NCCC-134

APPLIED COMMODITY PRICE ANALYSIS, FORECASTING AND MARKET RISK MANAGEMENT

Can Cattle Basis Forecasts Be Improved? A Bayesian Model Averaging Approach

by

Nicholas Payne and Berna Karali

Can Cattle Basis Forecasts Be Improved?
A Bayesian Model Averaging Approach

Nicholas Payne

and

Berna Karali

Paper presented at the NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management
St. Louis, Missouri, April 21-22, 2015

Graduate Student and Associate Professor in the Department of Agricultural and Applied Economics at The University of Georgia.

Can Cattle Basis Forecasts Be Improved?
A Bayesian Model Averaging Approach

Basis forecasts aid producers and consumers of agricultural commodities in price risk management. A simple historical moving average of nearby basis on a specific date is the most common forecast approach; however, in previous evaluations of forecast methods, the best prediction of basis has often been inconsistent. The best forecast also differs with respect to commodity and forecast horizon. Given this inconsistency, a Bayesian approach which addresses model uncertainty by combining forecasts from different models is taken. Various regression models are considered for combination, and simple moving averages are evaluated for comparison. We find that model performance differs by location and forecast horizon, but the average model typically performs favorably compared to regression models. However, except for very short-horizon forecasts, the simple moving averages have a lower out of sample forecast error than the regression models. We also examine using a basis series created using a specific month's futures contract as opposed to the nearby contract and find that basis forecasts calculated this way have lower forecast errors in the month of the contract examined.

Keywords: basis, basis forecasts, Bayesian model averaging, feeder cattle, futures, Georgia.

Introduction

Basis, the difference between the spot and futures prices, localizes the futures price for a given location and time and is therefore a key component of price risk management (Tomek, 1997; Taylor, Dhuyvetter and Kastens, 2006). When compared to current basis levels, historical basis puts current price levels in perspective; so if current basis deviates significantly from historical levels, market participants can address why by inquiring into local market circumstances. Having accurate basis forecasts aids in hedging with futures contracts by providing an estimate of the prospective price they will receive or pay if they enter into a futures contract. Accurate forecast of basis is also important in timing sales and purchases of commodities, as it generates a forecast of expected future cash prices (Tonsor, Dhuyvetter, and Mintert, 2004).

Given the importance of accurate basis information, we investigate in this paper if basis forecasts can be improved by combining different forecasting models in a Bayesian framework. Specifically, we estimate feeder cattle basis in three different Georgia locations using both nearby and September futures contracts.

The earliest studies examining basis forecasts used structural models to forecast basis. For instance, storage and transportation costs, excess local supply, existing stocks, and seasonality are found to affect basis (Martin, Groenewegen, and Pidgeon, 1980; Garcia and Good, 1983). Moving historical averages, commonly referred to as naïve forecasts, have been the most typical approach to forecast basis (Hauser, Garcia, and Tumblin, 1990). When compared to structural models, the naïve models are generally found to predict basis more accurately. A regression on structural variables such as transportation costs and local supply information as regressors along with a naïve forecast is found to somewhat improve on the performance of the best naïve model (Jiang and Hayenga, 1997).

The number of years to include into a moving average has been a key issue in the literature. Looking at wheat, corn, milo, and soybeans, Dhuyvetter and Kastens (1998), for instance, find that longer moving averages work best, whereas Taylor, Dhuyvetter, and Kastens (2006), looking at the same commodities, find shorter moving averages perform better. Hatchett, Brorsen, and Anderson (2010) conclude that while using shorter moving averages yields better basis forecasts when a market has undergone a structural change, using longer moving averages performs better otherwise.

Compared to naïve models, more complex time-series models such as ARIMA and VAR models modestly outperform naïve approaches, but only for short forecast horizons (Sanders and Manfredo, 2006; Sanders and Baker, 2012). Additionally, some studies indicate basis may be getting harder to predict over time (Irwin et al., 2009). This observation has led to recent investigations of regime switching models (Sanders and Baker, 2012); however these too only perform well over short forecast horizons.

Our paper reexamines many variations of the models in previous work and gauges their effectiveness at forecasting basis. Given the ambiguity in the literature on the best forecasting model of basis and the importance of accurate forecasts of basis for market participants, we take a Bayesian Model Averaging (BMA) approach. By averaging over different competing forecasting models, BMA incorporates model uncertainty into inferences about parameters and has been shown to improve predictive performance in many instances (Hoeting et al., 1999; Koop, 2003). The main focus of this paper is on evaluating the potential of BMA as a method for producing a consistently performing forecasts and potentially providing an improvement over current methods of basis forecasts. For this purpose, we consider variations of previously examined models in our BMA framework. Specifically, autoregressive models and regressions on moving historical averages of varying lengths with and without additional fundamental variables such as current market information, transportation costs, and input costs are considered as explanatory variables. Each selected model's forecasting performance are compared to that of naïve forecasts and an average forecast obtained with estimated Bayesian model probabilities.

Most previous studies of basis forecasts do so with a nearby basis series: basis in a given month is computed with futures contracts that correspond to that month or the nearest months' contract. This kind of forecast concerns itself primarily with what basis will be in the near future, specifically at the time the nearby contract expires. However, a producer may be also interested in what basis will be in the far distant future, when sale or purchase of products will occur. This type of forecast is most likely to be used by an agricultural producer who has a fixed timeline for making marketing decisions. Thus, We deviate from the convention of computing basis with the nearby futures price and examine the potential of using a series consisting entirely of futures contracts from a particular month to improve forecasts made for that month at any time of the year. We consider the September futures contract because most feeder cattle in Georgia are sold in August for September delivery to feedlots in the West.

Results of nearby basis forecasts for individual models are similar to those of previous studies: autoregressive models perform well for short horizon forecasts, but for longer horizon forecasts a multiyear historical moving average performs better than regressions. The best performing individual regression model varies by location and forecast horizon, but at all locations and forecast horizons considered, the BMA forecasting model performs well comparably to the best individual models considered for averaging. However, BMA never performs better than the best naïve forecast except for the shortest forecast horizon considered.

For a one week ahead forecast of September basis, autoregressive models, a naïve forecast consisting of the last observed September basis, and the average model all perform best and quite similarly. As with the nearby series, a multiyear historical average of September basis becomes the front runner for the longer forecasts examined. Unlike with the nearby basis forecasts, the best regression models computed with the September series tend to perform noticeably better than the average model.

Previous Work

Dating back several decades, early empirical studies examining basis for agricultural commodities examine fundamental supply and demand factors believed to affect basis as set forth in the theory of storage by Working (1949). This theory states that carrying and delivery charges, existing stocks, and flows of commodities to or from outside markets determine the difference between the cash and futures prices. The earliest studies do not explicitly attempt to forecast basis, but they do identify important determinants of basis. The first studies to forecast basis focus on intuitive technical models, primarily variants of a moving/rolling historical average. These studies emphasized the optimal number of years to include in the forecast for a certain commodity. More recent research examines more complicated time series models' ability to forecast basis. Most of the earlier studies focus on grains with a few exceptions examining livestock basis. In general, variants of the simple moving averages, commonly called the naïve approach, has been found to be the best forecast of basis.

Martin, Groenewegen, and Pidgeon (1980) model nearby corn basis in Ontario and find that local supply and demand conditions explain a substantial amount of basis that is not explained by spatial costs such as loading, tariffs, and rail charges. Garcia and Good (1983) examine the combined impact of spatial and supply and demand factors on corn basis. They find the effect these fundamental factors are significant in explaining basis variation and that their impact on basis varies across seasons.

Hauser, Garcia, and Tumblin (1990) forecast nearby soybean basis in Illinois using several different approaches. They consider various moving averages and a no change naïve approach where expected basis is the last observed basis at the time of forecast. They further consider a model which incorporates current market information measured as an implied basis based on price spreads between distant futures contracts. Finally, they examine a regression model based on the fundamental factors. In their study, the simple no change and three year historical average approaches provide the best forecasts before and during harvest while the futures spread approach provides the best forecast after harvest—the implication being that simpler forecasting methods provide better results than more complicated regressions on fundamental variables.

Jiang and Hayenga (1997) examine corn and soybean basis at several locations and find that a three year average performs well compared to other models. They also introduce fundamental variables into simple moving averages. They find the fundamental model with a three year historical average plus additional information on supply and demand factors performs better than a naïve average. They also find that an ARIMA outperforms the simple three year average for short horizon forecasts, but the improvement on the simpler models is not large. They also consider a state space model and artificial neural networks; however these perform no better than their simple historical moving averages models. Sanders and Manfredo (2006) also estimate autoregressive models. They look at VAR and ARMA models, and find that while for short forecast horizons these models show promise, the improvement is not substantial.

Dhuyvetter and Kastens (1998) compare the forecast performance of the no change model and historical averages of up to seven years for milo, corn, and soybeans. They also incorporate current market information in the form of the futures price spread as well as a new method which is a three year average including an adjustment for how much current basis deviates from its historic average. They find the forecast performance differs by crop and varies within a year with forecasts generally being the worst during planting periods. They determine a four year average to be the best forecast for wheat basis. On the other hand, longer averages perform better for corn and milo. The forecasts which incorporate current market information outperform the simpler averages by a small amount, but only for short forecast horizons of less than 12 weeks.

Tonsor, Dhuyvetter, and Mintert (2004) forecast livestock basis. They study one to five year historical averages for feeder cattle between 1979 and 2002 and for live cattle between 1982 and 2002. For feeder cattle, they find a three year average performs best over the sample period, but that a four year average performs best for the most recent five years of their sample. Contrarily, for live cattle, they find a four year average for the entire sample period and a two year average for the most recent five years of the sample period as the best forecasts. They find a benefit to including current market information in the form of an adjustment for how much current basis deviates from its historic average for forecasts of less than 16 weeks. They conclude that for shorter forecasts, naïve basis forecasts should be adjusted for that current market information.

Taylor, Dhuyvetter, and Kastens (2006) build on previous work examining the inclusion of current market information. For wheat, soybeans, corn, and milo, they compare simple historical averages of varying lengths to the same length averages supplemented with current market information in the form of current basis’ deviation from its historical average. The authors find that shorter averages perform best. Specifically, for pre-harvest forecasting, they find the basis lagged by one year to be the best predictor for corn, milo, and soy and a four year average to be the best for wheat. For post-harvest forecasts, they find a one year lagged basis supplemented with current market information performs best. They also report higher prediction errors than previous studies suggesting basis might be becoming harder to predict. Irwin, et al. (2009) echo the sentiment that grain basis is becoming harder to predict.

Hatchett, Brorsen, and Anderson (2010) acknowledge differences in the length of historical moving average found to be optimal in the literature, and they examine hard and soft wheat, corn, and soybeans over a large timespan to determine the optimal length of a moving average. They find a shorter average to be optimal, particularly in the face of a structural change, and like the previous research, they find current market information on basis to improve post-harvest forecasts. They recommend using shorter averages in the presence of structural change and longer averages otherwise. They also conclude that the differences in the forecast accuracy of varying lengths of historical averages are typically quite small for all crops examined.

In this paper, we combine these various forecasting models in a Bayesian framework to investigate the potential improvement over existing models. We also examine the potential of a basis series consisting of a specific month's futures contract to improve forecasts for that month.

Models

Models which consist of simple moving averages of basis on a specific date and various regression models are considered. These models are then weighted and combined using BMA method which is described below. The models that are included the BMA process are variations of regression models previously examined in the literature.

Because the no change and moving average forecasts do not have parameters, these models are not incorporated into the BMA framework as it is currently formulated. However, the performances of these models are evaluated so that models which *are* looked at and the average model can be benchmarked against them.

Two naïve models are considered: The model in equation (1) is a no change model where predicted basis is the last observed basis at the time of forecast. The model in equation (2) is a historical moving average; where *i* refers to the number of years included in the average, ranging from between one and five years. In all models, basis is defined as the cash price minus the futures price; *k* refers to location, *j* refers to week, *t* refers to year, and *h* refers to the forecast horizon.

$$\widehat{Basis}_{k,j,t} = Basis_{k,j-h,t} \quad (1)$$

$$\widehat{Basis}_{k,j,t} = \frac{1}{i}\sum_{l=t-i}^{t-1} Basis_{k,j,l} \quad (2)$$

$$i = 1, \ldots, 5;\ h = 1,\ 4,\ 8$$

We consider several regression models which we use to form an average model. Models are calculated with and without fundamental variables added as regressors in order to see how the addition of these regressors affects forecast performance.

The first set of regression models considered are regressions on the historical moving averages from (2), so they will be referred to as historical moving average regressions, HMA(*i*) or the HMA(*i*)+, where the "+" indicates the addition of fundamental variables, *F*. The HMA(*i*) and the HMA(*i*)+ are shown in equations (3) and (4) respectively.

$$\widehat{Basis}_{k,j,t} = \alpha + \beta\left(\frac{1}{i}\sum_{l=t-i}^{t-1} Basis_{k,j,l}\right) \quad (3)$$

$$\widehat{Basis}_{k,j,t} = \alpha + \beta\left(\frac{1}{i}\sum_{l=t-i}^{t-1} Basis_{k,j,l}\right) + \gamma F_{j-h,t} \quad (4)$$

$$i = 1, \ldots, 5;\ h = 1,\ 4,\ 8$$

The models in equations (5) and (6) include yearly lagged values of basis as separate independent variables. Thus, in these models, instead of every lag having the same weight in the average, the estimation procedure determines the weights. These models will be referred to as the yearly autoregressive model, YAR(*i*), or the YAR(*i*)+ if it includes fundamental explanatory variables.

$$\widehat{Basis}_{k,j,t} = \alpha + \left(\sum_{l=t-i}^{t-1} \beta_i Basis_{k,j,l}\right) \tag{5}$$

$$\widehat{Basis}_{k,j,t} = \alpha + \left(\sum_{l=t-i}^{t-1} \beta_i Basis_{k,j,l}\right) + \gamma F_{j-h,t} \tag{6}$$

$$i = 1, \ldots, 5;\ h = 1,\ 4,\ 8$$

The models in (7) and (8) are typical autoregressive models. As the data are weekly, the lagged values now refer to basis in previous weeks. This model will be referred to as an AR(*i*) and an AR(*i*)+ when explanatory variables are included.

$$\widehat{Basis}_{k,j,t} = \alpha + \left(\sum_{l=j-i-(h-1)}^{j-h} \beta_i Basis_{k,l,t}\right) \tag{7}$$

$$\widehat{Basis}_{k,j,t} = \alpha + \left(\sum_{l=j-i-(h-1)}^{j-h} \beta_i Basis_{k,l,t}\right) + \gamma F_{j-h,t} \tag{8}$$

$$i = 1,2,3;\ h = 1,\ 4,\ 8$$

The models in (9) and (10) are a combination of historical moving average and current market information in the form of the difference between basis at the time of forecast and its historical average. This model is an attempt to put the current market information model of Taylor, Dhuyvetter, and Kastens (2006) in a regression format. Without fundamental variables model will be referred to as CMI(*i*) and with them it will be called CMI(*i*)+.

$$\widehat{Basis}_{k,j,t} = \alpha + \beta\left(\frac{1}{i}\sum_{l=t-i}^{t-1} Basis_{k,j,l}\right) + \delta\left(Basis_{k,j-h,t} - \frac{1}{i}\sum_{l=t-i}^{t-1} Basis_{k,j-h,l}\right) \tag{9}$$

$$\widehat{Basis}_{k,j,t} = \alpha + \beta\left(\frac{1}{i}\sum_{l=t-i}^{t-1} Basis_{k,j,l}\right) + \delta\left(Basis_{k,j-h,t} - \frac{1}{i}\sum_{l=t-i}^{t-1} Basis_{k,j-h,l}\right) + \gamma F_{j-h,t} \tag{10}$$

$$i = 1, \ldots, 5;\ h = 1,\ 4,\ 8$$

We use a Bayesian approach to estimate all the linear models presented in (3)-(10) and Bayesian model averaging to handle model uncertainty. First, each model is estimated using Bayesian estimation. Then the support of each model, in the form of posterior model probabilities, is computed, and the predictions of each model are weighted by these posterior model probabilities. For the estimation of each model, consider a general regression equation in matrix form for model *m*:

$$y = X_m\theta_m + \varepsilon_m, \qquad m = 1, \ldots, M \tag{11}$$

where y is the dependent variable which is the same for all models at a given forecast horizon, and X_m is the matrix of independent variables for each model, θ_m is the vector of regression parameters to be estimated, ε_m is the vector of random error terms, and *M* is the total number of models evaluated.

Prior distributions on the regression parameters are specified to be normally distributed while that of the error variance, σ_m^2, is specified to be an inverse gamma distribution:

$$p(\theta_m) \sim N(\theta_{0m}, \sigma_m^2 V_{0m}), \tag{12}$$

$$p(\sigma_m^{-2}) \sim G(s_{0m}^{-2}, d_{0m}), \tag{13}$$

where *N* and *G* denote the multivariate normal distribution and gamma distribution, respectively.

$$L_m(y|\theta_{0m}, \sigma_m^2, X_m) = (2\pi\sigma_m^2)^{-n/2} exp\{-0.5(y - X_m\theta_m)'\sigma_m^{-2}(y - X_m\theta_m)\}, m = 1, \dots, M. \tag{14}$$

With these priors and the likelihood function, the joint posterior distribution of the parameters and the error variance can be derived with Bayes Theorem, and can be shown to follow a normal gamma distribution as:

$$p(\theta_m, \sigma_m^2|y, X_m) \sim NG\big(\theta_{pm}, V_{pm}, s_{pm}^2, d_{pm}\big),\ m = 1, \dots, M, \tag{15}$$

where

$$V_{pm} = (V_{0m}^{-1} + X_m' X_m)^{-1}, \tag{16}$$

$$\theta_{pm} = V_{pm}\big(V_{0m}^{-1}\theta_{0m} + (X_m' X_m)\hat{\theta}_m\big), \tag{17}$$

$$d_{pm} = d_{0m} + n_m, \tag{18}$$

$$s_{pm}^2 = d_{pm}^{-1}\left[d_{0m}s_{0m}^2 + (n_m - k_m)s_m^2 + \big(\hat{\theta}_m - \theta_{0m}\big)'(V_{0m} + (X_m' X_m)^{-1})^{-1}\big(\hat{\theta}_m - \theta_{0m}\big)\right] \tag{19}$$

In the above equations, *NG* represents the normal gamma distribution, $\theta_{pm}$is the posterior mean of the coefficients, $s_{pm}^2 V_{pm}$ is the posterior mean of the variance, d_{pm} is the posterior degrees of freedom, $\hat{\theta}_m$ and s_m^2 are standard OLS quantities, n_m and k_m are the number of rows and columns of X_m, respectively. After all models are estimated, uncertainty about which model is best supported by the data is addressed by weighting models by their posterior model probabilities.

To address model uncertainty, we first define prior weights for each model to be uninformative so that each model is treated equally:

$$p(M_m) \equiv \mu_m = \frac{1}{M}, \textstyle\sum_{m=1}^{M} \mu_m = 1, \quad \forall\, m = 1, \dots, M \tag{20}$$

Using the above posterior distribution from equation (15), the marginal likelihood for each model can be derived by integrating out the parameter uncertainty resulting in:

$$p(y|M_m) = c_m\big[|V_{pm}|/|V_{0m}|\big]^{1/2}\big(d_{pm}s_{pm}^2\big)^{-d_{pm}/2}. \tag{21}$$

where

$$c_m = \frac{\Gamma(d_{pm}/2)(d_{0m}s_{0m}^2)^{d_{0m}/2}}{\Gamma(d_{0m}/2)\pi^{n/2}}. \qquad (22)$$

Combining (21) and (22) by the Bayes Theorem, the posterior probability of each model can be shown as:

$$p(M_m|y) \propto \mu_m \left[\frac{|V_{pm}|}{|V_{0m}|}\right]^{\frac{1}{2}} \left(d_{pm}s_{pm}^2\right)^{-\frac{d_{pm}}{2}} = \mu_m p(y|M_m),\ m = 1, \ldots, M \qquad (23)$$

Finally, the posterior probabilities are normalized to sum to one by dividing each value in equation (23) by the sum of the unnormalized posterior probabilities across all models. The normalized posterior model probabilities are given as:

$$\omega_m = \frac{\mu_m p(y|M_m)}{\sum_{m=1}^{M} \mu_m p(y|M_m)},\ m = 1, \ldots, M. \qquad (24)$$

These normalized posterior model probabilities indicate how well a model is supported by the data with a higher value indicating a better model fit.

After each model is estimated and its normalized posterior probability is computed for the in-sample period, we create out-of-sample forecasts using those estimated parameters and weight these forecasts by their model probabilities. We evaluate the models based on their out-of-sample mean absolute error (MAE) and root mean squared error (RMSE). For the one step ahead forecasts, a total of 34 models are considered. Because some of these models use yearly lags as their sole explanatory variables, the HMA and YAR models do not produce unique forecasts for multiple step ahead forecasts. Therefore, they are not included among the models for the four and eight step ahead forecasts, resulting in a total of 25 models for these forecasts horizons.

Data

Cash price data on feeder cattle auctions held weekly are obtained from the USDA Agricultural Marketing Service for the period of January 1, 2004 to December 31, 2013 for three high volume Georgia auctions: Ashford, Calhoun, and Carnesville. The data are for steers with grade and weight descriptions matching the futures contracts. Specifications of cattle in the series are a mix of medium and large, number one and two cattle weighing between 650 and 700 pounds. Auctions at the chosen locations take place on either Tuesdays or Thursdays.

Daily feeder cattle futures settlement price data are obtained from Livestock Marketing Information Service. Because auctions occur on Tuesdays or Thursdays, feeder futures on Wednesdays are chosen as the weekly futures price in computing basis. A nearby contract series is created with the nearby contract being that closest to expiration. The contract rolls over at the end of the month the current nearby contract expires. For example, the March contract would be nearby in February since there is no February futures contract for feeder cattle, and this contract would be nearby until March 31, at which point the nearby contract would rollover to the next contract, April. A series consisting only of September futures contracts is also created. This series rolls over to the next year's September contract at the end of September in the current year.

Daily December corn futures settlement prices are chosen as a proxy for winter feed cost and obtained from CRB Trader. Wednesday's closing price is chosen as the representative weekly price to match that of the cattle futures. Weekly U.S. No 2 diesel retail prices are chosen as a proxy for transportation cost and these data are obtained from the U.S. Energy Information Administration.

All series are tested for stationarity using a single mean augmented Dickey-Fuller test with two augmenting lags (see Table 1). Both December corn futures and diesel price series show evidence of non-stationarity, so first-differenced, log price series, which are stationary, are used.

Given the great deal of upheaval in financial markets during the sample period, the basis series are examined for structural breaks. First, to test the most recent recession caused a structural break, the second week of September, 2008, the week Lehman Brothers filed for bankruptcy, is chosen as a break point. No break is implied visually in the plot of either basis series and no statistical evidence is found with Chow tests conducted at this break point. Second, a point in the spring of 2012 is chosen based on a disturbance in the plot of nearby basis. For nearby basis, this point tested positive for a break using Chow tests at all three locations. We believe this is due to the severe drought around this time. For consistency and to avoid complications of potential structural breaks, both the nearby and September basis series are cut off at the end of 2011. The years 2004-2010 are used as in-sample period to create the yearly moving averages and estimate model parameters. The year 2011 is reserved for out-of-sample forecast evaluation, so both nearby and September basis forecasts errors are examined over the whole year of 2011.

Figure 1 shows a plot of the nearby futures contract and cash price series at all locations separately. Prices begin to rise around 2010 and appear to stabilize at a higher level towards the end of the sample period. Figure 2 shows the nearby basis series at all locations. This series looks much more stable than the cash series, but around the beginning of 2012, basis does appear to have a lower mean value than in preceding years. Figure 3 shows plots of the September basis series.

Results

Results for nearby basis forecasts are presented in Tables 2-4. Each table presents the five best performing regression models based on out-of-sample mean absolute error (MAE) and root mean squared error (RMSE) along with their estimated posterior model probabilities. In all locations the top five models are the same whether chosen based on the lowest MAE or RMSE, and the ranking of the remaining models generally does not change either. The tables also present the performance of the average model denoted as BMA, the best naïve model, and the model with the highest posterior model probability.

For one step ahead forecasts (Table 2), a weekly autoregressive model performs better than other regression models considered at all locations. The inclusion of fundamental variables improves these models' forecast performance inconsistently across locations, but where it does improve performance, the improvement is typically very small compared to a model without fundamental variables. The average model does not outperform the best regression model or the best naïve, but the MAE and RMSE are in the same ballpark as the regression models and the naïve models. At all locations the best naïve model is the no change model for this short horizon forecast, and at two of the three locations considered (Calhoun and Carnesville), the best regression model and the average model do better than the naïve. The models with the highest posterior probabilities are AR(2)+ or AR(3)+ models.

For the four week ahead forecasts (Table 3), the best performing regression models vary across locations. There are also many different types of models within a location; however, an AR(2) does appear in the top five at all locations. The average model still produces forecast errors in the ballpark of the top five regression models, but, as can be seen by looking at Table 3, at four steps ahead, a naïve model is now outperforming the best regression models by around roughly $1/ctw. The best naïve is still a no change model at two locations (Ashford and Calhoun) while a four year average performs best at the third location (Carnesville). Although not presented in the table, the difference between a no change and a three or four year average is quite small for this forecast horizon. It is interesting that at this forecast horizon, the model with the highest posterior model probability does not appear in the top five even though the average model performs comparably with these models. The model with the highest posterior model probability is different at each location, but they do all contain the additional fundamental variables.

For eight weeks ahead forecasts (Table 4), the best performing naïve model outperforms the best regression and the average model at all locations. At this forecast horizon, a three or four year moving average performs best of the naïve models considered while the no change model now performs dismally. The best performing regression models are mostly HMA(*i*)+ or YAR(*i*)+ models. Majority of the best regression models contain the fundamental explanatory variables at this horizon. Models with the highest posterior probability appear in the top five regression models once again.

The results of the one step ahead September basis forecasts are presented in Table 5. For this forecast horizon, the autoregressive models and the average model perform better than the naïve forecast in one location (Carnesville). Weekly autoregressive models perform in the top five at all locations at this forecast horizon; although the specific model differs by location.

For four step ahead September basis forecasts (Table 6), regression models and the average model, in fact, perform worse than the naïve forecast. The average model also performs noticeably worse than the top five regression models. This appears to be because the model with the highest model probability at this forecast horizon performs relatively poorly compared to the best regression models. At this forecast horizon, the same model uncertainty observed in the four step ahead nearby basis forecast appear as many different types of models appear in the top five for the three locations considered.

For the eight step ahead September basis forecasts (Table 7), the variation in the best regression models vary considerably as well. At this forecast horizon the best naïve forecast performs better than the best regression model and the average model. At this forecast horizon, the average model also performs slightly worse than the best regression models; though not as starkly as with the four step ahead forecasts.

Conclusions

This paper examines feeder cattle basis forecasts at three Georgia locations. Both nearby basis and basis forecasts for a specific contract are examined. This second type of forecast would be useful in making decisions on distant future marketing plans. Forecasts are created using a naïve moving average and many regression based models. The regression models are used to form a combined, average model using Bayesian model averaging. Models are evaluated based on their out-of-sample forecast performance.

For both the forecast made with the nearby basis series and the September basis series, the best performing regression models at a given forecast horizon vary across locations, and, as in previous studies, the best model differs by forecast horizon. Autoregressive models do well for the shortest horizon while models with yearly lags do better for longer forecasts. The addition of fundamental variables does not consistently improve the performance of one and four week forecasts. However, all the best regression models for the eight step forecast contain the fundamental variables. At all locations and all forecast horizons, the model with the highest posterior model probability contains the fundamental variables. This shows that these fundamental variables are important factors in forecasting basis.

For both the naïve and September basis forecasts, the average model and the regression models perform comparably to the best naïve model for the one step ahead forecast, but for longer horizon forecasts, the best naïve model produces a substantially lower forecast error, generally around $1/ctw. For the shortest forecast horizon, a no change naïve model performs best. For the longest forecast horizon a three or four year moving average performs best among the naïve models considered. For the four step ahead forecast, a no change model and a three or four year moving average perform quite similarly. The naïve models' performance generally fell in line with previous studies' findings with no change models doing well for the shortest forecasts and three or four year moving average doing best for longer forecasts.

The regression models and the average model as currently formulated either did not perform much better or performed worse than the benchmark naïve models at all forecast horizons considered. In no case did averaging regression models produce lower forecast error than the best individual regression model. However, the set of regression models and explanatory variables included in the averaging process can be extended for further investigation of potential improvements in basis forecasting.

The disparity in the best performing regression model across locations and over different forecast horizons highlights the presence of model uncertainty in using regression models to forecast basis. Though performance of regression models varied substantially, the average model always performed comparably regardless of location or forecast horizon. While it may not be possible to generalize without further analysis of more locations and commodities, these results suggest that basis forecasts which use regressions should account for model uncertainty. Bayesian model averaging as it is implemented in this paper does so at the locations and for the commodities considered and it produces forecast errors in line with the best regression models.

References

Dhuyvetter, K.C. and T.L. Kastens. 1998. "Forecasting Crop Basis: Practical Alternatives." Proceedings of the NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management. St. Louis, MO. [http://www.farmdoc.illinois.edu/nccc134].

Garcia, P. and D. Good. 1983. "An Analysis of the Factors Influencing the Illinois Corn Basis, 1971-1981." Proceedings of the NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management. St. Louis, MO. [http://www.farmdoc.illinois.edu/nccc134].

Hatchett, R.B., B.W. Brorsen, and K.B. Anderson. 2010. "Optimal Length of Moving Average to Forecast Futures Basis." *Journal of Agricultural and Resource Economics* 35(1):18-33.

Hauser, R.J., P. Garcia, and A.D. Tumblin. 1990. "Basis Expectations and Soybean Hedging Effectiveness." North Central Journal of Agricultural Economics 12(1):125-136.

Hoeting, J.A., D. Madigan, A.E. Raftery, and C.T. Volinsky. 1999. "Bayesian Model Averaging: A Tutorial." *Statistical Science* 14(4):382-417.

Irwin, S. H., P. Garcia, D. L. Good, and E. L. Kunda. "Poor Convergence Performance of CBT Corn, Soybean, and Wheat Futures Contracts: Causes and Solutions." Marketing and Outlook Res. Rep. No. 2009-02, Dept. of Agr. and Consumer Econ., University of Illinois at Urbana Champaign, March 2009. Online. Available at http://www.farmdoc.uiuc.edu/marketing/reports.

Jiang, B. and M. Hayenga. 1997. "Corn and Soybean Basis Behavior and Forecasting: Fundamental and Alternative Approaches." Proceedings of the NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management. St. Louis, MO. [http://www.farmdoc.illinois.edu/nccc134].

Koop, G. Bayesian Econometrics. Chichester, UK: Wiley, 2003.

Martin, L., J.L. Groenewegen, and E. Pidgeon. 1980. "Factors Affecting Corn Basis in Southwestern Ontario." *American Journal of Agricultural Economics* 62(1):107-112.

Sanders, D.R. and M.R. Manfredo. 2006. "Forecasting Basis Levels in the Soybean Complex: A Comparison of Time Series Models." *Journal of Agricultural and Applied Economics* 38(3):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. [http://www.farmdoc.illinois.edu/nccc134].

Taylor, M.R., K.C. Dhuyvetter, and T.L. Kastens. 2006. "Forecasting Crop Basis Using Historical Averages Supplemented with Current Market Information." *Journal of Agricultural and Resource Economics* 31(3):549-567.

Tomek, W.G. 1997. "Commodity Futures Prices as Forecasts." *Review of Agricultural Economics* 19(1):23-44.

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.

Working, H. 1949. "The Theory of Price and Storage." *The American Economic Review* 39(6): 1254-1262.

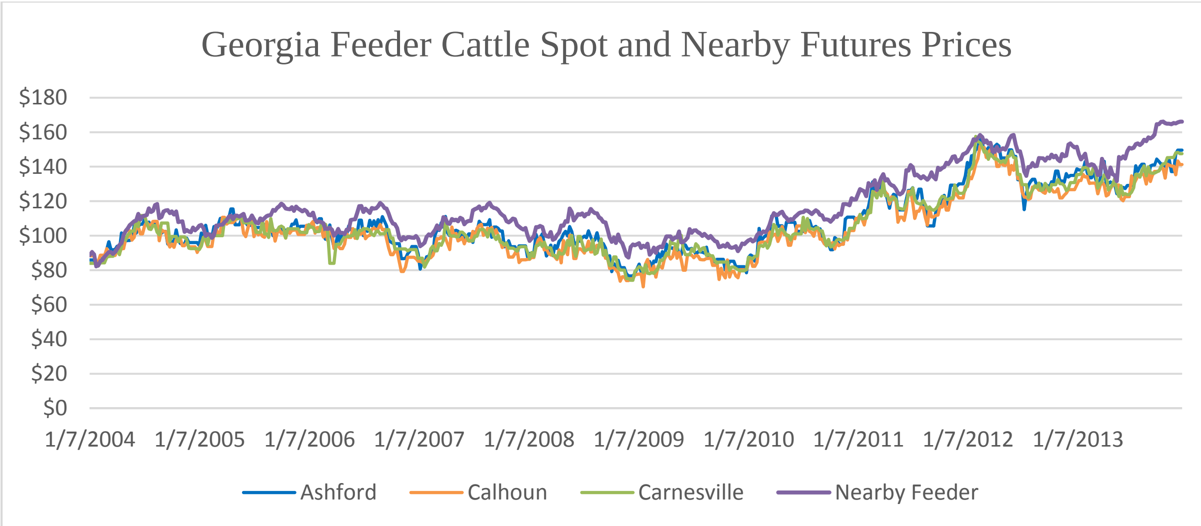

Figure 1. Spot and nearby futures prices at three Georgia locations from 2004 to 2012

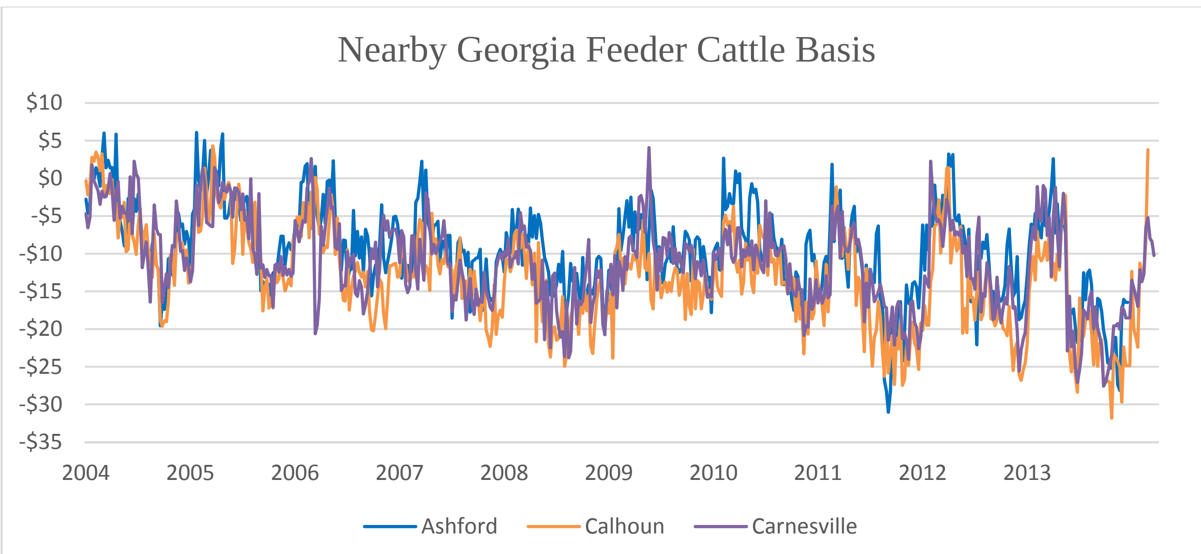

Figure 2. Nearby feeder cattle basis at three Georgia locations from 2004 to 2012

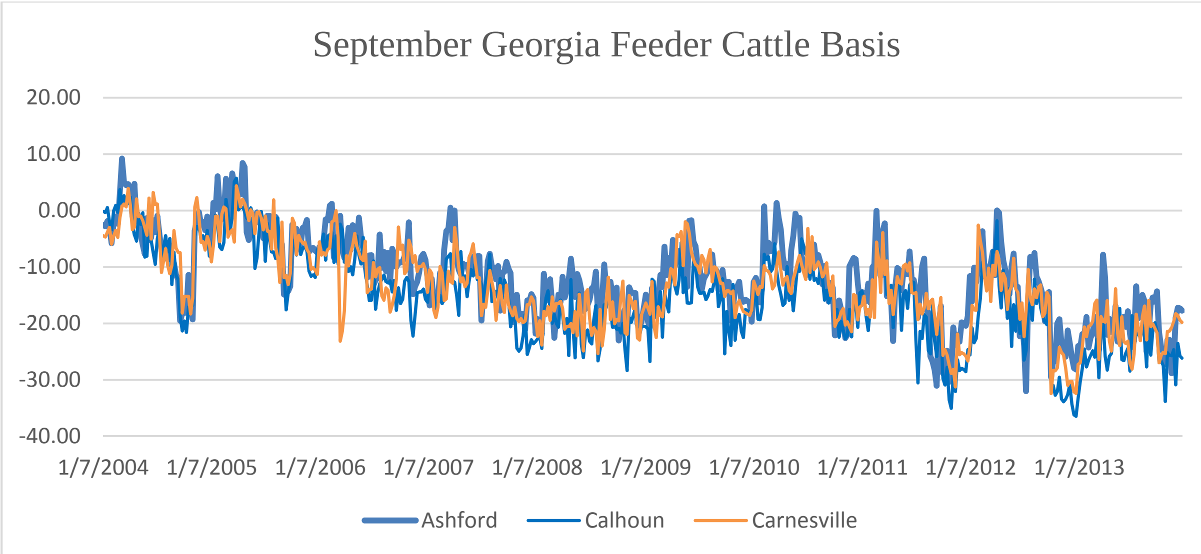

Figure 3. September feeder cattle basis at three Georgia locations from 2004 to 2012

Table 1. Augmented Dickey-Fuller Test Results

Series	Test statistic	P-value
Ashford Nearby Basis	-5.25	0.00
Calhoun Nearby Basis	-4.62	0.00
Carnesville Nearby Basis	-5.35	0.00
December Corn Futures Price	-1.87	0.34
Diesel Price	-2.01	0.28
Differenced, Log December Corn Futures Price	-11.39	0.00
Differenced, Log Diesel Price	-8.71	0.00

Table 2. One Step Ahead Nearby Basis Forecast Results

Ashford				Calhoun				Carnesville			
Model	Post Prob.	MAE	RMSE	Model	Post Prob.	MAE	RMSE	Model	Post Prob.	MAE	RMSE
AR(2)	0.77%	3.85	4.78	AR(3)+	55.32%	3.74	4.62	AR(3)	1.72%	2.65	3.41
AR(3)	0.67%	3.86	4.78	AR(3)	3.86%	3.79	4.65	AR(2)	1.62%	2.65	3.41
AR(3)+	36.23%	3.87	4.79	AR(2)+	27.03%	3.84	4.75	AR(3)+	50.03%	2.68	3.4
AR(2)+	41.80%	3.87	4.79	AR(2)	2.03%	3.87	4.77	AR(2)+	46.20%	2.69	4.42
AR(1)	0.24%	3.97	4.93	AR(1)+	0.09%	4.47	5.33	AR(1)	0.01%	3.04	3.76
BMA		3.89	4.82	BMA		3.87	4.87	BMA		2.68	3.4
Best Naïve: Last Observed		3.72	4.71	Best Naïve: Last Observed		4.08	4.79	Best Naïve: Last Observed		2.68	3.53
Highest Prob: AR(2)+	41.80%	3.87	4.78	Highest Prob: AR(3)+	55.32%	3.74	4.62	Highest Prob: AR(3)+	50.03%	2.68	3.41
M=34											

Table 3. Four Step Ahead Nearby Basis Forecast Results

Ashford				Calhoun				Carnesville			
Model	Post Prob.	MAE	RMSE	Model	Post Prob.	MAE	RMSE	Model	Post Prob.	MAE	RMSE
HMA(4)+	0.12%	5.19	6.89	AR(2)	0.12%	5.42	6.43	AR(2)	0.04%	3.64	4.61
HMA(5)+	0.02%	5.23	6.9	HMA(4)+	0.02%	5.42	6.43	AR(3)	0.04%	3.66	463
AR(2)	0.00%	5.42	7.22	HMA(2)+	0.00%	5.66	6.73	AR(2)+	37.94%	3.86	4.99
AR(3)	0.00%	5.46	7.24	CMI(2)	0.00%	5.66	6.50	AR(3)+	44.61%	3.88	5.01
AR(1)	0.00%	5.54	7.24	YAR(1)+	0.00%	5.69	6.77	AR(1)	0.00%	4.09	4.99
BMA		5.97	7.67	BMA		5.66	6.51	BMA		4.01	5.05
Best Naïve: Last Observed		4.76	5.68	Best Naïve: Last Observed		4.71	5.93	Best Naïve: 4 Year Avg		3.72	4.68
Highest Prob: YAR(4)+	96.34%	5.97	7.67	Highest Prob: CMI(2)+	91.00%	5.66	6.50	Highest Prob: AR(3)+	44.61%	3.88	5.01
M=25											

Table 4. Eight Step Ahead Nearby Basis Forecast Results

Ashford				Calhoun				Carnesville			
Model	Post Prob.	MAE	RMSE	Model	Post Prob.	MAE	RMSE	Model	Post Prob.	MAE	RMSE
HMA(5)+	0.51%	5.25	6.71	YAR(4)+	89.68%	5.72	6.55	HMA(5)+	0.02%	4.97	5.85
HMA(4)+	0.06%	5.26	6.90	YAR(3)+	8.50%	5.80	6.66	HMA(4)+	99.95%	5.00	6.11
HMA(3)+	0.03%	5.75	7.42	HMA(5)+	0.76%	5.87	6.75	HMA(3)+	0.03%	5.07	5.92
YAR(4)+	96.70%	5.79	7.46	YAR(2)+	0.80%	5.93	6.90	AR(3)	0.00%	5.16	5.93
YAR(5)+	2.25%	5.79	7.52	HMA(4)+	0.17%	6.04	6.99	AR(2)	0.00%	5.17	5.93
BMA		5.78	7.67	BMA		5.73	6.65	BMA		5.00	6.11
Best Naïve:				Best Naïve:				Best Naïve:			
3 Year Avg		4.76	6.52	4 Year Avg		4.71	6.03	4 Year Avg		3.72	4.68
Highest Prob:				Highest Prob:				Highest Prob:			
YAR(4)+	96.70%	5.79	7.46	YAR(4)+	89.68%	5.72	5.72	HMA(4)+	99.95%	5.00	6.11

M=25

Table 5. One Step Ahead September Basis Forecast Results

Ashford				Calhoun				Carnesville			
Model	Post Prob.	MAE	RMSE	Model	Post Prob.	MAE	RMSE	Model	Post Prob.	MAE	RMSE
AR(3)+	42.47%	3.96	4.98	AR(2)	2.30%	3.98	5.02	AR(2)	0.98%	2.91	3.79
AR(1)+	25.00%	4.02	5.08	AR(2)+	51.44%	3.99	5.03	AR(3)	2.32%	2.93	3.83
AR(3)	0.25%	4.03	4.98	AR(3)+	43.41%	3.99	5.03	AR(2)+	28.27%	3.03	3.93
AR(2)+	31.89%	4.04	5.02	AR(3)	1.91%	3.99	5.03	AR(3)+	68.00%	3.05	3.97
AR(1)	0.17%	4.05	5.07	AR(1)	0.04%	4.28	5.36	AR(1)	0.01%	3.26	4.23
BMA		3.99	5.00	BMA		3.99	5.03	BMA		3.02	3.95
Best Naïve:				Best Naïve:				Best Naïve:			
Last Observed		3.84	5.06	Last Observed		3.78	4.84	Last Observed		3.14	4.32
Highest Prob:				Highest Prob:				Highest Prob:			
AR(3)+	42.47%	3.96	4.98	AR(2)+	51.44%	3.99	5.03	AR(3)+	68.00%	3.05	3.97

M=34

Table 6. Four Step Ahead September Basis Forecast Results

Ashford				Calhoun				Carnesville			
Model	Post Prob.	MAE	RMSE	Model	Post Prob.	MAE	RMSE	Model	Post Prob.	MAE	RMSE
HMA(5)+	0.02%	5.35	7.06	AR(2)	0.05%	5.60	7.09	AR(2)+	33.94%	4.16	5.59
HMA(4)+	0.00%	5.74	7.36	AR(3)	0.04%	5.66	7.16	AR(2)	0.02%	4.18	5.28
AR(3)	0.00%	5.74	7.40	AR(1)	0.01%	5.77	7.38	AR(3)+	42.23%	4.19	5.59
AR(2)	0.00%	5.88	7.53	AR(2)+	1.54%	5.78	7.28	AR(1)+	23.77%	4.21	5.69
AR(3)+	0.00%	5.92	7.45	AR(3)+	1.39%	5.81	7.32	AR(3)	0.03%	4.21	5.28
BMA		6.86	8.34	BMA		7.17	9.12	BMA		4.17	5.61
Best Naïve:				Best Naïve:				Best Naïve:			
4 Year Avg		5.52	7.12	Last Observed		5.07	6.45	4 Year Avg		4.32	5.42
Highest Prob:				Highest Prob:				Highest Prob:			
YAR(5)+	96.40%	6.86	8.33	YAR(5)+	95.16%	7.23	9.19	AR(3)+	42.23%	4.19	5.59
M=25											

Table 7. Eight Step Ahead September Basis Forecast Results

Ashford				**Calhoun**				**Carnesville**			
Model	Post Prob.	MAE	RMSE	Model	Post Prob.	MAE	RMSE	Model	Post Prob.	MAE	RMSE
HMA(5)+	0.14%	5.31	7.07	HMA(5)+	0.06%	6.42	8.10	HMA(5)+	1.34%	5.00	6.36
HMA(4)+	0.01%	5.66	7.37	AR(2)	0.00%	6.46	8.27	HMA(4)+	0.48%	5.12	6.42
CMI(1)+	0.00%	6.46	7.81	AR(1)	0.00%	6.49	8.31	CMI(1)	0.00%	5.73	7.08
YAR(5)+	97.21%	6.47	7.97	AR(3)	0.00%	6.52	8.34	AR(1)	0.00%	5.75	7.14
CMI(1)	0.00%	6.53	7.93	AR(2)+	0.00%	6.58	8.43	AR(2)	0.00%	5.80	7.20
BMA		6.47	7.98	BMA		7.00	8.91	BMA		5.89	7.13
Best Naïve:				Best Naïve:				Best Naïve:			
4 Year Avg		5.52	7.12	Last Observed		5.85	7.33	4 Year Avg		4.32	5.42
Highest Prob:				Highest Prob:				Highest Prob:			
YAR(5)+	97.21%	6.47	7.97	YAR(5)+	98.58%	7.00	8.91	YAR(5)+	86.58%	5.88	7.11

M=25