NASA-CR-200097

FINAL
IN-61-CR
OCIT
7682

FINAL TECHNICAL REPORT

Period Covered by Report: 10/1/91-9/30/95

"Inverse sequential procedures for the monitoring of time series"

NASA Project NAGW-2706

Principal investigator: Uwe Radok
CIRES, Campus Box 449 University of Colorado
Boulder CO 80309
303-492-5562 (FAX -2468)

P. 79

Winter solstice 1995

Preamble.

The early progress reports on this project (Radok 1991,1992; Radok and Brown 1993) borrowed concepts and terms from the Neyman-Pearson theory of hypotheses and spoke of "inverting" the sequential analysis of Wald (1947) by determining decision probabilities from the data, rather than prescribing them. A reformulation of the project's basic idea for the fourth progress report (Radok and Brown, 1995) has led to the "change index", a simpler concept without equivalent in previous work. This development traded theoretical sophistication for expanded usefulness in statistical exploration and is reflected in the simpler title of this final report:

"Detecting change in progress"

by Uwe Radok and Timothy J.Brown*

Contents

Preamble
Summary
1. Introduction
2. Probability argument
3. Properties of the Change Index (CI)
 3.1 CIs from Poisson samples
 3.2 CIs from Gaussian samples
 3.3 CIs for regression coefficients
 3.4 CIs from Chi-square samples
 3.5 Limitations and scope
4. Examples
 4.1 Retrospective tests
 4.2 Real-time monitoring
5. Alternatives
6. Conclusion
Acknowledgements, references.
Appendix A: Formulae
Appendix B: Fortran code "Sequitor".

[1]

[1]*Present affiliation: Western Regional Climate Center,Desert Research Institute, Reno NV 80506-0220

(NASA-CR-200097) INVERSE SEQUENTIAL PROCEDURES FOR THE MONITORING OF TIME SERIES Final Report, 1 Oct. 1991 - 30 Sep. 1995 (Colorado Univ.) 79 p

N96-18397

Unclas

Summary

When one or more new values are added to a developing time series, they change its descriptive parameters (mean, variance, trend, coherence). A "change index (CI)" is developed as a quantitative indicator that the changed parameters remain compatible with the existing "base" data. CI formulae are derived, in terms of normalized likelihood ratios, for small samples from Poisson, Gaussian, and Chi-Square distributions, and for regression coefficients measuring linear or exponential trends. A substantial parameter change creates a rapid or abrupt CI decrease which persists when the length of the bases is changed.

Except for a special Gaussian case, the CI has no simple explicit frequency distribution that could be used to delineate critical regions for tests of hypotheses. However, its design ensures that the series sampled need not conform strictly to the distribution form assumed for the parameter estimates. The use of the CI is illustrated with both constructed and observed data samples, processed with a Fortran code "Sequitor" (Appendix B).

1. Introduction

Modern statistics came into being and developed, on the fields of the British agricultural research station Rothamsted, from carefully designed randomized experiments and results obeying strictly valid frequency distributions with well-defined population parameters. But from the start there was another reality, consisting of uncontrollable observational data from the natural and human environments - an untidy mess of scattered, trended, abruply changing numbers with stretches of "statistically controlled" variation around quasi-constant central values. An example is shown in fig.1a. When a sufficiently large number of such data have been assembled an apparently orderly behavior may be constructed a posteriori (fig. 1b), but often the past is no reliable guide to a developing future, which then needs to be examined "sequentially" anew with each accruing observation.

A full analysis of time series with such short-lived quasi-steady regimes demands consideration of two disparate scales. A long time scale serves to define frequency distributions of regime lengths, regime sequences, and transitions between different regimes (cf. e.g.Olberg 1977). Here we shall be concerned exclusively with the other time scale: that of individual regimes, each of which needs to be considered on its own because its sample parameters contitute the full information available.

2.Probability argument.

The basic assumption made is that each observation has a probability $p(x;\Theta,\Theta',\Theta''...)$ of occurring where the parameters Θ remain valid or change as new observations are added. For a sample of n observations the product of the n individual probabilities define its "likelihood" $L(n;\Theta,\Theta',\Theta'',...) = p_1p_2...p_n$. In the simplest situation, two different estimates for a single parameter Θ constitute a "null hypothesis" H_0 $(\Theta=\Theta_0)$ which will be rejected in favor of the alternative hypothesis $H_1(\Theta=\Theta_1)$ when the data fall into a "critical region" of the

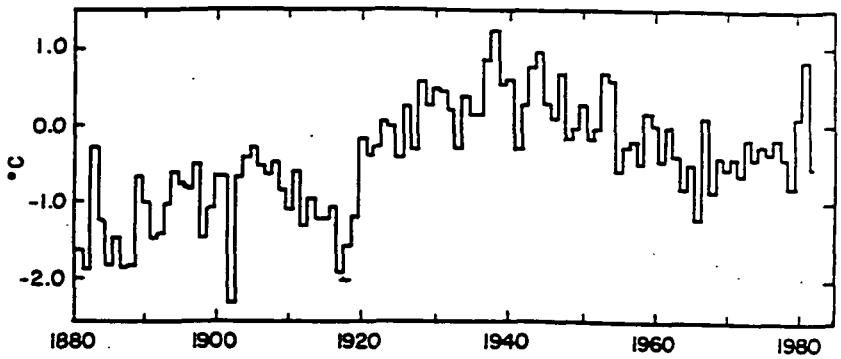


Fig. 1a: Deviations of the Arctic annual mean temperature from its 1946-1960 mean (adapted from Raper et al. 1983). Note the abrupt changes of mean around 1920,1955, and 1980; the trend changes around 1905 and 1940; and a variance change around 1968.

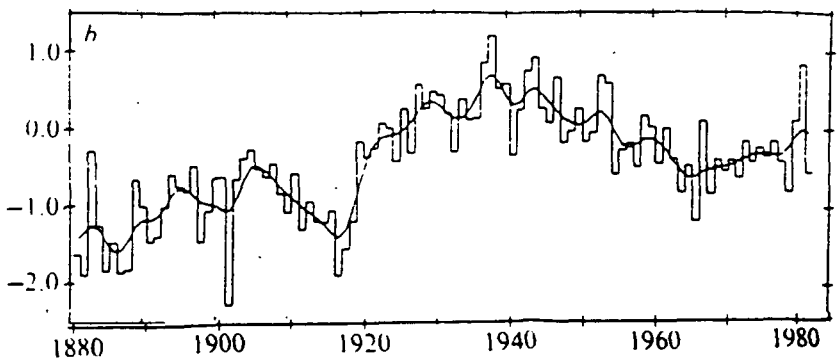


Fig. 1b: The original continuous interpretation of the same data.

n-dimensional sample space. The critical region is constructed to allow only a small probability α for H_0 in fact being true; α represents an "error of the first kind" and is also known as the "size" of the critical region. In the remainder of the sample space the acceptance of H_0, when in fact H_1 is true occurs with another probability β , known as "error of the second kind".

The likelihood ratio $L(n;\theta_1)/L(n;\theta_0)$ serves to define a best critical region if it exists. Its choice is to ensure that, for a prescribed small error of the first kind α (probability of rejecting H_0 though true), the error of the second kind (probability of rejecting H_1 and accepting H_0 in the remainder of the sample space) is as small as possible. This implies also that the probability of correctly acccpting H_1 in the critical region becomes $1 - \beta$.

A possible optimum critical region can be found, according to the Neyman-Pearson theory of hypotheses, by making the likelihood ratio $L(n\ ;\theta_1)/L(n\ ;\theta_0)$ larger than a constant k in the critical region, and smaller than k outside it. Intuitively this makes sense since the denominator is small in the critical region while the numerator (representing the probability $1 - \beta$) should be as large as possible to minimize β .

The distribution of the likelihood ratio directly defines best critical regions but, as shown by Kendall and Stuart (1967, ch. 24), its form in most cases is known for large n only. For one special case of practical importance (see Kendall and Stuart, example 24.1), however, the distribution of the likelihood ratio is an explicit function of Student's t ;this will be further discussed below. Without such an explicit distribution function the likelihood ratio can still serve for exploratory tests which confirm H_0 when the likelihood ratio is near 1 and rejects it when the ratio is large. A sequential "change index" for such a test will now be described.

It employs four likelihoods $L(n;\theta_n)$, where θ_n is a parameter estimate derived from n observations x_i, $i = 1,2,3...$ with probability $p(x_i\ ;\theta_n)$ of occurring. Two of the likelihoods

use optimal ("maximum likelihood") parameter estimates, θ_m for a base sample of m and θ_{m+j} for one of a series of "augmented" samples of n =m+j, j=1,2,.... The remaining two use θ_m for the augmented sample, and θ_{m+j} for the base sample. Thus $L(m;\theta_m)=L_1$, say; $L(m+j, \theta_{m+j}) = L_2$; $L(m+j,\theta_m) = L_3$; and $L(m,\theta_{m+j}) = L_4$.

The likelihood ratios $q(m+j) = L_2/L_3$ and $q(m)=L_1/L_4$ compare maximum likelihood estimates of θ, using the full samples, with "cross-over" likelihoods using parameter estimates based on more (L_4) or fewer (L_3) data, respectively. If the j new values represent a regime different from that of the initial m "base" values, both q(m+j) and q(m) will be larger than 1 and increase rapidly with increasing j. Without a regime change both ratios remain near 1 or increase slowly as the parameter estimates for the augmented samples drift towards some population values ("return to normal").

The sum of the two likelihoods for a sample of m or m+j represents a measure of the probability that either parameter estimate is acceptable. For comparisons of different sample sizes these sums can be used to normalize their likelihoods. With $L_4/(L_1+ L_4) =\gamma_m$, say, the other base likelihood becomes $L_1/(L_1+L_4) = 1 -\gamma_m$; that normalization leaves their ratio intact as $L_1/L_4 = (1 -\gamma_m)/\gamma_m$. The corresponding normalization of the augmented-sample likelihoods L_2 and L_3 leads to $L_2/L_3 = (1 -\gamma_{m+j}) /\gamma_{m+j})$. Writing the product of the two ratios as $Q = L_1 L_2 /L_3 L_4 \equiv (1 -\gamma)/\gamma$ finally defines the "change index (CI)" [2] in terms of the geometric average $Q^{1/2} = [q(m)q(m+j)]^{1/2}$ as

$$\gamma = (1 + Q^{1/2})^{-1} \qquad (2.1)$$

Its maximum value , in absence of any parameter change, is 0.5, except in some special cases, discussed in section 3.3, when incompatible parameters can create values between .5 and 1.

[2]Alternative names used during the project include "compatibility index" or "consistency index", and "no-change probability(NCP)"

3. Properties of the change index (CI).

Likelihood ratio products Q are derived in Appendix A for the major probability laws (Poisson, Gaussian, Chi-square). For small samples of observational data the choice of law is somewhat arbitrary; it is then useful that the normalized likelihood ratios involved in Q are less sensitive to the distribution form assumed than the likelihoods themselves. Except in the special case mentioned earlier and discussed in 3.2 below, the frequency distribution of the CI is unknown, but some of its properties have been determined experimentally, as described in the following sub-sections.

3.1. CIs from Poisson samples.

For Poisson samples of m and m+j , with means $\bar{x}_m$ and $\bar{x}_{m+j}$, respectively, the normalized likelihood ration product Q is derived in Appendix A (equation A.1.6) as

$$Q = \exp\left[\bar{x}_m\left\{[(m+j)y + j]\log_e(1+y) - jy\right\}\right] . \qquad (3.1)$$

where $y = (\bar{x}_{m+j} - \bar{x}_m)/\bar{x}_m$. This dependenc of Q on the base mean and the means difference scaled with the same base mean has been used by Radok and Brown (1993) to demonstrate the general dependence of the CI on the base sample mean and on the sample sizes m and m+j when all the new values x_j j=1,2... equal the average of a Poisson distribution with mean $\bar{x}' = \bar{x}_m + \Delta = x_j$. Then with $g = \Delta/\bar{x}_m y$, say, y = jg/(m+j), and equ. 3.1 takes the form

$$Q' = \exp\{j\bar{x}_m (1+g)\log_e[1+jg/(m+j)] - jg/(m+j)\} \qquad (3.1a)$$

Change indices have been calculated with (3.1a) for different combinations of the parameters m, x_s, j, and g. The results are shown in figs. 3.1a and b. Longer bases (larger m) slow the response of the no-change probability to the change in mean but leave the curves essentially unaltered in shape. The decline in the change index starts at the change of mean and accelerates down to values of the order of 0.3 - 0.2 before becoming more gradual as small CIs are approached.

While of theoretical interest, these curves do not offer any help for assessing small samples which at best can be <u>believed</u> to have come from a Poisson distribution. The curves do however illustrate the basic tendency of the CI to decrease as a (Poisson) sample is being augmented.

3.2 CIs derived from Gaussian samples.

The product Q of the normalized likelihood ratios for Gaussian samples is derived in Appendix A (equation A.2.8) as

$$Q = \exp\left[\frac{m+j}{2}\log_e\frac{\sigma_m^2}{\sigma_{m+j}^2} + \frac{m}{2}\log_e\frac{\sigma_{m+j}^2}{\sigma_m^2} - \frac{m}{2}\left[1-\frac{\sigma_m^2+(\Delta\mu)^2}{\sigma_{m+j}^2}\right] - \frac{m+j}{2}\left[1-\frac{\sigma_{m+j}^2+(\Delta\mu^2)}{\sigma_m^2}\right]\right] \qquad (3.2)$$

The first two terms of (3.2) could be combined, but the form given will prove convenient for a development later in this section. The dependance of (3.2) on sample sizes and parameters cannot be evaluated in a manner analogous to that used for Poisson samples. Instead CI calculations have been carried out for small sets of independent values drawn at random from Gaussian populations with mean 10 and variances 4 and 25 {N(10;4), N(10;25)} as well as from the same data rendered

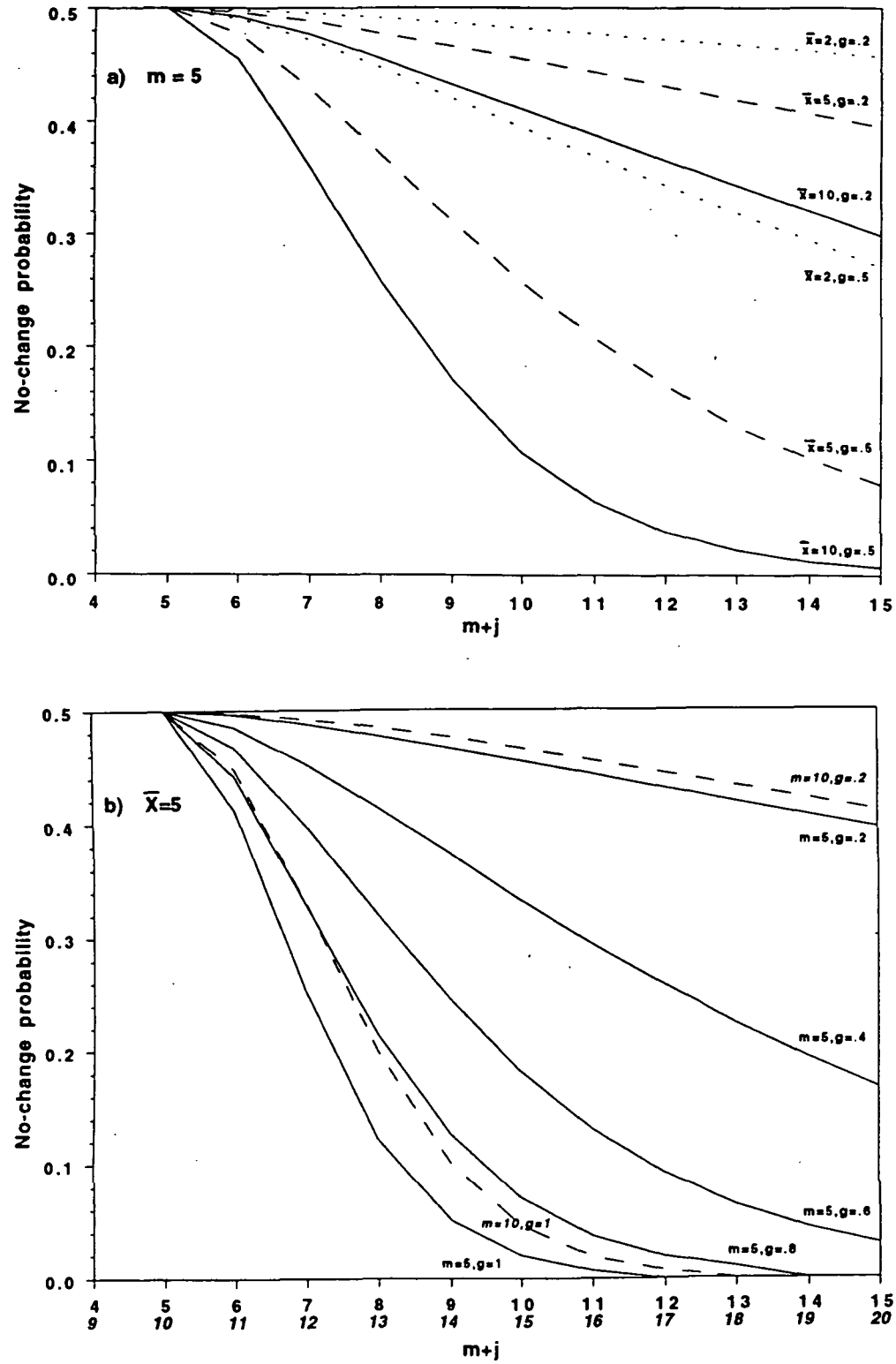


Fig. 3.1 a and b: The change index (CI; here called "no-change probability") of Poisson samples that result when the same new value $\bar{x} + \Delta$ is added j times to a base sample of m with mean $\bar{x}$. The curves represent different values of $\bar{x}$, m, and $g = \Delta / \bar{x}$ (from Radok and Brown, 1993).

autoregressive with moving averages of lengths $\Lambda = 3$ and $\Lambda = 7$ {N(10;4/3),N(1);4/7), etc.}, creating non-zero lag correlations $r_\lambda = (\Lambda - \lambda)/\Lambda$, $\lambda = 1,2...(\Lambda - 1)$.

In order to describe the CI decrease with the number j of added new data, the medians M_m of the CI as functions of j for different base lengths m were fitted with linear regression lines of the form $M_m = a_m - b_m j$. Close approximations to the actual median CIs resulted in each case. The regression slopes b_m, representing the incremental CI change per new value added to the sample, are given in table 3.1 :

Table 3.1 : Change rate $-b_m$ per added datum of the CI medians $M_m = a_m - b_m j$ as functions of base length m:

Gaussian data set parameters			m = 5	7	10	13
mean	variance	mov aver. Λ				
10	4	0	.021	.014	.011	
10	4	3	.050	.045	.037	
10	4	7	.079	.073	.073	
10	25	0	.035	.019	.013	.009
10	25	3	.108	.058	.034	
10	25	7	.144	.080	.045	.052

The entries in the table were obtained from 50 random samples of 20 values each. As expected, the CI decreases in a regular way when nmore values are added to the base sample, and the rates of decrease are enhanced for the autocorrelated series of moving averages. For quantiles lower (higher) than the median the decrease is faster (slower), as demonstrated by the following trends for samples from the N(10,25) data set and from the 7-term moving averages of the same data, with 13-term bases:

Data set	Quantile:	20%	50%	80%
N(10,25)	$-b_m$ =	.015	.009	0
7-term mov.aver	$-b_m$ =	.091	.052	.025

A more general and complete description of Gaussian CI properties can be given for a test which allows for changes in means only. In that case, when μ_{m+j} is used as alternative mean for the base sample of m values, the alternative variance σ_{m+j} becomes $\sigma_m^2 + (\Delta\mu)^2$, where $\Delta\mu = \mu_m - \mu_{m+j}$. Similarly, when μ_m is to be tested as a potential alternative mean for the augmented sample of m+j, the alternative variance becomes $\sigma_{m+j}^2 + (\Delta\mu)^2$. With these variances inserted in (3.2), the last two terms cancel, and the remaining terms yield as likelihood ratio product

$$Q' = [1+ (\Delta\mu)^2/\sigma_{m+j}^2]^{(m+j)/2}[1+(\Delta\mu)^2/\sigma_m^2)^{m/2} \qquad (3.2a)$$

With sample means $\bar{x}_n$ and variances $s_n^2 = [n/(n-1)]\sigma_n^2$, the variable $(\Delta\bar{x})^2/[s_n^2 (n-1)]$ represents the square of a "Student" variable t_{n-1} with n-1 degrees of freedom. Thus the product of the likelihood ratios becomes

$$Q' = [1+(m-1)t_{m-1}^2]^{m/2} \; [1+(m+j-1)t_{m+j-1}^2 \;]^{(m+j)/2} \qquad (3.2b)$$

The CI for this special case then has the explicit frequency distribution

$$f(\gamma)=[1+\{1+(n-1)t_{n-1}^2\}^{n/2}]^{-1} \qquad (3.2c)$$

in terms of a "Student" variate with m-1< n < m+j-1 degrees of freedom.

Fig. 3.2a shows the cumulative form of equ. (3.2c) to be

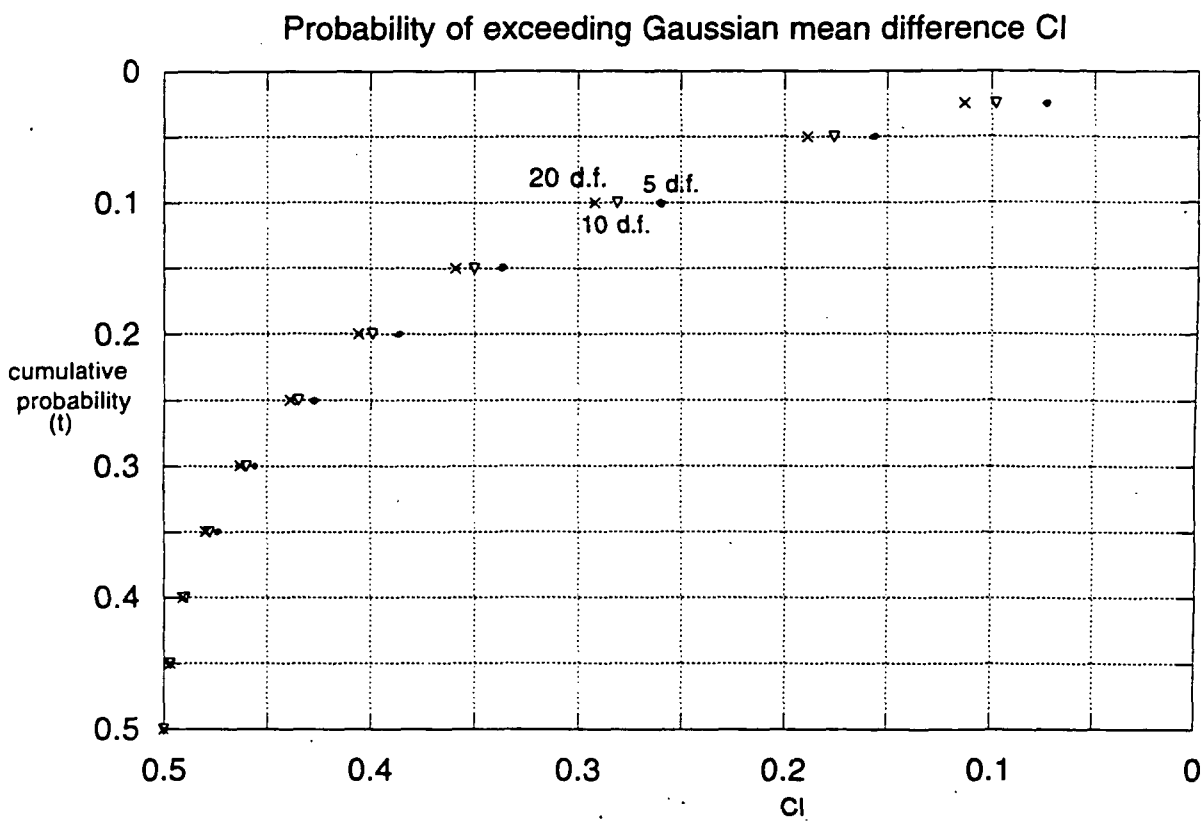


Fig. 3.2a: Chance probabilities of exceeding change indices (CI) of Gaussian samples with the same variance but different means. The scale of cumulative probability has been determined with equ. (3.2c) from "Student" t distributions with different degrees of freedom (d.f.).

similar for different degrees of freedom. With the probabilities for n=10 in fig. 3.2a, a scale has been constructed on which those probabilities fall along a straight line. That scale is used in figs. 3.2b and 3.2c to examine the distribution of some of the CIs obtained from random samples with the full Gaussian equation (3.2). All the plots are approximately linear and distorted only for high values of the CI. Critical regions for small values of the change index might then be constructed with the distribution function (3.2c) even when both mean and variance are allowed to vary. But as in the case of Poisson samples, such regions will be have little practical value when a small data sample constitutes the complete information available.

3.3 CIs for regression coefficients.

The formulae derived in Appendix A.3 for the change index of regression coefficients again clearly have no simple interpretations. They have therefore been evaluated for some of the Gaussian samples described in the preceding section. The results are summarized in the following table for the independent data with population mean 10 and population variance 25 (N(10;25)}, as well as for the autocorrelated 7-term moving averages of the same data, N (10;25/7):

Table 3.2 Frequency (%) of regression-coefficient change indices for Gaussian samples. Frequencies in parentheses relate to CIs < 0.4.

CI Range	Sample size	m=5 j=5	m=5 10	m=5 15	m=10 j=5	m=10 10
	Data set					
0.5(.4) >CI>0:	N(10;25)	62(36)	44(24)	36(18)	88(38)	72(42)
	N(10;25/7)	32(12)	36(20)	28(18)	44(26)	44(30)
CI>0.5	N(10;25)	38	56	64	12	28
	N(10;25/7)	68	64	72	56	56

The first two rows of frequencies in the table represent CIs in the range 0.5<CI<0. For the longer of the two base lengths tested, a majority of trend change indices for random samples

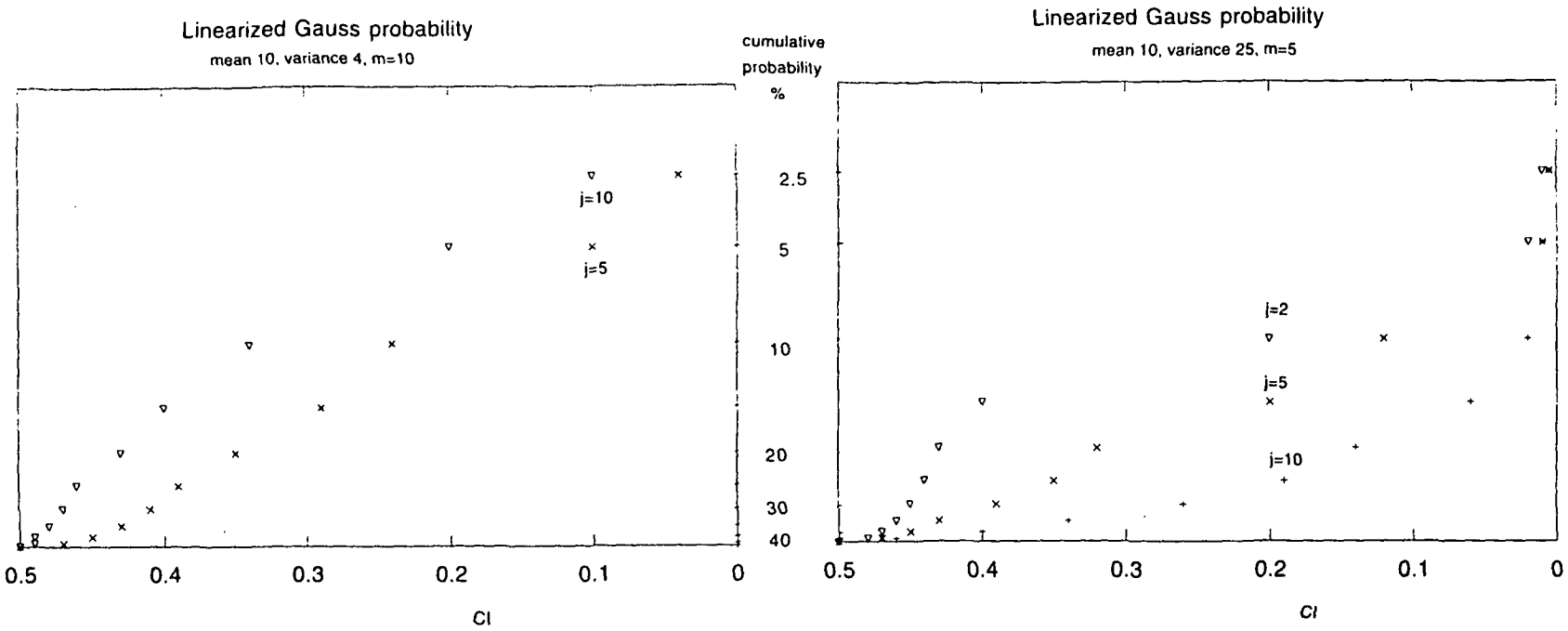


Fig. 3.2b and 3.2c: Cumulative probabilities of change indices determined with equ. (3.2) for 50 Gaussian samples with changing means and variances. Probabilities of CIs with changing means only (satisfying equ. (3.2c)) would form straight lines.

remain above 0.4. The corresponding frequencies for the moving-average (autocorrelated) series are smaller but increase in the same way for longer base samples.

The remaining two lines of the table refer to CIs that are larger than 0.5. Such unrealistic CIs arise from the violation of a limiting condition noted in the derivation of the regression formulae (A.3.7) and (A.3.8) in Appendix A.3 : the use of the regression coefficient b_{m+j} with the base sample of m values, and of b_m with an augmented sample of m+j values must not result in a zero or negative residual sum of squares. This implies that a good regression fit with small residuals will narrow the limits for permissible alternative regression coefficients, increasing the sensitivity of the change index.

3.4 CIs from chi-square samples.

The product Q of the normalized likelihood ratios for samples from a chi-square distribution with ν degrees of freedom (d.f., which represent also the mean and one half the variance of the distribution) is derived in appendix A (equation A.4.8) as

$$Q = \exp\left[\frac{j}{2}(\nu_m - \nu_{m+j})\log_e 2 + j\log_e \frac{\Gamma\left(\frac{\nu_m}{2}\right)}{\Gamma\left(\frac{\nu_{m+j}}{2}\right)} + \frac{\nu_{m+j} - \nu_m}{2}\sum_{m+1}^{m+j}\log_e \chi^2\right] \quad (3.4)$$

Some of the characteristics of this expression have again been determined experimentally for chi-square values calculated from the Gaussian data sets N(10;25) and N(10;25/7) introduced in section 3.2. For sample variances computed from 5 values, the chi-sqares of the first data set should have 4 d.f. and those of the autocorrelated set, 1.14 d.f. (cf Radok 1992). The following

table gives the percentage frequencies of CIs calculated for 50 samples of 100 in each case:

Table 3.3 Frequency (%) of change indices for chi-square samples. Bracketed frequencies relate to CIs < 0.4 .

	Sample size	m=5			m=10	
CI Range		j=5	10	15	j=5	10
	Data set					
0.5(.4) > CI >0:	(N10;25)	76(58)	78(46)	74(50)	86(72)	78(62)
	N(10;25/7)	72(48)	70(34)	54(18)	48(38)	50(30)
CI>0.5	N(10;25)	24	22	26	14	22
	N(10;25/7)	28	30	36	52	50

These frequencies show a concentration of random-sample CIs in the 0.4 to 0.5 range but also a substantial fraction of unrealistic values exceeding 0.5, the largest value possible if the estimates $_{m}$ and $_{m+j}$ represent maximum-likelihood estimates for their respective samples of m and m+j. The Gamma functions in the chi-square likelihoods dominate the value of Q, especially in the region of small x where $\Gamma(x)$ has a minimum around x= 1.46 and goes to infinity as x goes to zero. This suggests that chi-square CIs require samples larger than those used for table 3.3.

3.5 Limitations and scope.

The experimental results presented in the preceding sections despite their limited extent establish clearly that the change index and the underlying likelihood ratios have downward trends for growing samples even when the data have been drawn at random from a single population. The change index is not intended for finding critical regions for rejecting the null hypothesis - i.e. that the base and augmented sample estimates are equivalent, within a prescribed error probability. That represents a "confirmatory" task, as defined by Flueck and Brown (1993) for which standard tests (e.g. the "Student" t test) can be used. The

change index instead serves the "exploratory" purpose of detecting and measuring parameter changes that have resulted from the addition of even a single new observation, leaving the implications to be assessed from whatever information is contained in the small sample(s) under consideration.

The difference between the two types of statistical problem sketched above can be illustrated for one of the random Gaussian samples introduced in section 3.2 and plotted in fig. 3.3a. Inspection suggests that its first eleven values have a smaller variance than the nine values that follow, while both sections vary around similar means (as permitted by the characteristic independence of the two parameters in a Gaussian population).

To <u>confirm</u> that the data in fig. 3.3a come indeed from the same Gaussian population, the sample means and variance of m values from the first section and j values from the second have been compared with a t-variable of the form

$$t_{diff.} = (\overline{x}_j - \overline{x}_m)\{(m+j-2)mj\}^{1/2}/\{(m+j)[(m-1)s_m^2+(j-1)s_j^2\}^{1/2} \quad (3.5)$$

Probabilities P of exceeding these t values of the differences between the <u>independent</u> means $\overline{x}_j$ of three and more new observations and different base means $\overline{x}_m$ have been calculated with the corresponding sample variance estimates s_j^2 and s_m^2. The values of P (+1 or 2 or 3, in order to separate the curves) are shown in fig. 3.3a for different bases m, with j increasing from 3 to 20-m in each case. Except for m=5 the probabilities of exceeding t lie well above the 5% significance level and vary little for the different m; a sudden temporary <u>increase</u> in P occurs as the more variable section (and hence a smaller t value) is encountered. The t test therefore confirms, within

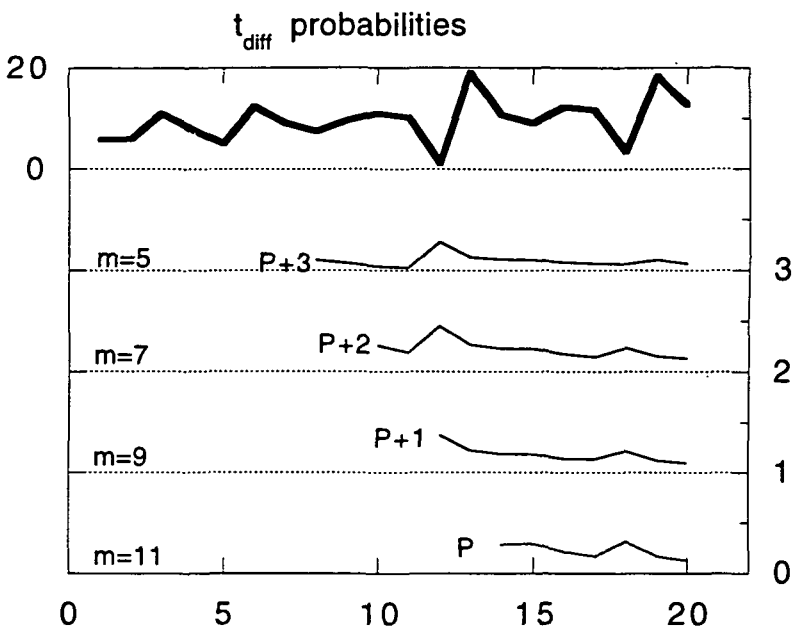


Fig. 3.3a: A Gaussian sample (heavy line), and probabilities that differences between the means of the initial m values and means of j=3,4...20-m subsequent values exceed tdiff. determined with equ.(3.5).

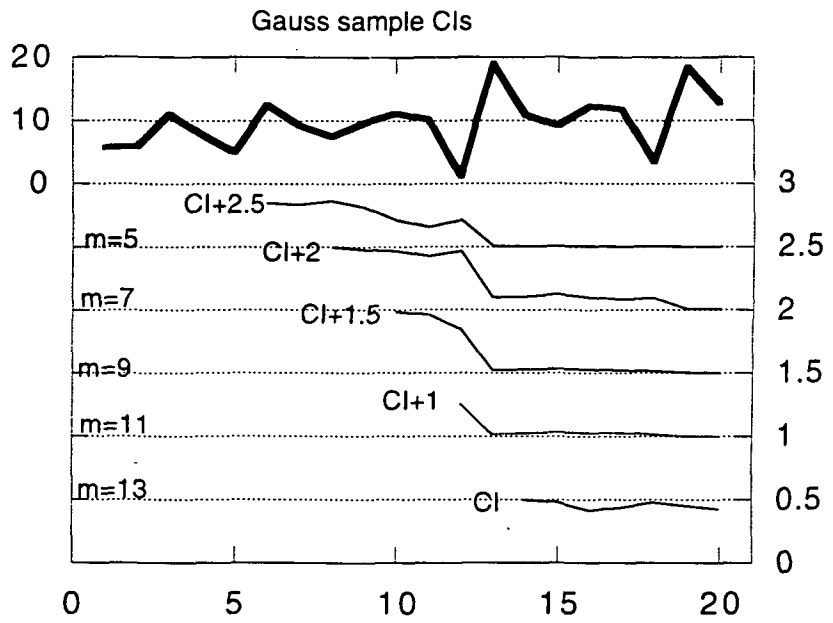


Fig. 3.3b: Change indices (CI) taking account of both means and variances of the Gaussian sample in fig. 3.3a (heavy line). Note the CI decrease signalling the onset of more variable values.

the limitations of a single 20-term sample, that the data in fig. 3.3a could indeed come from a single Gaussian population.

The CIs for the same data are shown in fig. 3.3b and tell a very different story. For the same bases as before and j starting with 1 rather than 3 (reflecting an advantage the CI has over the t test), the CI drops abruptly to almost zero as soon as the more variable section is encountered with the 12th value. But after the first two values of that section have become part of the base, with m=13, the CI remains near 0.5 for the rest of the series. This exploratory test therefore responds sensitively to the sudden substantial increase in variance. It leaves open the question of whether that increase is compatible with a single Gaussian population or perhaps points to some other type of distribution - a question that may not have a firm answer if the data in question represent the full information available.

Fig. 3.4 serves to illustrate this type of uncertainty. Its data form come from a series of annual numbers of North Atlantic hurricanes (Case 1988) which will be shown in section 4.1 to approximate a Poisson distribution. As a consequence, sample means and variances would be expected to vary in a similar way, but this clearly is not the case in figure 3.4: the mean number for the first ten years (4.9) differs little from that for the remaining 5 years (5.2), whereas the sample variance increases from 0.89 for the first ten years to 6.56 for the last five years. The CI formula for Poisson samples considers the means only, and the Poisson CIs (full lines in fig. 3.4) remain in the 0.4 to 0.5 range throughout.

However, as noted earlier, independence of mean and variance is a defining characteristic of the Gaussian distribution; the data in fig. 3.4 in fact bear a close resemblance to the Gaussian sample of fig. 3.3. They have therefore also been analyzed with the Gaussian formulae of section 3.2. A set of Gaussian CIs calculated with equation (3.2a) considering only means changes is shown as a dotted curve and resembles the Poisson curves. By contrast the broken CI curves, calculated with equation (3.2),

take into consideration both the sample means and sample variances and drop to small values in 1980, when the first of the more variable numbers is encountered. After that year becomes itself part of the base set, the full Gaussian CIs approximate those derived with the Poisson formula.

This underlines the need for flexibility in the choice of the distribution form for the CI calculation when alternative forms are not firmly ruled out by the limited data available. The examples of CI tests to be presented next will therefore not be concerned with confirming the distribution forms assumed for the parameter estimates and likelihoods, but rather with the parameter changes themselves that resulted from the addition of extra observations, and with possible physical implications.

4. Examples.

The examples in this section have been chosen to illustrate potential alternative uses of the CI. The index is uniquely suited for assessing a single new observation which seems out of line with the quasi-controlled regime of its predecessors. In existing longer series the index can also be used to identify transition points between regimes, created by unknown causes or by suspected events such as a station shift or change in observational routine. Currrent data can be subjected to real-time monitoring with CIs calculated both retrospectively for the immediate past and progressively for each new observation that comes to hand.

4.1 Retroactive use of the CI.

The data of the first example are annual numbers of tropical hurricanes, reported for the North Atlantic by Case (1988) and updated to 1990; some of them were already used in section 3.5. Their frequency distribution (fig. 4.1) approximates a Poisson distribution (light shaded bars) with mean (and variance) 5.6. The annual numbers themselves are shown in fig. 4.2a and have been used to calculate sequences of progressive means shown in

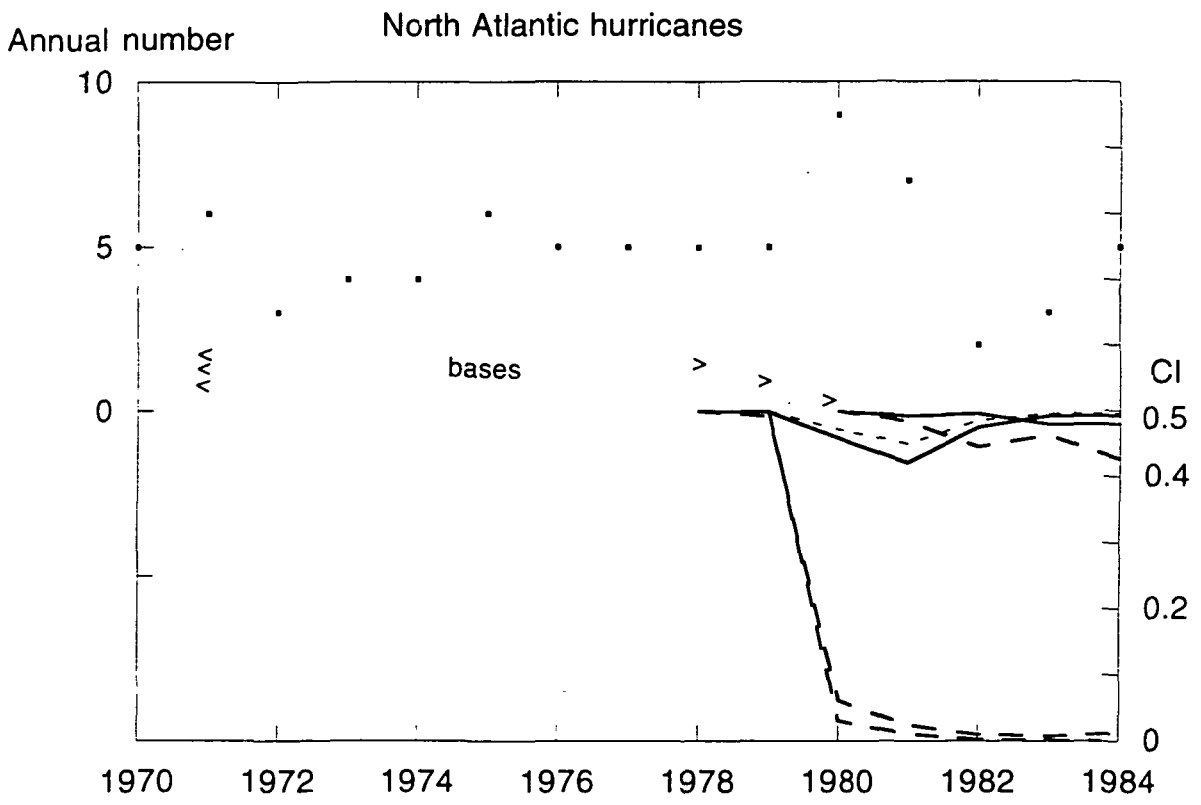


Fig. 3.4: Annual numbers of North Atlantic tropical hurricanes and change indices assuming a Poisson distribution (full lines), a Gaussian distribution with constant variance (dotted line), and a Gaussian distribution with changing mean and variance (broken lines).

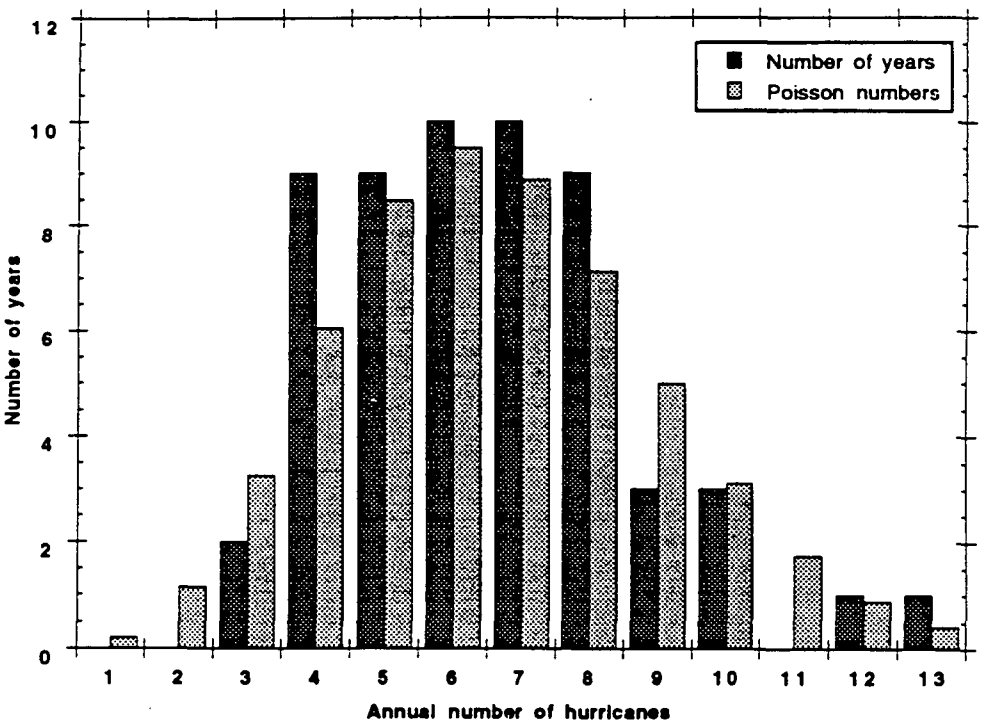


Fig. 4.1: Histogram of annual numbers of North Atlantic tropical hurricanes (Case 1988, updated). The light-shaded bars give the frequencies of a Poisson distribution with the observed mean number (5.6).

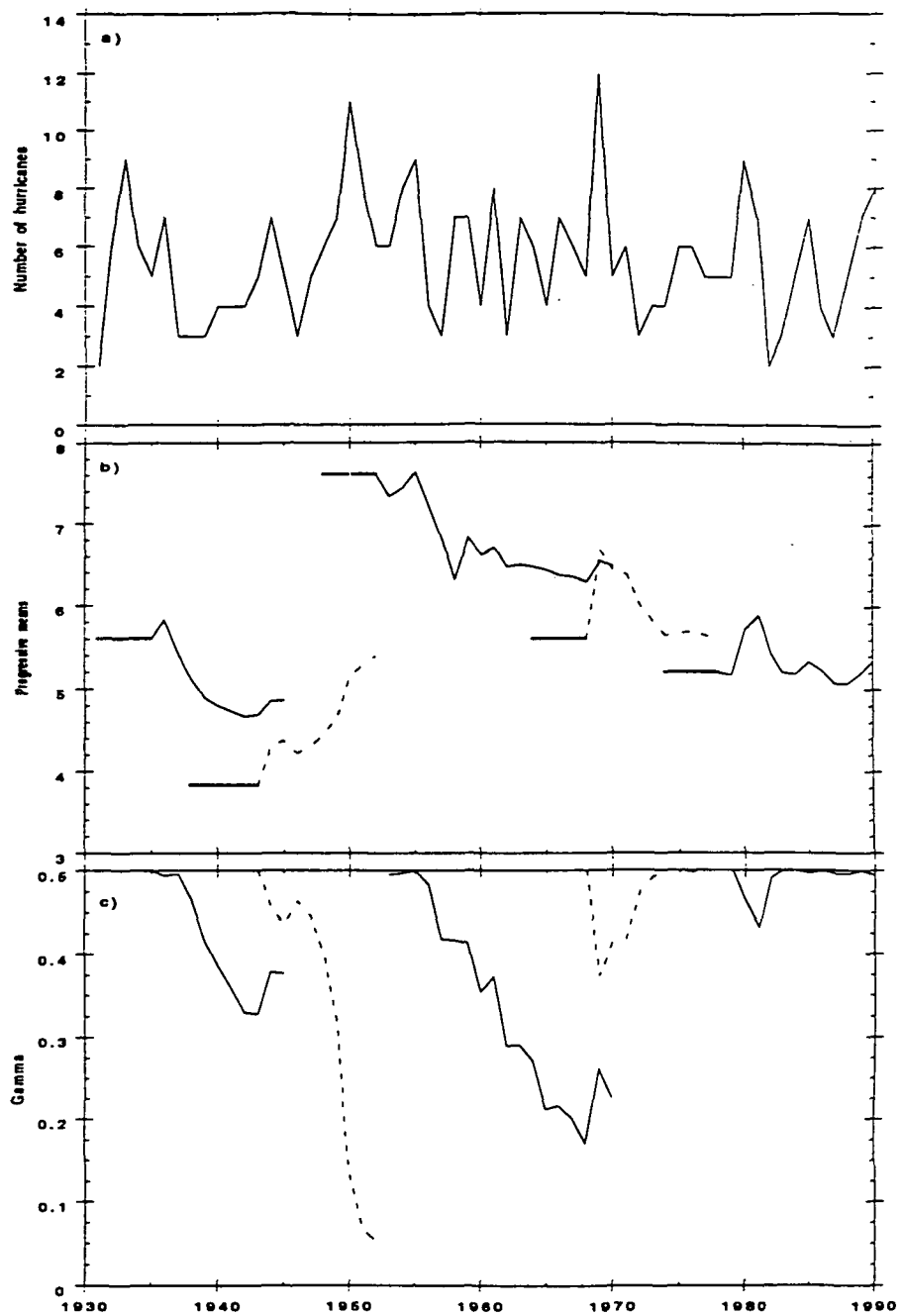


Fig. 4.2: (a) Time series of the annual number of North Atlantic tropical hurricanes; (b) Progressive means of the numbers in (a) starting with those of different base periods; (c) Change indices ("Gamma") starting from the ends of the base periods.

Fig. 4.2b. Each of these starts with the mean of a "base" section and continues until a parameter change is suggested by a substantial decrease in the change index CI (here called "no-change probability"), computed with equation 3.1 and shown in fig. 4.2c. This illustrate the use of the CI for delineating plausible "regimes" in an existing series, whereas the analysis of the 1970-1984 subset in section 3.5 simulated a real-time test initiated on the occurrence of a single anomalous value.

The first base period (1931-1935) had a mean value of 5.6 and was followed by several years with small hurricane numbers which produced a steady slow CI decrease. Starting anew with the numbers for 1937-1042 as base (mean 3.8) produced a sharper CI signal of a return to larger numbers in the late 1940s. The next base period (1948-1952) had an annual mean number of 7.6 hurricanes which decreased in the late 1960s to the original mean number of 5.6 for the next base period (1961-1965). Subsequent CIs showed no further change of control even after a new base period was adopted for the 1970s to sharpen the test.

The next illustration uses a composite annual mean temperature for four Alaskan airfields (Fairbanks, Anchorage, Nome, Barrow) reported by Bowling (1991). Fig. 4.3 shows the (cumulative) frequency distribution of the initial 23 values to approximate the Gaussian form, a straight line in the coordinate system of fig. 4.3. The annual mean temperatures themselves are shown in fig.4.4a. To detect changes in mean and/or variance, three base periods of 7 values each are used, ending in 1960, 1967, and 1974. Cumulative means and variances following each of the three base periods are shown in fig. 4.4b and Fig.4.4c, respectively.

The cumulative variance decreased steadily from 1960, the end of the first base period, while the cumulative mean remained near its base level. Both mean and variance increased rapidly after the 1977 change to higher winter temperatures reported by Bowling (1991).

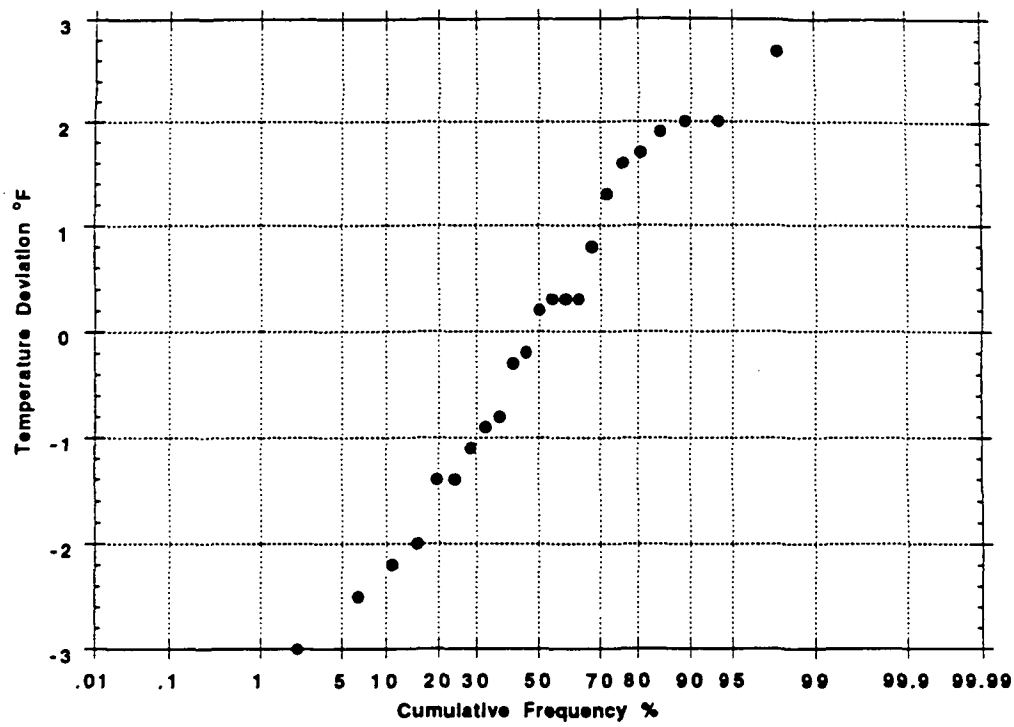


Fig. 4.3: Probabilities of exceeding different composite annual mean temperatures at four Alaskan airfields (deviations from the 1954-1976 mean; Bowling 1991). The probability scale turns a Gausian distribution into a straight line (cf. e.g.Radok 1992).

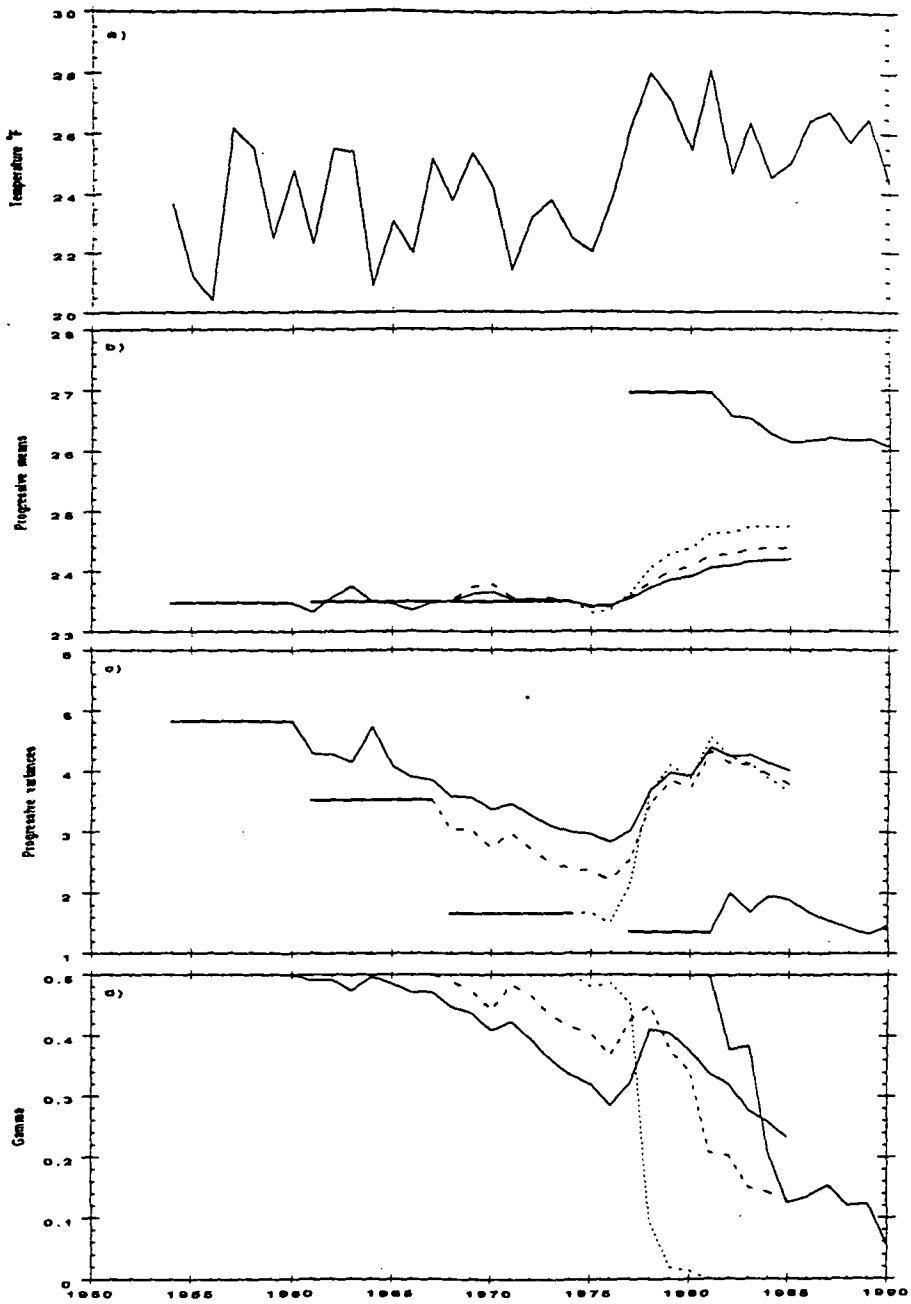


Fig. 4.4: (a) Time series of the Alaskan airfields mean temperature deviations in fig. 4.3, updated; (b) Progressive means starting with those of different base periods; (c) progressive variances; (d) Means/variances change indices ("Gamma") starting from the ends of base periods.

The corresponding CI values are shown in fig. 4.4d. The initial progressive decline in variance was reflected in a similarly gradual CI decline after each of the first two base periods. That CI decline was reversed temporarily by the 1977 temperature increase before continuing as a steeper CI descent. With the base period ending in 1974, the combined increases in both cumulative mean and variance produced a much steeper CI decrease after 1977, which became only a little more gradual when the entire set of 21 years was used as base period. CIs calculated with the temperatures for 1977-1981 as a new base suggested a return to slightly lower temperatures and unchanged small variance values during the 1980s, which have persisted through 1993 (Radok and Brown 1995).

The detection of trend changes can be illustrated with the first part of the same Alaska data, repeated in fig. 4.5 from fig.4.4a. Fig. 4.5 shows progressive regression coefficients starting with that for the 7-year base 1954-1960. The CIs in fig. 4.5 indicate the beginning of a distinctly different slope regime in 1964. The series of CI values is continued until two CI=1 are encountered, to illustrate what happens when the base regression coefficient b_n exceeds the upper limit $b_{n\ max}$ (the dotted line) for its use in an augmented sample,

None of the regression lines in this example explains more than 10% (= r^2, where r is the temperature-time correlation) of the total variance. For a good regression fit the permissible range for b_n as alternative for b_{n+1} becomes very narrow. This is demonstrated in table 4.1 for an exponential approximation to annual mean atmospheric carbon dioxide conentrations C(ppm) observed at Mauna Loa, Hawaii (data from Boden et al, 1990):

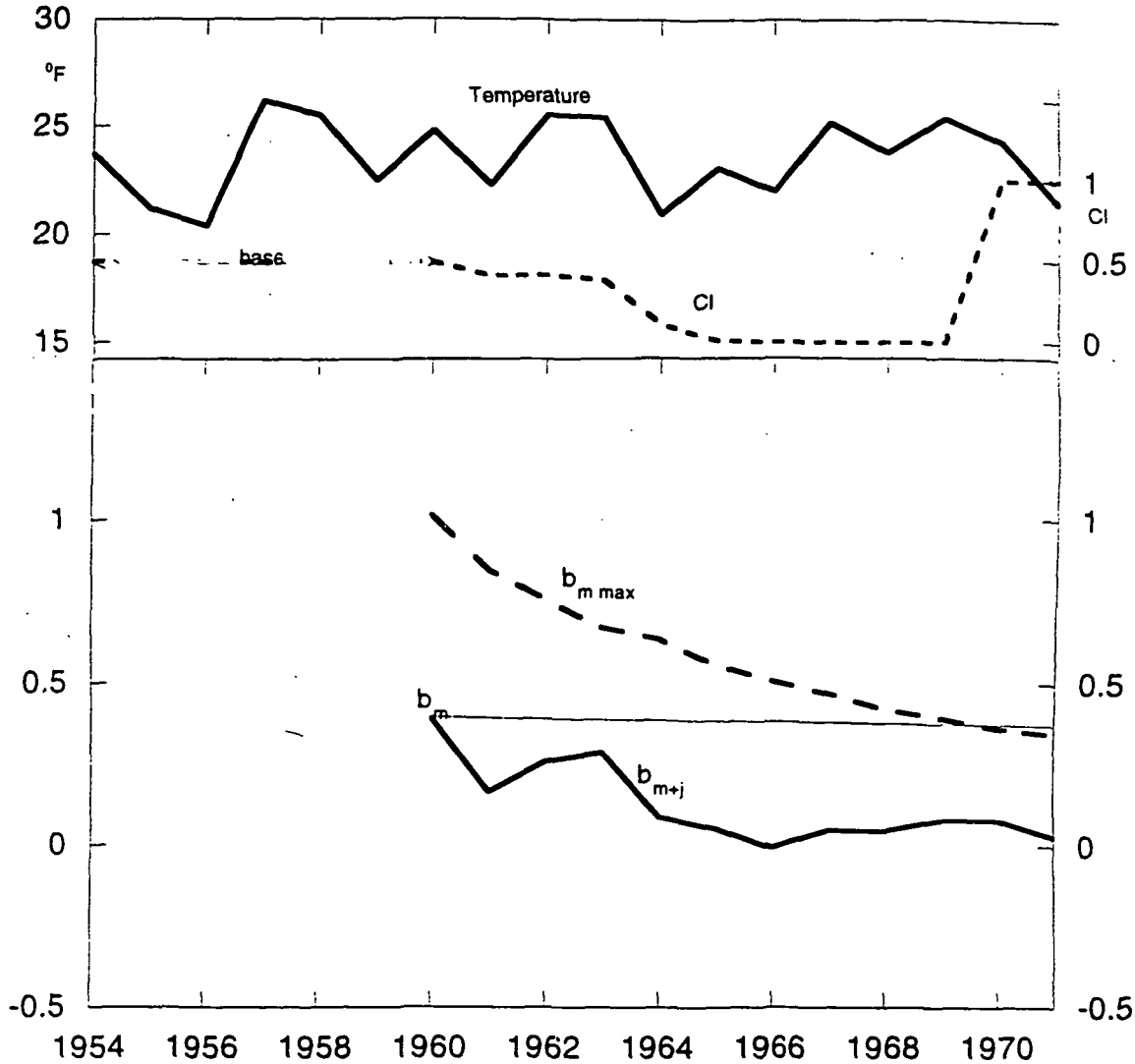


Fig. 4.5: Change indices (CI) for the progressive trend b_{m+j} in the Alaskan airfields mean temperature deviations (top curve) from the base trend b_m. Unrealistic values CI=1 arise after the b_{m+j} curve descends below the value of b (dotted line).

Table 4.1: Regression analysis of CO_2 concentrations C (ppm) represented by y = 100 (logC-5.8):

Base period 1983-1988, b_B = 4.83

year	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993
y	37	41	46	49	54	61	66	70	74	75	77
Change index CI									0.95	0.85	0.97
Progressive regression coefficient b_{B+j}									4.78	4.55	4.31
upper limit $b_{B\,max}$ for use in augm. sample:									4.80	4.57	4.35
residual variance $(1 - r^2)$ %								0.7	0.5	1.1	1.8

The base regression of 4.83 leaves a mere 0.7 of the total variance unaccounted for. When the observations for the years 1991 through 1993 are added, this percentage remains almost the same, while the base regression coefficient immediately begins to exceed its permissible limits.

The final example, from Brown and Radok (1995), illustrates how the CI can be used to detect inhomogeneities in a climate series, caused by a station move or by a change in observational routine. Such discontinuities resemble sudden regime changes, or appear as outliers if the change was of short duration. An example is provided by temperatures observed at the San Francisco, California, National Weather Service airport station, which is known to have been moved further inland during the 1940s. By reducing the number of maritime air incursions this should have raised the mean temperatures for the summer months.

Fig.4.6 shows those mean summer temperatures for the years 1933 through 1956. The CI calculations in this case used a moving ten-year base period advancing by two-year steps. For each of the first three bases, the gradual CI declines changed to a rapid drop with the 1948 temperature; this drop vanished as soon as the moving ten-year base reached and included the 1948 temperature itself. The station move in fact took place that year (Jon Eischeid, pers. comm.).

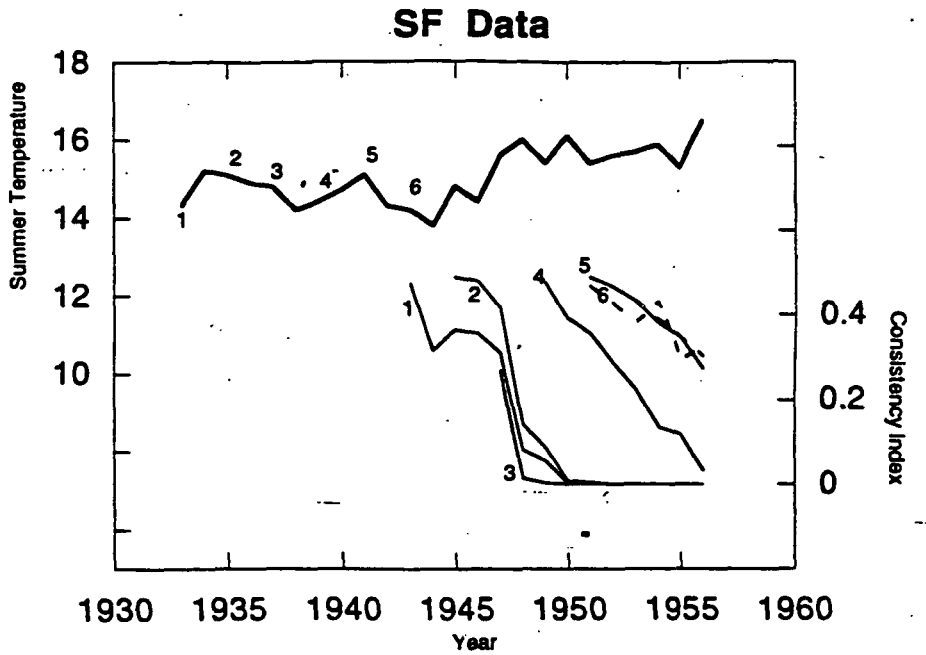


Fig. 4.6: Summer mean temperatures observed at the San Francisco Airport, and change indices for base periods extending between corresponding numbers. The CI decrease in 1948 arises from a change of station location.

4.2 Real-time CI analysis.

The first example of an (almost) real-time CI analysis also serves to demonstrate its broad applicability. The data used are shown as dots in Fig. 4.7; they represent Dow Jones 65-stock averages, as reported by the Wall Street Journal prior to and during the 1994 Congressional elections. Dow values of November 7 are used as an expanding base with m= 4 through 8. All the CIs decreased slowly during November 7 and early on November 8, until a noon increase in the Dow started a precipitous CI decrease to values near zero.

A possible form of real-time climatic monitoring is illustrated in tables 4.1 and 4.2 for global mean temperature anomalies y from their 1951-1980 averages, using data kindly provided by Dr. Henry Diaz. For each season 5 different bases extend backward from 1990, to provide retrospective CIs testing the statistical control or lack of it during the early 1980's. The same bases provide progressive CIs for 1991 through 1993. With the years 1985 through 1990 as base, the earlier anomalies for winter and, to a lesser degree, for summer appear to belong to different regimes, but all the most recent anomalies have remained within the framework of the statistical control established during the 1980s.

5. Alternatives.

Several other procedures have been proposed over the years for the specific purpose of detecting abrupt changes in time series (e.g.Oerlemans , 1979; Epstein, 1982; Goossens and Berger, 1987; Howell, 1995). All of them postulate some form of change to be confirmed by filtering. In the example shown in fig. 5.1a, Epstein (1982) postulated three possible change scenarios to account for the rise at the end of the 1970s of the annual mean temperatures recorded by a global network of radiosonde stations (Angell and Korshover 1977, updated to 1981). The parameter values of the scenarios were chosen to maximize Gaussian likelihood ratios, shown in fig. 5.1b. These ratios are

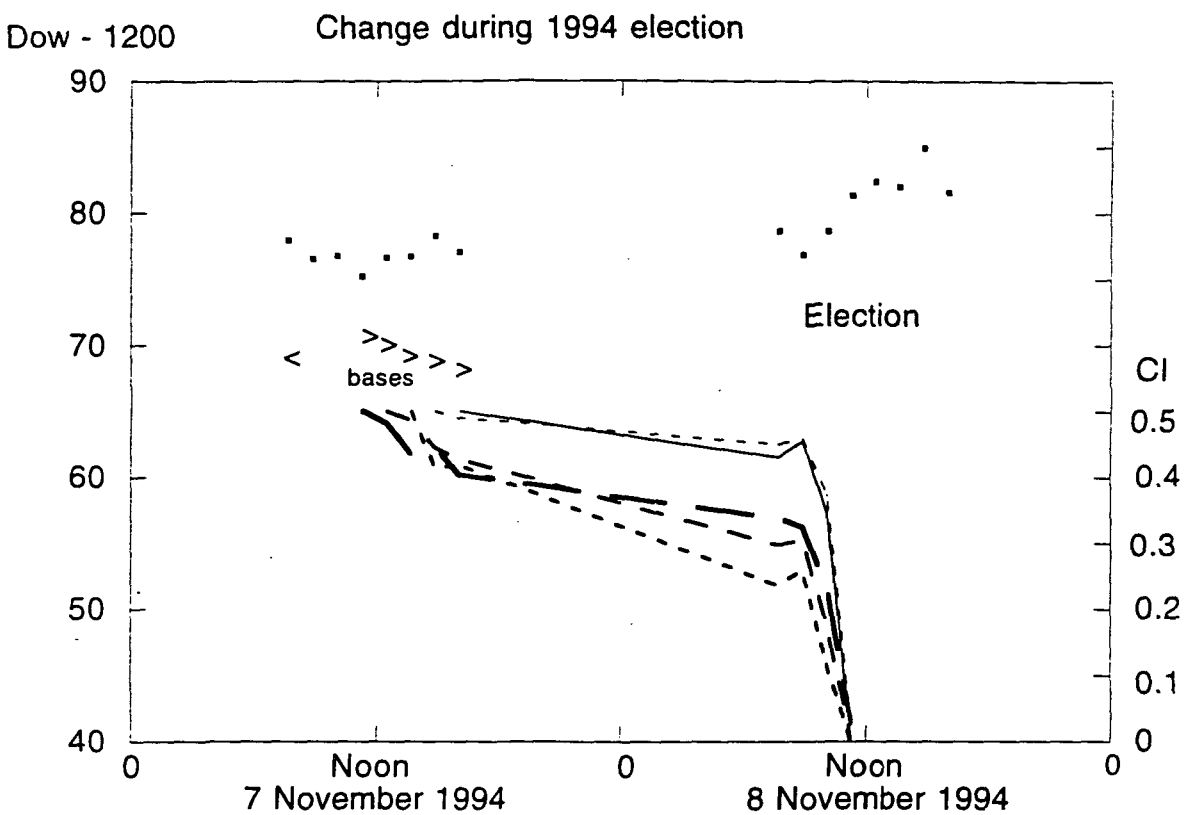


Fig. 4.7: Dow-Jones 65-stock averages (dots) and change indices before and during the November 1994 Congressional elections.

Table 4.1 Change in current global land temperature anomalies y °C from 1951-1980 (H.Diaz, pers. comm). Numbers in table are change indices CI*10^3

1) Winter (DJF)

y	yr	base: 5	6	7	8	9	10	
.437	81	000	465	454	488	488		
.093	82	000	477	485	491			
.449	83	000	493	484				
-.001	84	000	488					
-.380	85	000						
							.210	= base mean
.159	86					.185	81-90	= base years
.247	87				.197	82-90		
.426	88			.161	83-90			
.291	89		.188	84-90				
.382	90	.301	85-90					
		86-90						
.530	91	369	477	472	480	477	482	
.391	92	353	448	442	457	453	463	
.329	93	368	413	469	429	428	441	

2) Spring (MAM)

y	yr	base: 5	6	7	8	9	10	
.356	81	315	429	476	491	492		
-.025	82	301	427	478	478			
.264	83	391	468	490				
.070	84	408	485					
.026	85	465						
							.242	= base mean
.215	86					.221	81-90	= base years
.112	87				.262	82-90		
.397	88			.261	83-90			
.269	89		.293	84-90				
.739	90	.346	85-90					
		86-90						
.442	91	486	488	489	489	489	490	
.276	92	462	467	471	475	476	479	
.296	93	419	430	439	447	450	456	

Table 4.(cont'd): Change in current global land temperature anomalies y°C from 1951-1980 (H.Diaz, pers. comm).
Numbers in table are change indices CI*10^3

3) Summer (JJA)

y	yr	base: 5	6	7	8	9	10			
.186	81	177	406	466	477	492				
-.016	82	182	421	483	478					
.226	83	305	475	490						
.004	84	285	474							
-.014	85	408								
							.174	=	base mean	
.047	86					.173	81-90	=	base years	
.269	87				.196	82-90				
.421	88			.192	83-90					
.224	89		.223	84-90						
.393	90	.271	85-90							
		86-90								
.439	91	476	475	471	469	466	462			
.012	92	463	494	494	490	489	485			
.156	93	463	498	499	499	497	496			

4) Fall (SON)

y	yr	base: 5	6	7	8	9	10			
.086	81	354	470	492	472	492				
-.070	82	372	479	498	487					
.367	83	436	497	481						
-.116	84	372	473							
-.081	85	462								
							.120	=	base mean	
-.108	86					.124	81-90	=	base years	
.203	87				.148	82-90				
.273	88			.117	83-90					
.182	89		.156	84-90						
.467	90	.203	85-90							
		86-90								
.275	91	486	488	489	491	491	492			
-.185	92	454	483	494	484	494	493			
-.075	93	391	450	475	464	476	476			

equivalent to those in equ. (3.2a); however Bayesian scruples prevented the author from drawing firm conclusions about their significance.

Fig. 5.2 shows CI profiles for the same temperatures. CIs calculated with equ. (3.2a) (dotted line) suggested no change when means are considered alone; by contrast when the variance is taken into account as well, with equ. (3.2), the CIs in fig. 5.2 decrease rapidly after 1977 for a 6-year base extending from 1971 through 1976, and more gradually for a base including all data prior to 1977. The change index therefore would have provided an objective basis for the initial decision to postulate different change scenarios, but no guidance to their possible forms.

6. Conclusion.

The likelihoods computed for an existing sample, and for the same data augmented by a single new value or by a small number of such values, indicate changes in progress that can represent the end of a "regime" or a new beginning. The "change index (CI)" provides a quantitative measure of differences between one or more parameters estimated from the original and augmented samples. The normalisation of the likelihood ratios used in the CI reduce its sensitivity to the form of distribution assumed to govern the data, and makes the index an effective exploratory tool when a small amount of data represents the full information available.

Acknowledgement. Support for this work has been provided by NASA Grant NAGW-2706. Partial support for the second author was provided by NOAA's Climate and Global Change Program. We are grateful to Barbara Sloan who produced appendix A.

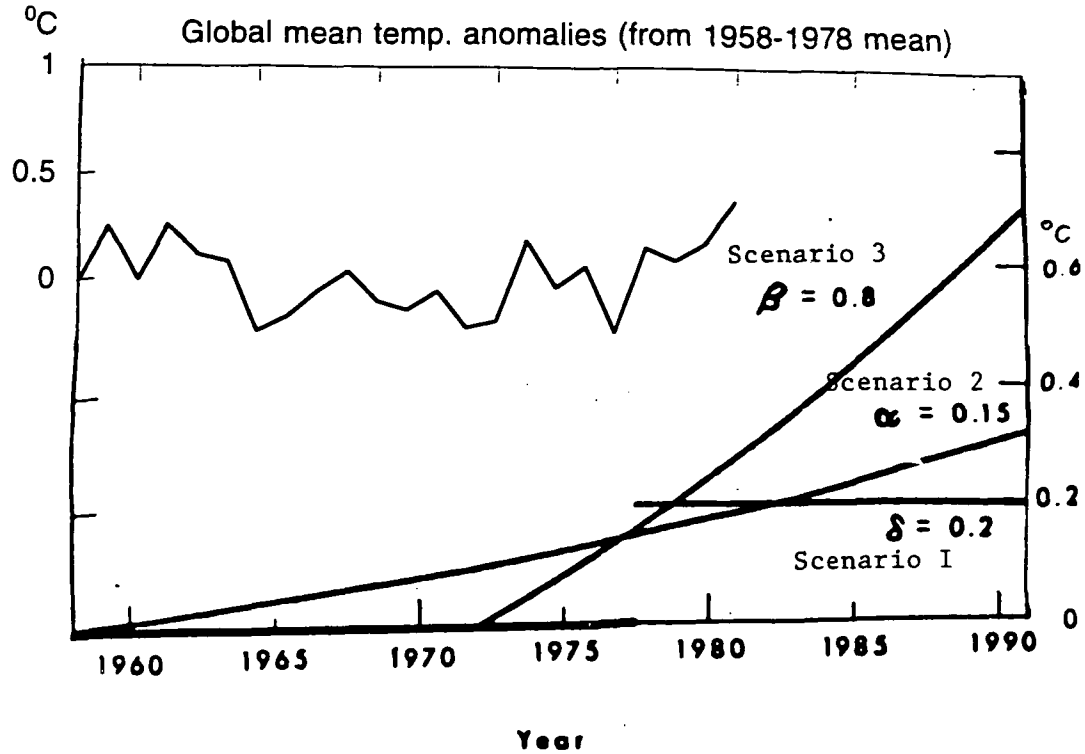


Fig. 5.1a: Global mean surface temperature anomalies (Angell and Korshover, 1977, top), and climatic change scenarios postulated by Epstein (1982).

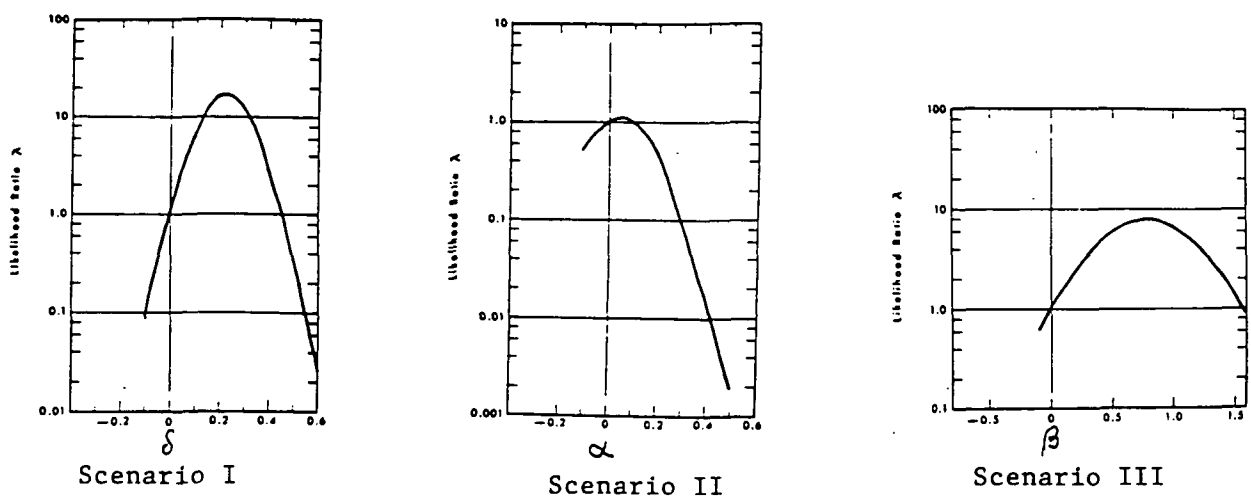


Fig. 5.1b: Gaussian likelihood ratios for different parameter values of the scenarios in fig. 5.1a (Epstein 1982).

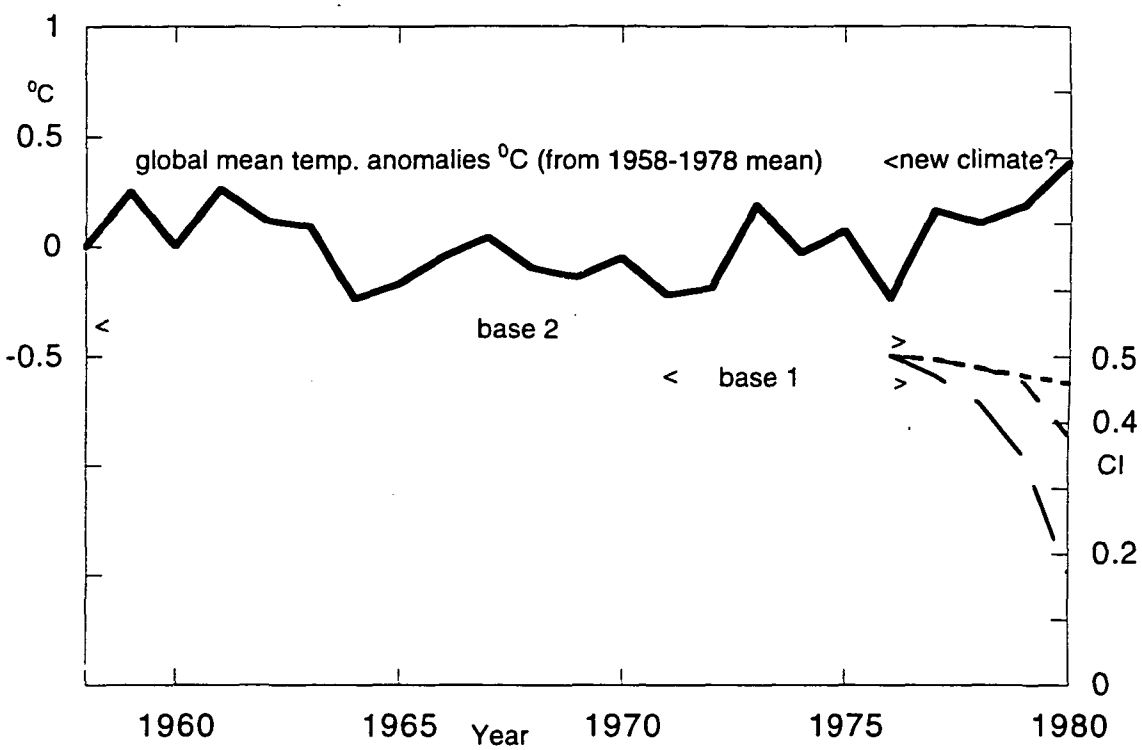


Fig. 5.2 The global mean surface temperature anomalies of fig. 5.1, and change indices testing the higher temperatures after 1976 against the previous 6 years (base 1) and against all observations prior to 1977 (base 2). The dotted CI line considers changes of mean only, the other two changes in mean and variance.

References .

Angell, J.K., and Korshover, J., 1977: Estimate of the global change in temperature, surface to 100 mb, between 1958 and 1975. Mon. Wea. Rev. 105, 375-385.

Boden, T.A., Kanciruk, P., Farrell, M.P., 1990: Trends '90: a compendium of data on global change. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratry, TN, 257 pp.

Bowling, S.A.,1991:Problems with the use of climatological data to detect climatic change at high latitudes. Proc. Internat. Conference on the Role of the Polar Regions in Global Change, Fairbanks AK 11-15 June 1990, pp.206-209.

Brown, T.J. and Radok, U., 1995: Detecting climate change in progress. Proceedings of the 6th International Conference on Statistical Climatology, Galway, Ireland, 19-23 June 1995, pp.13-15.

Case, R.A., 1988: North Atlantic tropical cyclones 1987. Mariners Weather Log, 32(1), 1-16

Epstein, E.S., 1982: Detecting climate change. J.Appl. Meteor.21, 1172-1182.

Flueck, J.A., and Brown, T.J., 1993: Criteria and methods for performing and evaluating solar weather studies. J.Climate 6, 373-385.

Goossens, C., and Berger, A, 1987: How to recognize an abrupt climatic change. In: W.H.Berger and L.D.Laeyrie, editors: Abrupt climatic change. Reidel.

Howell, J.F.,1995: Identifying sudden changes in data. Mon. Wea. Rev. 123(4), 1207-1212.

Kendall, M.G., and Stuart, A., 1967: The advanced theory of statistics , Vol 2: Inference and Relationships. Charles Griffin & Co., London.

Oerlemans, J., 1979: An objective approach to breaks in the weather. Mon. Wea. Rev. 106(12), 1672-1679.

Olberg,M., 1977: Zum Verteilungsgestz des Andauern von zonalen und meridionalen Zirkulationsformen der Nordhemisphäre. Zeitschr. f. Meteorologie 27, 43-48.

Press, W.H.,Flannery, B.P.,Teukolsky, S.A.,Vetterling, W.T.,1986 Numerical Recipes. Cambridge University Press, 156-157.

Radok, U., 1992: Assessing climate persistence from climate "noise". Climatic Change 20,155-167.

Radok, U., 1991,1992: Progress Reports #1, #2 on NASA Grant NAGW-2706. Unpublished.

Radok, U., and Brown, T.J.,1993, 1995: Progress Reports #3,#4 on NASA Grant NAGW-2706. Unpublished.

Radok, U., and Brown, T.J., unpublished: In search of polar warming. Paper submitted to AAAS Arctic Science Division Meeting, Fairbanks AK September 1995.

Raper, S.C.B., Wigley, T.M.L., Jones, P.D., Mayes, P.R., and Limbert, D.W.S., 1983: Recent temperature changes in the Arctic and Antarctic. Nature 306, 458-459.

Wald, A. 1947: Sequential Analysis. Wiley , 212 pp.

Appendix A: Formulae

Subscripts will be used to indicate the number of values used for parameter estimates, and bracketed symbols the numbers used to calculate the likelihoods and their ratios. Thus $L(m\,;\theta_m) = L_1$ becomes $L_m(m)$, $L(m+j\,,\theta_m) = L_3 = L_m(m+j)$, etc.

A.1 Poisson mean (= variance)

With the mean number $\bar{x}$ of occurrences the basic probability is

$$p = \frac{\bar{x}^x}{x!exp(\bar{x})} \quad , \tag{A.1.1}$$

The logarithmic likelihood functions are

$$\log_e L_m(m) = m\bar{x}_m \log_e \bar{x}_m - \sum_1^m \log_e x! - m\bar{x}_m \quad ; \tag{A.1.2a}$$

$$\log_e L_{m+j}(m) = m\bar{x}_m \log_e \bar{x}_{m+j} - \sum_1^m \log_e x! - m\bar{x}_{m+j} \quad ; \tag{A.1.2b}$$

$$\log_e L_{m+j}(m+j) = (m+j)\bar{x}_{m+j} \log_e \bar{x}_{m+j} - \sum_1^{m+j} \log_e x! - (m+j)\bar{x}_{m+j} \quad ; \tag{A.1.2c}$$

$$\log_e L_m(m+j) = (m+j)\bar{x}_{m+j} \log_e \bar{x}_m - \sum_1^{m+j} \log_e x! - (m+j)\bar{x}_m \quad , \tag{A.1.2d}$$

likelihood ratios

$$q(m) = \exp\left[m\left\{\bar{x}_m \log_e \frac{\bar{x}_m}{\bar{x}_{m+j}} + (\bar{x}_{m+j} - \bar{x}_m)\right\}\right] \quad ; \tag{A.1.3}$$

$$q(m+j) = \exp\left[(m+j)\left\{\bar{x}_{m+j} \log_e \frac{\bar{x}_{m+j}}{\bar{x}_m} - (\bar{x}_{m+j} - \bar{x}_m)\right\}\right] \quad , \tag{A.1.4}$$

so that

$$Q = \exp\left[\left\{(m+j)\bar{x}_{m+j} - m\bar{x}_m\right\}\log_e \frac{\bar{x}_{m+j}}{\bar{x}_m} - j(\bar{x}_{m+j} - \bar{x}_m)\right] . \tag{A.1.5}$$

with $y = (\bar{x}_{m+j} - \bar{x}_m)/\bar{x}_m$

$$Q = \exp\left[\bar{x}_m\left\{[(m+j)y + j]\log_e(1+y) - jy\right\}\right] . \tag{A.1.6}$$

A.2 Gaussian means and variances

Gaussian distribution.

$$p = (2\pi\sigma^2)^{-1/2} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right] . \tag{A.2.1}$$

For the population mean μ and variance σ^2 we use the sample estimates $\bar{x} = \frac{\Sigma x}{m}$ and $s^2 = \frac{1}{(n-1)}\Sigma(x-\bar{x})^2$.

$$\log_e L_m(m) = -\frac{m}{2}\log_e 2\pi - \frac{m}{2}\log_e \sigma_m^2 - \sum_1^m \frac{(x-\mu_m)^2}{2\sigma_m^2} , \tag{A.2.2}$$

with $\sum_1^m (x-\mu)^2 = m\sigma_m^2$, the last term reduces to $-m/2$.

Proceeding in the same way for $L_4 = L_{m+j}(m)$ leads to

$$\log_e L_{m+j}(m) = -\frac{m}{2}\log_e 2\pi - \frac{m}{2}\log_e \sigma_{m+j}^2 - \sum_1^m \frac{(x-\mu_{m+j})^2}{2\sigma_{m+j}^2} ; \tag{A.2.3a}$$

with $\mu_{m+j} - \mu_m = \Delta\mu$, the numerator of the last term can be written as

$$-\sum_1^m (x - (\mu_m + \Delta\mu))^2 = -\sum_1^m (x-\mu_m)^2 - \sum_1^m (-2x\Delta\mu + 2\mu_m\Delta\mu + \Delta\mu^2)$$

$$= -m\sigma_m^2 - 2m\mu_m\Delta\mu + 2m\mu_m\Delta\mu - m(\Delta\mu)^2 ,$$

so that equation (A.2.3a) becomes

$$\log_e L_{m+j}(m) = -\frac{m}{2}\log 2\pi - \frac{m}{2}\log_e \sigma_{m+j}^2 - \frac{m(\sigma_m^2 + (\Delta\mu)^2)}{2\sigma_{m+j}^2} \tag{A.2.3b}$$

subtracting equation (A.2.2b) from equation (A.2.3) gives the log likelihood ratio

$$\log_e q(m) = \log_e \frac{L_m(m)}{L_{m+j}(m)} = \frac{m}{2}\log_e \frac{\sigma_{m+j}^2}{\sigma_m} - \frac{m}{2}\left[1 - \frac{\sigma_m^2 + \Delta\mu^2}{\sigma_{m+j}^2}\right]. \tag{A.2.4}$$

augmented set of $m+j$ values proceeding in the same way for the yields first

$$\log_e L_{m+j}(m+j) = -\frac{m+j}{2}\log_e 2\pi - \frac{m+j}{2}\log_e \sigma_{m+j}^2 - \sum_1^{m+j}\frac{(x-\mu_{m+j})^2}{2\sigma_{m+j}^2}, \tag{A.2.5}$$

where the last term, with $\sum_1^{m+j}(x-\mu_{m+j})^2 = (m+j)\,\sigma_{m+j}^2$, reduces to $-(m+j)/2$. Next,

$$\log_e L_m(m+j) = -\frac{m+j}{2}\log_e 2\pi - \frac{m+j}{2}\log_e \sigma_m^2 - \sum_1^{m+j}\frac{(x-(\mu_{m+j}-\Delta\mu))^2}{2\sigma_m^2}. \tag{A.2.6a}$$

Expanding the last term as before yields

$$\log_e L_m(m+j) = -\frac{m+j}{2}\log_e 2\pi - \frac{m+j}{2}\log_e \sigma_m^2 - \frac{m+j}{2}\left[\frac{\sigma_{m+j}^2 + (\Delta\mu)^2}{\sigma_m^2}\right]. \tag{A.2.6b}$$

Subtracting equation (A.2.6b) from equation (A.2.5) we obtain the second log likelihood ratio

$$\log_e q(m+j) = \frac{m+j}{2}\log_e \frac{\sigma_m^2}{\sigma_{m+j}^2} - \frac{m+j}{2}\left[1 - \frac{\sigma_{m+j}^2 + (\Delta\mu)^2)}{\sigma_m^2}\right]. \tag{A.2.7}$$

$$\begin{aligned} Q = q(m)q(m+j) = \exp\Bigg[&\frac{m+j}{2}\log_e \frac{\sigma_m^2}{\sigma_{m+j}^2} \\ &+ \frac{m}{2}\log_e \frac{\sigma_{m+j}^2}{\sigma_m^2} - \frac{m}{2}\left[1 - \frac{\sigma_m^2 + (\Delta\mu)^2}{\sigma_{m+j}^2}\right] - \frac{m+j}{2}\left[1 - \frac{\sigma_{m+j}^2 + (\Delta\mu^2)}{\sigma_m^2}\right]\Bigg] \end{aligned} \tag{A.2.8}$$

A.3 Trends (least-square regression)

Observations made at equally spaced times $t = 1,2\cdots n$ ($n = m$ or $m+j$) are represented by

$$y = A + Bt + e \quad . \tag{A.3.1}$$

This also covers the case of exponential regression when $y = \log x$. The residuals e are assumed to be normally distributed with zero mean and variance σ^2. Sample estimates of the regression coefficients A and B satisfying least-square requirements are

$$a = \bar{y} + b\bar{t};\ b = \frac{\sum_1^n (y - \bar{y})(t - \bar{t})}{\sum_1^n (t - \bar{t})^2} \quad , \tag{A.3.2}$$

where

$$\bar{t} = \frac{(n+1)}{2}\ ;\ \sum_1^n (t - \bar{t})^2 = \frac{n(n^2-1)}{12} \quad .$$

The regression estimate $\hat{y}'$ for a given t' is then $\hat{y}' = \bar{y} + b(t' - \bar{t})$, and the corresponding residual $e_t = y'_t - \hat{y}'_t$ has a Gaussian distribution with zero mean and variance

$$s_{t_n}^2 = \frac{1}{n-2}\left[\sum_{i=1}^n (y_t - \bar{y}_t)^2 - b_n^2\,\frac{n(n^2-1)}{12}\right]\left[\frac{n+1}{n} + \frac{\left[t - \frac{(n+1)}{2}\right]^2}{\frac{n(n^2-1)}{12}}\right] \quad . \tag{A.3.3}$$

The general form of the likelihood functions defined by the n residuals e_t, $t = 1,2,\cdots n\,(= m \text{ or } m + j)$ is

$$L = \exp \sum_{t=1}^n \left[-\frac{(y_t - \hat{y}_t)^2}{2s_{t_n}^2}\right] \quad . \tag{A.3.4}$$

$$\log_e L_m = \log_e L_1 = -\left[\frac{m-2}{2\left[SS(m) - b_m^2 d(m)\right]}\right] \cdot \sum_{t=1}^m \frac{(y - \hat{y}_m)_t^2}{\left[\frac{m+1}{m} + \frac{\left[t - \frac{m+1}{2}\right]^2}{d(m)}\right]} \quad , \tag{A.3.5}$$

$$\log_e L_m(m+j) = \log_e L_3 = -\left[\frac{m+j-2}{2\left[SS(m+j) - b_{m+j}^2 d(m+j)\right]}\right] \cdot \sum_{t=1}^{m+j} \frac{(y-\hat{y}_m)_t^2}{\left[\frac{m+j+1}{m+j} + \frac{\left[t - \frac{m+j+1}{2}\right]^2}{d(m+j)}\right]} \; ; \tag{A.3.6}$$

$$\log_e L_m(m+j) = \log_e L_3 = -\left[\frac{m+j-2}{2\left[SS(m+j) - b_m^2 d(m+j)\right]}\right] \cdot \sum_{t=1}^{m+j} \frac{(y-\hat{y}_m)_t^2}{\left[\frac{m+j+1}{m+j} + \frac{\left[t - \frac{m+j+1}{2}\right]^2}{d(m+j)}\right]} \; ; \tag{A.3.7}$$

valid only with $b_m < \sqrt{\frac{SS(m+j)}{d(m+j)}}$

$$\log_e L_{m+j}(m) = \log_e L_4 = -\left[\frac{m-2}{2\left[SS(m) - b_{m+j}^2 d(m)\right]}\right] \cdot \sum_{t=1}^{m} \frac{(y-\hat{y}_{m+j})_t^2}{\left[\frac{m+1}{m} - \frac{\left[t - \frac{m+1}{2}\right]^2}{d(m)}\right]} \; ; \tag{A.3.8}$$

valid only with $b_{m+j} < \sqrt{\frac{SS(m)}{d(m)}}$,

$$\left.\begin{array}{ll} \text{where } SS(n) = \sum_{t=1}^{n}(y-\bar{y})^2 \; ; & d(n) = \frac{n(n^2-1)}{12} \; ; \\ \hat{y}_n = a_n + b_n t & n = m \text{ or } m+j \end{array}\right\} . \tag{A.3.9}$$

Finally as before

$$Q = L_1 L_2 / L_3 L_4 \tag{A.3.10}$$

A.4 Chi-square

The variances s^2 of samples from a Gaussian population with variance σ^2 define a chi-square variate

$$\chi^2 = \frac{hs^2}{\sigma^2} \ , \tag{A.4.1}$$

where h is the number of values in each sample. If these values are independent of one another, χ^2 has $h-1$ degrees of freedom (d.f.). For "coherent" (autocorrelated) series the d.f. number (which also represents the mean of the chi-square distribution as well as one half its variance) is reduced (Radok, 1992) to

$$\overline{\chi}^2 = \nu = h - \varepsilon(h) \ , \tag{A.4.2}$$

where

$$\varepsilon(h) = 1 + \frac{2(h-1)}{h} r_1 + \frac{2(h-2)}{h} r_2 + , \ldots , + \frac{2}{h} r_{h-1} \ . \tag{A.4.3}$$

Here the r_i are the autocorrelations of observations i values apart, and $h-\varepsilon(h)$ represents the number of independent observations in each section, which equals $h-1$ when all autocorrelations are zero.

The basic probability for the chi-square distribution is

$$p = \left[2^{\nu/2} \ \Gamma\left(\frac{\nu}{2}\right)\right]^{-1} (\chi^2)^{(\nu/2)-1} \ \exp\left(-\frac{\chi^2}{2}\right) \ . \tag{A.4.4}$$

Here ν is the number of degrees of freedom which equals the mean as well as half the variance of the distribution. Then the logarithmic likelihood functions take the form

$$\log_e L_m(m) = -\frac{m}{2} \nu_m \log_e 2 - m \log_e \Gamma\left(\frac{\nu_m}{2}\right) + \left(\frac{\nu_m}{2} - 1\right) \sum_1^m \log_e \chi^2 - \frac{1}{2} \sum_1^m \chi^2 \ ; \tag{A.4.5a}$$

$$\log_e L_{m+j}(m) = -\frac{m}{2} \nu_{m+j} \log_e 2 - m \log_e \Gamma\left(\frac{\nu_{m+j}}{2}\right) + \left(\frac{\nu_{m+j}}{2} - 1\right) \sum_1^m \log_e \chi^2 - \frac{1}{2} \sum_1^m \chi^2 \ ; \tag{A.4.5b}$$

$$\log_e L_{m+j}(m+j) = -\frac{m+j}{2}\nu_{m+j}\log_e 2 - (m+j)\log_e \Gamma\left[\frac{\nu_{m+j}}{2}\right] + \left[\frac{\nu_{m+j}}{2} - 1\right]\sum_1^{m+j}\log_e \chi^2 - \frac{1}{2}\sum_1^{m+j}\chi^2 ; \tag{A.4.5c}$$

$$\log_e L_m(m+j) = -\frac{m+j}{2}\nu_m\log_e 2 - (m+j)\log_e \Gamma\left[\frac{\nu_m}{2}\right] + \left[\frac{\nu_m}{2} - 1\right]\sum_1^{m+j}\log_e \chi^2 - \frac{1}{2}\sum_1^{m+j}\chi^2 . \tag{A.4.5d}$$

The likelihood ratios become

$$q(m) = 2^{\frac{m}{2}(\nu_{m+j} - \nu_m)}\left[\frac{\Gamma\left[\frac{\nu_{m+j}}{2}\right]}{\Gamma\left[\frac{\nu_m}{2}\right]}\right]^m \exp\left[\frac{1}{2}(\nu_m - \nu_{m+j})\sum_1^m \log_e \chi^2\right] , \tag{A.4.6}$$

and

$$q(m+j) = 2^{\frac{m+j}{2}(\nu_m - \nu_{m+j})}\left[\frac{\Gamma\left[\frac{\nu_m}{2}\right]}{\Gamma\left[\frac{\nu_{m+j}}{2}\right]}\right]^{m+j} \exp\left[\frac{1}{2}(\nu_{m+j} - \nu_m)\sum_1^m \log_e \chi^2\right] , \tag{A.4.7}$$

so that

$$Q = \exp\left[\frac{j}{2}(\nu_m - \nu_{m+j})\log_e 2 + j\log_e \frac{\Gamma\left[\frac{\nu_m}{2}\right]}{\Gamma\left[\frac{\nu_{m+j}}{2}\right]} + \frac{\nu_{m+j} - \nu_m}{2}\sum_{m+1}^{m+j}\log_e \chi^2\right] . \tag{A.4.8}$$

The gamma functions can be evaluated with the Euler relation

$$\Gamma(z)^{-1} = (z)e^{cz}\prod_{i=1}^{\infty}\left[1 + \frac{z}{i}\right]e^{-z/i} , \tag{A.4.9}$$

where

$$c = \lim_{i\to\infty}\left[1 + \frac{1}{2} + \frac{1}{3} +, \ldots, + \frac{1}{i}\log_e i\right] = 0.5772156649\cdots \tag{A.4.10}$$

However, a more efficient alternative procedure for determining the Gamma function has been provided by Press et al. (1986, 6.1). It has the form

$$\Gamma(z+1) = \left[z+\gamma+\frac{1}{2}\right]^{z+\frac{1}{2}} \exp\left[-z+\gamma+\frac{1}{2}\right] \sqrt{2\pi}\left[c_0 + 1 + \frac{c_1}{z+1} + \frac{c_2}{z+2} + \cdots \frac{c_N}{z+N} + \varepsilon\right] (z > 0) \tag{A.4.11}$$

The constants c are given in the Fortran routine, Appendix B.

Appendix B

Program "SEQUITOR" is designed to be an interactive program for analysis of Gaussian mean and variance, Poisson mean, chi-square (coherence), and linear (or exponential) trend changes in a sequential time series. The user typically will receive FORTRAN source code, providing an opportunity to make code changes as desired. For example, in the original code, data input is assumed to be free format. However, the user may desire to change this to a specific format. It may also be desirable for the user to add write statements that exclude headings, such that the results can then be easily imported into a graphics package.

An input/output flowchart is included in this appendix. Each square box represents an input step by the user, and an oval represents results output. A brief description of the input steps follows:

Enter input filename: This is the input data filename up to 80 characters.

Enter descriptive title: This a descriptive header of the data and/or the analysis up to 80 characters.

Enter number of values in series: This is the total number of rows in the input data file. It is assumed that the input data file contains a column of x-values (column 1) which represent an index or year for example, followed by n columns of y-values containing the actual series for analysis.

Enter missing value: This program allows for missing values. Enter a unique number (e.g., -999.) to represent missing values.

Enter column number: Input data files may contain multiple y-value columns. This entry should be from 1 to n depending upon which y-value column is desired for analysis. The very first column in the input data file is considered column 0.

Enter 0=continue, 1=reverse data input order: Often it is desirable to do the sequential monitoring analysis beginning with the most recent value and working backwards. This helps identify "regimes" in the time series. Enter either a 0 or 1.

Énter beginning and ending x-range values: Enter values separated by a comma. This range corresponds to the x-values in the very first column of the input data file. Since the x-values might represent an index or year, examples would be 10,18 or 1985,1992. Note that these values can represent a sub-set of the input data file.

Enter window size for sub-samples: In determining a regime, it is useful to examine smaller sub-sets of values. A typical sub-sample might contain 5 values. If the total number of cases in the series is not evenly divisible by the window size, the remaining values will be ignored in only the sub-sample analysis.

Enter analysis type: Here there are several options. Entering 1 through 4 places the user in the desired sequential analysis routine. Other options include changing the sub-sample size, changing

the column number, changing the data range, reversing the data order, or simply quitting the program.

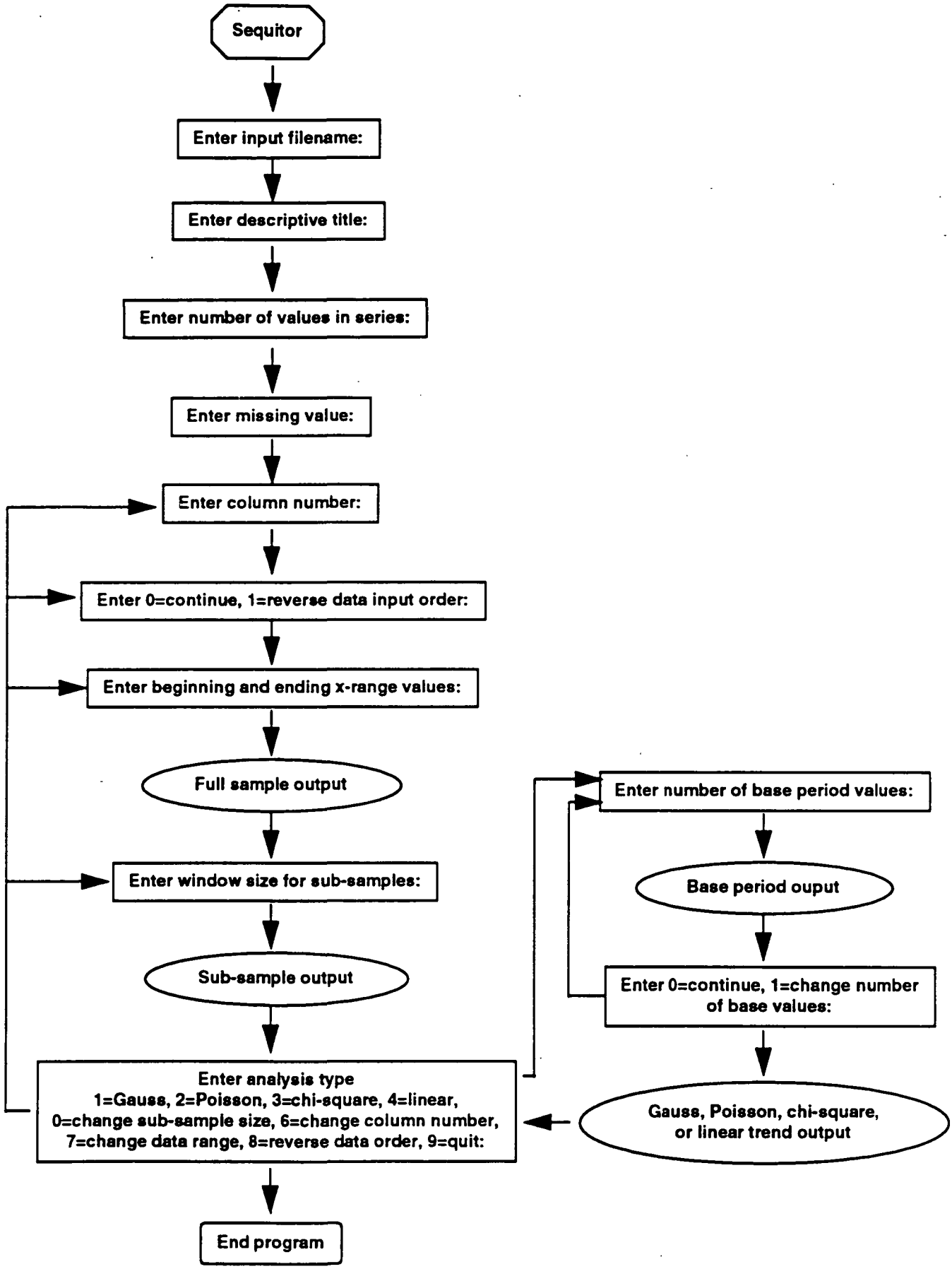
Enter number of base period values: Within each analysis routine, the user is prompted for the number of base period values. These should typically be small, say 5 to 15 or so. Base period results is ouput at this point.

Enter 0=continue, 1=change number of base values: Upon examining the base period results, the user is given the option to change the base period size, or continue with the final analysis.

This program was written interactively because it is intended to be exploratory in nature. An attempt was made to allow the user to make changes during the analysis, instead of having to restart the program several times. Results are output to the screen and to a file named "sequitor.out", which is replaced each time the program is run.

The program contains minimal comments, but variables are defined at the beginning of each subroutine to help the user understand the program; a sample analysis is included in this appendix.

Flowchart of input/output for program SEQUITOR



```
 ENTER INPUT FILE NAME:
gauss.dat
 ENTER DESCRIPTIVE TITLE:
Detection of change in Gaussian mean and variance, Test I
INVERSE SEQUENTIAL MONITORING (PROGRAM <SEQUITOR>)
 ENTER NUMBER OF VALUES IN SERIES:
15
 ENTER MISSING VALUE:
-999
 ENTER COLUMN NUMBER:
1
 ENTER 0=CONTINUE, 1=REVERSE DATA INPUT ORDER:
0
 ENTER BEGINNING AND ENDING X-RANGE VALUES:
1978,1992

FULL SAMPLE UNIVARIATE STATISTICS:
X-VALUE RANGE = 1978.-1992.
NUMBER OF VALUES =  15
NUMBER OF MISSING VALUES =   0
SAMPLE MEAN =    6.240
SAMPLE VARIANCE =   40.034
SAMPLE SLOPE =    0.466

 ENTER WINDOW SIZE FOR SUB-SAMPLES:
5
```

gauss.dat

```
1978. -5.0 -5.0 10.2
1979. -1.3 -1.3 9.2
1980. 10.2 10.2 6.8
1981. 14.9 14.9 8.8
1982. -0.6 -0.6 5.1
1983. 2.6 2.6 6.7
1984. 6.6 6.6 6.9
1985. 2.5 2.5 5.1
1986. 20.2 12.4 9.7
1987. 9.2 10.5 14.0
1988. 6.8 8.9 10.4
1989. 8.8 8.7 6.1
1990. 5.1 7.7 6.5
1991. 6.7 8.6 8.6
1992. 6.9 8.8 6.8
```

SUB-SAMPLE PARAMETERS:

INDEX RANGE	X-VALUE RANGE	NUMBER OF VALUES	MISSING VALUES	MEAN	VARIANCE	CHI-SQUARE
1- 5	1978.-1982.	5	0	3.640	71.713	7.165
6- 10	1983.-1987.	5	0	8.220	52.852	5.281
11- 15	1988.-1992.	5	0	6.860	1.723	0.172

```
 ENTER ANALYSIS TYPE
 1=GAUSS, 2=POISSON, 3=CHI-SQUARE, 4=LINEAR,
 0=CHANGE SUB-SAMPLE SIZE, 6=CHANGE COLUMN NUMBER,
 7=CHANGE DATA RANGE, 8=REVERSE DATA ORDER, 9=QUIT:
1

+-------------------------------------------------+
|  TEST FOR CHANGE IN GAUSSIAN MEAN AND VARIANCE  |
+-------------------------------------------------+

 ENTER NUMBER OF VALUES IN BASE PERIOD:
5

BASE PERIOD PARAMETERS:
X-VALUE RANGE = 1978.-1982.
NUMBER OF VALUES =   5
NUMBER OF MISSING VALUES =   0
BASE MEAN =    3.640
BASE VARIANCE =   71.713


 ENTER 0=CONTINUE, 1=CHANGE NUMBER OF BASE VALUES:
0

PROGRESSIVE PARAMETERS:
          X        Y                                  DELTA
INDEX    OBSERVATIONS      MEAN    VARIANCE   GAMMA   GAMMA
```

=====	==========	========	========	======	======
6	1983.000	2.60	3.467	57.551	0.486	
7	1984.000	6.60	3.914	49.361	0.456	0.030
8	1985.000	2.50	3.737	42.560	0.410	0.047
9	1986.000	20.20	5.567	67.353	0.453	-0.044
10	1987.000	9.20	5.930	61.189	0.422	0.031
11	1988.000	6.80	6.009	55.139	0.394	0.028
12	1989.000	8.80	6.242	50.775	0.351	0.043
13	1990.000	5.10	6.154	46.644	0.322	0.029
14	1991.000	6.70	6.193	43.078	0.280	0.042
15	1992.000	6.90	6.240	40.034	0.236	0.044

```
 ENTER 0=CONTINUE, 1=CHANGE NUMBER OF BASE VALUES:
0

 ENTER ANALYSIS TYPE
 1=GAUSS, 2=POISSON, 3=CHI-SQUARE, 4=LINEAR,
 0=CHANGE SUB-SAMPLE SIZE, 6=CHANGE COLUMN NUMBER,
 7=CHANGE DATA RANGE, 8=REVERSE DATA ORDER, 9=QUIT:
9


END OF SEQUITOR RUN
```

```
      PROGRAM SEQUITOR
C-----------------------------------------------------------------------]
C           AUTHOR: TIMOTHY J. BROWN                                    |
C                                                                       |
C  INVERSE SEQUENTIAL PROGRAM                                           |
C                                                                       |
C-----------REVISION HISTORY-----------                                 |
C LEVEL AUTHOR DATE       DESCRIPTION                                   |
C ===== ====== ========== ====================                          |
C .01A. TJB    93/04/28.  ORIGINAL VERSION.                             |
C .01B. TJB    93/05/07.  REMOVED UNUSED VARIABLE 'CFLAG' FROM OUTPUT   |
C                         IN ROUTINES SGAUSS, SCHI, AND SPOISS.         |
C .01C. TJB    93/08/26.  CHANGED ALGORITHM FOR COMPUTING CHI2;         |
C                         CHANGED ALGORITHM FOR COMPUTING GAMMA         |
C                         FUNCTION IN CHI2 CALCULATION.                 |
C                         ADDED OUTPUT OF BASE CHI-SQUARE MEAN.         |
C                                                                       |
C-----------------------------------------------------------------------]
C
C  SIX SUBROUTINES ARE ATTACHED TO THE MAIN PROGRAM:
C  'SGAUSS' COMPUTES CHANGE IN GAUSSIAN MEAN AND VARIANCE.
C  'SPOISS' COMPUTES CHANGE IN POISSION MEAN.
C  'SCHI' COMPUTES CHANGE IN CHI-SQUARE DEGREES OF FREEDOM.
C  'SLINEAR' COMPUTES CHANGE IN LINEAR TREND.
C  'UNIVAR' COMPUTES UNIVARIATE STATISTICS MEAN, VARIANCE, AND SUM.
C  'RCOEFF' COMPUTES LLS REGRESSION B0 AND B1 COEFFICIENTS.
C
C  INPUT IS ASSUMED TO BE FREE-FORMAT, BUT USER CAN CHANGE AS DESIRED.
C
C  THE PARAMETER STATEMENT AND COMMON BLOCK IS LOCATED IN ALL SUBROUTINES.
C  THE USER SHOULD CHANGE 'NDIM' AS REQUIRED.
C
C  THE FOLLOWING ARRAYS AND VARIABLES ARE USED IN THE COMMON BLOCK:
C  'FINDEX'  INDEX VALUE (1, 2,...N) ASSOCIATED WITH EACH Y-VALUE.
C  'FXVAL'   INPUT X-VALUES.
C  'FYVAL'   INPUT Y-VALUES.
C  'XDATA'   WORK ARRAY FOR X-VALUES.
C  'YDATA'   WORK ARRAY FOR Y-VALUES.
C
C  'FMISS'   NUMBER REPRESENTING MISSING VALUES.
C  'NDIM'    DIMENSION SIZE FOR DATA AND WORK ARRAYS.
C  'NCASE'   NUMBER OF FULL SAMPLE VALUES WITHIN INDEX RANGE.
C  'OUNIT'   OUTPUT UNIT NUMBER.
C
C  THE FOLLOWING ARRAYS AND VARIABLES ARE USED IN THE MAIN PROGRAM:
C  'FDATA'   HOLDS THE Y-VALUES WHEN THEY ARE INPUT; SHOULD BE
C            DIMENSIONED >= NUMBER OF COLUMNS IN INPUT FILE.
C  'FXDATA'  HOLDS THE ORIGINAL X-VALUES OR REVERSED ORDER VALUES.
C  'FYDATA'  HOLDS THE ORIGINAL Y-VALUES OR REVERSED ORDER VALUES.
C  'XWORK'   WORK ARRAY FOR X-VALUES.
C  'YWORK'   WORK ARRAY FOR Y-VALUES.
C
C  'B0'      INTERCEPT FROM LLS REGRESSION.
C  'B1'      SLOPE FROM LLS REGRESSION.
C  'CFILE'   INPUT DATA FILE NAME.
C  'CTITLE'  DESCRIPTIVE TITLE.
C  'H'       NUMBER OF VALUES WITHIN EACH SUB-SAMPLE.
C  'I'       DO LOOP COUNTER.
C  'II'      INDEX COUNTER.
C  'IBEG'    INDEX COUNTER.
```

```
C  'IEND'     INDEX COUNTER.
C  'IDIR'     DATA DIRECTION FLAG (1=REVERSE DATA ORDER, 0=CONTINUE).
C  'ITYPE'    ANALYSIS TYPE.
C  'IUNIT'    INPUT UNIT NUMBER.
C  'J'        DO LOOP COUNTER.
C  'K'        COUNTER.
C  'N'        DO LOOP COUNTER.
C  'NBEG'     BEGINNING INDEX NUMBER FOR INDEX RANGE.
C  'NEND'     ENDING INDEX RANGE FOR INDEX RANGE.
C  'NN'       COUNTER.
C  'NCOL'     COLUMN NUMBER OF Y-VALUES TO BE ANALYZED.
C             THIS IS USEFUL FOR FILES CONTAINING MULTIPLE COLUMNS OF DATA.
C             X-VALUES ARE ASSUMED TO BE IN COLUMN ONE.
C  'NMISS'    NUMBER OF MISSING VALUES.
C  'NPOP'     NUMBER OF POPULATION VALUES.
C  'NSUB'     NUMBER OF SUB-SAMPLES.
C  'NVALS'    NUMBER OF NON-MISSING VALUES.
C  'POPVAR'   POPULATION VARIANCE FROM FULL SAMPLE.
C  'XBEG'     BEGINNING VALUE OF X-RANGE.
C  'XEND'     ENDING VALUE OF X-RANGE.
C  'YBAR'     MEAN OF Y-VALUES.
C  'YSUM'     SUM OF Y-VALUES.
C  'YVAR'     VARIANCE OF Y-VALUES.
C
      PARAMETER (NDIM=10000)
      COMMON /WORK/ FXVAL(NDIM), FYVAL(NDIM), XDATA(NDIM),
     +              YDATA(NDIM), FINDEX(NDIM), NCASE, NSUB, H,
     +              OUNIT, FMISS
      REAL FDATA(15)
      REAL FXDATA(NDIM), FYDATA(NDIM), XWORK(NDIM), YWORK(NDIM)
      INTEGER H, OUNIT
      CHARACTER*80 CTITLE, CFILE

      DATA IUNIT, OUNIT / 1, 2 /
C-----------------------------------------------------------------------
C
C  THIS SECTION REQUESTS THE INPUT INFORMATION, OPENS FILES, INPUTS
C  THE DATA, AND COMPUTES FULL SAMPLE UNIVARIATE STATISTICS.
C
      WRITE(*,801)
      READ(*,101) CFILE

      OPEN(IUNIT,FILE=CFILE,STATUS='OLD')
      OPEN(OUNIT,FILE='sequitor.out')

      WRITE(*,802)
      READ(*,101) CTITLE

      WRITE(OUNIT,900)
      WRITE(*,900)
      WRITE(OUNIT,901) CTITLE

      WRITE(*,803)
      READ(*,*) NPOP
      WRITE(*,804)
      READ(*,*) FMISS

 3    CONTINUE

      WRITE(*,805)
```

```
      READ(*,*) NCOL
      REWIND IUNIT
C
C  INPUT THE DATA AND FILL WORK ARRAYS. REVERSE DATA ORDER IF REQUESTED.
C
C--------
      DO 13 I = 1, NPOP

         READ(IUNIT,*) FXDATA(I), (FDATA(J),J=1,NCOL)

         FINDEX(I) = FLOAT(I)
         FYDATA(I) = FDATA(NCOL)

 13   CONTINUE
C--------
 5    CONTINUE

      WRITE(*,806)
      READ(*,*) IDIR
C--------
      IF( IDIR .EQ. 1 ) THEN

         K = 0
C-----------
         DO 14 I = NPOP, 1, -1

            K = K + 1
            XWORK(K) = FXDATA(I)
            YWORK(K) = FYDATA(I)

 14      CONTINUE
C-----------
C-----------
         DO 15 I = 1, NPOP

            FXDATA(I) = XWORK(I)
            FYDATA(I) = YWORK(I)

 15      CONTINUE
C-----------
      END IF
C--------
C
C  FILL WORK ARRAYS WITH DATA WITHIN SELECTED INDEX RANGE AND COLUMN.
C
 4    CONTINUE

      WRITE(*,807)
      READ(*,*) XBEG, XEND
C--------
      DO 18 I = 1, NPOP

         IF( FXDATA(I) .EQ. XBEG ) NBEG = I
         IF( FXDATA(I) .EQ. XEND ) NEND = I

 18   CONTINUE
C--------
C--------
      IF( NBEG .LT. 1 ) THEN
```

```
            WRITE(*,810)
            GOTO 4

         END IF
C--------
C--------
         IF( NEND .GT. NPOP ) THEN

            WRITE(*,810)
            GOTO 4

         END IF
C--------
         NCASE = 0
C--------
         DO 16 I = 1, NPOP
C-----------
            IF( I .GE. NBEG .AND. I .LE. NEND ) THEN

               NCASE = NCASE + 1

               FXVAL(NCASE) = FXDATA(I)
               FYVAL(NCASE) = FYDATA(I)
               XDATA(NCASE) = FINDEX(I)
               YDATA(NCASE) = FYVAL(NCASE)

            END IF
C-----------
 16      CONTINUE
C--------
C
C  COMPUTE FULL SAMPLE STATISTICS AND OUTPUT RESULTS
C
         CALL UNIVAR( NCASE, NVALS, NMISS, YBAR, YVAR, YSUM )
         CALL RCOEFF( NCASE, B0, B1 )

         IRANGE = NEND - NBEG + 1

         WRITE(OUNIT,902) FXVAL(1), FXVAL(IRANGE), NVALS, NMISS, YBAR,
     +                    YVAR, B1
         WRITE(*,902) FXVAL(1), FXVAL(IRANGE), NVALS, NMISS, YBAR,
     +                YVAR, B1

 1       CONTINUE

         WRITE(*,808)
         READ(*,*) H

         WRITE(OUNIT,903)
         WRITE(*,903)
C
C  COMPUTE SUB-SAMPLE STATISTICS.
C
         POPVAR = YVAR
         NN = 0
         K = 0
         N = 0
         NSUB = 0
         IBEG = -(H) + 1
C--------
```

```
      DO 17 I = NBEG, NEND

         K = K + 1
         IBEG = IBEG + 1
         N = N + 1
C-----------
         IF( FYVAL(N) .GT. FMISS ) THEN

            NN = NN + 1
            YDATA(NN) = FYVAL(N)

         END IF
C-----------
C-----------
         IF( K .EQ. H ) THEN

            CALL UNIVAR( NN, NVALS, NMISS, YBAR, YVAR, YSUM )
C
C  OUTPUT SUB-SAMPLE STATISTICS.
C
            II = (I-H) + 1
            IEND = IBEG + H - 1

            WRITE(OUNIT,904) II, I, FXVAL(IBEG), FXVAL(IEND), NVALS,
     +                       NMISS, YBAR, YVAR
            WRITE(*,904) II, I, FXVAL(IBEG), FXVAL(IEND), NVALS, NMISS,
     +                   YBAR, YVAR

            NSUB = NSUB + 1
            NN = 0
            K = 0

         END IF
C-----------
 17    CONTINUE
C--------
C
C  BRANCH OFF TO APPROPRIATE SUBROUTINE, CHANGE SUB-SAMPLE SIZE,
C  CHANGE COLUMN NUMBER, OR STOP PROGRAM.
C
 2     CONTINUE

       WRITE(*,809)
       READ(*,*) ITYPE
C--------
       IF( ITYPE .EQ. 0 ) THEN

          GOTO 1

       ELSE IF( ITYPE .EQ. 1 ) THEN

          WRITE(OUNIT,1001)
          WRITE(*,1001)

          CALL SGAUSS

       ELSE IF( ITYPE .EQ. 2 ) THEN

          WRITE(OUNIT,1002)
          WRITE(*,1002)
```

```
         CALL SPOISS

      ELSE IF( ITYPE .EQ. 3 ) THEN

         WRITE(OUNIT,1003)
         WRITE(*,1003)

         CALL SCHI

      ELSE IF( ITYPE .EQ. 4 ) THEN

         WRITE(OUNIT,1004)
         WRITE(*,1004)

         CALL SLINEAR

      ELSE IF( ITYPE .EQ. 6 ) THEN

         GOTO 3

      ELSE IF( ITYPE .EQ. 7 ) THEN

         GOTO 4

      ELSE IF( ITYPE .EQ. 8 ) THEN

         GOTO 5

      ELSE IF( ITYPE .EQ. 9 ) THEN

         WRITE(OUNIT,907)
         WRITE(*,907)

         GOTO 999

      ELSE

         WRITE(*,905)
         GOTO 2

      END IF
C--------
      GOTO 2

 101  FORMAT(A)

 801  FORMAT(' ENTER INPUT FILE NAME:')
 802  FORMAT(' ENTER DESCRIPTIVE TITLE:')
 803  FORMAT(' ENTER NUMBER OF VALUES IN SERIES:')
 804  FORMAT(' ENTER MISSING VALUE:')
 805  FORMAT(' ENTER COLUMN NUMBER:')
 806  FORMAT(' ENTER 0=CONTINUE, 1=REVERSE DATA INPUT ORDER:')
 807  FORMAT(' ENTER BEGINNING AND ENDING X-RANGE VALUES:')
 808  FORMAT(/,' ENTER WINDOW SIZE FOR SUB-SAMPLES:')
 809  FORMAT(/' ENTER ANALYSIS TYPE',/,
     +        ' 1=GAUSS, 2=POISSON, 3=CHI-SQUARE, 4=LINEAR,',/,
     +        ' 0=CHANGE SUB-SAMPLE SIZE, 6=CHANGE COLUMN NUMBER,',/,
     +        ' 7=CHANGE DATA RANGE, 8=REVERSE DATA ORDER, 9=QUIT:')
 810  FORMAT(/'RANGE EXCEEDS TOTAL NUMBER OF CASES')
```

```
 900      FORMAT('INVERSE SEQUENTIAL MONITORING (PROGRAM <SEQUITOR>)')

 901      FORMAT(//,A)

 902      FORMAT(/,'FULL SAMPLE UNIVARIATE STATISTICS:',/,
     +           'X-VALUE RANGE = ',F5.0,'-',F5.0,/,
     +           'NUMBER OF VALUES = ',I3,/,
     +           'NUMBER OF MISSING VALUES = ',I3,/,
     +           'SAMPLE MEAN = ',F8.3,/,
     +           'SAMPLE VARIANCE = ',F8.3,/,
     +           'SAMPLE SLOPE = ',F8.3)

 903      FORMAT(/,'SUB-SAMPLE PARAMETERS:',/,
     +    ' INDEX      X-VALUE    NUMBER OF  MISSING',/,' ',
     +    'RANGE       RANGE      VALUES     VALUES     MEAN     VARIANCE',/,
     +    '=======  ===========  =========  =======  ========  ========')

 904      FORMAT(I3,'-',I3,2X,F5.0,'-',F5.0,2X,I9,2X,I7,2(2X,F8.3))

 905      FORMAT(/,'>>> NOT A VALID SELECTION <<<',/)

 907      FORMAT(//'END OF SEQUITOR RUN')

 1001     FORMAT(/,'+---------------------------------------------------+',
     +           /,'|  TEST FOR CHANGE IN GAUSSIAN MEAN AND VARIANCE   |',
     +           /,'+---------------------------------------------------+')

 1002     FORMAT(/,'+-----------------------------------+',
     +           /,'|  TEST FOR CHANGE IN POISSON MEAN  |',
     +           /,'+-----------------------------------+')

 1003     FORMAT(
     +         /,'+---------------------------------------------------+',
     +         /,'|  TEST FOR CHANGE IN CHI-SQUARE DEGREES OF FREEDOM  |',
     +         /,'+---------------------------------------------------+')

 1004     FORMAT(/,'+-----------------------------------+',
     +           /,'|  TEST FOR CHANGE IN LINEAR TREND  |',
     +           /,'+-----------------------------------+')

 999      STOP
          END
CSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS
          SUBROUTINE SGAUSS
C
C   SUBROUTINE TO COMPUTE CHANGE IN GAUSSIAN MEAN AND VARIANCE.
C
C   VARIABLES USED IN SUBROUTINE NOT DESCRIBED IN MAIN PROGRAM:
C   'BARM'     MEAN OF BASE PERIOD VALUES.
C   'BARMJ'    PROGRESSIVE MEAN.
C   'DGAMMA'   CHANGE IN GAMMA FROM PREVIOUS VALUE.
C   'GAMMA'    GAMMA VALUE.
C   'GAMOLD'   PREVIOUS VALUE OF GAMMA.
C   'J'        PROGRESSIVE VALUE INDEX.
C   'JJ'       DO LOOP COUNTER.
C   'J1'       STARTING VALUE FOR DO LOOP COMPUTING PROGRESSIVE VALUES.
C   'J2'       ENDING VALUE FOR DO LOOP COMPUTING PROGRESSIVE VALUES.
C   'M'        REAL VALUE OF NUMBER OF BASE VALUES.
C   'MBASE'    NUMBER OF BASE PERIOD VALUES.
```

```
C  'Q'        RATIO OF QMJ/QM.
C  'QM'       BASE PERIOD LIKELIHOOD RATIO.
C  'QMJ'      PROGRESSIVE PERIOD LIKELIHOOD RATIO.
C  'SDM'      STANDARD DEVIATION OF BASE PERIOD VARIANCE.
C  'SDM2'     BASE PERIOD VARIANCE.
C  'SDMJ'     PROGRESSIVE STANDARD DEVIATION.
C  'SDMJ2'    PROGRESSIVE VARIANCE.
C  'T1'       WORK VARIABLE; FIRST TERM IN EITHER Q(M) OR Q(M+J) EQUATIONS.
C  'T2'       WORK VARIABLE; SECOND TERM IN EITHER Q(M) OR Q(M+J) EQUATIONS.
C  'T3'       WORK VARIABLE; THIRD TERM IN EITHER Q(M) OR Q(M+J) EQUATIONS.
C
      PARAMETER (NDIM=10000)
      COMMON /WORK/ FXVAL(NDIM), FYVAL(NDIM), XDATA(NDIM),
     +              YDATA(NDIM), FINDEX(NDIM), NCASE, NSUB, H,
     +              OUNIT, FMISS
      REAL J, M
      INTEGER OUNIT, H
C
C  FILL WORK ARRAYS WITH BASE PERIOD VALUES, COMPUTE BASE PERIOD
C  STATISTICS, AND OUTPUT RESULTS.
C
 1    CONTINUE

      WRITE(*,801)
      READ(*,*) MBASE
      WRITE(OUNIT,1000)
      WRITE(*,1000)
C--------
      DO 20 I = 1, MBASE

         XDATA(I) = FINDEX(I)
         YDATA(I) = FYVAL(I)

         WRITE(OUNIT,1001) I, XDATA(I), YDATA(I)
         WRITE(*,1001) I, XDATA(I), YDATA(I)

 20   CONTINUE
C--------
      CALL UNIVAR( MBASE, NVALS, NMISS, YBAR, YVAR, YSUM )
C--------
      IF( NVALS .GT. 0 ) THEN

         WRITE(OUNIT,1002) FXVAL(1), FXVAL(MBASE), NVALS, NMISS, YBAR,
     +                     YVAR
         WRITE(*,1002) FXVAL(1), FXVAL(MBASE), NVALS, NMISS, YBAR,
     +                 YVAR

      ELSE

         WRITE(*,1003)
         RETURN

      END IF
C--------
      WRITE(*,802)
      READ(*,*) ITYPE
      IF( ITYPE .EQ. 1 ) GOTO 1

      WRITE(OUNIT,1004)
      WRITE(*,1004)
```

```
C
C  INITIALIZE AND COMPUTE FIXED VALUES; BEGIN PROGRESSIVE VALUE LOOP.
C
      BARM = YBAR
      SDM = SQRT(YVAR)
      SDM2 = YVAR
      M = FLOAT(MBASE)
      GAMOLD = FMISS
      J1 = MBASE + 1
      J2 = NCASE
C
C  FILL PROGRESSIVE WORK ARRAYS, COMPUTE PROGRESSIVE STATISTICS, AND
C  OUTPUT RESULTS.
C
C--------
      DO 22 JJ = J1, J2

         XDATA(JJ) = FINDEX(JJ)
         YDATA(JJ) = FYVAL(JJ)

         CALL UNIVAR( JJ, NVALS, NMISS, YBAR, YVAR, YSUM )
C
C  COMPUTE GAMMA.  SEE TEXT FOR EQUATION DETAILS.
C
C-----------
         IF( NVALS .GT. 0 ) THEN

            BARMJ = YBAR
            SDMJ = SQRT(YVAR)
            SDMJ2 = YVAR
            J = FLOAT(JJ) - M

            T1 = M * ALOG(SDM / SDMJ )
            T2 = ((M - 1.) / 2.) * (1. - (SDM2 / SDMJ2))
            T3 = (M / (2. * SDM2)) * (BARMJ - BARM)**2
            QM = EXP(T1 + T2 - T3)

            T1 = (M + J)* ALOG(SDM / SDMJ )
            T2 = ((M + J - 1.) / 2.) * ((SDMJ2 / SDM2) - 1.)
            T3 = ((M + J) / (2. * SDM2)) * (BARMJ - BARM)**2
            QMJ = EXP(T1 + T2 + T3)

            Q = QMJ / QM
            GAMMA = 1. / (1. + SQRT(Q))
            DGAMMA = GAMOLD - GAMMA
C--------------
            IF( GAMOLD .EQ. FMISS ) THEN

               WRITE(OUNIT,1005) JJ, FXVAL(JJ), YDATA(JJ), YBAR, YVAR,
     +                           GAMMA
               WRITE(*,1005) JJ, FXVAL(JJ), YDATA(JJ), YBAR, YVAR,
     +                       GAMMA

            ELSE

               WRITE(OUNIT,1005) JJ, FXVAL(JJ), YDATA(JJ), YBAR, YVAR,
     +                           GAMMA, DGAMMA
               WRITE(*,1005) JJ, FXVAL(JJ), YDATA(JJ), YBAR, YVAR,
     +                       GAMMA, DGAMMA
```

```
            END IF
C-------------
            GAMOLD = GAMMA

         ELSE

            WRITE(OUNIT,1006) JJ, FXVAL(JJ), YDATA(JJ)
            WRITE(*,1006) JJ, FXVAL(JJ), YDATA(JJ)
            GAMOLD = FMISS

         END IF
C----------
 22     CONTINUE
C--------
        WRITE(*,802)
        READ(*,*) ITYPE
        IF( ITYPE .EQ. 1 ) GOTO 1

 801    FORMAT(/,' ENTER NUMBER OF VALUES IN BASE PERIOD:')
 802    FORMAT(/,' ENTER 0=CONTINUE, 1=CHANGE NUMBER OF BASE VALUES:')

 1000   FORMAT(/'BASE PERIOD INPUT VALUES:'
     +         /'INDEX      X          Y'
     +         /'=====  ========  ========')

 1001   FORMAT(I5,2X,F8.3,2X,F8.2)

 1002   FORMAT(/,'BASE PERIOD PARAMETERS:',/,
     +         'X-VALUE RANGE = ',F5.0,'-',F5.0,/,
     +         'NUMBER OF VALUES = ',I3,/,
     +         'NUMBER OF MISSING VALUES = ',I3,/,
     +         'BASE MEAN = ',F8.3,/,
     +         'BASE VARIANCE = ',F8.3,/)

 1003   FORMAT(' ALL VALUES IN BASE PERIOD MISSING')

 1004   FORMAT(/,'PROGRESSIVE PARAMETERS:',/,
     +  '            X         Y                                  DELTA',/,
     +  'INDEX    OBSERVATIONS      MEAN    VARIANCE   GAMMA   GAMMA',/,
     +  '=====  ================  ========  ========  ======  ======')

 1005   FORMAT(I5,2X,F8.3,1X,F7.2,2(2X,F8.3),2X,F6.3,2X,F6.3)

 1006   FORMAT(I5,2X,F8.3,1X,F7.2)

        RETURN
        END
C-----------------------------------------------------------------------
        SUBROUTINE SPOISS
C
C  SUBROUTINE TO COMPUTE CHANGE IN POISSON MEAN.
C
C  VARIABLES USED IN SUBROUTINE NOT DESCRIBED IN MAIN PROGRAM:
C  'BARM'    MEAN OF BASE PERIOD VALUES.
C  'BARMJ'   PROGRESSIVE MEAN.
C  'DGAMMA'  CHANGE IN GAMMA FROM PREVIOUS VALUE.
C  'GAMMA'   GAMMA VALUE.
C  'GAMOLD'  PREVIOUS VALUE OF GAMMA.
C  'J'       PROGRESSIVE VALUE INDEX.
C  'JJ'      DO LOOP COUNTER.
```

```
C  'J1'       STARTING VALUE FOR DO LOOP COMPUTING PROGRESSIVE VALUES.
C  'J2'       ENDING VALUE FOR DO LOOP COMPUTING PROGRESSIVE VALUES.
C  'M'        REAL VALUE OF NUMBER OF BASE VALUES.
C  'MBASE'    NUMBER OF BASE PERIOD VALUES.
C  'Q'        COMBINED LIKELIHOOD RATIOS.
C  'T1'       WORK VARIABLE; FIRST TERM IN EITHER Q EQUATION.
C  'T2'       WORK VARIABLE; SECOND TERM IN EITHER Q EQUATION.
C  'T3'       WORK VARIABLE; THIRD TERM IN EITHER Q EQUATION.
C
      PARAMETER (NDIM=10000)
      COMMON /WORK/ FXVAL(NDIM), FYVAL(NDIM), XDATA(NDIM),
     +              YDATA(NDIM), FINDEX(NDIM), NCASE, NSUB, H,
     +              OUNIT, FMISS
      REAL J, M
      INTEGER OUNIT, H
C
C  FILL WORK ARRAYS WITH BASE PERIOD VALUES, COMPUTE BASE PERIOD
C  STATISTICS, AND OUTPUT RESULTS.
C
 1    CONTINUE

      WRITE(*,801)
      READ(*,*) MBASE
      WRITE(OUNIT,2000)
      WRITE(*,2000)
C--------
      DO 20 I = 1, MBASE

         XDATA(I) = FINDEX(I)
         YDATA(I) = FYVAL(I)

         WRITE(OUNIT,2001) I, XDATA(I), YDATA(I)
         WRITE(*,2001) I, XDATA(I), YDATA(I)

 20   CONTINUE
C--------
      CALL UNIVAR( MBASE, NVALS, NMISS, YBAR, YVAR, YSUM )
C--------
      IF( NVALS .GT. 0 ) THEN

         WRITE(OUNIT,2002) FXVAL(1), FXVAL(MBASE), NVALS, NMISS,
     +                     YBAR, YVAR, YSUM
         WRITE(*,2002) FXVAL(1), FXVAL(MBASE), NVALS, NMISS, YBAR,
     +                 YVAR, YSUM

      ELSE

         WRITE(*,2003)
         RETURN

      END IF
C--------
      WRITE(*,802)
      READ(*,*) ITYPE
      IF( ITYPE .EQ. 1 ) GOTO 1

      WRITE(OUNIT,2004)
      WRITE(*,2004)
C
C  INITIALIZE AND COMPUTE FIXED VALUES; BEGIN PROGRESSIVE VALUE LOOP.
```

```
C
      BARM = YBAR
      M = FLOAT(MBASE)
      GAMOLD = FMISS
      J1 = MBASE + 1
      J2 = NCASE
C
C  FILL PROGRESSIVE WORK ARRAYS, COMPUTE PROGRESSIVE STATISTICS, AND
C  OUTPUT RESULTS.
C
C--------
      DO 22 JJ = J1, J2

         XDATA(JJ) = FINDEX(JJ)
         YDATA(JJ) = FYVAL(JJ)

         CALL UNIVAR( JJ, NVALS, NMISS, YBAR, YVAR, YSUM )
C
C  COMPUTE GAMMA.  SEE TEXT FOR EQUATION DETAILS.
C
C-----------
         IF( NVALS .GT. 0 ) THEN

            BARMJ = YBAR
            J = FLOAT(JJ) - M

            T1 = (BARMJ * (M + J)) - (BARM * M)
            T2 = ALOG(BARMJ / BARM)
            T3 = (BARMJ - BARM) * J
            Q = EXP(T1 * T2 - T3)
            GAMMA = 1. / (1. + SQRT(Q))
            DGAMMA = GAMOLD - GAMMA
C--------------
            IF( GAMOLD .EQ. FMISS ) THEN

               WRITE(OUNIT,2005) JJ, FXVAL(JJ), YDATA(JJ), BARMJ, YVAR,
     +                           YSUM, GAMMA
               WRITE(*,2005) JJ, FXVAL(JJ), YDATA(JJ), BARMJ, YVAR,
     +                       YSUM, GAMMA

            ELSE

               WRITE(OUNIT,2005) JJ, FXVAL(JJ), YDATA(JJ), BARMJ, YVAR,
     +                           YSUM, GAMMA, DGAMMA
               WRITE(*,2005) JJ, FXVAL(JJ), YDATA(JJ), BARMJ, YVAR,
     +                       YSUM, GAMMA, DGAMMA

            END IF
C--------------
            GAMOLD = GAMMA

         ELSE

            WRITE(OUNIT,2006) JJ, FXVAL(JJ), YDATA(JJ)
            WRITE(*,2006) JJ, FXVAL(JJ), YDATA(JJ)
            GAMOLD = FMISS

         END IF
C-----------
 22   CONTINUE
```

```
C--------
      WRITE(*,802)
      READ(*,*) ITYPE
      IF( ITYPE .EQ. 1 ) GOTO 1

 801  FORMAT(/,' ENTER NUMBER OF VALUES IN BASE PERIOD:')
 802  FORMAT(/,' ENTER 0=CONTINUE, 1=CHANGE NUMBER OF BASE VALUES:')

 2000 FORMAT(/'BASE PERIOD INPUT VALUES:'
     +        /'INDEX      X          Y'
     +        /'=====  ========  ========')

 2001 FORMAT(I5,2X,F8.3,2X,F8.2)

 2002 FORMAT(/,'BASE PERIOD PARAMETERS:',/,
     +        'X-VALUE RANGE = ',F5.0,'-',F5.0,/,
     +        'NUMBER OF VALUES = ',I3,/,
     +        'NUMBER OF MISSING VALUES = ',I3,/,
     +        'BASE MEAN = ',F8.3,/,
     +        'BASE VARIANCE = ',F8.3,/,
     +        'BASE SUM = ',F6.0,/)

 2003 FORMAT(' ALL VALUES IN BASE PERIOD MISSING')

 2004 FORMAT(/,'PROGRESSIVE PARAMETERS:',/,
     + '            X        Y                                    ',
     + '   DELTA',/,
     + 'INDEX    OBSERVATIONS      MEAN    VARIANCE    SUM    GAMMA',
     + '   GAMMA',/,
     + '=====  ================  ========  ========  ======  ======',
     + '  ======')

 2005 FORMAT(I5,2X,F8.3,1X,F7.0,2(2X,F8.3),2X,f6.0,2X,F6.3,2X,F6.3)

 2006 FORMAT(I5,2X,F8.3,1X,F7.0)

      RETURN
      END
C-----------------------------------------------------------------------
      SUBROUTINE SCHI
C
C  SUBROUTINE TO COMPUTE CHANGE IN CHI-SQUARE DEGREES OF FREEDOM.
C
C  VARIABLES USED IN SUBROUTINE NOT DESCRIBED IN MAIN PROGRAM:
C  'DGAMMA'  CHANGE IN GAMMA FROM PREVIOUS VALUE.
C  'FCHI2'   CHI-SQUARE VALUE FOR EACH SUB-SAMPLE.
C  'GAMMA'   GAMMA VALUE.
C  'GAMNUM'  VALUE OF THE GAMMA FUNCTION FOR THE BASE PERIOD.
C  'GAMNUMJ' VALUE OF THE GAMMA FUNCTION FOR THE PROGRESSIVE PERIOD.
C  'GAMOLD'  PREVIOUS VALUE OF GAMMA.
C  'GNUM'    BASE PERIOD MEAN FOR GAMMA FUNCTION.
C  'GNUMJ'   PROGRESSIVE MEAN FOR GAMMA FUNCTION.
C  'H'       NUMBER OF SUB-SAMPLES.
C  'J'       PROGRESSIVE VALUE INDEX.
C  'JJ'      DO LOOP COUNTER.
C  'J1'      STARTING VALUE FOR DO LOOP COMPUTING PROGRESSIVE VALUES.
C  'J2'      ENDING VALUE FOR DO LOOP COMPUTING PROGRESSIVE VALUES.
C  'K'       INDEX COUNTER.
C  'KK'      INDEX COUNTER.
C  'M'       REAL VALUE OF NUMBER OF BASE VALUES.
```

```
C  'MBASE'   NUMBER OF BASE PERIOD VALUES.
C  'NBASE'   ACTUAL NUMBER OF CASES IN BASE PERIOD.
C  'NUM'     BASE PERIOD MEAN.
C  'NUMJ'    PROGRESSIVE MEAN.
C  'Q'       COMBINED LIKELIHOOD RATIOS.
C  'SUMCHI2' PROGRESSIVE SUM OF CHI-SQUARE VALUES.
C  'T1'      WORK VARIABLE; FIRST TERM IN EITHER Q EQUATION.
C  'T2'      WORK VARIABLE; SECOND TERM IN EITHER Q EQUATION.
C  'T3'      WORK VARIABLE; THIRD TERM IN EITHER Q EQUATION.
C  'YVARC'   VARIANCE FOR EACH CHI2 VALUE (SUB-SAMPLE).
C  'YVARM'   VARIANCE FOR BASE PERIOD.
C

      PARAMETER (NDIM=10000)
      COMMON /WORK/ FXVAL(NDIM), FYVAL(NDIM), XDATA(NDIM),
     +              YDATA(NDIM), FINDEX(NDIM), NCASE, NSUB, H,
     +              OUNIT, FMISS
      REAL J, M, NUM, NUMJ, FCHI2(NDIM)
      INTEGER OUNIT, H
C
C  FILL WORK ARRAYS WITH BASE PERIOD VALUES, COMPUTE BASE PERIOD
C  STATISTICS, AND OUTPUT RESULTS.
C
 1    CONTINUE

      WRITE(*,801)
      READ(*,*) MBASE
      WRITE(OUNIT,3000)
      WRITE(*,3000)
C--------
      DO 20 I = 1, MBASE

         XDATA(I) = FINDEX(I)
         YDATA(I) = FYVAL(I)

         WRITE(OUNIT,3001) I, XDATA(I), YDATA(I)
         WRITE(*,3001) I, XDATA(I), YDATA(I)

 20   CONTINUE
C--------

      NBASE = MBASE * H
C--------
      DO 21 I = 1, NBASE

         YDATA(I) = FYVAL(I)

 21   CONTINUE
C--------
      CALL UNIVAR( NBASE, NVALS, NMISS, YBAR, YVAR, YSUM )
C--------
      IF( NVALS .GT. 0 ) THEN

         WRITE(OUNIT,3002) FXVAL(1), FXVAL(NBASE), NVALS, NMISS,
     +                     YBAR, YVAR
         WRITE(*,3002) FXVAL(1), FXVAL(NBASE), NVALS, NMISS, YBAR,
     +                 YVAR

         YVARM = YVAR
```

```
      ELSE

         WRITE(*,3003)
         RETURN

      END IF
C--------
C
C  COMPUTE CHI2 VALUES.
C
      KK = 0
C--------
      DO 22 JJ = 1, NSUB

         K = 0
         XDATA(JJ) = FLOAT(JJ)
C-----------
         DO 24 I = 1, H

            K = K + 1
            KK = KK + 1
            YDATA(K) = FYVAL(KK)

 24      CONTINUE
C-----------
         CALL UNIVAR( H, NVALS, NMISS, YBAR, YVAR, YSUM )

         YVARC = YVAR
C-----------
         IF( JJ .GT. MBASE ) THEN

            KK = JJ * H
C--------------
            DO 26 I = 1, KK

               YDATA(I) = FYVAL(I)

 26         CONTINUE
C--------------
            CALL UNIVAR( KK, NVALS, NMISS, YBAR, YVAR, YSUM )

            FCHI2(JJ) = (H - 1.) * YVARC / YVAR

         ELSE

            FCHI2(JJ) = (H - 1.) * YVARC / YVARM

         END IF
C-----------
 22   CONTINUE
C--------
C--------
      DO 28 I = 1, MBASE

         YDATA(I) = FCHI2(I)

 28   CONTINUE
C--------
      CALL UNIVAR( MBASE, NVALS, NMISS, YBAR, YVAR, YSUM )
C
```

```
C  COMPUTE GAMMA FUNCTION FOR THE BASE PERIOD.
C
        NUM = YBAR

        WRITE(OUNIT,3004) NUM
        WRITE(*,3004) NUM
        WRITE(*,802)
        READ(*,*) ITYPE
        IF( ITYPE .EQ. 1 ) GOTO 1

        WRITE(OUNIT,3005)
        WRITE(*,3005)

        GNUM = YBAR / 2.
        CALL GAMMLN( GNUM, GAMNUM )
C
C  INITIALIZE AND COMPUTE FIXED VALUES; BEGIN PROGRESSIVE VALUE LOOP.
C
        J1 = MBASE + 1
        J2 = NSUB
        M = FLOAT(MBASE)
        GAMOLD = FMISS
        SUMCHI2 = 0.
C
C  FILL PROGRESSIVE WORK ARRAYS, COMPUTE PROGRESSIVE STATISTICS, AND
C  OUTPUT RESULTS.
C
C--------
        DO 30 JJ = J1, J2

           SUMCHI2 = SUMCHI2 + ALOG(FCHI2(JJ))
C-----------
           DO 32 I = 1, JJ

              YDATA(I) = FCHI2(I)

 32        CONTINUE
C-----------
           CALL UNIVAR( JJ, NVALS, NMISS, YBAR, YVAR, YSUM )
C-----------
           IF( NVALS .GT. 0 ) THEN
C
C  COMPUTE GAMMA FUNCTION FOR PROGRESSIVE VALUES.
C
              NUMJ = YBAR
              GNUMJ = YBAR / 2.
              CALL GAMMLN( GNUMJ, GAMNUMJ )
              J = FLOAT(JJ) - M
C
C  COMPUTE GAMMA.  SEE TEXT FOR EQUATION DETAILS.
C
              T1 = (J * (NUM - NUMJ) * ALOG(2.)) / 2.
              T2 = J * ALOG(GAMNUM / GAMNUMJ)
              T3 = ((NUMJ - NUM) / 2.) * SUMCHI2
              Q = EXP(T1 + T2 + T3)

              GAMMA = 1. / (1. + SQRT(Q))
              DGAMMA = GAMOLD - GAMMA
              KK = JJ * H
C--------------
```

```
            DO 34 I = 1, KK

               YDATA(I) = FYVAL(I)

 34         CONTINUE
C-------------
            CALL UNIVAR( KK, NVALS, NMISS, YBAR, YVAR, YSUM )
C-------------
            IF( GAMOLD .EQ. FMISS ) THEN

               WRITE(OUNIT,3006) JJ, FCHI2(JJ), NUMJ, YVAR, GAMMA
               WRITE(*,3006) JJ, FCHI2(JJ), NUMJ, YVAR, GAMMA

            ELSE

               WRITE(OUNIT,3006) JJ, FCHI2(JJ), NUMJ, YVAR, GAMMA,
     +                           DGAMMA
               WRITE(*,3006) JJ, FCHI2(JJ), NUMJ, YVAR, GAMMA, DGAMMA

            END IF
C-------------
            GAMOLD = GAMMA

         ELSE

            WRITE(OUNIT,3007) JJ, YDATA(JJ)
            WRITE(*,3007) JJ, YDATA(JJ)
            GAMOLD = FMISS

         END IF
C----------
 30     CONTINUE
C--------
        WRITE(*,802)
        READ(*,*) ITYPE
        IF( ITYPE .EQ. 1 ) GOTO 1

 801    FORMAT(/,' ENTER NUMBER OF VALUES IN BASE PERIOD:')
 802    FORMAT(/,' ENTER 0=CONTINUE, 1=CHANGE NUMBER OF BASE VALUES:')

 3000   FORMAT(/'BASE PERIOD INPUT VALUES:'
     +          /'INDEX      X          Y'
     +          /'=====  ========  ========')

 3001   FORMAT(I5,2X,F8.3,2X,F8.2)

 3002   FORMAT(/,'BASE PERIOD PARAMETERS:',/,
     +          'X-VALUE RANGE = ',F5.0,'-',F5.0,/,
     +          'NUMBER OF VALUES = ',I3,/,
     +          'NUMBER OF MISSING VALUES = ',I3,/,
     +          'BASE MEAN = ',F8.3,/,
     +          'BASE VARIANCE = ',F8.3)

 3003   FORMAT(' ALL VALUES IN BASE PERIOD MISSING')

 3004   FORMAT('BASE CHI-SQUARE = ',F8.3,/)

 3005   FORMAT(/,'PROGRESSIVE PARAMETERS:',/,
     +  '         CHI-SQUARE                                DELTA',/,
     +  'INDEX  OBSERVATIONS    MEAN    VARIANCE    GAMMA   GAMMA',/,
```

```
     +  '=====  ============  ========  ========  ======  ======')

 3006   FORMAT(I5,2X,F12.3,2(2X,F8.3),2(2X,F6.3))

 3007   FORMAT(I5,2X,F12.3)

        RETURN
        END
C-----------------------------------------------------------------------
        SUBROUTINE SLINEAR
C
C  SUBROUTINE TO COMPUTE CHANGE IN LINEAR TREND.
C
C  VARIABLES USED IN SUBROUTINE NOT DESCRIBED IN MAIN PROGRAM:
C  'AM'       BASE PERIOD INTERCEPT FROM LLS REGRESSION.
C  'BM'       BASE PERIOD SLOPE FROM LLS REGRESSION.
C  'DM'       CONSTANT USED IN Q(M) LIKELIHOOD RATIO.
C  'DMJ'      CONSTANT USED IN Q(M+J) LIKELIHOOD RATIO.
C  'DENOM'    DENOMINATOR IN 'T21' AND 'T22' TERMS.
C  'DGAMMA'   CHANGE IN GAMMA FROM PREVIOUS VALUE.
C  'GAMMA'    GAMMA VALUE.
C  'GAMOLD'   PREVIOUS VALUE OF GAMMA.
C  'I'        DO LOOP COUNTER.
C  'J'        PROGRESSIVE VALUE INDEX.
C  'JJ'       DO LOOP COUNTER.
C  'J1'       STARTING VALUE FOR DO LOOP COMPUTING PROGRESSIVE VALUES.
C  'J2'       ENDING VALUE FOR DO LOOP COMPUTING PROGRESSIVE VALUES.
C  'LNL1'     NATURAL LOG LIKELIHOOD 1.
C  'LNL2'     NATURAL LOG LIKELIHOOD 2.
C  'LNL3'     NATURAL LOG LIKELIHOOD 3.
C  'LNL4'     NATURAL LOG LIKELIHOOD 4.
C  'M'        REAL VALUE OF NUMBER OF BASE VALUES.
C  'MBASE'    NUMBER OF BASE PERIOD VALUES.
C  'PREDM'    PREDICTED VALUES FROM REGRESSION EQUATION USING 'M' VALUES.
C  'PREDMJ'   PREDICTED VALUES FROM REGRESSION EQUATION USING 'M+J' VALUES.
C  'Q'        RATIO OF QMJ/QM.
C  'QM'       BASE PERIOD LIKELIHOOD RATIO.
C  'QMJ'      PROGRESSIVE PERIOD LIKELIHOOD RATIO.
C  'RESIDM'   ARRAY OF BASE PERIOD RESIDUALS FROM LLS REGRESSION.
C  'RESIDMJ'  ARRAY OF PROGRESSIVE VALUE RESIDUALS FROM LLS REGRESSION.
C  'SSM'      SUM OF SQUARES FOR BASE PERIOD.
C  'SSMJ'     SUM OF SQUARES FOR 'M+J' VALUES.
C  'SUMT2'    SUM OF 'T2' TERM IN 'LNL1' AND 'LNL4' FORMULAE.
C  'SUMT21'   SUM OF 'T2' TERM IN 'LNL2' FORMULAE.
C  'SUMT22'   SUM OF 'T2' TERM IN 'LNL3' FORMULAE.
C  'T'        REAL VALUE OF BASE PERIOD LOOP INDEX.
C  'T1'       WORK VARIABLE FOR FIRST TERM IN 'LNL1' AND 'LNL4' FORMULAE.
C  'T2'       WORK VARIABLE FOR SECOND TERM IN 'LNL1' AND 'LNL4' FORMULAE.
C  'T11'      WORK VARIABLE FOR FIRST TERM IN 'LNL2' FORMULAE.
C  'T12'      WORK VARIABLE FOR FIRST TERM IN 'LNL3' FORMULAE.
C  'T21'      WORK VARIABLE FOR SECOND TERM IN 'LNL2' FORMULAE.
C  'T22'      WORK VARIABLE FOR SECOND TERM IN 'LNL3' FORMULAE.
C
        PARAMETER (NDIM=10000)
        COMMON /WORK/ FXVAL(NDIM), FYVAL(NDIM), XDATA(NDIM),
     +                YDATA(NDIM), FINDEX(NDIM), NCASE, NSUB, H,
     +                OUNIT, FMISS
        REAL RESIDM(NDIM), RESIDMJ(NDIM), PREDM(NDIM), PREDMJ(NDIM)
        REAL M, J, LNL1, LNL2, LNL3, LNL4
        INTEGER OUNIT, H
```

```
C
C  FILL WORK ARRAYS WITH BASE PERIOD VALUES, COMPUTE BASE PERIOD
C  STATISTICS, AND OUTPUT RESULTS.
C
 1      CONTINUE

        WRITE(*,801)
        READ(*,*) MBASE
        WRITE(OUNIT,4000)
        WRITE(*,4000)
C--------
        DO 20 I = 1, MBASE

           WRITE(OUNIT,4001) I, FXVAL(I), FYVAL(I)
           WRITE(*,4001) I, FXVAL(I), FYVAL(I)

 20     CONTINUE
C--------
C--------
        DO 21 I = 1, NCASE

           XDATA(I) = FINDEX(I)
           YDATA(I) = FYVAL(I)

 21     CONTINUE
C--------
        CALL UNIVAR( MBASE, NVALS, NMISS, YBAR, YVAR, YSUM )
C--------
        IF( NVALS .GT. 0 ) THEN

           CALL RCOEFF( MBASE, AM, BM )

           SDM = SQRT(YVAR*(12./(FLOAT(MBASE)*(FLOAT(MBASE**2)-1.))))
           SDM2 = SDM**2
           BARM = BM

           WRITE(OUNIT,4002) FXVAL(1), FXVAL(MBASE), NVALS, NMISS,
     +                       YBAR, YVAR, BM
           WRITE(*,4002) FXVAL(1), FXVAL(MBASE), NVALS, NMISS, YBAR,
     +                   YVAR, BM

        ELSE

           WRITE(*,4003)
           RETURN

        END IF
C--------
        WRITE(*,802)
        READ(*,*) ITYPE
        IF( ITYPE .EQ. 1 ) GOTO 1

        WRITE(OUNIT,4004)
        WRITE(*,4004)

        M = FLOAT(MBASE)
        DM = (M * (M**2 - 1.)) / 12.
        SSM = 0.
C--------
        DO 22 I = 1, MBASE
```

```
C-----------
            IF( YDATA(I) .NE. FMISS ) THEN

               SSM = SSM + (YDATA(I) - YBAR)**2

            END IF
C-----------
 22     CONTINUE
C--------
C--------
        DO 23 I = 1, NCASE
C-----------
            IF( YDATA(I) .NE. FMISS ) THEN

               PREDM(I) = (XDATA(I) * BM + AM)
               RESIDM(I) = (YDATA(I) - PREDM(I))

            ELSE

               PREDM(I) = FMISS
               RESIDM(I) = FMISS

            END IF
C-----------
 23     CONTINUE
C--------
        SUMT2 = 0.
        T1 = (-1.) * (M - 2.) / (2. * (SSM - (BM**2 * DM)))
C--------
        DO 24 I = 1, MBASE
C-----------
            IF( YDATA(I) .NE. FMISS ) THEN

               T = FLOAT(I)

               T2 = RESIDM(I)**2 / (((M + 1.) / M)
     +              + (T - ((M + 1.) / 2.))**2 / DM)

               SUMT2 = SUMT2 + T2

            END IF
C-----------
 24     CONTINUE

        LNL1 = T1 * SUMT2
C--------
C
C  INITIALIZE AND COMPUTE FIXED VALUES; BEGIN PROGRESSIVE VALUE LOOP.
C
        GAMOLD = FMISS
        J1 = MBASE + 1
        J2 = NCASE
C
C  FILL PROGRESSIVE WORK ARRAYS, COMPUTE PROGRESSIVE STATISTICS, AND
C  OUTPUT RESULTS.
C
C--------
        DO 26 JJ = J1, J2

            XDATA(JJ) = FINDEX(JJ)
```

```
      YDATA(JJ) = FYVAL(JJ)

      CALL UNIVAR( JJ, NVALS, NMISS, YBAR, YVAR, YSUM )

      IF( YDATA(JJ) .EQ. FMISS ) NVALS = 0
C
C  COMPUTE EQUATION TERMS AND GAMMA.  SEE TEXT FOR EQUATION DETAILS.
C
C-----------
      IF( NVALS .GT. 0 ) THEN

         CALL RCOEFF( JJ, AMJ, BMJ )

         SDMJ = SQRT(YVAR*(12./(FLOAT(JJ)*(FLOAT(JJ**2)-1.))))
         SDMJ2 = SDMJ**2
         BARMJ = BMJ

         SSMJ = 0.
         J = FLOAT(JJ) - M
         DMJ = ((M + J) * ((M + J)**2 - 1.)) / 12.
C--------------
         DO 30 I = 1, JJ

            SSMJ = SSMJ + (YDATA(I) - YBAR)**2
            PREDMJ(I) = (XDATA(I) * BMJ + AMJ)
            RESIDMJ(I) = (YDATA(I) - PREDMJ(I))

 30      CONTINUE
C--------------
         SUMT2 = 0.
         T1 = (-1.) * (M - 2.) / (2. * (SSM - (BMJ**2 * DM)))
C--------------
         DO 32 I = 1, MBASE

            T = FLOAT(I)

            T2 = RESIDMJ(I)**2 / (((M + 1) / M) +
     +           (T - ((M + 1.) / 2.))**2 / DM)

            SUMT2 = SUMT2 + T2

 32      CONTINUE
C--------------
         LNL4 = T1 * SUMT2

         SUMT21 = 0.
         SUMT22 = 0.

         T11 = (-1.) * (M + J - 2.)
     +         / (2. * (SSMJ - (BMJ**2 * DMJ)))
         T12 = (-1.) * (M + J - 2.)
     +         / (2. * (SSMJ - (BM**2 * DMJ)))
C--------------
         DO 34 I = 1, JJ

            T = FLOAT(I)

            DENOM = ((M + J + 1.) / (M + J))
     +              + ((T - ((M + J + 1.) / 2.))**2 / DMJ)
```

```
                T21 = RESIDMJ(I)**2 / DENOM
                T22 = RESIDM(I)**2 / DENOM

                SUMT21 = SUMT21 + T21
                SUMT22 = SUMT22 + T22

 34           CONTINUE
C-------------
              LNL2 = T11 * SUMT21
              LNL3 = T12 * SUMT22

              QM = EXP(LNL4 - LNL1)
              QMJ = EXP(LNL2 - LNL3)
              Q = QMJ / QM
              GAMMA = 1. / (1. + SQRT(Q))
              DGAMMA = GAMOLD - GAMMA
CNNNNNNNNNNNN
              T1 = M * ALOG(SDM / SDMJ )
              T2 = ((M - 1.) / 2.) * (1. - (SDM2 / SDMJ2))
              T3 = (M / (2. * SDM2)) * (BARMJ - BARM)**2
              QM = EXP(T1 + T2 - T3)

              T1 = (M + J)* ALOG(SDM / SDMJ )
              T2 = ((M + J - 1.) / 2.) * ((SDMJ2 / SDM2) - 1.)
              T3 = ((M + J) / (2. * SDM2)) * (BARMJ - BARM)**2
              QMJ = EXP(T1 + T2 + T3)

              Q = QMJ / QM
C             GAMMA = 1. / (1. + SQRT(Q))
              DGAMMA = GAMOLD - GAMMA
CNNNNNNNNNNNN
C-------------
              IF( GAMOLD .EQ. FMISS ) THEN

                 WRITE(OUNIT,4005) JJ, FXVAL(JJ), YDATA(JJ), YBAR, YVAR,
     +                             BMJ, GAMMA
                 WRITE(*,4005) JJ, FXVAL(JJ), YDATA(JJ), YBAR, YVAR,
     +                         BMJ, GAMMA

              ELSE

                 WRITE(OUNIT,4005) JJ, FXVAL(JJ), YDATA(JJ), YBAR, YVAR,
     +                             BMJ, GAMMA, DGAMMA
                 WRITE(*,4005) JJ, FXVAL(JJ), YDATA(JJ), YBAR, YVAR,
     +                         BMJ, GAMMA, DGAMMA

              END IF
C-------------
              GAMOLD = GAMMA

           ELSE

              WRITE(OUNIT,4006) JJ, FXVAL(JJ), YDATA(JJ)
              WRITE(*,4006) JJ, FXVAL(JJ), YDATA(JJ)

           END IF
C-----------
 26      CONTINUE
C--------
        WRITE(*,802)
```

```
      READ(*,*) ITYPE
      IF( ITYPE .EQ. 1 ) GOTO 1

 801  FORMAT(/,' ENTER NUMBER OF VALUES IN BASE PERIOD:')
 802  FORMAT(/,' ENTER 0=CONTINUE, 1=CHANGE NUMBER OF BASE VALUES:')

 4000 FORMAT(/'BASE PERIOD INPUT VALUES:'
     +        /'INDEX      X          Y'
     +        /'=====  ========  ========')

 4001 FORMAT(I5,2X,F8.3,2X,F8.2)

 4002 FORMAT(/,'BASE PERIOD PARAMETERS:',/,
     +        'X-VALUE RANGE = ',F5.0,'-',F5.0,/,
     +        'NUMBER OF VALUES = ',I3,/,
     +        'NUMBER OF MISSING VALUES = ',I3,/,
     +        'BASE MEAN = ',F8.3,/,
     +        'BASE VARIANCE = ',F8.3,/,
     +        'BASE SLOPE = ',F6.3,/)

 4003 FORMAT(' ALL VALUES IN BASE PERIOD MISSING')

 4004 FORMAT(/,'PROGRESSIVE PARAMETERS:',/,
     + '            X        Y                                        ',
     + '   DELTA',/,
     + 'INDEX    OBSERVATIONS      MEAN    VARIANCE   SLOPE   GAMMA',
     + '   GAMMA',/,
     + '=====  ================  ========  ========  ======  ======',
     + '  ======')

 4005 FORMAT(I5,2X,F8.3,1X,F7.2,2(2X,F8.3),2(2X,F6.3),2X,F6.3)

 4006 FORMAT(I5,2X,F8.3,1X,F7.2)

      RETURN
      END
C-----------------------------------------------------------------------
      SUBROUTINE RCOEFF( N, B0, B1 )
C
C  SUBROUTINE TO COMPUTE LINEAR LEAST SQUARES (LLS) INTERCEPT AND SLOPE.
C
C  VARIABLES USED IN SUBROUTINE NOT DESCRIBED IN MAIN PROGRAM:
C  'I'      DO LOOP COUNTER.
C  'N'      NUMBER OF VALUES.
C  'NN'     COUNTER FOR NON-MISSING VALUES.
C  'SX2'    SUMS OF SQUARED DEVIATIONS FROM THE MEAN X-VALUE.
C  'SXY'    SUM OF THE CROSS PRODUCTS OF DEVIATIONS.
C  'SUMX'   SUM OF NON-MISSING X-VALUES.
C  'SUMY'   SUM OF NON-MISSING Y-VALUES.
C  'SUMSX2' SUM OF X-VALUES SQUARED.
C  'SUMSXY' SUM OF X TIMES Y SQUARED.
C  'XBAR'   MEAN OF NON-MISSING X-VALUES.
C  'XBAR2'  NUMBER OF NON-MISSING VALUES TIMES 'XBAR' SQUARED.
C  'XWORK'  WORK ARRAY FOR X-VALUES.
C  'XYN'    'XBAR' TIMES 'YBAR' TIMES NUMBER OF NON-MISSING VALUES.
C  'YBAR'   MEAN OF NON-MISSING Y-VALUES.
C  'YWORK'  WORK ARRAY FOR Y-VALUES.
C
      PARAMETER (NDIM=10000)
      COMMON /WORK/ FXVAL(NDIM), FYVAL(NDIM), XDATA(NDIM),
```

```
     +                YDATA(NDIM), FINDEX(NDIM), NCASE, NSUB, H,
     +                OUNIT, FMISS
        REAL XWORK(NDIM), YWORK(NDIM)
        INTEGER OUNIT, H

        SUMX = 0.
        SUMY = 0.
        NN = 0
C
C  FILL WORK ARRAYS WITH NON-MISSING VALUES.
C
C--------
        DO 15 I = 1, N
C-----------
           IF( YDATA(I) .GT. FMISS ) THEN

              NN = NN + 1
              XWORK(NN) = XDATA(I)
              YWORK(NN) = YDATA(I)

           END IF
C-----------
 15     CONTINUE
C--------
C
C  COMPUTE SUMS OF X- AND Y-VALUES.
C
C--------
        DO 20 I = 1, NN

           SUMX = SUMX + XWORK(I)
           SUMY = SUMY + YWORK(I)

 20     CONTINUE
C--------
C
C  COMPUTE MEANS OF X- AND Y-VALUES.
C
        XBAR = SUMX / FLOAT(NN)
        YBAR = SUMY / FLOAT(NN)

        SUMSX2 = 0.
        SUMSXY = 0.
C
C  COMPUTE SUMS OF SQUARED DEVIATIONS AND CROSS PRODUCTS.
C
C--------
        DO 22 I = 1, NN

           SUMSX2 = SUMSX2 + XWORK(I)**2
           SUMSXY = SUMSXY + (XWORK(I) * YWORK(I))

 22     CONTINUE
C--------
        XBAR2 = XBAR**2 * FLOAT(NN)
        XYN = XBAR * YBAR * FLOAT(NN)

        SX2 = SUMSX2 - XBAR2
        SXY = SUMSXY - XYN
C
```

```
C  COMPUTE COEFFICIENTS.
C
         B1 = SXY / SX2
         B0 = YBAR - B1 * XBAR

         RETURN
         END
C-----------------------------------------------------------------------
         SUBROUTINE UNIVAR( N, NVALS, NMISS, YBAR, YVAR, YSUM )
C
C  SUBROUTINE TO COMPUTE LINEAR LEAST SQUARES (LLS) INTERCEPT AND SLOPE.
C
C  VARIABLES USED IN SUBROUTINE NOT DESCRIBED IN MAIN PROGRAM:
C  'I'       DO LOOP COUNTER.
C  'N'       NUMBER OF VALUES.
C  'SUMS2'   SUM OF SQUARED DIFFERENCES.
C  'YWORK'   WORK ARRAY FOR NON-MISSING Y-VALUES.
C
         PARAMETER (NDIM=10000)
         COMMON /WORK/ FXVAL(NDIM), FYVAL(NDIM), XDATA(NDIM),
     +                 YDATA(NDIM), FINDEX(NDIM), NCASE, NSUB, H,
     +                 OUNIT, FMISS
         REAL YWORK(NDIM)
         INTEGER OUNIT, H

         YSUM = 0.
         NVALS = 0
         NMISS = 0
C
C  FILL WORK ARRAY WITH NON-MISSING VALUES.
C
C--------
         DO 15 I = 1, N
C-----------
            IF( YDATA(I) .GT. FMISS ) THEN

               NVALS = NVALS + 1
               YWORK(NVALS) = YDATA(I)

            ELSE

               NMISS = NMISS + 1

            END IF
C-----------
 15      CONTINUE
C--------
C
C  COMPUTE SUM OF Y-VALUES.
C
C--------
         DO 20 I = 1, NVALS

            YSUM = YSUM + YWORK(I)

 20      CONTINUE
C--------
C
C  COMPUTE MEAN OF Y-VALUES.
C
```

```
      YBAR = YSUM / FLOAT(NVALS)
C
C  COMPUTE SUM OF SQUARED DIFFERENCES.
C
      SUMS2 = 0.
C--------
      DO 22 I = 1, NVALS

         SUMS2 = SUMS2 + (YWORK(I) - YBAR)**2

 22   CONTINUE
C--------
C
C  COMPUTE VARIANCE OF Y-VALUES.
C
      YVAR = SUMS2 / (FLOAT(NVALS) - 1)

      RETURN
      END
C-----------------------------------------------------------------------
      SUBROUTINE GAMMLN( XX, GAMEXP )
C
C  RETURNS THE VALUE LN[GAMMA(XX)] FOR XX > 0.
C  FROM NUMERICAL RECIPES IN FORTRAN, 2ND EDITION, 1992, PP. 206-207.
C
      REAL COF(6)

      DATA COF, STP / 76.18009172947146, -86.50532032941677,
     +                24.01409824083091, -1.231739572450155,
     +                .1208650973866179E-2, -.5395239384953E-5,
     +                2.5066282746310005 /

      X = XX
      Y = X
      TMP = X + 5.5
      TMP = (X + 0.5) * ALOG(TMP)-TMP
      SER = 1.000000000190015
C--------
      DO 20 J = 1, 6

         Y = Y + 1.
         SER =SER + COF(J) / Y

 20   CONTINUE
C--------
      GAMMALN = TMP + ALOG(STP*SER/X)
      GAMEXP = EXP(GAMMALN)

      RETURN
      END
```