Journal of Agricultural and Applied Economics, 38,3(December 2006):513–523

Forecasting Basis Levels in the Soybean Complex: A Comparison of Time Series Methods

Dwight R. Sanders and Mark R. Manfredo

A battery of time series methods are compared for forecasting basis levels in the soybean futures complex: soybeans, soybean meal, and soybean oil. Specifically, nearby basis forecasts are generated with exponential smoothing techniques, autoregression moving average (ARMA), and vector autoregression (VAR) models. The forecasts are compared to those of the 5-year average, year ago, and no change methods. Using the 5-year average as the benchmark method, the forecast evaluation results suggest that alternative naïve techniques may produce better forecasts, and the improvement gained by time series modeling is relatively small. In this sample, there is little evidence that the basis has become systematically more difficult to forecast in recent years.

Key Words: basis forecasts, time series models, soybean complex

JEL Classifications: C53, Q13

Accurate cash-futures basis forecasts are important for the successful marketing and procurement of agricultural commodities. In particular, an accurate understanding of basis and basis predictability is critical for the execution of successful hedging strategies. Indeed, researchers have recognized that understanding and forecasting basis relationships is an important component in agricultural price risk management (Tomek and Peterson). Researchers have primarily focused on explaining basis behavior in livestock (Leuthold and Peterson; Naik and Leuthold; Garcia, Leuthold, and Sarhan) and grain markets (Garcia and Good). Several studies have also focused specifically on forecasting the basis in these markets (Liu et al.; Garcia and Sanders; Hauser, Garcia, and Tumblin; Jiang and Hayenga; Dhuyvetter and Kastens). At one end of the spectrum, researchers have attempted to find the "best" sample length at which to calculate simple historic averages for basis forecasts (Dhuyvetter and Kastens). At the other end, researchers have compared econometric models with more complex time series procedures such as neural networks and state-space models (Jiang and Hayenga). In general, these basis forecasting studies have often found that time series models work as well as structural econometric models, and they are less costly to develop and maintain.

While basis forecasts do indeed play a crucial role in agricultural price risk management, a second important reason for improved basis forecasting has emerged in recent years. In an effort to increase the relevance of research related to price forecasting and commodity marketing, Brorsen and Irwin suggest that economists should move away from predicting prices—currently a major function of many extension programs. Kastens, Jones, and Schroeder echo this advice and recommend that price forecasts be formulated using pre-

Dwight R. Sanders is assistant professor, Department of Agribusiness Economics, Southern Illinois University, Carbondale, Illinois. Mark R. Manfredo is associate professor, Morrison School of Agribusiness and Resource Management, Arizona State University, Mesa, AZ.

The authors wish to thank two anonymous reviewers of the *Journal* for their useful comments and insights.

vailing futures prices and the expected basis. The clear implication of this futures-based price forecasting methodology is a renewed focus on forming basis expectations. In developing this idea, Kastens, Jones, and Schroeder examine different ways to define the relationship between the cash and futures price. They recommend using the futures price plus the simple average of the prior five years basis as the futures-based price forecast. Hence, the 5-year average is serving as the basis forecast. Given the futures price, any improvement in this average basis forecast will result in a more accurate price forecast. Therefore, it is important to understand if better basis forecasting procedures are available.

This research expands the basis forecasting literature on three fronts. First, it examines the relative performance of a number of time series methods that are easy to implement and maintain. Second, this research examines the basis relationships for a set of closely related markets—the soybean complex—and explores the benefits of modeling these related markets as a system. For the soybean complex, this entails the examination of a seasonally produced storable market (soybeans), a continual production storable product (soybean oil), and a continual production semi-storable product (soybean meal). It is plausible that a method which performs well for seasonally produced products such as soybeans may not perform as well for a continually produced product such as soybean oil. Furthermore, basis forecasts for the soybean complex are of interest to a diverse group of participants in the agribusiness sector, including soybean producers, crushers, livestock feeders, and food manufacturers. Finally, this research starts to address Tomek's (1993) concern about a lack of academic attention to basis forecasting in general and the changing predictability of the basis in particular. Collectively, a number of time series techniques are compared in an attempt to identify one that may assist risk managers in lifting or placing hedges, and assist extension economists and agribusinesses in making futures-based forecasts.

Data and Models

The soybean complex provides a unique opportunity to examine three closely related products (soybeans, soybean meal, and soybean oil) that have decidedly individual characteristics and to note any potential disparity in forecasting performance across products. Forecasting performance focuses on month-end (last business day of each month) nearby or spot basis levels. The Illinois Agricultural Statistics Service provides Central Illinois price quotes for soybeans, soybean meal, and soybean oil. Futures prices reflect the nearest-to-expiration futures contract for which the delivery month has not been entered. For instance, on April 30, the nearby soybean basis is the cash price quote for Central Illinois minus the May futures price. Similar prices are used to calculate the nearby soybean meal and soybean oil basis levels. Data are collected from January of 1975 through April of 2004. Models are estimated from 1975 through 1998 (288 monthly observations), and data from January of 2000 through April of 2004 (52 observations) are used to evaluate out-of-sample forecasting performance.[1]

The basis levels are forecasted using six alternative time series methods.[2] The methods are chosen based on their varying levels of complexity; yet each method has relatively low maintenance requirements in terms of model specification, programming, and database management. Therefore, if one particular model outperforms the others, the marginal cost of implementing that model is relatively modest. In this vein, the first three models represent naïve approaches. The first model,

[1] The data for 1999 are also used for making out-of-sample forecasts. However, because a common sample is desired for both 1-and 12-month-ahead forecasts, the forecasts for 1999 are removed from the out-of-sample tests.

[2] The basis levels for all three markets, soybeans, soybean meal, and soybean oil, were tested for stationarity over the estimation sample from 1975 through 1999 (300 observations). The augmented Dickey-Fuller test and Phillips-Perron procedures reject the null of a unit root at the 1% level for all series. Therefore, the bases series are stationary and can be modeled in levels.

which serves as the benchmark model, is the historical 5-year average basis for the month being forecast (5YR). This approach is commonly used in the literature (e.g., Dhuyvetter and Kastens) and is consistent with the recommendation by Kastens, Jones, and Schroeder for use by agribusinesses. The second model is a seasonal "no change" forecast, where the forecast equals the basis that existed at the same time one year ago (YAG). The third model is a "no change" assumption for which the forecasted basis equals the most recently observed basis (NC).

In addition to these naïve forecasting methods, three progressively more complex modeling procedures are used. First, the basis levels are forecast using the smoothing technique of Holt. Holt's adaptive k-step-ahead forecasting procedure assumes a smoothed series:

$$(1) \qquad y^s_{t+k} = a + bk + c_{t+k},$$

where y is the series being smoothed, a is the intercept or permanent component, b is the trend, and c_t is the additive seasonal factor. The parameters a, b, and c are calculated over the previous s observations using recursion. Consistent with the adaptive nature of the Holt's smoothing method, the parameters are calculated over the most recent 24 observations ($s = 24$).[3]

The second forecasting technique is to model each basis series individually as an autoregression moving average (ARMA) process. Over the in-sample period, 1975 through 1999, model specification is determined using the search procedure advocated by Beveridge and Oickle. That is, the basis levels are modeled using monthly dummy variables and all possible combinations of autoregressive and moving average terms up to twelve lags. Then, the model that minimizes Akaike's information criteria (AIC) is selected. For all three markets, the selected model was a relatively simple ARMA(1,1) specification. In out-of-sample forecasting, the models are continually reestimated using data up to the forecast date. However, the models are not respecified.

The third statistical procedure models the three series as a vector autoregression (VAR) process, which is a natural way to model a system of closely related markets. Again, lag selection is determined over the estimation sample by performing a search over all possible models, with each basis series allowed up to 12 lags. The model that minimizes AIC is picked to generate the out-of-sample forecasts. Exogenous variables in the final specification include the set of monthly binary variables and three lags of each endogenous basis level. The models are not respecified, but all available data at each forecasting date are used in estimating the models.

The out-of-sample forecasting period is chosen such that it reflects relatively current performance. This is especially important given the steady rise in South American soybean production in recent years. Each of the six forecasting procedures defined are used to generate forecasts from January of 2000 through April of 2004, resulting in 52 out-of-sample forecasts for each model. The number of out-of-sample forecasts is sufficiently large to examine how forecasting accuracy has changed, if at all, over this period.

Forecasts are made over horizons ranging from 1 to 12 months, which allows for a meaningful examination of how performance changes with the length of the forecast. This simulation produces out-of-sample forecasts for 12 horizons over a 52-month interval generated by six models, for a total of 3,744 individual forecasts for each market. In the following section, we examine these forecasts in detail to discern performance characteristics through time, across horizons, between markets, and among the models.

Forecast Accuracy

Mean absolute forecast errors (MAE) are calculated for each market, forecast method, and horizon. The results are presented in Table 1. Some similar results can be seen across markets. For soybeans, soybean meal, and soybean oil, the 5-year average (5YR) and year

[3] For a complete explanation of the Holt-Winters method and exponential smoothing, see Granger and Newbold (pp.165–78).

Table 1. Mean Absolute Forecast Errors, January 2000–April 2004

	Forecast Horizon (Months)												
Method	1	2	3	4	5	6	7	8	9	10	11	12	Average
Soybean basis, cents per bushel													
5YR	7.29	7.29	7.29	7.29	7.29	7.29	7.29	7.29	7.29	7.29	7.29	7.29	7.29
YAG	6.18	6.18	6.18	6.18	6.18	6.18	6.18	6.18	6.18	6.18	6.18	6.18	6.18
NC	7.15	8.67	8.42	8.77	9.46	7.91	8.34	8.22	8.54	7.62	6.97	6.18	8.02
Holt	6.00	6.18	5.99	6.59	7.01	7.11	7.46	7.92	8.23	8.55	9.08	9.46	7.47
ARMA	4.99*	5.15*	5.12*	5.18*	5.25*	5.12*	5.11*	5.22*	5.23*	5.25*	5.27*	5.28*	5.18
VAR	6.03*	6.45	6.42	6.82	7.15	7.25	6.98	7.14	6.97	6.66	6.77	6.62*	6.77
Soybean meal basis, dollars per ton													
5YR	3.79	3.79	3.79	3.79	3.79	3.79	3.79	3.79	3.79	3.79	3.79	3.79	3.79
YAG	2.78*	2.78*	2.78*	2.78*	2.78*	2.78*	2.78*	2.78*	2.78*	2.78*	2.78*	2.78*	2.78
NC	2.06*	2.68*	3.38	3.40	3.53	3.53	3.76	3.66	3.54	3.14	2.99*	2.78*	3.20
Holt	3.24	3.52	3.92	4.15	4.08	4.24	4.34	4.48	4.19	4.07	4.21	4.35	4.06
ARMA	2.27*	2.66*	3.06	3.10	3.14	3.04	3.23	3.20	3.25	3.23	3.32	3.38	3.07
VAR	2.34*	2.79*	3.21	3.29	3.27	3.34	3.31	3.33	3.29	3.14*	3.07*	3.07*	3.12
Soybean oil basis, dollars per hundredweight													
5YR	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98
YAG	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88
NC	0.24*	0.34*	0.47*	0.53	0.61	0.67	0.73	0.78	0.81	0.85	0.86	0.88	0.65
Holt	0.55*	0.66*	0.74	0.78	0.80	0.80	0.82	0.87	0.89	0.93	0.95	0.99	0.81
ARMA	0.27*	0.38*	0.49*	0.58*	0.65	0.70	0.75	0.79	0.82	0.85	0.87	0.88	0.67
VAR	0.28*	0.40*	0.48*	0.54*	0.60	0.65	0.69	0.74	0.78	0.81	0.84	0.85	0.64

Note: The average in the far right column is the simple average across all forecast horizons.
* An asterisk denotes that the mean absolute error is statistically different from that produced by the five-year average (5YR) at the 5% level.

ago (YAG) forecasts do not exhibit declining accuracy with the length of the forecast horizon. Because of the nature of the 5YR and YAG methods, where a particular month's forecast is the same at the 1-month horizon and 12-month horizon, the accuracy of these forecasting methods is identical across horizons. For soybeans the no change (NC) forecast declines in accuracy up to the 5-month horizon, and then improves. This pattern is best explained by seasonality in the soybean basis, where the 5-month-ahead forecast is counter-seasonal to the current month and the 12-month forecast is seasonally in sync with the current month. This pattern is also observed with the soybean meal basis forecasts, but forecast accuracy of the NC model declines until the 7-month horizon and then improves. This pattern is not seen, however, with the soybean oil basis where the NC forecast accuracy continues to decline as the forecast horizon increases.[4] The remaining forecasting methods, Holt, ARMA, and VAR, generally display some of the expected decline in accuracy beyond the 1-month horizon for each of the commodities.

The average MAE across all horizons for each of the commodities examined suggests that forecast performance is indeed commodity specific. For soybeans, the average MAE across all horizons suggests that the most accurate forecasting method is the ARMA model, with an average MAE of 5.18, followed by the YAG model with an average MAE of 6.18.

[4] The 12-month-ahead forecasts with the 5YR and YAG methods are immediately available following a given month, and the forecast does not change as the horizon shortens. Also, at the 12-month horizon, the NC forecast and the YAG forecast are the same.

The least accurate method for soybeans is the NC method with an MAE of 8.02. For soybean meal, the accuracy rankings differ. The most accurate method is the YAG with an average MAE of 2.78 followed by ARMA at 3.07, with the Holt method being the least accurate, having an average MAE of 4.06. For soybean oil, the VAR performs best based on average MAE (0.64), while the NC and ARMA methods follow closely behind (0.65 and 0.67 respectively). The 5YR method with an average MAE of 0.98 is the least accurate method for forecasting soybean oil basis.

While these rankings are useful in comparing average accuracy across forecast horizons, it is equally important to determine how the various forecasts perform relative to a simple benchmark forecast, and if the performance is statistically better (or worse) than the benchmark (Manfredo, Leuthold, and Irwin). Using the 5YR as the benchmark forecast (Kastens, Schroeder, and Plain), the statistical differences in MAEs of 5YR versus those produced by all the other methods across horizons is tested using the modified Diebold-Mariano (MDM) test proposed by Harvey, Leybourne, and Newbold. Specifically, the modified Diebold-Mariano test (MDM) considers two time series of h-step-ahead forecast errors (e_{1t}, e_{2t}), for $t = 1, \ldots, n$, and a specified loss function $g(e)$, with the null hypothesis of equal expected forecast performance being $E[g(e_{1t}) - g(e_{2t})] = 0$. For h-step-ahead forecasts, the MDM test is based on the sample mean ($\bar{d}$) of $d_t = g(e_{1t}) - g(e_{2t})$ with appropriate adjustments for $h - 1$ autocorrelation.

$$(2)\qquad \mathrm{MDM} = \left[\frac{n + 1 - 2h + n^{-1}h(h - 1)}{n}\right] \times \left[n^{-1}\left(\hat{\gamma}_0 + 2\sum_{k=1}^{h-1} \hat{\gamma}_k\right)\right]^{-1/2} \bar{d},$$

where $\hat{\gamma}_k = n^{-1} \sum_{t=k+1}^{n} (d_t - \bar{d})(d_{t-k} - \bar{d})$ is the estimated kth autocovariance of d_t, and $\bar{d}$ is the sample mean of d_t. In this application, the loss function, $g(e)$, is the absolute value function. The MDM statistic is compared with the critical values from a t-distribution with $n - 1$ degrees of freedom. These results are also presented in Table 1.

For soybeans (Table 1), only the ARMA model and the VAR model generate statistically smaller forecast errors than the 5YR. The ARMA model generates a statistically smaller MAE at all horizons, while the VAR only does so at the 1- and 12-month horizons. For soybeans, forecast accuracy can be improved over the 5YR by employing a time series technique. Moreover, the ARMA performs better than the 5YR at all horizons. Surprisingly, the VAR model, which incorporates information from the three related markets, does not provide better forecasts than the univariate ARMA model.

Turning to the soybean meal results in Table 1, the results are markedly different from those for the soybean basis. The YAG method, which is the most accurate model based on average MAE, also produces a statistically smaller MAE than 5YR at all horizons. The second most accurate model, ARMA, produces a statistically smaller MAE than the 5YR at only the 1- and 2-month forecast horizons. The NC and VAR models both produce statistically smaller MAE at short horizons (1 and 2 months) and long horizons (11 and 12 months). Unlike for soybeans, time series techniques offer only limited improvement in forecasting the soybean meal basis. Indeed, the most accurate forecast on average is provided by the simple YAG model.

For soybean oil (Panel C), the most accurate forecasts are produced by the VAR, NC, and ARMA methods, respectively. Notably, the benchmark 5YR is the least accurate forecasting method, especially at short horizons. All methods, except the YAG, produced a statistically smaller MAE at the 1- and 2-month horizons. However, no method statistically outperformed 5YR at a horizon greater than 4 months. As with soybean meal, the time series modeling techniques provide some marked improvement over the 5YR at short horizons. Again, however, there is a naïve method, the NC, that performs at a comparable level to the time series models. In particular, NC has the smallest MAE at the 1-, 2-, 3-, and 4-month horizons, and the MAE is statistically smaller than the base case 5YR at the 1-, 2-, and 3-month horizons.

While the MAE results suggest that forecast performance is commodity and horizon specific, a few observations can be made regarding the general performance of the alternative forecasting procedures across commodities. For instance, across the three markets, the ARMA model is marked by consistently good performance. The ARMA model produces statistically smaller MAEs than 5YR at all horizons for the soybean basis. The ARMA model also produces statistically smaller MAEs than 5YR at shorter horizons for soybean meal (1 and 2 months) and soybean oil (less than 5 months). However, for soybean meal and soybean oil, the NC model performs comparable to the ARMA model. At short horizons (1 and 2 months for soybean meal and 1, 2, and 3 months for soybean oil), NC displays smaller MAEs than ARMA, and the MAEs are statistically smaller than those produced by the base case 5YR model. Thus in these cases, a naïve method performs comparably to the more complex time series techniques. While simple methods may indeed produce accurate forecasts, the most suitable method can differ across markets.

To further highlight the relative performance across markets, for each forecast the percent reduction in forecast error versus the 5YR benchmark is calculated and presented in Table 2. In each market, the most accurate forecasting method at each horizon provides a considerable reduction in MAE. For instance, the ARMA model for the soybean basis (Table 1), produces a MAE that is approximately 30% smaller than that of the 5YR model across forecast horizons. In the soybean products, the short-horizon gains are quite large, with a 46% reduction in soybean meal MAE at the 1-month horizon, and 76% reduction in soybean oil MAE at the same forecast horizon. Indeed, the information in Table 2 highlights three important points relative to using a 5-year average basis forecast (5YR). First, the relative improvement using alternative forecasting methods is quite large, with a range of 13% to 76% decrease in MAE across commodities. Second, the largest improvements in performance are concentrated at shorter horizons. Finally, there is some evidence that the 5YR method is a relatively poorer forecast for the continually produced product markets than with the seasonally produced soybean market, in particular at shorter horizons.

Forecast Characteristics

While the results in Tables 1 and 2 hint at the performance of the forecasts across horizons and methods, a further understanding of the forecast characteristics is pursued using the regression-based methodology of Kastens, Schroeder, and Plain. Their methodology specifically allows for an investigation of forecast performance across methods, horizons, and time. The absolute forecast errors are modeled in the following OLS regression:

$$
\begin{aligned}
(3)\qquad \text{MAE}_t = {} & \alpha + \alpha_i\text{Method}_{i,t} + \beta\text{Horizon}_t \\
& + \beta_i(\text{Horizon}_t \times \text{Method}_{i,t}) \\
& + \lambda\text{Trend}_t + \lambda_i(\text{Trend}_t \times \text{Method}_{i,t}) \\
& + \varphi_j\text{Month}_j + \varepsilon_t
\end{aligned}
$$

where $\text{Method}_{i,t}$ is a set of dummy variables that equal zero for the 5YR method and equal one for each of the other models (YAG, NC, Holt, ARMA, and VAR). Horizon_t takes a value from 0 to 11 to capture the 12 forecast horizons. Trend_t is a linear time trend, with the first monthly observation (January 2000) set equal to zero. Month_j is a set of monthly dummy variables with a base of January. Simply, the regression explains the MAE across horizons, through time, and across months while allowing each forecast model to interact with the intercept, horizon, and trend. In this framework, the base case is the 5YR model at the 1-month horizon in January of 2000. The model is estimated with OLS across all 3,744 out-of-sample forecasts. Following Kastens, Schroeder, and Plain, the standard errors are corrected for heteroskedasticity and serial correlation using the heteroskedastic and serially correlated consistent covariance estimator of Newey and West. The estimation results for the soybean, soybean meal, and soybean oil basis are presented in Table 3.

Table 2. Percent Error Reduction versus the Five-Year Average Benchmark Forecast, January 2000–April 2004

	Forecast Horizon (Months)												
Method	1	2	3	4	5	6	7	8	9	10	11	12	Average
Soybean basis													
YAG	−0.15	−0.15	−0.15	−0.15	−0.15	−0.15	−0.15	−0.15	−0.15	−0.15	−0.15	−0.15	−0.15
NC	−0.02	0.19	0.16	0.20	0.30	0.09	0.14	0.13	0.17	0.05	−0.04	−0.15	0.10
Holt	−0.18	−0.15	−0.18	−0.10	−0.04	−0.02	0.02	0.09	0.13	0.17	0.25	0.30	0.02
ARMA	−0.31	−0.29	−0.30	−0.29	−0.28	−0.30	−0.30	−0.28	−0.28	−0.28	−0.28	−0.27	−0.29
VAR	−0.17	−0.12	−0.12	−0.06	−0.02	0.00	−0.04	−0.02	−0.04	−0.09	−0.07	−0.09	−0.07
Soybean meal basis													
YAG	−0.26	−0.26	−0.26	−0.26	−0.26	−0.26	−0.26	−0.26	−0.26	−0.26	−0.26	−0.26	−0.26
NC	−0.46	−0.29	−0.11	−0.10	−0.07	−0.07	−0.01	−0.03	−0.07	−0.17	−0.21	−0.26	−0.15
Holt	−0.14	−0.07	0.04	0.10	0.08	0.12	0.15	0.18	0.11	0.07	0.11	0.15	0.07
ARMA	−0.40	−0.30	−0.19	−0.18	−0.17	−0.20	−0.15	−0.15	−0.14	−0.15	−0.12	−0.11	−0.19
VAR	−0.38	−0.26	−0.15	−0.13	−0.14	−0.12	−0.13	−0.12	−0.13	−0.17	−0.19	−0.19	−0.18
Soybean oil basis													
YAG	−0.10	−0.10	−0.10	−0.10	−0.10	−0.10	−0.10	−0.10	−0.10	−0.10	−0.10	−0.10	−0.10
NC	−0.76	−0.65	−0.52	−0.46	−0.37	−0.32	−0.25	−0.20	−0.17	−0.13	−0.12	−0.10	−0.34
Holt	−0.44	−0.33	−0.24	−0.20	−0.18	−0.18	−0.16	−0.11	−0.09	−0.05	−0.02	0.01	−0.17
ARMA	−0.72	−0.61	−0.50	−0.41	−0.33	−0.28	−0.23	−0.19	−0.15	−0.13	−0.11	−0.09	−0.31
VAR	−0.71	−0.59	−0.51	−0.44	−0.38	−0.34	−0.29	−0.25	−0.20	−0.17	−0.14	−0.13	−0.35

Note: The average in the far right column is the simple average across all forecast horizons.

Table 3. Absolute Forecasting Error Models, Soybean Complex Basis, January 2000–April 2004

Variables	Parameter	Soybean	Soybean Meal	Soybean Oil
Intercept	α	6.916**	2.483**	0.535**
	α_{YAG}	−0.415	−0.427	0.252*
	α_{NC}	−0.331	−1.886**	−0.128
	α_{Holt}	1.150	1.466**	0.031
	α_{ARMA}	−3.189**	−0.579	−0.075
	α_{VAR}	−1.868**	−0.893**	−0.074
Horizon	β	−0.041	−0.020	0.004
	β_{YAG}	0.000	0.000	0.000
	β_{NC}	−0.085	0.056	0.054**
	β_{Holt}	0.327**	0.073	0.033**
	β_{ARMA}	0.018	0.072	0.053**
	β_{VAR}	0.041	0.035	0.050**
Trend	λ	−0.015	0.037**	0.020**
	λ_{YAG}	−0.027	−0.023**	−0.014**
	λ_{NC}	0.057**	0.037**	−0.019**
	λ_{Holt}	−0.109**	−0.062**	−0.015**
	λ_{ARMA}	0.039	−0.021	−0.021**
	λ_{VAR}	0.044*	0.001	−0.021**
Month	$\phi_{Feb.}$	−1.427**	0.189	0.098**
	$\phi_{Mar.}$	0.540*	0.416**	−0.103**
	$\phi_{Apr.}$	1.404**	−0.100	−0.085**
	ϕ_{May}	0.641*	0.298*	−0.343**
	$\phi_{Jun.}$	0.311	−0.854**	−0.360**
	$\phi_{Jly.}$	2.690**	0.810**	−0.317**
	$\phi_{Aug.}$	1.084**	2.613**	−0.245**
	$\phi_{Sep.}$	5.262**	1.181**	−0.022
	$\phi_{Oct.}$	0.559	1.298**	0.107**
	$\phi_{Nov.}$	4.592**	0.463**	−0.034
	$\phi_{Dec.}$	−1.944**	0.317**	0.125**
	$\bar{R}^2$	0.14	0.13	0.25

Notes: The parameter estimates are from the model, $MAE_t = \alpha + \alpha_i Method_{i,t} + \beta Horizon_t + \beta_i(Horizon_t{*}Method_{i,t}) + \lambda Trend_t + \lambda_i(Trend_t{*}Method_{i,t}) + \phi_j Month_j + \epsilon_t$, where the subscript i denotes forecast method i. The model is estimated over 3,744 observations in each market.

* Statistically different from zero at the 10% level using a t-test.

** Statistically different from zero at the 5% level using a t-test.

For soybeans, the baseline 5YR forecast results are characterized by an intercept of 6.916 cents per bushel. The MAE for the 5YR method does not display statistically significant trends across horizons or through time. Relative to this base case, the method intercept shifters suggest that both the ARMA and VAR models generate a MAE smaller than the base 5YR method (−3.189 and −1.868 respectively). These results are consistent with the MDM test results in Table 1. Across horizons, only the Holt method shows a relative decline in accuracy, with the MAE increasing 0.327 cents per bushel per month. Through time, the 5YR method's accuracy has not changed. However, the accuracy of the NC and VAR methods has gotten progressively worse (0.057 and 0.044 respectively). In contrast, the Holt method has produced smaller MAEs through time (−0.109). It is not clear what (if any) underlying characteristic may be driving this trend. Finally, across months, there is a fairly strong seasonality in the MAE for the soybean basis. In particular, the February and December errors are statistically smaller, while the April, August, September, and November are larger and statistically significant relative to the base January error. This finding is likely due to the known greater overall price and supply uncertainty during the growing season. Alternatively, the transition from old crop soybeans (August) to new crop (September) may generate larger forecast errors. However, the forecast errors for August and September are not materially larger than those for April or November.

The 5YR method for soybean meal has a conditional average MAE of 2.483 dollars per ton. Consistent with the results in Table 1, the 5YR method's accuracy does not decline across horizons, but it does decline across time (trend), with the MAE increasing an average of 0.037 per month. Relative to the 5YR method, the soybean meal basis forecasts produced by the VAR and NC models are statistically more accurate by 1.886 and 0.893, respectively, all else constant. The Holt model produces a MAE that is statistically larger by 1.466 relative to the 5YR. These results are consistent with the MDM test results in Table 1, which also showed the YAG and VAR models producing a smaller MAE at certain horizons. None of the soybean meal basis forecasts display statistically significant accuracy degradation at longer horizons. However, over the sample period, the 5YR forecast accuracy statistically declined (0.037). Importantly, the NC method's accuracy declined at nearly twice the rate, indicating that its relative outperformance is concentrated early in the sample. So, there is some evidence that the soybean meal basis has become more difficult to forecast with naïve methods. Yet, YAG and Holt accuracy declined statistically slower than the 5YR method, with their net declines equal to 0.014 (0.037 − 0.023) and −0.025 (0.0357 − 0.062), respectively. Seasonally, the soybean meal basis in March, July, August, September, October, November, and December is more difficult to forecast than in January, while June has a smaller MAE. Again, part of the increase in errors could stem from growing season uncertainty or from the shift to new crop pricing in late summer.

For soybean oil, the base 5YR soybean oil model has a conditional average MAE of 0.535 dollars per hundredweight. The 5YR model's MAE is stable across horizons, but the trend intercept estimate suggests that it statistically increases by 0.020 per month. Only the YAG method's conditional MAE is different from that of the 5YR. All else equal, the YAG method produces a MAE that is 0.252 larger than that of the 5YR. This seems at odds with the results in Table 1, but the YAG MAE decreases with time, which implies it performed much better late in the sample. Consistent with the results presented in Table 1, the NC, Holt, ARMA, and VAR forecasts display a strong tendency for accuracy to decline with the horizon, relative to the 5YR forecasts. Relative to the 5YR method's MAE, the other methods show a trend towards improved forecast performance over time; although the net trend is near zero. For instance, the ARMA model's net change through time is a negligible −0.001 (0.020 − 0.019). So, other than the 5YR method, the soybean oil forecast errors show no discernable trend through time. Finally, relative to January, the months of March, April, May, June, July, and August have smaller forecast errors, and February, October, and December have larger forecast errors. There is no clear reason for the pattern of seasonality in the forecastibility of the soybean oil basis.

Collectively the results in Table 3 suggest that relative to the 5YR average base, the other forecasting techniques provide no clear benefit in terms of accuracy improvement at longer horizons. Indeed, at longer horizons it is difficult to outperform the 5YR. However, in other aspects, forecast performance relative to the 5YR is highly dependent on the method and market. For instance, in this sample, the performance of the 5YR declined in soybean meal as it did for the VAR and NC methods, while the Holt forecasts actually improved. Across the markets, it appears that any consistent benefit from using more rigorous VAR and ARMA time series procedures are concentrated in soybeans and at short forecast horizons.

Summary and Conclusions

This research considers the performance of five alternative models for forecasting the cash-futures basis for soybeans, soybean meal, and soybean oil. The models vary in degree of complexity from the use of naïve models, such as a no-change model, to time series methods such as vector autoregression (VAR). Following the suggestions of Tomek (1993), this research also considers how the alternative forecasts perform across varying forecast horizons (1 to 12 months) and examines if the forecastibilty of the basis changed over the period. Thus the methods used here provide a unique perspective on forecasting the cash-futures basis in the soybean complex. Given the importance of basis forecasting for risk management purposes, and for developing futures-based price forecasts, the results from this research should provide both practitioners and academics alike with critical insights into both alternative procedures and the general forecastibiltiy of the basis in the soybean complex.

In general, the results suggest that recommending a 5-year average basis forecast, a method routinely used by agribusiness practitioners and recommended by extension economists, may not always be the best advice. For each basis examined in the soybean complex, at least two alternative methods produce a statistically smaller mean absolute error (MAE) than the 5-year average (5YR) forecasting method. For instance, in soybean meal the even-simpler year ago (YAG) and no change (NC) methods produce smaller forecast errors. Generally speaking, the 5YR method is not the most accurate forecasting method, even among the naïve alternatives (5YR, YAG, and NC). ARMA models generally produce smaller forecast errors at short horizons relative to 5YR. However, there seems to be little gained by more complicated time series techniques such as the VAR. Moreover, rarely did a modeling procedure produce dramatically better results than the best naïve method. For example, in soybean meal, the ARMA model produced a MAE of 2.27, which is statistically smaller than the 5YR, but the NC method has a MAE of 2.06 and is an easier method to implement. Collectively, the results suggest that modeling efforts may be worthwhile at short horizons, but the results must be carefully weighed against the performance of an array of naïve methods.

The regression analysis further suggests that modeling efforts relative to using the 5YR method are most rewarding at short horizons and appear to perform better later in the sample period. This is especially true for soybean oil, where the 5YR method's accuracy actually got worse through time relative to alternative methods. Tomek's concerns about basis forecastability declining are a complicated issue: it depends on the forecasting model used. In this specific sample, some models certainly show declining forecast accuracy, but generally this is not a systematic problem across all markets and methods.

Indeed, general statements about forecasting performance are difficult to make. The results do not suggest that a single method dominates across these markets, even though the commodities are closely related. Forecasting performance through time and across horizons depends on the forecasting method used and market examined. It is clear that there are methods, both simple and more complex, that statistically outperform the 5-year average (5YR). For soybeans, the ARMA model appears to be the best basis forecaster, with the smallest MAE and no degradation across time or horizons. In soybean meal, the YAG method is a likely choice for practitioners, as it produces the smallest average MAE and performance has improved in recent periods. The NC forecasts provide good performance for soybean oil basis with a smaller MAE than the 5YR and consistent performance through time.

These results do not preclude the use of the 5YR method by practitioners. In particular, as the forecast horizon lengthens, the performance gap between the 5YR and more complicated methods tends to decline. The 5YR method is likely to be robust over a number of different markets, allowing it to be easily implemented across a number of locations and commodities. Indeed, for producers—who may not have the time or skill set to implement more formal modeling—the 5YR may be a preferred approach. But for analysts and extensions economists, the presented results indicate that improvements may be obtained by implementing simple time series models or looking for a most accurate naïve method.

While forecasting performance clearly varies across methods and markets, simple modeling techniques (e.g., ARMA) are consistently good performers vis-à-vis the naïve 5-year average. However, practitioners should be cautious about applying a generic recipe for basis forecasting. Even within closely related markets, such as the soybean complex, there is no single method that can be uniformly applied to generate the best forecasts.

[Received March 2005; Accepted February 2006.]

References

Beveridge, S., and C. Oickle. "A Comparison of Box-Jenkins and Objective Methods for Determining the Order of a Nonseasonal ARMA Model." *Journal of Forecasting* 13(1994):419–34.

Brorsen, B.W., and S.H. Irwin. "Improving the Relevance of Research on Price Forecasting and Marketing Strategies." *Agriculture and Resource Economics Review* 25(1996):68–75.

Dhuyvetter, K.C., and T.L. Kastens. "Forecasting Crop Basis: Practical Alternatives." *Applied Commodity Price Analysis, Forecasting, and Market Risk Management,* Proceedings of the NCR-134 Conference, 1998, pp. 49–67.

Diebold, F.X., and R.S. Mariano. "Comparing Predictive Accuracy." *Journal of Business and Economic Statistics* 13(1995):253–63.

Garcia, P., and D.L. Good. "An Analysis of the Factors Influencing the Illinois Corn Basis, 1971–1981." *Applied Commodity Price Analysis, Forecasting, and Market Risk Management,* Proceedings of the NCR-134 Conference, 1983, pp. 306–26.

Garcia, P., R.M. Leuthold, and M.E. Sarhan. "Basis Risk: Measurement and Analysis of Basis Fluctuations for Selected Livestock Markets." *American Journal of Agricultural Economics* 66(1984):499–504.

Garcia, P., and D.R. Sanders. "Ex Ante Basis Risk in the Live Hog Futures Contract: Has Hedgers' Risk Increased?" *Journal of Futures Markets* 16(1996):421–40.

Granger, C.W.J., and P. Newbold. *Forecasting Economic Time Series,* 2nd ed. New York: Academic Press, 1986.

Harvey, D., S. Leybourne, and P. Newbold. "Testing the Equality of Prediction Mean Squared Errors." *International Journal of Forecasting* 13(1997): 281–91.

Hauser, R.J., P. Garcia, and A.D. Tumblin. "Basis Expectations and Soybean Hedging Effectiveness." *North Central Journal of Agricultural Economics* 12(1990):125–34.

Holt, C.C. "Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages." *International Journal of Forecasting* 20(2004):5–10.

Jiang, B., and M. Hayenga. "Corn and Soybean Basis Behavior and Forecasting: Fundamental and Alternative Approaches." *Applied Commodity Price Analysis, Forecasting, and Market Risk Management,* Proceedings of the NCR-134 Conference, 1997, pp. 125–40.

Kastens, T.L., R. Jones, and T.C. Schroeder. "Futures-Based Price Forecasts for Agricultural Producers and Businesses." *Journal of Agricultural and Resource Economics* 23(1998):294–307.

Kastens, T. L., T.C. Schroeder, and R. Plain. "Evaluation of Extension and USDA Price and Production Forecasts. *Journal of Agricultural and Resource Economics* 23(1998):244–61.

Leuthold, R.M., and P.E. Peterson. "The Cash-Futures Price Spread for Live Hogs." *North Central Journal of Agricultural Economics* 5(1983): 25–29.

Liu, S.M., B.W. Brorsen, C.M. Oellermann, and P.L. Farris. "Forecasting the Nearby Basis of Live Cattle." *Journal of Futures Markets* 14(1994):259–73.

Manfredo, M.R., R. M. Leuthold, and S.H. Irwin. "Forecasting Fed Cattle, Feeder Cattle, and Corn Cash Price Volatility: The Accuracy of Time Series, Implied Volatility, and Composite Approaches." *Journal of Agricultural and Applied Economics* 33(2001):523–38.

Naik, G., and R.M. Leuthold. "Cash and Futures Price Relationships for Nonstorable Commodities: An Empirical Analysis Using a General Theory." *Western Journal of Agricultural Economics* 13(1988):327–38.

Tomek, W.G. "Dynamics of Price Changes: Implications for Agricultural Futures Markets." *Research Frontiers in Futures Markets and Options: An Exchange of Ideas. A Symposium in Recognition of Thomas A. Hieronymus.* Urbana, IL: University of Illinois, 1993, pp. 45–55.

Tomek, W.G., and H.H. Peterson. "Risk Management in Agricultural Markets: A Review." *Journal of Futures Markets* 21(2001):953–85.