STATISTICAL RESEARCH REPORT No 3
Institute of Mathematics 1972
University of Oslo

ON THE PROBLEM OF SHIFT IN THE REGRESSION STRUCTURE

by

Tore Schweder

1. INTRODUCTION AND SUMMARY

When using the ordinary regression model

$$(1) \qquad EY_t = \sum_{i=1}^{q} x_{ti}\beta_i = x_t\beta$$

one may be faced with one of the problems:

1. There may be outliers in the material. The regression structure is assumed to be essentially stable, but for some reason or other, the absolute difference
$|Y_t - \Sigma\, x_{ti}\beta_i|$
is feared to be abnormally high for some indices t. How can the outliers be identified and excluded?
2. There may have been a shift in the regression parameters. How can the time of occurrence of such a shift be estimated?

In the following we shall present two simultaneous methods, both are in a sense optimal. The first of them is a coordinated sequence of ordinary student tests obtained by reasoning similar to the idea basic to Anderson's test for the complexity of a regression model, see [1]. The second method is a cumulative sum test obtained by maximizing minimum average power, see [2]. The two methods are based on statistics obtained by stepwise regression in the direction of observations.

2. STEPWISE REGRESSION IN THE DIRECTION OF OBSERVATIONS.

Let (y_t, x_t) ; $t = 1,\ldots,s$ be observations in the regression model (1). The least square estimator of the parameter vector β is

$$\hat{\beta}_{(1)} = [X'_{(1)}X_{(1)}]^{-1}X'_{(1)}y_{(1)} \quad ,$$

where $y_{(1)}$, $X_{(1)}$ is the material on matrix form. Let $y_{(2)}$, $X_{(2)}$ be another material satisfying (1). The least square estimator of the combined material may now be computed according to the formula

$$\hat{\beta} = \hat{\beta}_{(1)} + (X'_{(1)}X_{(1)})^{-1}X'_{(2)}He^{*}$$

where

$$e^{*} = y_{(2)} - X_{(2)}\hat{\beta}_{(1)}$$

and

$$H = [I + X_{(2)}(X'_{(1)}X_{(1)})^{-1}X'_{(2)}$$

If furthermore $e_{(1)} = y_{(1)} - X_{(1)}\hat{\beta}_{(1)}$ and the residualvector for the combined material is $e = y - X\hat{\beta}$, then the sum of squared residuals may be calculated by

(2) $$e'e = e'_{(1)}e_{(1)} + e^{*'}He^{*}$$

These formulae were derived by Placket [3]. Väliaho [5] has put these ideas into a computational scheme wherein also ordinary stepwise regression (stepwise regression in the direction of variables) is possible. Proofs of the formulae and extensions of them is found in the two references.

3. THE PROBLEM

In the following, we shall assume the model

$$Y_t = x_t\beta + U_t \qquad ; \; t = -r+1, \ldots, 0 \; ,$$

$$Y_t = x_t\beta + \gamma_t + U_t \qquad ; \; t = 1, \ldots, s \; .$$

U_t, $t = -r+1, \ldots, s$ are independent $N(0,\sigma)$. The q-dimensional regressor vectors x_t are non-stochastic, β is the q-dimensional column vector of regression parameters, and γ_t is the difference between the true mean of the regressand Y_t and $x_t\beta$; $-\infty < \gamma_t < \infty$, $t = 1, \ldots, s$.

We want to determine at what time the first shift in the regression structure occurs. That is, we want a method which determies the least k for which $\gamma_k \neq 0$. This problem may be formulated as a test problem with s alternatives:

$$
(3) \qquad
\begin{aligned}
H_0 &: \gamma_1 = \gamma_2 = \dots = \gamma_s = 0 \\
A_1 &: \gamma_1 \neq 0 \\
A_2 &: \gamma_1 = 0 \ , \ \gamma_2 \neq 0 \\
&\ \ \vdots \\
A_s &: \gamma_1 = \gamma_2 = \dots = \gamma_{s-1} = 0 \ , \ \gamma_s \neq 0
\end{aligned}
$$

A test for this is an s-dimensional vectorfunction $(\psi_1(y),\dots,\psi_s(y))$ of the vector of observations, y, satisfying

$$\psi_i(y) \geq 0 \ ; \ i = 1,\dots,s \ ; \ \sum_{i=1}^{s} \psi_i(y) \leq 1 \ .$$

If y is observed, H_0 is rejected with probability $\sum_{i=1}^{s} \psi_i(y)$ and alternative A_i is accepted with probability $\psi_i(y)$. At most one of the alternatives is accepted, and if A_i is accepted then our conclusion is that the first shift in the regression structure occurs at time $t = i$. If H_0 is not rejected, the material gives no reason to claim any shift in the regression structure.

Denote by θ the entire parametervector (representing σ, β, $\gamma_1,\dots,\gamma_s$), and by Ω the space of a priori possible θ. Ω is partitioned into the $s + 1$ disjoint subsets

$$\Omega_{s+1} = \{\theta \mid \gamma_1 = \dots = \gamma_s = 0\}$$

$$\Omega_k = \{\theta \mid \gamma_1 = \dots = \gamma_{k-1} = 0, \ \gamma_k \neq 0\} \ ; \ k = 1,\dots,s \ .$$

The problem may now be formulated

$$H_o : \theta \in \Omega_{s+1} \; ; \; A_k : \theta \in \Omega_k \; ; \; k = 1,\dots,s \; .$$

A test ψ is called a level α test if

$$\sup_{\theta\in\Omega_{s+1}} \sum_{i=1}^{s} E_\theta \psi_i(Y) \leq \alpha$$

Let

$$\beta(\theta, \psi_i) = E_\theta \psi_i(Y) \; ; \; i = 1,\dots,s \; .$$

The power function of ψ is the vector function $(\beta(\theta, \psi_1),\dots,\beta(\theta, \psi_s))$. We shall construct two level α tests which maximizes the power function with respect to two different optimality criterions.

When $\theta \in \Omega_k$, i.e. the regression structure is stable up to time $t = k-1$ but a shift occurs at time $t = k$, then β and σ should be estimated by $\hat{\beta}_{k-1}$, $\hat{\sigma}_{k-1}$ which are the least square estimators based on the material $Y_{-r+1},\dots,Y_{k-1}$.

$$\hat{\beta}_{k-1} = (X'_{k-1}X_{k-1})^{-1}X'_{k-1}\underline{Y}_{k-1}$$

(4)

$$(r+k-q-1)\hat{\sigma}^2_{k-1} = \sum_{i=-r+1}^{k-1} (Y_i - x_i\hat{\beta}_{k-1})^2$$

where X_{k-1} is the regressor matrix and $\underline{Y}_{k-1}$ the regressand vector of the material (Y_t, x_t) ; $t = -r+1,\dots,k-1$. The model error at time $t = k$, γ_k, should in this case be estimated by the observed residual

$$e^*_k = Y_k - x_k\hat{\beta}_{k-1}$$

For the residuals we have

Lemma 1

$e_1^*, e_2^*, \dots, e_s^*$ are stochastically independent and normally distributed with

$$Ee_k^* = \xi_k = \gamma_k + x_k(\beta - E\hat{\beta}_{k-1})$$

$$\sigma^{-2}\mathrm{var}\ e_k^* = \tau_k^2 = 1 + x_k(X_{k-1}'X_{k-1})^{-1}x_k' \tag{5}$$

Proof.

The only montrivial thing to show is that e_i^* and e_j^* are uncorrelated when $i < j$. With $\underline{U}_k$ defined analogous to $\underline{Y}_k$, see (4), we have

$$\begin{aligned}\mathrm{covar}(e_i^*, e_j^*) &= -\ \mathrm{covar}(Y_i,\ x_j\hat{\beta}_{j-1}) + \mathrm{covar}(x_j\hat{\beta}_{j-1},\ x_i\hat{\beta}_{i-1}) \\ &= -\ EU_i \circ (x_j(X_{j-1}'X_{j-1})^{-1}X_{j-1}'\underline{U}_{j-1}) \\ &+ E(x_j(X_{j-1}'X_{j-1})^{-1}X_{j-1}'\underline{U}_{j-1})(\underline{U}'_{i-1}X_{i-1}(X'_{i-1}X_{i-1})^{-1}x_i').\end{aligned}$$

But

$$X_{j-1}'E(\underline{U}_{j-1} \circ U_i) = x_i'\sigma^2$$

and

$$X_{j-1}'E(\underline{U}_{j-1}\underline{U}_{i-1}') = X_{i-1}'\sigma^2$$

so e_i^* and e_j^* are uncorrelated.

4. A COORDINATED SEQUENCE OF STUDENT TESTS

Let p_k ; $k = 1,\dots,s$ be probabilities

$$\alpha = \sum_{k=1}^{s} p_k < 1 .$$

and let

$$\alpha_k = \frac{p_k}{1-p_1-\dots-p_{k-1}} .$$

Denote by δ_k the testfunction of the ordinary student test at level α_k for the problem

$$H_k : \gamma_1 = \ldots = \gamma_k = 0, \text{ against } A_k : \gamma_1 = \ldots = \gamma_{k-1} = 0, \; \gamma_k \neq 0$$

That is

$$(6) \quad \begin{cases} \delta_k(y) = 1 & \text{when } \left| \dfrac{Y_k - x_k \hat{\beta}_{k-1}}{\hat{\sigma}_{k-1} \tau_k} \right| > t_k \\ \delta_k(y) = 0 & \text{when } \left| \dfrac{Y_k - x_k \hat{\beta}_{k-1}}{\hat{\sigma}_{k-1} \tau_k} \right| \leq t_k \end{cases}$$

where t_k is the upper $\frac{1}{2}\alpha_k$ fractile in the student distribution with $r+k-q-1$ degrees of freedom.

Consider the test

$$(7) \qquad \psi_k = \delta_k \prod_{j=1}^{k-1} (1-\delta_j) \; ; \; k = 1, \ldots, s \; .$$

This test is an optimal level α test in the following sense:

<u>Theorem 1</u>. For $k = 1,2,\ldots,s$, we have

(i) $E\,\psi_k = p_k$ when $\gamma_1 = \ldots = \gamma_k = 0$

(ii) $E\,\psi_k \geq p_k$ when $\gamma_1 = \ldots = \gamma_{k-1} = 0$

(iii) when $\gamma_1 = \ldots = \gamma_{k-1} = 0$, $E\,\psi_k$ is maximized among all tests satisfying (i) and (ii).

<u>Proof.</u>

Introduce $\omega_{k-1} = \bigcup_{i=k}^{s+1} \Omega_i$; $k = 1, \ldots, s+1$. The sequence of testing problems introduced above may be written

$$H_k : \theta \in \omega_k \quad \text{against} \quad A_k : \theta \in \omega_{k-1} - \omega_k \; .$$

Relative to ω_{k-1} , the statistic

$$D_k = (\hat{\sigma}_{k-1}, \hat{\beta}_{k-1}, Y_k, \dots, Y_s)$$

is sufficient and complete, since the family of distributions of D_k when θ varies over ω_{k-1} is a regular exponential one; $k = 1, \dots, s$.

But according to Sverdrup [4], the test ψ defined by (7) satisfies (i)-(iii) whenever the underlying test δ_k is uniformly most powerful unbiased test for H_k against A_k at level α_k ; $k = 1, \dots, s$.

To see that δ_k is uniformly most powerful unbiased level α_k test for H_k against A_k, it is convenient to write the model (ω_{k-1}) on standard regression form:

$$Y_i = x_i\beta + \sum_{j=k}^{s} \delta_{ij}\gamma_j + U_i \ ; \quad i = -r+1, \dots, s \ ,$$

δ_{ij} is the kroenecker delta.

The least square estimators of the parameters are

$$\tilde{\beta} = \hat{\beta}_{k-1} \quad \text{and} \quad \tilde{\gamma}_i = Y_i - x_i\tilde{\beta} \qquad i = k, \dots, s \ ,$$

and the sum of squared residuals is

$$Q_o = \sum_{i=-r+1}^{k-1} (Y_i - x_i\tilde{\beta})^2 = (r+k-q-1)\hat{\sigma}^2_{k-1} \ .$$

Consequently, by ordinary regression analysis, the mentioned student-test is the uniformly most powerful unbiased level α_k' test. Note that the test is based on the statistic $(\hat{\beta}_{k-1}, \hat{\sigma}_{k-1}, Y_k)$ or on D_k. □

The test ψ defined above is very simple. One tests sequentially $H_k : \gamma_k = 0$ against $A_k : \gamma_k \neq 0$ assuming $\gamma_1 = \dots = \gamma_{k-1} = 0$ by ordinary student tests with coordinated levels α_k; $k = 1, \dots, s$, and when a H_k is rejected for the first time, say H_j, it is stateded that the first shift in the regression structure occurs at time $t = j$.

By lemma 1 and the following lemma, it is easy to characterize the power function of ψ further.

Lemma 2.

When $\gamma_1 = \dots = \gamma_k = 0$, then $\delta_1, \dots, \delta_k$ are independent and independent of $\delta_{k+1}, \dots, \delta_s$.

Proof

for each $j \leq k$, $\delta_{j+1}, \dots, \delta_s$ is a function of D_{j+1} which is sufficient and complete relative to ω_j. But the distribution of $\delta_1, \dots \delta_j$ does not vary with $\theta \in \omega_j$, so by Basu's theorem $\delta_1, \dots, \delta_j$ is independent of $\delta_{j+1}, \dots, \delta_s$ for all $\theta \in \omega_j$. When $\gamma_1 = \dots = \gamma_k = 0$, $\theta \in \omega_j$ for $j = 1, \dots, k$ and we have the conclusion.

By the recursion formula (2) and the definition of τ_k^2 , (5), it is seen that

$$\hat{\sigma}_j^2 = (r+j-q)^{-1}[(r+j-q-1)\hat{\sigma}_{j-1}^2 + \tau_j^{-2} e_j^{*2}]$$

and by iteration

$$\frac{r+j-q}{\sigma^2}\hat{\sigma}_j^2 = \frac{r-q}{\sigma^2}\hat{\sigma}_o^2 + \sum_{i=1}^{j}\left(\frac{e_i^*}{\tau_i \sigma}\right)^2 ,$$

which by lemma 1 is seen to be eccentric χ^2 distributed with $r+i-q$ degrees of freedom and eccentricity

$$\rho_j = \sum_{i=1}^{j}\left(\frac{\xi_i}{\sigma\tau_i}\right)^2 \quad ; \quad j = 1, \dots, s .$$

But an eccentric χ^2_ν distributed stochastic variable W with eccentricity ρ may be written

$$W = Y^2 + Z$$

where Y and Z are independent, Y is $N(\sqrt{\rho}, 1)$ and Z is $\chi^2_{\nu-1}$. It is seen that as $r+j-q \to \infty$, $\hat{\sigma}^2_j - \sigma^2(1 + \frac{1}{r+j-q}\rho_j) \to 0$ in probability, and consequently for $r+k-q$ large enough,

$$\frac{Y_j - x_j\hat{\beta}_{j-1}}{\overset{*}{\sigma}_{j-1}\ \tau_j} \; ; \quad j = k+1,\dots,l$$

has approximately the same simultaneous distribution as

$$\frac{e^*_j}{\sigma\tau_j}\,(1 + (r+j-q-1)^{-1}\rho_{j-1})^{-\frac{1}{2}} \; ; \quad j = k+1,\dots,l ,$$

i.e. they are approximately independent and normally distributed.

By combining the two lemmas we then have since

$\psi_l = \prod_{j=1}^{l-1}(1-\delta_j)\delta_l$, the

<u>approximation</u>:

When $r+k-q$ is not too small, then for $\gamma_1,\dots,\gamma_k = 0$, $k < l \leq s$,

$$\beta(\theta,\psi_l) = (1 - \sum_{j=1}^{k} p_j) \cdot P(\bigcap_{j=k+1}^{l-1} \delta_j = 0,\ \delta_l = 1)$$

$$\approx (1 - \sum_{j=1}^{k} p_j)(\prod_{j=k+1}^{l-1} q_j)(1-q_l)$$

where

$$q_j = \Phi(t_j\sqrt{1+(r+j-q-1)^{-1}\rho_{j-1}} - \frac{\xi_j}{\sigma\tau_j}) - \Phi(-t_j\sqrt{1+(r+j-q-1)^{-1}\rho_{j-1}} - \frac{\xi_j}{\sigma\tau_j})$$

t_j is defined by (14) and Φ is the c.df. of the standard normal distribution.

5. A CUSUM TEST

Assume σ known. A sufficient statistic is then

$$(\hat{\beta}_o, e_1^*, \ldots, e_s^*) .$$

Our problem defined by (3) is invariant to the group of transformations: $(\hat{\beta}_o, e_1^*, \ldots, e_s^*) \to (\hat{\beta}_o - b, a_1 e_1^*, \ldots, a_s e_s^*)$ where b is any q dimensional vector and $a_i \in \{-1, 1\}$; $i = 1, \ldots, s$.

A maximal invariant is

$$(e_1^{*2}, \ldots, e_i^{*2}) ,$$

on which we, by the principle of invariance, may base our inference. To simplify notation, we write

$$Z_i = \frac{e_i^*}{\sigma \tau_i} ; \quad i = 1, \ldots, s .$$

By lemma 1 the problem is now: $Z_1^2, \ldots, Z_s^2$ are stochastically independent, Z_i^2 is χ^2 distributed with one degree of freedom and eccentricity.

$$\lambda_i = \left(\frac{\xi_i}{\sigma \tau_i} \right)^2 ; \ i = 1, \ldots, s .$$

H_o : $\lambda_1 = \ldots = \lambda_s = 0$

A_i : $\lambda_1 = \ldots = \lambda_{i-1} = 0$, $\lambda_i \neq 0$; $i = 1, \ldots, s$.

Following Pfanzagl [2], we shall construct a level α test maximizing minimum average power over the sub-alternatives

$\Omega_i' = \{\theta \mid \lambda_1 = \ldots = \lambda_{i-1} = 0, \ \lambda_i \geq \lambda, \ldots, \lambda_s \geq \lambda\}$; $i = 1, \ldots, s$, $\lambda > 0$.

I.e.

$$\inf \{ \sum_{i=1}^{s} \beta(\theta_i, \psi_i) ; \quad \theta_i \in \Omega_i' , \quad i = 1, \ldots, s \}$$

is to be maximized with respect to ψ among all level α tests.

The density of $\underline{Z} = [Z_1^2, \ldots, Z_s^2]$ belonging to the sub-alternative Ω_i' "closest" to H_o, is

$$f(\underline{z};\overline{\theta}_i) = (2\pi)^{-\frac{s}{2}} \left(\prod_{j=1}^{s} z_j^2\right)^{-\frac{1}{2}} e^{-\frac{1}{2}\sum\limits_{j=1}^{s} z_j^2} e^{-\frac{s-i+1}{2}\lambda} \cdot 2^{-(s-i+1)}$$

$$\prod_{j=i}^{s} \left[e^{|z_j|\sqrt{\lambda}} + e^{-|z_j|\sqrt{\lambda}} \right],$$

where $\overline{\theta}_i : \lambda_1 = \ldots = \lambda_{i-1} = 0$, $\lambda_i = \ldots = \lambda_s = \lambda$; $i = 1, \ldots, s$.

Let $\theta_o : \lambda_1 = \ldots = \lambda_s = 0$.

Pfanzagl's result is: if the level α test $(\psi_1, \ldots, \psi_s)$ has the structure

$$\psi_i(\underline{z}) = 1 \quad \text{when} \quad f(\underline{z};\ \overline{\theta}_i) > kf(\underline{z};\ \theta_o) \quad \text{and}$$

$$f(\underline{z};\ \overline{\theta}_i) = \max_{j=1,\ldots,s} f(\underline{z};\ \overline{\theta}_j)$$

$$\psi_i(\underline{z}) = 0 \quad \text{otherwise,}$$

then it maximizes

$$\sum_{i=1}^{s} \int \varphi_i(\underline{z}) f(\underline{z};\ \overline{\theta}_i) d\underline{z}$$

among all level α tests $(\varphi_1, \ldots, \varphi_s)$.

But

$$f(\underline{z};\ \overline{\theta}_i) > k\ f\ (\underline{z};\theta_o)$$

$$\Updownarrow$$

$$\sum_{j=i}^{s} \left[\ln\left(e^{|z_j|\sqrt{\lambda}} + e^{-|z_j|\sqrt{\lambda}}\right) - \left(\frac{\lambda}{2} + \ln 2\right) \right] > \ln k .$$

Let

$$S_i' = \sum_{j=i}^{s} \left[\ln\left(e^{|Z_j|\sqrt{\lambda}} + e^{-|Z_j|\sqrt{\lambda}}\right) - \left(\frac{\lambda}{2} + \ln 2\right) \right] .$$

Then

$$f(\underline{z};\overline{\theta}_k) = \max_{i=1,\dots,s} f(\underline{z};\overline{\theta}_i) > k\, f(\underline{z};\theta_o)$$

$$\Updownarrow$$

$$S_k' = \max_{i=1,\dots,s} S_i' > \ln k \quad .$$

The test obtained is thus:

$\psi_k(\underline{z}) = 1$ if S_k' is the largest of the cumulative sums $S_1',\dots,S_s'$, and if it ecceeds c',

$\psi_k(\underline{z}) = 0$ otherwise.

c' is to be determined to make the test a level α test.

Since Z_i^2 is stochastically increasing in λ_i and S_i' is increasing in Z_j^2 $j = i,\dots,s$, then for any $\theta \in \Omega_i'$, $\theta \neq \overline{\theta}_i$ S_i' is stochastically larger than if $\overline{\theta}_i$ were the true value of the parameter. It is thus seen that

$$\sum_{i=1}^{s} \beta(\overline{\theta}_i,\psi_i) = \inf\left\{ \sum_{i=1}^{s} \beta(\theta_i,\psi_i) \,|\, \theta_i \in \Omega_i' \;;\; i = 1,\dots,s \right\}$$

and hence, our test maximizes minimum average power over the sub alternatives Ω_i' ; $i = 1,\dots,s$, among all tests invariant to the above group of transformations, which have level α .

When λ is not too small, the following approximation is reasonable

$$\ln(e^{|Z_j|\sqrt{\lambda}} + e^{-|Z_j|\sqrt{\lambda}}) = |Z_j|\sqrt{\lambda} + \ln(1+e^{-2|Z_j|\sqrt{\lambda}}) \approx |Z_j|\sqrt{\lambda}$$

With

$$S_i = \sum_{j=i}^{s} (|Z_j| - k) , \quad k = \frac{\lambda+2\ln 2}{2\sqrt{\lambda}} ,$$

the following is a simple approximation to the above test:

$\varphi_k(\underline{z}) = 1$ if S_k is the largest of the cumulative sums $S_1,\ldots,S_s$, and if it ecceeds c

$\varphi_k(\underline{z}) = 0$ otherwise

where c is to be determined to make the test a level α test.

For moderate s, c is perhaps best determined by simulation. For large s, an approximate value of c is obtained by random walk theory. The problem is to choose the barriere c such that the process S_s, $S_{s-1},\ldots,S_1$ with probability α crosses the barriere when $\theta = \theta_o$. If $\theta = \theta_o$, then the process is a random walk.

Let now

$$T_i = \sum_{j=1}^{i} (|Z_j|-k) ; \quad i = 1,\ldots$$

where $Z_1,\ldots$ are independent $N(0,1)$ distributed. The moment generating function of the steps is

$$f^*(\omega) = E\, e^{-\omega(|Z|-k)} = 2\, e^{\omega k+\frac{\omega^2}{2}} [1-\Phi(\omega)]$$

where Φ is the c.d.f. of the standard normal distribution.

The equation

$$f^*(\omega) = 2e^{\omega k+\frac{\omega^2}{2}}[1-\Phi(\omega)] = 1$$

has a unique solution $\omega_0 (< 0)$ different from zero when $E|Z| = \sqrt{\frac{2}{\pi}} < k$. In this case it is known that the probability that the random walk $T_1, T_2, \ldots$ ever shall cross the barriere c is approximately $e^{\omega_0 c}$. Consequently, we may use the critical value

$$c = \omega_0^{-1} \ln \alpha .$$

It should be noted that the function

$$k = \frac{\lambda+2\ln 2}{2\sqrt{\lambda}}$$

attains its minimum $k_0 = \sqrt{2\ln 2} \approx 1.18$ at $\lambda = 2\ln 2$.

6. APPLICATIONS OF THE TWO METHODS

For problem situation 1, with possible outliers in the material, the first method should be well suited. The observations are re-indexed from $-r+1$ to s in such a way that we are sure about having no outliers among the observations with non-positive indices, but we are suspicious about the positively indexed observations. It is perhaps wise to let the most suspicious observations have the highest indices.

The procedure should then be to run the t-tests in sequence, perhaps with $p_i = \frac{\alpha}{s}$. Each time we get a rejection, we just delete the observation, and carry on/to the next test with unaltered estimates of β and σ. The procedure results in a purged material, which one can hope does not include serious outliers, and in estimates of β and σ based on this material.

If there are no outliers presented in the original material, then we will with probability $1-\alpha$ accept it as it is. Property (iii) indicates that we have not much hope in finding another method satisfying (i) and (ii) with a nicer performance function.

Problem 2 is met with in several situations. When we are investigating a historical process, we ask how long a structure, which in our case is a regression structure, remains unchanged. We may ask: when did the stable period start? or: when did the stable period come to an end? In prediction and control situations we want to use a material as modern and as extensive as possible. In such a situation, we may ask: from what time has the present structure remained stable? In situations like this both the methods seem to be suitable. They should be applied with the material indexed chronologically or anti-chronologically.

The coordinated sequence of t-tests is favourable because it is so simple and because we have exact knowledge of important features of its power function, also when σ is unknown. If σ is unknown but r is large, the cusum test may be used with σ replaced by $\hat{\sigma}_o$. Despite the approximations done for the cusum test, it is felt that it is superior to the first method in the present situation. It is more relevant to maximize minimum average power than to maximize what could be called marginal powers in this case.

The first method is easily generalized to a coordinated sequence of F-tests. Assume that it is possible to partition the index set into stable subperiods $T_o, T_1, \dots, T_s$, T_k includes ν_k indices ; $k = 0,\dots,s$. Let Y_k be the ν_k dimensional regressand vector and X_k the $\nu_k \times q$ dimensional regressor matrix referring to subperiod T_k ; $k = 0,\dots,s$. Then the model is

$Y_o, \ldots, Y_s$ are independent and normal

$EY_k = X_k \beta_k \quad , \quad V(Y_k) = \sigma^2 I \quad ; k = u, \ldots, s.$

(V is the coveriance matrix operator.)

With $\beta_o = \beta$ the problem may be to test

$$H_o : \beta_1 = \ldots = \beta_s = \beta$$

$$A_k : \beta_1 = \ldots = \beta_{k-1} = \beta \ , \ \beta_k \neq \beta \quad ; k = 1, \ldots, s \ .$$

With δ_k beeing the usual F test at level α_k for testing $H_k : \beta_k = \beta$ against $A_k : \beta_k \neq \beta$ assuming $\beta_1 = \ldots = \beta_{k-1} = \beta$, then

$$\psi_k = \prod_{j=1}^{k-1} (1-\delta_j)\delta_k$$

has the properties

(i) $E\psi_k = p_k$ when $\beta_1 = \ldots = \beta_k = \beta$

(ii) $E\psi_k > p_k$ when $\beta_1 = \ldots = \beta_{k-1} = \beta \ , \ \beta_k \neq \beta$

with

$$p_k = \prod_{j=1}^{k-1} (1-\alpha_j)\alpha_k \quad ; k = 1, \ldots, s \ .$$

This is seen in exactly the same manner as for the coordinented sequence of student tests.

In other situations, however, it may be known that one specific part of the regressionstructure is stable throughout the whole period while the other part may be subject to a shift. Let us assume stable subperiods $T_o, \ldots, T_s$, so that we have the model defined above. Now

$$\beta_k = \begin{bmatrix} \beta_k^1 \\ \hline \beta^2 \end{bmatrix} \quad \text{and} \quad X_k = \left[X_k^1 \ \vdots \ X_k^2 \right]$$

where β_k^1 is the $p < q$ dimensional part of the vector of regression parameters which corresponds to the substructure which possibly is subject to a shift, and X_k^1 is the $\nu_k \times p$ dimensional part of the regressor matrix X_k corresponding to the possibly unstable substructure.

If one is interested in testing whether there has been a shift in β^1, and if so, estimating the time when the first shift occurred, then one is led to the testing of the following problem

$$H_o : \beta_1^1 = \ldots = \beta_s^1 = \beta^1$$

$$A_k : \beta_1^1 = \ldots = \beta_{k-1}^1 = \beta^1 \quad , \quad \beta_k^1 \neq \beta^1 \; ; \; k = 1,\ldots,s \; .$$

Let

$$\Delta_k^1 = \beta_k^1 - \beta^1 \quad ; \; k = 1,\ldots,s \; ,$$

and $\omega_k = \{\theta \,|\, \Delta_1 = \ldots = \Delta_k = 0\} \; ; \; k = 1,\ldots,s \quad , \quad \omega_o = \Omega \quad ,$

(θ is the vector of all the parameters in the model).

As previously we may formulate the problem

$$H_o : \theta \in \omega_s \quad , \quad A_k : \theta \in \omega_{k-1} - \omega_k \quad ; \; k = 1,\ldots,s \, ,$$

where $\omega_o \supset \ldots \supset \omega_s$. The test we propose is based on sequential testing of

$$H_k : \theta \in \omega_k \quad \text{against} \quad A_k : \theta \in \omega_{k-1} - \omega_k \quad ; \; k = 1,\ldots,s \; .$$

Under ω_{k-1} , we have the regressionstructure

$$EY_t = \begin{cases} X_t\beta & ; \; t \leq k-1 \\ X_t\beta + X_t^1 \Delta_t & ; \; t \geq k \end{cases} \quad .$$

On matrix form, this is

$$\underline{EY} = \underline{X}_{k-1}\underline{\beta}_{k-1} = \left[\begin{array}{c|cccc} & & & 0 & \\ \hline & X_k^1 & 0 & \cdots & 0 \\ X & 0 & X_{k+1}^1 & & \\ & & & \ddots & \\ & & & & X_s^1 \end{array} \right] \left[\begin{array}{c} \beta \\ \hline \Delta_k \\ \hline \Delta_{k+1} \\ \hline \vdots \\ \hline \Delta_s \end{array} \right]$$

By the standard theory it is known that the usual F-test at level α_k for $H_k : \Delta_k = 0$ against $A_k : \Delta_k \neq 0$ in this model is an unbiased test based on a sufficient and complete statistic (relative to ω_{k-1}). Let this test be denoted by δ_k. Then the multiple test

$$\psi_k = \prod_{j=1}^{k-1} (1-\delta_j)\delta_k \quad ; k = 1,\dots,s \ ,$$

have properties (i) and (ii) mentioned above.

The computational work for the above test is conveniently done by stepwise regression in the direction of variables. The parameter vector $\hat{\underline{\beta}}_k$ is computed on basis of $\hat{\underline{\beta}}_{k-1}$ by deleting the p items $\Delta_{k,1},\dots,\Delta_{k,p}$ from the parameter vector when utilizing standard stepwise regression technique.

More on this method to construct multiple tests is found in Sverdrup [4], or in Das Gupta [6].

7. REFERENCES

[1] Anderson, T.W. (1962): The choice of degree of a polynomial regression as a multiple decision problem. Ann. Math. Stat. **33**, p. 255-265.

[2] Pfanzagl, J. (1959): Ein kombiniertes Test & Klassifikations - Problem. Metrika **2**, 11-45.

[3] Placket, R.L. (1950): Some theorems in least squares. Biometrika **37** , 149-157.

[4] Sverdrup, E. (1972): Stepwise statistical procedures with good performance in stepwise reduced parameter spaces. Statistical Research Report, Institute of Mathematics, University of Oslo **1**

[5] Väliaho, H. (1969): A synthetic approach to stepwise regression analysis. Soc. Sci. Fenn., Comm. Phys.-Math. **34**, No. 12.

[6] Das Gupta, S. (1970): Step-down multiple decision rules. Essays in Probability and Statistics, 229-250. Ed. R.C. Bose, I.M. Chahravarti, P.C. Mahalanobis, C.R. Rao and K.J. Smith, The University of North Carolina.