

# Staff Papers Series

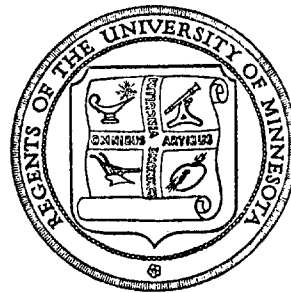
P83-8

February 1983

THE CONTRIBUTION OF RESEARCH TO U.S. SOYBEAN YIELDS

by

Alan G. Miner



**Department of Agricultural and Applied Economics**

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

THE CONTRIBUTION OF RESEARCH TO U.S. SOYBEAN YIELDS

by

Alan G. Miner

February 20, 1983

The study which led to this article was supported by the Minnesota Agricultural Experiment Station. I thank Vernon Ruttan, Burt Sundquist, Jean Lambert, Willis Peterson, Vernon Eidman, and Harald Jensen for their valuable comments.

## THE CONTRIBUTION OF RESEARCH TO U.S. SOYBEAN YIELDS

Recent economic events have refocused attention on agricultural research, productivity and technical change. Tighter state and federal budgets, increased export demand, forecasts of increased world food demand, and energy price increases have brought into question the direction and durability of agricultural productivity change.

Increasing agricultural productivity requires innovation brought about by research and development. There have been some indications that U.S. agricultural productivity growth has been lagging behind that of other countries. Some scientists have suggested that productivity limits are being approached. The productivity question has been approached in many different ways. This study differs from other agricultural and economic productivity studies in several respects. (1) A complete model including weather, prices and technology is specified. (2) It uses the indirect production form where factor demands replace input quantities, leading to a production function in terms of prices and other parameters. (3) The model is derived for and estimated with disaggregated yield data, and is crop specific for both production and research. (See Appendix A for the derivation of the yield model.)

This study addresses several questions. (1) Is a soybean yield limit being reached? (2) What is the rate of return to soybean research, and is the research effect stable over time? (3) Are technology spill-overs between states important? (4) Do input and output price fluctuations affect soybean yields? (5) What types of weather affect soybean yields?

### THE YIELD FUNCTION

Yield is a partial productivity measure in and of itself while total production is highly sensitive to the land input. Yield serves as a common denominator which is useful when comparing different locations. Ease of comparison and interpretation are also important when comparing farm production and experiment station production where the land area devoted to a crop is drastically different.

If a representative farmer maximizes expected profit over multiple outputs (corn and soybeans) and inputs, given uncertain output prices, and if the production function is well behaved, then one can derive factor demands in terms of the parameters which the farmer faces. These parameters are input and expected output prices, and technology. One can then substitute these factor demands into the yield (production) function to create a yield function in terms of prices and technology (Takayama, 1974, and Groenewegen 1980) (see Appendix A). To complete the stochastic specification, one needs to add weather and random effects.

#### a) PRICES

The farmer's expected price cannot be observed. Rather than assume that the expected price is some arbitrary weighted average of past prices, one can more fully use the information available to farmers at the time of their production decision. Each producer may not have complete information at hand, but he does have nearly costless access to the futures prices which embody information from both sides of the

market <sup>1</sup> -/ .

In this study factor prices are assumed to be known with certainty, and to be fixed to the producer. At the time he makes his production decision, the farmer knows the factor prices he faces and can obtain virtually any quantity at that price.

b) TECHNICAL CHANGE

Evenson (1968) introduced technical change directly into the production function by arguing that:

$$(1.1) Q = f(X_i, K)$$

where Q is output, X<sub>i</sub> are inputs and K is the output of a research unit relevant to the experiment station; and

$$(1.2) K = g(R)$$

where R is research inputs (expenditures). Then

$$(1.3) Q = f(X_i, g(R))$$

and one can use the research expenditures to represent technical change in the production function <sup>2</sup> -/ .

Then one can write the model generally as:

$$(1.4) Y = f(EP_s, EP_c, P_i, g(R)) + \text{weather} + \text{error}$$

where Y is yield, EP is the expected price, s is soybeans, c is corn, P<sub>i</sub> are input prices, and g(R) is technical change or the research

component.

In this study research is represented by the constant dollar (1967) levels of expenditures directly related to soybean research <sup>3</sup> / .

Until recently virtually all soybean breeding research has been at public experiment stations. Thus station expenditures should be a good measure of that research effort on soybeans.

Research investments may take many years to come into fruition. Many economic researchers have documented the length of time between the initiation of research and the impact on productivity (Griliches, 1958; Evenson, 1968; Cline, 1975; Evenson, 1978; Davis, 1979; Lu and Quance, 1979; Lu, Cline and Quance, 1979; Davis and Peterson, 1981; and Norton, 1981). In general these lags extend from about two to eighteen years. In this study a total lag of up to 17 years is allowed. A second order polynomial lag structure is used to allow "curvature" of the lagged coefficients <sup>4</sup> / .

The question of spillovers of research effects is addressed by testing for differences in effectiveness between research carried out within the state and research carried out in states lying in the same latitudes. Spill-ins from broader areas are not considered for reasons of simplicity and manageability. These spillovers between areas which grow dissimilar varieties under dissimilar conditions are likely to be much less important than the other spillovers.

Private research and development measures are not available directly. One measure of private research results which is sporadically

available is row spacing. Higher yields associated with planting in narrower rows have been largely the result of more effective chemical weed control. Narrow row spacing could represent a response to private R and D levels which have concentrated in areas other than varietal improvement, such as machinery and herbicides. (see note 3)

c) WEATHER

Further information on the treatment of weather in the models can be found in a separate article I am developing for the Journal of Agricultural Meteorology. I chose to incorporate weather measures such as precipitation and temperature directly into the models. Other methods had flaws which prevented their use <sup>5</sup> / .

RESULTS

Pooling the data into one group was not justified on statistical grounds (see Appendix A). Two groups of states were analyzed separately. One group is comprised roughly of those states on a line from Minnesota through Louisiana (the western group). The second group (the eastern group) is the other six states which are all east of the first group.

The linear regression models for the eastern and western groups of states are presented tables A and B, respectively. These results take into account first order serial correlation via a modified Cochran-Orcutt procedure, (Beach and McKinnon, 1978) and the possibility of multicollinearity.

a) WEATHER

Again, note that weather is discussed only briefly here.

There are a few things to note. First there are a number of significant weather coefficients for eastern group locations (Table A), while there are relatively few significant weather variables for the western group (Table B). There are no significant differences between states in the eastern group with respect to weather variables.

Weather variation has a measurable, substantial, and significant effect on county average soybean yields.

b) PRICES

Factor price (deflated) variation has played a small role in determining soybean yields. Little fertilizer has been applied to soybeans so it is not surprising that the fertilizer price coefficient is indistinguishable from zero. The insignificance of other factor price coefficients suggests that factor price variation has not been great enough to affect yields, and perhaps that some production elasticities of substitution are small.

Futures output prices (deflated) for corn and soybeans have the expected signs (negative and positive, respectively) and are significantly different from zero. Higher soybean futures prices are associated with higher soybean yields, and higher corn futures prices tend to be associated with lower soybean yields, all other things equal. The relative magnitudes of the effects differ between the two groups of



states. In Louisiana soybean yields seem to be unaffected by the variation in corn futures prices. This is not so surprising since corn is not highly competitive with soybeans in production in Louisiana.

Both groups have quadratic expected corn price terms with negative coefficients. However, the interpretations are different. For the western states there is no linear term so that the lower the corn futures price, the higher the soybean yields tend to be. For the eastern states, however, the linear term is positive and significant. This indicates that at some price level for corn, any change in that price will tend to reduce soybean yields. It is difficult to accept that a lower expected corn price will tend to boost soybean yields. However, there is no obvious irregularity in the data to account for this puzzle.

### c) RESEARCH

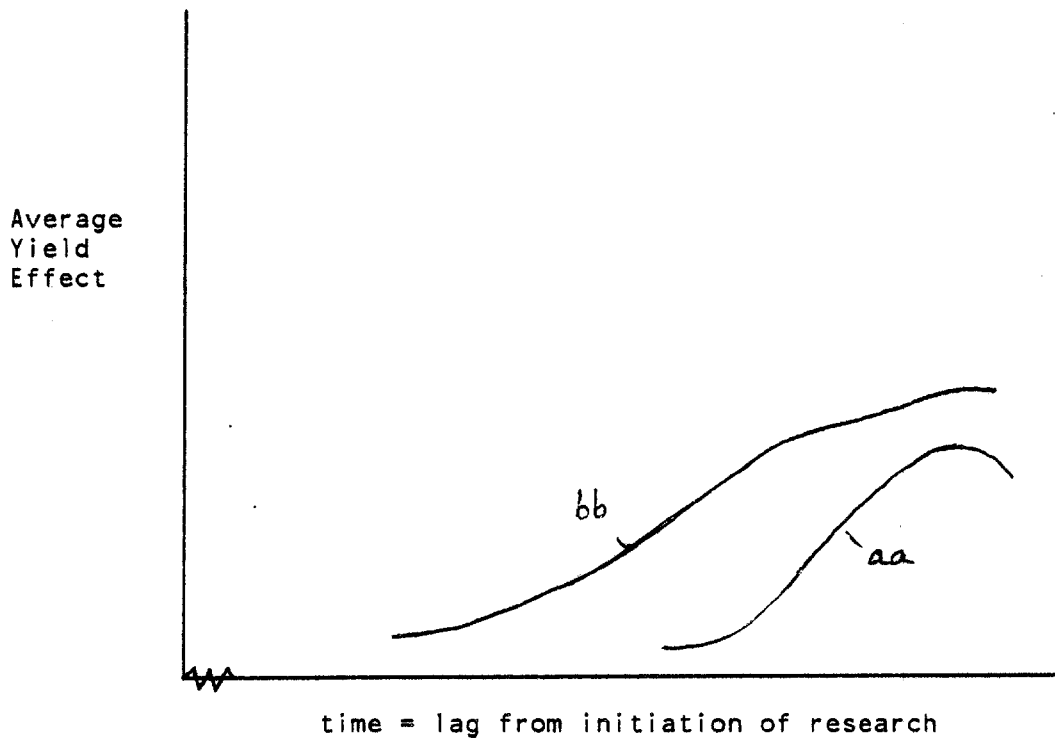
Research expenditures, in 1967 dollars, were combined for up to seventeen years to form a polynomial lag structure, an Almon type lag. The estimated lag structure suggests several things. Intuitively, one might expect the research to have an initial effect after some time, for that effect to grow as the results become widespread, and then for the effect to decay (see Figure 1.0). When many types of research overlap, the total research time shape may be quite different than the shape of any one of its constituents <sup>6</sup> / .

(1) The research coefficients for eastern and western groups of states are different (see Tables A and B). In eastern states the lag

structure is increasing and linear with respect to time. It suggests that there is a lag of three years before soybean research begins to affect soybean yields and that the innovations' effects build until the ninth year.

Figure 1.0

One-time Research Expenditure  
Effects over Time



aa - research with high maintenance research requirements - disease resistance for example.

bb - research with low maintenance research requirements - machinery improvements, for example.

(Note: relative yield magnitudes between aa and bb are purely arbitrary.)

Table A  
 Results of the County Model Estimation - Group I  
 Using Lagged Research Expenditures  
 Adjusted for Serial Correlation - Linear Form  
 (Arkansas, Illinois, Indiana, Kentucky, Ohio, Tennessee)

OLS:  $\bar{R}^2 = 0.76$   $\rho = 0.420$  RSS = 1452.6 d.f. = 178

Variable	OLS Coefficients	Ridge Regression Coefficient (d=.04)
Constant	0.480 ( 1.705)	-1.150 (1.77)
JUNEP	-0.137 ( 0.105)	-0.250 (.109) *
JULYP	1.150 ( 0.277) ***	1.318 (.293) **
JULYP2	-0.071 ( 0.022) ***	-0.084 (.024) **
AUGP	0.958 ( 0.319) **	0.727 (.336) *
AUGP2	-0.046 ( 0.030)	-0.027 (.032)
AUGT	-4.712 ( 0.681) ***	-2.168 (.517) **
AUGT2	0.032 ( 0.005) ***	0.014 (.004) **
SEASON	0.221 ( 0.103) *	0.389 (.104) **
LATITUDE	2.026 ( 0.271) ***	2.12 (.287) ***
SOY	0.528 ( 0.549)	0.593 (.584)
CORN	185.710 (32.140) ***	28.317 (13.416) *
CORN2	-74.437 (12.785) ***	-11.896 (5.37) *
ALMON RESEARCH	0.00024 (0.00004) ***	0.00013 (.00004) **

One, two and three asterisks refer to the 5%,  
 1% and 0.1% significance levels, respectively.  
 Numbers in parentheses are standard errors.

(Table A continued)

Definition of Variables:

JUNEP . . . is total rainfall in June in inches.

JULYP . . . is total rainfall in July in inches.

JULYP2 . . . is the square of JULYP.

AUGP . . . is the rainfall in August in inches.

AUGP2 . . . is the square of AUGP.

AUGT . . . is the mean of the daily high and low temperatures for the  
the month of August in degrees Fahrenheit.

AUGT2 . . . is the square of AUGT.

SEASON . . . is the length of the growing season in weeks.

LATITUDE . . . is the number of degrees North latitude.

SOY . . . is the April Chicago futures price of soybeans for  
September delivery deflated by the prices paid index.

CORN . . . is the April Chicago futures price of corn for December  
delivery deflated by the prices paid index.

CORN2 . . . is the square of CORN.

ALMON RESEARCH . . . is an Almon lag variable composed of the sum of  
research expenditures on soybeans in adjacent states deflated by the  
consumer price index (CPI) for the past seventeen years, with endpoint  
restrictions at two and ten year lags.

Table B

Results of the County Model Estimation - Group II  
 With Lagged Research Expenditures - Linear Form  
 (Iowa, Louisiana, Minnesota, Mississippi and Missouri)

$\bar{R}^2 = 0.72$   $\rho = 0$   $D-W = 2.2$   $RSS = 1591.62$   $d.f. = 131$

A) Results Common to All States

Variable	Coefficient
Constant	76.32 (20.38) ***
AUGP	-0.38 (0.13) **
JULYT	-0.39 (0.18) *
SEASON	-0.53 (0.13) *** (except Minnesota)
LATITUDE	-0.93 (0.29) **
SOY	2.08 (0.81) **
ENERGY	24.57 (3.37) *** (except Minnesota)
CORN2	-1.82 (0.70) ** (except Louisiana)
RESEARCH 1 (1968-73)	0.000219 (0.000054) *** (except Minnesota)
RESEARCH 2 (1974-79)	0.000079 (0.000025) ** (except Minnesota)

B) Individual State Effects

State	Coefficients
Iowa	+ 7.6 (1.32) ***
Minnesota	+ 0.000402*RESEARCH 1 + 0.000262*RESEARCH 2 (.00005) *** (.000025) *** + 7.65*LAND PRICE INDEX (2.01) ***

One, two and three asterisks refer to the 5%,  
 1% and 0.1% significance levels, respectively.  
 Numbers in parentheses are standard errors.

(Table B continued)

Definition of Variables

AUGP . . . is total August rainfall in inches.

JULYT . . . is the average of daily average temperatures for July in degrees Fahrenheit.

SEASON . . . is the number of weeks from the last killing frost in the Spring to the first killing frost in the Fall. It is the length of the potential growing season.

LATITUDE . . . is the latitude of the location in degrees North.

SOY . . . is the futures price of soybeans at Chicago in April for delivery in September, deflated by the index of prices paid (1967).

ENERGY . . . is the fuel and oils price index (1967) deflated by the index of prices paid.

CORN2 . . . is the square of the futures price of corn at Chicago in April for delivery in December, deflated by the index of prices paid.

RESEARCH . . . is the sum of the CPI deflated expenditures over the past seventeen years on soybean research for the reference state and for other states growing soybeans of the same maturity group as those grown at the reference location. RESEARCH 1 refers to the period 1968 to 1973. RESEARCH 2 refers to the period 1974 to 1979.

The lag structure in western states is different. The structure is invariant with respect to time. It suggests that there is only a lag of one year between initiation of research and beginning of an effect on soybean yields in those states. The constant lag coefficients suggest that the effect of the innovations does not build over time. However, the research is still effective more than seventeen years after the research was initiated.

(2) There are strong research spillovers between states. This suggests it is of little consequence whether the research occurs in that state or in adjacent states. As long as the total research effort is the same in those states, regardless of the distribution between states, the effect on yields should be the same. These positive externalities, from a state's point of view, are one incentive to underinvest in crop research.

(3) Continuing productivity growth requires increasing real research expenditures. If research expenditures are held at constant levels, then no yield increases are indicated.

(4) The effects of research are not the same across the sample states. The composition of the research affects the magnitude of the results, the length of the lag, and the time shape of the lags. One way to interpret the lag coefficients is to assume equal real expenditures in each year and then look at the sum of the coefficients. For \$250,000 (1967) spent in each of the past nine years for all states growing the relevant maturity group, between a 1.26 and a 0.7 bushel yield increase in eastern states is indicated. For \$250,000 (1967) spent in each of



the past seventeen years for all states growing the relevant maturity group, there is approximately a 1.1 bushel yield increase for Minnesota (the Jean Lambert effect), and a 0.34 bushel increase for the other western states. Other differences in the research coefficients between states were tested for but were not found. The seven states (AR, IL, IN, KY, MN, OH, TN) seem to have a high physical rate of return to research relative to the four states IA, LA, MS, and MO.

A structural test for stability of the research coefficient could not reject the null hypothesis that it was the same for the periods 1968-73 and 1974-79 for eastern states. However, the null hypothesis was rejected for western states. This is evidence to suggest that research expenditures became less effective in producing higher yields for western states. It is not conclusive that biological limits are the reason. Changes in the direction of research and just plain bad luck could also account for the decline in the research coefficient.

#### RETURNS TO RESEARCH

By using simplifying assumptions <sup>7</sup> / one can convert the physical return to investment into an approximate monetary internal rate of return (IRR). Assume that the results of research cause a proportionate shift in the supply function. Then the benefit measurement method devised by Hayami and Akino (1977) can be used (see Table 1.1).

The rate of shift in the supply function,  $h$ , can be approximated by  $(1 + \xi)k$ , in equilibrium, where  $k$  is the rate of shift in the

production function, and  $\xi$  is the elasticity of supply. Then the following approximation of net benefits will hold in equilibrium:

$$(1.8) \quad ABC = \frac{P Q^2 * h}{2 * (\xi + \eta)} \text{ net consumer's surplus}$$

$$(1.9) \quad ACO = \frac{p Q * h}{1 + \xi} \text{ net producer's surplus}$$

where ABC, ACO, and  $P Q$  are as in Figure 1.1,  $\xi$  is the elasticity of supply and  $\eta$  is the absolute value of the elasticity of demand.

$k_t$  was estimated for each state from the research coefficients and past expenditures for 1968 to 1979. The state  $k$ 's were then weighted by the state harvested acreage as a proportion of harvested acreage for all eleven states. The sum of the weighted  $k$ 's was assumed to represent the national  $k$ . Since these 11 states represent over 90% of soybean production, it does not seem likely that the estimated  $k$  will be biased much by leaving out the other states.

$P Q$  is taken as the value of U.S. soybean production on the farm from U.S. Agricultural Statistics.  $\xi = 0.84$ , and  $\eta = 0.29$  from Houck, Ryan, and Subotnik (1972).

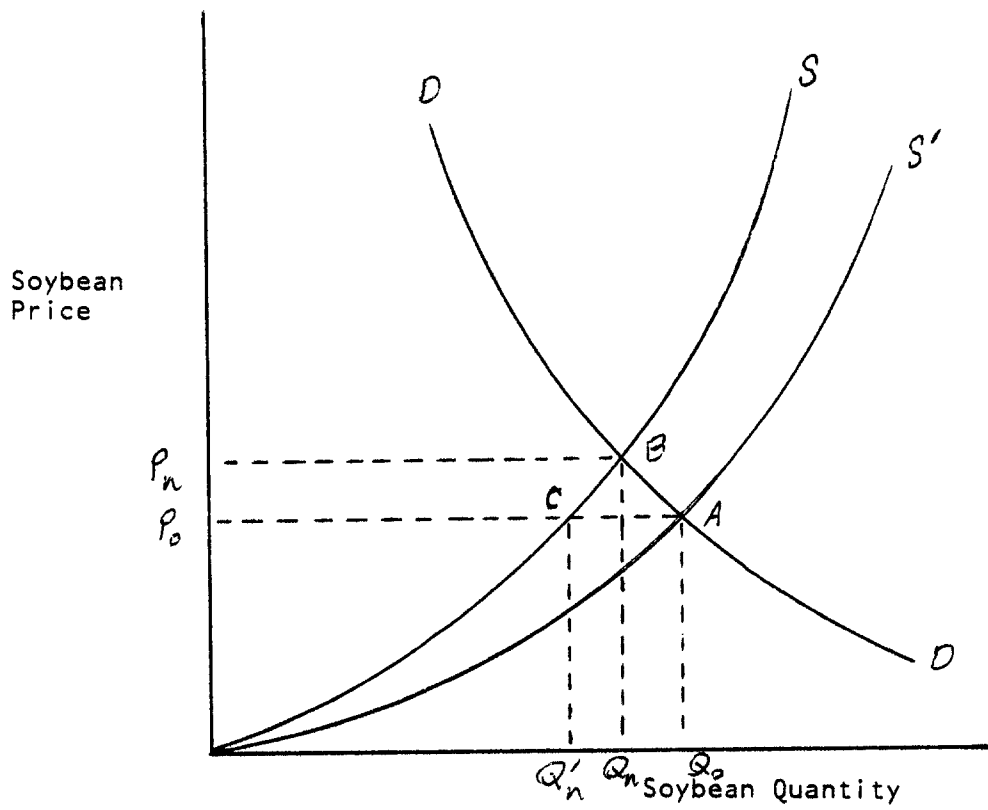
The IRR assumes that the benefits accruing in 1979 will accrue at the same rate to the year 2000 rather than terminate in 1980. In practice these two assumptions mean only a 1 or 2 point difference in the IRR.

Benefits were deflated by the consumer price index (CPI) to 1967 dollars. Costs are measured as the real (1967 CPI deflated) dollar amounts spent on public soybean research from 1951 to 1978.

The IRR lies between 55 and 56%. The IRR estimate is rather coarse, but conservative. 1) The IRR does not include returns to maintenance research or research which does not lead to higher yields. 2) If the supply shift is not strictly proportionate, but involves a positive constant as well, then the net change in producers' surplus will tend to be understated. 3) Real growth in demand is not reflected in the benefit estimates.

Although this is not a true marginal rate of return, it indicates that there has been a substantial return to soybean research, and that there may have been underinvestment in public soybean research.

Figure 1.1



Model of estimating social returns to soybean research (Hayami and Akino, 1977).

SUMMARY

1) There is a measurable physical return to research in the way of improved soybean yields.

2) The research coefficient is stable over time for some states and has declined for others. There are several potential reasons for the decline in the coefficient. Among them are a) yields reaching a biological limit, b) bad luck in research, and c) a change in the direction of research away from yield improving innovations.

3) Research spill-overs between states are very important. Taken at face value, a state typically acquires more than 80% of its yield improving technology from other states. This suggests that a) there may be substantial benefits to be had from collaboration and coordination of research between similar states, and b) for soybeans in particular, much of the yield improvement has been due to the cooperation between state research institutions.

4) The lag between initiation of research and an effect on county yields is short. The lag ranges between one and three years before a measurable impact occurs.

5) The additional length of time before the research is fully adopted also varies. In the eastern Soybean Belt states, adoption seems to be complete after about six years. In western Soybean Belt states, adoption seems to have been going on for over 15 years after initial adoption.

6) Factor price fluctuations do not appear to have been large enough to have had an effect on soybean yields generally. Expected output prices have had a substantial and significant effect on yields, and the signs have been as expected - positive for soybean price and negative for corn price.

7) Weather effects seem to differ between eastern and western Soybean Belt states. In the eastern states there are indications that rainfall has been both too high and too low, but too low on average. In western Soybean Belt states rainfall variation in different growing season months has not been as important. Potential season length and the general location (latitude) also seem to have some influence on soybean yields.

APPENDIX A

DERIVATION OF THE YIELD MODEL

The production function expresses the technical relation between inputs and outputs.

. . . the biological nature of the production process, the time lags involved between planting and harvest, and the generally extensive use of land and climate leads naturally to the separation of total crop production into acreage and yield components. (Houck and Gallagher, 1976)

One can write the production function for soybeans as the product of a yield function and acreage:

$$(A.1) \quad Q_s = Q[Y(I; T) * A]$$

Where  $Q_s$  is the output,  $Y$  is the yield function,  $I$  represents all physical inputs, and  $A$  is soybean acreage.

I assume that  $Q$  is continuous and twice differentiable for all inputs and technology. Technology is included as a parameter to the producer. Technology is often embodied within inputs and cannot be separated from them. For example new varieties and herbicides embody technology. The producer may have the choice of several substitute inputs, but he either uses the new technology or he does not. He cannot affect the level of available technology.

Now assume expected profit maximization where corn is the only crop competitive in production. For most of the locations in question this is not a bad assumption. The locations are in the Corn Belt/Soybean Belt. The assumption of expected profit maximization presumes technical efficiency.

Form the profit function and apply the expectations operator, E.

$$(A.2) \quad E(\pi) = E[P_s * A_s * Y_s(*) + P_c * A_c * Y_c(*) - P_i (I_{ic} + I_{is})] \quad \text{or}$$

$$(A.3) \quad E(\pi) = EP_s * A_s * EY_s(*) + COV(P_c, Y_c) * A_c + EP_c * A_c * EY_c(*)$$

$$+ COV(P_i, Y_i) * A_i - P_i (I_{ic} + I_{is})$$

Where  $Y_s(*)$ , or  $Y_c(*)$  is the yield function in (A.1), P is price, Y is the yield, s represents soybeans, c represents corn, and I represents the input quantity of input i (including land).

Actual prices and yields are assumed to be random variables. There is assumed to be no effect across locations producing an effect on the national price:  $COV(P, Y) = 0$  <sup>8</sup> / for both corn and soybeans. Input prices and quantities are assumed to be known with certainty. Then (A.3) is equivalent to (A.4).



$$(A.4) \quad E(\pi) = EP_{s s} * A_{s s} * EY_{s s} (*) + EP_{c c} * A_{c c} * EY_{c c} (*) - P_i (I_{i s} + I_{i c})$$

The expected natural environment is not included in the expected profit function because the effect of expected (average) weather is assumed to be zero and independent of the inputs.

Now maximize  $E(\pi)$  with respect to the decision variables  $I_{i s}$ ,  $I_{i c}$ ,  $A_{s s}$ , and  $A_{c c}$  (treating land separately). The first order conditions are:

$$\left. \begin{aligned} (A.5) \quad EP_{c c} * A_{c c} * EY'_{c c} - P_{i s} &= 0 \\ (A.6) \quad EP_{s s} * A_{s s} * EY'_{s s} - P_{i s} &= 0 \end{aligned} \right\} \text{other inputs}$$

$$\left. \begin{aligned} (A.7) \quad EP_{c c} (EY_{c c} (*) + EY'_{c c}) - P_{l c} &= 0 \\ (A.8) \quad EP_{s s} (EY_{s s} (*) + EY'_{s s}) - P_{l s} &= 0 \end{aligned} \right\} \text{land}$$

The  $P_{l s}$ 's are land prices. The second order sufficient conditions are that

$$(A.9) \quad (-1)^n \det \begin{matrix} A_{11} & \dots & A_{1n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ A_{n1} & \dots & A_{nn} \end{matrix} < 0 \text{ for } n = 1 \dots N$$

where  $A_{ij}$  is the partial derivative of the  $i^{\text{th}}$  first order condition with respect to the  $j^{\text{th}}$  decision variable.

The intuition behind the first order conditions is that the expected marginal value product of each factor equal its factor price. If not, expected profit could be increased by using more or less of the factor.

Assume that  $\det [A_{ij}] = 0$ , then we can apply directly the implicit function theorem. There exist explicit continuously differentiable functions of the decision variables expressed as functions of parameters faced by the producer, i.e. prices and technology (Takayama, 1974). These functions are factor demand equations, and satisfy the production function when substituted into the production function. Thus

$$(A.10) \quad EQ_s = E[Y_s(P_s, P_c, P_i, P; T) * A_s] \quad \text{and dividing by } A_s$$

$$(A.11) \quad EY_s(*) = EQ_s / A_s = E[Y_s(P_s, P_c, P_i, P; T) / A_s]$$

Expected yields are not observed, however. Actual yields can be expressed as the sum of the expected yield, plus environmental effects and a random disturbance component. That is,

$$(A.12) Y_s = EY_s (*) + \text{Environmental Effects} + \text{error}$$

where  $Y$  is observed yield. Part of the error is measurement or estimation error, which is assumed to be random.

#### Technical Change

Technical change on the farm comes about through the adoption of innovations that the decision maker expects to enhance his net revenue. Although it is possible for an exceptional farmer to discover new technology on his own, it is assumed that for the average farmer the experiment station and input supply firms are the sources of new technology.

Evenson (1968) introduced technical change directly into the production function by arguing that:

$$(A.13) Q = f(X_i, K)$$

$Q$  is output,  $X_i$  are inputs and  $K$  is the output of a research unit that is relevant to the experiment station, and

$$(A.14) K = g(R)$$

where R is research inputs, expenditures. So that,

$$(A.15) \quad Q = f(X_i, g(R))$$

If R are research expenditures, then K can be thought of as the increase in yields and improvements in other attributes that are a result of those expenditures. Thus one model of county yield change could include research expenditures, and another could include experiment station yields as independent variables.

The transmission of the innovation from the experiment station to the farm has been documented to involve a lag. The length of the lag between discovery and use of innovations on the farm may vary depending on the type of innovation.

Write the full models:

$$(A.16) \quad Y = EY(*;g(R)) + \text{Weather effects} + \text{error} \quad \text{or}$$

$$(A.17) \quad Y = EY(*;K) + \text{Weather effects} + \text{error}$$

The specification of the functional form need not be a problem. The functional form can be allowed to be determined by the data rather than imposed, by using a Box-Cox (1964) flexible functional form approach.

FUNCTIONAL FORM

Economic theory does not suggest a functional form for the direct yield function, and consequently does not prescribe a specific form for the indirect yield function. One can allow the data to determine the functional form by using a combination of Box-Cox transformations on the variables and a grid-search maximum likelihood estimation technique to find the right transformation.

A simple Box-Cox transformation for a variable is defined as:

$$(A.18) \quad y^{(\lambda)} = (Y^{(\lambda)} - 1) / \lambda \quad \text{for } \lambda \neq 0 \quad \text{and}$$

$$(A.19) \quad y^{(\lambda)} = \text{Ln}(Y) \quad \text{for } \lambda = 0 \quad (\text{Box and Cox, 1964})$$

The transformed linear model includes the linear, Generalized Leontief, and Cobb-Douglas/Trans-Log functional forms when  $\lambda = 1, 0.5,$  and  $0.0,$  respectively.

The maximized log-likelihood estimate for a given  $\lambda$  is:

$$(A.20) \quad L_{\max} = N(-0.5) \text{Ln}(\hat{\sigma}_{(\lambda)}^2) + (\lambda - 1) \sum_{i=1}^N \text{Ln}(Y_i) \quad (\text{Zarembka, 1974})$$

where  $Y$  is untransformed and  $N$  is the total number of observations.

An approximate  $100(1 - \alpha)$  per cent confidence region can be computed from:

$$(A.21) L_{\max}(\lambda^*) - L_{\max}(\lambda) < 0.5 \chi_{\tau}^2(\lambda) \quad (\text{Box and Cox, 1964})$$

where  $\tau$  is the number of independent components in  $\lambda$  (1 in this case).  
At the 5% level  $\chi^2/2 = 1.92$ .

For the eastern group (AR, IL, IN, KY, OH, TN) and for the western group (IA, LA, MN, MS, MO) the log-likelihood value was maximized for  $\lambda = 1.0$ . The 95% confidence interval on  $\lambda^*$  includes the Generalized Leontief form ( $\lambda = 0.5$ ) for the eastern group, but is much tighter around  $\lambda^*$  for the western group. (see Table C) This indicates that the linear-in-parameters form of the yield model is satisfactory, and that other forms, such as the Cobb-Douglas/Trans-Log, are unacceptable.

Table C  
 Box-Cox Transformation and Maximum Likelihood  
 Estimation of the County Yield Models  
 Using Lagged Research Expenditures

Lambda	Log-Likelihood Value	MSE
Group I AR, IL, IN, KY, OH, TN		
-2.00	-616.511	0.0000014
-1.50	-570.035	0.000015
-1.00	-531.867	0.000176
-0.50	-502.793	0.00230
0.00	-483.226	0.03223
0.50	-473.163	0.50618
1.00*	-472.203	8.74034
1.50	-479.615	164.66477
2.00	-494.433	3351.07950
Group II - MN, IA, MO, MS, LA		
0.00	-202.474	0.024
0.50	-183.965	0.494
1.00*	-170.732	10.711
1.50	-178.811	312.471
2.00	-187.222	9157.679

POOLING THE DATA

Since there was both time-series and cross-section data, it was desirable to test whether a pooled model was justified. A test of the equality of the error variances for the group of Corn Belt vs. Non-Corn Belt states rejected equality at the 5% level. Separate preliminary regressions for each state revealed a similarity of parameters and error variances which led to two groupings of states. The eastern group contains Arkansas, Illinois, Indiana, Kentucky, Ohio and Tennessee. The western group contains Iowa, Louisiana, Minnesota, Mississippi, and Missouri. Further tests within these two groups could not reject the equality of the residual error variances.

Seemingly unrelated regression tests for across-location correlation of the residuals by using an intercept dummy for each year rejected that possibility for the western group, but found significant effects for the eastern group. In other words this supported the hypothesis that  $E(u_{it} u_{jt}) = 0$  for location  $i \neq j$  in period  $t$  for the western group, but that  $E(u_{it} u_{jt}) \neq 0$  for the eastern group. Zellner's (1962) GLS method for more efficient parameter estimation was used for the eastern group. The results were not more efficient as measured by a comparison of the diagonal elements of the variance-covariance matrices for OLS and Zellner's GLS. This was probably the result of a loss of 25% of the observations in order to obtain equal numbers of observations for each location, and to the restriction that all locations have the same independent variables. The significance of the year dummies may also mean that some sort of weather variable was missing, in which case "When the correlations are due to common omitted variables, it is not clear whether Zellner's GLS method is superior to OLS." (Maddala, p. 331-2)

Given that the GLS parameter estimates were relatively close to those of OLS, that they were generally significant in both cases, and the potential for leaving out a common variable, the cross-sectional error correlation did not seem to be a severe problem. Consequently Zellner's GLS method was not used.

A test of forward lagged corn and soybean futures prices found them to be insignificant. This tended to confirm the direction of causality of the model and substantiated the use of a single equation.



LIST OF REFERENCES

Ayer, H.W. and G.E. Schuh. 1974. Social rates of return and other aspects of agricultural research: the case of cotton research in Sao Paulo, Brazil: reply. AJAE 56:178-9.

Barah, B.C. 1980. Combining Cross-section and Time-Series Data. ICRISAT (International Crops Research Institute for the Semi-Arid Tropics), Andhra Pradesh, India.

Beach, C.M. and J.G. MacKinnon. 1978. A maximum likelihood procedure for regression with autocorrelated errors. Econometrica 46: 51-8.

Bonner, J. 1962. The upper limit of crop yield. Science 137(3523).

Box, G.E.P. and D.R. Cox. 1964. An analysis of transformations. J. of the Royal Statistical Society 26(2), Series B; 211-43.

Cline, P.L. 1975. Sources of Productivity Change in U.S. Agriculture. Ph.D Dissertation, Oklahoma State University.

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

\_\_\_\_\_, and W. Peterson. 1981. The declining productivity of agricultural research. in Evaluation of Agricultural Research. Miscellaneous Publication 8-1981, Minnesota Agricultural Experiment Station, University of Minnesota.

Evenson, R.E. 1968. The Contribution of Agricultural Research and Extension to Agricultural Productivity. Ph.D Dissertation, University of Chicago.

\_\_\_\_\_. 1978. Research invention, extension and productivity change in U.S. agriculture: an historical decomposition analysis. Paper Prepared for the Symposium of Agricultural Research and Extension, Moscow, Idaho, May 21-23, 1978.

\_\_\_\_\_, P.E. Wagoner, and V.W. Ruttan. 1979. Economic benefits from research: an example from agriculture. Science 14 September, 1979.

Frey, K.J. 1971. "Improving Crop Yields Through Plant Breeding." in Moving Off the Yield Plateau. ASA Special Publication No. 20. J.D. Easton and R.D. Munson, eds.

Fuss, M. and D. McFadden, eds. Production Economics: A Dual Approach to Theory and Application. Amsterdam: North-Holland Publishing Co.

Griliches, Z. 1958. Research costs and social returns: hybrid corn and related innovations. JPE 66; 419-31.

Groenewegen, J.R. 1980. Corn and Soybean Acreage and Yield Response with Emphasis on Multiple Product Production, Uncertainty, and Commodity Programs. Ph.D. Dissertation, University of Minnesota.

Hayami, Y. and M. Akino. 1977. "Organization and Productivity of Agricultural Research Systems in Japan." in Resource Allocation and Productivity in National and International Agricultural Research. T.M. Arndt, D.G. Dalrymple, and V.W. Ruttan, eds. Minneapolis: University of Minnesota Press.

Houck J.P., M.E. Ryan, and A. Subotnik. 1972. Soybeans and Their Products: Markets, Models, and Policy. Minneapolis: University of Minnesota Press.

Loomis, R.S. and W.A. Williams. 1963. Maximum crop productivity: an estimate. Crop Science 3(1); 67-72.

Lu, Y. and L. Quance. 1979. Agricultural Productivity: Expanding the Limits. Agricultural Information Bulletin 431, U.S. Department of Agriculture. Washington, D.C.: Government Printing Office.

Lu, Y., P. Cline, and L. Quance. 1979. Prospects for Productivity Growth in U.S. Agriculture. Agricultural Economic Report No. 435, U.S. Department of Agriculture. Washington, D.C.: Government Printing Office.

Maddala, G.S. 1977. Econometrics. New York: McGraw-Hill Book Company.

Miner, A.G. 1981. The Contribution of Technology to Soybean Yields in Nerlove, M. 1958. The Dynamics of Supply: Estimation of Farmers' Response to Price. Baltimore: The Johns Hopkins Press.

Norton, G.W. 1981. The productivity and allocation of research: U.S. agricultural experiment stations, revisited. in Evaluation of Agricultural Research. Miscellaneous Publication 8-1981, Minnesota State Experiment Station, University of Minnesota.

Oury, B. 1965. Allowing for weather in crop production model building. JFE 47; 270-83.

Peterson, W.L. 1971. "The Returns to Investment in Agricultural Research in the United States." in Resource Allocation in Agricultural Research. W.L. Fishel, ed. Minneapolis: University of Minnesota Press.

Shaw, L.H. 1964. The effect of weather on agricultural output: a look at methodology. JFE 46; 218-30.

Stallings, J.L. 1960. Weather indexes. AJAE 42; 180-6.

Takayama, A. 1974. Mathematical Economics. Hinsdale, Ill.: The Dryden Press.

Thompson, L.M. 1970. Weather and technology in the production of soybeans in the central United States. Agronomy Journal 62; 232-6.

Tomek, W.G. and R.W. Gray. 1970. Temporal relationships among prices on commodity futures markets: their allocative and stabilizing roles. AJAE 52; 372-80.

U.S. Department of Agriculture. Inventory of Agricultural Research, Vol. 1, Table 1. Various years. Washington, D.C.: Government Printing Office.

White, F.C. and N. Garren. 1981. Effects of externalities on the efficient allocation of agricultural research expenditures. Unpublished Paper.

Zarembka, P. 1968. Functional form in the demand for money. JASA June, 1968, p. 502-11.

\_\_\_\_\_, ed. 1974. "Transformations of Variables in Econometrics." in Frontiers in Econometrics. New York: Academic Press.

Zellner, A. 1962. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. JASA June, 1962, p. 348-368.

FOOTNOTES

1. If trading volume is adequate and trading positions are diverse, the futures market is efficient in discovering future prices. A trader who has information that the futures price should be different can take a position whereby he can profit from that information. Tomek and Grey (1972) found that the futures prices of corn and soybeans were reasonably good predictors of the harvest cash price.

2. One should recognize at this point that there are types of technical change which are not reflected in higher yields. Changes in oil and protein content, taste, disease resistance, and managerial techniques may not increase yields. In some cases they may actually reduce yields while reducing costs or increasing the useable quantities of outputs. These types of technical change are not expected to be captured by this model, and so it is expected to underestimate the contribution of research to productivity change in this respect.

3. For practical purposes, research expenditures of a generic nature were not used in computations. That is not to say that the research results of nonspecific or of other crop research were ignored. There are severe practical problems with assigning non-specific research to soybeans. For example, weed control may have a substantial effect on any crop yield, but what portion, if any, of the research cost should be attributed to soybeans? Research conducted in the private sector presumably must be recouped in the market price, then the amortized research cost is already reflected in soybean production costs and prices. As for private varietal development, it was insignificant relative to public research over the time in question.

4. The second order polynomial allows the lagged coefficients to "curve" with respect to time. One might expect the research results to be small at first, to build and then to decay over an extended period. This lag structure allows that time shape to emerge from the data.

5. There seem to be three approaches to include weather in econometric models. One is the method used by Stallings (1960), the experimental plot index. In that index the ratio of the actual yield to the detrended, predicted yield for selected plots is assumed to represent the net effect of weather.

This type of index assumes that weather is the only source of deviations from the trend. That may not be the case. Detrending for a long series to obtain the predicted yields embodies the assumption that there has been no trend effect due to weather, and that the trend adequately accounts for technical change. It also assumes no technical-weather interaction. In essence it says all the variation about the trend is due to weather.

This method is not practical to use for several reasons. First, the published indexes are too highly aggregated across space, and less

aggregated indexes could not be found. Second, the collected experiment plot data could not meet the requirement for constancy of treatment factors to construct the index. Third, the cost of computing the indexes for so many locations would be quite high, and the necessary data was incomplete.

A second approach is written about by Oury (1965). In this method readily available weather measures (temperature and precipitation) are nonlinearly combined into "aridity" indexes. Many forms were proposed. In short, the indexes impose some structure on the weather variables which may be suitable for wheat or corn, but which may be unsuitable for soybeans. A few of these indexes were tried in earlier work, but results were disappointing and difficult to interpret.

The third approach is that of directly including weather variables as explanatory variables in regression equations.

The fact that precipitation is different from the amount of available soil moisture does not necessarily mean that it is not a good proxy for soil moisture. Use of such proxies may be better than none at all, and may substantially add to the reliability of the other coefficients.

Aggregation over time, a month for example, may tend to be misleading vis-a-vis the true relation. The effect of weather is of secondary importance and it is difficult to justify the enormously greater task of collecting daily data, constructing crop calendars, etc.

Agronomists, farmers, and others often are willing to characterize a crop year as good, bad or average. Crop years are frequently summarized in terms of wet/dry Spring, early/late Fall, and other aggregated weather characteristics. This suggests that using monthly data will also be useful statistically even though they may not be precise.

The model may suffer slightly from geographic aggregation, but the variables are likely to be representative since they are recorded for a location within the county.

Finally, the specification of the functional form may tend to be a problem. For example, quadratic treatment of temperature and precipitation doubles the number of weather variables: a loss of degrees of freedom. However, it allows changing marginal weather effects: additional realism. At the same time, the response to low precipitation may have a different shape than the response to high precipitation, while the quadratic form imposes symmetric effects. Although this may be a problem, there may also be a solution: the functional form can be determined by the data rather than imposed a priori.

The problems with the first two methods make them undesirable to use. The third approach, use of simple weather variables, although it

may not be precise, seems to be a reasonable approach.

6. The intuitive, smoothed time shapes in Figure 1.1 are a very few of the many possible time shapes for particular types of research. Curve aa might represent a shape for an innovation which requires a long period to develop, is disseminated quickly, but which shows rapid deterioration after some time in use. Breeding for disease resistance might fall into this category. After generations of selection and breeding, a new disease resistant variety might be introduced, quickly adopted in areas where it is needed, but after a few years the disease may adapt or overcome the disease resistance. The deterioration of the innovation is distinct, in theory, from the emergence of new and different problems - e.g. new diseases or pests - but it may be extremely difficult to separate their effects in practice.

Curve bb might represent an innovation which is developed relatively rapidly, is adopted over an intermediate period, and which does not deteriorate. New machinery might be such an innovation. The lag from idea to developed product may be short relative to biological innovations since development does not depend upon repeated crop production cycles. Adoption may be somewhat slower due to capital requirements and the vintage of existing machinery. As long as the innovation is designed into new equipment, the research effect should tend to stabilize at some level.

#### 7. Rate of Return Assumptions

i) The percentage change in yield = the percentage change in the production shift.

ii) Constant elasticity of demand.

iii) Derived demand implies that the compensated demand is irrelevant for measuring consumers' surplus. (Ayer and Schuh, 1974)

iv) Equilibrium.

v) No net soybean imports.

vi) Ignore distribution of benefits.

vii) Constant proportionate shift in the supply function relative to that which would have been without the innovations.

viii) Assume the supply curve is equivalent to the marginal cost curve.

8. Although Groenewegen (1980) assumed that  $COV(Y,P) = 0$ , this may not be a good assumption. County yields may indeed have no effect on the national or international price. There may be, however, a wide-spread weather effect which tends to influence all yields, thus

influencing price and making  $COV(Y,P)$  not equal to zero.

If experiment station yields respond to year to year weather variation as county yields do and if  $COV(Y,P)=0$  for station yields, then  $COV(Y,P)$  for county yields is also likely to be zero. Several results of experiment station estimation indicate that  $COV(Y,P)$  might be zero. (1) The current price of soybeans has an insignificant coefficient. The correlation coefficient of the experiment station yield residuals with the residuals of the regression of soybean price on the independent variables is not significant. The correlation coefficient is  $-0.0256$  and is significant at only the 22.5% level. (2) Individual year dummies were not significant indicating no pan-geographic effect not captured by the independent variables. With these results for the experiment station model, it seems that  $COV(Y,P)=0$  is not a bad assumption.

The dummy year variables were tested in specially constructed county yield equations. Variables which had the same value at more than one location because of lack of detailed data were dropped from the equation. This left only location specific variables and insignificant dummy year variables.

This indicates there is no pan-geographic effect which is not captured by the independent variables. This also suggests that the covariance term is zero.

9. In this paper it is hypothesized that extension does not contribute to technological change. Rather it is important in determining the length of time from discovery of an innovation to its adoption on the farm. Soybean/corn producers in Minnesota seem to be sophisticated in their ability to gather information (Miner, 1981) so that even without extension per se, new technology would eventually filter out to farmers. This does not mean that extension is valueless. Rather, it places the value of extension on the timeliness and extent of the dissemination of information. In less developed countries, extension may affect the flow of information in a more critical way.

Some extension may be embodied within experiment station expenditures, since researchers frequently summarize results and preliminary results in experiment station reports. Farmers may be able to draw conclusions and extract valuable information from these reports.

Inputs may embody new technology, in the form of disease and pest resistance in new varieties, for example. These types of innovations may be promoted by seed, fertilizer and other companies, reducing the need for extension per se. The costs of this type of extension are embodied in prices.

It is not the intention in this paper to test the hypothesis that extension does not contribute to technological change. The relevant data one would need to test this hypothesis does not exist to my knowledge.