Value of Weather Information in Cranberry Marketing Decisions

Bernard J. Morzuch and Cleve E. Willis
University of Massachusetts, Amherst, MA 01003

(Manuscript received 4 May 1981, in final form 3 February 1982)

ABSTRACT

Econometric techniques are used to establish a functional relationship between cranberry yields and important precipitation, temperature, and sunshine variables. Crop forecasts are derived from the model and are used to establish posterior probabilities to be used in a Bayesian decision context pertaining to leasing space for the storage of the berries.

1. Introduction

The purpose of this paper is to evaluate the usefulness of a cranberry yield model that has been developed for Massachusetts. The intent is to illustrate the manner in which this model can be used by a large marketer of cranberries to make better informed decisions pertaining to the firm's need for bin storage capacity as the berries are harvested. The plan of the paper is to present a brief review of the manner in which investigators have modeled the relationship between crop yields, economic inputs and weather, to present the empirical model itself, to illustrate the decision to be made and the corresponding Bayesian strategy to be implemented, and finally to evaluate this decision in relation to a decision that would have been made had information about the future been perfect.

2. Historical perspective

It is common for agricultural economists to express the physical relationship between output of an agricultural commodity and the economic inputs used in its production by way of a production function. The literature provides a fairly complete catalog of functional forms suitable for presenting this relationship (Heady and Dillon, 1972; Fuss and McFadden, 1978).

In addition to crops responding to alternative applications of economic inputs such as land, labor and capital, crop yields are sensitive to noncost inputs such as solar radiation, precipitation, and temperature. The relationship between crop output (yield) and economic and noncost inputs is presented here using a power function. This form is used because it has a long history of application in agricultural economics research. Letting Y denote output, N labor, L land, K capital, and W some appropriate weather measurement, we can write

$$Y = AL^{\alpha}K^{\beta}N^{\gamma}W^{\omega}, \tag{1}$$

where A is a constant or coefficient of technology and α, β, γ, and ω are the elasticities of output with respect to the corresponding inputs above.

Empirical studies by agronomists and agricultural economists have been concerned with representations of either the economic or noncost weather inputs, the amount of detail depending upon the inclination of the researcher, and the type of information available. For example, Thompson (1970) chose to account for economic inputs within a soybean yield model by compressing the effects of these inputs into a technology trend variable. The parabolic effect that various precipitation and temperature variables can have on yield was accounted for by considering each variable's departure from normal in addition to the square of its departure from normal. Oury (1965) likewise accounted for the current state of the art with respect to production by using a trend variable when analyzing corn yield functions. Rather than incorporating explicit measures of weather variables in his response functions, he used an aridity index approach to account for weather. Unlike the above two studies, which were highly aggregative in nature, Morzuch (1977) used specific economic input variables associated with 44 farms over a seven-year period in conjunction with precipitation and temperature measures to arrive at soybean and corn production functions for Missouri farms in the spirit of (1).

Once available data sources have been identified, the variables used to measure both meteorological phenomena and the physical production process have been constructed, appropriate statistical hypotheses have been established pertaining to the effect of the included variables of the model, and finally competing functional forms have been tested against each other, the result may be a crop yield model with

0021-8952/82/040499-06$05.50

explanatory capabilities and/or one which may have value in a predictive sense. Indeed, one way to evaluate the usefulness of the resulting model is to test whether or not it can be used by a decision maker to arrive at more enlightened decisions than if the model were not available. For example, if a private firm were able to take the yield model suggested above and use it to increase its expected profit on the basis of the predictive ability of the model, then the model has value. Halter and Dean (1971) provide several worthwhile expositions of the manner in which decision theory is used in conjunction with positively estimated models within agriculture and climatology.

3. Weather/yield model

Cranberries are harvested in October and November in Massachusetts. They grow on bushes in bogs and are harvested by dry picking methods or by flooding the bogs. They are extremely sensitive to frost damage, since the bogs are located in low-lying areas. One of the technologies of paramount significance has been the utilization of sprinkler systems by the growers to combat frost. Adoption of sprinkler systems by the growers took hold around 1962 (Cross *et al.*, 1976). Furthermore, the effects of sunshine, temperature and precipitation on cranberry production are well documented (Franklin and Stevens, 1946; Franklin and Cross, 1948; Cross *et al.*, 1976).

On the basis of this information, an aggregate cranberry yield model expressed as a function of various weather measurements and a dummy variable suggesting the adoption of sprinkler systems in 1962 was developed for the period 1932–79. A variant of the power function was selected (Morzuch *et al.*, 1982). It takes the form:

$$Y = \exp(\alpha D + \delta_1 S_{AP} + \delta_2 S_A + \delta_3 S_S) \times P_M^{\beta_1} P_A^{\beta_2} P_S^{\beta_3} T_D^{\gamma_1} T_F^{\gamma_2} T_M^{\gamma_3} T_J^{\gamma_4} \mu, \quad (2)$$

where:

Y = yearly cranberry yield in barrels per acre for 1932–79, inclusive;
exp = the base of natural logarithms;
D = a dummy variable taking on a value of zero between 1932 and 1961 and a value of one between 1962 and 1979;
P_M = total precipitation in May of the present crop year;
P_A = total precipitation in August of the present crop year;
P_S = total precipitation in September of the present crop year;
T_D = mean December temperature in the previous crop year;
T_F = mean February temperature in the present crop year;
T_M = mean May temperature in the present crop year;
T_J = mean July temperature in the present crop year;
S_{AP} = total sunshine hours in April of the previous crop year;
S_A = total sunshine hours in August of the previous crop year;
S_S = total sunshine hours in September of the previous crop year;
μ = a random disturbance term assumed to have zero expected value and constant variance;

and α, δ_1, δ_2, δ_3, β_1, β_2, β_3, γ_1, γ_2, γ_3, and γ_4 are parameters to be estimated. Furthermore, precipitation measurements are in inches and temperature measurements in degrees Fahrenheit. The monthly measures of precipitation and temperature were obtained from the East Wareham weather station, and sunshine at Boston (Logan Airport).

Taking the natural logarithms of both sides of (2), the function becomes:

$$\ln Y = \alpha D + \delta_1 S_{AP} + \delta_2 S_A + \delta_3 S_S + \beta_1 \ln P_M + \beta_2 \ln P_A + \beta_3 \ln P_S + \gamma_1 \ln T_D + \gamma_2 \ln T_F + \gamma_3 \ln T_M + \gamma_4 \ln T_J + \ln \mu. \quad (3)$$

Ordinary least-squares techniques were used to estimate the model. Table 1 reports the values of the estimated coefficients. The high t-values associated with the majority of coefficients suggest a fairly accurate structural model. The Durbin-Watson statistic of 1.68 does not suggest the presence of correlation among the disturbance terms. Finally, the R^2 value implies that 85% of the variation in yield is explained by this model.

4. Using the model in a decision theory framework

Given the model presented above, an important issue revolves around the potential economic value it may have to decision makers within a cranberry marketing firm. In the short run, it can be used to update strategies pertaining to berry shipments to alternative supply points, changes in advertising expenditures, and bin leasing commitments for the storing and freezing of berries. Given predictions of yield well in advance of the harvest period, these strategies can be amended depending upon the projections themselves. Changing any of these strategies from year to year has an impact on the average profitability of the firm.

To illustrate, consider those items that determine the net profit of the firm. On the input side, intensity of usage of certain inputs is extremely sensitive to the amount of berries delivered by growers to the firm, with the amount of berries delivered being a function of weather. For example, if the amount of berries delivered turns out to be a bumper crop, at some previous point in time plans should have been made to step up advertising, to assure that enough space is available to store the berries, and to consider the increased transportation needs in shipping this crop to its subsidiaries. The plant likewise incurs other costs that are totally invariant to the quantity of berries delivered. Examples of such costs are some taxes and depreciation on plant and equipment.

For present purposes, consider that bin leasing for storage purposes is the only decision that needs to be made before the harvest is actually in and is the only item on the input side that affects net income. Although this simplistic formulation of net income is unrealistic, it will facilitate the presentation of the Bayesian decision strategy. The resulting "profit" function can be expressed as

$$\pi = P_{\theta}(\theta) - P_L(L), \tag{4}$$

where:

θ = barrels of berries delivered to the firm;
P_θ = average price received per barrel;
L = bin space leased by the firm (converted to barrels);
P_L = bin space cost per barrel;
π = net profit.

It is perennially the case that, given a planning strategy, too many or too few bins will be leased. Erring in either direction represents a penalty to the firm. Thus, two additional cost components need to be added to (4) to account for the lack of certainty involved in leasing the correct number of bins. Letting:

E = excess space leased in barrels; occurs when space leased is greater than quantity delivered, i.e., $L > \theta$;
d = deficient space; occurs when quantity delivered is greater than space leased, i.e., $\theta > L$;
P_E = cost associated with leasing excess space; merely the bin space lease cost per barrel, i.e., $P_E \equiv P_L$;
P_d = cost associated with not leasing enough space; finding additional space to lease, transporting to more distant areas, and loss in berry quality; $P_d > P_L$;

(4) can be rewritten as:

$$\pi = P_{\theta}(\theta) - P_L(L) - P_L(E) - P_d(d). \tag{5}$$

Within a decision theory context, berries delivered (θ) represent the state of nature and bin space leased (L) the course of action. Assuming there to be m states of nature and n courses of action, a typical state is represented as θ_i, where $i = 1, 2, \ldots, m$, and a typical action as L_j, where $j = 1, 2, \ldots, n$. For each state of nature (θ_i) and resulting action (L_j), there is a corresponding consequence (π_{ij}). With m states of nature and n actions, there are $m \times n$ possible consequences. Also, each potential state of nature has a corresponding probability of occurrence $P(\theta_i)$. The expected value of each particular action can, in turn, be calculated by weighting the consequences of the action given the alternative states of nature by the respective prior probabilities of these states; i.e.,

$$E(L_j) = \sum_{i=1}^{m} \pi_{ij} \cdot P(\theta_i). \tag{6}$$

All of the possible actions, states of nature, consequences, prior probabilities, and corresponding expected value of each action can be conveniently arranged in the format of Table 2.

For the problem at hand, typical states of nature include high, medium and low yields, with shades in between. Typical actions include leasing high, me-

TABLE 1. Regression results of Massachusetts cranberry yield model.

Variable	Estimated coefficients*	Variable	Estimated coefficients
D	0.84 (12.58)	T_F	0.54 (2.29)
P_M	0.24 (4.47)	T_M	2.52 (3.28)
P_A	0.10 (2.23)	T_J	−2.74 (−3.51)
P_S	−0.09 (−2.28)	S_{AP}	0.002 (1.56)
T_D	.44 (1.65)	S_A	0.002 (2.13)
		S_S	0.002 (2.20)

* t-values appear in parentheses beneath estimated coefficients; Durbin Watson = 1.68; R^2 = 0.85.

TABLE 2. General representation of states of nature, actions, consequences and prior probabilities.

θ \ L	L_1	L_2	$\cdots$	L_j	$\cdots$	L_n	$P(\theta)$
θ_1	π_{11}	π_{12}	$\cdots$	π_{1j}	$\cdots$	π_{1n}	$P(\theta_1)$
θ_2	π_{21}	π_{22}	$\cdots$	π_{2j}	$\cdots$	π_{2n}	$P(\theta_2)$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
θ_i	π_{i1}	π_{i2}	$\cdots$	π_{ij}	$\cdots$	π_{in}	$P(\theta_i)$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
θ_m	π_{m1}	π_{m2}	$\cdot$	π_{mj}		π_{mn}	$P(\theta_m)$
$E(L_j)$	$E(L_1)$	$E(L_2)$	$\cdots$	$E(L_j)$	$\cdots$	$E(L_n)$	

TABLE 3. Consequences of bin leasing strategies.

States of nature	Actions: Yields (in barrels)	Bin leasing strategies (in barrel equivalents): L_1 572 000	L_2 684 000	L_3 796 000	L_4 908 000	L_5 1 020 000	Prior probabilities
θ_1	572 000	14 300 000	13 180 000	12 060 000	10 940 000	9 820 000	0.03
θ_2	684 000	16 764 000	17 100 000	15 980 000	14 860 000	13 740 000	0.07
θ_3	796 000	19 228 000	19 564 000	19 900 000	18 780 000	17 660 000	0.15
θ_4	908 000	21 692 000	22 028 000	22 364 000	22 700 000	21 580 000	0.45
θ_5	1 020 000	24 156 000	24 492 000	24 828 000	25 164 000	25 500 000	0.30
	Expected value of each strategy*	21 494 880	21 787 200	21 977 600	21 949 600	21 266 400	

* The expected value of each strategy is calculated by weighting each consequence element in that column by its associated prior probability and summing the five values.

dium and low numbers of bins, again with intermediate possibilities. Values associated with these categorical descriptions can be obtained by asking the decision maker and/or from historical records. The prior probabilities are obtainable similarly.

Consultation with marketing experts revealed five discrete representative values for θ and five corresponding values for L. Likewise, the experts were able to suggest the prior probability distribution. Furthermore, given the prices of P_θ, P_L, and P_d of \$30, 5 and 8 per barrel, respectively, (5) can be used to derive 25 alternative values of π_{ij} to correspond with the five possible leasing strategies (L_j) given the occurrence of one of the possible yield states of nature (θ_i). Finally, the expected value of each action can be calculated by taking the column associated with each L_j, weighting each of its elements by the corresponding prior probabilities, and summing these items. This information is presented in Table 3. On the basis of using only that information supplied by the experts, the bin leasing strategy that maximizes expected net revenue is L_3, with an expected net revenue of \$21 977 600.

It is useful at this point to compare the result which leads to maximum expected value using only this limited information with that expected value which would result if it were known for certain which state of nature would occur. This difference can be expressed as

$$\mathrm{EVPI} = \sum_{i=1}^{m} \max_{\pi_{ij}} j \cdot P(\theta_i) - \max_{E(L_j)} j, \qquad (7)$$

where EVPI is the expected value of perfect information and can be considered as providing an upper bound on the expected value of additional information that might be obtained using econometric forecasts based on weather and other observable information.

With respect to the problem at hand, if we were certain θ_1 were to occur, we would take action L_1, if θ_2 then action L_2, if θ_3 then L_3, if θ_4 then L_4 and if θ_5 then L_5. By weighting the corresponding consequences by the probabilities for the respective states of nature, summing these items, and then subtracting the expected value of that strategy that maximizes net revenue from Table 3, we obtain \$498 400. If we could forecast yield perfectly, and hence avoid all costs of leasing too few or too many bins, we could save on the average just under one-half million dollars.

Of course we never have weather and other information of sufficient quality to enable perfect forecasts. If we forecast yields on the basis of the current weather and other variables indicated in (2), using the regression coefficients provided in Table 1, we can improve our forecast of yields. The expected value of such forecast information depends upon the precision of the estimation of (2) as well as the values of the explanatory variables for the equation. In general, the expected values of forecasts based upon models like (2) are bounded by zero and one-half million dollars (the expected value of information if (2) were capable of perfect forecasts).

To illustrate, we have constructed confidence intervals for our predictions, given each of the five states of nature, using the mean values of the weather and other explanatory variables over the last five years in conjunction with the vector of coefficients from the regression model presented in Table 1. These conditional probabilities $P(z_k|\theta_i)$ are presented in Table 4. Technically, since final decisions on leas-

TABLE 4. Conditional probabilities $P(z_k/\theta_i)$.

	z_1	z_2	z_3	z_4	z_5
θ_1	0.84	0.1375	0.0210	0.0015	0
θ_2	0.16	0.68	0.1375	0.0210	0.0015
θ_3	0.0225	0.1375	0.68	0.1375	0.0225
θ_4	0.0015	0.0210	0.1375	0.68	0.16
θ_5	0	0.0015	0.0210	0.1375	0.84

TABLE 5. Posterior probabilities $P(\theta_i/z_k)$.

	z_1	z_2	z_3	z_4	z_5
θ_1	0.6229	0.0501	0.0034	0.0001	0
θ_2	0.2768	0.5787	0.0533	0.0039	0.0003
θ_3	0.0834	0.2507	0.5653	0.0558	0.0103
θ_4	0.0166	0.1148	0.3429	0.8283	0.2198
θ_5	0	0.0054	0.0349	0.1116	0.7694

ing are made in September, prior to the main harvest activity in mid-November, the September weather information is unavailable for the forecast. The model which gave rise to Table 4 was run without the September variables with virtually no change in forecast or forecast variance. In the interest of brevity, we include only these conditionals; the complete set of results for alternative models, as well as the formulas for calculating these conditionals are available upon request from the authors.

Next, if we multiply each of the conditionals by the appropriate $P(\theta_i)$, the set of joint probabilities $P(z_k \text{ and } \theta_i)$ is obtained. The updated or posterior probabilities are determined by using the above information and applying Bayes' rule

$$P(\theta_i|z_k) = \frac{P(z_k \text{ and } \theta_i)}{\sum_i P(\theta_i)P(z_k|\theta_i)}. \qquad (8)$$

These are presented in Table 5. To obtain the value of this additional information, we first calculate the expected value of each action, updated by $P(\theta_i|z_k)$ in place of $P(\theta_i)$, for each z_k. This can be expressed as

$$E(\pi)L_j|z_k = \sum_{i=1}^{m} \pi_{ij} \cdot P(\theta_i|z_k). \qquad (9)$$

For the problem at hand, this amounts to multiplying each column of the consequence matrix in Table 2 by each column of posterior probabilities in Table 5. The result is 25 new values which are presented in Table 6. These figures are the expected values of the consequences of each action given the jth value of the yield predictor as given by the regression model. For this application, the best actions given the predictions are found on the main diagonal—e.g., if z_1 occurs, L_1 has the highest expected net revenue in the first row, if z_2 obtains, L_2 is best, etc. Table 6 also supplies the probabilities of each z_k. These are calculated as $P(z_k) = \sum_i P(\theta_i)P(z_k|\theta_i)$, i.e., the denominator of (8).

To obtain the expected value of information supplied by the posterior distribution, the largest value of each row (the fact that successive rows contain larger values than their previous counterparts is simply a reflection of the higher expected yields for z_5 than z_4, z_4 than z_3, etc.) in Table 6 is multiplied by the probability of its occurrence $P(z_k)$. These products are then summed to arrive at the expected value of the Bayesian decision (EVBD). In this case, EVBD = \$22 304 019. The value of the additional information is the expected value resulting from using the updated information minus the maximum expected value associated with using only the primary information. Thus, the expected value of information (EVI) is:

$$\$22\,304\,019 - \$21\,977\,600 = \$326\,419.$$

5. Conclusions

We began with the observation that weather and other inputs influence the annual yields of cranberries. This relationship was estimated econometrically and these results can be used along with data on weather and other factors to provide reasonable forecasts of yields.

The second issue raised was that decisions which are conditional on yield forecasts and which affect profitability must be made. An example is bin leasing. Costs of wrong decisions (and forecasts) include the costs associated with leasing too few or too many bins. If the decision makers used only historical record [$P(\theta_i)$] to form their forecasts, the expected costs of wrong decisions are estimated to be \$498 400 annually, relative to a theoretically perfect yield forecast.

If the decision makers were to use both historical record [$P(\theta_i)$] and the forecasts supplied by the regression model (Table 1), given values of weather and other variables, in a Bayesian way, the reduced cost of wrong decisions is estimated to be \$326 419 per year on the average. This represents the expected value of the forecasts using the estimated yield model.

Finally, it is interesting to compare the expected value of our forecasts with that of theoretically perfect information. Our estimation of (2) provides

TABLE 6. Expected consequences given posterior probabilities.

	L_1	L_2	L_3	L_4	L_5	$P(z_k)$
$E(\pi)\ L_j|z_1$	15 511 447	14 940 405	13 966 340	12 870 846	11 751 182	0.04045
$E(\pi)\ L_j|z_2$	17 858 901	19 312 165	17 542 220	16 597 568	15 485 766	0.08225
$E(\pi)\ L_j|z_3$	20 092 960	20 423 943	20 677 321	20 107 622	19 038 660	0.18043
$E(\pi)\ L_j|z_4$	21 803 025	22 138 778	22 468 854	22 717 684	21 760 510	0.36939
$E(\pi)\ L_j|z_5$	23 556 605	23 892 537	24 228 034	24 548 533	24 549 004	0.32448

forecasts worth just under two-thirds (0.655) of the maximum theoretically attainable value of information. Given additional resources, a more accurate model than (2) could presumably be developed. One could calculate the costs of reducing the standard error of the forecasts and compare these with the marginal gains of reducing the expected costs of wrong leasing decisions in deciding whether additional work on the forecasting model were economic. Clearly, the upper limit to the value of improving the forecasting equation (2) is the difference between the expected reduced costs of wrong decisions if the forecasts from (2) were perfect (\$498 000) and the expected reductions in costs of wrong decisions if the current model (2) were used (\$326 419), or \$171 981 annually.

Acknowledgment. The authors wish to thank Robert Lautzenheiser, State Climatologist, New England Climatic Service for his helpful assistance in obtaining the weather data.

REFERENCES

Cross, C. E., J. S. Norton, I. E. Demoranville, W. E. Tomlinson, R. M. Devlin, K. H. Devbert, R. B. Gray and B. M. Zukerman, 1976: *The modern art of cranberry cultivation.* Cooperative Extension Service Publ., No. 112, University of Massachusetts, 84 pp.

Franklin, H. J., and C. E. Cross, 1948: Weather in Relation to Cranberry Production and Condition. Mass. Agric. Exp. Sta. Bull. 450, University of Massachusetts, 15 pp.

——, and N. E. Stevens, 1946: Weather and water as factors in cranberry production. Mass. Agric. Exp. Sta. Bull. 433, University of Massachusetts, 51 pp.

Fuss, M., and D. McFadden, 1978: *Production Economics: A Dual Approach to Theory and Applications. Vol. 1.* North-Holland, 217–268.

Halter, A. N., and G. W. Dean, 1971: *Decisions Under Uncertainty.* South-Western Publishing, 266 pp.

Heady, E. O., and J. L. Dillon, 1972: *Agricultural Production Functions.* Iowa State University Press, 667 pp.

Morzuch, B. J., 1977: Technology and climatic effects in aggregate production functions. Ph.D. dissertation, Dep. Agric. Econ., University of Missouri, 217 pp.

——, J. Kneip and D. C. Smith, 1982: An econometric approach to modeling the effects of weather and technology on cranberry yields. Mass. Agric. Exp. Sta. Bull., University of Massachusetts (forthcoming).

Oury, B., 1965: Allowing for weather in crop production model building. *J. Farm Econ.*, **47**, 250–282.

Thompson, L. M., 1970: Weather and technology in the production of soybeans in the central United States. *Agron. J.*, **62,** 232–236.