A TECHNIQUE TO ESTIMATE INPUT PRODUCTIVITY FROM FARM DATA

John R. Allison and David W. Parvin, Jr.

INTRODUCTION

Unfortunately, procedures are not available for handling variations induced by unquantifiable difference in location, soil, weather or management, particularly if the data source is farm survey information from relatively small samples for a single production season. Estimation problems occur regardless of whether classical, profit or trans-log approaches are used. Procedures suggested by Hoch and Hoch and Mundlak for handling these disturbances in classical production functions require *a priori* knowledge to devise a weighting system or observations over time to provide estimates of weights.

Profit functions as proposed by Lau and Yotopoulos require data to be of such nature that a production function can be specified either in the normal form or that the relationship between profit and input quantities can be specified and estimated. The price of the product is also required to be either a function of quality or of selling costs or that some common or average price is utilized.

Spann's procedure (using reformulation of logarithmic derivation of the trans-log production function to estimate payments to a factor of production as a fraction of total revenue) requires the same product price data as profit function estimation, as well as requiring the underlying production surface to be approximated by a trans-log production function. Trans-log cost functions as used by Binswanger can estimate elasticities of input demand when the production surface approaches that of a trans-log, and when weighting or estimating techniques for management, weather, location and soil differentials, and technology changes can be incorporated in the model.

This article presents a procedure for estimating average marginal productivity of inputs from a group or subgroup of farm operations without estimating or specifying production surface. The relationship between average cost per unit of output and the amount of or cost of individual inputs applied in the production process is estimated. The procedure does not require any assumptions concerning elasticities of production functions. It can be used in some situations where either incorrect production surface specifications or extreme variability of data prevent estimation of the usual production function.

PROCEDURE DEVELOPMENT

Given a production function:

$$Y_i = f(X_{i1}, \dots, X_{is}) \tag{1}$$

where

Y_i = production of Y from the i^{th} unit (may consist of a farm unit or farm subunit) and i=1 to r

X_{ij} = input j used in the production of Y_i and j=1 to s

and total cost of Y_i is

$$TC_i = \sum_{j=1}^{s} P_{ij}X_{ij} \tag{2}$$

where P_{ij}=price or cost per unit of X_{ij}. TC_i can also be expressed as a function of Y_i:

John R. Allison is Associate Professor, Department of Agricultural Economics, Georgia Experiment Station, University of Georgia and David W. Parvin, Jr. is Associate Professor, Department of Agricultural Economics, Mississippi State University.

$$TC_i = g(Y_i) = h(X_{il}, \dots, X_{is})$$

and

$$ATC_i = k(X_{il}, \dots, X_{is})$$

$$= 1(P_{il}X_{il}, \dots, P_{is}X_{is}) \quad (3)$$

We suggest that a simple linear approximation of the average relationships of ATC_i for a given point or small segment of the production surface or surfaces can be made in those situations where it is unfeasible to specify the production surface or surfaces because of mathematical complexity, or where data represent a multitude instead of a single surface or even a single family of production surfaces. One alternative is the following linear formulation in which:

$$ATC_i = a_{io} + a_{il}P_{il}X_{il} + \dots + a_{is}P_{is}X_{is} + e_i \quad (4)$$

where

- ATC_i = total cost of production divided by Y_{il} or total production
- a_{io} = cost per unit of output explained by the equation but not by individual a_{ij}'s
- a_{ij} = change in ATC_i associated with a change in $P_{ij}X_{ij}$
- e_i = random error for management, soil differences, differences in crop and environment, etc.[1]

Assuming P_{ij} is a constant, $\partial Y_i/\partial X_{ij}$ can be estimated from a_{ij};

$$a_{ij} = \frac{\partial\left(\frac{TC_i}{Y_i}\right)}{\partial P_{ij}X_{ij}} = \frac{Y_i\left(\frac{\partial TC_i}{\partial P_{ij}X_{ij}}\right) - TC_i\left(\frac{\partial Y_i}{\partial P_{ij}X_{ij}}\right)}{Y_i^2} \quad (5)$$

Since $\partial TC_i/\partial P_{ij}X_{ij}=1$; and $\partial Y_i/\partial P_{ij}X_{ij}=\partial Y_i/\partial X_{ij} \cdot 1/P_{ij}$ if P_{ij} is constant.

$$\frac{\partial Y_i}{\partial X_{ij}} = \frac{P_{ij}}{TC_i}\left[Y_i - Y_i^2\,(a_{ij})\right] \quad (6)$$

We are estimating $\partial Y_i/\partial X_{ij}$ from the average cost function using the relationship between marginal productivity and average unit cost. If:

$$\partial_{ij} < \frac{1}{Y_i},\ MP > 0$$

$$\partial_{ij} = \frac{1}{Y_i},\ MP = 0$$

$$\partial_{ij} > \frac{1}{Y_i},\ MP < 0$$

Productivity Per Unit of Land

Estimates of marginal productivity of land are useful, especially when comparing agricultural regions where differences in land cost reflect regional differences in climate and opportunity cost of land. Unfortunately, procedures to estimate land costs are very subjective. Differences in land costs within a region may be artificial. Using a land quantity variable will also reduce the ability to estimate the effect of size of farms since the land variable is also a common measure of size on these farms.

Since the land variable is fixed for many farm units during a given production season, inputs for crop production are evaluated or planned in terms of units of inputs per unit of land. The production function in equation (1) becomes:

$$Y_i = X_{is} \cdot h_i\left(\frac{X_{il}}{X_{is}}, \dots, \frac{X_{is-1}}{X_{is}}\right) \quad (7)$$

where X_{is}=land input and equation (2) becomes:

$$TC_i = X_{is}\left[P_{is} + \sum_{j=1}^{s-1} P_{ij}\right] \quad (8)$$

Transforming the relationship to a per unit of land basis modifies equation 4 to:

$$\frac{TC_{iL}}{Y_i} \text{ or } ATC_{iL} = b_{io} + b_{il}P_{il}\frac{X_{il}}{X_{is}} + \dots$$

$$+ b_{is-1}P_{is-1}\frac{X_{is-1}}{X_{is}} + e \quad (9)^{2}$$

[1] Error from differences in soil would be removed if the function was fitted over years using a farm unit consisting of a single soil.

[2] Equation (9) could be written:

$$\frac{\left(\frac{TC_{iL}}{X_{is}}\right)}{\left(\frac{Y_i}{X_{is}}\right)} = ATC_{iL} = b_{i0} + b_{il}P_{il}\frac{X_{il}}{X_{is}} + \dots + b_{is-1}P_{is-1}\frac{X_{is-1}}{X_{is}} + e$$

where

$$\frac{TC_{iL}}{Y_i} = \frac{TC_i - P_{is}X_{is}}{Y_i}, \text{ and}$$

$$ATC_{iL} = ATC_i - \frac{P_{is}X_{is}}{Y_i}$$

And equation 9 being linear:

$$\frac{\partial ATC_{iL}}{\partial\left(\frac{P_{ij}X_{ij}}{X_{is}}\right)} = b_{ij} \tag{10}$$

Assuming P_{ij}'s are constants:

$$\frac{\partial ATC_{iL}}{\partial\left(\frac{P_{ij}X_{ij}}{X_{is}}\right)} = \frac{\partial\left(\frac{TC_{iL}}{Y_i}\right)}{\partial\left(\frac{P_{ij}X_{ij}}{X_{is}}\right)} = \frac{\partial\left[\frac{\left(\frac{TC_{iL}}{X_{is}}\right)}{\left(\frac{Y_i}{X_{is}}\right)}\right]}{\partial\left(\frac{P_{ij}X_{ij}}{X_{is}}\right)}$$

$$= \frac{\frac{Y_i}{X_{is}}\left[\frac{\partial\left(\frac{TC_{iL}}{X_{is}}\right)}{\partial\left(\frac{P_{ij}X_{ij}}{X_{is}}\right)}\right] - \frac{TC_{iL}}{X_{is}}\left[\frac{\partial\left(\frac{Y_i}{X_{is}}\right)}{\partial\left(\frac{P_{ij}X_{ij}}{X_{is}}\right)}\right]}{\left(\frac{Y_i}{X_{is}}\right)^2} \tag{11}$$

and

$$\frac{\partial\left(\frac{Y_i}{X_{is}}\right)}{\partial\left(\frac{X_{ij}}{X_{is}}\right)} = \frac{P_{ij}}{\left(\frac{TC_{iL}}{X_{is}}\right)}\left[\frac{Y_i}{X_{is}} - \left(\frac{Y_i}{X_{is}}\right)^2 b_{ij}\right]. \tag{12}$$

Marginal productivities of inputs applied per unit of land—$\partial(Y_i/X_{is}) / \partial(X_{ij}/X_{is})$—are desirable measures of productivity. Unfortunately, in some classes, input must be aggregated over several physical forms because limited observations do not allow specification of every physical form used by farm operators, e.g., lime, fertilizer and gypsum. Additionally, for other input classes such as pesticides, there is no unique chemical compound or even physical form, i.e., the chemical compound(s) vary with the pest, its severity, and application techniques used by farm operators. Also, where price reflects differences in quality of an input, amount paid for the input is important.[3] Thus, in some analyses the measure $\partial(Y_i/X_{is}) / \partial(P_{ij}X_{ij}/X_{is})$ may be a more useful measure. Regrouping terms in equation 12 provides the relationship:

$$\frac{\partial\left(\frac{Y_i}{X_{is}}\right)}{\partial\left(\frac{P_{ij}X_{ij}}{X_{is}}\right)} = \frac{1}{\frac{TC_{iL}}{X_{is}}}\left[\frac{Y_i}{X_{is}} - \left(\frac{Y_i}{X_{is}}\right)^2 (b_{ij})\right] \tag{13}$$

In most analyses $\partial(Y_i/X_{is}) / \partial(X_{ij}/X_{is})$ or $\partial(Y_i/X_{is}) / \partial(P_{ij}X_{ij}/X_{is})$ are more useful measures than $\partial Y_i/\partial X_{ij}$ or $\partial Y_i/\partial P_{ij}X_{ij}$ since inputs are applied to land, i.e., the marginal value of one dollar of N applied to corn has meaning only if related to a specified land area.

Variation in Input-Output Relationships

Estimating $\partial ATC_{iL} / \partial(P_{ij}X_{ij}/X_{is})$ from cost relationships will not be successful if variations in cost per unit of output are not associated with variations in amount of inputs applied per unit of land, i.e., if all variations in cost per unit of output occur from random variations in soil, weather and/or management. In these situations, input productivity is a meaningless measurement. Some form of the variable, yield, would explain a high proportion of the variation in cost per unit of output. The variable, yield, becomes a proxy for net effects of variations in soil, weather and management.

The opposite occurs if variations in unspecified variables of soil, weather and management have no influence on cost per unit of output and all variations in output are explained by variations in specified inputs. If this occurs, there will be correlation between yield and the respective cost per land unit of inputs. In these situations, if yield is placed in equation (9), high multicollinearity between the yield variable and input variables would exist. If variations in cost per unit of output are explained entirely (or to a large extent) by those in cost per unit of land or the respective inputs, estimation of a production function should be possible and would be a more efficient procedure for estimation of input productivity.

[3]Differences paid for the same quality of input are management variations, as are differences in responses from the same combination of inputs.

If equation (9) was fitted excluding a variable for yield, and yield and one or more $P_{ij}X_{ij}/X_{is}$ were correlated, biased estimates would result.[4] The inclusion of the yield variable would reduce the standard error of one or more of the regression coefficients for costs variables if increase in the variance (due to multicollinearity between the yield variable and cost variables) were less than the decrease in residual variance from inclusion of the yield variable. Estimating the equation with and without the yield variable will indicate differences in standard errors caused by inclusion of the variable. In addition, measures of nonorthogonality, as proposed by Marquardt and Snee, can be made to determine the amount of multicolinearity in the equation.

Equation (10) is linear and $\partial ATC_{iL} / \partial(P_{ij}X_{ij}/X_{is})$ are marginal changes occurring in ATC_{iL} for incremental changes in $P_{ij}X_{ij}/X_{is}$. Although expectations are that $P_{ij}X_{ij}/X_{is}$ is related in some degree and form to the yield variable, this relationship does not influence the partial of ATC_{iL} with respect to $P_{ij}X_{ij}/X_{is}$. Conversely, the linear relationship assumes that within the range of observations $\partial ATC_{iL} / \partial(P_{ij}X_{ij}/X_{is})$ does not change over yield levels.

Excluding the effect of the yield variable, high multicollinearity will exist only if one or more input classes are highly correlated with another and would provide meaningless variable denotation by their separation. For example, if the cost of herbicide is a specified ratio of the cost of fertilizer, it would be more meaningful if these two inputs were combined.

EXAMPLES

Nitrogen Response

The average marginal productivity of nitrogen in a corn fertility response study was estimated using (1) an estimated cost per unit of output relationship, and (2) an estimated production function. Corn production response data were collected from research plots over several years; therefore, the production function included year variables along with first and second degree nitrogen variables and a moisture-nitrogen interaction variable. The production function provided a good fit with experimental data, the first and second degree nitrogen, moisture-nitrogen interaction and several of the year variables being significant. The estimated marginal response from a pound of nitrogen at midpoint rainfall and nitrogen levels was .39 bushel corn.[5]

The cost per unit of output equation was formulated by using a cost of $108.75 per acre plus the cost of nitrogen. The estimated equation was: cost/bushel of corn=$1.86 plus (−.0176) (cost of N/acre). The coefficient −.0176 was significant at the five percent level. The average marginal response from a pound of nitrogen, as estimated from the cost per acre relationship using equation 13, was .42 bushel corn.

The .42 bushel of corn estimate of marginal productivity of N component compares very favorably with the estimate calculated from the production function. That function, though, provides a marginal function, not just an estimate of average marginal response. The nitrogen response example is used for comparison of estimates, not to suggest that the cost per acre technique be used where sufficient data is available for estimation of production functions or surfaces.

Farm Cost Data

Input productivity was estimated from cost data obtained in a survey of Georgia peanut producers. Attempts using linear, quadratic and logarithmic equations to estimate an overall and yield group production functions from these data were unsuccessful. The following cost variables were used in the analyses:

Z_1 = cost per acre of lime, gypsum, and fertilizer
Z_2 = cost per acre of seed
Z_3 = cost per acre of seedbed preparation, planting and weed control
Z_4 = cost per acre of insect control
Z_5 = cost per acre of irrigation
Z_6 = cost per acre of harvesting and drying.

In addition to the cost variables, two other variable were included:

Z_7 = 1/yield per acre
Z_8 = 1/acres of peanuts.

Variable Z_7 was included as a proxy for weather, soil and management and thereby to measure their influence on variations in TC_{i2}/Y_i unassociated with

[4] Johnston, p. 168.

[5] Research analyses reporting the results of estimating corn response to nitrogen with production functions will be in a forthcoming manuscript by W. Lanny Bateman and Fred C. Boswell, Georgia Station, University of Georgia.

variations in the production input variates. To test desirability of including the yield variable, the model was also estimated with Z_7 excluded. The variable, Z_8, was included to test for economies of size.

Observations were divided into yield level subgroups—high, medium and low, on the basis of historical average yields. A separate equation was estimated for each subgroup and for the overall group.

Input Coefficients

The inclusion of Z_7 (1/yield per acre) made a large difference in the coefficients and their standard errors (Table 1). Even in the low yield subgroups were 1/yield per acre explain 95 percent of the variation, the addition of cost variables reduced the unexplained variation in cost per unit of output significantly and all coefficients were significant at the .005 level.[6]

Variance inflation factors were computed for the estimated relationships with and without the yield variable. All values were less than 4.00, which is the most conservative guideline suggested by Snee. (Variance inflation factor of 4.00 means that 75 percent of the variation of the variable is explained by variation of the other "independent" variables.)

Interdependence among explanatory variables is large enough that biased coefficients would result if either one or more cost variables or the yield variables were deleted from the estimation equation. The net effect of soil, weather and management variations are significant in all three yield subgroups.

The added explained variation from the use of three yield group equations as compared to use of one overall equation was significant at the one percent level. The most consistent relationship over all equations was the insignificance of the acreage or size variable.

TABLE 1. REGRESSION COEFFICIENTS AND STANDARD ERRORS WITH AND WITHOUT THE VARIATE 1/YIELD PER ACRE BY PEANUT YIELD SUBGROUPS

	Regression coefficients[bc]									
Yield Variate Excluded Yield Subgroup[a]	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	R^2	r_{yz7}[d]
Low	-.0059	.0030	.0018	.0031	.0017	-.0032	-	.0128	.323	-
	(.0029)	(.0024)	(.0023)	(.0016)	(.0025)	(.0024)	-	(.4987)		
Medium	.00043	.00050	.00066	.00048	.00021	.00014	-	-.0458	.529	-
	(.00023)	(.00022)	(.00033)	(.00016)	(.00028)	(.00013)	-	(.0448)		
High	-.00016	.00035	-.00007	.00009	.00135	.00048	-	-.0804	.468	-
	(.00079)	(.00059)	(.00054)	(.00055)	(.00068)	(.00032)	-	(.0889)		
Yield Variate Included Yield Subgroup[a]										
Low	.00067	.00100	.00057	.00084	.00065	.00056	117.84	.02281	.998	.974
	(.00017)	(.00013)	(.00012)	(.00009)	(.00013)	(.00014)	(1.54)	(.02688)		
Medium	.00041	.00040	.00043	.00035	.00040	.00040	144.85	.01153	.987	.609
	(.000039)	(.000039)	(.000058)	(.000028)	(.000050)	(.000024)	(5.34)	(.00792)		
High	.00049	.00030	.00036	.00028	.00047	.00025	152.70	.02461	.995	.732
	(.000081)	(.000059)	(.000056)	(.000056)	(.000072)	(.000033)	(4.08)	(.00928)		

[a]Number in parenthesis is standard error of the respective regression coefficient.

[b]Average yields of the subgroups were: low—1,243; medium—2,649; and high—2,887 pounds of peanuts per acre.

[c]Input classes for the coefficients are: b_1—lime, gypsum and fertilizer; b_2—seed; b_3—seedbed preparation, planting and weed control; b_4—insect control; b_5—irrigation; b_6—harvesting and drying; b_7—one/yield; b_8—one/yield.

[d]r_{yz7} is the simple correlation coefficient between cost per unit of output and 1/yield per acre. The regression coefficients and their standard errors for the simple relationship between cost per unit and 1/yield are: low—115.92 (5.95); medium—78.34 (35.50); and high—136.19 (28.37).

[6]A comparison was made by generating a series of random numbers for cost, yield and acreages and fitting the equation to the random numbers. The yield variable explained 95 percent of the variation in the cost per unit of output, but there was no relationship between the input cost variables and the cost of output.

Marginal Productivities

The marginal physical products of a dollar unit of the input classes were computed using equation 15 with average values or levels of yield and input costs per acre for the respective yield subgroups (Table 2).

The negative marginal physical productivities in Table 2 are not inconsistent with practices used during the year survey data were obtained. Some producers in the high yield subgroup applied relatively large quantities of fertilizer without realizing higher yields than others in the subgroup who used lower fertilizer rates (fertilizer was relatively cheap and price of peanuts relatively high). The response to gypsum application was not definable, the same range of yields occurring with and without gypsum application. The year in question had a very favorable rainfall distribution, thus peanut growers with irrigation equipment had high fixed costs of irrigation without a yield response. The negative value for seed in the low yield subgroup suggests managerial ability does not justify the same expenditure for seed that would be profitable in medium and high yield subgroups.

The similar marginal physical productivities support the hypothesis that the low yield subgroup farm operators are not facing the same production function or surface as the high or medium yield subgroups. They also suggest that capital rationing and/or unavailability of inputs were not dominant factors in the low yields of the subgroup.

CONCLUSIONS

Estimating productivity through the estimation of average cost functions is a feasible alternative in some situations where production functions cannot be estimated from the data. In these instances, the average cost procedure will provide estimates because point estimates do not require the production function or surface to be specified and because of the ease of adapting a proxy variable such as yield for hard-to-measure influences of soil, environment and management. This proxy variable becomes exceedingly important in those situations where variations of the unmeasured variable strongly overshadow influences of measured variables.

TABLE 2. MARGINAL PHYSICAL PRODUCTIVITIES PER DOLLAR OF INPUT FOR CLASSES OF INPUTS BY PEANUTS YIELD SUBGROUPS[a]

Yield subgroups	Input Variate					
	Lime, gypsum and fertilizer	Seed	Seedbed preparation, planting and weed control	Insect control	Irrigation	Harvesting and drying
	(b_1)	(b_2)	(b_3)	(b_4)	(b_5)	(b_6)
	pounds of peanuts per acre					
Low	1.66	-2.55	2.93	-.51	1.91	3.06
Medium	-.46	.00	-1.37	2.29	0.00	0.00
High	-8.40	2.24	-1.12	3.36	-7.28	5.04

[a]These were estimated for average values or levels of production and input costs per acre within yield subgroups using equation 16:

$$\frac{\partial\left(\frac{Y_i}{X_{is}}\right)}{\partial\left(\frac{P_{ij}X_{ij}}{X_{is}}\right)} = \frac{1}{\left(\frac{TC_{i2}}{X_{is}}\right)}\left[\frac{Y_i}{X_{is}} - \left(\frac{Y_i}{X_{is}}\right)^2\left(\frac{\partial ATC_{iL}}{\partial\frac{P_{ij}X_{ij}}{X_{is}}}\right)\right]$$

NOTE: $\frac{\partial ATC_{iL}}{\partial\left(\frac{P_{ij}X_{ij}}{X_{is}}\right)}$ equals the respective b_{ij}.

REFERENCES

[1] Binswanger, Hans P. "A Cost Function Approach to the Measurement of Elasticities of Factor Demand and Elasticities of Substition," *American Journal of Agricultural Economics*, Volume 56:2, May 1974, pp. 377-385.

[2] Hoch, Irving. "Simultaneous Equation Bias in the Context of the Cobb-Douglas Production Function," *Econometrica*, Volume 26, October 1958, pp. 566-578.

[3] Johnston, J. *Econometrics Methods*, 2nd Edition, New York: McGraw-Hill Book Company, 1972.

[4] Law, Lawrence J. and Pan A. Yotopoulos. "Profit, Supply, and Factor Demand Functions," *American Journal of Agricultural Economics*, Volume 54:1, February 1972, pp. 11-18.

[5] Marquardt, D. W. "Generalized Inverses, Ridge Regression, Biased Linear Estimation and Nonlinear Estimation," *Technometrics*, Volume 12:3, August 1970, pp. 591-612.

[6] Mundlak, Yair and Irving Hoch. "Consequences of Alternative Specifications in Estimation of Cobb-Douglas Production Functions," *Econometrica*, Volume 33, October 1965, pp. 814-828.

[7] Snee, Ronald D. "Some Aspects of Nonorthogonal Data Analysis, Part I. Developing Prediction Equations," *Journal of Quality Technology*, Volume 5:2, April 1973.

[8] Spann, Robert M. "Rate of Return Regulation and Efficiency in Production: An Empirical Test of the Averch-Johnson Thesis," *The Bulletin of Economics and Management Sciences*, Volume 5:1, Spring 1974, pp. 38-52.