FORECASTING SOUTH CAROLINA TOMATO PRICES PRIOR TO PLANTING

Gary J. Wells

South Carolina producers supply tomatoes on a national scale during an approximately five-week market window centered in June. This market window follows a six-month period dominated by Florida production and to a lesser extent Mexican exports. Even though South Carolina is the major East Coast supplier of tomatoes during its market window, the success or failure of South Carolina marketings depends on the stage set by Florida and Mexico. That is, the price level obtained by South Carolina producers for their tomatoes tends to be influenced by the volume of tomatoes delivered to market prior to South Carolina's harvest. South Carolina, Florida, and Mexico serve many of the same markets (i.e., Northeast terminal markets). As a result, the influence of the volume prior to South Carolina's harvest on the state's average tomato price can be used to provide farmers with a price estimate before planting. Specifically, a model is constructed to forecast the average spring tomato price received by South Carolina producers. This forecast must be available to producers by mid-February if it is to serve as a decision-making tool.

METHOD OF ANALYSIS

The model used is formulated with the objective of forecasting South Carolina's spring tomato prices as early as February. Consequently, the range of information that can be used is limited. The basic hypothesis is that South Carolina tomato prices are influenced predominantly in the following manner.

(1) PSC = f(QFLA, QMEX, QSC, RPEF)

where

PSC = South Carolina's average producer price ($/cwt)
QFLA = Florida early season tomato production
QMEX = Mexico early season tomato exports to the U.S.
QSC = S. C. spring tomato production
RPEF = U. S. expenditures on food.

The choice of specific measures of the independent variables depends on the availability of the measures when the forecast needs to be made, in February. Of the specific measures available in February, selection is based on the contribution that the selected measure makes to the objective of accurate forecasting.

The food expenditure measure, RPEF, selected is the previous year's fourth quarter real expenditure on food. To find this measure U.S. food purchases are adjusted by the GNP deflator for personal expenditures. The fourth quarter expenditure value is chosen instead of the previous year's spring quarter value because the expenditures are seasonally adjusted. In addition, the fourth quarter value has a greater probability than the previous spring measure of reflecting a trend in food expenditures that may have an impact on the upcoming spring tomato consumption.

Several variables could be used as a measure or proxy for Florida early season production, QFLA. For example, acres planted or acres harvested or production could be used. Production appears to be the most reasonable measure of volume to use but the inability to accurately predict Florida's early season production as early as February makes it a less desirable choice. The measure of Florida's volume used is the acres planted for harvest in the fall, winter, and spring quarters.[1] This measure is chosen because it is accurately predicted by February and it is statistically superior to other measures investigated.

Different estimates of South Carolina production, QSC, were evaluated but none proved satisfactory. Initially this outcome was believed to be due to the difficulty in estimating production prior to planting, but production information available after harvest indicated that QSC has a minor impact on PSC in relation to QFLA. A regression of South Carolina's price on QFLA, RPEF, and QSC increased the coefficient of determination from

Gary J. Wells is Assistant Professor, Department of Agricultural Economics and Rural Sociology, Clemson University.

[1]This volume measure is available in the January issue of *Vegetables* (USDA). The measure is of actual plantings for fall and winter harvest and an estimate of plantings for spring harvest. This spring estimate is generally very accurate because of the stability of season-to-season plantings and because the estimate is made only a month before actual plantings in Florida.

.8496 to .8548 over the same equation without S.C. production. In addition the coefficient estimate for QSC had an unexpected sign and was not significantly different from zero at the 10 percent level. As a result QSC was dropped from consideration.

Mexican shipments to the U.S. from the beginning of the early season (late December or early January) through January, MEX, were evaluated, but the estimated coefficient was insignificant at the 10 percent level.[2] Therefore, for the initial model estimated, Mexican shipments to the U.S. were dropped from consideration. However, Mexican shipment information is incorporated into a second model specification that is discussed hereafter.

The initial model estimated is

(2) PSC = g(QFLA, RPEF)

where

PSC = South Carolina average spring producer price ($/cwt)
QFLA = total acres planted in Florida for the fall, winter, and spring quarters (1000s),
RPEF = Real U. S. personal expenditures on food for the previous year's fourth quarter ($B, 1972=100, seasonally adjusted).

Data are from USDA's *Vegetables* for PSC and QFLA and the Commerce Department's *Survey of Current Business*, for RPEF.

EMPIRICAL RESULTS

Ordinary least squares regression is used to estimate the coefficients of equation 2 for the period 1960-1979. The resulting estimation is

$$(3)\quad PSC = \underset{(5.94)}{10.64} - \underset{(.00012)}{.00068}\,QFLA + \underset{(.024)}{.204}\,RPEF$$

R^2=.85 (standard errors in parentheses).

The estimated coefficients have the expected signs and are significant at the 1 percent level. The R^2 is somewhat low for a forecasting equation, but subsequent consideration of the impact of Mexican shipments to the U.S. improves the R^2.

The ability of equation 2 to forecast accurately depends in large measure on the assumption that the regression relationship is constant over time. That is, there are no structural changes. If the impact of Mexican shipments to the U.S. changed over time or if consumers began to view tomatoes differently with respect to their food budget, the coefficients would be expected to change as a response to these structural changes. As a check for a possible structural change, several techniques were used to test the constancy of the regression relationship in equation 2. A description of the techniques used to test for a structural change is given by Brown et al. These techniques are available in a program entitled TIMVAR which is designed to investigate the possibility of a gradual structural change as well as a structural change at one point in time.

The TIMVAR results for the 1960-1979 data suggest that a one-shot structural change did occur between the 1973 and 1974 seasons. A Chow test provides additional evidence of this structural change. The method used to incorporate the structural change is to add slope and intercept shifters to equation 2. Also, Mexican shipment totals from the beginning of the winter export season until the earliest report in February are included for years after 1973 (USDA 1968-1979). For 1973 and earlier, Mexican shipments are assumed to be zero.[3] This formulation allows consideration of the impact of Mexican exports to the U.S. in recent years while ignoring smaller and erratic early season shipments prior to 1974. The reformulated model is

(2′) PSC = j(QFLA_S, RPEF_S, MEX′, INT).

The S represents the shift component of the independent variables of equation 2, INT is the intercept shifter, and MEX′ is the adjustment for Mexican shipments. The S and INT variables equal one if the year is 1974 or later and otherwise they equal zero. The regression results for the 1960-1979 data are

$$(4)\quad PSC = -2.13 + QFLA\,(-\underset{(.0001)}{.0002} - \underset{(.0002)}{.0011}\,S) + RPEF(\underset{(.022)}{.158} + \underset{(.120)}{.478}\,S) - (MEX')\underset{(.001)}{.005} - 30.18\,INT$$

R^2 = .98 (standard errors in parentheses).

These results are encouraging, but the ability of the equation to forecast is the true measure of the equation's worth. Equations 2 and 2′ are considered in this light.

FORECAST EVALUATION

In this section we directly evaluate the usefulness of equation 2 as a one-period-ahead

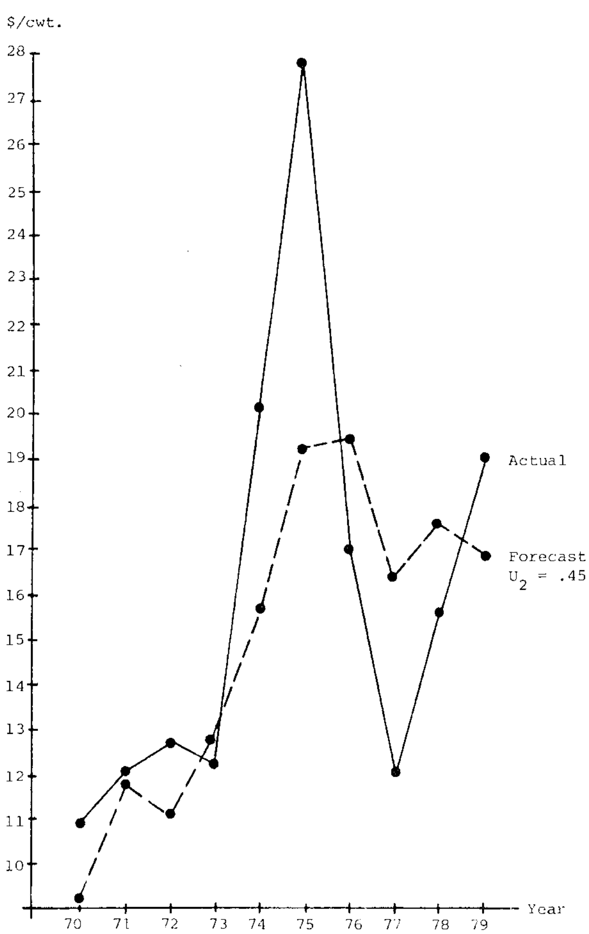
[2]Mexican shipments were evaluated by using a shorter data series, 1968-1979, because of the lack of comparable shipment information prior to 1968 (USDA 1968-1979).

[3]Mexican early season exports to the U.S. averaged 1194 thousand cwt between 1968 and 1973 and 1760 thousand cwt after 1974. The standard deviation during the earlier period was 575 thousand cwt compared with 449 thousand cwt during the later period.

forecasting tool. Equation 2′, however, cannot be evaluated directly because of the multicollinearity introduced by the dummy variables as recent observations are dropped off. That is, as more and more of the recent observations are dropped off to allow forecasts for the recent years to be made, the remaining dummy variables approach having only zero values. An indirect method of evaluation therefore is employed for evaluating equation 2′. Equation 2 is considered first.

The forecast ability of equation 2 is investigated by forecasting the one-period-ahead price of South Carolina tomatoes for the 10 most recent years. That is, the 1970 price is forecasted by using an equation estimated with observations up to 1969. The price for 1971 is then forecasted after the equation is reestimated to include the 1970 observation. This process is continued up to the forecasting of the 1979 price.[4] Figure 1 is a graph of these

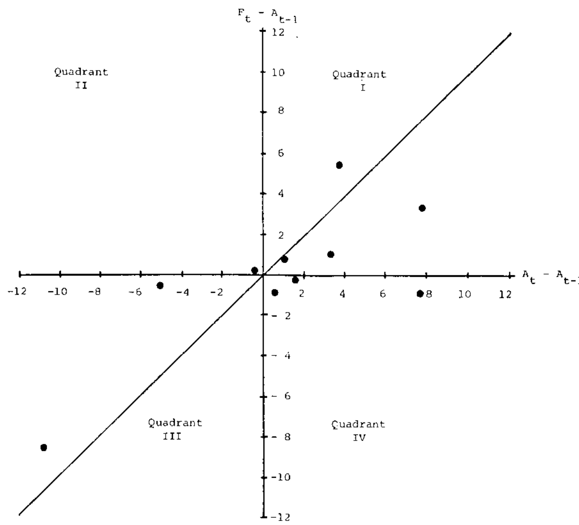
FIGURE 1. ONE-PERIOD-AHEAD FORECASTS 1970-1979



results and the actual prices. Theil's second inequality coefficient, U2, is also provided.[5] A value of one for U2 suggests that the prediction method does no better than a simple no-change model. A value of zero suggests perfect forecasting, and values greater than one suggest a forecasting technique worse than the naive no-change model. Therefore numbers closer to zero are preferred to numbers farther away. However, no test statistic is available. The U2 value of .45 indicates that the estimate contains 45 percent of the error that would have been observed if the forecaster had limited his forecasts to a no-change model (Theil).

As a further evaluation of equation 2, the change in year t's forecasted price, F_t, from year (t−1)'s actual price, A_{t-1}, is compared with the actual change from the previous year. As a result, $F_t - A_{t-1}$ is plotted against $A_t - A_{t-1}$. The results are shown in Figure 2 for the 10-

FIGURE 2. ACTUAL VERSUS PREDICTED CHANGES, 1970-1979



year period studied. The diagonal line drawn through quadrants I and III represents the line of perfect forecast. Points in quadrants I and III but not on the diagonal represent correct forecasts of the direction of change from the previous period but not correct forecasts of the magnitude of the change. Points in quadrants II and IV represent incorrect forecasts of direction and magnitude of the change. Six of the 10 forecasts are in quadrants I and III. The forecast in quadrant II is near the origin, suggesting that the model forecasted a small positive change when a small negative change occurred. Two of the three forecasts in quadrant IV are also for small changes (i.e., they are less than $2/cwt).

The results of equation 2′ should not be overlooked, however. The estimation results are encouraging. To investigate the forecasting

[4]The estimation equation for each forecast may be obtained from the author.

[5]U2 is calculated by using changes alluded to by Theil and discussed by Leuthold. U2 is defined as $\sqrt{\Sigma(F_t-A_t)^2} / \sqrt{\Sigma A_t^2}$ where A = acutal changes and F = forecasted changes.

potential of this equation, a technique entitled PRESS is used. Each year's price is predicted on the basis of all information available from other years. As an example, to predict 1965's price, observations from 1960-1964 and 1966-1979 are used to estimate the prediction equation. The predicted price is then compared with the actual value and the square of the difference is calculated. The sum of these squared differences is the PREdiction Sum of Squares. As a means of comparison, the PRESS results of equation 2 are also calculated. The PRESS for equation 2 and equation 2′ is 156.8 and 21.2, respectively. The reduction in sums of squares given equation 2′ appears to be substantial in comparison with equation 2. It is concluded, therefore, that equation 2′ has the potential of providing superior forecasts in relation to equation 2.

SUMMARY AND CONCLUSIONS

The objective of this study is to construct a one-period-ahead price forecasting model for South Carolina tomatoes. The forecasted price needs to be available for producer use prior to planting (e.g., in February). Florida plantings for the fall, winter, and spring quarters and real expenditures on food are found in an initial model to generate forecasts of the South Carolina average price superior to those from a naive no-change model. The U2 value based on forecasts for 10 years for the initial model is .45. It indicates that this model has only 45 percent of the error that would have been observed if the forecaster had limited his predictions to a no-change model. The key to success appears to be the impact of a dominant producing region like Florida on the prices of producing regions that harvest after the dominant region.

Statistical analysis with TIMVAR, a computer regression package, indicates that a structural change occurred between 1973 and 1974. Adjustments to the initial model are made by including slope and intercept shifters. Additionally early season Mexican shipments to the U.S. for years after 1973 are introduced. The results are not directly verifiable, but an indirect verification technique, PRESS, indicates that adjusting for the structural change may provide improved forecasts.

An approach similar to the one used here may prove valuable for other vegetable-producing regions that market their crops just after a dominant producer completes marketing. The price impact of the dominant producer may overshadow the impact that the region of concern is capable of initiating. If this is the case, construction of a price forecasting model is possible. Forecasts made by a method similar to the one described here would need to be provided yearly by the researcher. Continued researcher input is necessary because of the need to incorporate the most recent observation into the estimation equation.

REFERENCES

Brown, R. L., J. Durbin, and J. M. Evans. "Techniques for Testing the Constancy of Regression Relationships Over Time (with comments)." *J. Roy. Statist. Soc.* Ser. B 37(1975):149-92.

Leuthold, Raymond M. "On the Use of Theil's Inequality Coefficients." *Amer. J. Agr. Econ.* 57(1975):344-6.

Theil, Henri. *Applied Economic Forecasting.* Chicago, Ill.: Rand McNally & Co., 1966.

U. S. Department of Agriculture. *Agricultural Outlook,* Jan.-Feb. issue.

U. S. Department of Agriculture. *Vegetables (Fresh Market),* 1960-1979.

U. S. Department of Agriculture. *Weekly Summary Shipments - Unloads,* 1968-1979.

U. S. Department of Commerce. *Survey of Current Business,* 1960-1979.