[QUA]NTIFYING LONG RUN AGRICULTURAL RISKS AND EVALUATING FARMER RESPONSES TO RISK

Proceedings of a Seminar sponsored by
Southern Regional Project S-232
"Quantifying Long Run Agricultural Risks and Evaluating Farmer Reponses to Risk"
Orlando, Florida
March 22-25, 1992

Department of Agricultural and Resource Economics
The University of Arizona
Tucson, AZ 85721

September 1992

DOES FARM SIZE MATTER IN WHEAT YIELD VARIABILITY? A PROPOSED APPROACH

by

Michele Marra*
University of Maine
Orono, ME 04469

and

Bryan Schurle
Kansas State University
Manhattan, KS 66506

*Senior authorship is shared. Comments encouraged.

Contribution number 92-558-A from the Kansas Agricultural Experiment Station, Kansas State University, Manhattan, KS 66506.

DOES FARM SIZE MATTER IN WHEAT YIELD VARIABILITY?
A PROPOSED APPROACH

Introduction

Over the past 25 years, there have been many developments in the theory of individual producer behavior under uncertainty. Beginning with the works of Baron, Sandmo and Holthausen and continuing with more recent papers by Antle, Just and Zilberman, and Meyer, to name a few, our understanding of the important theoretical aspects of risky decisionmaking has made significant progress.

Several problems remain, however in moving toward empirical implementation of this work. Some of these relate to the measurement of risk attitudes, while others are concerned with the definition and measurement of the risk, itself. One important issue related to the latter is the general lack of sufficient data needed to measure the yield risk faced by individual producers. Collection of these data at the farm level is expensive, while yield data at more aggregated levels are readily available from USDA Crop Reporting Service Series and, now, from the S-232 dataset.

The purpose of this paper is to investigate whether the more aggregated data series can be adjusted in a systematic way to reflect the yield risk faced by producers at the farm level. Without compiling new data at the farm level for each application, there are two ways to approach this investigation using existing data. First, one could attempt to describe what has been reported in the literature about yield variability at various levels of aggregation in the conventional way. The results of this might be a table of results and a verbal description of them that might include some summary statistics. The second way to attempt to discover an adjustment parameter is to perform a meta-analysis of the existing information.

Meta-Analysis

A meta-analysis is, essentially, an analysis of analyses. It is an attempt to cumulate research findings in a more formal, statistical way so that, if there is some systematic, underlying "weight of evidence" in the research to date, it is more likely to be discovered. Meta-analysis can be performed across a number of studies, on multiple findings within one study, or both at once (Hunter, Schmidt and Jackson). It uses any one of a number of standard statistical procedures, including regression, to summarize the cumulative meaning of the results of past work on a particular subject. The first meta-analyses were performed in the areas of medicine and psychology and were generally concerned with cumulating correlations across a group of experiments performed by different researchers on the same subject (Glass, McGraw and Smith). It has been used to cumulate the research findings in such diverse areas as treatment of migraine and tension headaches (Blanchard, et al.) to teaching style and pupil achievement (Cohen).

The two major questions posed by meta-analysts are: 1) Is the effect of factor X on outcome Y significant? and 2) What is the size of the effect of factor X on outcome Y? Answering these questions through a descriptive review of existing literature can lead to startling errors. Hunter, Schmidt and Jackson describe an experiment they conducted in which a group of study outcomes was generated from an underlying distribution. The outcomes were then presented to several researchers in tabular form, and they were asked to summarize the study results. None of the researchers came close to the true mean effect, and some concluded that several factors contributed significantly to the results that were, in fact, randomly assigned to each study result! However, descriptive literature review is still the most popular way to summarize research findings today. Meta-analysis, while still a controversial approach, seems to hold promise for the cumulation of research results.

One method of meta-analysis, which can answer both of the above questions simultaneously (and is surely the method most familiar to economists) is least squares regression of the study outcomes on various characteristics of the studies, such as study location, time, type of subject (students, general public, hospital patients, etc.), published or unpublished work. Meta-analysis using regression techniques has been used in marketing research to analyze differences in consumer response to external stimuli, such as price, advertising, etc.

(Farley and Lehmann). More recently, Smith and Kaoru used it to cumulate the findings of the numerous studies of user benefits from recreation sites that employed the travel cost method of estimating value. One of their stated purposes was to determine if the current practice of adjusting the results of one or more existing studies and using them to value a particular resource that has not been studied (called benefits transfer) is valid. Their method was to regress the real consumer surplus per unit of use on several characteristics of the recreation site studies, several behavioral assumptions, such as how the opportunity cost of time is handled in the study, and several researcher judgements, such as functional form or estimator used. They found that many factors under the researchers' control, in addition to site characteristics, significantly affected the consumer surplus measure. They, therefore, conclude that caution should be used when transferring benefits from existing studies to another recreation site.

Our study is similar in purpose to the Smith and Kaoru study. We are investigating the potential existence of an adjustment parameter that could be used to adjust aggregated yield variability information to reflect the variability faced by farmers where farm level data are not available. In this initial analysis, we perform a meta-analysis of within-study results to avoid potential statistical problems of cumulating over time and space and to eliminate across study effects so that we can concentrate on the question of the existence of an adjustment parameter. It is a meta-analysis in the sense that the variability measures at various levels of aggregation are statistically cumulated, rather than just presented tabularly and discussed.

Wheat Yield Data

As a preliminary effort, the analysis was limited to dryland wheat in Kansas. Data from the Kansas Farm Management Association were organized for analysis. Only farms which had grown wheat for 16 consecutive years (1973 to 1988) were selected. The farm management data contains information on rented wheat acres and production, owned wheat acres and production and total wheat acres and production. After sorting farms, 339 farms had a complete series of wheat production. Of these, 171 had a complete series of wheat on owned acres, and 221 had a complete series of wheat on rented acres. Some farms had complete series on rented, owned and then the total acreage. We included in the analysis every complete series whether rented, owned or total.

Several methods of detrending the data were explored. Since farm yield variability can be substantial, particularly over a short period of time, detrending methods must be carefully considered so that over-fitting does not occur. Over-fitting would remove more of the variability than is logically acceptable.

Three measures of variability were calculated. First, standard deviations were calculated after no trend removal. Table 1 shows the mean of the standard deviations calculated without trend removal for three categories of wheat. For dryland wheat, the means of the standard deviations range from 8.72 to 9.25. These are likely over estimates of the variability since no trend is removed. Second, standard deviations were calculated after removing the linear time trends which best fit each individual farmer's data. Table 2 shows the mean of the standard deviations calculated in this fashion. For dryland wheat, the means of the standard deviations range from 8.36 to 8.87. This measure has the potential for over-fitting the data and thus underestimating the variability of yields. Examination of the trends indicates that the mean of the individual trends for the farms ranged from .25 to .30 bushels per year increase in yield. However, individual trends removed from farm data ranged from -1.07 bushels per year to 1.88 bushels per year. These trends provide some indication of the overfitting that occurs by allowing individual trends to be removed from each farmer's data. Finally, a common trend was removed from all the data from all the farms. Variability around this trend was then measured. Table 3 shows that the means of standard deviation ranged from 8.62 to 9.15 after removal of the common trend. This method could be viewed as a compromise approach given that a common trend is removed. However, this common trend is more than some farmers are experiencing and less than others are experiencing. As indicated by the means of the standard deviation, this method on average provides estimates of variability which fall between the other two methods.

The method finally selected for measurement of variability was the method which removed individual trends from each farm. This can be viewed as an underestimate of the variability due to overfitting each farmers data, but it was judged to not be a major underestimate given that it averaged only .35 less than the measure removing no trend and .26 less than the measures of variability calculated after removing a common trend. While this is an underestimate of the variability, it should alleviate potential criticism when farm variability is compared to more aggregated measures of variability to see to what extent aggregation reduces yield variability. Furthermore, trends need to be removed from more aggregated yield series which tend to be less volatile and a simple linear trend method of removing variability could be consistently used for the different series for the different levels of aggregation.

The main issue now is to investigate the relationship between aggregation level or size and yield variability, all 731 farm observations were included in the analysis, along with 105 counties, 9 regions and the total state observation. Some farms may have contributed three observations if they had a complete series on rented and owned land because one observation would come from the rented acres, one from the owned acres, and one from the combination of the acres. This automatically provides some data on aggregation even at the farm level. Table 4 provides information on the range of acres for the different aggregation levels. The mean of the mean acres per farm ranges from 202 to 408 and the range of the means across all farm observations ranges from 21 to 2,388 acres. The mean acres for counties ranged from 2,600 to 462,000. There was a small gap between the largest farm and the smallest county.

Estimates of county, region, and state yield variability were calculated after removing a linear trend. Table 5 provides information on the yield variability for the different aggregation units.

To explore the relationship between size or aggregation level and variability, a linear relationship was estimated for the farm data, the county data, and the farm and county data combined. Table 6 provides information on the resulting estimates. All size coefficients are highly significant and negative in size suggesting that variability decreases as size increases. The coefficients suggest a curvilinear relationship of some form with a much greater initial decline in variability as size increases and a more gradual decline in variability over the larger size ranges. Since there are two major clusters of data, farm and county, it is useful to examine the shapes of the estimated equations in the ranges of both clusters.

The Adjustment Parameter

Several functions were estimated using three sets of data, farm data, county data and all the data which includes farm, county, region and state. Table 7 provides information on the two forms of estimated equations. Quadratic equations and an equation with $1/acre^{.5}$ were also estimated but not reported since they did not fit as well. Efforts were made to visually identify equations which fit best by comparing equations estimated from the cluster of farm data with equations estimated using all data. Figure 1 shows the resulting equations which were estimated. Obviously, the best fits of the farm data are those estimated only using farm data. Trying to fit both farm and the rest of the data with the same equation results in a poorer fit of the farm data. Based on visual examination, the log models appeared to work well.

Similar comparisons were made using only the county data. Figure 2 illustrates equations estimated with county data and with all data. In addition, Table 7 provides information on the estimates. It appears again that the log models work fairly well to fit the data. The R^2 for the log model, adjusted to compare with the linear and reciprocal model R^2's is .14.

The implied adjustment parameter derived from the log model and calculated at the weighted mean of the county and farm level acreage is .000034. Using this parameter, the average county level variability in this example must be increased by 2.82 (or about 42%) to reflect the average farm level variability. It seems, then, at least for the case of Kansas dryland wheat over the last sixteen years, that the parameter results in a smaller difference between county level and farm level yield variability smaller than the two- to three-fold difference commonly assumed in the literature (see, for example, Marra and Carlson).

Conclusions and Further Work

The important result of this study is that a significant, non linear, relationship was found between wheat yield variability and the acreage over which the variability measure was calculated. This shows some promise for establishing adjustment parameters for moving from aggregate data to the farm level, although it is clearly a preliminary, first step.

Additional analyses of a broader range of data should be conducted in order to identify the generality of the variability/size relationship apparent in this study of wheat yield variability. The next step would be to see if the relationship for wheat is stable across states and over time using other existing county level and farm level datasets and by performing out-of-sample prediction error tests. If it is, then adjustments could be made from the county level data to reflect variability at the farm level with some confidence. Then, the approach could be tried with other crops. If successful, individual decision models under uncertainty could find many more empirical applications.

REFERENCES

Antle, J. 1987. Econometric estimation of producers' risk attitudes. Amer. J. of Ag. Econ. 69:509-22.

Baron, D. P. 1970. Price uncertainty, utility, and industry equilibrium in pure competition. Int. Econ. Rev. 3:463-479.

Blanchard, E. B., F. Andrasik, T.A. Ahles, S.J. Teders, and D. O'Keefe. 1980. Migraine and tension headache: A meta-analytic review. Albany: Headache Project, State Univ. of New York.

Cohen, P.A. 1980. A meta-analysis of the relationship between student ratings of instruction and student achievement. Unpublished PhD. dissertation, University of Michigan, Ann Arbor.

Farley, J. U. and D. R. Lehmann. 1986. Meta-Analysis in Marketing. Lexington Books: Lexington, MA.

Glass, G.V., B. McGaw, and M.L. Smith. 1981. Meta-Analysis in Social Research. Sage Publications: Beverly Hills.

Hunter, J.E. F. L. Schmidt, and G. B. Jackson. 1982. Meta-Analysis: Cumulating Research Findings Across Studies. Sage Publications: Beverly Hills.

Holthausen, D.M.1979. Hedging and the competitive firm under price uncertainty. Am. Econ. Rev. 69:989-95.

Just, R. and D. Zilberman. 1983. Stochastic structure, farm size and technology adoption in developing agriculture. Oxford Economic Papers. 35:307-28.

Marra, Michele C. and G.A. Carlson. 1990. "The decision to double-crop wheat and soybeans: An application of expected utility theory using Steins's theorem. Am. J. Ag. Econ. 72:337-345.

Meyer, J. 1987. Two-moment decision models and expected utility maximization. Amer. Econ. Rev. 77:421-30.

Sandmo,A. 1971. On the theory of the competitive firm under price uncertainty. Am. Econ. Rev. 61:65-73.

Smith, V.K. and Y. Kaoru. 1990. Signals or noise? Explaining the variation in recreation benefit estimates. Am. J. of Agric. Econ. May: 419-33.

Table 1. Farm Level Wheat Yield Variability Measures Calculated Without Removing Trend

Type	Owned Dryland Wheat	Rented Dryland Wheat	Total Dryland Wheat
# of farms	171	221	339
Mean of Means	35.20	34.19	34.42
Mean of Stand. Dev.	9.25	8.73	8.72
Minimum Stand. Dev.	3.55	4.62	3.20
Maximum Stand. Dev.	18.38	25.81	19.38
Mean of Coeff. of Var.	26.68	26.04	25.77

Table 2. Farm Level Wheat Yield Variability Measures Calculated after Individual Trends were Removed

Type of farm	Owned Dryland Wheat	Rented Dryland Wheat	Total Dryland Wheat
# of farms	171	221	339
Mean of St.Dev. of residuals	8.87	8.41	8.36
Minimum of Stand. Dev. of Residuals	3.21	4.25	3.15
Maximum of Stand. Dev. of Residuals	17.61	25.51	19.89
Mean of individ. trends removed	.30	.25	.28
Min. Ind. Trend	-1.00	-.96	-1.07
Max. Ind. Trend	1.83	1.68	1.88

Table 3. Farm Level Wheat Yield Variability Measures Calculated After Removing a Common Trend

Type	Owned Dryland Wheat	Rented Dryland Wheat	Total Dryland Wheat
# of farms	171	221	339
Mean of Stand. Dev. of Residuals	9.15	8.66	8.62
Min. of Stand. Dev. of Residuals	3.21	4.60	3.25
Max. of Stand. Dev. of Residuals	18.03	26.02	19.90
Common Trend	.30	.25	.28

Table 4. Acres of Wheat Production on Farms, Counties, Region and State.

	Mean Owned Wheat Acres per Farm	Mean Rented Wheat Acres per Farm	Mean Total Wheat Acres per Farm	Mean Wheat Acres per County	Mean Wheat Acres per Region	Mean Wheat Acres in Kansas
Mean	202	323	408	112,858	1,316,674	11,759,400
Min.	21	35	30	2,600	383,233	-
Max.	1,052	1,359	2,388	462,000	2,747,813	-

Table 5. Yield Variability Estimates for Different Aggregated Areas.

	County	Region	State
# of observations	105	9	1
Mean of standard deviation	6.63	5.53	3.88
Minimum standard deviation	4.13	3.59	-
Maximum standard deviation	9.75	6.55	-

Table 6. Linear Estimate of Relationship Between Standard Deviation of Detrended Yield and Size of Unit.

	Farm data	County data	Farm and county data
# of observations	731	105	836
Intercept	9.109	7.543	8.459
Size Coefficient	-.001838	-.000008114	-.00001369
t-statistic	-6.490	-5.085	-8.435
Significance level	.0001	.0001	.0001
R^2	.05	.20	.08

Table 7. Information on estimates of equation parameters derived from farm, county, and all the data combined.

	Reciprocal Equation			Log Equation		
	Farm Data	County Data	All Data	Farm Data	County Data	All Data
Intercept	7.87	6.49	7.59	2.63	3.06	2.387
T-statistic	68.30	42.70	76.50	44.04	12.23	96.72
Coefficient for Size	103.72	6821.64	127.61	-.096	-.105	-.0501
T-statistic	7.59	1.95	10.14	-8.89	-4.766	-13.621
R^2	.07	.03	.11	.10	.17	.18

FIGURE 1. EQUATIONS FITTING

FARM DATA

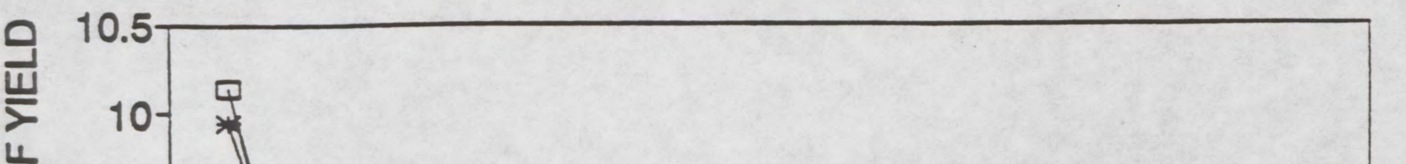

FIGURE 2. EQUATIONS FITTING

COUNTY DATA