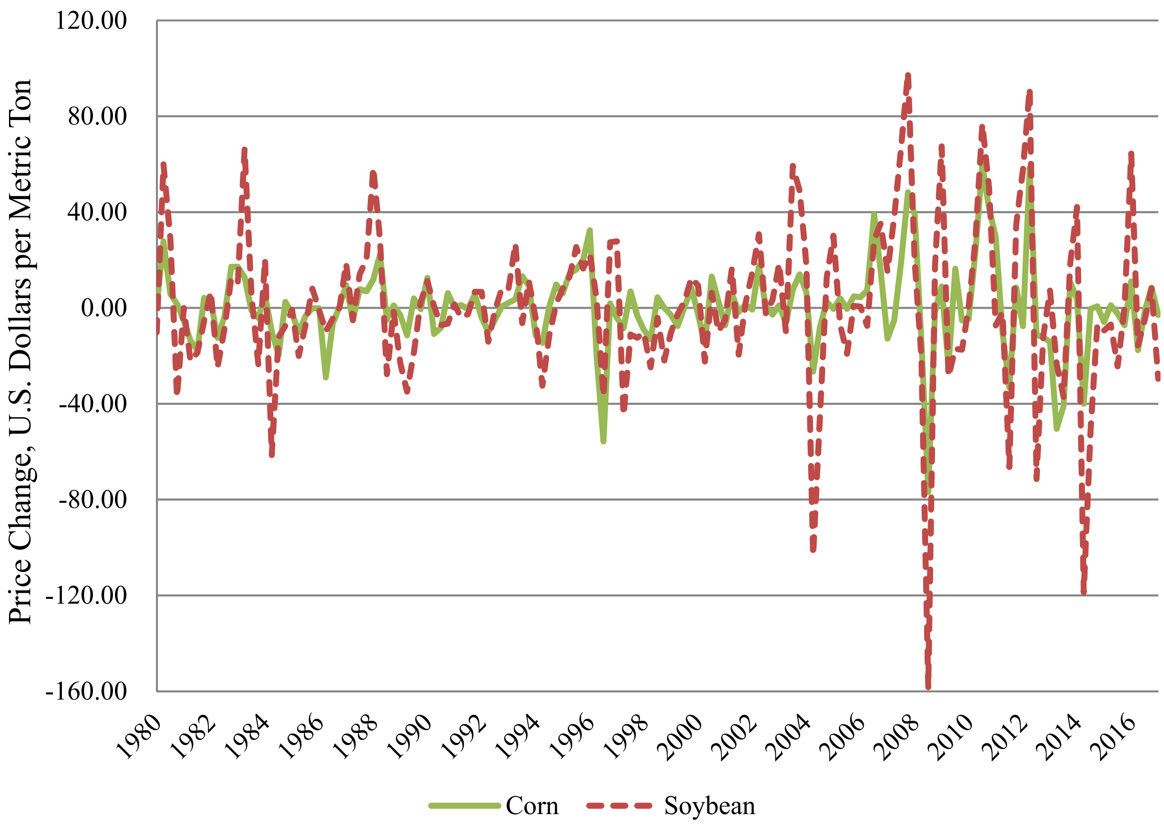


Figure 1. Changes in global corn and soybean prices.
Source: International Monetary Fund

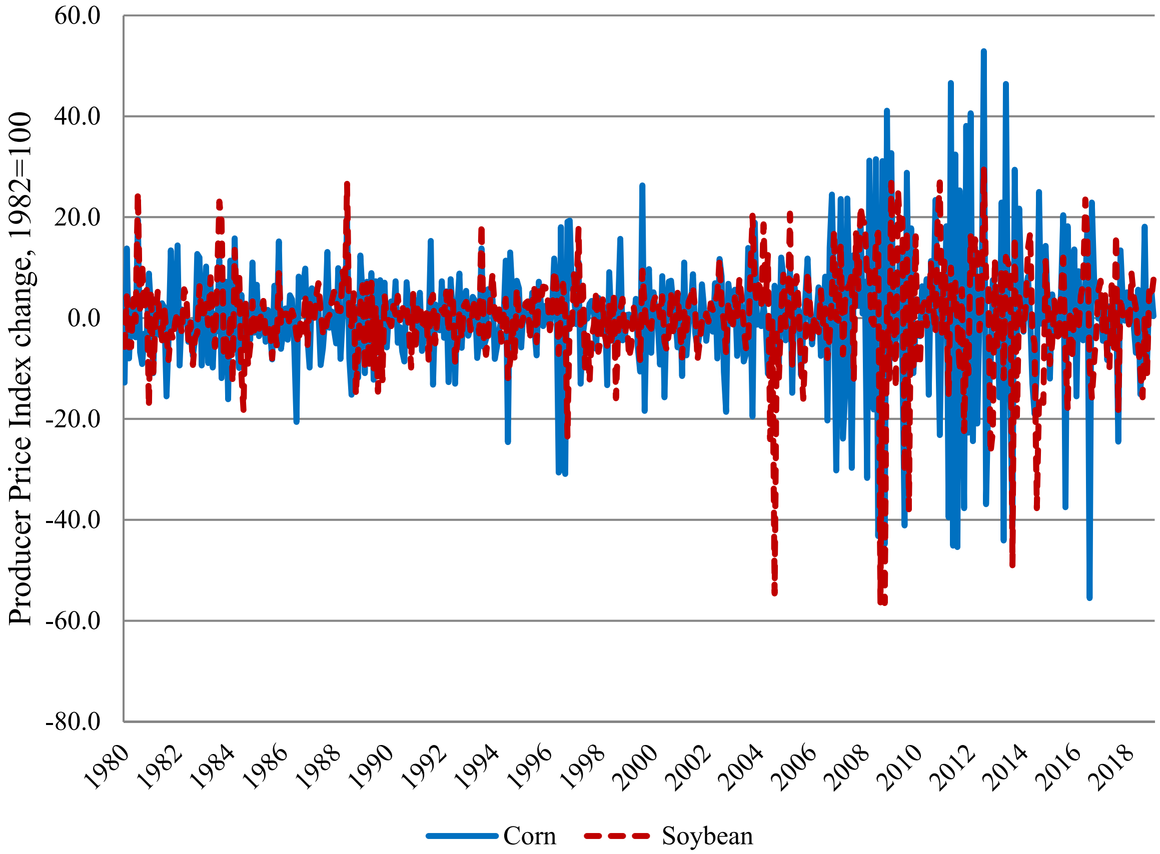


Figure 2. Changes in U.S. corn and soybean producer price indices.
Source: U.S. Bureau of Labor Statistics

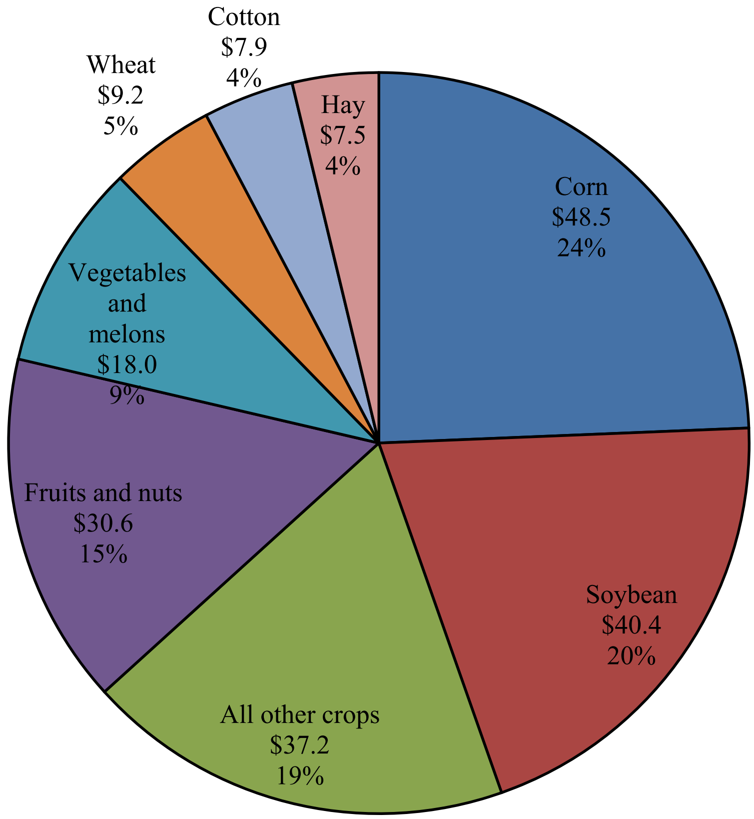
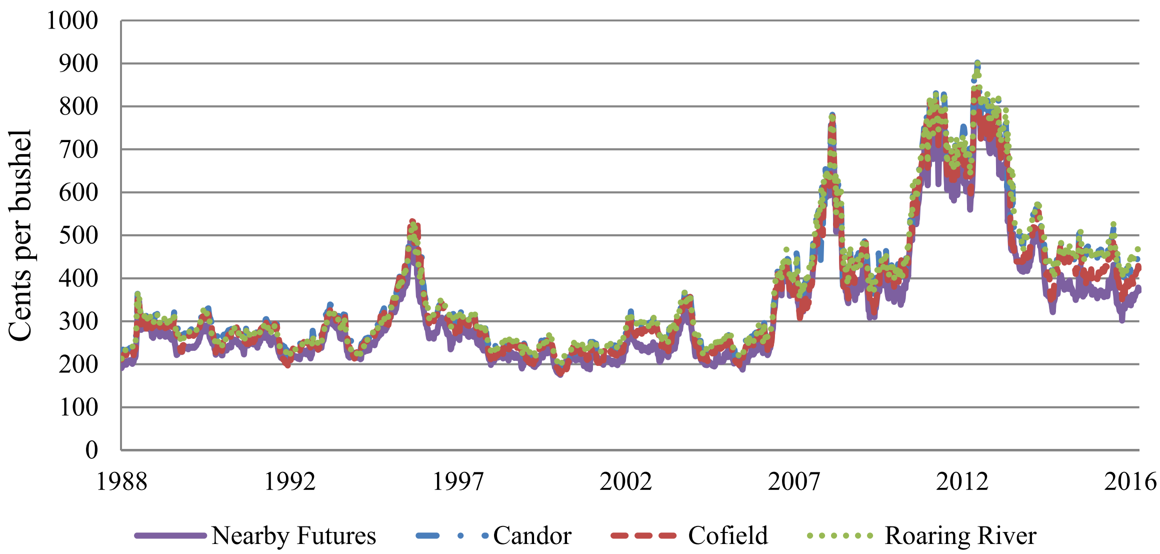
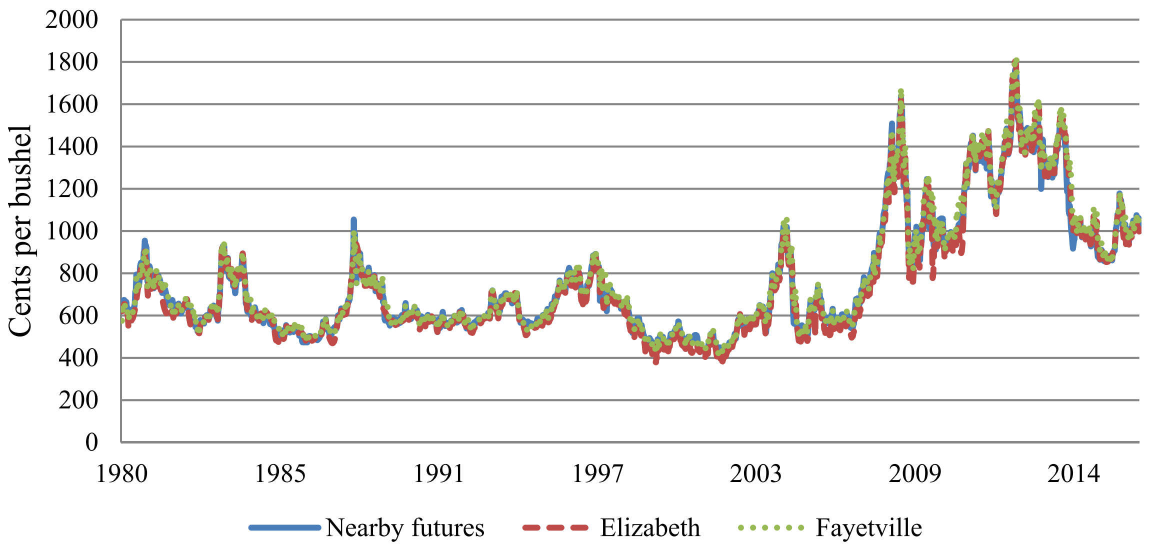


Figure 3. Crop cash receipts ($billion and %), 2018.
Source: USDA, Economic Research Service

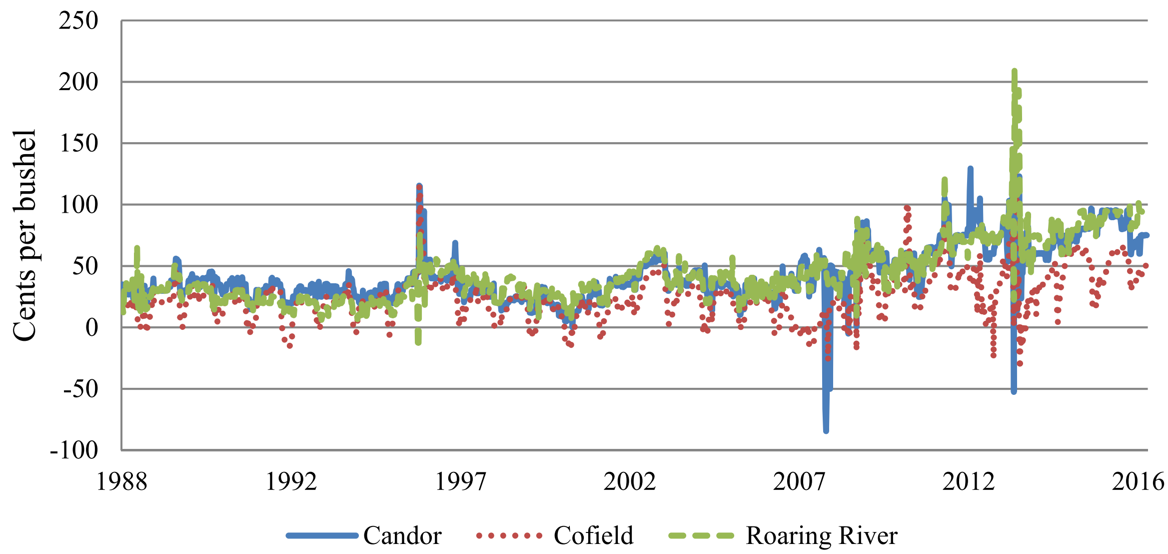


(a) Corn

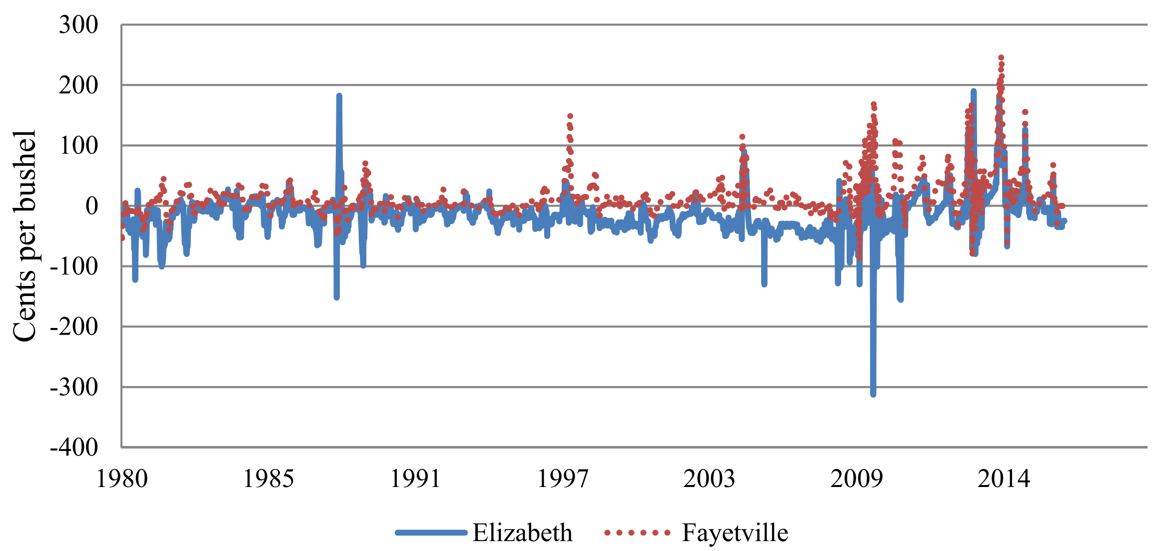


(b) Soybean

Figure 4. Cash and nearby futures prices.

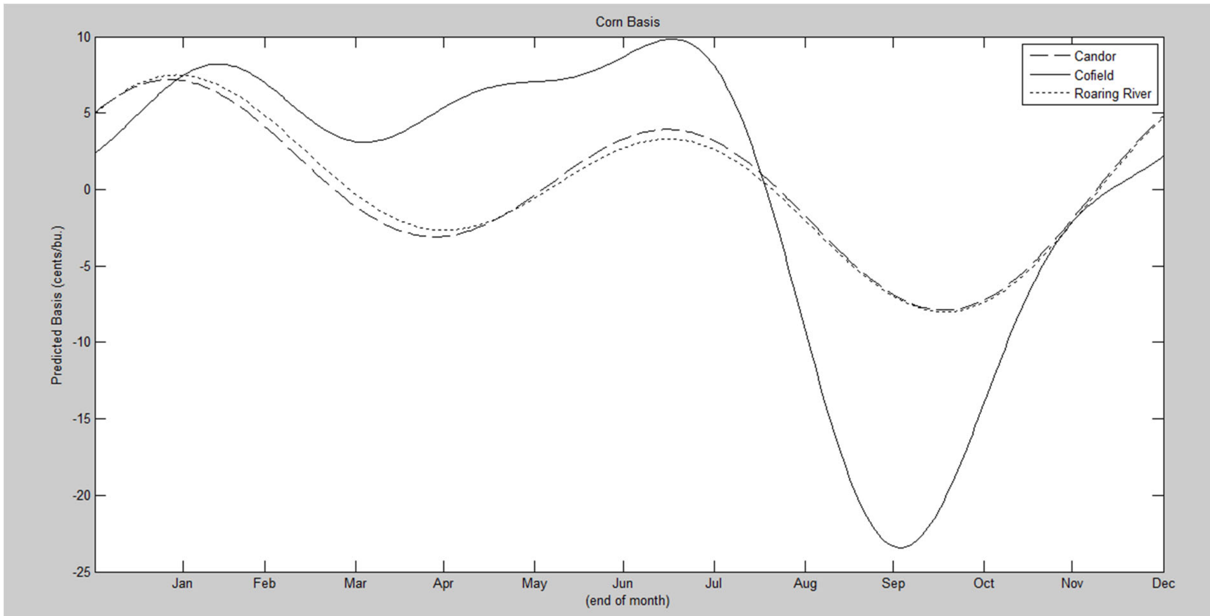


(a) Corn

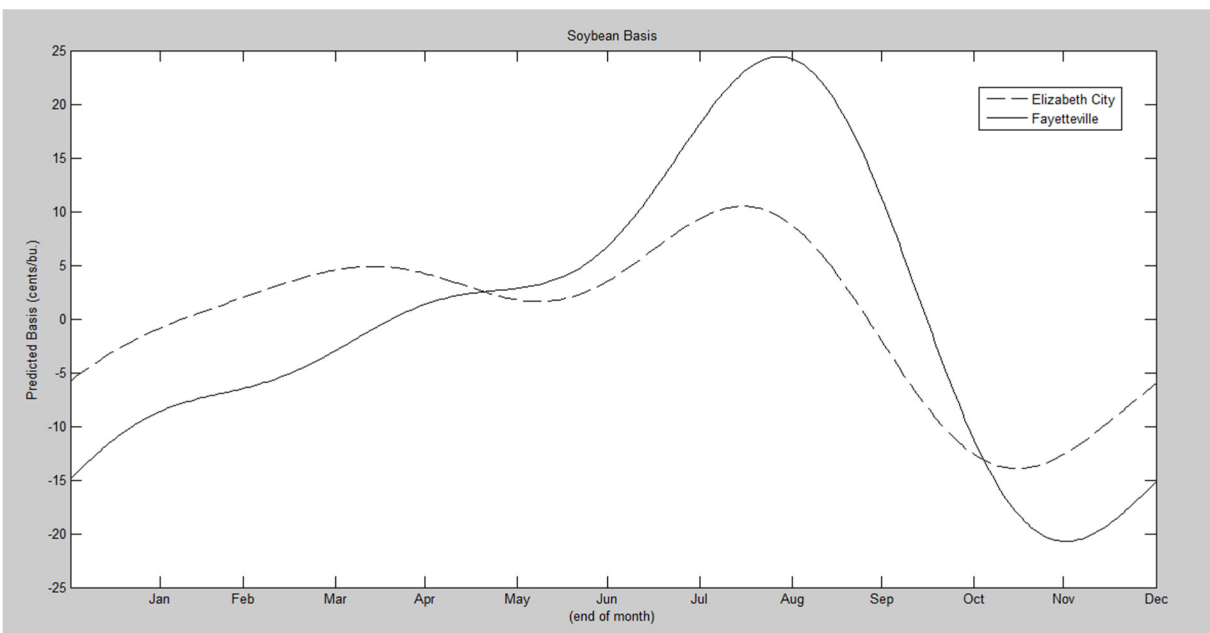


(b) Soybean

Figure 5. Basis in corn and soybean markets.

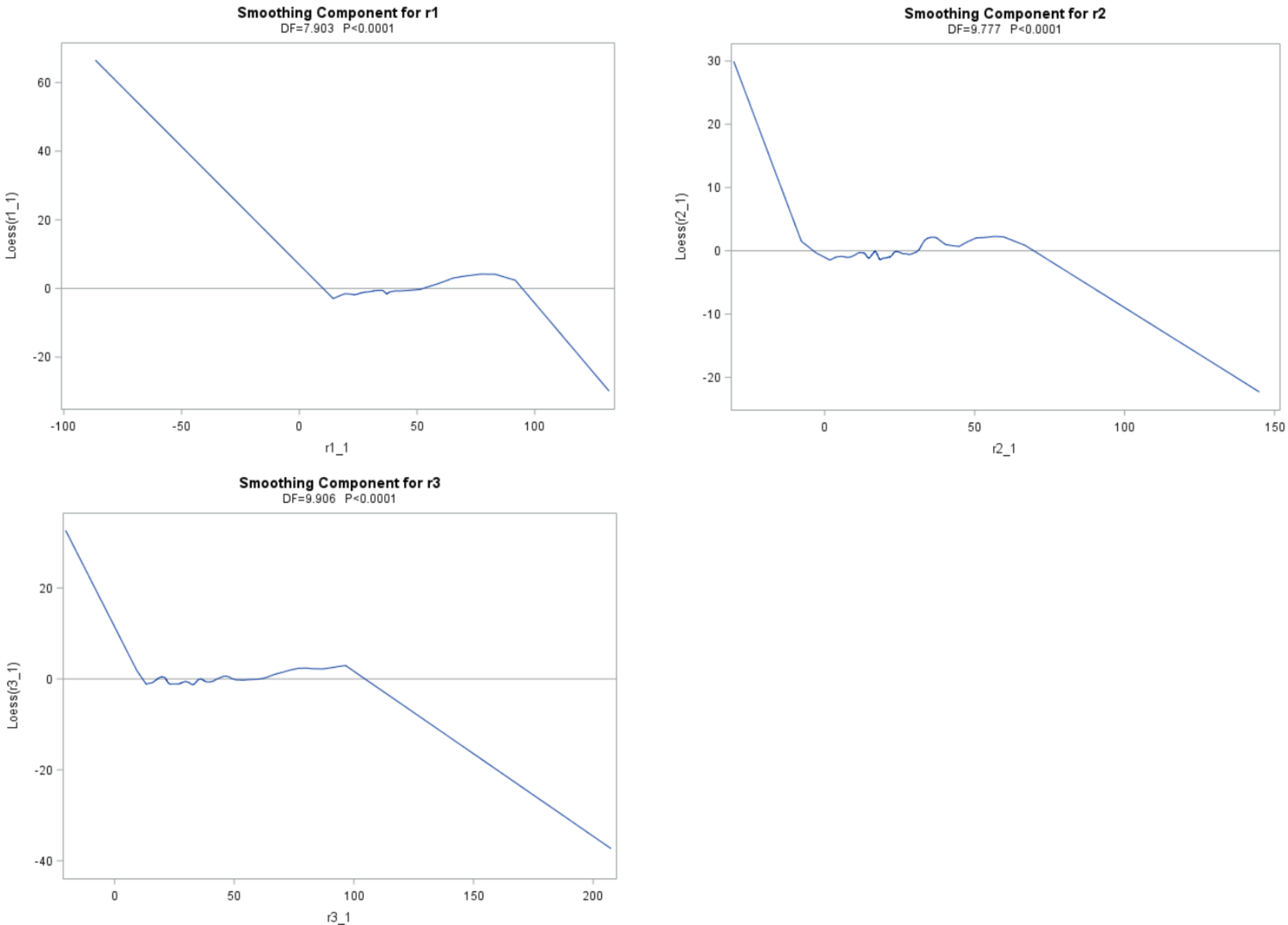


(a) Corn

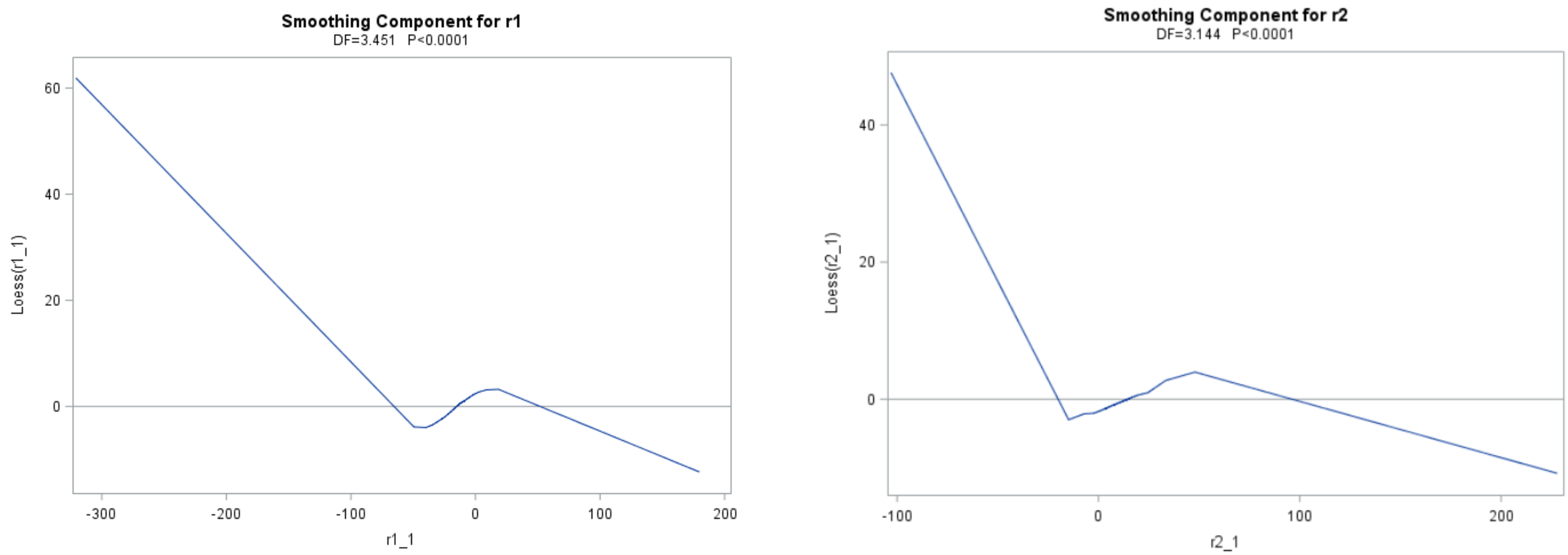


(b) Soybean

Figure 6. Predicted basis with sinusoidal polynomial functions.


Smoothing Component for r1
DF=7.903 P<0.0001
Loess(r1_1)
r1_1
Smoothing Component for r2
DF=9.777 P<0.0001
Loess(r2_1)
r2_1
Smoothing Component for r3
DF=9.906 P<0.0001
Loess(r3_1)
r3_1

(a) Corn



(b) Soybean

Figure 7. Nonlinear effects of lagged basis.

Note: In corn markets, r1, r2, r3 denote Candor, Cofield, and Roaring River, respectively. In soybean markets, r1, r2 denote Fayetteville, and Elizabeth City, respectively.

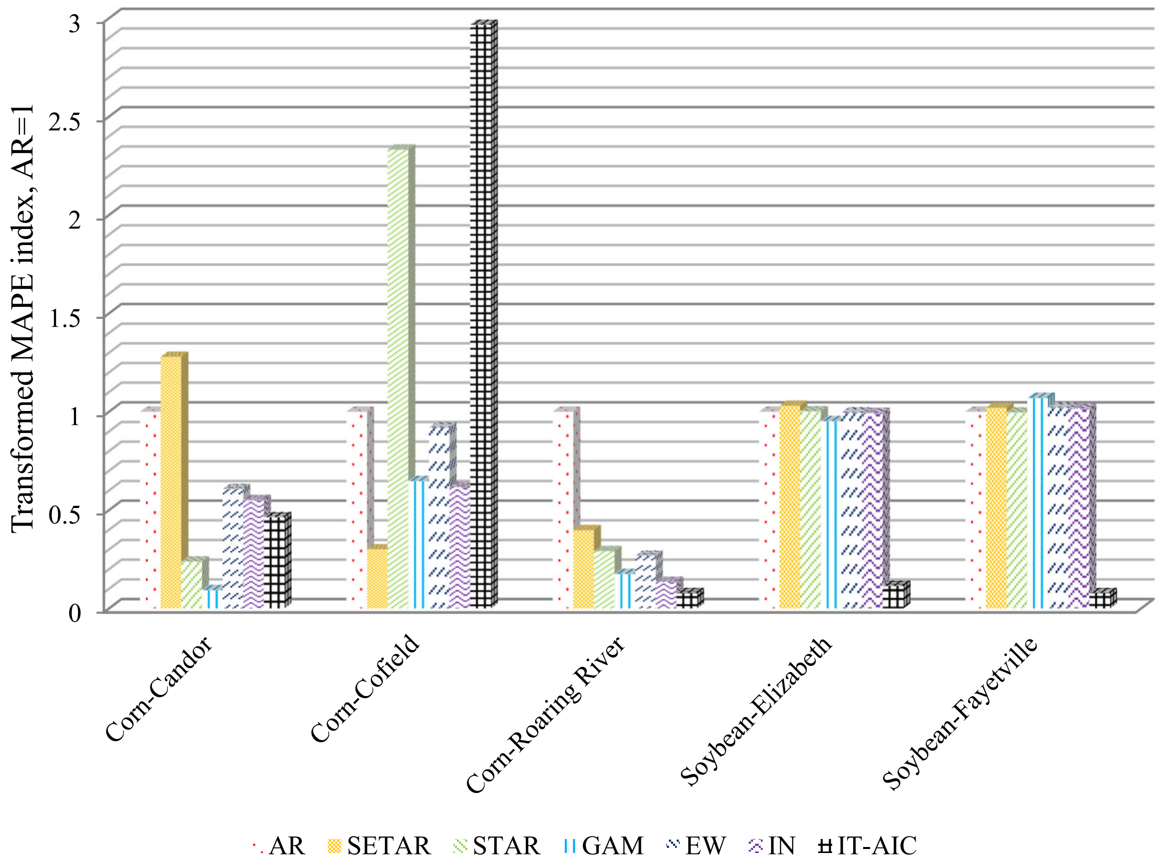


Figure 8. Relative performance of forecasting models, 1-week horizon.
Note: AR = autoregressive model; SETAR = self-exciting threshold autoregressive model; STAR = smooth-transitioning autoregressive model; GAM = generalized additive model; EW = equally-weighted forecast, IN = inverse mean-squared error forecast; IT-AIC = information-theoretic forecast combinations based on AIC.

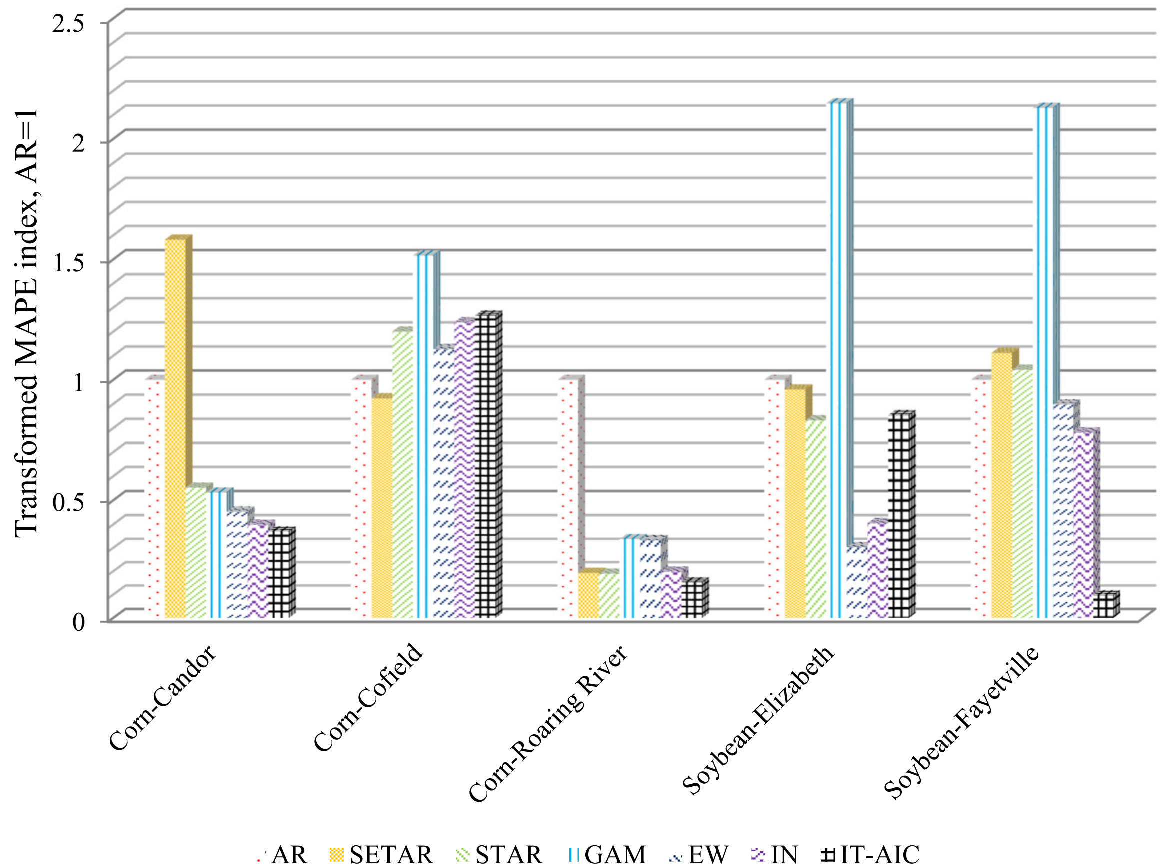


Figure 9. Relative performance of forecasting models, 2-week horizon.
Note: AR = autoregressive model; SETAR = self-exciting threshold autoregressive model; STAR = smooth-transitioning autoregressive model; GAM = generalized additive model; EW = equally-weighted forecast, IN = inverse mean-squared error forecast; IT-AIC = information-theoretic forecast combinations based on AIC.

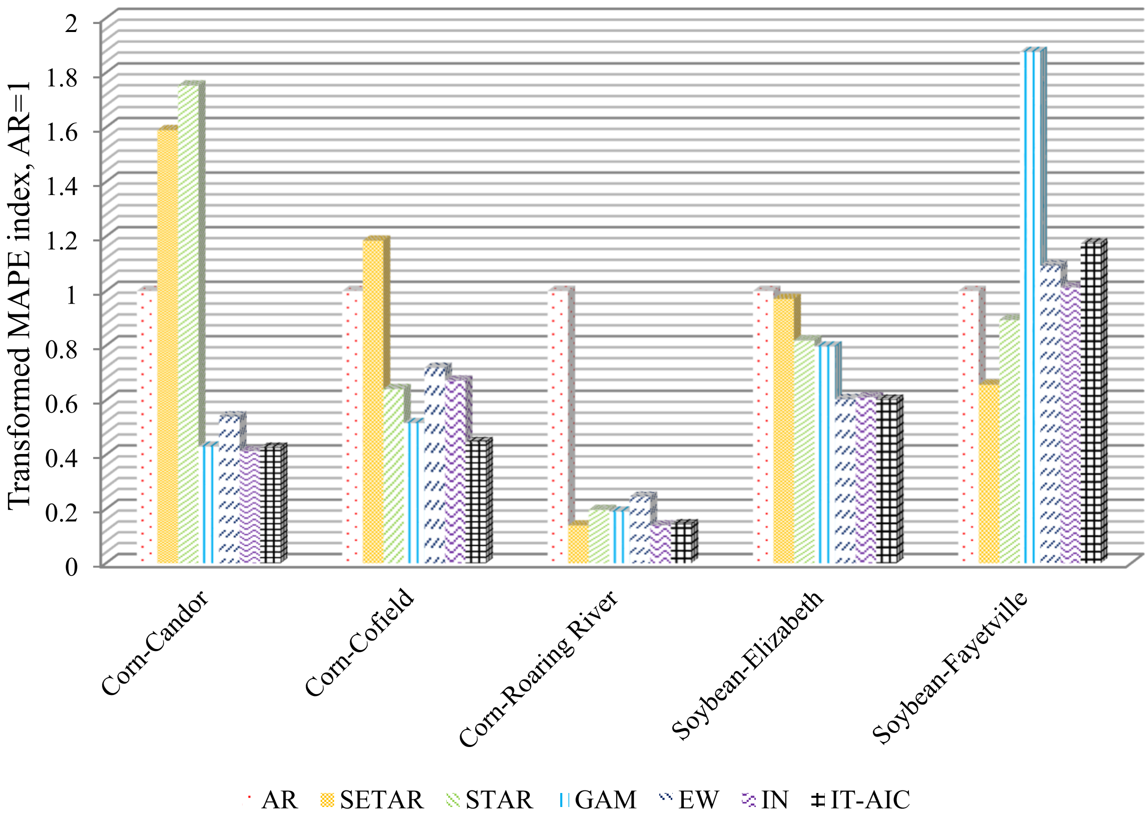


Figure 10. Relative performance of forecasting models, 6-week horizon.
Note: AR = autoregressive model; SETAR = self-exciting threshold autoregressive model; STAR = smooth-transitioning autoregressive model; GAM = generalized additive model; EW = equally-weighted forecast, IN = inverse mean-squared error forecast; IT-AIC = information-theoretic forecast combinations based on AIC.

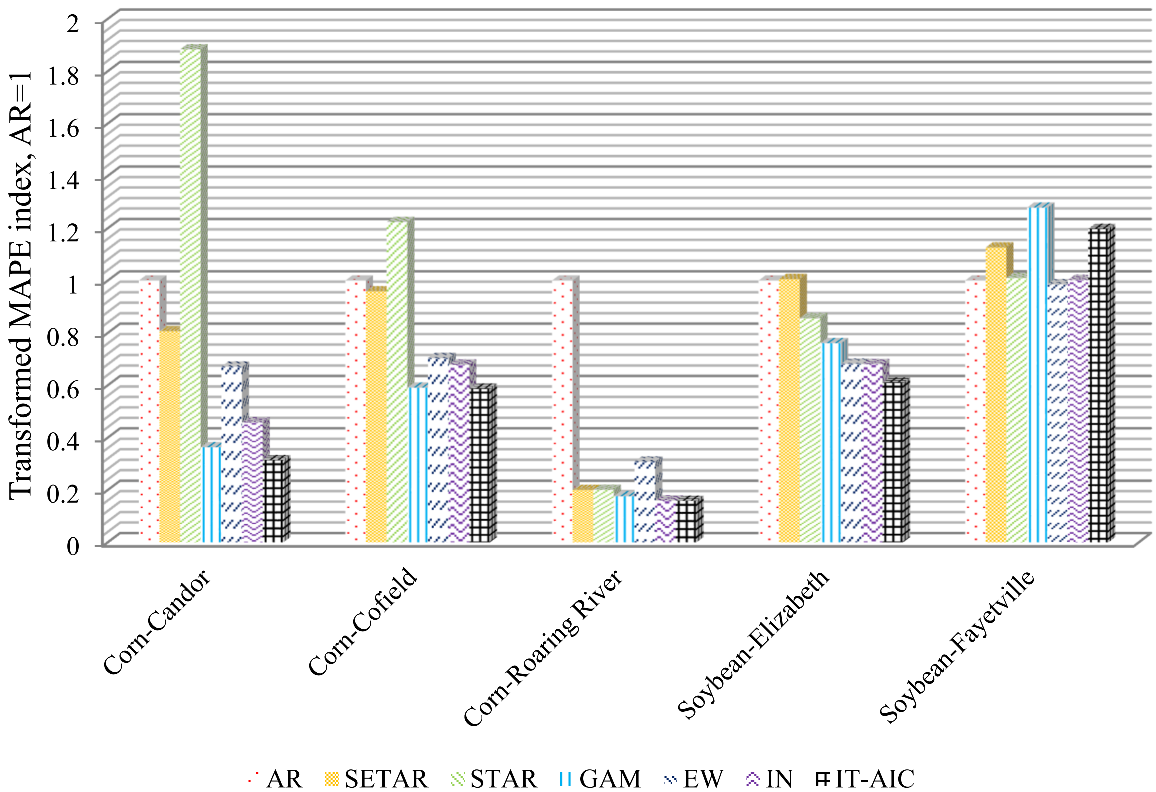


Figure 11. Relative performance of forecasting models, 12-week horizon.
Note: AR = autoregressive model; SETAR = self-exciting threshold autoregressive model; STAR = smooth-transitioning autoregressive model; GAM = generalized additive model; EW = equally-weighted forecast, IN = inverse mean-squared error forecast; IT-AIC = information-theoretic forecast combinations based on AIC.

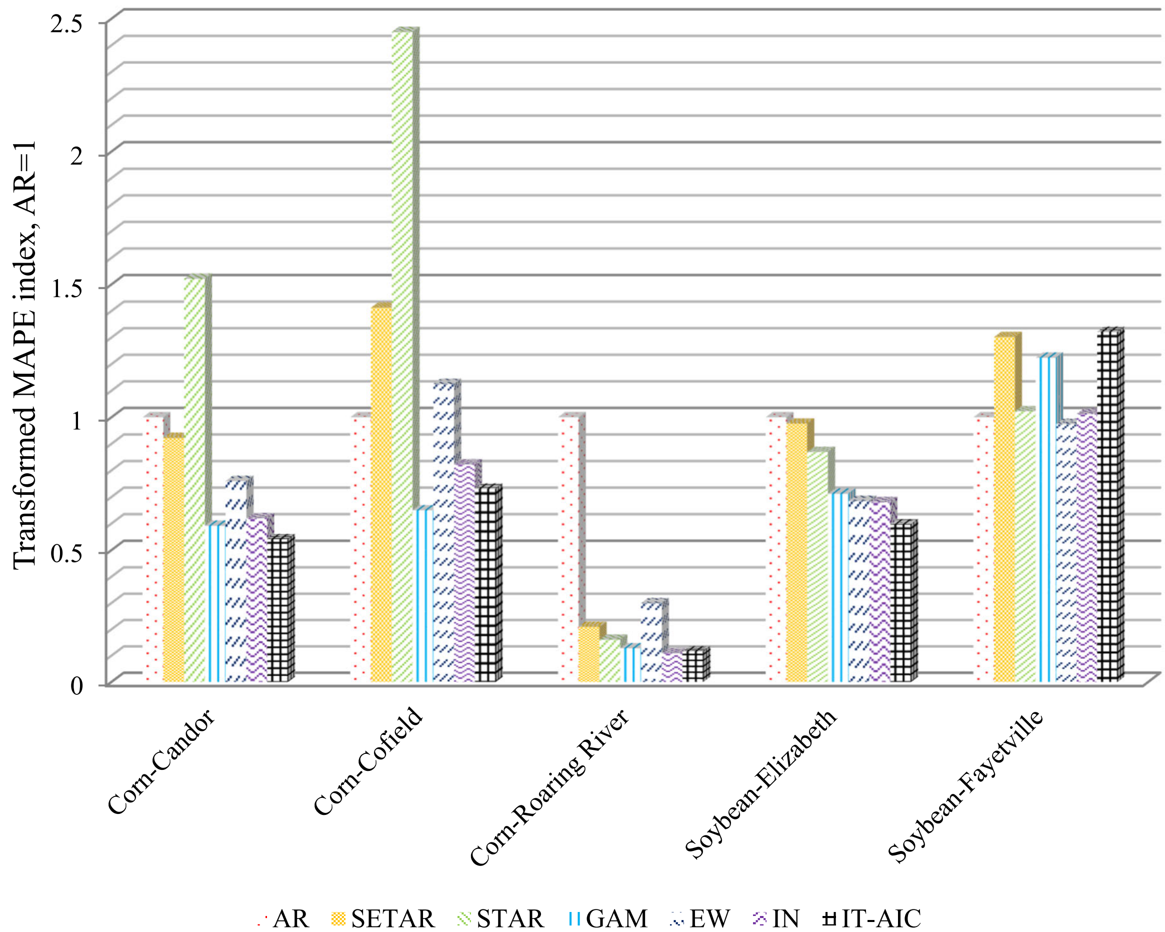


Figure 12. Relative performance of forecasting models, 18-week horizon.
Note: AR = autoregressive model; SETAR = self-exciting threshold autoregressive model; STAR = smooth-transitioning autoregressive model; GAM = generalized additive model; EW = equally-weighted forecast, IN = inverse mean-squared error forecast; IT-AIC = information-theoretic forecast combinations based on AIC.

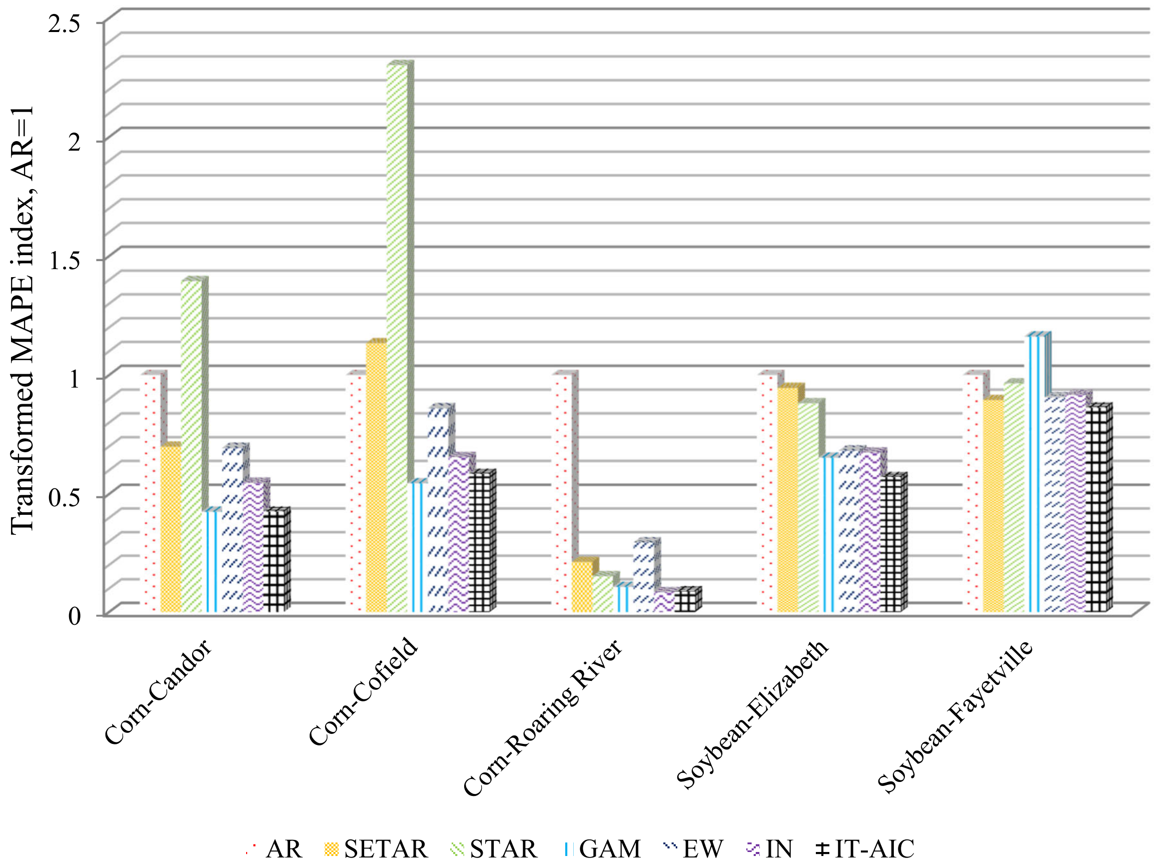


Figure 13. Relative performance of forecasting models, 24-week horizon.
Note: AR = autoregressive model; SETAR = self-exciting threshold autoregressive model; STAR = smooth-transitioning autoregressive model; GAM = generalized additive model; EW = equally-weighted forecast, IN = inverse mean-squared error forecast; IT-AIC = information-theoretic forecast combinations based on AIC.