CLIFFORD G. HILDRETH
North Carolina State College

Chapter 4

Discrete Models With Qualitative Restrictions

IN statistical analyses, as in many other human endeavors, the product of a particular undertaking is closely related to the input. At the final stage of a statistical application, what one puts in are some observations and a specification; what one gets out are some statistical inferences, i.e., estimates, tests, and/or optimal decisions. Ways in which good observations contribute to useful inferences are generally well understood and are quite properly stressed in most applied statistics courses. The possible contributions, positive or negative, of alternative specifications are not as easily understood and, for many problems, have not been adequately explored by statistical theorists.

Specifications

Since the rationale for the procedure to be outlined and illustrated depends entirely on considerations of specification, a few general remarks on these matters may be helpful. First, a statistical specification is defined as the complete set of assumptions which are accepted as a basis for a particular statistical investigation. Another way of putting this is to say that a specification includes all statements about the underlying statistical population which the investigator accepts a priori.

Specification and model are nearly synonymous terms. According to a fairly well accepted usage (1, 6) observed here, the model is the class of all statistical populations which are consistent with the specification, i.e., which satisfy the a priori assumptions.[1] For most of the discussion the terms will be used interchangeably.

In general, an investigator's situation is such that if he adds assumptions to his specification (narrows his model), the prospective accuracy of his inferences is increased, provided the added assumptions are realistic. However, if the assumptions are unrealistic, biased inferences will generally result. Thus, a researcher should clearly use in his specification all of the relevant a priori information that he is sure is realistic.[2] In doubtful cases, the investigator may be helped by

[1]This statement would require some modification in contexts in which one needs to distinguish the statistical population from the theoretical structure which explains it. Such instances have arisen mainly in economics and psychology and need not be taken into account in the following discussion.

[2]Sometimes a researcher may ignore potentially useful a priori information to simplify computations. This possibility is left aside to keep from diverting the discussion.

theoretical research indicating the extent to which a particular added assumption may improve the inferences to be drawn and, on the other hand, the biases to which particular errors in the assumption will lead. If the possible biases are large relative to potential gains, a doubtful assumption should, of course, be rejected. If the prospects are reversed, a doubtful assumption might well be utilized. Considerable reliance upon the judgment of the investigator is unavoidable in all but the most routine applications, and good judgment combined with technical skill is what makes a good applied statistician.

That the contribution of a priori information differs from one problem to another may be observed by considering estimates obtained from a random sample from a normal population. If the investigator is primarily interested in a good estimate of the population variance, he may improve his estimate by specifying the population mean a priori, if it is known. This specification will substantially improve his estimate of the variance, if he has only a few observations, but will only be a slight improvement if he has many observations. Thus, if he (a) has a fair a priori notion of the mean but does not know it exactly, and (b) has a small number of observations, he might very well use his best a priori value for the mean; otherwise he may neglect his a priori notion.[3] On the other hand, if the investigator is primarily interested in estimating the mean, a priori knowledge of the variance is not of any help.

Clearly the difficult case is the one in which an uncertain assumption (a) may improve the analysis significantly if correct, and (b) damage it badly if incorrect.[4] A thorough knowledge of the field of application should help the research worker to judge the likelihood of bias. Sometimes a test of significance can be developed as an additional aid to judgment. However, this precaution has often been pointed out; i.e., to test an assumption and then use it (if not rejected) as part of the specification on which subsequent estimates and tests are based complicates the interpretation of the traditional probability statements that are later made about test statistics or confidence regions. While this statement is undeniable, it should not seriously inhibit use of preliminary tests. The basic difficulty is not that a preliminary test is performed but that the investigator is under pressure to utilize an uncertain assumption. Proceeding without attest does not remedy this basic difficulty.

The particular specification problem with which we shall be concerned is that of formulating appropriate assumptions about the form of a response surface. For convenience, a certain observable response, y, depends upon the magnitudes of certain observable, and sometimes controllable, variables, $z_1, z_2 \ldots z_k$, and certain unobservable variables whose net effect may be approximately represented by a random variable, u. The unobserved variables may be partly controllable, especially in carefully conducted experiments. The assignment of the z's

[3] If his a priori information could be put in the form of a distribution function for the population mean, and the weight function for various possible errors in the estimate of the variance were taken into account, this could be handled as a statistical decision problem.

[4] A simple but suggestive example has been presented by Leonid Hurwicz (4).

may be randomized to assure that u is independent of the z's, and conditions can sometimes be held sufficiently stable from one observation to another that u will have a small variance.

Form of Equation

Familiar statistical procedures give the investigator two types of alternatives. He may assume a priori that an equation of a certain known form will represent the surface to a close approximation and use the observations to estimate several unknown parameters in the equation. Alternatively, he may forego the assumption as to form and regard each distinct combination of the z's as a different treatment, unrelated to the others in his statistical model. These alternatives correspond to the continuous and discrete models discussed by Anderson in Chapter 3. To use a discrete model it is necessary to make some specifications about the form of the function. Also, assumptions must be made about the way in which the random component, the u, enters. It is usually found desirable to make some assumptions about the interactions of the z's. There are, of course, an infinite number of models for each type of interaction from which an investigator might choose.

Advantages and Disadvantages of Continuous Models

Continuous models offer several potential advantages. There may be a substantial gain in efficiency in having a small number of parameters to estimate and in estimating response at a particular point (a particular combination of the z's) from all of the observations rather than just the observations at that point. The estimated equation provides a convenient means of interpolation and limited extrapolation. Furthermore, the form of the relation, once it is well established, may have interesting theoretical implications.

The principal disadvantage of continuous models lies in the biases which may accrue if an inappropriate form is used, and the difficulty of designing a satisfactory test of the appropriateness of a particular assumption regarding the form. It is particularly disconcerting that, in many instances in which several alternative assumptions have been investigated, alternative fitted equations have resulted which differ little in terms of conventional statistical criteria, such as multiple correlation coefficients or F tests of the deviation, but differ much in their economic implications (cf. 5, 9). It is also worth noting that bias due to inappropriate form does not decrease as sample size increases,[5] whereas inefficiencies in discrete or form-free methods become less important in large samples. In many contexts the convenience of interpolation offered by a continuous function may not be very important. Frequently the discrete alternatives analyzed will be sufficiently numerous to

[5]In general, bias will decrease if the range of the observations is increased along with sample size and, of course, can also be decreased by changing the assumed form as discrepancies become apparent.

determine an optimal decision to the degree of accuracy permitted by the data. In addition, when results of analyses are put to practice, there will always be relevant discrepancies between the conditions underlying the analysis and the conditions faced in commerical production on farms. Some judgment will of necessity be exercised at this stage; interpolation may be as effective as using a predetermined formula.

As noted earlier, there are many situations in which choosing a specification involves delicate judgment and a thorough knowledge of the particular field of application. Where judgment plays a large part, two different researchers may use somewhat different models and procedures without any existing way of labeling one, right, and the other, wrong. Instead of seeking "the" way to proceed in such instances, mathematical statisticians might better try to give the applied worker the means for employing any of a variety of models and procedures, thus enlarging the area over which judgment can be exercised.

Situations sometimes arise, for production economics analysis as elsewhere, in which the investigator does not find either the continuous or traditional discrete type of model to be ideal. He may feel that no particular form of function has been sufficiently well established in his area to give reasonable assurance against bias in a continuous model. He may have rather firm notions about some properties underlying the relation. These properties are ignored if he treats distinct input combinations as unrelated treatments. An economist might, for instance, strongly believe that a particular production function is characterized by diminishing returns; that a certain demand equation is homogeneous; that a certain supply curve slopes upward. To the extent that he knows these properties exist, it is wasteful to analyze statistical results that are inconsistent with them. For such situations it might be useful to have procedures enabling the researcher to include in his specification such qualitative properties as seem sufficiently well established, without forcing him to specify his relation as completely as a continuous model requires.

A Discrete Model

A possible approach is to formulate discrete models which include the appropriate qualitative restrictions and to work out appropriate statistical procedures for these models. Appropriate procedures can be found for a variety of such models. In an article by Hildreth (2), procedures were developed for obtaining estimates of points on a production surface under the assumption that inputs are subject to diminishing returns.[6] The work is now being extended and, while it is highly incomplete, a sketch of accomplishments may serve to suggest possibilities of the approach and the kinds of problems, mostly unsolved, which are encountered in using it.

[6] This exposition is marred by the inclusion of a hastily attempted generalization which can be shown to be false. A correction may be found in the December 1955 issue of the Jour. Amer. Stat. Assn. Fortunately, the false generalization does not affect the main result or the applications which have been developed.

The extensions have been worked out jointly by the author and A. P. Stemberger. They will be more fully reported in Stemberger's doctoral thesis. The data come from experiments on the response of corn yields to nitrogen, conducted by Krantz and Chandler (7).

The model initially employed was of the following form:

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

where the N-observed levels of nitrogen have been arranged in ascending order and z_n is the pounds per acre in the n-th level ($z_{n+1} > z_n$). y_{nt} is the observed yield for the t-th trial (observation) with application z_n. T_n is the number of plots to which z_n pounds have been applied. u_{nt} is a random disturbance assumed to be independent of z_n.

The algebraic form of the production or response function, ρ (z), is regarded as unknown except that successive increments of z are assumed to increase y at a nonincreasing rate. In other words ρ (z) is concave, or $\frac{d^2y}{dz^2} \leq 0$ if the derivative exists. With only this assumption regarding form it is not generally possible to estimate the response to levels of nitrogen other than those (N in number) for which observations are available.[7]

Since there is no loss of generality in taking E (u_{nt}) = O, the following may be written:

$$(2) \qquad \eta_n = E\ (y_{nt}) = \rho\ (z_n).$$

The assumption of diminishing returns then requires:

$$(3) \qquad \frac{\eta_{n+2} - \eta_{n+1}}{n+2 - n+1} \leq \frac{\eta_{n+1} - \eta_n}{z_{n+1} - z_n} \qquad n = 1,2\ \ldots\ N\text{-}2.$$

Regarding the η_n as the magnitudes to be estimated, the application of the method of maximum likelihood (if the u_{nt} are normally distributed) or the method of least squares leads to the problem of finding estimates, $\hat{\eta}_n$, which minimize the sum of squares:

$$(4) \qquad Q = \sum_{n=1}^{N} \sum_{t=1}^{T_n} (y_{nt} - \hat{\eta}_n)^2$$

[7] It is possible to estimate upper and lower bounds for all z such that $z_1 < z < z_N$; upper bounds can be estimated for $z > z_N$ or $z < z_1$.

when the restrictions, equation 3 above, believed to hold for the population parameters, are also required to hold for the estimates.

Thus the estimation problem is one of minimizing a positive definite quadratic form subject to constraints in the form of inequalities. Problems like this have been studied in activity analysis and in game theory. With the aid of a theorem by Kuhn and Tucker (8), it was possible to develop an iterative procedure for obtaining the required estimates.[8]

The use of this procedure to obtain yield estimates from the Krantz-Chandler data is described in the article mentioned previously. At the time of the estimates, only data pertaining to "good" weather and one type of soil were available. When access to the complete data was obtained, it was found that numerous other observations were available covering weather experience classified into three main categories: good, fair, and dry. Also, several soil types were available which could be placed in three fairly homogeneous classes: Piedmont, Coastal, and Drained Coastal.

The problem of using all of the data in a unified analysis was similar to problems sometimes encountered in combining data from different experiments. The model was modified to allow for soil and weather effects and could then be indicated:

(5) $$y_{ijnt} = \alpha + \beta_i + \gamma_j + \eta_n + u_{ijnt}$$

$$\begin{aligned} l &= 1,2,3 \\ j &= 1,2,3 \\ n &= 1,2 \ \ldots \ 12 \\ t &= 1,2 \ \ldots \ T_{ijn} \end{aligned}$$

where:

y_{ijnt} = the t-th yield observed on soil i with weather j and nitrogen level n.

α = a general constant

β_i = the contribution to yield of soil i

γ_j = the contribution to yield of weather j

η_n = the contribution to yield of applying z pounds of nitrogen

u_{ijnt} = a random disturbance

T_{ijnt} = the number of observations with soil i, weather j, and level of nitrogen n.

The twelve levels of nitrogen were in 20-pound intervals from 0 to 220, inclusive.

[8]The computing procedure developed may also be used to solve a number of nonlinear programming problems, including some involving monopoly and risk elements.

Interaction Among Soil, Water, and Fertilizer

The model indicated by 5 assumes no interaction among soil, weather, and nitrogen effects. With observed yield as the dependent variable, this would mean, for instance, that dry weather should cut yield the same number of bushels on heavily fertilized plots as on lightly fertilized plots, and similarly for other effects. This assumption is not entirely plausible. A somewhat more promising possibility is the assumption that a change in weather has the same percentage effect on plots with various combinations of soil and fertilizer. To modify these assumptions regarding interaction, log y_{ijnt} is substituted for y_{ijnt} in equation 5.

For convenient future reference, write:

$$Y_{ijnt} = \alpha + \beta_i + \gamma_j + \eta_n + u_{ijnt} \tag{6}$$

where Y_{ijnt} = log y_{ijnt} and other symbols have meanings similar to their meanings in equation 5, except that the constants are now logs of factors in an expression for observed yield. Equation 6 is equivalent to

$$y_{ijnt} = A^{\alpha + \beta_i + \gamma_j + \eta_n + u_{ijnt}} \tag{7}$$

where A is the base of the system of logarithms used.

For several reasons it seemed desirable to initially analyze both equations 5 and 6 without imposing restrictions on the η_n. Before doing this it seemed reasonable to test the interaction assumption in equation 5. The restrictions on the η_n in equation 6 which would express diminishing marginal productivity are nonlinear; direct estimation of the coefficients of equation 6, subject to restrictions, would be even more difficult. While the interaction assumption implicit in equation 6 seems more plausible a priori than that in equation 5, it still seemed desirable to test this assumption before deciding what other analyses might be worthwhile. The data on which the analyses are based are given in table 4.1.

The tests for interaction confirmed the a priori belief that equal percentage effects were more plausible than equal absolute effects. The test showed significant interaction in equation 5 at the 0.01 level,[9] whereas the test applied to equation 6 shows no significant interaction, as can be seen in table 4.2. Accordingly, further analysis was confined to equation 6. The estimates of coefficients for equation 6 are given in table 4.3.

All of the indicated F ratios are significant at the 0.001 level, except for interaction which is not significant at the 0.05 level. In testing for

[9]For equation 5, the interaction mean square was 364.08, within cells mean square was 189.24, giving an F of 1.92. Degrees of freedom are 39 and 182 as in equation 6. The assistance of R. L. Anderson in performing these tests is gratefully acknowledged.

TABLE 4.1 Corn Fertilization Data

	Levels of Nitrogen in Pounds											
	0	20	40	60	80	100	120	140	160	180	200	220
	A - Piedmont Soil											
	(Dry)	(Dry)	(Dry)	(Dry)	(Dry)	(Dry)	(Dry)	(Dry)				
	18.0	29.9	44.5	50.4	50.4	51.0	52.1	51.1	-	-	-	-
	33.3	41.4	51.1		62.3		60.2					
	35.6	48.5	37.0		37.7		39.2					
	31.7	42.0	54.9		54.5		57.8					
			63.8		67.1		74.5					
Cell means	29.6	40.4	50.3	50.4	54.4	51.0	56.8	51.1				
	B - Drained Coastal Soil											
	-	(Dry)	-	-	(Dry)	-	-	(Dry)	-	-	-	-
	-	27.6			50.2			61.4				
	-	(Fair)	-	-	(Fair)	-	-	(Fair)	-	-	-	-
		66.0			86.5			88.9				
	(Good)	(Good)	(Good)		(Good)	-	(Good)	(Good)	(Good)	-	-	(Good)
	50.1	39.8	62.9		77.7		89.0	102.8	86.6			90.7
		80.8	63.5		78.1		114.3		90.0			
		22.2	57.0		111.5				123.2			
		84.4			110.2				123.9			
					70.3				86.7			
					90.6				99.4			
					102.4				114.5			
Cell means	50.1	56.8	61.1	-	91.5	-	101.6	102.8	103.5	-	-	90.7
	C - Coastal Soil											
	(Dry)	(Dry)	(Dry)	(Dry)	(Dry)	(Dry)	(Dry)	(Dry)	(Dry)	(Dry)	(Dry)	(Dry)
	9.4	21.0	-	37.8	48.5	53.4	43.9	52.5	59.0	45.5	59.2	41.5
	21.3	44.6		52.6	60.8	60.5	61.3	40.0		68.6	63.0	81.2
	4.3	22.1		50.6	48.8	50.6	52.7	66.4		58.5	75.3	70.7
	14.4	29.6		42.0	57.5	68.2	80.9	62.6		50.7	61.8	70.7
	15.2	31.6		56.0		40.9	62.2			64.8		
	3.7			56.9								
	20.4											
	17.2											
	5.4											
	3.5											
	14.7											
	39.5											
Cell means	14.1	29.8	-	49.3	53.9	54.7	60.2	55.4	59.0	57.6	64.8	66.0
	(Fair)	(Fair)	(Fair)	(Fair)	(Fair)	(Fair)	(Fair)	(Fair)	(Fair)	-	(Fair)	-
	9.9	43.4	39.1	52.2	60.3	81.0	72.0	83.6	85.5		81.0	
	31.3		24.9	66.0	59.5		74.1	88.2			85.8	
	24.2			64.3	81.8		78.8					
	15.2			74.0	74.8							
	12.9											
Cell means	18.7	43.4	32.0	64.1	69.1	81.0	75.0	85.9	85.5		83.4	
	(Good)	(Good)	(Good)	(Good)	(Good)	(Good)	(Good)	(Good)	(Good)	(Good)	(Good)	(Good)
	19.8	35.7	40.2	59.0	59.8	77.2	80.8	86.9	81.5	115.7	72.9	107.8
	63.9	50.1	96.9	77.3	102.2	86.7	73.0		117.1	95.1	108.0	116.5
	22.8	56.0	52.1	58.5	81.7	83.6	117.0		102.3	110.3		74.7
	51.7	42.0	85.1	34.0	107.1	98.3	100.5		114.3	102.7		
	18.8	31.8	63.6	50.1	80.0	108.3	115.5		101.0	120.9		
	17.4	27.3	44.9	49.5	69.0		92.1		98.3	103.9		
	2.8	42.2	49.1	62.2	94.6		83.9		97.8	98.2		
	13.3	35.3			79.0		90.6		104.9	128.6		
	14.6				68.6		96.3		70.2			
	24.4				88.1		107.0					
	11.6				59.0		102.5					
	20.8				94.4		78.1					
	19.1				90.2		68.7					
	7.2				79.5		69.8					
	16.6						62.4					
	6.9											
	8.3											
	9.2											
	7.7											
	25.8											
Cell means	19.3	40.1	61.7	55.8	82.4	90.8	89.2	86.9	98.6	109.4	90.5	99.7
Treatment means	19.4	41.9	54.7	55.2	74.8	71.6	78.4	71.3	97.6	89.5	75.9	81.7

Soil means		Weather means		General means	
$\bar{y}_{1..}$	47.7	$\bar{y}_{\cdot 1.}$	44.5	$\bar{y}_{...}$	61.7
$\bar{y}_{2..}$	81.7	$\bar{y}_{\cdot 2.}$	55.0		
$\bar{y}_{3..}$	60.1	$\bar{y}_{\cdot 3.}$	72.3		

TABLE 4.2. Analysis of Variance for Equation 6

Source of Variation	d.f.	S.S.	M.S.	F. Ratio
Mean	1	691.358	-	-
Regression	15	17.882	1.192	47.11
Soils	(2)	(.364)	.182	7.28
Weather	(2)	(1.136)	.568	22.45
Nitrogen	(11)	(15.438)	1.403	55.45
Error	221	5.598	.0253	-
Interaction	(39)	(1.023)	.0261	1.04
Within cells	(182)	(4.576)	.0250	-
Total	237	714.839	-	-

TABLE 4.3. Estimates of Coefficients in Equation 6

Coefficient		Estimated Coefficient	Estimated Standard Error of Coefficient	Antilog of Coefficient
Symbol	Interpretation			
α	General constant	1.2408	.0270	17.41
β_1	Piedmont soil	.12090	.0397	1.321
β_2	Drained Coastal soil	.08394	.0333	1.213
β_3	Coastal soil	0	0	1.000
γ_1	Dry W_X	-.17836	.0270	.663
γ_2	Fair W_X	-.02589	.0344	.942
γ_3	Good W_X	0	0	1.000
η_1	0# Nitrogen	0	0	1.000
η_2	20# Nitrogen	.38941	.0415	2.451
η_3	40# Nitrogen	.48193	.0470	3.034
η_4	60# Nitrogen	.56049	.0449	3.635
η_5	80# Nitrogen	.63435	.0370	4.309
η_6	100# Nitrogen	.67796	.0521	4.764
η_7	120# Nitrogen	.67381	.0382	4.719
η_8	140# Nitrogen	.66649	.0552	4.640
η_9	160# Nitrogen	.71981	.0467	5.246
η_{10}	180# Nitrogen	.75580	.0507	5.699
η_{11}	200# Nitrogen	.72706	.0618	5.334
η_{12}	220# Nitrogen	.73117	.0617	5.385

interaction, the within cells sum of squares was placed in the denominator. It has 182 degrees of freedom because only 55 of the 72 cells have any observations from which to estimate cell means. The error mean square has been used as the denominator for the other F ratios. This has been done so that both the estimates and the tests, other than the test for interaction itself, would be based on the same specifications. The adjusted R^2 is 0.72.

The estimates of coefficients are given in table 4.3, along with estimated standard errors and antilogs. The coefficients of equation 6 are not unique. The meaning of the equation would be unchanged if a constant were added to α, and the same constant were subtracted from all of the β_i, or all of the γ_j, or all of the η_n. This makes it possible to select arbitrary values for one coefficient for each type of effect. B_3, γ_3 and η_1 were set equal to zero.

The antilogs indicate how estimated yields change from one soil-weather combination to another. In going from Coastal soil to Piedmont, 32.1 percent was added to the estimated yield regardless of weather and nitrogen; in going from good weather and Coastal soil to fair weather and drained Coastal soil, 14.3 percent was added (1.213 x .942 - 1 = .143), etc.

It was desirable to obtain an estimate of the nitrogen effects subject to the diminishing returns restrictions. This estimate was complicated by the fact that cell frequencies were highly disproportionate and by the fact that the restrictions on the log of yield are nonlinear. The first difficulty is perhaps not too serious. Since the restrictions apply only to the nitrogen effects and since interaction is not significant, it seems a reasonable conjecture that imposing the restrictions would affect the soil and weather coefficients very little. The estimates of these coefficients are, in any case, unbiased but would be somewhat more efficient if estimated subject to the restrictions.

One might proceed by correcting the original observations on logs of yield by the estimated soils and weather effects and then re-estimate the nitrogen effects, treating these corrected values as observations. This procedure would go quite smoothly except for the second complication — the nonlinearity of the restrictions on log of yield. While the estimates subject to nonlinear inequalities can be developed, time has not been available, and therefore the author will not speculate as to how much the computations would be increased.[10] An approximation to the results

[10]It appears that quadratic restrictions would suffice for this problem.

Let $y = \Upsilon(x)$, $Y = \log y$

$$\frac{dY}{dx} = y^{-1}\frac{dy}{dx}$$

$$\frac{d^2Y}{dx^2} = y^{-1}\frac{d^2y}{dx^2} - y^{-2}\left(\frac{dy}{dx}\right)^2 = y^{-1}\frac{d^2y}{dx^2} - \left(\frac{dY}{dx}\right)^2.$$

Since y is positive

$$\frac{d^2y}{dx^2} < 0 \Longleftrightarrow \frac{d^2y}{dx^2} + \left(\frac{dY}{dx}\right)^2 < 0.$$

Thus, imposing the condition on the right is equivalent to imposing the condition on the left. While this relation only holds exactly at a point, its interval analogue will be sufficiently close for practical purposes and this will involve quadratic restrictions on the treatment effects in the log form.

that would be obtained under this procedure can be found by converting the corrected logs of yields back to yields and proceeding as in the original problem cited.

(8) $$y^*_{nk} = 10\,(y_{ijnt} - \hat{\beta}_i - \hat{\gamma}_j)$$

where k runs from 1 to K_n, and K_n is the number of observations at the n-th level of nitrogen ($K_n = \sum_i \sum_j T_{ijn}$). Then, choose estimates $\breve{\eta}_n$ to minimize the sum of squares.

(9) $$Q^* = \sum_n \sum_k (y^*_{nk} - \breve{\eta}_n)^2$$ subject to the restrictions

(10) $$\frac{\breve{\eta}_{n+2} - \breve{\eta}_{n+1}}{z_{n+2} - z_{n+1}} = \frac{\breve{\eta}_{n+1} - \breve{\eta}_n}{z_{n+1} - z_n} \qquad n = 1,2 \ldots N-2.$$

This procedure is not quite consistent with the assumptions implicit in equation 6 since the sums of squares of deviations are minimized in yields rather than in logs of yields. However, a comparison of restricted and unrestricted estimates in table 4.4 confirms that the error is not large. Estimates are presented for good and dry weather and Coastal soil. To obtain the estimate, either restricted or unrestricted, for any other soil-weather class and for any level of nitrogen, one could

TABLE 4.4. Estimates Responses to Nitrogen for Coastal Soil and Two Types of Weather

Nitrogen Level (inPounds)	Good Weather Estimates		Dry Weather Estimates	
	Unrestricted	Restricted	Unrestricted	Restricted
0	17.41	21.75	11.54	14.42
20	42.68	44.30	28.30	29.37
40	52.81	54.91	35.01	36.41
60	63.11	65.30	41.84	43.29
80	75.01	75.70	49.73	50.19
100	82.95	80.85	55.00	53.60
120	82.15	84.78	54.47	56.21
140	80.78	88.70	53.56	58.81
160	91.33	92.63	60.55	61.41
180	99.22	96.56	65.78	64.02
200	92.88	95.95	61.58	63.61
220	93.80	95.34	62.19	63.21

multiply the estimate from table 4.4 for good weather by the product of the antilogs, from table 4.3, of the coefficients for the desired soil-weather combination.

To become a generally useful tool, estimation subject to qualitative restrictions needs to be developed in several directions. Better procedures for handling transformation of variables are needed. It would be useful to have confidence regions and tests which take account of the restrictions.[11] As more variables are restricted, improved computational procedures will be needed.

Even when these developments take place, the procedures should be regarded as supplementing rather than supplanting existing techniques. There will still be the advantages of efficiency and convenience attached to continuous models when the appropriateness of a particular algebraic form can be rigorously established. However, criteria for goodness of fit are needed that take account of the implications to be drawn from fitted relations.

Certain other improvements in statistical capabilities are needed irrespective of the type of model chosen. In crop production studies, more effective procedures are needed for incorporating data on the initial condition of the soil into models and for relating response to specific observable weather variables.

There is one additional topic that should be mentioned, viz., the drawing of economic implications from our results. After estimating a continuous production surface for an economic unit, the natural procedure is to form a net revenue function with prices of inputs and outputs appearing as variables. This can be maximized with respect to inputs and outputs yielding the optimal quantity as a function of all of the prices. When the economic unit is a firm, these equations are the individual firm's supply and demand functions. More generally, these might be designated as the optimal decision relations.

When the analysis takes the form of estimation of response to a set of discrete alternatives, the natural analogue to the functions described above is a construction of a price map (3). If all possible prices of inputs and outputs are considered as points in a multidimensional Euclidean space, then the price map is a partitioning of this price space into regions which correspond to the production alternatives in such a way that a particular alternative (or combination of alternatives in extended analyses) is optimal whenever the actual price combination falls inside the corresponding region. A price map corresponding to the restricted estimates in table 4.4 is shown in figure 4.1. The procedure for determining regions is the same as that used for cotton fertilization data in the reference cited previously. Crosses show the price combinations which actually prevailed (on the average) in North Carolina in the indicated years.

[11] It should be recognized that conventional tests which ignore the restrictions are unbiased even when the restrictions are known to hold. Utilizing the restrictions would generally increase the power of our tests.

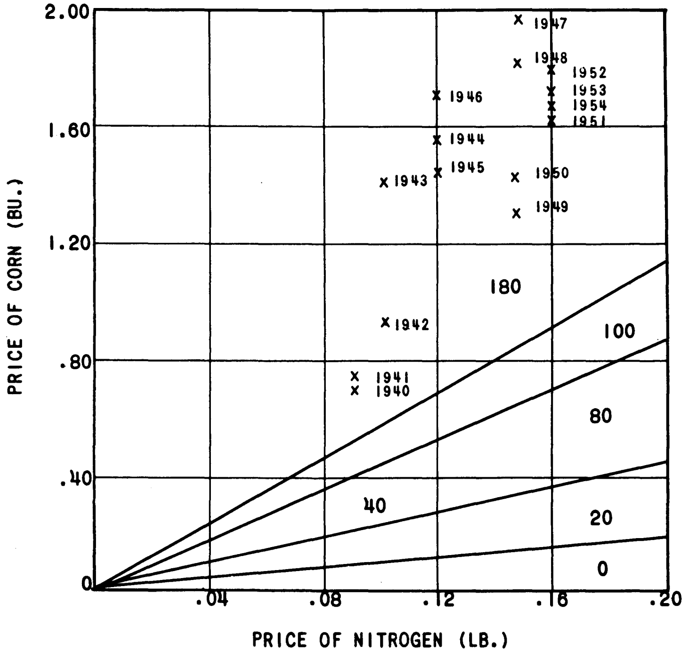

Fig. 4.1 — Corn-nitrogen price map.

Experimenters have increasingly accepted the desirability of taking statistical considerations into account, in planning their investigations, and of examining the statistical implications of their results. It now appears that a good start is being made toward assigning economic considerations and implications of their proper role. Actually, a set of optimal decision relations or its discrete counterpart, a price map, might well be regarded as just as necessary to a complete report of an investigation as the analysis of variance table.

References Cited

1. HAAVELMO, T., 1944. The probability approach in econometrics, Econometrica Suppl. 12:31-47.

2. HILDRETH, C., 1954. Point estimates of ordinates of concave functions. Jour. Amer. Stat. Assn. 49:598-619.

3. ———, 1955. Economic implications of some cotton fertilizer experiments. Econometrica 23:88-98.

4. HURWICZ, L., 1951. Some specification problems and applications to econometric models. (Abstract.) Econometrica 19:343.

5. JOHNSON, P. R., 1953. Alternative functions for analyzing a fertilizer-yield relationship. Jour. Farm. Econ. 35:519-21.

6. KOOPMANS, T. C., 1949. Identification problems in economic model construction. Econometrica 17:52-63.

7. KRANTZ, B. A., and CHANDLER, W. V., 1954. Fertilize corn for higher yields. N. C. Agr. Exp. Sta. Bul. 366 (revised).

8. KUHN, H. W., and TUCKER, A. W., 1951. Nonlinear programming. Proc. 2nd Berkeley Symposium on Math. Stat. and Probability. University of California Press.

9. PRAIS, S. J., 1953. Nonlinear estimation of the Engel curves. Rev. Econ. Stu. 20:102.