TANDEM FORECASTING OF PRICE AND PROBABILITY -- THE CASE OF WATERMELON

by

James E. Epperson and Stanley M. Fletcher*

*Senior authorship is not assigned, associate and assistant professors, respectively, Agricultural Economics Department, Georgia Experiment Station, University of Georgia, Experiment, GA 30212.

Tandem Forecasting of Price and Probability -- the Case of Watermelon

James E. Epperson and Stanley M. Fletcher

Abstract

The purpose of this paper is to show how probability prediction can be incorporated with price prediction to enhance the usefulness of forecast information and provide greater intuitive appeal in its use. Empirical application encompasses forecasting in the watermelon industry to demonstrate the power and appeal of the approach.

TANDEM FORECASTING OF PRICE
AND PROBABILITY -- THE CASE OF WATERMELON

In recent years, price forecasting has been accomplished using traditional causal models, noncausal models such as the auto-regressive-integrated moving-average model (ARIMA), and even composites of causal and noncausal models (Just and Rausser; Martin and Garcia; Oliveira, et al.; Naylor, et al.; Nelson; Leuthold, et al.; Bechter and Rutner; Zellner; and Bates and Granger). Review of the literature revealed two interesting advances in empirical price forecasting. Menkhaus and Adams employed a predicted discrete variable exogenously in an effort to improve the accuracy of predicting turning points with a price forecasting model. The discrete variable was predicted using discriminant analysis. Ferris reports that the Michigan State University Agriculture Model predicts a price array with an associated probability for each price in the array. Each price in the array is dependent on level of yield; and each yield category has an associated probability of occurrence.

The approach used in this paper is somewhat akin to the two procedures just described, yet quite different. The idea behind the approach is that those who benefit from price forecasts would also benefit from knowing the probability that the predicted price will cross some predetermined threshold of importance (e.g., trigger price). This is analogous to the weatherman's forecast of quantity of rain and the associated probability of rain. Thus, most people understand the concept of probability in this context which suggests that price forecasting in this vein would be useful.

Model

The model used for forecasting the price and the probability that the predicted price will cross some predetermined threshold of importance follows the model from Heckman's work :

(1) $Y_1^* = \gamma_1 Y_2 + \beta_1 X_1 + u_1$

(2) $Y_2 = \beta_2 X_2 + u_2$

where

Y_1^* = a latent variable indicating the propensity of the price forecast to exceed or be equal to the predetermined threshold of interest;[1]

Y_2 = price in t+1 (t = year);

X_1 = a vector of explanatory variables for Y_1^*;

X_2 = a vector of explanatory variables for Y_2;

γ_1 = a scalar coefficient;

β_1, β_2 = vectors of coefficients; and

u_1, u_2 = random error terms with a bivariate normal distribution.

The reduce form of the above equations can be written as

(3) $Y_1^* = X\pi_1 + \nu_1$

(4) $Y_2 = X\pi_2 + \nu_2$

where

X = all the exogenous variables in X_1 and X_2;

π_1, π_2 = vectors of coefficients; and

ν_1, ν_2 = random error terms with a bivariate normal distribution.

For this model, the conditions for identification are that u_1 and u_2 be independent, or else there is at least one variable in X_2 not included in X_1 (Maddala, p. 120).[2]

Two procedures exist to obtain consistent estimates for equations (1) and (2). These are maximum likelihood estimation and the two-stage procedure specified by Maddala (pp. 121, 122, 244, 245). The maximum likelihood procedure provides asymptotically efficient estimates but the procedure is computationally more cumbersome than the two-stage procedure. Given the goal is forecasting, either procedure is viable since both provide consistent estimates. Thus, both procedures were used in this study for comparison. The Davidson-Fletcher-Powell algorithm was used for maximum likelihood estimation as provided in the numerical optimization computer package of Goldfeld and Quandt.

The likelihood function for a sample is the product of the likelihood functions for the individual observations. Thus, letting n_1 be the set of observations when Y_1^* equals zero and n_2 the set of observations when Y_1^* equals one, the likelihood function for the model is

$$L(\gamma_1/\sigma_1, \beta_1/\sigma_1, \beta_2, \sigma_2) = \prod_{n_1} (1/\sigma_2) \int_{X\pi_1/\sigma_1}^{\infty} g(\nu_1, X_2\beta_2/\sigma_2 - Y_2/\sigma_2) d\nu_1$$

(5)

$$* \prod_{n_2} (1/\sigma_2) \int_{-\infty}^{X\pi_1/\sigma_1} g(\nu_1, X_2\beta_2/\sigma_2 - Y_2/\sigma_2)\, d\nu_1$$

where g (,) is a bivariate normal distribution.

Maximizing the above likelihood function, one obtains the estimates of the coefficients of the structural equations.

The two-stage procedure specified by Maddala for estimating the model is much simplier than the likelihood function approach. In this model one should note that the reduced form (equation 4) and the structural form (equation 2) for the price equation are the same. The procedure consists of estimating π_2 and β_2 by ordinary least squares. The predicted value of Y_2, $\tilde{Y}_2 = X\tilde{\pi}_2$, is substituted into equation (1). Because Y_1^* is observed only as a dichotomous variable, the parameter estimates for equation (1) can only be estimated up to a scale, γ_1/σ_1 and β_1/σ_1, where $\sigma_1^2 = Var(\nu_1)$. Thus, equation (1) is estimated using probit ML in the form

(6) $$Y_1^* = \frac{\gamma_1}{\sigma_1}\tilde{Y}_2 + \frac{\beta_1}{\sigma_1}X_1 + \frac{u_1}{\sigma_1}.$$

The asymptotic covariance matrix for equation (2) is the same covariance matrix from the ordinary least squares estimation of equation (2). However, the asymptotic covariance matrix for equation (6) is not the estimate from the probit ML. Rather, the formula for the covariance matrix is

(7) $$Var(\tilde{\gamma_1}/\sigma_1, \tilde{\beta_1}/\sigma_1) = (G'V_0^{-1}G)^{-1} + d(G'V_0^{-1}G)^{-1}\, G'V_0^{-1}(X'X)^{-1}V_0^{-1}G(G'V_0^{-1}G)^{-1}$$

where

V_0 = covariance matrix of the probit ML estimate of π_1 (equation 3);

$G = (\pi_2, J_1)$;

J_1 = matrix consisting of 1's and 0's such that $XJ_1 = X_1$; and

$d = (\gamma_1/\sigma_1)^2\sigma_2^2 - 2(\gamma_1/\sigma_1)(\sigma_{12}/\sigma_1)$.

Empirical Example

The tandem forecasting procedure just described may be applied to any commodity. But for purposes of this inquiry the commodity of focus is watermelon produced in the U.S. and marketed during the summer season, July through September. Our intent is to provide price forecast information at a time when producers of summer watermelons could perhaps benefit the most -- just prior to planting, say, February.

For a summer forecast delivered in February, our hypothesis was that the season average price for summer watermelon is a function of previous prices or perhaps values per acre of watermelon and possible competing crops, cost of production, price forecasts available for competing crops, income, population, and possible previous summer weather in the major population areas of the North.

Observed prices and values per acre for watermelon and possible competing crops as explanatory variables were employed by year and two and three-year moving averages.[3] Investigation in this vein was considered since Wall and Tilley found that current price affects the magnitude of watermelon production in Florida for three years hence. Possible competing crops for watermelon considered in this study include cantaloupe for all watermelon producing regions of the U.S., corn and soybeans for the eastern U.S., grain sorghum for the Southwest, and tomatoes for processing in the West.

Available price forecasts for possible competing crops encompassed futures prices for corn and soybeans in September, quoted in the previous February.

Weather in the highly populated northern U.S. was considered as an explanatory variable because some watermelon shippers have indicated that hot, dry summer weather in the North encourages watermelon consumption, while, conversely, cool, rainy weather dampens consumption.

Data used for equation estimation were for the periods 1954 through 1983. The price prediction equation shown in table 1 was found to possess the greatest prediction capability within the confines of our hypothesis. The price equation results from both the two-stage procedure and the maximum likelihood procedure are identical which is not unexpected given the formulation of the model. However, the standard errors differ between the two procedures.

In table 1, WP_{t+1} is the season average price of watermelon in year t+1. $WP2A_t$ is a two-year moving average of the season average price of watermelons. UWP_t is the season average price of watermelon in year t. CPI_t is the consumer price index. DPI_t is U.S. disposable income in year t. CSS_t represents shipments of cantaloupe in the summer season in year t. Accuracy of shipment data for watermelon is questionable. Indeed, cantaloupe shipments as an exogenous variable performed much better than watermelon shipments. Shipment data from the Agricultural Marketing Service, USDA, were used in the final analysis since the Crop Reporting Service discontinued reporting production statistics on watermelon and cantaloupe after 1981.

The exogenous variables considered for the price prediction equation were also considered for the probability equation (table 1).

Table 1. Coefficient Estimates for the Price and Probability Prediction Models, 1954-1983

Variable	Two-Stage Procedure: Price Equation	Two-Stage Procedure: Probability Equation[a]	Maximum Likelihood Procedure: Price Equation	Maximum Likelihood Procedure: Probability Equation
	(asymptotic t-statistics are in parentheses)			
WP_{t+1}		5.1426 (1.664)		8.4278 (9.973)
$WP2A_t$	0.3421 (1.466)	2.5422 (0.594)	0.3421 (1.609)	0.9548 (1.118)
UWP_t		-7.5156 (-1.213)		-8.8371 (-10.763)
CPI_t	-0.0638 (-4.791)		-0.0638 (-5.251)	
DPI_t	0.854E-3 (5.511)		0.854E-3 (6.036)	
CSS_t	-0.167E-3 (-1.940)		-0.167E-3 (-2.127)	
Constant	4.2087 (5.255)	-0.0734 (-0.0491)	4.2088 (5.759)	-1.6708 (-1.730)
R^2	0.930			
F^b	82.367			
Likelihood Ratio Test[b]		26.013	115.055	

[a]Predicted WP in t+1 was used in estimating the probability equation with probit ML per Maddala's two-stage procedure outlined in the model section of this paper.

[b]Both the F and the Likelihood Ratio test statistic were statistically significant at $\alpha < .05$.

The threshold criterion for the dummy dependent variable of the probability equation for this inquiry is price in t+1 relative to price in t. Other threshold criteria could have been chosen. The threshold criterion for the dummy dependent variable of the probability equation could have been price in t+1 relative to breakeven price for example.[4]

For the threshold criterion used in this analysis, if the price in t+1 is greater than or equal to the price in t, the dummy variable is assigned a value of 1, 0 otherwise. In other words, the probability equation projects the odds that the price this summer will be greater than or equal to the price of last summer. Thus, the probability equation is incomplete without the explanatory power of WP_{t+1}. In essence, the probability is another measure of how well price is predicted; yet, it is in a form that is intuitively appealing relative to some threshold or trigger price of interest.

The prediction results of the estimated price and probability equations are shown in table 2. Estimation was over the periods from 1954 through 1982, while forecasting was for 1979 through 1983. All available observations were used in the forecasting procedure. Data through 1978 were used to estimate the price and probability models to forecast for 1979. Data through 1979 were used to forecast price and probability for 1980, and so on. Parameter estimates of the models were stable as the sample was updated.

For the purpose of evaluating the forecasting power of the estimated price model <u>ex post</u>, two deterministic models were employed, a no-change and a trend model. The measure used for evaluation is

Table 2. Actual Values and Predictions Using the Price and Probability Equations, 1979-1983

	Price		Probability		
Year	Actual	Predicted[a]	Actual	Two-Stage Predicted	Maximum Likelihood Predicted
	(dollars per cwt)				
1979	4.70	4.45	1	0.70	0.999
1980	5.67	4.90	1	0.96	1.000
1981	4.56	5.59	0	0.39	0.947
1982	4.60	5.11	1	1.00	1.000
1983	4.16	4.53	0	0.80	0.188
MSE (OLS)		0.42			
MSE (No Change)[b]		0.50			
MSE (Trend)[c]		1.32			
Theil U		.92			

[a] Same for two-stage and maximum likelihood.

[b] $WP_{t+1} = WP_t$.

[c] $WP_{t+1} = WP_t + (WP_t - WP_{t-1})$.

known as mean square error (MSE) ex post, table 2 (Granger). Using this criterion, the estimated price model compares favorably. The Theil U statistic was also used as an alternative forecasting evaluation procedure. The value of this statistic (.92) implies that the price forecast equation is better than a naive model which collaborates the MSE ex post measure.[5]

The prediction power of the probability equation is evaluated in table 3. A criterion of 60-40 means that if the predicted probability is 0.60 or greater and the actual value for the dichotomous variable is 1, the probability prediction is correct. If the predicted value is 0.40 or less and the actual value is 0, the probability prediction is also correct. If the predicted and actual values do not conform as described, the probability predictions are deemed incorrect.

In table 3 the percentage of accurate probability predictions remains relatively high as the classification criterion becomes more restrictive. Even for the 70-30 criterion, probability prediction accuracy is better than 50 percent for the two-stage procedure. However, the probability predictions based on the maximum likelihood procedure seemed to be more accurate than the two-stage estimation procedure. Furthermore, the maximum likelihood predictions seem to be more robust across classification criteria. Thus, there seems to be a trade-off between computation complexity and prediction robustness.

Conclusion

This paper endeavored to show how probability prediction can be incorporated with price prediction to enhance the usefulness of forecasting results. This tandem forecasting application is

Table 3. Probability Prediction Accuracy

Classification Criterion[a]	Percentage of Accurate Probability Predictions	
	Two-Stage Procedure	Maximum Likelihood Procedure
50-50	80	80
60-40	80	80
70-30	60	80
80-20	40	80
90-10	40	60

[a]Using the 70-30 criterion as an example, the predicted probability was correct 60 percent of the time (two-stage procedure) during the prediction interval, 1979 through 1983, where the predicted probability was 0.70 or greater when the actual price in year t+1 was greater than or equal to the actual price in year t and 0.30 or less when the actual price in year t+1 was less than the actual price in year t.

intuitively appealing and provides another vantage point from which to evaluate alternative actions. Certainly, the demonstrated procedure is subject to the same specification requirements as other referenced forecasting frameworks.

Footnotes

[1]The latent variable will empirically be observed as one or a zero.

[2]For this study, u_1 and u_2 are not assumed to be independent which implies that at least one variable in X_2 be not included in X_1 for identification.

[3]Given the hypothesized model for watermelon prices includes a two-year moving average of past watermelon prices, this implies watermelon prices follow some sort of an autoregressive process. The process is adequately specified since a hypothesis test on the error term for the equation rejects the null hypothesis of an autoregressive process. Thus, sample separation is possible and maximum likelihood estimation is feasible.

[4]The use of a breakeven price or cost of production as a "trigger price" would have been more intuitively appealing; however, due to data limitations these other options were not available.

[5]If the U statistic lies between zero and one, the forecast is better than a forecast from a naive model. A value of one implies equivalency in forecasting ability while a value greater than one indicates a naive model's forecast is better.

References

Bates, J. M. and C. W. J. Granger. "The Combination of Forecasts," Operations Research Quarterly. 20(1969):451-68.

Bechter, D. M. and J. L. Rutner. "Forecasting with Statistical Models and A Case Study of Retail Sales," Economic Review, Fed. Res. Bank of K.C. 63(1978):3-11.

Ferris, John. "Probability Forecasts on U.S. Corn Prices." Department of Agricultural Economics Staff Paper 81-1, Michigan State University, January 1981.

Granger, C. W. J. Forecasting in Business and Economics. New York: Academic Press, 1980.

Heckman, James J. "Dummy Endogenous Variables in a Simultaneous Equation System." Econometrica. 46(1978):931-959.

Just, R. E. and G. C. Rausser. "Commodity Price Forecasting with Large-Scale Econometric Models and the Futures Market." American Journal of Agricultural Economics. 63(1981):197-208.

Leuthold, R. M., A. J. MacCormick, A. Schmitz, and D. C. Watts. "Forecasting Daily Hog Prices and Quantities: A Study of Alternative Forecasting Techniques." Journal of the American Statistical Association. 65(1970):90-107.

Maddala, G. S. Limited-Dependent and Qualitative Variables in Econometrics. Cambridge: Cambridge University Press, 1983.

Martin, L. and P. Garcia. "The Price-Forecasting Performance of Futures Markets for Live Cattle and Hogs: A Disaggregated Analysis." American Journal of Agricultural Economics. 63(1981):209-15.

Menkhaus, D. J. and R. M. Adams. "Forecasting Price Movements: An Application of Discriminant Analysis." Western Journal of Agricultural Economics. 6(1981):229-238.

Naylor, T. G., T. G. Seaks, and D. W. Wichern. "Box-Jenkins Methods: An Alternative to Econometric Models." Inst. Stat. Rev. 40(1972):123-137.

Nelson, C. R. "The Prediction Performance of the FRB-MIT-PENN Model of the U.S. Economy." American Economic Review. 62(1972):902-17.

Oliveira, R. A., C. W. O'Conner, and G. W. Smith. "Short-Run Forecasting Models of Beef Prices." Western Journal of Agricultural Economics. 4(1979):45-55.

U.S. Department of Agriculture, AMS, CRB. Vegetables for Fresh Market, Acreage, Production and Value 1949-55. Statistical Bull. 212.

U.S. Department of Agriculture, AMS. Fresh Fruit and Vegetable Shipments, Annual Summaries. 1952-1983.

U.S. Department of Agriculture, AMS. Fresh Fruit and Vegetable Prices, Wholesale Chicago and New York City, F.O.B. Leading Shipping Points, Annual Summaries. 1952-1983.

U.S. Department of Agriculture, SRS, CRB. Vegetables for Fresh Market, Acreage, Production and Value 1954-59. Statistical Bull. 300.

U.S. Department of Agriculture, SRS, CRB. Vegetables for Fresh Market, Acreage, Production and Value 1959-65. Statistical Bull.412.

U.S. Department of Agriculture, SRS, CRB. Vegetables for Fresh Market, Acreage, Production and Value 1964-70. Statistical Bull. 495.

U.S. Department of Agriculture, SRS, CRB. Vegetables for Fresh Market, Acreage, Yield, Production and Value. Annual Summaries. 1971-73.

U.S. Department of Agriculture, ESS, CRB. Vegetable Estimates by Seasonal Groups and States 1974-78, Acreage, Yield, Production, and Value. Statistical Bull. 665.

U.S. Department of Agriculture, SRS, CRB. Vegetable Acreage, Yield, Production, and Value, 1981 Annual Summary.

U.S. Department of Agriculture. U.S. Foreign Agricultural Trade. Annual Statistical Reports. 1952-59.

U.S. Department of Agriculture, ERS. U.S. Fresh Market Vegetables Statistics, 1949-80. Statistical Bull. 688.

U.S. Department of Agriculture, SRS, CRB. Agricultural Prices, Annual Summaries. 1952-81.

U.S. Department of Agriculture, ERS, SRS, AMS. Livestock and Meat Statistics. Statistical Bull. 522 and Supplement for 1980. 1952-81.

U.S. Department of Agriculture, AMS, CRB. Field Crops. Statistical Bull. 185. 1952-53.

U.S. Department of Agriculture, AMS, CRB. Crop Production, Annual Summaries. 1954-81.

U.S. Department of Agriculture, AMS, CRB. Field and Seed Crops by States, 1949-54, Farm Disposition, Season Average Price, Value of Production, Value of Sales, Revised Estimates. Statistical Bull. 208.

U.S. Department of Agriculture, SRS, CRB. Field and Seed Crops, Production, Farm Use, Sales, Value, Revised Estimates. 1954-59. Statistical Bull. 311.

U.S. Department of Agriculture, SRS, CRB. Field and Seed Crops, Production, Farm Use, Sales, Value, Revised Estimates, 1959-64. Statistical Bull. 404.

U.S. Department of Agriculture, SRS, CRB. Crop Values, Annual Summaries. 1964-81.

U.S. Department of Agriculture, AMS, CRB. Vegetables for Processing, Acreage, Production, Value by States 1949-55, Revised Estimates. Statistical Bull. 210.

U.S. Department of Agriculture, SRS, CRB. Vegetables for Processing, Acreage, Production, Value by State 1954-59, Revised Estimates. Statistical Bull. 299.

U.S. Department of Agriculture, SRS, CRB. Vegetables for Processing, Acreage, Production, Value by States 1959-65, Revised Estimates. Statistical Bull. 411.

U.S. Department of Agriculture, SRS, CRB. Vegetables - Processing, Acreage, Production and Value of Principal Commercial Crops by States, Annual Summaries. 1966-76.

U.S. Department of Agriculture, ESCS, ESS, SRS, CRB. Vegetables, Acreage, Yield, Production, and Value, Annual Summaries. 1977-81.

U.S. Department of Agriculture, AMS, CRB. Soybeans, Cowpeas, and Velvet Beans by States, 1924-53, Acreage, Yield, Production, Price, Value.

U.S. Department of Agriculture, SRS, CRB. Field Crops by States 1954-59, Acreage, Yield, Production. Statistical Bull. 290.

U.S. Department of Agriculture, SRS, CRB. Field Crops by States, 1960-64, Acreage, Yield, Production. Statistical Bull. 384.

U.S. Department of Agriculture, SRS, CRB. Crop Production, Acreage, Yield, Production, Annual Summaries. 1965-81.

U.S. Department of Agriculture, AMS, CRB. Field Crops. Statistical Bull. 185, 1954.

U.S. Department of Agriculture, AMS, CRB. Crop Production, Annual Summaries. 1954-81.

U.S. Department of Commerce. Survey of Current Business, monthly issues. 1952-83.

U.S. Department of Commerce, Weather Bureau. Climatological Data, State Data. 1948, 49, and 81-83.

U.S. Department of Commerce, Weather Bureau. Climatological Data, National Summaries. 1950-83.

Wall, G. B. and D. S. Tilley. "Production Responses and Price Determination in the Florida Watermelon Industry." Southern Journal of Agricultural Economics. 11(1979):153-56.

Wall Street Journal. Commodities, various issues in February. 1952-83.

Zellner, A. "Statistical Analysis of Econometric Models." Journal of the American Statistical Association. 74(1979):628-643.