Papers of the 1991 Annual Meeting

Western Agricultural Economics Association

Portland, Oregon
July 7 – 10, 1991

Donna J. Lee
University of Hawaii

Evaluating Risky Input Decisions in Crop Response Analysis

The relationship between input use and farm risk is an important issue in crop response analysis. Empirical risk analysis, however, is often hindered by a dearth of adequate data. This research presents a method for generating empirical probability distributions from small samples. Results for corn response to nitrogen fertilizer and irrigation water are presented.

Agricultural pollution externalities have become a prevalent issue in national and regional policy. In many areas, agricultural use of fertilizers, pesticides, and irrigation water, degrades the environment and creates social externalities. Agricultural pollution problems can often be alleviated through policy that regulates the use of farm inputs or provides farmers with incentives to reduce input use.

In agricultural production, variable inputs enhance crop yields and reduce the risk inherent to farming. Production models have been developed to predict farm response to policy proposals and to estimate policy costs. Models that capture the influence of input use on yield response but neglect the effect of inputs on yield variability, may fail to account for the additional risk incurred when input use is reduced. Thus, to adequately assess the value of inputs in production, one must consider both the contribution of inputs to expected yields and the effect of inputs on yield variability.

In stochastic production analysis, the problem that typically arises is the lack of sufficient data to generate an acceptable empirical probability distribution. The properties of the traditional estimators simply do not hold for very small samples. This paper describes a method for generating probability distributions from small samples for empirical risk analysis. The method, applied to western corn yield data, was used to determine the influence of fertilizer and irrigation water on productivity, and to assess the riskiness of reducing input use.

Crop yield distribution

For two primary reasons, the distribution of annual crop yields tends to be skewed. One reason why distributions are skewed is because crop yields are lower bounded by zero. The other, less obvious reason, is that very high yields can only be achieved when growing conditions are excellent for the entire season. Low yields, in contrast, can result from a spite of bad weather at any time during the season (Day). Thus, crop yields tend to be positively skewed, and lower than average yields occur more frequently than above average yields. Day (1972) also pointed out that yield estimates from a model that approximates a positively skewed distribution with a central or uniform distribution will overpredict yields more often than it will underpredict yields.

We can estimate an empirical probability distribution by weighting the observed values according to their frequency of occurrence. The basic approach assigns equal weights to each individual observation. For a sample with N observations, the assigned weights will be 1/N or 1/(N+1) (Shlaifer's rule; Shlaifer, 1959). If the frequency values are plotted onto a probability histogram, the shape of the distribution will become apparent. When the sample size is very small or when the data are very precise, the histogram will reveal a distribution that is essentially uniform.

Anderson (1973) acknowledged the need for an estimation procedure that could be applied to cases with very little data. He developed a six-step technique for generating an empirical probability distribution that involved: Shlaifer's rule for assigning frequencies, hand smoothing of the cumulative distribution function, and estimation of fractiles. Using a Monte Carlo approach, the technique was found to produce fairly consistent results (Anderson, 1974). The subjective nature of the approach (due to the hand smoothing and the selection of fractiles) diminished as sample size increased. Also, the quality of the results improved with sample size. Anderson (1974) noted that if numerous alternatives need to be analyzed, the laborious nature of the estimation procedure could present a drawback.

This study proposes an alternative method for weighting observations. The method, which generates an empirical probability distribution, does not entail hand smoothing, can quickly be executed with a simple computer algorithm, and can easily accommodate prior information.

Grouping Method

For a set of N observations $z_j \in Z$, an empirical probability distribution can be estimated in the following manner. First, sort the data set and compute the range of values (R) spanned by the observations.

(1) $$R = z_{max} - z_{min}$$

Second, select a value G, where G is any integer greater than or equal to 1, to divide the range into equal bands of width (w).

(2) $$w = R/G$$

Denote the group of observations within each band g_i (for i=1 through G), and count the number of observations (n_i) in each group. The sample frequency (f_j) for each $z_j \in g_i$ is defined

(3) $$f_j = n_i / \Sigma n_i^2$$

By definition of f_j in equations (1) through (3) the sum of the sample frequencies is equal to one.

(4) $$\Sigma f_j = 1$$

The sample frequency assigned to each observation will be a function of the group in which the observation lies, the relative number of observations in all other groups, and the total number of observations in the sample. With the estimated sample frequencies, the moment estimators are defined as follows. The mean of z is

(5) $$\hat{\mu} = \Sigma f_j z_j$$

The variance (σ^2) is defined

(6) $$\hat{\sigma}^2 = \sum_{j=1}^{N} f_j (z_j - \hat{\mu})^2$$

Relative skewness (ρ) is expressed

(7)
$$\hat{\rho} = \sum_{j=1}^{N} f_j (z_j - \hat{\mu})^3 / \hat{\sigma}^3$$

And the cumulative distribution is written

(8)
$$F = Prob\ (z \leq z_j) = \sum_{k=1}^{j} f_k$$

If prior information about the random variable is available, it can easily be incorporated into the estimation process. Possible sources of priors are econometric models, time series analyses, and secondary information from previous works. This research prescribes ordering the data set (as described above), and then varying the number of groups (G) that divide the data range. In this way the sample frequencies can be altered until the estimated moments converge on the prior moments. The objective then is to choose G to minimize the deviation between the estimated moments and the known priors. Ideally, the objective function should be specified to reflect the quality of the prior information and the relative importance of each moment to the particular study. If, for example, mean and variance priors are available, the objective might choose G to minimize the sum of the squared deviation from the prior mean and the absolute deviation from the prior variance.

Empirical Analysis

Experimental data published in Hexem and Heady (1978) were obtained to illustrate the properties of the estimation procedure and to evaluate the production risk of reducing input use. This study analyzed corn grain data from a series of field trials in Yuma Mesa, Arizona. Originally, the trials were conducted to test the effect of water and nitrogen application on corn yield. Water application rates ranged from 2.4 to 2.9 acre-feet per acre and nitrogen applications were 75, 225, and 375 pounds per acre. Results from six replicates of each input combination was available.

Hexem and Heady (1978) specified a quadratic model of corn yield response to water and nitrogen and estimated the model with least squares. Yield estimates from the quadratic model provided the information priors for this study. The yield response model is shown below. In equation (9), Z is corn yield in bushels per acre, W is water applied in acre-feet per acre, and N is nitrogen application in pounds per acre.

(9)
$$Z = \underset{(253.6)}{-1869} + \underset{(192.2)}{1447W} - \underset{(.1675)}{.6698N} - \underset{(36.33)}{270.0W^2} - \underset{(8.9*10^{-5})}{.0001N^2} + \underset{(.0598)}{.2666WN}$$

$$R^2 = .581 \quad F=16.65$$

This study applied the grouping method to the corn yield response data for nine different combinations of inputs. In the analysis, the number of groups was varied (between 1 and 7) to minimize the absolute deviation between the mean of the empirical distribution and the mean yield predicted by the econometric model (equation 9). The resulting empirical distribution was used to assess the riskiness of input use in terms of the probability of a crop failure. A failed crop was defined to be a harvest below the minimum yield required for breakeven returns. For the study region, the minimum corn yield for breakeven returns was deemed 73.2 bushels per acre. For comparison, the uniform distribution was estimated and its results were presented along with the results from the grouped distribution.

Results

For a single grouping (G=1), the grouped distribution was identical to the uniform distribution. As the number of groups increased, the mode of the grouped distribution became more peaked and moved toward the range of values that were most prevalent. Further increasing the number of groups caused the grouped distribution to flatten out and again to resemble the uniform distribution. For this data set, the number of groupings required to converge the empirical mean on the mean priors varied with input combination and ranged between 2 and 5. Interestingly, the number of groups required for convergence increased with both nitrogen and water application rate.

With minimal nitrogen application of 75 pounds and a water application rate of 2.7 acre-feet, the probability of crop failure was estimated to be very small (ε). Reducing the water application to 2.4 acre-feet increased the probability of crop failure to .99. For the input combination of 225 pounds of nitrogen and 2.9 acre-feet of water, both the grouped and the uniform distributions estimated the probability of a failed crop to be negligible. The grouped distribution indicated that reducing the water application rate to 2.6 acre-feet would increase the probability of failure to .62. The uniform distribution predicted that the water reduction would increase the risk of failure to .54. By applying 375 pounds of nitrogen and 2.8 acre-feet of water, both distributions suggested that the chance of a failed crop would be very small. The grouped distribution showed that a .10 acre-foot reduction in water use would raise the risk of a crop failure to .37. The uniform distribution predicted a slightly lower risk of .25. At 2.9 acre-feet of water, a reduction in the nitrogen application from a hefty 225 pounds to a meager 75 pounds per acre would increase the risk of crop failure from negligible to .58, according to the grouped distribution, and from negligible to .56 as given by the uniform distribution. The empirical results are displayed in Table 1.

Table 1. Estimated Probability of Crop Failure

Nitrogen N lb/ac	Water W af/ac	$Prob(z \leq 73.2)$ Uniform Distribution	$Prob(z \leq 73.2)$ Grouped Distribution
75			
	2.4	.99	.99
	2.7	ε	ε
	2.9	.56	.58
225			
	2.5	.17	.20
	2.6	.54	.62
	2.9	ε	ε
375			
	2.6	.85	.98
	2.7	.25	.37
	2.8	ε	ε

Note: Crop failure is defined as average corn yield below 73.2 bushels per acre. Epsilon, written as ε denotes small probabilities less than .01.

Three important conclusions can be drawn from the empirical results. The first is that at each nitrogen application level, a reduction in the water application rate increased the risk of crop failure. There were some inconsistencies, particularly at the lower nitrogen levels, but on the whole, the results indicated that inputs are an important factor in reducing farm risk. The second point to note is that the risk reducing role of water became more important as the nitrogen rate increased. It was most apparent at the high nitrogen level of 375 pounds per acre. The third and most important point, is that for every input combination, the uniform distribution predicted a lower risk of crop failure than the grouped distribution. Since we are quite certain that crop yields were sampled from a skewed parent distribution, we can be assured that the uniform distribution overestimates the probability of high yields and underestimates the risk of failure. The result nicely illustrates how a uniform distribution might mislead decision makers by underestimating the riskiness of input use. In this example, the differences between the two distributions were, admittedly, small. This suggests, that in applications where accuracy is less critical, the simpler uniform approximation may suffice. In general, however, the grouped distribution is likely to provide a more accurate representation of the true distribution and the additional effort required for the increase in accuracy is small.

Final Remarks

Because of the limitations involved in working with small data sets, the final estimates from the proposed methodology are not presumed to be completely free of bias. Instead, this paper suggests a method for assigning sample frequencies to observations in small data sets in a manner that preserves the relative frequency of observations between groups. The method is less subjective than hand-smoothing, more elegant than direct histogram construction, and accommodates prior information into the estimation process with relative ease. Because values are grouped, the empirical distribution is more likely to reveal the presence of skewness in small samples, and risk estimates are therefore likely to be more accurate. For empirical risk analysis, the grouped distribution provides information that is both necessary and useful.

References

Anderson, Jock R. "Sparse Data, Estimation Reliability and Risk-Efficient Decisions." *Amer. J. Agr. Econ.* 56(1974):564-72.

------. "Sparse Data, Climatic Variability and Yield Uncertainty in Response Analysis." *Amer. J. Agr. Econ.* 55(1973):77-82.

Day, Richard H. "Probability Distributions of Field Crop Yields." *J. Farm Econ.* 47(Aug. 1965):713-41.

Feinerman, E., J. Letey, and H.J. Vaux, Jr. "The Economics of Irrigation With Nonuniform Infiltration." *Water Resources Research.* 19(Dec. 1983):1210-14.

Hexem, Roger W. and Earl O. Heady. *Water Production Functions for Irrigated Agriculture.* Microfiche Table 6.19. Iowa State University Press. Ames, Iowa. 1978.

Pindyck, Robert S. and Daniel L. Rubinfeld. *Econometric Models and Economic Forecasts. 3rd Ed.* McGraw-Hill,Inc. 1991.

Schlaifer, Robert. *Probability and Statistics for Business Decisions.* pp102-9. McGraw-Hill Book Co., 1959.