CCDP Economics 260

PROBLEMS IN THE ESTIMATION OF AGRICULTURAL PRODUCTION FUNCTIONS

by
Clifford Hildreth

The present discussion will be largely confined to a report of a study recently conducted at Iowa State College. In this study it was desired to estimate the production effects of various resources used by farmers in a fairly homogeneous area in Southern Iowa. Data were available on the operations of a sample group of 161 individual farms in the area during the calendar year 1947. Since the observed variables related to behavior of actual economic units, it was to be expected that they were subject to a number of technical and economic relationships. An approach to the statistical treatment of variables generated by a system of simultaneous relations has been developed by Haavelmo and others. This approach involves three steps. The first is the formulation of a model specifying the a priori assumptions the investigator is willing to make about the variables that enter the equations used to represent the simultaneous relations, the algebraic form of the equations, and the characteristics of error terms in the equations. The second step is the consideration of the extent to which unknown characteristics of the equations can be inferred from knowledge of the distribution of the observed variables. This is called the problem of identification and a parameter of the model is called identifiable if complete knowledge of the distribution of the observed variables would determine the parameter uniquely. These steps are followed by estimation of such of the identifiable parameters as the investigator wishes to estimate. This approach was used to estimate agricultural production functions for Southern Iowa using the data mentioned above. The process is sketched below along with a comparison of the production functions estimated by this approach with regression equations estimated by least squares.

I. The Model

This was developed in collaboration with Wallace E. Ogg of Iowa State College. It consists of a crop production function, a livestock production function and five input-decision relations. The two production functions are -

(1) $y_1 = y_1(w_1, e_1, l_1)$

(2) $y_2 = y_2(w_2, e_2, f_2, l_2)$ where

y_1: Crop output

y_2: livestock output

w_1: crop labor

w_2: livestock labor

e_1: cost of services of crop equipment used

e_2: cost of services of livestock equipment and costs (other than feed) of maintaining the livestock herd

f_2: feed fed to livestock

l_1: cost of using cropland during period

l_2: cost of using livestock land during period

These variables and those introduced below were measured in dollars at average 1947 prices. However, since a dollar's worth of anything is a definite physical quantity so long as a constant set of prices are used, the relationships may be viewed as physical relationships.

It is assumed that l_1 and l_2 are predetermined variables and that the others are jointly determined by the relations of the model. The decision relations have not been written out explicitly but the variables that enter them were derived from the following decision making model -

(i) $U = U(\tilde{I}, E')$

(ii) $\tilde{y}_1 = \tilde{y}_1(w_1, e_1, l_1)$

(iii) $\tilde{y}_2 = \tilde{y}_2(w_2, e_2, f_2, l_2)$

(iv) $\tilde{I} = \tilde{y}_1 + \tilde{y}_2 - (w_1 + w_2 - w_f) - (e_1 - \rho_1 E_1) - (e_2 - \rho_2 E_2) - f_2$
$- \frac{\tilde{y}_1}{2} \left(\frac{\alpha l_1 - L_1}{\alpha l_1}\right) - \tau L_1 - (l_2 - \rho_3 L_2) - \rho_4 D + \rho_5 J$

$$\text{(v)} \quad R' = \eta_1 E_1 + \eta_2 E_2 + \eta_3 F_2 + \eta_4 L_1 + \eta_5 L_2 + \eta_6 J + C - D$$

$$\text{(vi)} \quad \tilde{R}' = R' - \gamma_1(w_1 + w_2 - w_f) - \gamma_2(e_1 - \xi E_1) - \gamma_3 f_2 + (1 - \eta_3)F_2 - \frac{\tau}{2}(L_1 + L_2) + \eta_2 \gamma_3 \tilde{y}_2 + (1 - \eta_2)\gamma_4 \tilde{y}_2 + \gamma_5(\rho_2 J - \rho_1 D)$$

where

U: utility function of entrepreneur during period when input decisions are made

$\tilde{I}$: anticipated income of entrepreneur when input decisions are made

$\tilde{R}'$: anticipated minimum liquidity position during production period

$\tilde{y}_1$: anticipated crop production

$\tilde{y}_2$: anticipated livestock production

w_f: family labor available to entreprenuer during year (includes own labor)

E_1: value of stock of crop equipment owned at beginning of year

E_2: value of livestock equipment and permanent herd at beginning of year

L_1: value of cropland owned

L_2: value of livestock land owned

D: debts of entrepreneur

J: investments of entrepreneur outside the farm enterprise

R': liquidity value of assets at beginning of year

ρ_i, $i=1, \ldots, 5$: rates of interest on investments in various assets. Interest may be received explicitly or imputed.

τ: tax rate on land

α: $\dfrac{1}{\text{interest on land } (\rho_3) + \tau}$; αl_1 is value of cropland used

η_i, $i=1, \ldots, 6$: liquidity or loan ratios for various assets, e.g. $\eta_1 E_1$ is the amount of cash that can readily be borrowed on the entrepreneur's crop equipment.

γ_i, $i=1, \ldots, 5$: proportions of the year's expenses or production to accrue before small grain harvest when the minimum liquidity position for the year is normally reached.

$\xi = \rho_1 + (1 - \gamma_1)\delta$ where δ is the rate of depreciation on crop equipment.

The decision making process consists of choosing w_1, w_2, e_1, e_2, f_2 so as to maximize utility subject to restrictions (ii) to (vi). If the derivatives of U with respect to each of the decision variables are set equal to zero, five decision relations result. If (i) is not assumed to be linear the decision relations will, in general, contain the following variables.-

w_1, w_2, e_1, e_2, f_2 - endogenous

$\left.\begin{array}{l} l_1, l_2, w_f, E_1, E_2, F_2 \\ L_1, L_2, J, C, D \end{array}\right\}$ predetermined

The five predetermined variables in the last row were not observed in this study. J, C, and D were omitted from consideration. Observations were available for the total acres owned and total acres used by each entrepreneur. A variable $T = 1 - \frac{\text{acres owned}}{\text{acres used}}$ was introduced in the belief that it would contain some of the information lost thru not observing L_1 and L_2. Thus the model actually employed contained seven predetermined variables and seven jointly determined variables. Neglect of some of the predetermined variables does not prevent the investigator from obtaining consistent estimates of parameters providing the neglected variables are not needed to establish identifiability of the parameters. (see Koopmans, T. C., "Statistical Methods of Measuring Economic Relationships", Cowles Commission Discussion Papers: Statistics, No. 310, p. 115)

To consider identification and estimation for this model it is convenient to revise the notation as follows -

y_1 - y_1	l_1 - z_1
y_2 - y_2	l_2 - z_2
w_1 - y_3	w_f - z_3
e_1 - y_4	E_1 - z_4
w_2 - y_5	E_2 - z_5
e_2 - y_6	F_2 - z_6
f_2 - y_7	T - z_7

Making a linearity assumption and including parameters and error terms in the production equations puts them in the form -

$$(1') \; y_1 + \beta_{13}y_3 + \beta_{14}y_4 + \gamma_{11}z_1 + \beta_{10} = u_1$$

$$(2') \; y_2 + \beta_{25}y_5 + \beta_{26}y_6 + \beta_{27}y_7 + \gamma_{22}z_2 + \beta_{20} = u_2$$

where the u_1 and u_2 are error terms assumed to have zero means and finite variances and the β's and γ's are parameters to be estimated.

II. Identification

For equation (1') to be identifiable a necessary and sufficient condition (see Koopmans pp. 30-31) is that the matrix containing the coefficients of y_2, y_5, y_6, y_7 and z_2 to z_7 in the six equations other than (1') have rank 6. This matrix has six rows and ten columns. While it is possible that the rows are not linearly independent it seems a reasonable assumption that they are in view of the meaning of the elements and the fact that each row contains ten elements. Identifiability of the equation was, therefore, assumed. Similar considerations apply to equation (2').

III. Estimation

The statistical theory of simultaneous economic relationships as developed in existing literature treats linear models (linear in the coefficients and the dependent variables, the y's). However the arguments can still be applied if it is assumed that the equations are linear in some functions of the observed variables. It has often been assumed that production functions are linear in logs of the observed variables. It seems quite probable that to assume either that the production equations are linear in the observed values or linear in the logs is a fairly severe oversimplification of the true relationship and the most we can hope is that the simplifying assumption will leave us with a reasonable approximation within the range of our data and the range we will contemplate in applying the results.

The assumption of linearity in the logs (constant elasticity) did not seem necessarily better to the investigator than assuming linearity in observed values (constant marginal productivity). Consequently estimates were first obtained under the latter assumption and then under the former. The method used for computing estimates has been called the limited information method by Koopmans (op. cit. pp. 82-125). It had previously been referred to as the reduced form method (see Girshick, M.A. and Haavelmo, Trygve. "Statistical Analysis of the Demand for Food", Econometrica 15-2, p. 79). With this method we are able to obtain consistent estimates of our production equations without writing out our decision relations in detail. The production equations with the estimated parameters included are -

Production Functions Linear

$$(1'')\quad \hat{y}_1 = 405 - .034y_3 + 384y_4 + 2.920z_1$$

$$(2'')\quad \hat{y}_2 = -366 + 1.078y_5 + 5.114y_6 + .432y_7 + .563z_2$$

Production Functions Linear in Logs

$$(1''')\quad \hat{y}_1 = 6.658y_3^{.011}y_4^{.435}z_1^{.358}$$

$$(2''')\quad \hat{y}_2 = 2.646y_5^{.165}y_6^{.494}y_7^{.413}z_2^{.025}$$

IV. Comparison With Single Equation Estimates

Since the above estimates required noticeably longer computations than would have been needed to fit ordinary regressions of the outputs upon their respective inputs, it may be of interest to see if the results differ much from the estimated regression equations. The regression equations obtained from the same data on imputs and outputs are -

Linear Regression Equations

$$(1^*)\ \hat{y}_1 = 91 + .517\, y_3 + .324\, y_4 + 2.671\, z_1$$

$$(2^*)\ \hat{y}_2 = 5434 - .711\, y_5 + .042\, y_6 + .184\, y_7 + 2.177\, z_2$$

Regression Equations Linear in Logs

$$(1^{**})\ y_1 = 2.475\, y_3^{.427} y_4^{.270} z_1^{.328}$$

$$(2^{**})\ y_2 = 1.921\, y_5^{.262} y_6^{.382} y_7^{.380} z_2^{.159}$$

There is a strong general presumption that the true marginal productivities and elasticities are positive and that elasticities are less than unity. Farm management workers familiar with the area sampled also have rather strongly held conclusions about the returns to various resources in the area. Probably the most universally held conclusion is that there is an over supply of labor in the area and that marginal productivity of labor is small. Other fairly strong presumptions about this area are that returns to crop land (z_1), crop equipment (y_4), and livestock herd and equipment (y_6) are relatively high and that returns to livestock land (z_2) are low. Crop yields in 1947 were generally poor, however, and returns to crop inputs were probably substantially lower than in other years. On the whole, the results for equations linear in logs are more consistent with previous presumptions than those linear in observed variables and the structural production functions are more consistent than the regression equations. Because most of our presumptions are in terms of marginal productivities it is a little hard to consider equations (1'''), (2'''), (1**), and (2**) in their above form. They have been used to calculate estimates of marginal productivities of each resource on an "average" farm - i.e. a farm employing an average quantity of each resource. These results are tabulated below; the MP_s column contains marginal productivities estimated from the structural equations and the MP_r column contains those estimated from the regression equations.

Input	MP_s	MP_r
y_3	.019	1.212
y_4	.515	.527
z_1	.949	1.435
y_5	1.320	2.557
y_6	5.620	5.315
y_7	.734	.825
z_2	.326	2.579