THE CHOICE OF THE DEGREE OF A POLYNOMIAL REGRESSION AS A MULTIPLE DECISION PROBLEM[1]

By T. W. Anderson

Columbia University

0. Summary. On the basis of a sample of observations, an investigator wants to determine the appropriate degree of a polynomial in the index, say time, to represent the regression of the observable variable. This multiple decision problem is formulated in terms used in the theory of testing hypotheses. Given the degree of polynomial regression, the probability of deciding a higher degree is specified and does not depend on what the actual polynomial is (except its degree). Within the class of procedures satisfying these conditions and symmetry (or two-sidedness) conditions, the probabilities of correct decisions are maximized. The optimal procedure is to test in sequence whether coefficients are 0, starting with the highest (specified) degree. The procedure holds for other linear regression functions when the independent variates are ordered. The problem and its solution can be generalized to the multivariate case and to other cases with a certain structure of sufficient statistics.

1. The problem. A frequent problem in regression analysis is to determine how many independent variables to include in the fitted regression function. In some cases the independent variables are ordered in importance or usefulness. We consider here the example in which the independent variables are successive powers of the observation index, say time. Usually, if a particular power is included in the regression function, all lower powers are included. In this note we study the problem of determining how many powers to include, that is, the degree of the polynomial regression.

This problem is typical of some other multiple decision problems in which the alternatives are ordered or ranked. After formulating and solving the problem of the choice of polynomial regression, we shall see that the ideas apply to other situations in which there is a certain structure of sufficient statistics.

Let $y_1, \cdots, y_T$ be the observed "dependent" variables, normally and independently distributed with common variance σ^2 and expected values

$$\mathcal{E}y_t = \gamma_0\phi_0(t) + \gamma_1\phi_1(t) + \cdots + \gamma_q\phi_q(t), \qquad t = 1, \cdots, T, \tag{1}$$

where $\phi_i(t)$ is a polynomial of degree i. In case the independent variables are ordered such as powers of t, one can orthogonalize them without effecting the regression function (except changing the coefficients). We take the polynomials

Received May 8, 1961; revised October 18, 1961.

[1] This research was sponsored by the Office of Naval Research under Contract Number Nonr-266(33), Project Number NR 042-034, and Contract Number Nonr-3279(00). Reproduction in whole or in part is permitted for any purpose of the United States Government.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Mathematical Statistics.* www.jstor.org

orthogonal and normalized; that is,

$$\sum_{t=1}^{T} \phi_i^2(t) = 1, \tag{2}$$
$$\sum_{t=1}^{T} \phi_i(t)\phi_j(t) = 0, \qquad i \neq j.$$

The estimates of the coefficients are

$$c_i = \sum_{t=1}^{T} y_t\phi_i(t), \qquad i = 0, \cdots, q. \tag{3}$$

These are independently normally distributed with common variance σ^2 and means $\mathcal{E}c_i = \gamma_i$. The estimate of σ^2 is based on

$$\begin{aligned} S &= \sum_{t=1}^{T} [y_t - c_0\phi_0(t) - \cdots - c_q\phi_q(t)]^2 \\ &= \sum_{t=1}^{T} y_t^2 - \sum_{i=0}^{q} c_i^2, \end{aligned} \tag{4}$$

and S/σ^2 is distributed according to the χ^2-distribution with $T - (q + 1)$ degrees of freedom. These $q + 2$ quantities constitute a sufficient set of statistics; the problem could be formulated in these terms (which constitute essentially the "canonical form").

The investigator may assume he needs a polynomial of degree at least m (which may be 0), and thus he does not question whether any of $\gamma_0, \gamma_1, \cdots, \gamma_m$ are 0. He may assume the maximum degree he will need is q. It is an advantage to represent the trend by a polynomial of low degree because the curve is smoother, the presumed "explanation" is simpler and the function is more economical. However, if the underlying mean value of the observed variable is not approximately a polynomial of low degree the investigator will want to use a polynomial of higher degree. When the degree has been established, the investigator will usually not ask whether coefficients of lower degree terms are 0. (If he does question the value of some coefficient after the degree has been established, it is for some other purpose, such as testing a theory, rather than for the purpose considered here, which is finding the simplest polynomial representation. See [6], p. 49, for example.)

The investigator has a multiple decision problem, namely, choosing whether the degree is $m, m + 1, \cdots, q - 1$, or q. We can formalize this by saying that he wants to decide to which of the following mutually exclusive sets the parameter point $(\gamma_{m+1}, \cdots, \gamma_q)$ belongs:

$$\begin{aligned} &H_q: \gamma_q \neq 0, \\ &H_{q-1}: \gamma_q = 0,\ \gamma_{q-1} \neq 0, \\ &\vdots \\ &H_{m+1}: \gamma_q = \cdots = \gamma_{m+2} = 0, \gamma_{m+1} \neq 0, \\ &H_m: \gamma_q = \cdots = \gamma_{m+1} = 0. \end{aligned} \tag{5}$$

The set H_i implies that the polynomial is of degree i. An alternative formulation is that he wishes to decide which (if any) of the following null hypotheses are true:

$$
\begin{aligned}
H_q^* &: \gamma_q = 0, \\
H_{q,q-1}^* &: \gamma_q = \gamma_{q-1} = 0, \\
&\vdots \\
H_{q,q-1,\cdots,m+1}^* &: \gamma_q = \cdots = \gamma_{m+1} = 0.
\end{aligned} \tag{6}
$$

If any hypothesis of (6) is true, the preceding hypotheses must be true, and if any hypothesis is false the succeeding ones must be false; that is,

$$H_{q,q-1,\cdots,m}^* \subset \cdots \subset H_q^* .$$

The two sets of sets are related by

$$
\begin{aligned}
H_q^* &= H_{q-1} \cup \cdots \cup H_m , \\
H_{q,q-1}^* &= H_{q-2} \cup \cdots \cup H_m , \\
&\vdots \\
H_{q,q-1,\cdots,m+1}^* &= H_m .
\end{aligned} \tag{7}
$$

We suppose that the investigator wants to control directly the probabilities of errors of saying coefficients are not zero when they are or correspondingly of choosing a higher degree than necessary. We assume that the investigator assigns a significance level to each null hypothesis:

$$
\begin{aligned}
p_q &= \Pr\{\text{Reject } H_q^* \mid H_q^*\}, \\
p_q + p_{q-1} &= \Pr\{\text{Reject } H_{q,q-1}^* \mid H_{q,q-1}^*\}, \\
&\vdots \\
p_q + \cdots + p_{m+1} &= \Pr\{\text{Reject } H_{q,q-1,\cdots,m+1}^* \mid H_{q,q-1,\cdots,m+1}^*\} \\
&= \Pr\{\text{Reject } H_m \mid H_m\},
\end{aligned} \tag{8}
$$

where $p_i \geqq 0$ and $p_q + \cdots + p_{m+1} \leqq 1$. Since one null hypothesis includes the next (that is, each subsequent null hypothesis is more stringent), the probabilities of rejection are taken monotonically nondecreasing (that is, the probability of rejecting a more restricted null hypothesis when it is true is greater than that of rejecting a less restricted one). In terms of the mutually exclusive categories the specification is

$$
\begin{aligned}
p_q &= \Pr\{\text{Accept } H_q \mid H_{q-1} \cup \cdots \cup H_m\}, \\
p_{q-1} &= \Pr\{\text{Accept } H_{q-1} \mid H_{q-2} \cup \cdots \cup H_m\}, \\
&\vdots \\
p_{m+1} &= \Pr\{\text{Accept } H_{m+1} \mid H_m\}.
\end{aligned} \tag{9}
$$

Let $p_m = 1 - p_q - \cdots - p_{m+1} = \Pr\{\text{Accept } H_m \mid H_m\}$. Another way of putting the formulation is that when the degree of polynomial needed is less than i, the investigator specifies the probability of deciding to use a polynomial of degree at least i (for each i).

A statistical procedure for this multiple decision problem consists of a set of $q - m + 1$ (mutually exclusive and exhaustive) regions in the space of $c_0, \cdots, c_q$ and S (or in the original space of $y_1, \cdots, y_T$), say $R_m, R_{m+1}, \cdots, R_q$ if the sample point falls in R_i then H_i is accepted. The assignment of significance levels implies that these regions are "similar regions" in the sense that when $\gamma_i = \cdots = \gamma_q = 0$ the probabilities of the sample point falling in $R_i, \cdots, R_q$ are $p_i, \cdots, p_q$, respectively (independent of $\gamma_0, \cdots, \gamma_{i-1}$ and σ^2); that is, when the degree of the polynomial needed is less than i the probability of making an error of saying the degree should be at least i does not depend on what that lower degree polynomial is.

Since it seems reasonable that it does not matter to the investigator whether a nonzero coefficient is positive or negative, we ask that the probabilities associated with the procedure not be affected by changing the sign of any of the coefficients in question; that is, the probabilities depend on $|\gamma_{m+1}|, \cdots, |\gamma_q|$.

Subject to the above requirements we ask for the best regions in the sense that we want to maximize the probability of R_i when H_i is true, $i = m + 1, \cdots, q$. It should be noticed that we are trying to maximize simultaneously the probabilities of $q - m$ different regions (each for all nonzero values of the corresponding parameter). It will have to be shown that under the above conditions the maximized probabilities of one region are not affected by the choice of another. This fact permits us to optimize $R_{m+1}, \cdots, R_q$ simultaneously.

2. The optimal procedure. First, we note that when $\gamma_{i+1} = \cdots = \gamma_q = 0$ the best test of the hypothesis $\gamma_i = 0$ at a given significance level α_i has the rejection region

$$(10) \qquad r_i = \frac{c_i^2}{c_{i+1}^2 + \cdots + c_q^2 + S} > \frac{t_{T-i-1}^2(\alpha_i)}{T - i - 1},$$

where $t_{T-i-1}(\alpha_i)$ is the significance point of the t-distribution with $T - i - 1$ degrees of freedom corresponding to the (two-sided) significance level α_i. The test is best in the sense that it is the uniformly most powerful test based on a similar region with power depending on $|\gamma_i|$, that is, not depending on the sign of γ_i.

Secondly, we note that if T_i is a similar region for testing the hypothesis $\gamma_i = 0$, that is,

$$(11) \qquad \Pr\{T_i \mid \gamma_i = \gamma_{i+1} = \cdots = \gamma_q = 0\} = \alpha_i$$

for all $\gamma_0, \gamma_1, \cdots, \gamma_{i-1}$ and $\sigma^2 (>0)$, then T_i cuts out conditional probability α_i on almost every combination of specified values of $c_0, c_1, \cdots, c_{i-1}$ and $c_i^2 + \cdots + c_q^2 + S$ (the sufficient statistics for $\gamma_0, \gamma_1, \cdots, \gamma_{i-1}$ and σ^2 when $\gamma_i = \cdots = \gamma_q = 0$); that is,

$$(12) \quad \Pr\{T_i \mid c_0, c_1, \cdots, c_{i-1}, c_i^2 + \cdots + c_q^2 + S; \gamma_i = \cdots = \gamma_q = 0\} = \alpha_i$$

almost everywhere (or with probability 1). In fact, some of the optimum properties of the t-test noted above are based on this "Neyman structure" of similar regions (see Lehmann [3], Sec. 4.3, for example). The requirements (9) (or equivalently (8)) imply that each R_i has such structure, in fact

$$R_i \cup R_{i+1} \cup \cdots \cup R_q ,$$

the region of rejection of $H^*_{q,q-1,\cdots,i}$, has such a structure for each $i(i = m + 1, \cdots, q)$. We make use of this fact to show that the choice of $R_q, \cdots, R_{i+1}$ (subject to (9)) does not affect the choice of R_i in the sense that the probability of R_i (a function of $|\gamma_i|$) when $\gamma_{i+1} = \cdots = \gamma_q = 0$ does not depend on how $R_q, \cdots, R_{i+1}$ are chosen. Note that we are interested in γ_i when $\gamma_{i+1} = \cdots = \gamma_q = 0$, and if one of $\gamma_{i+1}, \cdots, \gamma_q$ is not 0 (that is the degree is greater than i) we are not interested in γ_i and may assume $\gamma_i \neq 0$.

Lemma 1. *Let S_{i+1} be a set in the space of $c_0, \cdots, c_q$ and S such that*

$$\text{(13)} \qquad \Pr\{S_{i+1} \mid \gamma_{i+1} = \cdots = \gamma_q = 0\} = p_q + \cdots + p_{i+1},$$

and let T_i be a set defined by $c_0, \cdots, c_i$ and $c_{i+1}^2 + \cdots + c_q^2 + S$. Then

$$\text{(14)} \qquad \begin{aligned} &\Pr\{\bar{S}_{i+1} \cap T_i \mid \gamma_{i+1} = \cdots = \gamma_q = 0\} \\ &\qquad = (1 - p_q - \cdots - p_{i+1}) \Pr\{T_i \mid \gamma_{i+1} = \cdots = \gamma_q = 0\}, \end{aligned}$$

where $\bar{S}_{i+1}$ is the complement of S_{i+1}.

Proof. The requirement (13) is that S_{i+1} is a similar region (with respect to $\gamma_0, \cdots, \gamma_i$ and σ^2) and therefore

$$\text{(15)} \qquad \begin{aligned} &\Pr\{S_{i+1} \mid c_0, \cdots, c_i, c_{i+1}^2 + \cdots + c_q^2 + S; \gamma_{i+1} = \cdots = \gamma_q = 0\} \\ &\qquad\qquad = p_q + \cdots + p_{i+1} \end{aligned}$$

for almost all $c_0, \cdots, c_i$ and $c_{i+1}^2 + \cdots + c_q^2 + S$. Let

$$f_i(c_0, \cdots, c_i, c_{i+1}^2 + \cdots + c_q^2 + S)$$

be the characteristic set function of $T_i (f = 1$ if the point is in the set and 0 otherwise). Then

$$\text{(16)} \qquad \begin{aligned} &\Pr\{\bar{S}_{i+1} \cap T_i\} \\ &\qquad = \mathcal{E}[\Pr\{\bar{S}_{i+1} \mid c_0, \cdots, c_i, c_{i+1}^2 + \cdots + S\} \\ &\qquad\qquad f_i(c_0, \cdots, c_i, c_{i+1}^2 + \cdots + S)] \\ &\qquad = \mathcal{E}[(1 - p_q - \cdots - p_{i+1}) f_i(c_0, \cdots, c_i, c_{i+1}^2 + \cdots + S)] \end{aligned}$$

which is (14). This proves the lemma.

The point of the lemma is that however $R_q, \cdots, R_{i+1}$ are chosen subject to (9), which implies (13) for $S_{i+1} = R_q \cup \cdots \cup R_{i+1}$, the probability of a region R_i defined as an intersection $\bar{S}_{i+1} \cap T_i$ depends only on T_i and not on S_{i+1} (when $H^*_{q,\cdots,i+1}$ is true).

Now let T_i^* be the region defined by (10) for $\alpha_i = p_i/(1 - p_q - \cdots - p_{i+1})$.

LEMMA 2. *Given S_{i+1} satisfying* (13) *and any R_i disjoint with S_{i+1} such that*

$$\Pr\{R_i \mid \gamma_i = \gamma_{i+1} = \cdots = \gamma_q = 0\} = p_i, \tag{17}$$

$$\Pr\{R_i \mid \gamma_i, \gamma_{i+1} = \cdots = \gamma_q = 0\} = \Pr\{R_i \mid -\gamma_i, \gamma_{i+1} = \cdots = \gamma_q = 0\}, \tag{18}$$

then

$$\Pr\{\bar{S}_{i+1} \cap T_i^* \mid \gamma_{i+1} = \cdots = \gamma_q = 0\} \geqq \Pr\{R_i \mid \gamma_{i+1} = \cdots = \gamma_q = 0\}. \tag{19}$$

PROOF. Suppose there were some value of γ_i so that the inequality (19) were violated. We shall show that this contradicts the previous assertion that

$$T_i^* = (S_{i+1} \cap T_i^*) \cup (\bar{S}_{i+1} \cap T_i^*)$$

is the uniformly most powerful test of the hypothesis $\gamma_i = 0$ (when $\gamma_{i+1} = \cdots = \gamma_q = 0$) with power independent of the sign of γ_i. For T_i^* and

$$R_i \cup (S_{i+1} \cap T_i^*)$$

are critical regions of the same size for testing $\gamma_i = 0$ with power independent of the sign of γ_i but the power of the second region at this special value of γ_i would be greater than that of T_i^*. Since this is false, the lemma must be true.

The implication of the two lemmas is that whatever $R_q, \cdots, R_{i+1}$ are, the best choice of R_i is the part of T_i^* disjoint from $R_q, \cdots, R_{i+1}$ and for such a choice

$$\Pr\{R_i \mid \gamma_{i+1} = \cdots = \gamma_q = 0\} = (1 - p_q - \cdots - p_{i+1}) \Pr\{T_i^*\}. \tag{20}$$

This does not depend on the choice of $R_q, \cdots, R_{i+1}$. Incidentally, this proves that $r_q, \cdots, r_i$ given by (10) are independently distributed when

$$\gamma_q = \cdots = \gamma_{i+1} = 0.$$

THEOREM. *Let $R_m, R_{m+1}, \cdots, R_q$ be $q - m + 1$ disjoint regions in the sample space such that*

$$\begin{aligned} \Pr\{R_i \mid \gamma_i = \gamma_{i+1} = \cdots = \gamma_q = 0\} &= p_i, \qquad i = m+1, \cdots, q, \\ \Pr\{R_m \mid \gamma_{m+1} = \gamma_{m+2} = \cdots = \gamma_q = 0\} &= p_m, \end{aligned} \tag{21}$$

where $p_m + p_{m+1} + \cdots + p_q = 1$, and

$$\Pr\{R_i \mid \gamma_i, \gamma_{i+1} = \cdots = \gamma_q = 0\} = \Pr\{R_i \mid -\gamma_i, \gamma_{i+1} = \cdots = \gamma_q = 0\}, \qquad i = m+1, \cdots, q. \tag{22}$$

Then for every value of γ_i (22) *is maximized by R_i defined by* (10) *for*

$$\alpha_i = p_i/(1 - p_q - \cdots - p_{i+1})$$

complementary to $R_q \cup \cdots \cup R_{i+1}$, $i = m + 1, \cdots, q$.

The optimum procedure is, therefore,

$$
\begin{aligned}
R_q &= T_q^*, \\
R_{q-1} &= \bar{T}_q^* \cap T_{q-1}^*, \\
&\vdots \\
R_i &= \bar{T}_q^* \cap \bar{T}_{q-1}^* \cap \cdots \cap \bar{T}_{i+1}^* \cap T_i^*, \\
&\vdots \\
R_{m+1} &= \bar{T}_q^* \cap \bar{T}_{q-1}^* \cap \cdots \cap \bar{T}_{m+2}^* \cap T_{m+1}^*, \\
R_m &= \bar{T}_q^* \cap \bar{T}_{q-1}^* \cap \cdots \cap \bar{T}_{m+2}^* \cap \bar{T}_{m+1}^*.
\end{aligned} \tag{23}
$$

What the procedure amounts to is to test $\gamma_q = 0$, $\gamma_{q-1} = 0$, $\cdots$ in turn until either one rejects such a hypothesis, say rejects $\gamma_i = 0$ and hence decides H_i, or one accepts all such hypotheses and eventually H_m. Thus the procedure is a sequence of hypothesis tests; this result is a consequence of the requirement that the probability of deciding that the degree of the polynomial is less than a given integer when that is the case should not depend on what the polynomial is.

The optimum properties of the t-test of $\gamma_i = 0$ include being uniformly most powerful among (i) the class of tests based on similar regions which are symmetric (in this c_i or in all y_t), (ii) the class of tests with powers depending only on γ_i^2/σ^2, (iii) the class of tests invariant under (scale and reflection) transformations, $c_j \rightarrow kc_j$, $S \rightarrow k^2S$ (equivalently $y_t \rightarrow ky_t$) and (location) transformations, $c_j \rightarrow c_j + a_j$, $j = 1, \cdots, i - 1$ (equivalently adding an arbitrary polynomial of degree $i - 1$ to y_t), and (iv) the class of unbiased tests. Each stated desirable property of the t-test leads to a corresponding formulation of the theorem of this paper. For example, the similarity and symmetry requirements, (21 and (22), could be replaced by unbiasedness requirements, namely, that

$$
\Pr\{\text{Reject } H_{q,\cdots,i}^* \mid H_{q,\cdots,i}^*\} \leqq p_q + \cdots + p_i, \qquad i = m + 1, \cdots, q; \tag{24}
$$

for unbiasedness implies similarity and independence of the signs of $\gamma_q, \cdots, \gamma_{m+1}$.

We have not stated here how $p_m, p_{m+1}, \cdots, p_q$ are to be chosen. If α_i is fixed, say α, then $p_q = \alpha$, $p_{q-1} = \alpha(1 - \alpha)$, $\cdots$, $p_i = \alpha(1 - \alpha)^{q-i}$, $\cdots$, $p_{m+1} = \alpha(1 - \alpha)^{q-m-1}$, $p_m = (1 - \alpha)^{q-m}$. In general one wants to balance the desirability of not overestimating the degree with sensitivity to nonzero coefficients.

While the approach of this paper is (presumably) novel, the procedure derived is not a new one. For example, a sequence of significance tests, starting with the coefficient of the highest permitted degree, after determining q by inspection of the data, has been suggested by Plackett ([4], p. 92), and this series of tests after a preliminary F-test of all coefficients has been recommended by Williams ([6], p. 41). Another procedure for deciding the degree of the polynomial is a sequence of significance tests, but in the reverse order. (This method seems to be widely used; see Snedecor [5], Section 14.6, for example.) First, test $\gamma_{m+1} = 0$ by using r_{m+1}; then, if this hypothesis is rejected, test $\gamma_{m+2} = 0$ by using r_{m+2}; and continue until some hypothesis is accepted or until $\gamma_q = 0$ has been rejected. A disadvantage of this procedure is that if some γ_i is very large, the probability of deciding on a polynomial of too low a degree is large. For example, if γ_2 is large and γ_1 is small, the probability of

$$\frac{c_1^2}{c_2^2 + \cdots + c_q^2 + S} < k \tag{25}$$

is large, that is, of accepting $\gamma_1 = 0$ and deciding that the degree of the polynomial is 0. (The indicated difficulty is partly the failure to satisfy (8) or (9) for $\gamma_3 = \cdots = \gamma_q = 0$.)

A practical disadvantage of the procedure, which is theoretically best in the formulation of this paper, is that it requires computation of S and hence of $c_0, c_1, \cdots, c_q$ for a value of q chosen in advance whereas in the other sequential procedure one computes $c_1, c_2, \cdots$, in sequence (with $c_2^2 + \cdots + c_q^2 + S = \sum y_i^2 - c_0^2 - c_1^2$, etc.) and one needs to compute only as long as the hypotheses are rejected. However, this disadvantage is limited because the regression coefficients of orthogonal polynomials are relatively easy to compute since tables of the polynomials are available and because usually one would choose q small since orthogonal polynomials are not very useful if the situation calls for a high degree. (In practice, if one accepted H_q one might be tempted to compute c_{q+1} as a check on the choice of q.)

The problem of deciding the degree of polynomial could, of course, be formulated in other ways. Perhaps one could take into account explicitly the cost of computation. The objection to the other sequential procedure might be affected by an assumption about the permissible relation between γ_i and γ_{i+1} (such as $\gamma_{i+1}^2 < k\gamma_i^2$ for a preassigned k). Another approach to the entire problem is in terms of prediction. Lehmann [2] has also formulated this problem as a multiple decision problem which involves the concepts and results of the theory of testing hypotheses, but his formulation and solution are different. He assumes explicitly that the procedure is sequential, whereas here the sequential feature is a consequence of the conditions of the solution. The critical difference, however, is that he requires that if the investigator has decided $\gamma_q = \cdots = \gamma_{i+1} = 0$ then the probability of deciding $\gamma_i = 0$ (when this is true) should not depend on $\gamma_{i+1}, \cdots, \gamma_q$ which he now permits to be different from 0. This requirement does not seem to be appropriate in many situations, however, because the question of $\gamma_i = 0$ is not of interest if any of $\gamma_{i+1}, \cdots, \gamma_q$ is nonzero and there is, therefore, no reason to control the probability of deciding $\gamma_i = 0$ when the polynomial is of higher degree.

3. Generalizations. It is clear that the formulation of the problem in terms of orthogonal polynomials is simply a convenience. The mean value function (1) could be written equivalently

$$\mathcal{E}y_t = \beta_0 + \beta_1 t + \cdots + \beta_q t^q, \qquad t = 1, \cdots, T. \tag{26}$$

Then β_i is a linear combination of $\gamma_i, \gamma_{i+1}, \cdots, \gamma_q$ and the hypothesis

$$\gamma_i = \cdots = \gamma_q = 0$$

is equivalent to $\beta_i = \cdots = \beta_q = 0$; that is, the degree in one formulation is the same as the other. (Consideration of this relationship emphasizes that if $\gamma_i \neq 0$ one is not interested in whether $\gamma_j = 0$ for $j < i$.)

The problem is formally unchanged if the powers of t in (26) are replaced by other "independent variates." The important point is that the independent variates are ordered in the sense that in the regression function we are interested in whether $\beta_i = 0$ only if $\beta_{i+1} = \cdots = \beta_q = 0$. When the independent variates are not polynomials in t it is unlikely that the independent variates are orthogonal. In that case the computations can be done by use of the forward solution of the normal equations.

Lehmann [2] has pointed out a direct generalization. Let

$$c_i = \begin{pmatrix} c_{1i} \\ \vdots \\ c_{n_i i} \end{pmatrix}, \qquad \mathcal{E}c_i = \gamma_i = \begin{pmatrix} \gamma_{1i} \\ \vdots \\ \gamma_{n_i i} \end{pmatrix}, \qquad i = 0, 1, \cdots, q. \tag{27}$$

Suppose the c_{gi} are independently and normally distributed with common variance σ^2 and let S/σ^2 be independently distributed as χ^2 with n degrees of freedom. Then again we can ask the questions indicated by (5) or (6) where we interpret $\gamma_{i+1} = \cdots = \gamma_q = 0$ to mean $\gamma_{gj} = 0$, $g = 1, \cdots, n_j$, $j = i + 1, \cdots, q$. (The vectors $\gamma_{i+1}, \cdots, \gamma_q$ may not have the same number of coordinates.) We can again require (8) or (9). Here we can ask that when $\gamma_{i+1} = 0, \cdots, \gamma_q = 0$ the probability of R_i depend on $\gamma_i'\gamma_i = \gamma_{1i}^2 + \cdots + \gamma_{n_i i}^2$. Then R_i with maximum probability for $\gamma_i'\gamma_i \neq 0$ (corresponding to the optimum property of (28) below) is the intersection of the complement of $R_q \cup \cdots \cup R_{i+1}$ and

$$r_i = \frac{c_i' c_i}{c_{i+1}' c_{i+1} + \cdots + S} > \frac{n_i}{n_{i+1} + \cdots + n_q + n} F_{n_i, n_{i+1}+\cdots+n}(\alpha_i), \tag{28}$$

where $F_{n_i, n_{i+1}+\cdots+n}(\alpha_i)$ is the significance point of the F-distribution with n_i and $n_{i+1} + \cdots + n$ degrees of freedom corresponding to significance level $\alpha_i = p_i/(1 - p_q - \cdots - p_{i+1})$. This formulation is in the canonical form of a regression function

$$\mathcal{E}y_t = \sum_{j=0}^{q} \sum_{g=1}^{n_j} \beta_{gj} z_{gjt} . \tag{29}$$

In the multivariate case y_t can be taken as a vector with coordinates $y_{1t}, \cdots, y_{pt}$. In the regression function (1) the coefficient γ_i can be taken as a vector with coordinates $\gamma_{1i}, \cdots, \gamma_{pi}$ or in (26) $\beta_i' = (\beta_{1i}, \cdots, \beta_{pi})$. We assume $y_1, \cdots, y_T$ to be independently distributed with y_t having a multivariate normal distribution with covariance matrix Σ. The estimate of γ_i is c_i given by

(3), and the estimate of Σ is $1/(T - q - 1)$ times

$$S = \sum_{t=1}^{T} [y_t - c_0\phi_0(t) - \cdots - c_q\phi_q(t)][y_t - c_0\phi(t) - \cdots - c_q\phi_q(t)]' \tag{30}$$
$$= \sum_{t=1}^{T} y_t y_t' - \sum_{i=0}^{q} c_i c_i' .$$

The sets (5) and (6) are now in terms of the vectors γ_i. When $\gamma_{i+1} = \cdots = \gamma_q = 0$, the best test of the hypothesis $\gamma_i = 0$ at a given significance level α_i has the rejection region (based on Hotelling's T^2)

$$c_i'(c_{i+1}c_{i+1}' + \cdots + S)^{-1}c_i > [p/(T - i - p)]F_{p,T-i-p}(\alpha_i). \tag{31}$$

The test is best in the sense of being the uniformly most powerful test with power depending on $\gamma_i'\Sigma^{-1}\gamma_i$ as well as being the uniformly most powerful invariant test ($c_j \rightarrow Kc_j$, $S \rightarrow KSK'$ for K nonsingular and $c_j \rightarrow c_j + a_j$, $j = 1, \cdots, i - 1$) (see [1], Chapter 5). If we assign probabilities (8) or (9) and ask for the best procedure with probability depending on $\gamma_i'\Sigma^{-1}\gamma_i$ (when $\gamma_{i+1} = \cdots = \gamma_q = 0$) we find that R_i is the intersection of (31) and the complement of $R_q \cup \cdots \cup R_{i+1}$.

It will be observed that the formulation of the problem has a solution because of the Neyman structure of the regions and the existence of a best procedure for each component hypothesis. The Neyman structure depends on having sufficient statistics for families of distributions that are boundedly complete. We can make more general statements. Suppose we observe a vector y with a distribution $F(\gamma_1, \cdots, \gamma_q, \sigma)$, where each of $\gamma_1, \cdots, \gamma_q$ and σ may be vectors. We formulate the sets of parameter points (5) or (6) with $m = 0$ (the "nuisance parameters" forming σ). Suppose there are statistics $c_1, \cdots, c_q, s$ such that $c_1, \cdots, c_i, s$ are sufficient when $\gamma_{i+1} = 0, \cdots, \gamma_q = 0$, $i = 1, \cdots, q$. (For the original problem the vector s would consist of c_0 and $c_1^2 + \cdots + c_q^2 + S$.) The decision regions may be defined in the space of $c_1, \cdots, c_q, s$; and a decision about γ_i can be made on the basis of $c_1, \cdots, c_i, s$. A similar region T_i has Neyman structure

$$\Pr\{T_i \mid c_1, \cdots, c_{i-1}, s; \gamma_i = 0, \cdots, \gamma_q = 0\} = \alpha \tag{32}$$

with probability 1 (with respect to all $F(\gamma_1, \cdots, \gamma_{i-1}, 0, \cdots, 0, \sigma)$) if the family of distributions $F(\gamma_1, \cdots, \gamma_{i-1}, 0, \cdots, 0, \sigma)$ is boundedly complete. Then Lemma 1 holds.

Suppose that when $\gamma_{i+1} = 0, \cdots, \gamma_q = 0$ there is a uniformly most powerful test of $\gamma_i = 0$ at significance level α_i within a class of tests, the class being defined in terms of the power as a function of γ_i. In view of Lemma 1 the intersection of the rejection region of such a test with the complement of $R_q \cup \cdots \cup R_{i+1}$ will have a corresponding property, and the corresponding form of Lemma 2 will hold. Then the corresponding Theorem holds for the set of regions $R_1, \cdots, R_q$. An application of this general theory can be made to determining the appropriate order of a stochastic difference equation in a certain class of nearly stationary stochastic processes.

The theory in this paper has been developed in terms of nonrandomized tests; only direct modifications are needed to make the results valid for randomized tests.

REFERENCES

[1] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, John Wiley and Sons, New York, 1958.
[2] E. L. Lehmann, "A theory of some multiple decision problems II," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 547–572.
[3] E. L. Lehmann, *Testing Statistical Hypothesis*, John Wiley and Sons, New York, 1959.
[4] R. L. Plackett, *Principles of Regression Analysis*, Oxford University Press, 1960.
[5] George W. Snedecor, *Statistical Methods*, Iowa State College Press, Ames, 1937.
[6] E. J. Williams, *Regression Analysis*, John Wiley and Sons, New York, 1959.