1986 c

Pooling Cross-Sectional and Time-Series Yield Data for Risk Analysis*

by

Joseph Atwood, Glenn A. Helmers, Kent M. Eskridge, John M. Morrill, and Michael R. Langemeier**

Risk -- Mathematical models

---

* Paper for presentation at the Annual Meeting of the American Agricultural Economics Association, Reno, Nevada, July 27-30, 1986.

** Assistant Professor of Agricultural Economics, Auburn, University; Professor of Agricultural Economics, University of Nebraska; Consultant, Biometrics and Information Systems Center, University of Nebraska; and Graduate Research Assistants, Department of Agricultural Economics, University of Nebraska, respectively.

Pooling Cross-Sectional and Time-Series Yield Data for Risk Analysis

Introduction and Objectives

There has long been concern about the adequacy of crop yield data in agricultural production risk research. Time series of crop yields are used extensively in agricultural risk analysis for both statistical (mean-variance) comparisons and diversification (risk-income) models.

First, as discussed by Eisgruber and Schuman, aggregate (such as county) yield series understate the variability experienced by individual farms. However, aggregate yield data are readily available while individual farm yield data are much more difficult to secure. Further, an individual farm may not have a yield history on all crops being analyzed.

Next, assuming a yield time series of 10-15 years exists for an individual farm, it can be questioned if that series has experienced the potential range of outcomes which exist from farm to farm. Some events such as disease, hail, and drouth occur in such an irregular pattern that individual farms may escape serious impact of such events over a given time period while other farms may repeatedly experience such incidence.

Because of the above problems, it can be suggested that lengthening the time series will remedy this problem. However, it is difficult to locate individual farm yield series for a long (say 30 years or more) time period. Next, there may have been major changes in management, technology, input use, and other factors such that complex adjustments are required to evaluate the pure variability characteristics of the yield series.

Finally, there remains the issue of the purpose of the analysis. That is, is the risk analysis directed to a specific farm or a general area analysis? Usually risk studies have been focused at general risk relationships which are experienced for a large group of farms. This does not suggest the use of average yields but it does suggest that a data series from an individual farm may be incomplete to represent an area. Yield relationships among crops may be considerably different from farm to farm, hence the choice of any one farm to represent the area may result in too much confidence placed in specific farm results. Thus, whether specific farm or area oriented, should yield relationships between crops vary, a wider assembly of yield data beyond one individual farm may be warranted.

With these concerns in mind the objectives of this paper are to:

(1) Examine farm to farm crop yield variability characteristics for alternative crops with particular emphasis placed on differences in yield correlation coefficients among farms,

(2) Develop a pooled yield data set from several farms with attention paid to time (technology) and farm (soil-management) adjustments.

(3) Examine differences in risk-income frontiers for

  (a) individual farms using individual farm data,

  (b) an average yield setting, and

  (c) a pooled cross-sectional and time-series data set, and

(4) Assess the use of using pooled yield data in risk analysis.

Yield Data

Yield data for this study were taken from Nebraska Farm Business records for eight individual farms across four East-central Nebraska counties. The eight farms fall within a forty-mile radius and were chosen because they each provided yields for seven full years of the regionally representative crops corn, soybeans, and alfalfa.

The yield data are presented in Table 1. It can be observed that considerable variability in yields exist for 1) across years for a given farm, 2) average (over-time) yields among the eight farms for a crop, 3) high and

Table 1. Yields of Corn, Soybeans and Alfalfa for Eight Farms for 1977-1983.

Corn - bu. per acre

	Year							
Farm	1977	1978	1979	1980	1981	1982	1983	Ave.
1	70	134	113	60	116	103	90	98.0
2	104	103	88	54	112	91	60	87.4
3	76	91	89	60	101	88	84	84.1
4	68	125	108	83	103	43	63	84.7
5	45	132	114	70	59	103	66	84.1
6	91	108	81	67	78	90	64	82.7
7	65	102	114	82	86	49	68	80.9
8	77	52	101	65	98	80	57	75.7
Ave.	74.5	105.9	101.0	67.6	94.1	80.9	69.0	84.7

Soybeans - bu. per acre

	Year							
Farm	1977	1978	1979	1980	1981	1982	1983	Ave.
1	41	34	33	42	49	35	24	36.9
2	42	28	26	26	39	32	27	31.4
3	43	35	35	33	38	40	21	35.0
4	42	45	39	35	44	32	31	38.3
5	37	43	35	31	42	38	30	36.6
6	28	24	36	35	37	37	27	32.0
7	40	39	40	28	45	27	31	35.7
8	35	21	39	26	38	36	28	31.9
Ave.	38.5	33.6	35.4	32.0	41.5	34.6	27.4	34.7

Alfalfa - tons per acre

	Year							
Farm	1977	1978	1979	1980	1981	1982	1983	Ave.
1	1	6	6	5	5	5	5	4.7
2	3	3	2	4	3	2	2	2.7
3	2	3	2	3	3	5	5	3.3
4	4	5	4	6	4	4	4	4.4
5	3	5	5	6	3	3	5	4.3
6	5	3	3	4	3	3	3	3.4
7	5	5	5	4	4	5	4	4.6
8	6	1	5	2	2	4	4	3.4
Ave.	3.6	3.9	4.0	4.25	3.4	3.9	4.0	3.9

low-year average (across farm) yields by crop, and 4) high and low average (overtime) yields by farm between crops.

Statistical Model

The structure of the data is of cross sectional-time series nature with farms representing cross sections. Initially, a simple dummy variable model was fit for each crop yield with linear and quadratic slope coefficients for time trend varying across farm. That is:

$$Y_{iT} = b_{0i} + b_{1i}T + b_{2i}T^2 + e_{iT}$$

where Y_{iT} = yield on the i'th farm at time T

b_{0i} = intercept for the i'th farm

b_{1i} = linear slope coefficient for i'th farm

b_{2i} = quadratic slope coefficient for i'th farm

e_{iT} = residual error for i'th farm at time T.

Different intercepts were assumed to represent different management practices, soil types and other environmental influences on yield which did not interact with time. Different slope coefficients were fit anticipating the possibility of differences in technology over time.

Because contemporaneous correlation across farms was considered important, a seemingly unrelated regressions (SUR) approach was considered. However because the independent variable time was identical for all farms, SUR was equivalent to ordinary least squares, the estimation method used here.

Overall F tests for heterogeneous slopes and intercepts for all crops failed to reject the zero impact hypothesis at the .01 level of significance. Refitting the model with common intercept and common linear and quadratic time effects failed to detect a significant linear or quadratic time trend (Table 2). Individual farm models with linear quadratic terms were also fit for each crop with similar results. The results of Table 2 should be interpreted as

Table 2. P Values for Pooled Model.

heterogeneous intercepts and slopes model

	Y_{iT} = b_{0i}	+ $b_{1i}T$	+ $b_{2i}T^2$
corn	(.8538)*	(.9261)*	(.9280)*
hay	(.0174)*	(.2332)*	(.1625)*
soybeans	(.3968)*	(.7160)*	(.8372)*

homogeneous slopes model

	Y_{iT} = b_0	+ b_1T	+ b_2T^2
corn		(.0744)**	(.0527)**
hay		(.8423)**	(.8183)**
soybeans		(.0372)**	(.1588)**

* Probability > F when $b_{K1} = b_{K2} = \ldots = b_{K8}$, K = 0-2

** Probability > F when $b_P = 0$, P = 1, 2

---

the probability of obtaining an F value larger than the calculated F value when the null hypothesis is true. The higher the value the less evidence against the acceptance of the null hypothesis. Thus, each crop yield for all farms was assumed to be generated by a process common to all farms.

Risk Programming Assumptions

Nominal per acre net returns for 1977-1983 were calculated for the three crops using the yield data, historical grain prices, and estimated production costs. The base year used to generate the variable costs for each crop was 1983. Relevant costs included fertilizer, herbicide, seed, fuel, labor charges, and repair and maintenance. A cost for establishing alfalfa was not included. Assuming constant input use for the study period, appropriate USDA indexes were then used to develop cost estimates for 1977-1982. A real interest charge on operating capital was included in developing costs.

The nominal return series was deflated to a real series (base 1973) using the consumer price index. The resulting series represents a real return to land, machinery and management. The effective constraint on land is assumed to be 1,000 acres. Labor and capital are not constrained.

MOTAD was used to analyze the tradeoff between expected returns and risk (in this case negative deviations below the mean). Tauer (1983) indicated that solutions on the MOTAD E-A frontier are not necessarily part of the SSD set. Watts et al. (1984) indicated that solutions within the efficient set would not be reached by MOTAD. However, because individuals are familiar with the MOTAD model, it is used as a comparative method in this analysis.

Results

Variance and Covariance Analysis

The ten settings ((a) eight individual farms for seven years, (b) average yields of eight farms for seven years, and (c) pooled yields with 56 observations) are presented in Table 3 with respect to yield dispersion characteristics. Considerable difference exists among farms with respect to yield standard deviation. The range of standard deviations for corn is from 13.0 for farm 3 to 32.2 for farm 5. The range for soybean yields is 5.0 (farm 5) to 8.0 for farm 1. For alfalfa the range is from .5 on farm 7 to 1.8 on farm 8. Of course, these standard deviations cannot be compared directly among crops because it is a yield analysis, not a returns analysis. Note that the above high and low standard deviations each exist on a different farm. Very little consistency exists in overall yield variability for a given farm. The possible exceptions to that are farms 1 and 6 (farm 1 having consistently high variability and farm 6 having consistently low variability).

As mentioned earlier, yield variability based on average yields is expected to be lower than individual farm variability. This is obviously the case here. For soybeans and alfalfa, variability based on average yields is lower than for any of the individual eight farms. Pooled yield variability

Table 3. Yield Correlation Coefficients and Standard Deviations for Corn, Soybeans, and Alfalfa for Eight Farms, Average Yields, and Pooled Yields.

	Standard Deviation			Correlation Coefficients		
Farm	Corn	Soybeans	Alfalfa	Corn-Soybeans	Corn-Alfalfa	Soybeans-Alfalfa
1	26.3	8.0	1.7	-.127	.605	-.285
2	22.4	6.6	.8	.690	-.100	.129
3	13.0	7.0	1.3	.100	.109	-.491
4	28.8	5.7	.8	.765	.271	.005
5	32.2	5.0	1.3	.344	.317	-.536
6	15.2	5.5	.8	-.367	.012	-.193
7	22.4	7.0	.5	.556	.092	.141
8	19.1	6.9	1.8	.931	.366	.608
Average Annual Yields	15.6	4.5	.3	.413	-.320	-.803
Pooled Yields	22.6	6.6	1.3	.366	.266	.122

tends to be representative of "average" variability individual farms.

Dramatic differences in covariance relationships are observed among farms. Correlation coefficients for corn - soybeans range from -.367 to .931 (farms 6 and 8 respectively). Corn - alfalfa yield correlation coefficients range from -.100 (farm 2) to .605 (farm 1). For soybeans - alfalfa the range of -.536 to .608 is experienced (farms 5 and 8 respectively). The most startling covariance relationships are observed for average yields. For corn-alfalfa and soybeans-alfalfa these lie outside the range of levels experienced by individual farms! For example, only one farm (farm 2) had a negative yield correlation coefficient between corn and alfalfa (-.100). However, the coefficient based on average yields was -.320. Only for corn-soybeans does the correlation coefficient for average yields appear to be representative of the farms as a group. It can be seen that the pooled yield data result in correlation coefficients roughly "average" of those experienced on the eight farms for the three crops.

Risk Programming Analysis

After converting the yield series to net returns series, risk-income frontiers were estimated for each of the ten settings. Because average product prices and average input costs were used, the income variability resulting from the analysis is due strictly to yield variability, the focus of the analysis. The frontiers presented in Table 4 are largely confined to the region in which no scaling down of land use is experienced. For each setting, only the first of the $5,000 expected income increments in which land goes slack is presented, in addition to frontier points lying above.

Quite obviously the individual frontiers differ in expected income due to differences in average yield. For this reason the frontiers cannot be directly compared, although organizational tendencies can be observed. For all settings, the LP solution was 1,000 acres of soybeans. Farm 4 has the highest expected income for the LP solution because farm 4's average soybean yield was highest of all farms. Thus, even with the high level of yield variability for these farms, the LP solutions were stable.

The pattern of organizational change starting at the LP solution and moving downward varies widely among the ten settings. For farms 2, 5, 6, and 8 the crop organization shifts from soybeans to soybeans and corn. For farm 4 the organization changes from soybeans to varying levels of soybeans and alfalfa. A third pattern is observed for farm 1 where crop acreage first moves from soybeans to soybeans and alfalfa and then "scales down" to soybeans and corn. For farms 3 and 7 as well as the setting based on average yields, the frontier shifts from soybeans to soybeans and alfalfa and then diversifies in all three crops. The fifth pattern is exhibited only by the pooled setting. There the organization shifts from soybeans to soybeans and corn and then diversifies in the three crops.

Examining diversification potential, farms 1, 3, 6, and 7 exhibit large

Table 4. MOTAD Solutions for Eight Individual Farms Average Yields, and Pooled-Data.[1/]

Farm	Expected Income	Negative Deviations	Organization-Acres Corn	Soybeans	Alfalfa
1	[2/] 87,569	94,260		1000	
	85,000	68,988		788	212
	80,000	52,613	484	512	
2	[2/] 69,149	60,329		1000	
	69,000	60,115	16	984	
	65,000	56,607	19	923	
3	[2/] 81,227	86,039		1000	
	80,000	81,437		969	31
	75,000	63,044	42	829	129
	70,000	45,666	202	644	153
	65,000	28,288	363	459	178
	60,000	13,720	453	300	246
	55,000	7,390	387	240	336
4	[2/] 93,510	93,190		1000	
	90,000	85,181		833	167
	85,000	74,563		596	404
	80,000	66,644		358	642
	75,000	58,900		125	873
5	[2/] 86,290	67,640		1000	
	86,000	67,150	10	990	
	85,000	65,956	25	969	
6	[2/] 70,137	52,432		1000	
	70,000	52,070	9	991	
	65,000	42,558	211	761	
7	[2/] 84,469	82,441		1000	
	80,000	62,548		432	568
	75,000	52,386	123	141	736
	70,000	47,378	290		707
8	[2/] 69,110	56,760		1000	
	68,000	55,812	48	952	
	65,000	53,346	50	907	
Average Yields	[2/] 80,180	65,420		1000	
	80,000	65,072		992	8
	75,000	55,413		768	232
	70,000	45,754		543	457
	65,000	37,973	109	321	570
	60,000	31,185	275	100	626
	55,000	26,794	333		620
Pooled	[2/] 80,180	661,928		1000	
	80,000	658,785	8	992	
	75,000	583,775	229	771	
	70,000	528,126	384	550	67
	65,000	489,817	370	463	114

1/ Solutions presented are for the LP solution at least one full-cropland use solution, and the first interval solution in which land is slack.

2/ LP solution.

decreases in negative deviations per dollar decline in expected income. Conversely, farms 2 and 8 demonstrate little decline in variability per dollar decline in expected income. These phenomena are consistent with the correlation coefficients in Table 3 under the circumstance when soybeans diversifies with the next entering crop. For example, farm 8 first diversifies with corn moving away from the LP solution. The correlation coefficient between corn and soybeans for farm 8 is .931, hence little diversification potential exists. The settings of average yields and pooled yields show intermediate impacts of diversification but different organizational patterns in diversifying.

Fig. 1 brings together the risk programming results into a framework which allows direct comparison. Organizations for individual frontier points are forced through the income-deviation matrix for the pooled setting. The result shows how organizations selected from each individual frontier differ in relation to each other and the base (the pooled frontier). Any setting could be used as the comparison setting, however, here the pooled setting was selected at the most representation of the risk-income relationships for the group. The nine non-pooled frontiers may cross each other. This is because the organization for an individual frontier may lie further from the base (pooled) organization at one point along the map than at another point on the map. The distance from the pooled frontier for any income level represents the level of non-optimality. For example, at $60,000 income farm 1's organization is very close in organization (hence income and deviations) to the pooled frontier, farms 2, 3, 5, 6, and 8 are further away, and farms 4, 7, and the average yield farm are very different from the pooled frontier.

The reason for departures of individual frontiers from the pooled frontier lies in yield, standard deviation of yield, and covariance relationship differences among farms. Farms 2, 5, 6, and 8 were found previously to have a similar diversification pattern. These are seen here to perform comparably, the only difference is that farm 6 has lower levels of deviations for given income compared to farm 2, 5, and 8. The reason for this is the negative correlation coefficient for corn and soybeans for farm 6 (-.367 from Table 3).

It should be noted that the frontier based on average yields results in a frontier widely different not only from the pooled frontier, but also from the frontiers for farms 1, 2, 3, 5, 6, and 8. Again, this leads to the rejection of average-yield risk studies in that frontiers based on average yields can depart widely from those farms used in forming the averages.

Assessment of Pooled Yield Data

Because of the very high degree of instability of yield relationships between farms, the use of pooled yield data appears to be useful in risk analyses. This data treatment is judged to be useful both when general area risk studies are undertaken as well as individual farm studies. For individual farms, benefit is gained by adding yield observations from other farms. Clearly, the use of average yields cannot be recommended in crop yield risk analysis. Further, because of yield instability among farms, little confidence can be placed in yield series from individual farm as representative for an area or for analyses of that individual farm.

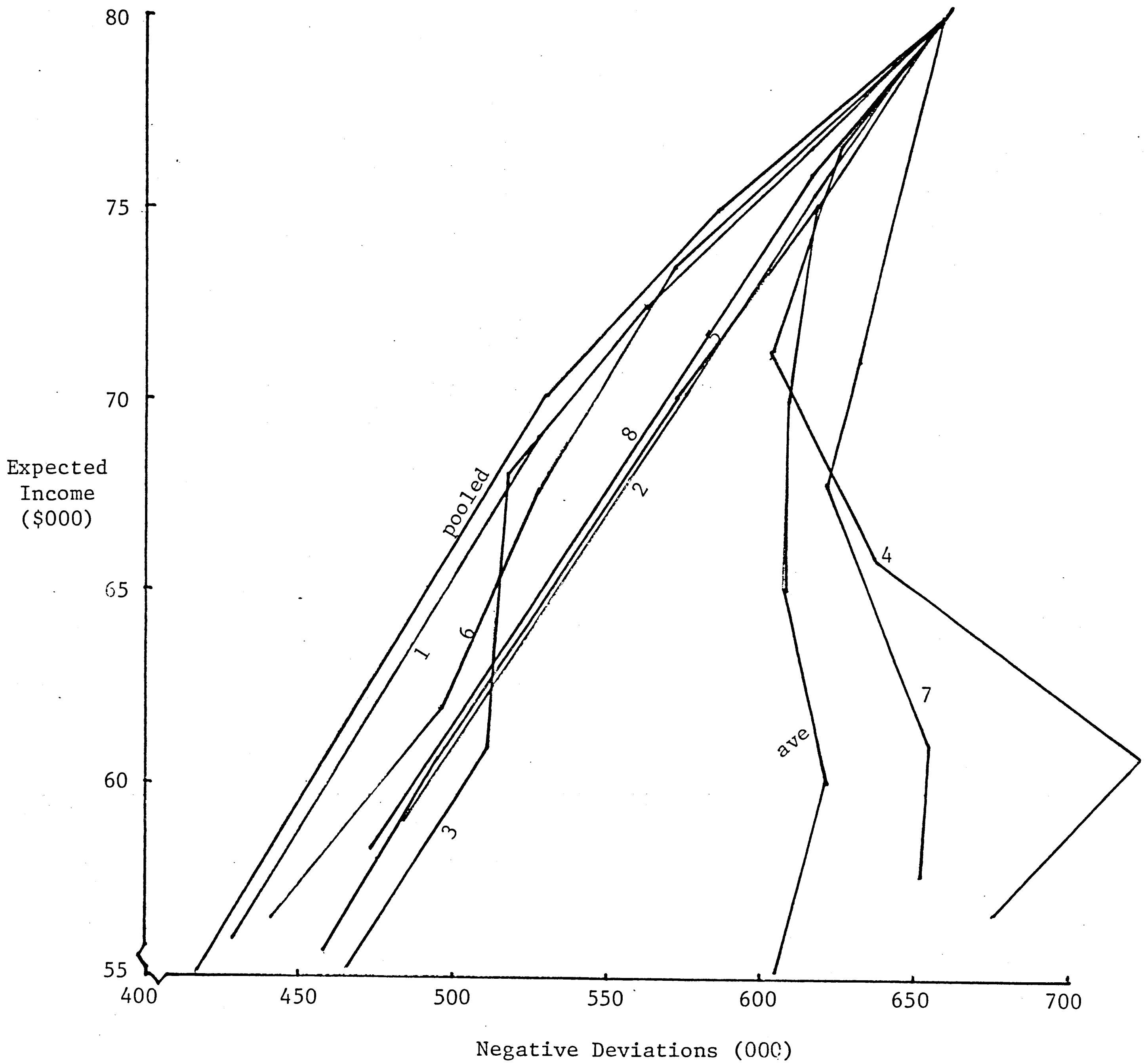

Fig. 1. Direct Comparison of Ten MOTAD Risk-Income Frontiers Using Pooled Yield Data As Comparison Base.

References

1. Eisgruber, L.M. and L.S. Schuman. "The Usefulness of Aggregate Data in the Analysis of Farm Income Variability and Resource Allocation." J. Farm Econ. 45(1963): 587-591.

2. Tauer, L.W. "Target MOTAD." Am. J. Agr. Econ. 65(1983): 606-610.

3. Watts, M.J., L.J. Held, and G.A. Helmers. "A Comparison of MOTAD to Target-MOTAD." Can. J. Agr. Econ. 32(1984): 175-186.