



AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
<http://ageconsearch.umn.edu>
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

Staff Papers Series

Combining Farm and County Data to Construct
Farm Level Yield Distributions

by

Joan R. Fulton
Robert P. King
Paul L. Fackler



Department of Agricultural and Applied Economics

University of Minnesota
Institute of Agriculture, Forestry and Home Economics
St. Paul, Minnesota 55108

Combining Farm and County Data to Construct
Farm Level Yield Distributions*

by

Joan R. Fulton**

Robert P. King

Paul L. Fackler

* Selected paper to be presented at the annual meeting of the American Agricultural Economics Association, Knoxville, Tennessee, August, 1988.

** Joan R. Fulton is a Graduate Research Assistant and Robert P. King is an Associate Professor, Department of Agricultural and Applied Economics, University of Minnesota. Paul L. Fackler is an Assistant Professor, Department of Economics and Business, North Carolina State University.

Staff papers are published without formal review within the Department of Agricultural and Applied Economics.

The University of Minnesota is committed to the policy that all persons shall have equal access to its programs, facilities, and employment without regard to race, religion, color, sex, national origin, handicap, age, veteran status or sexual orientation.

Probability distributions of crop yields are an important input in risk management analyses by farmers. Farm yield data are one obvious source of information regarding yield distributions. Often, however, farmers do not have yield data for enough years to construct reliable representations of their yield distributions. Data at a more aggregated level, such as county average yields, may be available for a much longer time period and should provide some insights on the nature of farm yield distributions. As Eisgruber and Schuman demonstrate, however, the variance of county average yields will, in general, be less than that of farm level yields, and unadjusted county average yields cannot simply be used along with farm level yields to create a larger data set.

In this paper we consider the case where a farmer has a short time series of farm level data and a longer time series of county level data. The goal is to adjust the county level data and combine it with the farm level data to obtain a probability distribution for the farm. The literature on the "expert problem," which deals with the issue of combining probability assessments from different sources, provides insight for this problem.

In the sections which follow, we first review some of the issues raised in the expert problem literature. We then describe the data used in this study and discuss one approach to the important problem of adjusting county level data so that it is appropriate for use in characterizing farm level probability distributions. Next, we describe the alternative methods for combining the farm level and county level data to be considered in this analysis. We then report the results of empirical tests of the performance of each method. In the concluding section we summarize our findings and identify areas for further research.

The Expert Problem

Decision makers often seek information about uncertain quantities from external sources. In the literature on combining probability assessments, these external sources -- which may include people, statistical forecasting models, and sample data -- are termed "experts." The problem facing the decision maker is to combine his or her own beliefs regarding the uncertain quantity with information from experts in a manner that takes into account knowledge about the reliability of the expert information and yields a probability distributional representation that is consistent with the laws of probability.

If the expert information and the decision maker's prior knowledge are specified in terms of probability density functions, Bayes theorem provides a framework for addressing this problem. For continuous random variables Bayes theorem can be written as:

$$(1) \quad p(\theta | X) = \frac{p(\theta)p(X | \theta)}{p(X)} \quad \text{if } p(X) > 0$$

The usual interpretation of Bayes theorem is that the posterior probability of a random variable θ conditional on X , $p(\theta | X)$, is equal to the product of the prior probability density of θ , $p(\theta)$, and the likelihood of X conditional on θ , $p(X | \theta)$, divided by the probability of X , $p(X)$. Since the posterior probability is conditional on X , the term $p(X)$ is a constant. When factored out of (1), the inverse of $p(X)$ is referred to as a normalizing constant, since it ensures that the total integral of the density function equals one.

In a practical setting, decision makers find it particularly difficult to formalize their knowledge of $p(X | \theta)$, the likelihood function that is required for implementation of Bayes theorem. Research on this problem has focused on the identification of conditions under which simpler rules can be used to combine probabilities in a manner consistent with Bayes rule.

One notable result is the multiplicative rule derived by Morris. Assuming that information on only the variance of the expert's probability distribution has no effect on the decision maker's assessment (invariance to scale) and that the decision maker's assessment of the true value of the random variable will be in the tail of the expert's distribution is not conditional on the true value of the random variable (invariance to shift), Morris shows that the decision maker's posterior density is the product of his prior density, the expert's density, a calibration function, and a normalizing constant. When expert judgements are independent, this multiplicative rule can easily be extended to combine probability assessments from several experts.

The calibration function in the multiplicative rule ensures that the expert's probability assessment is well calibrated -- i.e. that "...over the long run, for all propositions assigned a given probability, the proportion that is true equals the probability assigned" (Lichtenstein, Fischhoff, and Phillips, p. 307). In this study, a probability distribution based on unadjusted county average data would not be calibrated for farm level yields if it underestimated their variance. Therefore, one practical problem to be addressed in this study is that of calibrating the county level data. When the calibration function is to be empirically based, the lack of farm level data complicates this problem.

Another difficulty encountered when combining probability assessments is dependence among the experts. It is frequently the case that probability assessments from experts are not independent of each other or of the decision maker's prior. Probability assessments can be dependent even if the experts have formulated the assessments independently and do not know each other. Similar training and experience or the use of common estimation techniques or common data is sufficient to render them dependent. In this study, farm

and county level data would be expected to be correlated, since both are subject to the effects of weather, widespread disease and pest problems. When such dependence exists, the decision maker should allow for it in formulating a posterior density. Winkler (1981) considers the case of dependence among experts when the probability distributions are normally distributed. His results will be introduced and applied later in this paper.

Given the difficulty of applying Bayes theorem in practical settings, other researchers have suggested the use of "rule of thumb" or "weighted average" methods of combining probability assessments. There is evidence in the literature to support the use of simpler methods. Winkler (1986, p. 301) questions "whether modeling in the expert resolution problem can be effective enough in practice to justify the time and effort that it requires." von Winterfeldt and Edwards (p. 134) note that procedures implementing Bayes theorem involve greater mathematical complexity than simple averaging. They conclude that "no reason exists for expecting such procedures to lead to better estimates." Therefore, we also consider weighted average schemes.

Design of the Empirical Analysis

In this study, data on per acre corn yields for five farms in Nobles County of southwestern Minnesota were used to investigate alternative strategies for constructing farm level yield distributions. For each farm, data for the years 1958-1982 were available from farm record data. Average per acre corn yields in Nobles County for the same time period were also available. To simplify the analysis, the effects of technological change over time were eliminated by detrending each data series with coefficients determined from an ordinary least squares regression of yield on time. The methods developed in this study can, however, be modified to include consideration of yield trends.

The detrended yield series are presented in Table 1. As predicted by Eisgruber and Schuman, the value of the standard deviation is lower for the county than for any one of the individual farms. Given the unique characteristics of individual farms the mean yield is expected to differ across farms, a point which is confirmed by this data series. The average county yield will not, in general, be directly relevant for an individual farm.

Construction of farm level yield distributions requires assumptions regarding the distributional form of the farm and county level data and the resulting posterior distribution. Consistent with much of the work done in the area of risk management, the normal distribution for the prior and posterior distributions is considered. In addition a nonparametric approach is considered, with the construction of an empirical CDF. The latter approach is supported by the findings of Pope and Ziemer that "... the empirical distribution function performs favorably relative to appropriate ML methods, especially in small samples and regardless of the underlying parent distributions" (p. 40).

Calibrating County Level Data

As evident from Table 1, adjustments to the location and dispersion of the county data are required for it to be a meaningful input into the creation of a farm yield distribution. The design of this study considers two cases; one where the farmer has three years of farm level data to combine with 15 years of county level data and one where 10 years of farm level data are combined with 15 years of county level data. In these cases, there are then three and 10 common years of farm and county data. The objective is to use the common years of farm and county data to determine the relationship between the particular farm yields and the county average yields. The 15-year county data time series is then adjusted to reflect this relationship.

Table 1: Detrended Yield Series

| Year | Farm 1 | Farm 2 | Farm 3 | Farm 4 | Farm 5 | County |
|-----------|--------|--------|--------|--------|--------|--------|
| 1958 | 122.28 | 106.99 | 114.93 | 108.80 | 120.94 | 97.73 |
| 1959 | 96.15 | 119.10 | 106.22 | 102.50 | 110.19 | 94.71 |
| 1960 | 125.08 | 131.72 | 128.20 | 107.75 | 124.43 | 99.69 |
| 1961 | 126.40 | 137.68 | 121.03 | 116.07 | 132.88 | 108.67 |
| 1962 | 116.65 | 114.47 | 115.99 | 109.00 | 115.60 | 98.65 |
| 1963 | 130.50 | 140.62 | 110.97 | 114.16 | 147.85 | 109.63 |
| 1964 | 93.18 | 102.66 | 114.01 | 53.90 | 103.54 | 98.61 |
| 1965 | 100.87 | 84.54 | 102.46 | 84.88 | 98.96 | 90.59 |
| 1966 | 120.92 | 136.68 | 116.17 | 98.66 | 122.75 | 106.57 |
| 1967 | 112.94 | 102.04 | 122.40 | 109.93 | 109.20 | 98.55 |
| 1968 | 132.98 | 116.04 | 126.77 | 97.63 | 108.30 | 103.53 |
| 1969 | 128.70 | 122.02 | 133.52 | 120.73 | 135.36 | 117.51 |
| 1970 | 134.57 | 141.34 | 133.28 | 42.49 | 138.02 | 108.49 |
| 1971 | 111.24 | 113.79 | 121.19 | 123.87 | 126.83 | 99.47 |
| 1972 | 132.51 | 139.11 | 151.54 | 124.50 | 150.12 | 118.45 |
| 1973 | 115.95 | 130.05 | 125.94 | 100.35 | 126.72 | 104.43 |
| 1974 | 68.81 | 84.71 | 95.47 | 78.96 | 93.24 | 63.41 |
| 1975 | 94.92 | 106.01 | 96.66 | 65.61 | 105.96 | 89.39 |
| 1976 | 67.20 | 63.29 | 72.54 | 56.37 | 83.68 | 61.37 |
| 1977 | 115.21 | 127.80 | 89.36 | 110.26 | 106.74 | 107.35 |
| 1978 | 140.27 | 138.12 | 135.67 | 131.03 | 119.12 | 111.33 |
| 1979 | 130.65 | 108.72 | 108.71 | 107.35 | 122.51 | 106.31 |
| 1980 | 110.95 | 110.74 | 100.42 | 112.11 | 126.29 | 89.29 |
| 1981 | 139.73 | 151.58 | 158.05 | 108.86 | 141.27 | 118.27 |
| 1982 | 117.85 | 126.01 | 125.84 | 105.01 | 126.58 | 111.25 |
| Mean | 115.46 | 118.23 | 117.09 | 99.63 | 119.88 | 100.53 |
| Std. Dev. | 19.13 | 20.42 | 18.49 | 22.76 | 16.20 | 13.89 |

For the years of common data, the standard deviation for the farm, σ_f , and for the county, σ_c , and the difference between the farm mean and the county mean, d , are calculated. Each element of the calibrated county series X_{ci} , is obtained from:

$$(2) X_{ci} = (m + (\sigma_f / \sigma_c)(X_i - m)) + d,$$

where X_i is the i th observation of the uncalibrated data and m is the mean of the 15-year county series. The calibrated county data are given in Table 2.

This method of calibration ensures that the adjusted county data is relevant in creating farm level yield distributions and also preserves the additional information contained in the longer county time series. This latter point is seen in Table 2 where the mean and standard deviation for the calibrated county differ from those for the farm.

Methods for Combining Farm and Calibrated County Data

Two base strategies in this study are to use farm level data only and calibrated county level data only. These are reported as methods one and two in the empirical results section. It is interesting in itself to compare these strategies since it provides insight into the question of when there is enough farm level data to make it unnecessary to use county level data. Three additional strategies involving a weighted average of farm and calibrated county level data are considered for the nonparametric case. These strategies, reported as methods three, four, and five, are to use equal weights, weights based on sample size with no consideration of dependence, and weights based on sample size with dependence considered. The weights assigned to the farm level data for methods four and five respectively are:

$$(3) w_f = \frac{2n_f}{n_f + n_c} \text{ and,}$$

$$(4) w_{fd} = [1 - (w_f - 0.5)^p] w_f.$$

Table 2: Calibrated County Level Data

| Year | Uncalibrated County | Three Years of Farm Level Data | | | | | Ten Years of Farm Level Data | | | | |
|--|---------------------|--|--------|--------|--------|--------|------------------------------|--------|--------|--------|--------|
| | | Farm 1 | Farm 2 | Farm 3 | Farm 4 | Farm 5 | Farm 1 | Farm 2 | Farm 3 | Farm 4 | Farm 5 |
| 1958 | 97.73 | 113.00 | 116.88 | 120.82 | 63.42 | 125.97 | 108.65 | 105.51 | 112.53 | 76.77 | 110.40 |
| 1959 | 94.71 | 108.89 | 112.01 | 115.96 | 48.42 | 122.26 | 103.62 | 98.77 | 107.75 | 66.87 | 104.03 |
| 1960 | 99.69 | 115.67 | 120.04 | 123.98 | 73.16 | 128.37 | 111.92 | 109.88 | 115.63 | 83.19 | 114.54 |
| 1961 | 108.67 | 127.88 | 134.51 | 138.43 | 117.78 | 139.39 | 126.89 | 129.91 | 129.83 | 112.64 | 133.48 |
| 1962 | 98.65 | 114.25 | 118.36 | 122.30 | 67.99 | 127.09 | 110.19 | 107.56 | 113.98 | 79.79 | 112.34 |
| 1963 | 109.63 | 129.19 | 136.06 | 139.97 | 122.54 | 140.57 | 128.50 | 132.05 | 131.35 | 115.78 | 135.50 |
| 1964 | 98.61 | 114.20 | 118.30 | 122.24 | 67.80 | 127.05 | 110.12 | 107.47 | 113.92 | 79.65 | 112.26 |
| 1965 | 90.59 | 103.29 | 105.37 | 109.33 | 27.95 | 117.21 | 96.75 | 89.58 | 101.23 | 53.36 | 95.34 |
| 1966 | 106.57 | 125.03 | 131.13 | 135.05 | 107.34 | 136.81 | 123.39 | 125.22 | 126.51 | 105.75 | 129.05 |
| 1967 | 98.55 | 114.12 | 118.20 | 122.14 | 67.50 | 126.97 | 110.02 | 107.33 | 113.82 | 79.46 | 112.13 |
| 1968 | 103.53 | 120.89 | 126.23 | 130.16 | 92.24 | 133.08 | 118.32 | 118.44 | 121.70 | 95.78 | 122.64 |
| 1969 | 117.51 | 139.91 | 148.76 | 152.65 | 161.69 | 150.23 | 141.64 | 149.62 | 143.82 | 141.62 | 152.12 |
| 1970 | 108.49 | 127.64 | 134.22 | 138.14 | 116.88 | 139.17 | 126.59 | 129.51 | 129.55 | 112.05 | 133.10 |
| 1971 | 99.47 | 115.37 | 119.68 | 123.62 | 72.07 | 128.10 | 111.55 | 109.39 | 115.28 | 82.47 | 114.07 |
| 1972 | 118.45 | 141.19 | 150.28 | 154.17 | 166.36 | 151.39 | 143.20 | 151.72 | 145.31 | 144.70 | 154.11 |
| Mean | 103.39 | 120.70 | 126.00 | 129.93 | 91.54 | 132.91 | 118.09 | 118.13 | 121.48 | 95.33 | 122.34 |
| Std.Dev. | 8.01 | 10.90 | 12.91 | 12.89 | 39.79 | 9.83 | 13.56 | 17.87 | 12.67 | 26.26 | 16.90 |
| Mean for the three years 1970-1972 | | Mean for the ten years 1963-1972 | | | | | | | | | |
| Farm | | 126.11 | 131.41 | 135.34 | 96.95 | 138.32 | 119.84 | 119.88 | 123.23 | 97.08 | 124.09 |
| County | | 108.80 | 108.80 | 108.80 | 108.80 | 108.80 | 105.14 | 105.14 | 105.14 | 105.14 | 105.14 |
| Standard Deviation for the three years 1970-1972 | | Standard Deviation for the ten years 1963-1972 | | | | | | | | | |
| Farm | | 12.91 | 12.30 | 15.28 | 47.17 | 11.65 | 14.67 | 19.62 | 13.91 | 28.83 | 18.55 |
| County | | 9.49 | 9.49 | 9.49 | 9.49 | 9.49 | 8.79 | 8.79 | 8.79 | 8.79 | 8.79 |

In these equations n is the number of observations, ρ is the sample correlation coefficient, and the subscripts f and c denote the farm and county respectively, while the subscript d denotes dependence. In each case, the weight assigned to the calibrated county data is one minus the weight assigned to the farm data. This weighting scheme assumes that the farmer is able to obtain at least as much county level data as farm level data. If the number of observations is the same and there is no dependence the weight assigned to the farm will be one. If there is positive correlation between the farm and county data, the weight assigned to the farm data will be reduced if, in the absence of dependence, more than half of the weight would have been assigned to the farm data.

Since a normal distribution is fully described by its mean and variance, the posterior distribution for the assumption of normal distributions can be represented by mean, μ^* , and variance, σ^{2*} . As derived by Winkler (1981, p. 484)

$$(5) \quad \mu^* = \frac{\frac{\sigma_f^2}{2} - \rho\sigma_f\sigma_c}{\frac{\sigma_f^2}{2} + \frac{\sigma_c^2}{2} - 2\rho\sigma_f\sigma_c} \mu_c + \frac{\frac{\sigma_c^2}{2} - \rho\sigma_f\sigma_c}{\frac{\sigma_f^2}{2} + \frac{\sigma_c^2}{2} - 2\rho\sigma_f\sigma_c} \mu_f \quad \text{and}$$

$$(6) \quad \sigma^{2*} = \frac{(1-\rho^2)\sigma_f^2\sigma_c^2}{\frac{\sigma_f^2}{2} + \frac{\sigma_c^2}{2} - 2\rho\sigma_f\sigma_c},$$

where the subscripts f and c represent farm and county. We consider the possibility of ignoring dependence ($\rho=0$) and accounting for dependence ($\rho \neq 0$), reported as methods four and five in the empirical results section.

Empirical Results

Our criterion for evaluating the methods of constructing farm yield distributions is calibration. Empirically, an assessment process is well calibrated if the CDF values associated with a sample of realizations are

uniformly distributed. In this analysis we calculate CDF values, U_i , associated with yields in the 10-year post sample period. Testing for calibration involves using goodness of fit tests for the hypothesis that the sample U_i 's come from a uniform $[0,1]$ distribution. Stephens identifies five nonparametric tests for uniformity. They are the Kolmogorov, D , the Cramér-von Mises, W^2 , the Kuiper, V , the Watson, U^2 , and the Anderson-Darling, A^2 , statistics. In this study the Cramér-von Mises, W^2 , and the Watson, U^2 , tests are used. Stephens evaluates the power of each of these tests and shows that the W^2 statistic is effective at detecting a change in the mean while the U^2 statistic is effective at detecting a change in the variance.

The results of the tests for calibration are presented in Table 3. The null hypothesis is that the distribution is well calibrated or that the U_i 's are uniformly distributed on the interval $[0,1]$. Cases where we fail to reject the null hypotheses at the 0.15 level of significance are denoted by the symbol, $*$, while cases where we fail to reject the null hypothesis at the 0.10 level of significance are denoted by the symbol, $*$. Since it is easier to reject the null hypothesis at the 0.15 level than at the 0.10 level of significance the former provides a stronger test for calibration.

As one would expect, yield distributions derived using 10 years of farm level data are better representations than those derived using only three years of farm level data. The evidence from the analysis with the five farms considered here indicates that if a farmer has 10 years of farm level data, there is no advantage to combining it with calibrated county data.

The distributions derived from the calibrated county data consistently perform well, while the use of weighted average methods yields mixed performance. This may be due to the way in which we use the farm data to calibrate the county data. Since the calibration method already makes use

Table 3: Calibration Test Results

| Farm Method | | Three Years of Farm Data | | | | Ten Years of Farm Data | | | |
|-------------|---|--------------------------|---------|----------------|---------|------------------------|---------|----------------|---------|
| | | W ² | | U ² | | W ² | | U ² | |
| | | N | NP | N | NP | N | NP | N | NP |
| 1 | 1 | 0.502 | 0.598 | 0.154 | 0.314 | 0.136** | 0.175** | 0.080** | 0.130** |
| | 2 | 0.257** | 0.153** | 0.165 | 0.119** | 0.110** | 0.117** | 0.097** | 0.124** |
| | 3 | - | 0.344* | - | 0.077** | - | 0.209** | - | 0.093** |
| | 4 | 0.833 | 0.311* | 0.531 | 0.048** | 0.320* | 0.118** | 0.263 | 0.048** |
| | 5 | 0.466 | 0.316* | 0.263 | 0.052** | 0.197** | 0.173** | 0.162 | 0.087** |
| 2 | 1 | 0.506 | 1.075 | 0.099** | 0.314 | 0.021** | 0.065** | 0.033** | 0.064** |
| | 2 | 0.265** | 0.317* | 0.119** | 0.183 | 0.026** | 0.031** | 0.051** | 0.057** |
| | 3 | - | 0.737 | - | 0.129** | - | 0.220** | - | 0.100** |
| | 4 | 0.532 | 0.661 | 0.250 | 0.100** | 0.285* | 0.125** | 0.305 | 0.087** |
| | 5 | 0.438 | 0.673 | 0.172 | 0.103** | 0.124** | 0.212** | 0.145* | 0.153 |
| 3 | 1 | 1.237 | 1.099 | 0.295 | 0.338 | 0.563 | 0.681 | 0.237 | 0.346 |
| | 2 | 0.954 | 0.865 | 0.279 | 0.321 | 0.536 | 0.647 | 0.275 | 0.312 |
| | 3 | - | 1.488 | - | 0.239 | - | 0.972 | - | 0.154 |
| | 4 | 1.402 | 1.769 | 0.452 | 0.329 | 0.760 | 0.610 | 0.499 | 0.109** |
| | 5 | 1.159 | 1.714 | 0.333 | 0.315 | 0.665 | 0.689 | 0.370 | 0.075** |
| 4 | 1 | 0.187** | 0.769 | 0.209 | 0.674 | 0.088** | 0.153** | 0.109** | 0.173 |
| | 2 | 0.178** | 0.231** | 0.165 | 0.215 | 0.107** | 0.121** | 0.113** | 0.144* |
| | 3 | - | 0.324* | - | 0.258 | - | 0.107** | - | 0.101** |
| | 4 | 0.170** | 0.179** | 0.141* | 0.121** | 0.301* | 0.136** | 0.251 | 0.148* |
| | 5 | 0.168** | 0.179** | 0.140* | 0.122** | 0.248** | 0.133** | 0.214 | 0.144* |
| 5 | 1 | 1.931 | 2.725 | 0.372 | 0.674 | 0.204** | 0.307* | 0.057** | 0.086** |
| | 2 | 1.370 | 1.583 | 0.273 | 0.381 | 0.133** | 0.178** | 0.047** | 0.093** |
| | 3 | - | 1.657 | - | 0.254 | - | 0.437 | - | 0.138* |
| | 4 | 2.400 | 1.450 | 0.672 | 0.235 | 0.244** | 0.410 | 0.105** | 0.158 |
| | 5 | 1.629 | 1.494 | 0.335 | 0.229 | 0.199** | 0.463 | 0.063** | 0.195 |

* indicates failure at the 0.10 level of significance to reject the null hypothesis that the distribution is well calibrated.

* indicates failure at the 0.15 level of significance to reject the null hypothesis that the distribution is well calibrated.

N - Normal Distribution

NP - Nonparametric

of the information in the farm level data, further combining of the farm and calibrated county data does not improve the results.

Under the assumption of the normal distribution, the methods where farm and calibrated county level data are combined tend to perform worse than the two base strategies. A possible explanation lies in the formula used to combine the distributions, which assigns weights based on the variance of the distribution. It does not allow for the possibility that there may be other indicators of the degree of confidence of the expert information, such as the number of sample observations.

For the nonparametric case, different weighting methods have little effect on the performance of the resulting distributions.

These results provide no strong indication regarding the choice of the nonparametric or normal distribution. In practical usage, the choice will most likely depend on the context of the problem. If one has reason to believe that the yield distribution for the farm in question is not normal, one would want to use the empirical CDF. However, the normal distribution may be the choice if its use simplifies the analysis and one does not have prior beliefs regarding the form of the distribution.

Conclusions

Probability yield distributions were constructed for five farms in southwestern Minnesota assuming a normal distribution and an empirical CDF or nonparametric case. The purpose was to consider the possibility of calibrating county level data to combine with farm level data when only a small amount of farm level data is available. Five different methods for combining the distributions were used for the empirical CDF and four different methods were used for the normal distribution.

Using the criterion of calibration we found that more years of farm level data produced better yield distributions. Yield distributions constructed using calibrated county data only generally performed well, indicating that an effective method of calibrating county level data may have been identified.

The results of this study, although they must be considered exploratory, indicate potential for the use of county level data in the creation of farm yield distributions. Further study with more farms in different regions will explore this potential further.

References

- Eisgruber, L.M. and L.S. Schuman. "The Usefulness of Aggregated Data in the Analysis of Farm Income Variability and Resource Allocation." Journal of Farm Economics 45(1963):587-591.
- Lichtenstein, S., B. Fischhoff, and L.D. Phillips. "Calibration of Probabilities: The State of the Art to 1980" in Judgment Under Uncertainty: Heuristics and Biases, D. Kahneman, P. Slovic and A. Tversky (eds.) Cambridge University Press, Cambridge, 1982.
- Minnesota Crop and Livestock Reporting Service, U.S. Department of Agriculture Statistical Reporting Service. Minnesota Agricultural Statistics. 1976-1982
- Morris, P.A. "Combining Expert Judgements A Bayesian Approach." Management Science. 23(1977):679-693
- Pope, R.D. and R. F. Ziemer. "Stochastic Efficiency, Normality, and Sampling Errors in Agricultural Risk Analysis." American Journal of Agricultural Economics. 66(1984):31-40
- State-Federal Crop and Livestock Reporting Service, Minnesota Department of Agriculture. Minnesota Agricultural Statistics. 1958-1975
- Stephens, M.A. "EDF Statistics for Goodness of Fit and Some Comparisons." Journal of the American Statistical Association. 69(1974):730-737
- von Winterfeldt, D. & W. Edwards. Decision Analysis and Behavioral Research. Cambridge University Press, Cambridge. 1986.
- Winkler, R.L. "Combining Probability Distributions from Dependent Information Sources." Management Science. 27(1981):479-488
- Winkler, R.L. "Expert Resolution." Management Science. 32(1986):298-303