Staff Papers Series

P82-23

November 1982

MEASURING THE RETURNS TO EXPERIMENT STATION RESEARCH

FOR CORN, SOYBEANS AND WHEAT - AN APPLICATION OF

RIDGE REGRESSION

BY

SHELLEY HENDRICKSON AND W. BURT SUNDQUIST

Department of Agricultural and Applied Economics

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

MEASURING THE RETURNS TO EXPERIMENT STATION RESEARCH FOR CORN, SOYBEANS AND WHEAT - AN APPLICATION OF RIDGE REGRESSION

BY

SHELLEY HENDRICKSON AND W. BURT SUNDQUIST*

November 1982

University of Minnesota

* Graduate Research Assistant and Professor, respectively, Department of Agricultural and Applied Economics, University of Minnesota. Earlier contributions to this research by Cheng G. Cheng and George Norton are acknowledged.

Measuring the Returns to Experiment Station Research for Corn, Soybeans and Wheat - An Application of Ridge Regression*

Shelley Hendrickson and W. Burt Sundquist**

Introduction

A major body of literature reports the research which has been conducted to evaluate the returns to agricultural research. A bibliography of publications which reports much of this work is presented by Norton and Davis (1981) in their recent review paper. Several major research evaluation studies have utilized a production function approach (cross-sectional or time series) and estimated research returns for aggregate agricultural production or for groups of agricultural commodities. Bredahl and Peterson (1976) conducted extensive cross-sectional analysis to measure ex post returns to agricultural research in the U.S. for commodity groupings which included cash crops, poultry, dairy, and other livestock. They used published data series on input categories including land, machinery, labor, fertilizer, chemicals, seed, feed and pasture from the 1969 U.S. Census of Agriculture and data on state research expenditures from the USDA annual Inventory of Agricultural Research (CRIS). Davis (1979)

---

* This report is a contribution to Minnesota Agricultural Experiment Station Project 14-038 and to IR-6, "National and Regional Research Planning, Evaluation, Analysis and Coordination".

** Graduate Research Assistant and Professor, respectively, Department of Agricultural and Applied Economics, University of Minnesota. Earlier contributions to this research by Cheng G. Cheng and George Norton are acknowledged.

evaluated the stability of the research coefficient over time in production function-type analyses. Norton (1981) updated the Bredahl-Peterson work using 1974 census data and also examined the temporal stability of the research coefficient for the several commodity groups. All of the above cited analyses used state-level expenditures as their research variable but used a "representative farm" concept to normalize other input variables.

The above cited analyses and others show returns to agricultural research in the U.S. to be high, historically, and allocated fairly efficiently among aggregated commodity groups when returns are computed as a marginal internal rate of return to research investment.

Thus a wealth of information on returns to agricultural experiment station research expenditures exists for aggregate agriculture and for commodity groups of major commercial importance in the U.S. But many of the investments made in agricultural research, and, consequently, many of the allocative decisions on research funding, are commodity-specific. Other research expenditures are specific to such areas as plant breeding, pest control, mechanization, etc. It would, as a result, be highly desirable to obtain expanded perspective on and measurement of the returns to research expenditures made for individual agricultural commodities and for individual research areas. Since a major potential use of these estimates is to project the pay-off for future research expenditures, it would also be desirable to have such estimates for the most recent time period(s) possible. Among the commodities of particular interest are corn, soybeans and wheat for which very substantial research investments are being made by the State Agricultural Experiment Station.

Also of interest are other major crops such as grain sorghum, rice and cotton and some "newer" crops with potential economic importance such as sunflower.

Commodity-Specific Research Evaluations

There are a number of examples of commodity-specific research evaluations in the literature. Griliches' (1958) work on the impact of hybrid seed corn represents a pioneering work on commodity-specific technology evaluation. Sim and Araji (1981) measured the economic impact of wheat research in the Western Region of the U.S. And, there are others. Most past commodity-specific studies have employed an index number and consumer-producer surplus approach to measuring benefits and have calculated an average rate of return based on an estimate of the value of inputs saved, past supply shifts, or scientists' estimates of future productivity increases resulting from additional research funding. More recently, Otto and Havlicek (1981) used a combination of cross sectional and time series data to estimate individual supply response functions for corn, wheat, soybeans and sorghum. Working from these response functions, they then estimated annual internal rates of return both for research expenditures made within individual states and for expenditures made in nearby states which might be expected to have "spill in" effects.

In general, a cross sectional production function approach of the type used by Bredahl and Peterson (1976) and Norton (1981) appears to be a preferred procedure when one is interested in estimating a marginal rate of return to research while holding fertilizer, land, labor and/or other inputs constant. Marginal internal rates of return to research investments for individual commodities can then be estimated and the implications for efficiency of allocation of research funds examined. But, a major problem in using this approach centers on the difficulty in obtaining commodity-specific measures of the several input variables. Even research expenditures pose some problems since much research is multi-purpose. Data for commodity-specific input variables are scarce because certain inputs such as machinery and labor are typically employed on several enterprises on the same farm. For example, many farms raise both soybeans and corn, and often other crops as well, and there is limited information on how machinery and labor use is split between these crops.

Since the U.S. Census of Agriculture does not report production input categories for specific agricultural commodities, we undertook to find alternative data sources from which to develop commodity-specific production function formulations using individual states as observations. One source of such data is that provided in the nationwide set of farm enterprise budgets developed by Krenz and others in the National Economics Division, ERS, USDA. These so-called "FEDS" budget data have been developed annually since 1974 for all production areas in the U.S. for which the several major farm commodities are produced commercially. The FEDS enterprise budget data are developed drawing heavily on survey data for each major substate production area, though in some cases such areas are specified as an entire state. We have weighted and aggregated these enterprise data for 1979 so as to develop commodity-specific input totals for each state included in our analysis.

While the FEDS budgets are readily available as a data source, they are not without shortcomings for production function analysis. First, though they provide rather reliable survey based estimates for deriving state aggregate input totals, there is no valid procedure for normalizing them on a "per farm" basis. Second, several input categories have a high degree of multicollinearity because budgets are typically based on a fairly similar complement of machinery and a fairly common set of production practices. Moreover, the machinery and labor input categories are highly correlated with land because each acre of land used for production of a specific crop has a similar package of machinery and labor inputs applied to it. These high intercorrelations among input variables produce unstable (and statistically unreliable) regression coefficients. It is for this reason, primarily, that we have used the procedure of ridge regression analysis to derive a set of regression estimates with stable coefficients.

As a practical aside, much of the per acre variance in input categories between state subregions, states and multi-state regions has disappeared over time as commercial farmers have developed farming operations which are highly mechanized and fairly standardized using the same chemical fertilizers, pesticides, etc. And, very little unemployed or underemployed labor remains on U.S. farms. This suggests that much of the earlier day variance between production areas in the use of at least some farm inputs has vanished as the shift to modernized production methods on commercial farms in the U.S. has become virtually complete. And, this phenomenon will be reflected in the farm input data whatever their source.[1/]

1/ For example, the following simple correlations exist between states for land, labor and machinery inputs in the 1969 and 1974 Agricultural Census data:

1969	Machinery	Land	1974	Machinery	Land
Land	0.77		Land	0.90	
Labor	0.46	0.80	Labor	0.73	0.61

The research variable included in our production functions for 1979 is the 1973 research expenditure for each crop from the CRIS data. The annual research expenditures are thus effectively lagged six years for the 1979 production function estimates.[2/]

Research expenditures for corn averaged $346 thousand with a range of more than $1 million for the 27 states included in the analysis. Iowa, Illinois, and Nebraska are the top three states in corn research expenditures. Illinois, Iowa, Minnesota and Missouri are the top soybean research states. Soybean research expenditures averaged $225 thousand for the 26 states included in the analysis with a range of $606 thousand. Kansas and North Dakota are the top wheat research states in terms of expenditures. The mean wheat research expenditure is $183 thousand for the 34 states included in the analysis, with a range of $841 thousand. Because the benefits of research are not neatly contained within the boundaries of the states in which the research is performed, we undertake later to measure the spillover effects of research conducted in other states.

2/ Actually, of course, the returns which we estimated and attributed to these annual research expenditures are predicated in part on prior-period based knowledge. Thus, one could not expect a similar flow of returns to annual research expenditures in the absence of these prior-year investments. And, it is important to retain the perspective of a stock of research capital that is being serviced (and to some extent, augmented) by a set of annual expenditures. In our analysis, it is to variance among states in these annual research expenditures that variance in the value of output is being associated.

Initial aggregate production functions were specified for each crop for 1979 using individual state aggregates as observations in order to estimate total crop values as a function of land, rainfall, fertilizer, chemicals, labor, machinery expenses, and research expenditures.[3/]

Additional slope dummy variables were then included to account for any major differences in land quality or general climatic influences. States included in the dummy variables for each crop are listed in Table 1. We found that the land quality for corn was significantly lower in the Eastern and Delta states (slope 1 group) than in the Central regions. Similarly, the land was of lower quality for soybeans in the Eastern and Southeastern states (slope groups 1 and 2). For wheat, the Eastern and Delta states (slope groups 1 and 2) were significantly lower while Colorado, Kansas, Oklahoma, and Texas (slope group 3) were significantly higher in quality than the remaining states.

The variables included in the final regression equation for each commodity were chosen based on the results of a ridge trace, which is a plot of all the initial ridge regression coefficients against the ridge bias parameter k, where k is plotted on the x-axis with values ranging from 0.0 to 1.0. As k increases, the expected bias of the estimates increase. It is therefore desirable to choose the lowest k at which the estimates are stable. Variables were eliminated using certain rules outlined by Hoerl and Kennard (1970), i.e., eliminate variables whose coefficients are stable but small, and eliminate those variables with unstable coefficients. Using this procedure, certain slope dummy variables as well as the rainfall variable for all three crops

---

3/ Sources or input and output data are listed in Appendix 1.

were eliminated due to their insignificant coefficients. The k parameters chosen for the final regressions (the value at which included estimates became stable) were .5 for corn, .5 for soybeans and .4 for wheat. The resulting regression equations included the above mentioned statistically significant slope dummies for each crop, and variables for the logarithms of land (in acres), labor (in hours) and the logarithms of dollar expenditures on research, chemicals, machinery, fertilizer and research spill in.

The three production functions finally selected all yielded equations with high R^2's and statistically significant coefficients for almost all variables. The regression for corn is shown in equation 1:

<u>Corn Equation</u>

$$(1)\quad \log X_1 = \underset{(.77251)}{-5.2487} - \underset{(.06326)}{.22428}\,(DCI) + \underset{(.01398)}{.19289}\log X_2 + \underset{(.02860)}{.06956}\log X_3$$

$$+ \underset{(.01845)}{.18603}\log X_4 + \underset{(.02389)}{.15683}\log X_5 + \underset{(.01778)}{.18724}\log X_6 + \underset{(.01787)}{.19693}\log X_7$$

$$+ \underset{(.03460)}{.12436}\log X_8$$

$$R^2 = .96$$

This equation has 18 degrees of freedom and a standard error of .25696. Included variables are as follows:

X_1 = corn production value in dollars

X_2 = land in acres

X_3 = within state research expenditures in dollars

X_4 = chemical expenditures in dollars

X_5 = machinery expense in dollars

X_6 = fertilizer expenditures in dollars

X_7 = labor expense in dollars

X_8 = research expenditures in neighboring states in dollars (spill in)

DCI = land slope dummy variable 1

The numbers in parantheses are the standard errors of the regression coefficients.

Over 96 percent of the variance in corn output value was associated with variance in the equation variables. Land, labor, fertilizer and chemicals, all of which had t-values well above the 99 percent level of significance, are the most significant variables. DCI and the machinery and spillover variables were significant at the 95 percent level.

The final soybean regression is shown in equation 2:

Soybean Equation

$$\text{(2)} \quad \log X_1 = \underset{(1.0162)}{-3.4418} - \underset{(.09774)}{.38255}\,(DCI) - \underset{(.09859)}{.22739}\,(DC2) + \underset{(.01667)}{.20443} \log X_2$$

$$+ \underset{(.04090)}{.15000} \log X_3 + \underset{(.01880)}{.14633} \log X_4 + \underset{(.01639)}{.18479} \log X_5 + \underset{(.02893)}{.08866} \log X_6$$

$$+ \underset{(.03754)}{.17578} \log X_7 + \underset{(.04512)}{.09219} \log X_8$$

$$R^2 = .93$$

with a standard error of .32717 and 16 degrees of freedom and where variables X_1 to X_8 are comparable to those for equation 1 and where DC1 and DC2 are land slope dummy variables 1 and 2.

Ninety-three percent of the variance in soybean output was associated with variance in the independent variables in equation 2. Land, machinery, chemicals, labor, research, fertilizer and DC1 were all significant at the 99 percent level. DC2 was significant at the 95 percent level while research spill in was significant at the 90 percent level.

Table 1. States Included in Land Dummies

Wheat:	34 States
Slope 1.	KY, MD, NY, NC, PA, TN, VA. = DC1
Slope 2.	AL, AR, GA, MS, SC. = DC2
Slope 3.	CO,KS,OK, TX = DC3
Slope 4.	MT, NB, ND, SD. = DC4
Slope 5.	AZ, CA, ID, NM, OR, WA. = DC5
Land	IL, IN, IA, MI, MN, MO, OH.
Corn:	27 States
Slope 1.	AL, DE, FL, GA, KY, MD, NJ, NY, NC, PA, SC, TN, VA. = DC1
Slope 2.	CO, KS, NB, TX. = DC2
Land	IL, IN, IA, MI, MN, MO, ND, OH, SD, WI.
Soybeans:	26 States
Slope 1.	AL, GA, OK, SC, TN, TX. = DC1
Slope 2.	DE, KY, MD, NJ, NC, VA. = DC2
Slope 3.	AR, LA, MS. = DC3
Land	IL, IN, IA, KS, MI, MN, MO, NB, OH, SD, WI.

Finally, the wheat regression equation is shown as equation 3:

Wheat Equation

$$\text{(3)} \quad \log X_1 = \underset{(.52141)}{-1.9306} - \underset{(.07399)}{.23007}\,(DC1) - \underset{(.08594)}{.27676}\,(DC2)$$

$$+ \underset{(.09667)}{.25833}\,(DC3) + \underset{(.01381)}{.13686} \log X_2 + \underset{(.02191)}{.14177} \log X_3 - \underset{(.01046)}{.04215} \log X_4$$

$$+ \underset{(.01101)}{.15773} \log X_5 + \underset{(.02074)}{.18748} \log X_6 + \underset{(.01267)}{.20439} \log X_7 + \underset{(.01998)}{.02893} \log X_8$$

$$R^2 = .97$$

This equation has a standard error of .25152 and 23 degrees of freedom. Variables X_1 and X_8 are comparable to the same variables for corn and soybeans and DC1, DC2 and DC3 are land slope dummy variables 1 through 3.

A very high 97 percent of the variance in wheat output was associated with the independent variables used in this equation. Labor and machinery again had t-values significant at well over the 99 percent level and land, fertilizer, research, chemicals, DC1 and DC2 were also significant at this level. DC3 was significant at the 95 percent level. Research spill in for wheat was the only variable (in all three regressions) which was not statistically significant at even the 90 percent level, indicating that our crude measure of research spillover probably does not adequately capture a rather complex research spill in phenomenon.

Economic Interpretation of the Regression Results

Marginal Products:

Marginal products for research expenditures were estimated using the following formula:

$$VMP\ X_i = \frac{b_i\ E(Y)}{X_i}$$

Where E(Y) = estimated output at mean input values

X_i = mean input value

This resulted in marginal products of about \$150, \$180 and \$360 respectively for corn, wheat and soybeans. A major expansion in export demand for food grains, feed grains and oil seed crops in the 1970's helped to generate extremely high productivity rates for public research expenditures for these crops. Other basic and non-commodity-specific research also undergirded these commodity-specific research investments. The major conclusion is, however, clear. Big gains in output value for corn, wheat and soybeans in 1979 were associated with earlier period public research investments for these crops.

Research Spillover:

A high portion of the research conducted in an individual State Agricultural Experiment Station has productivity impacts in other states as well. This is particularily true for scientific research, but also for technology-oriented research relating to crop and livestock commodities. Some credible estimates indicate that only about one-third of the productivity from science-oriented research and perhaps up to two-thirds of the productivity from technology-related research is realized within the state undertaking the research.[4/] Our own analysis for corn and soybeans suggests that, in very

---

4/ See, for example, Robert E. Evenson, Paul E. Waggoner, and Vernon W. Ruttan, "Economic Benefits from Research: An Example from Agriculture", Science 205 (September 14, 1979). Recent unpublished analysis by Garren and White also indicates that nationally about two-thirds of the total marginal product from research on cash grains is associated with research within the state where research is done and about one-third from research in other states. They found a smaller portion of spillover, however, for dairy research.

general terms, three-fifths and four-fifths, respectively, of the research related productivity for these crops results from research conducted within the state where utilized and the balance is spilled in from research conducted in other states.

The spillover (spill out and spill in) of research benefits between states is a complex phenomenon and complicates the process of research planning and funding. Moreover, it contributes to a hesitancy by individual states to fund research (1) in the expectation of losing some of the benefits of this research to other states and (2) in the hope that other states might provide the needed research. But, it also points up the importance of research related planning, coordination and communication on an interstate basis if the total pay-off from agricultural research is to be as great as possible.

Rates of Return for Research:

Annual rates of return for research expenditures made for cash grains have, in general, been high in the 1960s and 1970s. Norton (1981) estimated that in-state research expenditures for cash grains in 1968 had a lagged return of about 70 percent in 1974. Otto and Havlicek (1981) using a supply response model and 1967-1977 research expenditures, estimated national-level returns ranging from 81 percent for wheat to more than 175 percent for corn and soybeans. They calculated even higher rates for research investments in the North Central Region. Our estimated rates are generally still higher at 207 percent for wheat, 155 percent for corn and 285 percent for soybeans.[5/]

It is our judgement that considerable caution should be excercised in interpreting these and other estimates of rates-of-return.[6/] Individually the estimates are crude and subject to considerable error. Moreover, cross sectional estimates reflect the unique price and weather conditions for a single year. Collectively, however, they provide strong evidence that: (1) rates of return are high for agricultural research generally and (2) they are particularily high for research on those "large-volume" agricultural commodities over which research benefits can be broadly spread and particularily if those commodities are faced with substantial growth in demand. Corn, soybeans and wheat are commodities which enjoyed both attributes in the 1970s.

5/ Following Norton (1981), internal rates of return were estimated by the formula:

$$MPR \sum_{i=1}^{n} = \left[\frac{Wi}{(1+r)^{i}} \right] - 1 = 0$$

where $W_i = \dfrac{2i-1}{2s^2}$ for i = 1 to s and

$$W_i - \frac{2n-(2i-1)}{2s^2} \quad \text{for } i = s + 1 \text{ to } n$$

n = total number of years over which past research has an impact on research

$s = \frac{n}{2}$ = mean lag

r = marginal internal rate of return

MPR = marginal product of research discounted by two-thirds to account for the contribution of extension and private research.

6/ For a comprehensive listing of previously estimated rates of return see Evenson, Ruttan and Waggoner (1979).

Selected References

Bredahl, M., and W. Peterson. "The Productivity and Allocation of Research: v.s. Agricultural Experiment Stations", American Journal of Agricultural Economics, 58 (1976): 684-692.

Davis, J. Stability of the Research Production Coefficient for U.S. Agriculture, Unpublished Ph.D. Dissertation (1979), University of Minnesota.

Evenson, R.E., V.W. Ruttan and P.E. Waggoner (1979). "Economic Benefits from Research: An Example from Agriculture". Science 205: 1101-1107.

Griliches, W. "Research Costs and Social Returns: Hybrid Corn and Related Innovations", Journal of Political Economy, Vol. 66, October 1958: 419-431.

Hoerl, A.E. and R.W. Kennard. "Ridge Regression: Biased Estimation for Non-othogonal Problems", Technometrics, Vol. 12, February 1970: 55-67.

Norton, G. "Productivity of Agricultural Research", North Central Journal of Agricultural Economics, Vol. 3, January 1981: 1-12.

Norton, G. and J. Davis. "Evaluating Returns to Agricultural Research", American Journal of Agricultural Economics, 63 (1981): 685-699.

Otto, Daniel M. and Joseph Havlicek, Jr. "An Economic Assessment of Research and Extension Investments in Corn, Wheat, Soybeans and Sorghum, SP-81-8, Department of Agricultural Economics, VPI, July, 1981.

Sim, R. and A. Araji. "The Economic Impact of Public Investment in Wheat Research in the Western Region", Research Bulletin 116, University of Idaho, Agricultural Experiment Station, June 1981.

Appendix 1: Data Sources

1979 Crop Production, Annual Summary, Crop Reporting Board, ESCS, USDA, January, 1980.

Costs of Producing Selected Crops in the United States - 1977, 1978 and Projections for 1979, ESCS, USDA, June 1979.

Inventory of Agricultural Research FY 1973, Vol. 2, Science and Education Staff.

Agricultural Prices, Annual Summary 1979, Crop Reporting Board, ESCS, USDA, June 1980.

Federal Enterprise Data System, 1979 budgets, ESCS, USDA.

Crop Values 1978-1979-1980 Season Average Prices Received by Farmers and Value of Production, Crop Repating Board, ESCS, USDA, January 1981.