MANAGING RISK FOR THE GRAPE HARVEST AT WELCH'S, INC.

STUART J. ALLEN
Penn State-Erie, Station Road, Erie, PA 16563

EDMUND W. SCHUSTER
Center for Transportation Studies, Massachusetts Institute of Technology, 77 Massachusetts Avenue, 1-235, Cambridge, MA 02139-4307

Welch's, Inc., is the world's largest processor of Concord and Niagara grapes with annual sales approaching $600 million. It is the production, distribution, and marketing arm of the National Grape Cooperative Association (NGCA), headquartered in Westfield, New York. The membership of NGCA includes 1,400 growers who cultivate 41,000 acres of vineyards in western New York, northern Ohio and Pennsylvania, western Michigan, and south-central Washington.

Welch's operates raw grape processing plants near the growing areas. These plants provide the juice for consumer products: bottled and frozen juices, concentrated juices, jellies, and jams. Settled, pasteurized juice from the annual harvest of some 300,000 tons of grapes is stored in refrigerated tank farms as "single strength" juice or in various degrees of concentration. Concentrated juice conserves storage capacity and lowers interplant shipping costs incurred for purposes of juice blending.

The grape harvest varies in size for each growing region each year, depending on spring and summer weather conditions. The exact size is not known until the end of the harvest process. Weather and agricultural practice directly affect the rate at which grapes reach the minimum percentage of water-soluble solids (%WSS) consistent with Welch's strict quality standards. The *start date* of the harvest process is the date that grapes within a growing region reach the proper %WSS.

The *frost-date* for a growing region, the end of the harvest process, occurs within five to seven days of the first 28°F day for that region. The cold causes tissue damage to the stems, and grapes fall to the ground when disturbed by the mechanized pickers. The length of the *harvest season* is then determined by the time interval between two randomly occurring events: the start date and the frost date. Temperature data gathered from state experiment stations are a fair representation of temperatures over the growing region.

IMPORTANCE OF HARVEST PLANNING

The harvest season is obviously a critical period each year for Welch's, and considerable resources are devoted to planning for that event. The planners must contend with several difficult issues. Each of the growers under contract to Welch's would like to limit any loss due to cold weather, but that expectation must be tempered by economic realities. Both harvest size and length of the harvest season are random variables. A policy that required a 100% harvest every year regardless of harvest size and time available would imply a prohibitive capital investment in equipment for harvesting and processing grapes.

Clearly it is imperative to have in place a rational means of balancing the growers' desires against the capital expenditures required. A well-designed planning tool could also be used to address long-term planning issues: the impact of proposed new harvest scheduling methods, expansion or contraction of the grower base, new plantings within a region, and changes in yield due to new agricultural practice.

Historically, those issues were addressed by estimating a fixed length of season to determine the harvest rate and capital investment requirements. That approach ignored the risk associated with the variability in the length of the harvest season and size of the harvest in each growing region.

Welch's has a history of applying modeling techniques to solve management problems. In 1994 we implemented the juice logistics model (JLM) to plan the proper timing and amount of raw material transfers, and the proper recipes for each manufacturing plant (see Schuster and Allen [4]). However, the JLM deals with grape juice only after it is processed. We need to model the harvest of grapes and determine the best rate to process fresh grapes into juice. The variability of frost dates for a growing area greatly influences the desired processing rate for grapes. Hence, calculating the correct capital investment plan, and

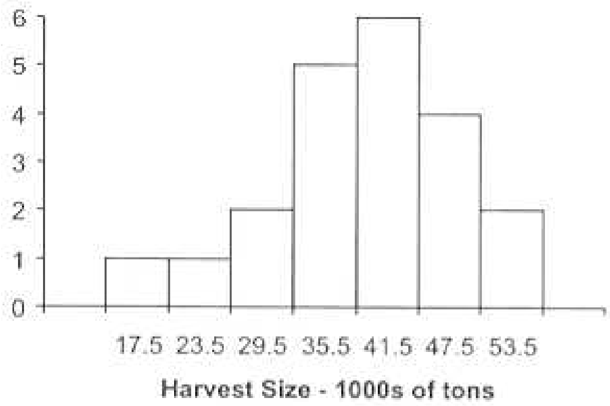

FIGURE 1: Histogram of harvest size

processing rate, for harvesting operations becomes an important goal for Welch's.

The operations management literature lists few applications of modeling to fruit harvesting operations in which a quantifiable trade-off exists between capital investment and the risk of crop loss. Porteus [2, 3], one of the few references we found, developed a case study of the National Cranberry Cooperative. In the case, he examines a complex study of capital investment and capacity for the processing of cranberries. The National Cranberry Cooperative had two main problems. First, trucks and drivers spent too much time waiting to unload cranberries during harvest. Second, overtime costs and absenteeism were out of control. But although the processing of cranberries shares common elements with grape processing, Porteus did not address the risk of crop loss from a frost. The case mentioned no formal model for evaluating the trade-off between capital investment and the risk of crop loss.

In what follows, we will describe a new method recently instituted at Welch's that recognizes harvest size and length of the harvest season as unique random

TABLE 1: Current Statistical Properties for Three Selected Regions

Growing Region	*Risk-Free Harvest Rate, Ro*	*Coefficient of Variation for Length of Harvest, k(L)*	*Coefficient of Variation for Harvest Size, k(H)*
Northeast	2326	0.35	0.15
Lawton	1667	0.45	0.13
Kennewick	2222	0.41	0.26

variables within each growing region. This new planning tool permits analysis of the capital expenditure consequences for any specified policy of the probability of harvesting the entire grape crop for a given region. It has as its basis, historical data on the size and duration of the harvest. We examine these data next.

HARVEST CHARACTERISTICS

For many years the NGCA has maintained an extensive database on all aspects of the grape harvest. This database provides an indispensable foundation for our efforts to rationalize risk management in harvest planning.

Harvest Size

Potential yields per acre are tracked throughout the growing season using random sampling of cluster counts, average berry count (per cluster), and vine weight. A multiple regression model using these quantitative variables and growing region as a qualitative variable provides midsummer yield estimates with a coefficient of determination of 86%. Nevertheless, the total harvest remains a random variable until the harvest is completed. Historical data then provide the characteristics of the total harvest size of each region.

Figure 1 illustrates the distribution of harvest size, H, for a typical region. The distribution is mound-shaped and justifies our assumption that the population distribution is approximately normal. The mean and standard deviation of this distribution are defined by $\mu(H)$ and $s\ (H)$ for each region and are estimated from historical data. Current estimates of harvest size parameters are given in table 1 for three of the growing regions. These data are given in non-dimensional ratio format for proprietary reasons. We note in passing that it may be possible to sharpen these estimates using the multiple regression predictions just before harvest. This refinement has not been implemented as yet.

Length of Harvest Season

The available harvest interval is not recorded as primary data, but start date and frost date are. The length of harvest season, L, is determined from these dates; a typical histogram for L is illustrated in figure 2. Again we see that the distribution is mound-shaped and we assume that the population is normally distributed with parameters $\mu(L)$ and $s\ (L)$. Estimates of these coefficients of variation are provided in table 1.

As a final note, tests indicate that the length of harvest season, L, is not correlated with harvest yield and hence harvest size, H. For example, one of the grow-

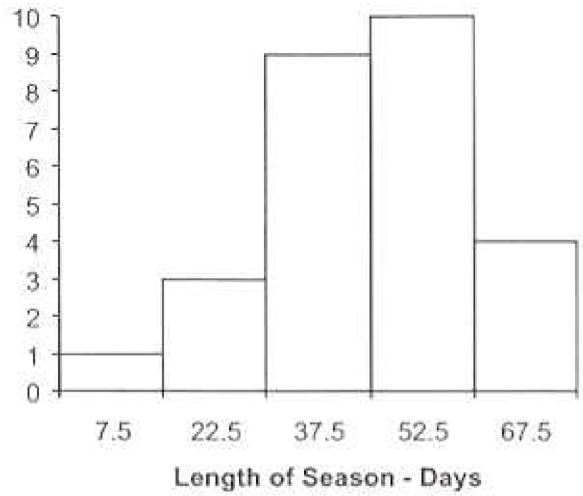

FIGURE 2: Histogram for length of harvest season

ing regions has a correlation coefficient of 0.14 between yield and length of harvest season based on 22 years of data. We cannot reject the hypothesis that the population correlation coefficient is zero, with an observed significance of 53%. That has important consequences for model development.

RISK MODEL FOR HARVEST PLANNING

In this section we summarize the development of an improved model for harvest planning. The model provides the means to balance the costs of capital investment in process equipment against the risks of crop loss due to frost. The model determines a mean harvest rate as a function of the length of the harvest season and the probability that 100% of the grape crop will be harvested. The harvest rate requirement, in turn, determines the equipment requirements for harvesting and processing grapes into juice. The processing rate determines equipment needs for pressing grapes, juice storage, and juice concentration (to create storage).

To avoid notational clutter, we do not distinguish the growing areas, but remind the reader that the resulting model must be applied separately to each area.

We will need to define some additional notation:

R = harvest rate in tons per day, a decision variable
T = time required to harvest the entire crop in days

The time required to harvest the entire crop is related to the harvest size and harvest rate by

$$T = H / R \qquad (1)$$

and we can define the slack time in days, S, as the difference between the length of harvest season, L, and time required to harvest 100% of the crop. Then,

$$S = L - T \qquad (2)$$

Slack time, S, is a random variable and is the foundation of the required harvest model. If we know its mean and standard deviation and the shape of its distribution, we can enforce a policy requirement that the entire crop be harvested with stated probability, p:

$$P(S \geq 0) = p \qquad (3)$$

where p might range, for example, from 85% to 98%. In subsequent work, we will refer to the value of p as the *percent recovery policy*.

The estimated expected value of slack can be expressed, using equations 1 and 2 as

$$\mu(S) = \mu(L) - \mu(H) / R \qquad (4)$$

To estimate the standard deviation of the slack time, we invoke the earlier finding of statistical independence between length of harvest season and harvest yield. Under statistical independence, the slack time variance is simply the sum of the variances of length of harvest season, L, and the time required to harvest, T (see, for example DeGroot [1, p. 159]).

Again using equations 1 and 2 we have:

$$\sigma^2(S) = \sigma^2(L) + \sigma^2(H) / R^2 \qquad (5)$$

Finally, we assume that the distribution of slack time, S, is approximately normal based on empirical evidence of figures 1 and 2. All pieces are now in place to use equations 3–5 to determine the required harvest rate, R (tons/day), for specified recovery policy, p, and known means $\mu(L)$ and $\mu(H)$ and variances $\sigma^2(L)$ and $\sigma^2(H)$. However, in their present form, equations 3–5 are not easy to work with.

We can obtain an explicit form for the required harvest rate, R, if we first define two coefficients of variation

$$k(L) = \sigma(L) / \mu(L), \quad k(H) = \sigma(H) / \mu(H) \qquad (6)$$

and denote the standard normal variate associated with an upper-tailed probability p by Z_p. Now define the *risk-free* harvest rate, R_0, as that harvest rate that would be used if there were no uncertainty in harvest size and length of harvest seasons. Then

$$R_0 = \mu(H) / \mu(L) \qquad (7)$$

Any uncertainty in H or L will cause the required harvest rate, R, to exceed R_0 so it is convenient to define the harvest rate inflation factor, r, as

$$r = R / R_0 \qquad (8)$$

TABLE 2: Harvest Rate Inflation Factors, r

		85% Harvest Recovery Policy ($Zp = -1.04$) Harvest Size Coefficient of Variation, K(H)				
		0.05	*0.10*	*0.15*	*0.20*	*0.25*
Length of	0.25	1.36	1.37	1.39	1.42	1.46
Harvest Season	0.30	1.46	1.47	1.49	1.52	1.55
Coefficient	0.35	1.58	1.59	1.60	1.63	1.66
of Variation,	0.40	1.72	1.73	1.74	1.76	1.79
K(L)	0.45	1.88	1.89	1.91	1.92	1.95
		90% Harvest Recovery Policy ($Zp = -1.28$) Harvest Size Coefficient of Variation, K(H)				
		0.05	*0.10*	*0.15*	*0.20*	*0.25*
Length of	0.25	1.48	1.50	1.52	1.56	1.61
Harvest Season	0.30	1.63	1.64	1.67	1.70	1.74
Coefficient	0.35	1.82	1.83	1.85	1.88	1.92
of Variation,	0.40	2.05	2.06	2.08	2.11	2.14
K(L)	0.45	2.36	2.37	2.39	2.41	2.44
		95% Harvest Recovery Policy ($Zp = -1.645$) Harvest Size Coefficient of Variation, K(H)				
		0.05	*0.10*	*0.15*	*0.20*	*0.25*
Length of	0.25	1.71	1.73	1.77	1.82	1.87
Harvest Season	0.30	1.98	2.00	2.03	2.08	2.13
Coefficient	0.35	2.36	2.38	2.41	2.45	2.49
of Variation,	0.40	2.93	2.94	2.97	3.00	3.05
K(L)	0.45	3.85	3.87	3.89	3.92	3.96
		98% Harvest Recovery Policy ($Zp = -2.05$) Harvest Size Coefficient of Variation, K(H)				
		0.05	*0.10*	*0.15*	*0.20*	*0.25*
Length of	0.25	2.06	2.09	2.14	2.20	2.27
Harvest Season	0.30	2.61	2.63	2.67	2.73	2.79
Coefficient	0.35	3.55	3.57	3.60	3.65	3.71
of Variation,	0.40	5.56	5.58	5.61	5.66	5.71
K(L)	0.45	12.91	12.93	12.95	12.99	13.04

This not only provides a more general result but also has an explicit physical interpretation: the inflation of the risk-free harvest rate caused by uncertainty.

After some algebra, carried out in the appendix, we obtain the required harvest rate ratio:

$$r = \{1 + [1 - (1 - Z^2_p k^2(L))\ (1 - Z^2_p k^2(H))]^{1/2}\} \ / \ (1 - Z^2_p k^2(L)) \qquad (9)$$

In equation 9, we must require that $Z^2_p k^2(L) < 1$ to ensure finite values of r. In addition, the term in square brackets must be non-negative. Note that as the two coefficients of variation become very small, the harvest rate inflation factor, r, approaches 1.0 so that the harvest rate, R, approaches the risk-free rate, R_0.

Example Application

From table 1, for the Northeast processing plant, we find that $k(L) = 0.35$, $k(H) = 0.15$. Using an 85% recovery policy with $Z_p = -1.04$, we obtain from equation 9, $r = 1.6045$. Then the physical harvest rate can be found from equation 8 using table 1: $R = r \times R_0 = 1.6045(2326)$ = 3732 tons per day. This in turn implies specific capital investment requirements for harvesting equipment, grape pressing capacity, and storage and concentration capacities.

Sensitivity Analysis of Harvest Rate Inflation Factor

The harvest rate inflation factor is influenced by three parameters: the recovery rate policy (as measured by Z_p) and the two coefficients of variation $k(L)$ and $k(H)$. It is important to know how *changes* in each of these parameters affect r. Ideally this could be done algebraically, but we have been unable to obtain simple, transparent expressions for purposes of sensitivity analysis.

We take the direct approach of evaluating equation 9 over ranges of each of the parameters of interest. For the recovery policy we have chosen four values: 85%, 90%, 95%, and 98%. We varied the coefficient of variation of the length of harvest season, $k(L)$, from 0.25 to 0.45 in steps of 0.05. The coefficient of variation of harvest size, $k(H)$, ranged from 0.05 to 0.25 in steps of 0.05.

The two-way DATA TABLE function in EXCEL was used to compute the dimensionless harvest rate, r, over the ranges of $k(L)$ and $k(H)$ exhibited in table 2 for each of the four recovery policies. These results demonstrate that the harvest rate requirements increase with increases in the two coefficients of variation and the percent recovery policy. What is perhaps more important is the extreme influence of increasing variability of the length of harvest season as measured by $k(L)$. This suggests that resources should be devoted to an attempt at more accurate predictions of dates of grape maturation and frost.

USING THE HARVEST MODEL FOR PLANNING

The harvest model was introduced into the planning process on a limited basis in 1996. After the model had been used as an adjunct to existing planning tools, it became clear that the model provided a much better representation of the risks associated with the harvest season. Confidence in the model grew with use, and it is now fully incorporated into all harvest related planning activities. In this section we summarize a few of the recent uses of the new risk model.

Proposed New Harvesting Strategy

Traditionally, two grape varieties, Niagara and Concord, were harvested sequentially in each region. A "dual-stream" procedure had been proposed; it would permit delay of the Niagara harvest start date but require simultaneous harvesting of both varieties. Delay of the Niagara harvest date would yield a higher value crop because of higher % WSS and higher re-

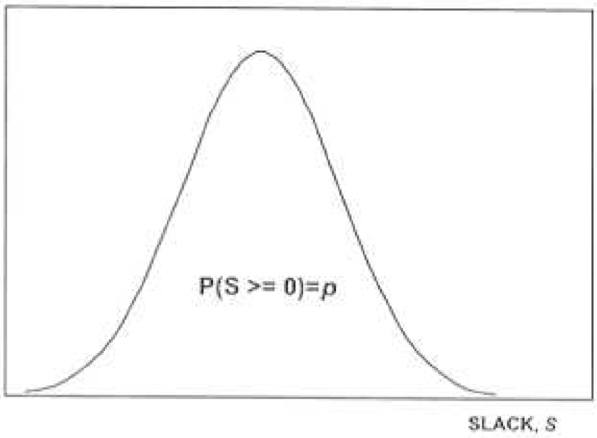

FIGURE 3: Normal distribution of *S*

covery. The dual-stream capability would accomplish that without extending the total season length. The price of dual-stream is an increased grape processing rate and subsequent capital equipment expenditures.

The new risk model was an indispensable aid in analyzing the cost/benefit trade-offs over a range of crop recovery policies.

Proposed New Plantings

Five-year estimates of consumer product sales led to a proposal that new plantings may be required in certain growing areas. These would lead in turn to projected increases in harvest rate requirements. The harvest risk model provided the foundation for assessing the costs of new plantings and additional processing equipment against the return on consumer product revenues.

Optimal Harvest Recovery Policy

Is there some "best" harvest recovery policy that balances the cost of capital expenditures against the cost of lost grapes? The development of the risk model for grape harvest provided the opportunity to answer that question.

Reexamination of table 2 reveals that a 98% harvest recovery will demand prohibitive capital outlays on processing equipment. At the same time, a recovery policy below 80% is ruled out as completely unacceptable to the growers. During this past year data were gathered on the harvest losses in dollars as a function of past recovery percentages. This permitted an incremental cost benefit analysis to be carried out that indicated the "best" percentage recovery policy.

CONCLUSION

The harvest risk model developed in this article has served Welch's in a variety of applications and has become an accepted part of the harvest planning process. We expect that this planning tool could easily be extended to other capital intensive tasks in which the start time, finish time, or both are random variables.

APPENDIX – DERIVATIONS

Consider the normally distributed random variable slack time, S, shown in figure 3. For a specified recovery policy, p, a corresponding value Z_p of the standard normal variate can be determined. For example, if p = 85%, then $Z_p = -1.04$. Points along line S are related to points along Z by

$$Z = (S - \mu(S)) \,/\, \sigma(S)$$

so we can solve for the value of S at Z_p as

$$S = \mu(S) + Z_p \cdot \sigma(S) \geq 0$$

where we have required that slack time, S, cannot be negative.

This last then required that

$$\mu(S) \geq -Z_p \cdot \sigma(S)$$

Using equation 4 in the expression above, we get

$$\mu(L) - \mu(H) \,/\, R \geq -Z_p \sigma(S)$$

Square both sides of the last expression and substitute for $\sigma^2(S)$ from equation 5 giving

$$[\mu(L) - \mu(H) \,/\, R]^2 \geq Z^2_p[\sigma^2(L) + \sigma^2(H) \,/\, R^2]$$

Gather the left side over R and the right side over R^2 and cancel the common R^2.

$$[R\mu(L) - \mu(H)]^2 \geq Z^2_p[R^2\, \sigma^2(L) + \sigma^2(H)]$$

Expand the left side and gather like powers of R:

$$(\mu^2(L) - Z^2_p\, \sigma^2(L))R^2 - 2\mu(L)\mu(H)R + (\mu^2(H) - Z^2_p\, \sigma^2(H)) \geq 0$$

Factor $\mu^2(L)$ from the first term and $\mu^2(H)$ from the last term and use definitions from equation 6:

$$\mu^2(L)(1 - Z^2_p k^2(L))R^2 - 2\mu^2(L)\mu(H)R + \mu^2(H)(1 - Z^2_p k^2(H)) \geq 0$$

Divide throughout by $\mu^2(H)$:

$$(1 - Z^2_p k^2(L))\left(\frac{\mu(L)R}{\mu(H)}\right)^2 - \frac{2\,\mu(L)}{\mu(H)} R + (1 - Z^2_p k^2(H)) \geq 0$$

Now from equations 7 and 8

$$r = R \,/\, R_0 = \frac{R\mu(L)}{\mu(H)}$$

so that

$$(1 - Z^2_p k^2(L))r^2 - 2r + (1 - Z^2_p k^2(H)) \geq 0$$

This is now of the form: $ar^2 + br + c$, with solution given by $r = \{-b \pm [b^2 - 4ac]^{1/2}\} \,/\, (2a)$, and equation 9 follows. We have discarded the negative sign in front of the term in square brackets in equation 9 since uncertainty should never act to decrease the harvest rate ratio, r.

ACKNOWLEDGMENT

The authors would like to thank two anonymous referees for their constructive suggestions.

REFERENCES

1. DeGroot, M.H. *Probability and Statistics*. Reading, Mass.: Addison-Wesley, 1975.
2. Porteus, E.L. "Case Analysis: Analyses of the National Cranberry Cooperative—1. Tactical Options." *Interfaces* 23, no. 4 (July–August 1993): 21–39.
3. Porteus, E.L. "Case Analysis of the National Cranberry Cooperative—2. Environmental Changes and Implementation." *Interfaces* 23, no. 6 (November–December 1993): 81–92.
4. Schuster, E.W., and S.J. Allen. "Raw Material Management at Welch's, Inc." *Interfaces* 28, no. 5 (September–October, 1998): 13–24.

About the Authors—

STUART J. ALLEN teaches management science and is a member of the Center for Process Manufacturing at Penn State-Erie. He works on the application of management science tools to real manufacturing problems.

EDMUND W. SCHUSTER is director, Affiliates Program in Logistics at the Massachusetts Institute of Technology in Cambridge. His interests are the application of models to logistical and planning problems experienced by the process industries.