COWLES COMMISSION DISCUSSION PAPER: STATISTICS NO. 379

Point Estimates of Ordinates of Concave Functions

Clifford Hildreth

June 10, 1953

I. Introduction.

Economists are frequently in the position of having fairly strong presumptions that relations among variables with which they deal satisfy certain qualitative restrictions, but they seldom have very good grounds for saying that a particular algebraic form is appropriate for representing a given relation. Diminishing marginal productivity of inputs in production relations, downward slope of demand relations and homogeneity of certain demand and production relations are examples of properties of relations which economists often assume.

It is unfortunate that statistical procedures available to economists typically force them either to rather arbitrarily assume that a particular algebraic form is appropriate or to completely ignore their a priori presumptions in their statistical analyses. In this paper a procedure is developed for estimating points on a production surface from data on outputs produced by various combinations of inputs when the inputs are subject to diminishing returns. The procedure is then applied to an example. It is clear that a similar approach could be applied to other types of relations and restrictions.

The present discussion is confined to the problem of obtaining maximum likelihood estimates of points satisfying a relation which contains only one current endogenous variable. Other variables are assumed to be predetermined and independent of any random disturbances which affect the observations. This makes the approach more appropriate for analyzing data from controlled experiments than for analyzing market data or data from surveys of firms or households. In addition to the estimation problem considered here it would be useful to have approaches to testing hypotheses, designing experiments and constructing statistical decision functions1/ when the investigator is concerned with a relation that is somewhat restricted but cannot be assumed to have any simple parametric form.

II. Production Functions with One Variable Input.

Consider a production relation of the form

$$y = \varphi(z) + u \tag{2.1}$$

where y represents output, z represents variable input and u is a random disturbance. In general z may be a vector with as many components as there are types of variable input but for the present we assume that all inputs except one are held constant, or as nearly constant as physical conditions permit. If inputs are defined broadly enough to include all factors influencing output, then variations in u may be attributed to unavoidable and unobserved variations in some of the constant inputs. It is assumed that variations in u approximate independent drawings from a normal distribution with zero mean and finite variance.

An investigator is considered to have observations on y and z for N selected values of z. Let the values of z be arranged in increasing order and denoted by $z_1, z_2, \ldots, z_n, \ldots, z_N$. For each level of input there may be several trials and corresponding observations of output. Let T_n be the number of trials at level of input z_n and let y_{nt} be the observed output for the t-th trial at this level. We have

$$y_{nt} = \varphi(z_n) + u_{nt} \qquad \begin{array}{l} n = 1, 2, \ldots, N \\ t = 1, 2, \ldots, T_n . \end{array} \tag{2.2}$$

An economist with such data is principally interested in the possibility of drawing inferences about which levels of input are most profitable for various combinations of prices of output and of variable input (or conditions of demand for output and supply of input). Frequently such inferences have been drawn by assuming that the function $\varphi(z)$ can be approximated by some given algebraic form with several unknown parameters to be estimated from the data. Estimates of the parameters are inserted into the form to obtain an estimated relation and this estimated relation is then used to calculate most profitable levels of input for chosen combinations of prices.

The chief difficulty with this procedure is that the inferences often depend critically upon the algebraic form chosen. It is not uncommon to find that alternative forms fit the data almost equally well but have very different implications for the most profitable level of input.[2/] It seems useful, therefore, to consider the possibility of drawing inferences without assuming a specific parametric form for $\varphi(z)$.

One thing seems intuitively clear. If the investigator does not wish to make an assumption about the form of the relation and he has data only for selected levels of input, then inferences that he may draw about the profitability of levels of input for which no observations exist will depend on assumptions in addition to those needed to compare the profitability of levels for which data are available. Two approaches are then possible. The investigator can try to formulate reasonable assumptions that will enable him to compare additional levels or he can restrict himself to choosing among the input levels for which data exist and work with a smaller set of assumptions.

The latter approach is adopted here. On this approach an economist who wished to make finer comparisons would encourage further experimentation or data gathering to provide information about additional levels of input. Let η_n be the expected value of output at input level z_n.

(2.3) $$\eta_n = \varphi(z_n) \qquad n = 1, 2, \ldots, N.$$

Our approach will be to try to construct a reasonable procedure for obtaining estimates of the η_n; these can then be translated into estimates of expected profitability of the z_n for any given price combinations. Our estimates will be derived by the method of maximum likelihood. However the same results would have been obtained by least squares, minimum x^2, and possibly other methods. Of course, if there were no a priori restrictions on φ, then the maximum likelihood estimates of the ordinates η_n would just be the means of observed outputs for the appropriate levels of input, i.e.

(2.4) $$\tilde{\eta}_n = \bar{y}_n = \frac{1}{T_n} \sum_{t=1}^{T_n} y_{nt}$$

where the $\tilde{\eta}_n$ might be called limited information maximum likelihood estimates.[3/] To obtain full maximum likelihood estimates (here denoted by $\hat{\eta}_n$) the investigator must maximize the likelihood function subject to all of the a priori restrictions he feels justified in imposing.

As was indicated earlier, in many production situations the investigator will feel that inputs are subject to decreasing returns. This is equivalent to assuming that φ is concave and yields the following restrictions on the ordinates[4/] —

$$(2.5) \qquad \frac{\eta_{n+1} - \eta_n}{z_{n+1} - z_n} \geq \frac{\eta_{n+2} - \eta_{n+1}}{z_{n+2} - z_{n+1}} \qquad n = 1, 2, \ldots, N-2.$$

For such cases it is desired to maximize the likelihood function subject to (2.5). The logarithm of the likelihood function is given by —

$$(2.6) \qquad L(\eta, \sigma^2) = -T/2 \log 2\pi\sigma^2 - \frac{1}{2\sigma^2} \sum_{n=1}^{N} \sum_{t=1}^{T_n} (y_{nt} - \eta_n)^2$$

where $T = \sum_{n=1}^{N} T_n$ and $\eta = (\eta_1 \ \eta_2 \cdots \eta_n)$. Since σ^2 is not restricted its estimator can be obtained by differentiation, yielding[5/] —

$$(2.7) \qquad \hat{\sigma}^2 = \frac{1}{T} \sum_{n=1}^{N} \sum_{t=1}^{T_n} (y_{nt} - \hat{\eta}_n)^2 .$$

The value of η which maximizes L is that which minimizes the double sum on the right of (2.6). We may write

$$(2.8) \qquad \sum_{n=1}^{N} \sum_{t=1}^{T_n} (y_{nt} - \eta_n)^2 = \sum_{n=1}^{N} \sum_{t=1}^{T_n} (y_{nt} - \bar{y}_n)^2 + \sum_{n=1}^{N} T_n (\eta_n - \bar{y}_n)^2$$

The terms of the first summation over n on the right do not depend on η and may be neglected. Let x be an N-dimensional vector with elements

$$(2.9) \qquad x_n = \eta_n - \bar{y}_n \qquad n = 1, 2, \ldots, N.$$

The problem is then to minimize the weighted (by the T_n) sum of squares of the x_n subject to the requirements that the $\hat{\eta}_n$ satisfy (2.5). Equivalent restrictions on the x_n are given by

(2.10)
$$\frac{1}{\Delta_{n+1}} x_n + \left(\frac{1}{\Delta_{n+2}} + \frac{1}{\Delta_{n+1}}\right) x_{n+1} - \frac{1}{\Delta_{n+2}} x_{n+2} - \frac{1}{\Delta_{n+1}} \bar{y}_n + \left(\frac{1}{\Delta_{n+2}} + \frac{1}{\Delta_{n+1}}\right) \bar{y}_{n+1} - \frac{1}{\Delta_{n+2}} \bar{y}_{n+2} \geqq 0$$

where $\Delta_{n+2} = z_{n+2} - z_{n+1}$, $\Delta_{n+1} = z_{n+1} - z_n$, and $n = 1, 2, \ldots, N-2$.

It will simplify the discussion to state the problem in matrix notation

(2.11) Find an $\hat{x} \varepsilon \mathcal{A}$ such that $\hat{x}\, D\, x' \leqq x\, D\, x'$ for all $x \varepsilon \mathcal{A}$ where[6]

(i) $\mathcal{A} : x \ni A\, x' + b' \geqq 0$

(ii)
$$A = \begin{pmatrix} \frac{-1}{\Delta_2} & \left(\frac{1}{\Delta_3} + \frac{1}{\Delta_2}\right) & \frac{-1}{\Delta_3} & 0 & \cdots & 0 & 0 & 0 \\ 0 & -1/\Delta_3 & \frac{1}{\Delta_4} + \frac{1}{\Delta_3} & -1/\Delta_4 & \cdots & 0 & 0 & 0 \\ & & \vdots & & & & & \\ 0 & 0 & 0 & 0 & \cdots & -1/\Delta_{N-1} & \left(\frac{1}{\Delta_N} + \frac{1}{\Delta_{N-1}}\right) & -1/\Delta_N \end{pmatrix}$$

(iii) $b' = A\, \bar{y}'$

(iv)
$$D = \begin{pmatrix} T_1 & & & 0 \\ & T_2 & & \\ & & \ddots & \\ 0 & & & T_N \end{pmatrix}$$

The investigator can readily obtain D, A, b from his input-output data. We seek a way to compute $\hat{x}$ and thereby $\hat{\eta}$. Problems of extremization subject to inequalities have been studied in connection with Activity Analysis and have commonly been found to be equivalent to saddle point problems similar to those encountered in the Theory of Games. By a Theorem of Kuhn and Tucker,[7] the minimization problem stated in (2.11) is equivalent to the following saddle point problem -

(2.12) Find vectors $\hat{x}$, $\hat{v}$ such that $\hat{v} \geqq 0$ and

$$\psi(\hat{x}, v) \leqq \psi(\hat{x}, \hat{v}) = \psi(x, \hat{v})$$

for all x and for all $v \geqq 0$ where

(2.13) $\psi(x, v) = x\,D\,x' + v(A\,x' + b')$.

Some of the general methods that have been developed for minimax problems could doubtless be adapted to this case. However, the following procedure seems exceedingly simple and is used in the example of sec. III.

Clearly a saddle point exists so

(2.14) $\psi(\hat{x}, \hat{v}) = \min\limits_{x} \max\limits_{v} \psi(x, v) = \max\limits_{v} \min\limits_{x} \psi(x, v)$.

Since x is unrestricted, we may find $\min\limits_{x} \psi(x, v)$ by differentiation.

(2.15) $\dfrac{\partial \psi}{\partial x} = 2\,D\,x' + A'v'$

Setting the derivatives equal to zero yields

(2.16) $x' = -\frac{1}{2} D^{-1} A'v'$

and substituting into (2.13),

(2.17) $\min\limits_{x} \psi(x, v) = \psi^*(v) = -\frac{1}{4} v\,A\,D^{-1} A'v' + v\,b'$.

To find the non-negative $N-2$ dimensional vector $\hat{v}$ that maximizes this expression it is convenient to consider an equivalent minimum problem. Let

(2.18) $C = A\,D^{-1} A'$ and

(2.19) $\theta(v) = -2\ \psi^*(v) = \frac{1}{2} v\,C\,v' - 2\,v\,b'$

Clearly $\hat{v}$ minimizes $\theta(v)$. The procedure to be followed in finding $\hat{v}$ is an iterative one. An initial value, say $v^{(o)} = (v_1^{(o)}\ v_2^{(o)} \ldots v_{N-2}^{(o)})$, is chosen as a starting point for the iteration. Holding all components of v except the first fixed at their values given by $v^{(o)}$, the non-negative value of v_1 which minimizes $\theta(v)$ is found. Call this value $v_1^{(1)}$. $\theta(v)$ is then minimized with respect to admissable values of the second coordinate holding v_1 fixed at $v_1^{(1)}$ and v_3 to v_{N-2} fixed at $v_3^{(o)}$ to $v_{N-2}^{(o)}$. This process of minimizing with respect to each coordinate in turn while holding others fixed at their last obtained values is continued until the desired degree of stability is obtained.

Let v_k, $k = 1, 2, \ldots, K$, be a given component of v where, of course, $K = N - 2$. The procedure indicated above can be made more explicit by observing that the minimum of $\theta(v)$ with respect to any v_k is either attained where $v_k = 0$ or where $\frac{\partial \theta}{\partial v_k} = 0$. If the latter equation yields a non-negative value for v_k, then this is the minimizing value, otherwise $v_k = 0$ is the minimizing value. We note

$$(2.20) \qquad \frac{\partial \theta}{\partial v} = C v' - 2 b' .$$

At the p^{th} stage of the iteration, we define $w_k^{(p)}$ as the value of the k^{th} coordinate that would be obtained by setting $\frac{\partial \theta}{\partial v_k} = 0$, i.e.

$$(2.21) \qquad w_k^{(p)} = -\sum_{i=1}^{k-1} \frac{c_{ki}}{c_{kk}} v_i^{(p)} - \sum_{i=k+1}^{k} \frac{c_{ki}}{c_{kk}} v_i^{(p-1)} + 2 \frac{b_k}{c_{kk}}$$

where the c_{ki} are elements of C and $k = 1, 2, \ldots, K$.

The value of the k^{th} coordinate of v at the p^{th} stage of the iteration is then obtained by taking

(2.22) $$v_k^{(p)} = \max (w_k^{(p)}, 0).$$

For the production problem described above it is convenient to start the process by setting $v^{(o)} = 0$.

To show that this process converges to a unique minimum we supplement the notation slightly and note several properties.[8] Let the sequence of vectors be generated by the process be denoted by $\left\{v^{(m)}\right\}$, i.e.

(2.23)
$$
\begin{aligned}
v^{(1)} &= (v_1^{(1)} \quad 0 \quad 0 \ldots 0) \\
v^{(2)} &= (v_1^{(1)} \; v_2^{(1)} \quad 0 \ldots 0) \\
&\vdots \\
v^{(K)} &= (v_1^{(1)} \; v_2^{(1)} \; v_3^{(1)} \ldots v_K^{(1)}) \\
v^{(K+1)} &= (v_1^{(2)} \; v_2^{(1)} \; v_3^{(1)} \ldots v_K^{(1)}) \\
&\vdots \\
v^{(pK+k)} &= (v_1^{(p)} \ldots v_k^{(p)} \; v_{k+1}^{(p-1)} \ldots v_K^{(p-1)}) \\
&\vdots \\
&\text{etc.}
\end{aligned}
$$

Clearly the sequence $\left\{\theta(v^{(m)})\right\}$ converges. It is non-increasing and bounded. We know that it is bounded because C in (2.19) is positive definite (A has K linearly independent rows).

It is also clear that $\theta(v)$ has an unique minimum for v in the closed positive orthant. The minimum over the whole space is unique and might fall in the positive orthant. If not, then the line joining any two admissable minimum points is also admissable (the positive orthant is convex). Any point on this line gives a lower $\theta(v)$ than the extremes, hence the extremes could not be minimum points.

We should next like to show that

$$(2.24)\quad \lim_{p \to \infty} \left| v_k^{(p-1)} - v_k^{(p)} \right| = 0 \quad \text{for all } k.$$

Let $m = p\,K + k$. Then in passing from $v^{(m-1)}$ to $v^{(m)}$, only the k^{th} coordinate of v changes. Consider the following -

$$(2.25)\quad \theta(v^{(m-1)}) - \theta(v^{(m)}) = \tfrac{1}{2}\, c_{kk}\left(v_k^{(p-1)^2} - v_k^{(p)^2}\right)$$

$$+ \left(v_k^{(p-1)} - v_k^{(p)}\right)\left[\sum_{i=1}^{k-1} c_{ki}\, v_i^{(p)} + \sum_{i=k+1}^{K} c_{ki}\, v_i^{(p-1)} - 2\, b_k\right].$$

Using (2.21) we obtain

$$(2.26)\quad \theta(v^{(m-1)}) - \theta(v^{(m)}) = \frac{c_{kk}}{2}\left(v_k^{(p-1)} - v_k^{(p)}\right)\left(v_k^{(p-1)} + v_k^{(p)} - 2\, w_k^{(p)}\right)$$

$$\geqq \frac{c_{kk}}{2}\left(v_k^{(p-1)} - v_k^{(p)}\right)^2 .$$

To justify the mixed inequality we note that if $w_k^{(p)} \geqq 0$ then $v_k^{(p)} = w_k^{(p)}$ and the equality holds. If $w_k^{(p)} < 0$, then $v_k^{(p)} = 0$ and $\left(v_k^{(p-1)} - 2\, w_k^{(p)}\right) > v_k^{(p-1)} \geqq 0$ and the mixed inequality holds. (2.24) follows from (2.26) and the convergence of $\left\{\theta(v^{(m)})\right\}$.

Let $P_k(v)$ be the vector obtained from v by holding all but the k^{th} component fixed and minimising $\theta(v)$ with respect to v_k. P_k is a continuous mapping of K dimensional Euclidean space into itself. In our sequence, $\left\{v^{(m)}\right\}$ we have

$$(2.27)\quad v^{(m)} = P_{m \sim K}(v^{(m-1)}) \quad \text{where } m \sim K \text{ means } m \text{ modulo } K.$$

Let $\theta_{min} = \theta(v_{min})$ be the value of our function at its minimum for non-negative v. Let θ_∞ be the limit of our sequence $\left\{\theta(v^{(m)})\right\}$. We should like to show that $\theta_\infty = \theta_{min}$. The vector sequence $\left\{v^{(m)}\right\}$ is bounded. In particular the ellipsoid given by $\theta(v) = \theta(v^{(o)})$ contains the ellipsoids given by $\theta(v) = \theta(v^{(m)})$ and thus bounds the sequence. $\left\{v^{(m)}\right\}$ therefore has at least one limit point and contains a subsequence which converges to this limit. Let v^∞ be a limit of $\left\{v^{(m)}\right\}$ and let $\left\{v^{(r)}\right\}$ be a subsequence approaching v^∞. For each r identifying an element of the subsequence, let $m(r)$ identify the same element in the original sequence.

From the continuity of θ,

(2.28) $$\theta_\infty = \theta(v^\infty).$$

We shall show that to suppose $\theta(v^\infty) \neq \theta_{min.}$ involves a contradiction.

If the supposition is true then it is possible to reduce $\theta(v)$ by changing a coordinate of v^∞, i.e.

(2.29) $$\exists\, k \ni P_k(v^\infty) \neq v^\infty.$$

Let $\underline{K}$ be the set of all such k and let

(2.30) $$\varepsilon = \min_{k \in \underline{K}} \left[\theta(v^\infty) - \theta(P_k(v^\infty))\right].$$

Let $\underline{V}$ be the set of all v in the convex set bounded by the ellipsoid $\theta(v) = \theta(v^o)$. $\theta(v)$ is uniformly continuous over $\underline{V}$, hence

(2.31) $$\exists\, \delta \ni \left\|v - v^*\right\| > \delta \implies \left|\theta(v) - \theta(v^*)\right| < \varepsilon$$

for all $v, v^* \in \underline{V}$.

We proceed to show that our original vector sequence, $\left\{v^{(m)}\right\}$, contains an element $P_k(v^{(m)})$ within δ of $P_k(v^\infty)$ for a $k \in \underline{K}$. It will then follow from (2.31) that $\left|\theta(P_k(v^{(m)})) - \theta(P_k(v^\infty))\right| < \varepsilon$. Since $\theta(P_k(v^\infty))$ is at least ε below θ_∞, $\theta(P_k(v^{(m)}))$ must also be less than θ_∞. But since θ_∞ is the limit of $\theta(v^{(m)})$, this contradicts the supposition that $\theta(v^\infty) \neq \theta_{min}$.

From the continuity of the P_k we know

$$(2.32) \qquad \exists \ \rho \ni \left\| v - v^\infty \right\| < \rho \Longrightarrow \left\| P_k(v) - P_k(v^\infty) \right\| < \delta$$

for all k.

From (2.24) we know that successive elements of $\left\{v^{(m)}\right\}$ can be made arbitrarily close together by making m sufficiently large. We also know that if r is sufficiently large, elements of the subsequence $\left\{ v^{(r)} \right\}$ lie arbitrarily close to v^∞. Specifically we may say

$$(2.33) \quad \exists \ M \ni m > M \Longrightarrow \left\| v^{(m+1)} - v^{(m)} \right\| < \frac{\rho}{K+1}$$

$$\exists \ R \ni r > R \Longrightarrow \left\| v^{(r)} - v^\infty \right\| < \frac{\rho}{K+1}.$$

Now consider an $\mathring{r}$ such that $\mathring{r} > R$, $\mathring{m} = m(\mathring{r}) > M$. The K elements of $\left\{v^{(m)}\right\}$ immediately following $v^{(\mathring{m})}$ are all within ρ of v^∞. At least one of them is obtained from the preceeding by applying P_K with $k \in \underline{K}$. Such an element, $P_k(v^{(\mathring{m}+i)})$ with i an integer between 0 and $K-1$, lies within δ of $P_k(v^\infty)$ and reduces $\theta(v)$ below θ_∞, i.e.

$$(2.34) \quad \left\| v^{(\mathring{m}+i)} - v^\infty \right\| < \rho \qquad \text{for some } 0 \leq i \leq K-1$$

such that $v^{(\mathring{m}+i+1)} = P_k(v^{(\mathring{m}+i)})$ for $k \in \underline{K}$.

$$(2.35) \quad \left\| P_k(v^{(\mathring{m}+i)}) - P_k(v^\infty) \right\| < \delta \qquad \text{by (2.32)}$$

$$(2.36) \quad \left| \theta(P_k(v^{(\mathring{m}+i)}) - \theta(P_k(v^\infty)) \right| < \varepsilon \qquad \text{by (2.31)}$$

$$(2.37) \quad \theta(P_k(v^{(\mathring{m}+i)}) < \theta_\infty \qquad \text{by (2.30).}$$

III. An Illustrative Computation.

The primary purpose of the illustration is to show how the computing procedure developed in the previous section can be applied. While data from actual experiments are used, I have not examined the original reports of these experiments and do not have any firm judgment about the appropriateness of combining data from these various experiments in the simple analysis proposed here. For this reason I do not try to discuss the economic implications of the data but merely use them to illustrate a computing procedure.

The data are taken from corn fertility experiments conducted at North Carolina State College.[9/] Corn yields that resulted from various applications of nitrogen fertiliser are available. Paul R. Johnson [10/] has used the results for fitting production functions under several alternative assumptions about the algebraic form of the function. Prior to his analysis Johnson made an attempt to select experiments that would yield observations of yields under fairly similar conditions in all respects except level of nitrogen applied. Only results from plots with closely related soil types and from years of "good" weather were used.[11/] Johnson's data consisted partly of direct observations and partly of interpolated values. Only the direct observations are used in the present calculation.

In addition to approximating the underlying functional relation, Johnson was interested in the optimal application of nitrogen when nitrogen costs 0.137 per lb. and corn sells for $1.75 per bushel. While all of his equations fitted the data reasonable well, they differed substantially in their implications for the most profitable level of nitrogen. This is shown in Table I where the first column shows the relations with estimated parameters filled in and the second column indicates the corresponding level of nitrogen for highest profit.

Table I

Fitted Equations and Optimal Inputs from Johnson Study

Fitted Equation	Nitrogen Application to Maximize Profit
$y = 4.504\ (z + 20)^{0.59}$	5340
$y = 25.16 + .7595\ z - .00209\ z^2$	164
$y = 108 - 82.48\ (.9897)^z$	230
$y = 109 - 85.53\ e^{-.0108\ z}$	228

y represents yield in bushels, z represents nitrogen in lbs.

Unless one has considerable a priori confidence in the appropriateness of a particular algebraic form, there is considerable uncertainty about the implication of the data for the decision as to level of input. To illustrate the alternative procedure developed in sec. II, the observations are first listed in Table II.

TABLE II

OBSERVED YIELDS AT SPECIFIED LEVELS OF NITROGEN

NITROGEN (lbs./acre) (z_n)	0	20	40	60	80	120	160	180
	9.9	43.4	44.9	52.2	79.0	72.0	81.5	74.7
	31.3	27.3	40.2	66.0	68.6	74.1	72.9	110.3
	32.0	35.3	96.9	74.0	59.8	78.8	117.1	102.7
	24.2	42.2	52.1	64.3	81.7	107.0	102.3	120.9
YIELDS	18.8	35.7	85.1	77.3	107.1	102.5	114.3	103.9
(Bu./Acre)	25.0	50.1	63.6	34.0	48.5	68.7	70.2	98.2
(y_{nt})	2.8	56.0	77.3	58.5	94.6	78.1	83.9	70.7
	17.4	42.1	63.6	49.5	101.8	12.4	104.9	70.7
	14.6	42.1		62.2	94.6	74.0	113.9	
	25.8			50.1		69.8	104.9	
	13.3					85.0		
	33.0					80.8		
	19.8					73.0		
	25.0					100.5		
	63.9					115.5		
	24.4					92.1		
	22.8					83.9		
	51.7					96.3		
	11.6					96.3		
	14.4							
	20.8							
	19.1							
	7.2							
	16.6							
	55.2							
	3.7							
	15.2							
SUM	619.5	374.2	523.7	588.1	735.7	1560.8	965.9	752.1
No. of Obs. (T_n)	27	9	8	10	9	19	10	8
Mean ($\bar{y}_n$)	22.94	41.58	65.46	58.81	81.74	82.15	96.59	94.01

From these data, the following are readily computed -

$$(3.1)\quad D^{-1} = \begin{pmatrix} \frac{1}{T_1} & 0 & - - - & 0 \\ 0 & \frac{1}{T_2} & - - - & 0 \\ & & \vdots & \\ 0 & 0 & - - - & \frac{1}{T_N} \end{pmatrix} = \begin{pmatrix} .0370 & & & & & \bigcirc & & \\ & .1111 & & & & & & \\ & & .1250 & & & & & \\ & & & .1000 & & & & \\ & & & & .1111 & & & \\ & \bigcirc & & & & .0526 & & \\ & & & & & & .1000 & \\ & & & & & & & .1250 \end{pmatrix}$$

$$(3.2)\quad A = \begin{pmatrix} -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -2 & 3 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 3 & -2 \end{pmatrix}.$$

The meanings of the restrictions are not changed if any row of A is multiplied by any positive constant. This sometimes makes it possible to choose convenient numbers for elements. The A above differs from that of (2.11) in that the above has been obtained by multiplying the first three rows by 20 and the last three rows by 40. Continuing

$$(3.3)\qquad b' = A\,\bar{y}' = \begin{pmatrix} -5.24 \\ 30.53 \\ -29.58 \\ 75.03 \\ -14.03 \\ 19.60 \end{pmatrix}$$

$$(3.4)\quad C = AD^{-1}A' = \begin{pmatrix} .6064 & -.4722 & .1250 & 0 & 0 & 0 \\ -.4722 & .7111 & -.4500 & .2000 & 0 & 0 \\ .1250 & -.4500 & .6361 & -.7333 & .1111 & 0 \\ 0 & .2000 & -.7333 & 1.4525 & -.4385 & .0526 \\ 0 & 0 & .1111 & -.4385 & .4215 & -.4052 \\ 0 & 0 & 0 & .0526 & -.4052 & 1.4526 \end{pmatrix}$$

From these, formulae for the $w_k^{(p)}$ are readily obtained.

$$(3.5)\quad w_1^{(p)} = .7787\, w_2^{(p-1)} - .2061\, w_3^{(p-1)} + 17.28$$

$$w_2^{(p)} = .6640\, w_1^{(p)} + .6328\, w_3^{(p-1)} - .2812\, w_4^{(p-1)} - 85.86$$

$$w_3^{(p)} = -.1965\, w_1^{(p)} + .7074\, w_2^{(p)} + 1.1527\, w_4^{(p-1)} - .1746\, w_5^{(p-1)} + 93.00$$

$$w_4^{(p)} = -.1377\, w_2^{(p)} + .5048\, w_3^{(p)} + .3019\, w_5^{(p-1)} - .0362\, w_6^{(p-1)} - 103.30$$

$$w_5^{(p)} = -.2636\, w_3^{(p)} + 1.0403\, w_4^{(p)} + .9613\, w_6^{(p-1)} + 66.57$$

$$w_6^{(p)} = -.0362\, w_4^{(p)} + .2789\, w_5^{(p)} - 26.99$$

From (3.5) the values of $v_k^{(p)}$ shown in Table III resulted when $v^{(o)}$ was taken equal to the zero vector.

TABLE III

SUCCESSIVE VALUES OF $v_k^{(p)}$

k \ p	1	2	3	4	5
1	17.28	0	0	0	0
2	0	0	0	0	0
3	89.60	85.50	85.31	85.30	85.30
4	0	0	0	0	0
5	42.95	44.03	44.08	44.08	44.08
6	0	0	0	0	0

Applying (2.16) we have

$$(3.6) \quad x' = -\frac{1}{2} D^{-1} A' v' = \begin{pmatrix} 0 \\ 0 \\ -5.33 \\ 8.53 \\ -7.19 \\ 2.32 \\ -2.20 \\ 0 \end{pmatrix}$$

and recalling (2.9), the maximum likelihood estimates of the ordinates are given by

$$(3.7) \quad \hat{\eta}' = \hat{x}' + \bar{y}' = \begin{pmatrix} 22.94 \\ 41.58 \\ 60.13 \\ 67.34 \\ 74.55 \\ 84.47 \\ 94.39 \\ 94.01 \end{pmatrix}$$

These are shown together with the original observations and their means in Figure 1.

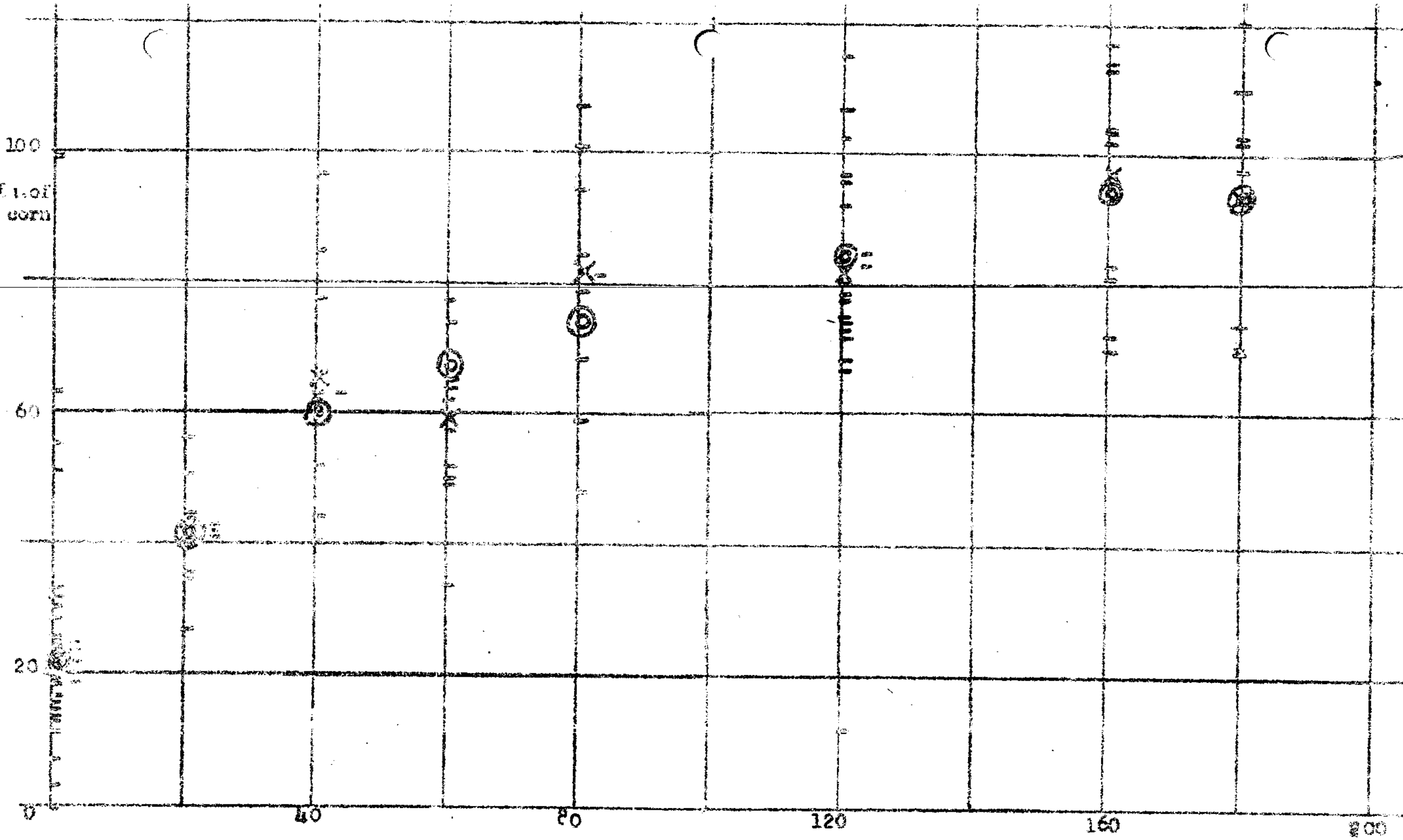

Figure 1. Observations and Estimates

- Observed yield (y_{nt})

* Mean yield ($\bar{y}_n$)

Maximum likelihood estimate of Expected yield ($\hat{\eta}_n$)

While a comparison of profitability of various applications is subject to the qualifications mentioned at the beginning of the section, it may be worthwhile to note that, at prices of $1.75 for corn and .137 for nitrogen, 160 lbs would be better than the other levels according to our estimates. To get a good determination of optimal input for about this price ratio there should clearly be more observations in the 120-220 lbs. interval, they should be more closely spaced, and weather effects should be taken into account (see fn. 11).

The approach developed in sec. II could, with minor modifications, be applied to a variety of problems of estimating an unknown coordinate of points on a surface about which the investigator has qualitative information. It is easy to imagine cases in which translation of the qualitative information into restrictions on the domain of the likelihood function would be rather complex. It is also possible that cases will arise in which the iterative procedure illustrated above does not converge to the minimum and more general computing procedures will have to be used.

The proof of convergence depended essentially on the existence of a unique (local and general) minimum, the boundedness of $\left\{v^{(m)}\right\}$, and the continuity of $\theta(v)$ and its first derivatives. The facts that the admissable set was convex and that the quadratic form in the minimand was positive definite made the first two properties easy to establish. If the number of restrictions on the η_n exceeds N, then the dual problem will involve minimization of a semi-definite quadratic form. Intuitively it seems that the procedure outlined should work for many such cases but they need to be studied and the necessary properties made explicit.

FOOTNOTES

1/ Approaches to design of experiments and construction of decision functions are needed for those situations in which the economist has a voice in the planning of the experiment and/or the gathering of the data. The present treatment is appropriate for those cases in which the economist can only analyze data that have already been assembled.

2/ The study by Paul R. Johnson cited in sec. III below offers one illustration. Others are not easy to find in the literature because many investigators try several forms for their relations but report only the one that in some sense "looks best" a posteriori.

3/ For an analogous use of this term in another context see T. C. Koopmans and William C. Hood, "The Estimation of Simultaneous Linear Economic Relations," Studies in Econometric Method, Cowles Commission Monograph 14, edited by T. C. Koopmans and W. C. Hood, John Wiley and Sons, New York.

4/ The investigator may typically feel justified in assuming strict concavity in which case the strict inequalities would hold in (2.5). However, if these were imposed the likelihood function would have no maximum. Using the restrictions given is equivalent to finding the least upper bound of the likelihood function in the region defined by corresponding strict inequalities.

5/ This estimator may be expected to have a substantial downward bias when the T_n are small. Its distribution has not yet been investigated.

6/ In the special case in which the input levels are equally spaced and the same number of trials exist for each level we may take $D = I$ and

$$A = \begin{pmatrix} -1 & 2 & -1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 & 0 & 0 \\ & & & \vdots & & & & \\ 0 & 0 & 0 & 0 & \cdots & -1 & 2 & -1 \end{pmatrix}$$

since multiplication of either D or A by a positive constant does not change the problem.

7/ H. W. Kuhn and A. W. Tucker, "Nonlinear Programming," Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, Jerzy Neyman, editor, University of California Press, 1951. For the parts relevant to the present application see p. 487 and the concluding section, pp. 491, 492.

8/ The main features of this proof were suggested by Roy Radner.

9/ The experiments are summarized in B. A. Krantz, "Fertilize corn for Higher Yields," Bulletin 366 of the North Carolina Agricultural Experiment Station, Raleigh, 1946.

10/ Johnson's work is reported in a mimeographed paper, "An Economic Analysis of Corn Fertilization in the Coastal Plain of North Carolina," issued by the North Carolina Agricultural Experiment Station. As material for his study Johnson used both direct observations of inputs and yields and some interpolations. The latter have been excluded from the computation presented here.

11/ "Good" weather is attributed by Johnson to years in which "the rainfall distribution was about normal and soil moisture conditions were not low enough to cause the leaves to roll during this period [a five-week critical period including the time of tasseling]." In principle one should take account of the weather effect in the statistical specification if it is believed to have significantly affected the observed yields. If this were done and the weather and input effects were assumed to be additive, (2.2) would be replaced by

$$(2.2') \qquad y_{nmt} = \eta_n + \delta_m + u_{nmt}$$

where $m = 1, 2, \ldots, M$ is an index of the year of a particular observation and δ_m is the weather effect in that year. Let T_{nm} be the number of observations at input level z_n in year m. If the T_{nm} are equal for all n and m, then the weather effect causes no significant complication. If we impose the natural requirement that $\sum_m \hat{\delta}_m = 0$, we have

$$\text{(i)} \qquad \hat{\delta}_m = \frac{\sum_n \sum_t y_{nmt}}{\sum_n T_{nm}} - \frac{\sum_n \sum_m \sum_t y_{nmt}}{\sum_n \sum_m T_{nm}}$$

from $m = 1, 2, ----, M$. The $\hat{\eta}_n$ can be obtained by minimizing

$$\text{(ii)} \qquad s = \sum_n \left(\eta_n - \frac{\sum_m \sum_t y_{nmt}}{\sum_m T_{nm}} \right)^2 \qquad \text{subject to the restrictions}$$

in (2.5) and the procedure developed in sec. II applies directly. If T_{nm} are unequal, as in the present case, the situation is more complicated. It seemed undesirable to introduce these complications in the present illustration especially since an effort had previously been made to select homogeneous years.