Forecasting the Basis for Corn in Western New York

Patricia D. Taylor and William G. Tomek

This study develops a simple model to forecast the basis for corn in a specific region. Improved forecasts can improve hedging decisions. Basis behavior, however, depends on explanatory variables that are themselves difficult to forecast with precision. This limits the usefulness of the basis model, but it does offer some benefit over naive forecasts.

The main objective of this article is to develop a model for forecasting the November basis behavior of corn in Batavia, New York. Although the empirical model is applicable only for Western New York corn growers, its specification may be relevant for similar corn markets. In addition, the discussion of the problems of using the model has general applicability.

The success or lack of success of hedges depends on the behavior of the basis faced by the hedger. If, for example, the local harvesttime basis is known to be $0.25 per bushel, then a corn grower selling December futures at $3.00 at planting time could confidently expect the hedge to assure a price of $2.75 per bushel.[1] In contrast, if the basis is variable and unpredictable, the return from placing a hedge also is uncertain.

In recent years, bases faced by corn producers in Western New York have been highly variable relative to their mean level. For instance, the basis at Batavia averaged $0.29 per bushel for the 11 years, 1972–82, with a standard deviation of $0.20. Moreover, selective hedging strategies using naive forecasts of basis do not appear to be very successful in achieving their objectives (Querin and Tomek). Thus, if better forecasts of a local basis were available, hedging programs could be improved.

The model is specified in the next section. Then, the empirical results are described and appraised. Finally, the usefulness of the results is analyzed.

Model Specification

Models that have attempted to explain basis behavior for corn for various regions and contract maturities have been complex and not especially successful, at least when evaluated from a forecasting perspective (Kahl; Garcia and Good; Martin, et al.). The econometric model developed in this paper is simple relative to those mentioned above (although more comlex than a naive approach) and is intended to explain the variability only of the November basis from year to year in Batavia, New York. Becuse the number of observations thought relevant to current conditions is small, the number of regressors were limited to maintain a reasonable number of degrees of freedom, but of course primary emphasis was on correct model specification.

Model specification is based on two general concepts. One is the difference between regional prices, the cash price in New York and the cash price in Chicago; the second is the difference between the Chicago cash and futures prices. In principle, regional prices depend on regional supplies and demands and on the transportation costs among regions. The New York price should not exceed the Chicago price by more than the cost of transportation between the two locations, and vice versa. Since, in fact, the historical series of New York prices have been within the band of

---

Patricia D. Taylor is with the Continental Grain Company and William G. Tomek is Professor of Agricultural Economics, Cornell University. This article is based on Taylor's M.S. thesis. We gratefully acknowledge the suggestions of K. L. Robinson, D. G. Sisler, and L. W. Tauer.

[1] In this paper, basis is defined as futures minus the cash price. This is a common definition in the academic literature, but the trade usually defines basis as cash minus the futures price.

Chicago plus or minus a transportation differential, these costs do not appear in the model.[2]

Local supply-demand conditions were represented by a "feed grain deficit" variable. This variable is estimated by the New York Crop Reporting Service as corn, barley and oats production in New York State minus estimated consumption by dairy cows.[3]

Economic conditions in the corn market outside of New York are modeled using national variables, which of course are heavily influenced by conditions in the Cornbelt. Supply was measured by production and by beginning stocks; stocks were included as a separate variable by or summing production and stocks. In both specifications, the statistical fits were poorer than using production alone as a measure of supply. These empirical results are surprising, and we were confronted with the choice of using logical, but poor fitting, models or a better fitting model that seems slightly misspecified. Given the emphasis on forecasting, we opted for the better fitting models, but a choice on empirical grounds can be dangerous. Clearly this specification will need to be checked as more observations become available.

The December futures prices observed on May 1 were considered as a variable to represent factors influencing national prices other than the supply variables mentioned above. This price presumably has the virtue of containing all of the information available on May 1 about the factors influencing corn prices. This specification also has the virtue of using only one variable to capture factors, other than production, that influence price and of using a variable that is observable at planting time when the forecast will be made. But, this specification consistently had small t-ratios and an illogical sign and was dropped from the final models. Thus, in this model, the regional differences in prices depend on local production adjusted for local consumption and on national production.

As mentioned above, the Chicago basis also may vary from year to year, thereby imparting variation to the New York (or other regional) bases. In principle, arbitrage between cash and futures markets should cause the two prices at Chicago to converge as maturity approaches. In practice, the basis will not exactly equal zero, because there are (usually small) costs associated with making and taking delivery (Paul), and in addition the basis would be expected to have some random variability. Still, aside from some small upward trend in the costs of delivery, the systematic variation in the basis at Chicago is likely to be small.

Occasionally, however, the costs of making delivery can be large, a so-called "natural squeeze" (Paul). There is a greater potential for such a squeeze when open interest is large relative to the deliverable supply of corn at acceptable delivery points. In this situation, the squeeze potential is defined by the cost to those short futures of moving corn into deliverable position. In this sense, national stocks of corn are unimportant in explaining the basis at maturity; the question is, how much corn is available for delivery and at what cost relative to the potential demand for delivery? Thus, Chicago stocks and open interest variables were used. The preferred specification was in terms of separate variables rather than as a ratio.

Both linear and semilogarithmic specifications were tried. Logically, a case can be made for a curvilinear relationship since limits exist on the magnitudes of bases; transportation costs limit regional price differences, and the "squeeze potential" is limited by the cost of making good delivery. In practice, the linear specification gave better R^2s, Durbin-Watson statistics, and t-ratios than the semilogarithmic specification for this sample period. Thus, within the range of observations, a straight line is the preferred specification.

Empirical Results

The results for two specifications fitted by OLS are given in Table 1; the specific definition of variables is reported in Table 2. The signs of the coefficients are consistent with expectations. The basis narrows (i.e., Chicago

[2] In examining regional prices in cash markets, it is more likely that Batavia, NY is linked to Toledo, Ohio than to Chicago. One could think of the Chicago-Batavia basis as having the two components Toledo cash minus Batavia cash and Chicago futures minus Toledo cash. The Western New York area, however, is a local island of surplus corn production within a region that has a feed grain deficit. Hence, there is some scope for the independent movement of Batavia prices relative to other prices.

[3] As alternative specifications to the feed grain deficit variable, corn production and on-farm stocks were considered. The production variable was defined for a 15 county area in Western New York that has a local surplus of corn for grain. Specifications based on such variables gave poorer statistical results; apparently the added information contained in the deficit variable is useful in explaining local prices.

Table 1. Corn basis equations for Western New York, 1972–1982

Equation	Inter	USPRO	NYDEF	DELCN	OPINT	$\bar{R}^2$	S_u	D-W
	Variables[b]							
(1)	1.58	−0.207	0.225	−0.009	0.0015	.85	.077	1.99
	(7.58)[a]	(5.93)	(2.79)	(1.21)	(3.63)			
(2)	1.51	−0.196	0.287	—	0.0013	.84	.079	1.31
	(7.29)	(5.64)	(4.50)		(3.31)			

[a] t-ratios in parentheses (H_0: $\beta_i = 0$).
[b] See Table 2 for variable definition.

prices decline relative to New York prices), *ceteris paribus,* as national production increases. The basis widens as the feed grain deficit in New York becomes less negative (i.e., as New York production grows relative to demand, New York prices drop relative to Chicago).

The evidence also supports the notion that a large open interest, other factors remaining the same, increases the basis, while a large deliverable supply of corn in Chicago reduces the basis. The coefficient of the latter variable has a t-ratio of only 1.2, but the inclusion of this variable improves the Durbin-Watson statistic and has little effect on the coefficients of the other variables. On balance, equation (1) was selected as the preferred alternative of the specifications considered.

Table 2. Definition of variables for basis analysis[a]

Symbol	Definition[b]	Units	Mean
B	Basis = Futures minus cash (monthly average using daily prices)	$ per bu.	0.29
USPRO	Corn production in U.S.	bil. bu.	6.61
NYDEF	Feed deficit in New York State	mil. ton	−0.762
DELCN	Chicago stocks of corn, first Friday in November	mil. bu.	8.765
OPINT	Open interest in the December corn futures on first trading day of November	mil. bu.	228.9

[a] Sample period is the 11 years, 1972–82.
[b] Data are from the following sources: Daily futures prices, Chicago stocks of corn, and open interest are from the Chicago Board of Trade, *Statistical Annual*, various issues. Daily cash prices were obtained from a major corn buyer in Batavia, NY and are unpublished. Corn production is from the USDA, *Crop Production Annual Summary*, various issues. Feed deficit variable is published by the New York Crop Reporting Service, *New York Crop and Livestock Report*, March issues.

Given the limited number of degrees of freedom and the data mining (pretest) procedures used, the t-ratios for the final model, no doubt, exaggerate the quality of results (see e.g. Wallace). We did, however, examine partial regression leverage plots for selected model specifications.[4] Based on these plots, several observations appeared to be influential, particularly the year 1980, and equations (1) and (2), as well as several other specifications, were refitted with this observation deleted. The deletion had little influence on the magnitude of the coefficients, while it did reduce the t-ratios. Thus, the effect of 1980 seems to be beneficial, and the results do not seem to reflect aberrant observations.

For the full data set, the coefficients of OPINT and USPRO were insensitive to alternative model specifications, and these variables had the largest t-ratios. The coefficients of NYDEF and DELCN were sensitive to alternative model specifications and had smaller t-ratios. Thus, while equation (1) seems to be the preferred specification, the correctness of this specification and the stability of results need to be re-examined as more observations become available.

Hedging decisions often are based on naive forecasts of the basis, and equation (1) is evaluated relative to such procedures. Specifically, this year's basis is forecast as equal to last year's basis and as equal to the average of the three previous years' bases. Evaluation statistics for the three methods, using the sample period observations, are shown in Table 3 (also see figure 1). Equation (1) does not have any turning point errors and has bet-

[4] These plots show the net scatter of observations for each slope coefficient, taking account of the other variables in the model, and hence help indicate whether the relationship depends on the bulk of the sample or is highly influenced by a single observation (Belsley, et al.).

Table 3. Evaluation statistics of forecast performance for New York basis, 1972–82

Forecast Method[a]	Turning-point error[b] Type I	Type II	RMSE	Theil's U_2
	———— % ————		(\$/cwt.)	
$\hat{B}_t = B_{t-1}$	0	100	0.187	1.00
$\hat{B}_t = \frac{B_{t-1} + B_{t-2} + B_{t-3}}{3}$	100	29	0.190	1.01
Equation (1), Table 1	0	0	0.057	0.30

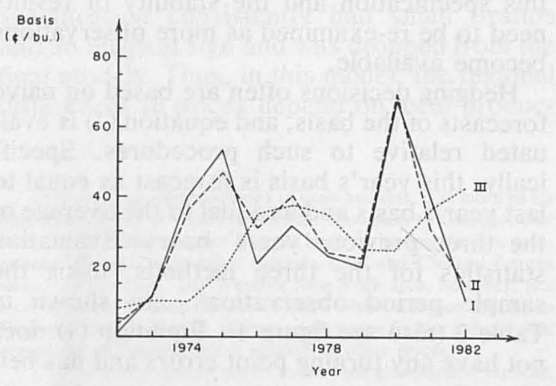
[a] B = basis. Hence, $\hat{B}_t = B_{t-1}$ is a naive forecast of basis.
[b] Type I error occurs when there is no turning point, but one is forecast; Type II error occurs when a turning point is observed, but none is forecast. A 100% erro rate means that the model was wrong in all the instances of that type.

ter goodness-of-fit measures than do the naive forecasts.

Usefulness of Results

Although equation (1) performs well in the sample period, the forecast of the November basis depends on variables whose values are unknown until early November. Unless high quality forecasts of the regressors for the forecast period (ancillary forecasts) can be obtained, actual forecasts of the basis can be seriously in error. Thus, the evaluation of the previous section overstates the usefulness of the results.

Another problem, which all of the forecasts have in common, is that the variable being forecast is a monthly average of daily prices, while hedges are placed and lifted at specific, daily prices. The forecasts do not take account of intramonthly variability in the basis, which had standard deviations as large as 13 cents (1977) and as low as 3.8 cents (1972).



I = actual basis, II = econometric model, III = three-year moving average

Figure 1. Basis in Western New York, November 1972–1982

Some relatively simple methods of making ancillary forecasts were developed, and they appear to be helpful and tractable. These forecasts, however, are subject to substantial error. Ancillary forecasts can be obtained from the following relations:

USPROD = USDA planting time crop estimate,

NYDEF = 11.1 − .135t, $r^2 = .63$, where t = 72, 73, etc.,

or NYDEF = .514 − 135t + .0248t², $\bar{R}^2 = .81$, where t = . . . , −1 in 1976, 0 in 1977, 1 in 1978, etc.,

OPINT = −15.2 + 1.43(OPINT May 1), $r^2 = .63$,

DELCN = sample mean.

Fitting the final crop size estimate to the planting time estimate suggests that the planting time estimate is an unbiased forecast. Thus, although this early season forecast is subject to large errors (see below), it appears to be the best available forecast.

Since the feed grain deficit has a definite downward trend, simply extending this trend seems like a useful approach to forecasting, especially since good early estimates of New York State production of feed grains are not available. A quadratic trend equation gives a better statistical fit than the linear form for the historical sample, but the quadratic function implies that the minimum deficit occurred in 1980 and that the deficit has increased since then. Such a reversal is not consistent with expected future changes in the deficit, and the linear approximation is perhaps the preferred alternative in estimating NYDEF. (NYDEF is treated as positive in the trend equations, and larger values of NYDEF imply larger deficits.) In any case, the trend equations should be updated as new observations become available.

Open interest on May 1 appears to be related with the open interest in early November. But the stocks of corn in Chicago in November had no easily predictable pattern, and the historical mean seems to be the best available forecast.

Thus, given equation (1), ancillary forecasts of the regressors are made, as explained above, to give a preliminary point forecast of the basis. As the season progresses and crop conditions become clearer, the forecast of the basis can be revised.

The effect of errors in the ancillary forecasts is illustrated for 1980 (see Table 4), a year of extreme basis behavior. The final, observed values of the regressors in equation (1) gave a close estimate of the actual basis—a forecast of $0.70 per bushel versus the actual $0.68. If, however, the early season forecasts of the regressors are used, then the forecast would have been $0.48, 20 cents below the actual value.[5]

The traditional standard error of forecast for 1980 is $0.116, but it is misleadingly small, because it does not take account of the potential errors in forecasting the regressors (Feldstein). However, a standard error that takes account of the errors in the regressors is difficult to compute and typically gives confidence interval so large that it is not useful for decision-making.

In contrast to the difficulty of obtaining accurate ancillary forecasts, the problem of using an average monthly forecast of the basis does not appear to be a serious limitation of the analysis. Indeed, the variability of daily prices around the average helps the hedger achieve the objective of the hedge and perhaps can help offset the effects of an erroneous forecast of the average. To illustrate whether forecasts of a monthly basis could help achieve a target return, a selective hedging program was simulated for the 10 years, 1972–81.[6] In this simulation, the hedger is assumed to sell futures if the futures price minus the forecast basis reached a pre-specified target price. This target price is based on the cost of production of corn, as estimated from Cornell cost accounts, plus $0.10 per bushel. The forecasts were derived from equation (1), above. Daily closing prices in futures were followed, and when the closing price indicated that the target could be met, futures were sold at the midpoint of the next day's opening range. If a hedge is placed, it is held until November.

In the 10 year period, seven hedging opportunities occurred based on the forecast of the average basis, and for these years, the subsequent daily prices in November did indeed permit the target to be obtained. In 1975, the targeted return could have been obtained by completing the hedge in just 2 of the 19 business days of the month, but in the other six years, the targeted return could have been obtained in half or more of the business days. That is, the variability of the daily bases around their mean provides a large probability that the hedge can be completed at the targeted level. Thus, given an accuract forecast of a monthly basis, the hedger likely can complete the hedge at a relatively favorable basis on a particular day provided that the hedger follows cash and futures price movements closely.

Table 4. Forecasts for 1980 basis

Variable[a]	Early forecasts	St. error forecast	Final Estimates
USPRO	7.2	0.705	6.648
NYDEF	−0.334	0.260	−0.328
DELCN	8.77	4.58	5.886
OPINT	364.8	81.2	418.68
B	.48	0.116	0.70[b]

[a] See Table 2 for units.

[b] Forecast value using the final reported values of regressors; the observed basis was 0.68.

Conclusions

It is feasible to develop a relatively simple model to explain the year-to-year variability of a specific, local basis. The harvesttime basis for corn in Western New York appears to be strongly related to the size of the national corn crop, the feed grain deficit in New York State, and the open interest in the December contract. The amount of corn in deliverable posi-

[5] The two naive forecasts, however, were $0.20 and $0.25, respectively, and using these forecasts, the disappointment would have been even larger. Indeed, if the hedger in 1980 had set a target return of $2.50 per bushel, futures would have been sold at $2.75, given a basis forecast of $0.25. But with an actual basis turns of 68 cents, then the hedge would have provided a return of only $2.07 ($2.75 minus $0.68).

[6] The year 1982 of the sample period was not included in the simulations, as some additional data needed for the simulations was not available at the time our analysis was done. For additional information about the simulations, see Taylor.

tion in Chicago in early November also may influence the size of the basis, although this conclusion is somewhat tenuous. In principle, forecasts from this model should aid in hedging decisions, but the basis depends on variables which also must be forecast if the model is going to be used for decision-making at planting time or during the growing season. Moreover, traditional estimates of the standard error of forecast will give confidence intervals around the point forecast that are misleadingly small. Thus, while the work reported in this paper offers a foundation for improving basis forecasts, the difficulty of making precise ancillary forecasts of the regressors is a serious limitation of models of basis behavior.

References

Belsey, David A., et al. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: John Wiley & Sons, 1980.

Feldstein, Martin S. "The Error of Forecast in Econometric Models when the Forecast-Period Exogenous Variables are Stochastic." *Econometrica* 39(1971): 55–60.

Garcia, Philip, and Darrel L. Good. "An Analysis of the Factors Influencing the Illinous Corn Basis, 1971–81," in *Applied Commodity Price Analysis, Forecasting, and Risk Management*. Des Moines, Iowa: Proceedings of the NCR-134 Conference, 1983.

Kahl, Kandice H. "Changes in the Chicago Corn Basis, 1960–1975." *Agricultural Economics Research* 34(1982):25–29.

Martin, Larry et al. "Factors Affecting Corn Basis in Western Ontario." *American Journal of Agricultural Economics* 62(1980):107–112.

Paul, Allen B. *Treatment of Hedging in Commodity Market Regulation*. USDA Technical Bulletin 1538, April 1976.

Querin, S. F., and W. G. Tomek. *Hedging Strategies Utilizing Technical Analysis: An Application to Corn in Western New York*. Cornell Univ. Agricultural Economics Research 83-12. February 1983.

Taylor, Patricia D. "The Western New York Basis for Corn at Harvest, 1972–1982," M.S. thesis, Cornell University, August 1983.

Wallace, T. Dudley. "Pretest Estimation in Regression: A Survey." *American Journal of Agricultural Economics* 59(1977):431–443.