INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI®

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

University of Alberta

ON SEQUENTIAL CHANGE-POINT DETECTION

by

Daniel Serban ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Statistics

Department of Mathematical and Statistical Sciences

Edmonton, Alberta

Spring 2005

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Library and Archives Canada

Published Heritage Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

0-494-08295-X

Your file Votre référence
ISBN:
Our file Notre reterence
ISBN:

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

To my parents

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Abstract

In this thesis new procedures for sequential testing of composite hypotheses are proposed. The procedures are based on large sample approximations of the efficient score vector.

In the first part of the thesis we consider the problem of change-point detection. Specifically, in Chapter 2, we will develop a new CUSUM-type procedure for sequential detection of a change-point in the distribution of a sequence of independent observations (not necessarily from the exponential family of distributions). In Chapter 3 the method is extended to the case of autocorrelated observations. In both cases we investigate the asymptotic distribution of the test statistics under the alternative hypothesis of change.

In Chapter 4 we compute the critical values for the tests of Gombay (2002) and show how these procedures can be used in clinical trials for comparison of three or more treatments. In the last part we adapt the CUSUM-test to solve the change-point ANOVA problem.

At the end of each chapter Monte Carlo experiments are conducted in order to evaluate the empirical power of the new procedures and for comparison to other methods.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Dr. Edit Gombay, for introducing me to the wonderful world of sequential analysis, and for her immense support and guidance during my graduate study and thesis preparation. Without her help, this thesis would not have come to the light. It has been a great pleasure to work under her supervision.

Special thanks go to Dr. D. Wiens, Dr. B. Schmuland, and Dr. R. Karunamuni for their valuable comments and support. I thank Profs R.J. Kulperger and Dr. M. Zuo for agreeing to be external members on my exam committee and for their careful reading of my thesis. Thank you very much to all my committee members for sacrificing some of their time to accommodate my exam date.

Thanks to the Department of Mathematical and Statistical Sciences for the financial support I received during my stay. Thanks are also due to the support staff members and especially to Dona Guelzow for her unforgettable support.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Contents

1 Introduction 1

1.1 Review and Problems 1

1.1.1 Sequential testing of composite hypotheses 1

1.1.2 Change-point problems 3

1.2 Overview of the thesis 7

2 A CUSUM-Type Sequential Test 9

2.1 Preliminaries 9

2.2 The CUSUM test 13

2.3 Applications 19

2.3.1 Normally distributed observations 19

2.3.2 The nested random effects model 21

2.4 Simulation Studies 26

2.4.1 The case of normally distributed observations 26

2.4.2 The case of the nested random effects model 29

2.5 Consistency of the test 36

3 Monitoring Parameter Change in AR(p) Models 42

3.1 Preliminaries 42

3.2 Monitoring the mean of an AR(p) process 51

3.3 Monitoring the variance of an AR(p) process 54

3.4 Monitoring the coefficients of an AR(p) process 58

3.4.1 AR(1) process 61

3.4.2 AR(2) process 62

3.5 Simulation Studies 64

3.5.1 The case of AR(1) process 64

3.5.2 The case of AR(2) process 73

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

3.6 Consistency of the test statistics 75

4 Sequential ANOVA 82
4.1 Preliminaries 82
4.2 Sequential F-test 83
4.3 Tests based on Rao's statistic 85
4.3.1 Critical values for Test 1 87
4.3.2 Critical values for Test 2 88
4.4 Simulation Studies 89
4.4.1 Comparison of three treatments 90
4.4.2 Comparison of four treatments 91
4.5 Sequential change-point ANOVA 94
4.5.1 The case of three groups 95
4.5.2 The case of three groups 97

5 Concluding Remarks 98

Bibliography 101

Appendices 106

A Strong Approximations for AR(p) Models 107

B The Distribution of the Maximum of the Bessel Process 117

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

List of Tables

2.1 Test for monitoring the mean μ of a normally distributed population with $\mu_0 = 0$ and $\sigma = 1$, and various μ_A. Simulated power (Power) and average stopping time (AVST) and its standard deviation(SD). The level of significance is $\alpha = 0.05$, and $C(\alpha) = 2.24$. 27

2.2 Test for monitoring variance σ^2 of a normally distributed population with $\mu = 0$ and $\sigma_0 = 1$, and various σ_A. Simulated power (Power) and average stopping time (AVST) and standard deviation(SD) of AVST. The level of significance is $\alpha = 0.05$, and $C(\alpha) = 2.24$ 28

2.3 Test for monitoring mean μ of a nested random effect model with $\mu_0 = 0$, $\sigma = 0.2$, $\sigma_w = 0.3$, and $\sigma_b = 0.6$, and various μ_A. Simulated power (Power) and average stopping time (AVST). The level of significance is $\alpha = 0.05$, $d = 1$, and $C(\alpha) = 2.24$. . 31

2.4 Test for monitoring variance component σ^2 of a nested random effects model with $\mu = 0$, $\sigma_0 = 0.2$, $\sigma_w = 0.3$, and $\sigma_b = 0.6$, and various σ_A. Simulated power (Power) and average stopping time (AVST). The level of significance is $\alpha = 0.05$, $d = 1$, and $C(\alpha) = 2.24$. 32

2.5 Test for monitoring variance component σ_w^2 of a nested random effects model with $\mu = 0$, $\sigma = 0.2$, $\sigma_{0w} = 0.3$, and $\sigma_b = 0.6$, and various σ_{Aw}. Simulated power (Power) and average stopping time (AVST). The level of significance is $\alpha = 0.05$, $d = 1$, and $C(\alpha) = 2.24$. 32

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

2.6 Test for monitoring variance component σ_b^2 of a nested random effects model with $\mu = 0$, $\sigma = 0.2$, $\sigma_w = 0.3$, and $\sigma_{0b} = 0.6$, and various σ_{Ab}. Simulated power (Power) and average stopping time (AVST). The level of significance is $\alpha = 0.05$, $d = 1$, and $C(\alpha) = 2.24$. 33
2.7 Test for simultaneously monitoring all variance components σ^2, σ_w^2, and σ_b^2 of a nested random effects model with $\mu = 0$, $\sigma_0 = 0.2$, $\sigma_{0w} = 0.3$, and $\sigma_{0b} = 0.6$. Only σ is changing. Simulated power (Power) and average stopping time (AVST).The level of significance is $\alpha = 0.05$, $d = 3$ and $C(\alpha) = 2.632$. 33
2.8 Test for simultaneously monitoring all variance components σ^2, σ_w^2, and σ_b^2 of a nested random effects model with $\mu = 0$, $\sigma_0 = 0.2$, $\sigma_{0w} = 0.3$, and $\sigma_{0b} = 0.6$. Only σ_w is changing. Simulated power (Power) and average stopping time (AVST). The level of significance is $\alpha = 0.05$, $d = 3$ and $C(\alpha) = 2.632$. 34
2.9 Test for simultaneously monitoring all variance components σ^2, σ_w^2, and σ_b^2 of a nested random effects model with $\mu = 0$, $\sigma_0 = 0.2$, $\sigma_{0w} = 0.3$, and $\sigma_{0b} = 0.6$. Only σ_b is changing. Simulated power (Power) and average stopping time (AVST). The level of significance is $\alpha = 0.05$, $d = 3$ and $C(\alpha) = 2.632$. 34
2.10 Test for simultaneously monitoring all four parameters of a nested random effects model with $\mu_0 = 0$, $\sigma_0 = 0.2$, $\sigma_{0w} = 0.3$, and $\sigma_{0b} = 0.6$. Only the mean is changing. Simulated power (Power) and average stopping time (AVST). The level of significance is $\alpha = 0.05$, $d = 4$ and $C(\alpha) = 2.727$. 35

3.1 Test for monitoring the mean of an AR(1) model $Y_i - \mu_0 = \phi(Y_{i-1} - \mu_0) + \varepsilon_i$. Simulated power (Power) and average stopping time (AVST). $\sigma^2 = 1$ and the in-control mean is $\mu_0 = 0$. The level of significance is $\alpha = 0.05$. 66
3.2 Test for monitoring σ of an AR(1) model $Y_i - \mu = \phi(Y_{i-1} - \mu) + \varepsilon_i$. Simulated power (Power) and average stopping time (AVST). $\mu = 0$ and the in-control value is $\sigma_0 = 1$. The level of significance is $\alpha = 0.05$. 69

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

3.3 Test for monitoring the coefficient ϕ_0 of the AR(1) model $Y_i - \mu = \phi_0(Y_{i-1} - \mu) + \varepsilon_i$. Simulated power (Power) and average stopping time (AVST). $\phi_0 = 0.1$, $\mu = 0$, $\sigma^2 = 1$ and various ϕ_A. The level of significance is $\alpha = 0.05$ and $C(\alpha) = 2.24$. 71
3.4 Test for monitoring the coefficient ϕ_0 of the AR(1) model $Y_i - \mu = \phi_0(Y_{i-1} - \mu) + \varepsilon_i$. Simulated power (Power) and average stopping time (AVST). $\phi_0 = -0.5$, $\mu = 0$, $\sigma^2 = 1$ and various ϕ_A. The level of significance is $\alpha = 0.05$ and $C(\alpha) = 2.24$. . . . 72
3.5 Test for monitoring the coefficient ϕ_0 of the AR(1) model $Y_i - \mu = \phi_0(Y_{i-1} - \mu) + \varepsilon_i$. Simulated power (Power) and average stopping time (AVST). $\phi_0 = 0.5$, $\mu = 0$, $\sigma^2 = 1$ and various ϕ_A. The level of significance is $\alpha = 0.05$ and $C(\alpha) = 2.24$. 73
3.6 Monitoring the coefficients ϕ_1 and ϕ_2 of the AR(2) model $Y_i - \mu = \phi_1(Y_{i-1} - \mu) + \phi_2(Y_{i-2} - \mu) + \varepsilon_i$. Simulated power (Power) and average stopping time (AVST). The in-control parameters are $\phi_{10} = -0.1$ and $\phi_{20} = 0.1$. Only ϕ_{10} is changing to ϕ_{1A}. . . 75

4.1 Critical value $C_1(\alpha, d)$ for different values of d and different levels of significance α. 89
4.2 Critical value $C_2'(\alpha, d, n_0)$ obtained using Vostrikova's formulae (4.14), for $\alpha = 0.05$ and different values of d and n_0 used in later simulations. 89
4.3 Comparison of three treatments. Simulated power (Power) and average stopping time (AVST) for Test 1, Test 2, and Sequential F-test with normally distributed outcomes for various δ. Level of significance $\alpha = 0.05$. 91
4.4 Comparison of four treatments. Simulated power (Power) and average stopping time (AVST) for Test 1, Test 2 and Sequential F-test, with normally distributed outcomes for various δ and n_0. Level of significance $\alpha = 0.05$. 93
4.5 Simulated power (Power) and average stopping time (AVST) for the CUSUM test with normally distributed outcomes. $\sigma^2 = 1$, $\mu_2 = \mu_3 = 0$ and for various $\mu_1^{(A)}$. Level of significance $\alpha = 0.05$. 96

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

4.6 Simulated power (Power) and average stopping time (AVST) for the CUSUM test with normally distributed outcomes. $\sigma^2 = 1$, $\mu_2 = \mu_3 = \mu_4 = 0$ and for various $\mu_1^{(A)}$. Level of significance $\alpha = 0.05$. 96

B.1 Maximum of the Bessel process probability in left-hand tail, $P(\sup_{0\leq t\leq 1} R^{(d)}(t) < z)$, for $d = 2$. 118
B.2 Maximum of the Bessel process probability in left-hand tail, $P(\sup_{0\leq t\leq 1} R^{(d)}(t) < z)$, for $d = 3$. 119
B.3 Maximum of the Bessel process probability in left-hand tail, $P(\sup_{0\leq t\leq 1} R^{(d)}(t) < z)$, for $d = 4$. 120
B.4 Maximum of the Bessel process probability in left-hand tail, $P(\sup_{0\leq t\leq 1} R^{(d)}(t) < z)$, for $d = 5$. 121
B.5 Maximum of the Bessel process probability in left-hand tail, $P(\sup_{0\leq t\leq 1} R^{(d)}(t) < z)$, for $d = 6$. 122
B.6 Maximum of the Bessel process probability in left-hand tail, $P(\sup_{0\leq t\leq 1} R^{(d)}(t) < z)$, for $d = 7$. 123
B.7 Maximum of the Bessel process probability in left-hand tail, $P(\sup_{0\leq t\leq 1} R^{(d)}(t) < z)$, for $d = 8$. 124

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

List of Figures

3.1 Power vs Coefficient when testing change in the mean of an AR(1) model. The truncation point is $n_0 = 200$ and the change point is $\tau = 100$. The in-control value is μ_0 and the drift after the change is $m = \mu_A - \mu_0$. 67

3.2 Power when testing for change in the coefficient for different AR(1) models ($b = \phi$). The truncation point is $n_0 = 200$ and the change point is $\tau = 100$. The initial coefficient value is -0.9, -0.5, 0, 0.5, 0.9. 67

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Conventions and notations:

The following notations will be used

Symbol	Meaning
$x_t = o(\phi(t))$ $a.s.$	$\limsup_{t\to\infty} \left\lvert \frac{x_t}{\phi(t)} \right\rvert = 0$ almost surely
$x_t \overset{a.s.}{=} O(\phi(t))$ or $x_t = O(\phi(t))$ $a.s.$	$\limsup_{t\to\infty} \left\lvert \frac{x_t}{\phi(t)} \right\rvert$ is almost surely finite
$x_t = O_p(\phi(t))$	$\frac{x_t}{\phi(t)}$ is bounded in probability
$\xrightarrow{a.s.}$	Almost sure convergence
$\xrightarrow{\mathcal{D}}$	Convergence in distribution
$\overset{\mathcal{D}}{=}$	Equality of finite dimensional distributions
LIL	Law of Iterated Logarithm
X^t	The transpose of vector or matrix X

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Chapter 1

Introduction

1.1 Review and Problems

1.1.1 Sequential testing of composite hypotheses

Sequential testing of hypotheses began in 1943 with the *Sequential Probability Ratio Test* (SPRT) proposed by Abraham Wald. Let $Y_1, Y_2, \ldots, Y_k, \ldots$, be a sequence of independent and identically distributed random variables with a common one-parameter density function $f(\cdot;\theta)$. Wald's SPRT procedure for testing the simple null hypothesis $H_0 : \theta = \theta_0$ versus the simple alternative $H_A : \theta = \theta_A$ is to stop sampling and accept H_0 as soon as $L_k \geq A$ or stop sampling and accept H_A as soon as $L_k \leq B$, where $0 < B < 1 < A$ are constant stopping boundaries dictated by error probabilities $\alpha = P_{\theta_0}(\text{Reject } H_0)$ and $\beta = P_{\theta_A}(\text{Reject } H_A)$, and

$$L_k = \frac{\prod_{i=1}^{k} f(y_i;\theta_0)}{\prod_{i=1}^{k} f(y_i;\theta_A)} \ ,$$

is the likelihood ratio based on the k observations available thus far. The sample size, N, at which the boundaries are crossed, is a random variable. The

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

mean of N is known as Average Sample Number (ASN) or average stopping time. Wald and Wolfowitz (1948) showed that the so defined SPRT procedure is optimal in the sense of minimizing the stopping times among all tests possessing a finite ASN and with error probabilities α and β.

The original SPRT did not deal with composite hypotheses, which were composite either because the null and/or alternative parameter spaces are not single points or because of the presence of nuisance parameters. Wald attempted without much success to adapt the SPRT to the composite hypotheses case by introducing the weight functions approach. Another attempt to extend the SPRT to the case of nuisance parameters produced the so-called *Invariant SPRT* (Ghosh 1970). This method consists of reducing the composite hypotheses to simple hypotheses by transforming the data as well as the hypotheses of interest and then applying Wald's SPRT procedure. According to Lai (2001), this approach has a few drawbacks that makes it difficult to implement.

In a case where the hypotheses of interest are composite because of the presence of nuisance parameters, a third approach suggested by Bartlett (1949), Cox (1963) and Breslow (1969) is based on using the likelihood ratio, or an asymptotically equivalent form of it, under the assumption of contiguity. The method replaces nuisance parameters in the likelihood ratio by their restricted maximum likelihood estimators and uses Wald's SPRT procedure. Based on the assumption that the parameters are close, i.e. $|\theta_A - \theta_0| = O(N^{-1/2})$, the Taylor expansion is truncated at the second order terms. Because the sample size N is a random variable with range $(0, \infty)$ the error made by truncation at the second term is not negligible. Alternatively, Gombay (1996, 1997) provided some tests based on the *generalized sequential likelihood*

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

ratio (GLR) along with their asymptotic critical values at significance level α.

Sequential testing of hypotheses was introduced into clinical trials during the 50's (Armitage 1960). As an alternative to the SPRTs, Armitage suggested the so-called *Repeated Significance Test* (RST). This method has some limitations which are removed in *group sequential analysis* introduced by Pocock (1977) and O'Brien and Fleming (1979). In this case, if no treatment difference is observed the trial will continue to its maximum sample size. This will be somewhat larger than the sample size required for a fixed-sample design of equivalent power.

The problem of comparing three or more treatments is frequently faced in clinical trials. Siegmund (1993) and Betensky (1996) studied the case of three treatments in the simplest situation of independent, normally distributed random variables with a common known variance. To make these procedures more useful in practice it is necessary to extend them to the case of unknown and perhaps unequal variances, and to the case of more than three treatments. It seems likely that the methods of Siegmund (1985), Section 5.4, could be applied although the analytic approximations will become more complicated.

1.1.2 Change-point problems

Sequential *change-point* detection problems have many important applications, including industrial quality control, reliability, fault detection, clinical trials, finance, signal detection, surveillance and security systems. Extensive research has been done in this field during the last few decades. For recent reviews, we refer readers to Csörgő and Horváth (1997), Basseville and Nikiforov (1993) and Lai (1995).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

In sequential change-point problems, one observes a sequence of independent observations $Y_1, Y_2, \ldots$ from some process. Initially, the process is *in control*, i.e., the Y's have some distribution f_0. At some unknown time τ, the process may go *out of control* and the Y's have another distribution f_A. The problem is to detect the change with a high power and as soon as possible while keeping false alarms as infrequent as possible.

When both the pre-change distribution f_0 and post-distribution f_A are completely specified, the problem is well understood and has been solved under a variety of criteria. Some popular procedures are Shewhart's control charts, Moving Average control charts, Page's CUSUM procedure, and the Shiryayev-Roberts procedure.

Perhaps the most successful algorithm of sequential change detection is Page's (1954) CUSUM test. If the interest is in the mean parameter, this test will monitor the partial sums

$$S_k - \min_{1 \leq j \leq k} S_j \ , \qquad k = 1, 2, \ldots, \tag{1.1}$$

where $S_k = \sum_{i=1}^{k} (Y_i - r)$ and r is a reference value, and will indicate change from the initial μ_0 mean value to μ_A, $\mu_A > \mu_0$, at time k, when statistic (1.1) is large enough.

This basic idea of change detection has several variations. The most famous is the case when the sums in (1.1) are not based on the initial observations Y_i, but on the ratios

$$r_i = \log \frac{f(Y_i; \theta_0)}{f(Y_i; \theta_A)} \ , \qquad i = 1, 2, \ldots, \tag{1.2}$$

and the purpose is to detect change from the initial parameter value θ_0 to the new value θ_A. Lorden (1971) and Moustakides (1986) have shown that

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

algorithms based on (1.1) and (1.2) are optimal in the sense of minimizing the expected value of the stopping time after change, while keeping the expected stopping time under the null hypothesis of no change at a desired level.

In practice, the assumption of known pre-change distribution f_0 and post distribution f_A is too restrictive. Motivated by applications in statistical quality control, the standard formulation of a more flexible model assumes that the pre-change distribution f_0 is given and the post-change distribution f_A involves unknown parameters. However, as shown by many examples, there are many situations in practice in which both the pre-change and the post-change distributions intrinsically involve unknown parameters.

So, for practical purposes this case has to be extended to the case of composite hypotheses. The natural idea is to use the generalized likelihood ratio in place of the sum of variables in (1.1). It has been tried, but computational difficulties (see Baseville and Nikiforov (1993)) have prevented its widespread use. Also, there are theoretical problems with its extension to cases involving nuisance parameters (see Gombay (2002, 2003b) for detailed discussion). The essence of the reason for these problems is that the generalized likelihood ratio is not like a partial sums sequence, but a quadratic form. There is, however, a statistic that is closely related to the generalized likelihood ratio and behaves approximately like a partial sums sequence. This is the efficient score vector, the main component of Rao's statistic, which is a quadratic form made with the help of the efficient score vector, and behaves, asymptotically the same way as the generalized likelihood ratio (see Gombay (1997) for detailed discussion).

In this thesis we assume that the pre-change distribution $f_0 = f(y; \theta_0, \eta)$ depends on two vectors of parameters θ_0 and η, where θ_0, the vector of parameters of interest is known, and η is the vector of nuisance parameters. The

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

post-change distribution is of the form $f_A = f(y; \theta_A, \eta)$ with θ_A unknown.

The methods discussed so far assume independent observations. As Wetherill (1977) has pointed out, observations from modern industrial processes are often autocorrelated and the process itself can behave like an autoregressive process. Such behaviour must be taken into account when setting up testing procedures.

Most procedures in the literature are based on residuals (see Kulperger (1985), Bai (1993), and Horváth (1993)) or on the one-step-ahead prediction errors (see Montgomery and Friedman (1989)). There are several extensions of CUSUM and GLR schemes to handle non-independent observations. In principle, Page's likelihood ratio CUSUM scheme (Page, 1955) can be easily extended to non-independent observations, simply by replacing $f(Y_i)$ by $f(Y_i|Y_1, ..., Y_{i-1})$. However, as Basseville and Nikiforov (1993) noted, practical implementation of the GLR algorithm is not always possible because the number of computations at time n grows very fast to infinity with n and maximization of the log-likelihood over $\theta \in \Theta$ must be carried out for each possible change time k between 1 and n. So these algorithms cannot be computed in realtime to support on-line decisions. Our procedure which is based on the efficient score vector has no such limitations.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

1.2 Overview of the thesis

As mentioned in the previous section, sequential testing of composite hypotheses has not yet received an adequate fully-sequential treatment. In this thesis we use large sample approximations of the efficient score vector to develop new CUSUM-type procedures for sequential testing of hypotheses in presence of nuisance parameters. The main attractive features of these procedures are:

- A generality allowing application of the methods to a wide class of distribution families including the exponential family.
- Simple accommodation of the nuisance parameters. In fact only the nuisance parameters have to be estimated giving a simpler algorithm.
- The simple structure of the efficient score vector allows the definition of the CUSUM-type test which is not feasible for the generalized likelihood ratio and the m.l.e. based on Wald's statistic sequence.
- Different test statistics are used for testing for a change in different parameters as the tests are based on the corresponding components of the efficient score vector.
- Easy-to-compute approximate boundaries (critical values) which do not require any numerical integration.

This thesis has five objectives:

1. Development of a class of CUSUM-type sequential procedures for testing for a change in the parameters of the distribution of a sequence of independent observations. The distribution may be from the non-exponential family. This extends the results of Gombay (2003a).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

2. Extension of the procedures to the case of non-independent observations. More specifically, we consider the case of testing for a change in the parameters of an autoregressive process of order p, AR(p).

3. To compute the critical values and show how the tests of Gombay (2003) based on Rao's statistic can be used for sequential comparison of three or more treatments.

4. To adapt the CUSUM test to the ANOVA change-point problem.

5. Empirical comparison of our test procedures to some other methods, in term of power and average stopping time.

Specifically, in Chapter 2 we shall consider the case of independent observations. As the efficient score vector behaves approximately as partial sum sequences, the CUSUM idea of Page (1954) can be used to improve performance for detection of later changes. Monte Carlo simulations are carried out in order to assess the power of the new CUSUM-type sequential test.

In Chapter 3, the same method is extended to the case of autocorrelated observations. When the observations come from an AR(p) process it turns out that the distribution and the rate of convergence of the test statistic is the same as in the case of independent observations.

In Chapter 4, we will present and compare some procedures based on Rao's statistic for testing equality of three or more treatments. We compare these procedures with the sequential F-test proposed by Siegmund (1980). At the end we show how the CUSUM test can be used for an ANOVA change-point problem.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Chapter 2

A CUSUM-Type Sequential Test

2.1 Preliminaries

One of the major aims of Statistical Process Control (SPC) is to achieve the condition where the parameters related to a given manufacturing, business, ecological or similar process, conform to some prescribed on-target behaviour. In many practical applications it is not reasonable to assume that the same model remains adequate as time progresses. Over time, something will inevitably change and possibly cause deterioration in process quality. Something that affects process quality is assumed to be reflected by a change in the parameters so the basic goal of process monitoring is to detect changes in the parameters that can occur at unknown time. In practical situations the relevant model involves not only the parameters of interest but also some nuisance parameters which are not monitored but they influence the functionality of the process.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

In this chapter a truncated CUSUM-type sequential test is proposed to detect an abrupt change in the distribution of a sequence of independent observations. The test is based on large sample approximations of the efficient score vector under the null hypothesis of no change and under the alternative hypothesis of change at an unknown time.

The problem can be described in general terms as follows. Suppose $Y_1, Y_2, \ldots,$ are independent random variables/vectors observed sequentially, i.e., one at a time, and let $f(y; \theta_i, \eta_i)$ be the density function with respect to a σ-finite measure ν. The distribution function is denoted by $F(\cdot\,; \theta, \eta)$. We assume that $\theta \in \Omega_1 \subset \mathbb{R}^d$, $d \geq 1$, and that $\eta \in \Omega_2 \subset \mathbb{R}^p$, $p \geq 0$, and the parameter space is $\Omega = \Omega_1 \times \Omega_2 \subset \mathbb{R}^{d+p}$. We use the notation $\xi = (\theta, \eta)$. In statistical terms θ will be the parameter of interest and η will be the nuisance parameter. We are interested in testing the composite null hypothesis

$$H_0 : \theta_i = \theta_0,\ \eta_i = \eta \in \Omega_2\ , \qquad i = 1, 2, \ldots,$$

against the alternative

$$H_A : \begin{cases} \theta_i = \theta_0\ , & \eta_i = \eta \in \Omega_2\ , \quad i < \tau\ , \\ \theta_i = \theta_A\ , & \eta_i = \eta \in \Omega_2\ , \quad i \geq \tau\ . \end{cases} \tag{2.1}$$

where τ is the unknown time of change, θ_0 is the known initial value of the parameter, and parameter values η, θ_A are also unknown. For example, in quality control the initial value θ_0 is the required measurement or target value, and change from it has to be detected.

Before presenting the test statistic and its asymptotic null distribution we will give some notations and regularity conditions that are needed. We denote by

$$I_{ij}(\xi) = -E_\xi \left[\frac{\partial^2}{\partial \xi_i \partial \xi_j} \log f(Y; \xi) \right]\ , \tag{2.2}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the entries of the Fisher information matrix, $\boldsymbol{I}$, where $i, j = 1, 2, \ldots, d+p$. We shall partition this matrix as

$$\boldsymbol{I} = \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix} ,$$

where

$$I_{11} = \left(-E\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log f(Y;\xi)\right)_{d\times d}, \qquad I_{22} = \left(-E\frac{\partial^2}{\partial\eta_i\partial\eta_j}\log f(Y;\xi)\right)_{p\times p},$$

and $I_{12} = I_{21}^t = \left(-E\frac{\partial^2}{\partial\theta_i\partial\eta_j}\log f(Y;\xi)\right)_{d\times p}$.

The inverse of $\boldsymbol{I}$ will also be partitioned and denoted by

$$\boldsymbol{I}^{-1} = \begin{pmatrix} I^{11} & I^{12} \\ I^{21} & I^{22} \end{pmatrix} .$$

Suppose that $\xi = (\theta, \eta)$ is a point in an open subset $\Omega \subset \mathbb{R}^{d+p}$. The following regularity and existence conditions are needed

C1. The distribution function $F(\cdot\,;\theta,\eta)$, is identifiable over Ω.

C2. There exists an open subset, $\Omega_0 \subset \Omega$, containing (θ_0, η), the true value of the parameter under H_0, where the partial derivatives,

$$\frac{\partial}{\partial\xi_i}\log f(y;\xi)\,, \quad \frac{\partial^2}{\partial\xi_i\partial\xi_j}\log f(y;\xi)\,, \quad \text{and} \quad \frac{\partial^3}{\partial\xi_i\partial\xi_j\partial\xi_k}\log f(y;\xi)\,,$$

exist and are continuous for all $y \in \mathbb{R}$ and $\xi \in \Omega_0$.

C3. For each $(\theta_0, \eta) \in \Omega_0$, $k = 1, 2, \ldots,$ and $j < k$ the score equations

$$\sum_{i=1}^{k} \nabla_\eta \log f(Y_i;\theta_0,\eta) = 0\,,$$

$$\sum_{i=j}^{k} \nabla_\eta \log f(Y_i;\theta_0,\eta) = 0\,,$$

have unique solutions, $\hat{\eta}_k$ and $\hat{\eta}_{k,j}$, respectively.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

C4. Under the setup of C2, there are functions $M_1(y)$, $M_2(y)$ such that

$$\int M_1(y)\nu(dy) < \infty \text{ , and } \quad E_\xi[M_2(Y)] < \infty \text{ ,}$$

with

$$\left|\frac{\partial}{\partial \xi_i}\log f(y;\xi)\right| \le M_1(y), \ \left|\frac{\partial^2}{\partial \xi_i \partial \xi_j}\log f(y;\xi)\right| \le M_2(y),$$

$$\text{and} \quad \left|\frac{\partial^3}{\partial \xi_i \partial \xi_j \partial \xi_k}\log f(y;\xi)\right| \le M_2(y),$$

for all $\xi \in \Omega_0$, and $1 \le i, j, k \le d + p$.

C5. $E_\xi[\frac{\partial}{\partial \xi_i}\log f(Y;\xi)] = 0$, for all $1 \le i \le d+p$, and $\xi \in \Omega_0$.

C6. $I_{ij}(\xi) = -E_\xi\left[\frac{\partial^2}{\partial \xi_i \partial \xi_j}\log f(Y;\xi)\right]$ and $\boldsymbol{I}^{-1}(\xi)$ exist and are continuous for all $\xi \in \Omega_0$, and $1 \le i, j \le d+p$.

C7. $Var_\xi\left[\frac{\partial^2}{\partial \xi_i \partial \xi_j}\log f(Y;\theta_0,\eta)\right] < \infty$, for $1 \le i, j \le d+p$.

C8. $E_{\theta_0,\eta}\left[\left|\frac{\partial^2}{\partial \xi_i^2}\log f(Y;\theta_0,\eta)\right|^{2+\delta}\right] < \infty$, for all $i = 1, 2, ..., d+p$, and some $\delta > 0$.

Remark 2.1 *Conditions C1 - C6 are the usual classical regularity conditions guaranteeing the existence and consistence of a sequence of MLE's (c.f. Lehmann 2001, Serfling 1980). The last two conditions are, respectively, required by the Law of the Iterated Logarithm (Serfling (1980)), and by the strong invariance principles of Csörgő and Révész (1981) that are used in this thesis.*

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

2.2 The CUSUM test

The truncated CUSUM-type sequential test proposed here is based on the efficient score vector which is defined as

$$V_k(\theta_0, \eta) = \sum_{i=1}^{k} \nabla_\xi \log f(Y_i; \theta_0, \eta) \ .$$

Because the nuisance parameter η is present we have to replace it by its restricted maximum likelihood estimator $\hat{\eta}_k$, that is, by the solution of equation

$$\sum_{i=1}^{k} \nabla_\eta \log f(Y_i, \theta_0, \eta) = 0 \ .$$

When we replace η by $\hat{\eta}_k$ V_k simplifies to

$$V_k = \sum_{i=1}^{k} \nabla_\xi \log f(Y_i; \theta_0, \hat{\eta}_k) = \sum_{i=1}^{k} \nabla_\theta \log f(Y_i; \theta_0, \hat{\eta}_k) \ . \tag{2.3}$$

The following theorem states that under H_0 V_k defined in (2.3) can be written approximately as a sum of independent d-dimensional random vectors.

Theorem 2.1 Under H_0, if C1-C8 hold, then V_k can be represented as

$$\begin{aligned} V_k &= \sum_{i=1}^{k} \nabla_\theta \log f(Y_i; \theta_0, \hat{\eta}_k) \\ &= \sum_{i=1}^{k} \left\{ \nabla_\theta \log f(Y_i; \theta_0, \eta) - \nabla_\eta \log f(Y_i; \theta_0, \eta) I_{22}^{-1}(\theta_0, \eta) I_{21}(\theta_0, \eta) \right\} \\ &\quad + O(\log\log k) \\ &= \sum_{i=1}^{k} Z_i + O(\log\log k) \quad a.s., \end{aligned} \tag{2.4}$$

as $k \longrightarrow \infty$, where Z_i are i.i.d.r.v's with $EZ_i = 0$, $Cov(Z_i) = \Gamma(\theta_0, \eta)$, $\Gamma(\theta_0, \eta) = I_{11} - I_{12} I_{22}^{-1} I_{21}$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Proof. We can write

$$
\begin{aligned}
V_k &= \sum_{i=1}^{k} \nabla_\theta \log f(Y_i; \theta_0, \hat{\eta}_k) \\
&= \sum_{i=1}^{k} \nabla_\theta \log f(Y_i; \theta_0, \eta) + \sum_{i=1}^{k} \left\{ \nabla_\theta \log f(Y_i; \theta_0, \hat{\eta}_k) - \nabla_\theta \log f(Y_i; \theta_0, \eta) \right\}.
\end{aligned}
\tag{2.5}
$$

The last term in the above sum represents the error committed in estimating η by its restricted MLE $\hat{\eta}_k$. The error can be linearized by using a three-term Taylor expansion of $\nabla_\theta \log f(Y_i; \theta_0, \hat{\eta}_k)$ around the true parameter value η, so that (2.5) can be rewritten as

$$
V_k = \sum_{i=1}^{k} \nabla_\theta \log f(Y_i, \theta_0, \eta) + (\hat{\eta}_k - \eta) \sum_{i=1}^{k} \nabla^2_{\theta\eta} \log f(Y_i, \theta_0, \eta) + R_k(\eta^*) \tag{2.6}
$$

where η^* is a point between $\hat{\eta}_k$ and η and $\nabla^2_{\theta\eta}$ denotes a $p \times d$ - matrix of second order partial derivatives, first with respect to components of θ and secondly with respect to components of η. The term $R_k(\eta^*)$ is a row vector whose r^{th} component has the form

$$
\frac{1}{2} \sum_{l=1}^{p} \sum_{q=1}^{p} (\hat{\eta}_{kq} - \eta_q)(\hat{\eta}_{kl} - \eta_l) \times \left\{ \sum_{i=1}^{k} \frac{\partial^3}{\partial \eta_q \partial \eta_l \partial \theta_r} \log f(Y_i; \theta_0, \eta^*) \right\}.
$$

By the Strong Law of Large Numbers and by C4, the terms in the curly brackets are almost surely $O(k)$. On the other hand, by Lemma 2.1 of Gombay and Horváth (1994),

$$
\left| (\hat{\eta}_{kq} - \eta_q)(\hat{\eta}_{kl} - \eta_l) \right| = O\left(\frac{\log \log k}{k} \right) \quad a.s.
$$

Hence we obtain

$$
R_k(\eta^*) = O(\log \log k) \quad a.s. \tag{2.7}
$$

In order to obtain an expression for $(\hat{\eta}_k - \eta)$, we shall analyze the following

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

three-term Taylor expansion

$$\sum_{i=1}^{k}\left\{\nabla_\eta \log f(Y_i,\theta_0,\hat\eta_k) - \nabla_\eta \log f(Y_i,\theta_0,\eta)\right\} = (\hat\eta_k - \eta)\sum_{i=1}^{k}\nabla^2_{\eta^2}\log f(Y_i,\theta_0,\eta) + R'_k(\eta^{**}) \qquad (2.8)$$

where $\nabla^2_{\eta^2}$ is the $p\times p$- matrix of partial derivatives with respect to the components of η. By the same arguments leading to (2.7), the error term above is almost surely of order $O(\log\log k)$. On the other hand, by C4 and the Law of Iterated Logarithm we have

$$\left\|\sum_{i=1}^{k}\nabla^2_{\eta^2}\log f(Y_i,\theta_0,\eta) + kI_{22}(\theta_0,\eta)\right\| = O\big(\sqrt{k\log\log k}\big) \quad a.s.$$

From (2.8) and by the definition of $\hat\eta_k$ we obtain

$$\begin{aligned}
-\sum_{i=1}^{k}\nabla_\eta \log f(Y_i;\theta_0,\eta) &= (\hat\eta_k-\eta)\left[\sum_{i=1}^{k}\nabla^2_{\eta^2}\log f(Y_i;\theta_0,\eta) + kI_{22} - kI_{22}\right] \\
&\quad + O(\log\log k) \\
&= (\hat\eta_k-\eta)\left[\sum_{i=1}^{k}\nabla^2_{\eta^2}\log f(Y_i;\theta_0,\eta) + kI_{22}\right] \\
&\quad - k(\hat\eta_k-\eta)I_{22} + O(\log\log k) \\
&= O(\log\log k) - k(\hat\eta_k-\eta)I_{22} + O(\log\log k)\ ,
\end{aligned}$$

and because I_{22}^{-1} exists we can write

$$(\hat\eta_k-\eta) = \frac{1}{k}\sum_{i=1}^{k}\nabla_\eta \log f(Y_i;\theta_0,\eta)I_{22}^{-1} + O\big(k^{-1}\log\log k\big)\ . \qquad (2.9)$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Now, collecting (2.6), (2.7) and (2.9) we obtain

$$\begin{aligned} V_k = & \sum_{i=1}^{k} \nabla_\theta \log f(Y_i;\theta_0,\eta) \\ & + \left[\frac{1}{k} \sum_{i=1}^{k} \nabla_\eta \log f(Y_i;\theta_0,\eta) I_{22}^{-1}(\theta_0,\eta) + O\left(\frac{\log\log k}{k}\right)\right] \\ & \times \left[\sum_{i=1}^{k} \nabla_{\theta\eta}^2 \log f(Y_i;\theta_0,\eta)\right] + O(\log\log k) \\ = & \sum_{i=1}^{k} \nabla_\theta \log f(Y_i;\theta_0,\eta) \\ & + \left[\sum_{i=1}^{k} \nabla_\eta \log f(Y_i;\theta_0,\eta) I_{22}^{-1}(\theta_0,\eta) + O(\log\log k)\right] \\ & \times \left[\frac{1}{k} \sum_{i=1}^{k} \nabla_{\theta\eta}^2 \log f(Y_i;\theta_0,\eta)\right] + O(\log\log k) \ . \end{aligned} \quad (2.10)$$

By using again C6 and LIL we have

$$\left\| \frac{1}{k} \sum_{i=1}^{k} \nabla_{\theta\eta}^2 \log f(Y_i;\theta_0,\eta) + I_{21}(\theta_0,\eta) \right\| = O\left(\sqrt{\frac{\log\log k}{k}}\right) \quad a.s.$$

Then it follows from (2.10) that

$$\begin{aligned} V_k = & \sum_{i=1}^{k} \nabla_\theta \log f(Y_i;\theta_0,\eta) \\ & + \left[\sum_{i=1}^{k} \nabla_\eta \log f(Y_i;\theta_0,\eta) I_{22}^{-1}(\theta_0,\eta) + O(\log\log k)\right] \\ & \times \left[-I_{21}(\theta_0,\eta) + O\left(\sqrt{\frac{\log\log k}{k}}\right)\right] + O(\log\log k) \ , \end{aligned}$$

and neglecting errors of amplitude less then $O(\log\log k)$ this is equivalent to

$$\begin{aligned} V_k = & \sum_{i=1}^{k} \left\{ \nabla_\theta \log f(Y_i;\theta_0,\eta) - \nabla_\eta \log f(Y_i;\theta_0,\eta) I_{22}^{-1}(\theta_0,\eta) I_{21}(\theta_0,\eta) \right\} \\ & + O(\log\log k) \\ = & \sum_{i=1}^{k} Z_i + O(\log\log k) \ , \end{aligned}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

that is, relation (2.4) with $Z_i = \nabla_\theta \log f(Y_i;\theta_0,\eta) - \nabla_\eta \log f(Y_i;\theta_0,\eta) I_{22}^{-1} I_{21}$. By C5 it follows that $E_{\theta_0,\,\eta}\big(Z_i\big) = 0$ for all i. If we denote $X_{1i} = \nabla_\theta \log f(Y_i;\theta_0,\eta)$ and $X_{2i} = \nabla_\eta \log f(Y_i,\theta_0,\eta)$ we can write $Z_i = X_{1i} - X_{2i} I_{22}^{-1} I_{21}$. Denote by the superscript t the transpose of a vector or matrix. Now we have

$$
\begin{aligned}
Cov(Z_i) &= E[Z_i^t Z_i] \\
&= E[(X_{1i}^t - I_{12} I_{22}^{-1} X_{2i}^t)(X_{1i} - X_{2i} I_{22}^{-1} I_{21})] \\
&= E[X_{1i}^t X_{1i}] - E[X_{1i}^t X_{2i} I_{22}^{-1} I_{21}] \\
&\quad - E[I_{12} I_{22}^{-1} X_{2i}^t X_{1i}] + E[I_{12} I_{22}^{-1} X_{2i}^t X_{2i} I_{22}^{-1} I_{21}] \\
&= Cov(X_{1i}) - E[X_{1i}^t X_{2i}] I_{22}^{-1} I_{21} \\
&\quad - I_{12} I_{22}^{-1} E[X_{2i}^t X_{1i}] + I_{12} I_{22}^{-1} Cov(X_{2i}) I_{22}^{-1} I_{21} \\
&= I_{11} - I_{12} I_{22}^{-1} I_{21} - I_{12} I_{22}^{-1} I_{21} + I_{12} I_{22}^{-1} I_{22} I_{22}^{-1} I_{21} \\
&= I_{11} - I_{12} I_{22}^{-1} I_{21} \ ,
\end{aligned}
$$

and the proof of the theorem is completed.

□

Based on Theorem 2.1 we can define a truncated CUSUM-type sequential test. For each $k \geq 2$, and each $1 \leq j < k$, the efficient score vector based on $Y_j, \ldots, Y_k$, is denoted by

$$V_{k,j} = \sum_{i=j}^{k} \nabla_\theta \log f(Y_i, \theta_0, \hat{\eta}_{k,j}) \ ,$$

where $\hat{\eta}_{k,j}$ is defined in C3. Under H_0, by Theorem 2.1, $V_{k,j}$ can be represented as

$$V_{k,j} = \sum_{i=j}^{k} Z_i + O\big(\log\log(k-j)\big) \ , \qquad (2.11)$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

with Z_i independent random vectors with mean zero and covariance matrix $\mathbf{\Gamma}(\theta_0, \eta)$ defined in Theorem 2.1. Then, when $d = 1$

$$W_{k,j} = \mathbf{\Gamma}^{-1/2}(\theta_0, \eta) V_{k,j}(\theta_0, \hat{\eta}_{k,j}) = \mathbf{\Gamma}^{-1/2}(\theta_0, \eta) \sum_{i=j}^{k} \nabla_\theta \log f(Y_i, \theta_0, \hat{\eta}_{k,j}) \;, \tag{2.12}$$

is approximately the sum of independent mean 0 and variance 1 r.v's.

Remark 2.2 *When $d > 1$ the vector $W_{k,j}$ has uncorrelated components and we can choose the approximating process with independent components. In this way we can monitor each component with a level α^* test which gives an overall level of significance $\alpha = 1 - (1 - \alpha^*)^d$. Hence it is enough to define a test for the case $d = 1$.*

It is easy to see that under H_0, if $d = 1$, as $n_0 \to \infty$

$$\max_{2 \le j < k \le n_0} \frac{1}{\sqrt{n_0}} \mathbf{\Gamma}^{-1/2}(\theta_0, \eta) V_{k,j}(\theta_0, \hat{\eta}_{k,j}) \xrightarrow{\mathcal{D}} \sup_{0 \le u < v \le 1} \left\{ W(v) - W(u) \right\} \stackrel{\mathcal{D}}{=} \sup_{0 \le t \le 1} |W(t)| \tag{2.13}$$

where $W(t)$ is a standard Wiener process and n_0 a fixed truncation point. An elementary proof of the last identity can be found in Gombay (1994).

In order for this result to be useful in testing hypotheses we need to replace the covariance matrix $\mathbf{\Gamma}(\theta_0, \eta)$ with an estimated version. By Lemma 2 of Gombay *et al.* (2001) the asymptotic distribution in (2.13) remains the same if η is replaced by its restricted m.l.e. $\hat{\eta}_k$ in $\mathbf{\Gamma}(\theta_0, \eta)$.

Remark 2.3 *$\hat{\eta}_k$ and $\hat{\eta}_{k,j}$ need not to be the unique solution of the restricted log-likelihood equation. All we need is that the estimator converges weakly to the parameter η with a rate of at least $k^{-1/2}$. This can be attained by, say, a one-step estimator based on the Newton-Raphson iterative procedure (see Lehmann 2001, p.475). Therefore condition C3 requiring that $\hat{\eta}_k$ and $\hat{\eta}_{k,j}$ be the unique MLE's can be relaxed accordingly.*

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The test is defined as follows.

CUSUM TEST. *For* $k = 2, 3, \ldots, n_0$ *compute*

$$T_k = \max_{1 \le j < k} \frac{1}{\sqrt{n_0}} \Gamma^{-1/2}(\theta_0, \hat{\eta}_k) V_{k,j}(\theta_0, \hat{\eta}_{k,j}) \ . \tag{2.14}$$

Stop and conclude that H_0 *is not supported by the data at the first* k *when* $T_k > C(\alpha)$. *Do not reject* H_0 *if it is not rejected by* $k = n_0$.

The critical value $C(\alpha)$ can be obtained from the well known formula (Borodin and Salminen (1996))

$$1 - \alpha = \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} \exp\left(-\frac{\pi^2(2k+1)^2}{8C(\alpha)^2}\right) ,$$

For example C(0.10)=1.96, C(0.05)=2.24, C(0.025)=2.50, and C(0.01)=2.80.

Remark 2.4 *As the partial sums have components* $Z_i = \nabla_\theta \log f(Y_i; \theta_0, \eta)$ - $\nabla_\eta \log f(Y_i, \theta_0, \eta) I_{22}^{-1} I_{21}$ *with* η *the true value of the parameter, alternatively, we may use* $V_{k,j}(\theta_0, \hat{\eta}_k)$ *in the test statistic* T_k. *It is the empirical properties that guide our choice in each case. Using* $\hat{\eta}_k$ *in* T_k *of (2.14) is computationally more efficient.*

2.3 Applications

2.3.1 Normally distributed observations

First we consider the simplest case of independent normally distributed observations. We present the tests for monitoring the mean and the variance. In the same way we can define a test for monitoring both parameters simultaneously.

a) Monitoring the mean of a normal distribution. Consider $Y_1, Y_2, \ldots,$ a sequence of independent normal random variables with mean μ and variance σ^2, and suppose we are interested in testing

$$H_0 : \mu_i = \mu_0, \ \sigma_i^2 = \sigma^2 \text{ unknown}, \quad i = 1, 2, \ldots,$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

against the alternative

$$H_A : \begin{cases} \mu_i = \mu_0 \ , \ \ \sigma_i^2 = \sigma^2 \text{ unknown}, & i < \tau, \\ \mu_i = \mu_A > \mu_0 \ , \ \ \sigma_i^2 = \sigma^2 \text{ unknown}, & i \geq \tau \ . \end{cases}$$

Without loss of generality we can assume $\mu_0 = 0$. For simplicity we consider the re-parametrization $\theta = \frac{\mu}{\sigma^2}$ and $\eta = -\frac{1}{2\sigma^2} < 0$. Under this re-parametrization, the hypothesis to be tested becomes

$$H_0 : \theta_i = 0, \ \eta_i = \eta \text{ unknown}, \quad i = 1, 2, \ldots,$$

against the alternative

$$H_A : \begin{cases} \theta_i = 0 \quad , \ \ \eta_i = \eta \text{ unknown}, & i < \tau, \\ \theta_i = \theta_A > 0 \ , \ \ \eta_i = \eta \text{ unknown}, & i \geq \tau \ . \end{cases}$$

In this case $d = p = 1$, and computations give $\hat{\eta}_k = -\frac{k}{2\sum_{i=1}^k Y_i^2}$, $\hat{\eta}_{k,j} = -\frac{k+1-j}{2\sum_{i=j}^k Y_i^2}$, and $V_{k,j}(0, \hat{\eta}_{k,j}) = \sum_{i=j}^k Y_i$. The information matrix is

$$\boldsymbol{I} = \begin{pmatrix} \eta & 0 \\ 0 & 2\eta \end{pmatrix} ,$$

which implies $\boldsymbol{\Gamma}(0, \eta) = \eta$. Then

$$\hat{W}_{k,j} = \boldsymbol{\Gamma}^{-1/2}(0, \hat{\eta}_k) V_{k,j}(0, \hat{\eta}_{k,j}) = \frac{\sum_{i=j}^k Y_i}{\sqrt{\frac{1}{k}\sum_{i=1}^k Y_i^2}} ,$$

and

$$T_k = \frac{1}{\sqrt{n_0}} \max_{1 \leq j < k} \hat{W}_{k,j} = \frac{1}{\sqrt{n_0}} \max_{1 \leq j < k} \frac{\sum_{i=j}^k Y_i}{\sqrt{\frac{1}{k}\sum_{i=1}^k Y_i^2}}.$$

b) Monitoring the variance of a normal distribution. Consider again $Y_1, Y_2, \ldots,$ a sequence of independent normal random variables with mean μ and variance σ^2, but now we are interested in testing

$$H_0 : \sigma_i^2 = \sigma_0^2 \ , \ \ \mu_i = \mu \text{ unknown}, \quad i = 1, 2, \ldots,$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

against the alternative

$$H_A : \begin{cases} \sigma_i^2 = \sigma_0^2 \ , \quad \mu_i = \mu \text{ unknown}, \quad i =< \tau, \\ \sigma_i{}^2 = \sigma_A^2 > \sigma_0^2 \ , \quad \mu_i = \mu \text{ unknown}, \quad i \geq \tau \ . \end{cases}$$

For simplicity we use the re-parametrization $\theta = \frac{-1}{2\sigma^2}$ and $\eta = \frac{-\mu}{\sigma^2}$. Again $d = p = 1$ and computations give $\hat{\eta}_k = -2\theta_0 \frac{\sum_{i=1}^k Y_i}{k}$, $\hat{\eta}_{k,j} = -2\theta_0 \frac{\sum_{i=j}^k Y_i}{k+1-j}$, and $V_{k,j}(\theta_0, \hat{\eta}_{k,j}) = \sum_{i=j}^k \left(Y_i^2 - \sigma_0^2 - \hat{\mu}_{k,j}^2\right)$. The information matrix is

$$\boldsymbol{I} = \begin{pmatrix} -\frac{\eta^2}{2\theta^3} + \frac{1}{2\theta^2} & \frac{\eta}{2\theta^2} \\ \frac{\eta}{2\theta^2} & -\frac{1}{2\theta} \end{pmatrix} .$$

Then

$$\hat{W}_{k,j} = \boldsymbol{\Gamma}^{-1/2}(\theta_0, \hat{\eta}_k) V_{k,j}(\theta_0, \hat{\eta}_{k,j}) = \sum_{i=j}^{k} \frac{\left(Y_i^2 - \sigma_0^2 - \hat{\mu}_{k,j}^2\right)}{\sigma_0^2\sqrt{2}} ,$$

where $\hat{\mu}_{k,j}^2 = \frac{\sum_{i=j}^k Y_i}{k-j+1}$. The test is based on

$$T_k = \frac{1}{\sqrt{n_0}} \max_{1 \leq j < k} \hat{W}_{k,j} = \frac{1}{\sqrt{n_0}} \max_{1 \leq j < k} \sum_{i=j}^{k} \frac{\left(Y_i^2 - \sigma_0^2 - \hat{\mu}_{k,j}^2\right)}{\sigma_0^2\sqrt{2}} .$$

2.3.2 The nested random effects model

In this subsection we consider the case of monitoring the parameters of a nested random effects model. We assume that observations come in as a sequence indexed by time i, and they have the following structure

$$X_{irn} = \mu + L_i + W_{r(i)} + E_{irn}, \qquad i = 1, 2, \ldots ,$$

$$r = 1, 2, \ldots, R \ , \quad \text{and} \quad n = 1, 2, \ldots, N \ ,$$

where we assume that $L_i \sim N(0, \sigma_b^2)$, $W_{r(i)} \sim N(0, \sigma_w^2)$, and $E_{irn} \sim N(0, \sigma^2)$ are independent normal random variables with corresponding variances as the second parameter. L_i is called factor A random effect, $W_{r(i)}$ nested random effect of factor B within the i^{th} level of factor A, while E_{irn} is the random

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

error term of observation n within the i^{th} level of factor A and the $r(i)^{th}$level of factor B.

Yashchin (1995) used this model to monitor electronic chip quality. In his example L_i corresponds to lot effect, $W_{r(i)}$ to wafer effect in wafer r of lot i, and E_{irn} is the random noise of the measurement of the n^{th} chip in wafer r of the lot i. Without loss of generality we can assume that R is the number of wafers randomly selected from each lot for the purpose of monitoring and let N be the size of a random sample chips taken from each wafer. In the above model σ_b, σ_w, and σ represent the lot-to-lot, wafer-to-wafer within-lot, and within-wafer components of variability. For each such a model, one will usually need to monitor not only the mean and the measure of total variance, but also its individual components. This is important for the following reasons. First, knowing which component of variance is out of control is important for diagnosing the problem, because different components are usually affected by different special causes . Second, the ability of screening procedures to improve the outgoing quality depends strongly on the individual variance components: for a fixed total variance, the higher the proportion of variance due to nested factors, the more difficult it is to screen out defective products (c.f., Yashchin 1995).

On the other hand, in cases where we deal with several sources of variability, the conventional control charts tend to produce an unacceptably high rate of false alarms and in general represent a rather weak diagnostic tool.

To solve the problem, Yashchin (1995) used the likelihood ratio in various forms to monitor change in the variance components, treating all other parameters of the model as nuisance. His approach has several drawbacks: the design, analysis, and implementation are relatively complex, and more important, the sensitivity is highly dependent on the levels of nuisance parameters. Atenafu and Gombay (2004) used the generalized likelihood ratio for the se-

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

quential procedures. Their procedure accommodates nuisance parameters but there is a decrease in power when the change is not at the beginning of the sequence.

Here we use the truncated CUSUM-type sequential test (2.14). From our simulations will be seen that the CUSUM-type sequential test is more powerful in detecting a change in the distribution.

Now we shall give the exact formulae of the test statistic T_k for monitoring the parameters indicated. Note that the extension of the monitoring process to the case of non-exponential family of distributions is needed when monitoring σ_w^2 and σ_b^2.

a) Monitoring the mean μ. The test statistic is based on the distribution of the estimator for the mean within i^{th} level of factor A

$$Y_i = \hat{\mu}_i = \frac{1}{RN} \sum_{r=1}^{R} \sum_{n=1}^{N} X_{irn} ,$$

which, under H_0 is normal with mean μ, and variance $\eta = \sigma_b^2 + \frac{\sigma_w^2}{R} + \frac{\sigma^2}{RN}$. The variance η is the nuisance parameter. When $\mu_0 = 0$, the maximum likelihood estimator is

$$\hat{\eta}_{k,j} = \frac{1}{k+1-j} \sum_{i=j}^{k} Y_i^2 ,$$

and

$$\hat{W}_{k,j} = \Gamma^{-1/2}(0, \hat{\eta}_k) V_{k,j}(0, \hat{\eta}_{k,j}) = \frac{1}{\sqrt{\frac{1}{k}\sum_{i=1}^{k} Y_i^2}} \sum_{i=j}^{k} Y_i .$$

The test is based on

$$T_k = \frac{1}{\sqrt{n_0}} \max_{1 \le j < k} \hat{W}_{k,j} = \frac{1}{\sqrt{n_0}} \max_{1 \le j < k} \frac{1}{\sqrt{\frac{1}{k}\sum_{i=1}^{k} Y_i^2}} \sum_{i=j}^{k} Y_i .$$

b) Monitoring σ^2. The estimator of σ^2 at the i^{th} level of factor A is $Z_i = \hat{\sigma}_i^2 = \frac{1}{R}\sum_{r=1}^{R} S_{ir}^2$ where $S_{ir}^2 = \frac{1}{N-1}\sum_{n=1}^{N}(X_{irn} - \bar{X}_{ir})^2$ and $\bar{X}_{ir}^2 =$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$\frac{1}{N}\sum_{n=1}^{N} X_{irn}$. $\frac{N-1}{\sigma^2}S_{ir}^2$ has a χ_{N-1}^2 distribution and S_{ir}^2 and S_{iq}^2 are independent if $r \neq q$, so the density of Z_i can be calculated by simple transformation methods from the chi-square distribution with $\nu_1 = R(N-1)$ degree of freedom. In this case there is no nuisance parameter, i.e. $p = 0$, and

$$\hat{W}_{k,j} = \mathbf{\Gamma}^{-1/2}(\sigma_0^2)V_{k,j}(\sigma_0^2) = \sqrt{\frac{\nu_1}{2}}\frac{1}{\sigma_0^2}\sum_{i=j}^{k}(Y_i - \sigma_0^2) \, .$$

The test is defined by

$$T_k = \frac{1}{\sqrt{n_0}}\max_{1\le j<k}\hat{W}_{k,j} = \sqrt{\frac{\nu_1}{2n_0}}\frac{1}{\sigma_0^2}\sum_{i=j}^{k}(Y_i - \sigma_0^2) \, .$$

c) Monitoring σ_w^2. Denote

$$U_i = \hat{\sigma}_{i.}^2 = \frac{1}{R-1}\sum_{r=1}^{R}(\bar{X}_{ir} - \hat{\mu}_i)^2 \, ,$$

which estimates $\sigma_w^2 + \frac{\sigma^2}{N}$ in the i^{th} lot, and $\hat{\mu}_i$, $\hat{\sigma}_i^2$ as above. Now we use the joint density of vector (Z_i, U_i) and the nuisance parameter is $\frac{\sigma^2}{N}$. With $\nu_2 = R - 1$, the computations give

$$\mathbf{\Gamma}(\sigma_{0w}^2, \hat{\eta}_k) = \frac{\nu_1\nu_2}{2\big[\nu_1(\hat{\eta}_k + \sigma_{0w}^2)^2 + \nu_2\hat{\eta}_k^2\big]} \, ,$$

$$V_{k,j}(\sigma_{0w}^2, \hat{\eta}_{k,j}) = \frac{\nu_2}{2(\hat{\eta}_{k,j} + \sigma_{0w}^2)^2}\sum_{i=j}^{k}\Big[U_i - (\hat{\eta}_{k,j} + \sigma_{0w}^2)\Big] \, ,$$

$$\hat{W}_{k,j} = \mathbf{\Gamma}^{-1/2}(\sigma_{0w}^2, \hat{\eta}_k)V_{k,j}(\sigma_{0w}^2, \hat{\eta}_{k,j}) \, .$$

Here $\hat{\eta}_k$ and $\hat{\eta}_{k-j}$ are the solution of two third-degree equations. The test is based on $T_k = \frac{1}{\sqrt{n_0}}\max_{1\le j<k}\hat{W}_{k,j}$.

d) Monitoring σ_b^2. Now we use the likelihood function of the bivariate statistics (Y_i, U_i) . The nuisance parameters are μ and $\xi = \sigma_w^2 + \frac{\sigma_w^2}{N}$. Under H_0 the maximum likelihood estimators are $\bar{\mu}_{k,j} = \frac{1}{k+1-j}\sum_{i=j}^{k} Y_i$, and $\hat{\xi}_{k,j}$ which

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

is the solution of a third-degree equation. In this case

$$\Gamma(\sigma_{0b}^2, \bar{\mu}_k, \hat{\xi}_k) = \frac{R^2(R-1)}{2\left[\xi_{0k}^2 + (R-1)(R\sigma_{0b}^2 + \hat{\xi}_k)^2\right]} ,$$

$$V_{k,j}(\sigma_{0b}^2, \bar{\mu}_{k,j}, \hat{\xi}_{k,j}) = \frac{R^2}{2(R\sigma_{0b}^2 + \hat{\xi}_{k,j})^2} \sum_{i=j}^{k} \left[(Y_i - \bar{\mu}_{k,j})^2 - \frac{R\sigma_{0b}^2 + \hat{\xi}_{k,j}}{R} \right] ,$$

$$\hat{W}_{k,j} = \Gamma^{-1/2}(\sigma_{0b}^2, \bar{\mu}_k, \hat{\xi}_k) V_{k,j}(\sigma_{0b}^2, \bar{\mu}_{k,j}, \hat{\xi}_{k,j}) ,$$

and we obtain: $T_k = \frac{1}{\sqrt{n_0}} \max_{1 \le j < k} \hat{W}_{k,j}$.

e) Simultaneous monitoring of the three variance components $\sigma^2, \sigma_w^2, \sigma_b^2$**.** In this case $d = 3$. The joint density of (Y_i, Z_i, U_i) can be used. The nuisance parameter is μ and the maximum likelihood estimator is $\hat{\mu}_{k,j} = \frac{1}{k+1-j} \sum_{i=j}^{k} Y_i$. Now $W_{k,j}$ has three independent components

$$\hat{W}_{k,j} = \Gamma^{-1/2}(\hat{\mu}_k, \sigma_{0b}^2, \sigma_{0w}^2, \sigma_0^2) V_{k,j}(\hat{\mu}_{k,j}, \sigma_{0b}^2, \sigma_{0w}^2, \sigma_0^2) = \begin{pmatrix} \sqrt{\frac{\nu_1}{2}} \frac{1}{\sigma_0^2} \sum_{i=j}^{k} (Z_i - \sigma_0^2) \\ \sqrt{\frac{\nu_2}{2}} \frac{1}{\xi_0} \sum_{i=j}^{k} (U_i - \xi_0) \\ \sum_{i=j}^{k} \frac{(Y_i^2 - \eta_0 - \hat{\mu}_{k,j}^2)}{\eta_0 \sqrt{2}} \end{pmatrix} .$$

The three independent components of the test are $T_k^i = \frac{1}{\sqrt{n_0}} \max_{1 \le j < k} \hat{W}_{k,j}^{(i)}$ for $i = 1, 2, 3$.

f) Simultaneous monitoring of all parameters. In this case $d = 4$ and there is no nuisance parameter. The four independent components of the efficient score vector are

$$\hat{W}_{k,j} = \Gamma^{-1/2}(\mu_0, \sigma_{0b}^2, \sigma_{0w}^2, \sigma_0^2) V_{k,j}(\mu_0, \sigma_{0b}^2, \sigma_{0w}^2, \sigma_0^2) = \begin{pmatrix} \sqrt{\frac{\nu_1}{2}} \frac{1}{\sigma_0^2} \sum_{i=j}^{k} (Z_i - \sigma_0^2) \\ \sqrt{\frac{\nu_2}{2}} \frac{1}{\xi_0} \sum_{i=j}^{k} (U_i - \xi_0) \\ \frac{1}{\eta_0 \sqrt{2}} \sum_{i=j}^{k} (Y_i^2 - \eta_0) \\ \frac{1}{\sqrt{\eta_0}} \sum_{i=j}^{k} Y_i \end{pmatrix} .$$

The four components of the test are $T_k^i = \frac{1}{\sqrt{n_0}} \max_{1 \le j < k} \hat{W}_{k,j}^{(i)}$, for $i = 1, 2, 3, 4$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

2.4 Simulation Studies

To assess the performance of the new truncated CUSUM-type sequential test we have carried out Monte Carlo experiments using the models described in the previous section. In all tables n_0 represents the truncation point, τ is the change point and the level of significance is $\alpha = 0.05$. Each scenario in these Monte Carlo simulations is based on 3000 replicates.

2.4.1 The case of normally distributed observations

In the case of monitoring the mean of a normal distribution, $\mu_0 = 0$ and μ_A was varied over the set $\{0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$. The variance (the nuisance parameter) was $\sigma^2 = 1$. Table 2.1 presents the simulated power and average stopping time(AVST) with its standard deviation(SD).

In the case of monitoring the variance of a normal distribution, $\sigma_0 = 1$ and σ_A was varied over the set $\{1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0\}$. The mean (the nuisance parameter in this case) was $\mu = 0$. Table 2.2 presents the simulated power and average stopping time(AVST) with its standard deviation(SD).

Tables 2.1 and 2.2 present three different situations $n_0 = 100$ and $\tau = 51$, $n_0 = 150$ and $\tau = 76$, $n_0 = 200$ and $\tau = 101$. These simulations suggest that the test procedure is consistent against fixed alternatives, a fact justified by Theorem 2.2. That is, as the truncation point n_0 grows to infinity, the power of the test goes to unity for any fixed alternative.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 2.1: Test for monitoring the mean μ of a normally distributed population with $\mu_0 = 0$ and $\sigma = 1$, and various μ_A. Simulated power (Power) and average stopping time (AVST) and its standard deviation(SD). The level of significance is $\alpha = 0.05$, and $C(\alpha) = 2.24$.

n_0	τ	μ_A	POWER	AVST	SD
100	51	0.00	0.039	99.24	0.53
		0.10	0.095	98.44	0.85
		0.20	0.219	96.81	1.26
		0.30	0.421	93.96	1.61
		0.40	0.641	90.13	1.84
		0.50	0.822	85.84	1.92
		0.60	0.944	81.61	1.82
		0.70	0.989	77.95	1.61
		0.80	0.998	75.00	1.37
		0.90	1.000	72.71	1.25
		1.00	1.000	70.87	1.09
150	76	0.00	0.044	148.85	1.25
		0.10	0.136	146.74	2.22
		0.20	0.312	142.73	2.88
		0.30	0.588	136.27	3.44
		0.40	0.833	127.99	3.63
		0.50	0.957	119.93	3.31
		0.60	0.994	113.37	2.80
		0.70	1.000	108.45	2.36
		0.80	1.000	104.79	1.98
		0.90	1.000	101.95	1.73
		1.00	1.000	99.68	1.52
200	101	0.00	0.041	198.30	1.31
		0.10	0.156	194.96	2.76
		0.20	0.396	187.68	4.68
		0.30	0.718	175.83	5.59
		0.40	0.928	162.48	5.35
		0.50	0.990	151.44	4.14
		0.60	0.999	143.56	3.26
		0.70	1.000	137.90	2.66
		0.80	1.000	133.66	2.26
		0.90	1.000	130.38	1.99
		1.00	1.000	127.80	1.82

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 2.2: Test for monitoring variance σ^2 of a normally distributed population with $\mu = 0$ and $\sigma_0 = 1$, and various σ_A. Simulated power (Power) and average stopping time (AVST) and standard deviation(SD) of AVST. The level of significance is $\alpha = 0.05$, and $C(\alpha) = 2.24$

n_0	τ	σ_A	POWER	AVST	SD
100	51	1.00	0.043	98.86	1.20
		1.10	0.180	96.65	1.82
		1.20	0.437	91.63	2.27
		1.30	0.720	84.98	2.62
		1.40	0.885	78.36	2.52
		1.50	0.969	73.20	2.06
		1.60	0.991	69.22	1.65
		1.70	0.998	66.26	1.32
		1.80	1.000	64.00	1.12
		1.90	1.000	62.25	1.04
		2.00	1.000	60.85	0.97
150	76	1.00	0.048	148.11	3.36
		1.10	0.237	143.51	3.76
		1.20	0.577	133.21	4.57
		1.30	0.852	120.79	4.62
		1.40	0.965	110.43	4.12
		1.50	0.993	103.03	3.34
		1.60	0.999	98.00	2.73
		1.70	1.000	94.34	2.41
		1.80	1.000	91.51	2.17
		1.90	1.000	89.44	1.95
		2.00	1.000	87.74	1.77
200	101	1.00	0.044	197.89	2.33
		1.10	0.275	190.14	4.34
		1.20	0.694	172.98	6.29
		1.30	0.929	153.87	6.26
		1.40	0.985	140.73	5.05
		1.50	0.999	131.80	4.05
		1.60	1.000	125.79	3.32
		1.70	1.000	121.55	2.84
		1.80	1.000	118.46	2.56
		1.90	1.000	116.06	2.18
		2.00	1.000	114.16	1.98

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

2.4.2 The case of the nested random effects model

In the case of the nested random effects model we first analyzed the Yashchin's (1995) published data (not presented in this thesis) on $n_0 = 30$ lots of chips manufactured at IBM. Each lot has two wafers, so $R = 2$, and from each wafer $N = 4$ chips were chosen randomly for measuring. The target value for mean is $\mu_0 = 1000\mathring{A}$; the historically acceptable variance components give $\sigma_0^2 = 400\mathring{A}$, $\sigma_{0w}^2 = 900\mathring{A}$, $\sigma_{0b}^2 = 3,600\mathring{A}$, respectively. We tested all four parameters for change individually and the tests for μ and for σ_w^2 were significant at $\alpha = 0.05$. When we monitor all parameters simultaneously the test is not significant at $\alpha = 0.05$. The reason for different conclusion from Yashchin (1995) is that his procedure was designed for a tolerance of $80\mathring{A}$ from the target whereas we can detect any size of change.

In the Tables 2.3-2.10 we present the simulated power (Power) and the average stopping time (AVST) of the CUSUM-type sequential test (2.14). The simulations were performed with level of significance $\alpha = 0.05$ on a model with $R = 2$ and $N = 4$. The first part of each of the tables 2.3 - 2.10 demonstrates the power of the test when change was at $\tau = 1$ and $n_0 = 30$, that is all observations come from the alternative distribution. The second part of each table demonstrates the power of the test when change was at $\tau = 31$ and $n_0 = 60$, that is, the first 30 observations come from the null distribution and the next 30 observations come from the alternative distribution. The in-control parameters were chosen to be $\mu = 0$, $\sigma = 0.2$, $\sigma_w = 0.3$, and $\sigma_b = 0.6$ as these give the same proportion as in the analyzed data of Yashchin(1995).

In Tables 2.3 to 2.6 we monitor only one parameter ($d = 1$) and the critical value is $C(0.05) = 2.24$. In tables 2.7, 2.8, and 2.9 we monitor three parameters simultaneously ($d = 3$) and the test statistic has three independent components. In this case, in order to obtain an overall level of significance of $\alpha = 0.05$ we monitor each component of the test statistic with a level of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

significance of

$$\alpha^* = 1 - (1-\alpha)^{1/d} = 1 - (1-0.05)^{1/3} = 0.0169 ,$$

(see *Remark 2.2*) and the critical value is $C(\alpha^* = 0.0169) = 2.632$. Similarly, in table 2.10 we monitor four parameters simultaneously ($d = 4$) and we monitor each component of the test statistic with a level of significance of

$$\alpha^* = 1 - (1-\alpha)^{1/d} = 1 - (1-0.05)^{1/4} = 0.0127 .$$

The critical value is now $C(\alpha^* = 0.0127) = 2.727$.

For comparison, in table 2.3 the entries in the brackets are the results of Atenafu and Gombay (2004) obtained by using the generalized likelihood ratio. We can see that the CUSUM-type sequential test has a higher power in the more realistic case where the first 30 observations come from the null hypothesis. This can be explained by the fact that the first 30 observation which come from the null hypothesis decrease the power of the test based on the generalized likelihood ratio but they do not affect in the same measure the power of the CUSUM-type test.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 2.3: Test for monitoring mean μ of a nested random effect model with $\mu_0 = 0$, $\sigma = 0.2$, $\sigma_w = 0.3$, and $\sigma_b = 0.6$, and various μ_A. Simulated power (Power) and average stopping time (AVST). The level of significance is $\alpha = 0.05$, $d = 1$, and $C(\alpha) = 2.24$.

n_0	τ	μ_A	POWER	AVST
30	1	0.0	0.034 (.049)	29.88 (29.7)
		0.1	0.130	29.38
		0.2	0.351 (.362)	28.07 (27.4)
		0.3	0.660	25.66
		0.4	0.885 (.887)	22.82 (21.0)
		0.5	0.980	20.26
		0.6	0.999 (.999)	18.37 (15.6)
		0.7	1.000	17.05
		0.8	1.000 (1.00)	16.15 (12.7)
		1.0	1.000 (1.00)	15.01 (11.1)
60	31	0.0	0.040 (.05)	59.57 (59.3)
		0.1	0.114	58.94
		0.2	0.273 (.18)	57.71 (58.4)
		0.3	0.524	55.66
		0.4	0.777 (.55)	52.81 (55.2)
		0.5	0.936	50.01
		0.6	0.989 (.87)	47.58 (51.1)
		0.7	0.999	45.75
		0.8	1.000 (.98)	44.35 (47.0)
		1.0	1.000 (1.0)	42.44 (44.6)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 2.4: Test for monitoring variance component σ^2 of a nested random effects model with $\mu = 0$, $\sigma_0 = 0.2$, $\sigma_w = 0.3$, and $\sigma_b = 0.6$, and various σ_A. Simulated power (Power) and average stopping time (AVST). The level of significance is $\alpha = 0.05$, $d = 1$, and $C(\alpha) = 2.24$.

n_0	τ	σ_A	POWER	AVST
30	1	0.20	0.037	29.78
		0.22	0.500	25.76
		0.25	0.986	14.68
		0.27	0.999	10.69
		0.30	1.000	7.61
60	31	0.20	0.034	59.55
		0.22	0.378	56.45
		0.25	0.962	46.11
		0.27	0.998	41.72
		0.30	1.000	38.26

Table 2.5: Test for monitoring variance component σ_w^2 of a nested random effects model with $\mu = 0$, $\sigma = 0.2$, $\sigma_{0w} = 0.3$, and $\sigma_b = 0.6$, and various σ_{Aw}. Simulated power (Power) and average stopping time (AVST). The level of significance is $\alpha = 0.05$, $d = 1$, and $C(\alpha) = 2.24$.

n_0	τ	σ_{Aw}	POWER	AVST
30	1	0.3	0.053	29.54
		0.4	0.674	21.77
		0.5	0.974	12.90
		0.6	0.999	8.54
		0.7	1.000	6.27
60	31	0.3	0.049	59.20
		0.4	0.569	53.05
		0.5	0.945	44.19
		0.6	0.997	39.30
		0.7	1.000	36.75

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 2.6: Test for monitoring variance component σ_b^2 of a nested random effects model with $\mu = 0$, $\sigma = 0.2$, $\sigma_w = 0.3$, and $\sigma_{0b} = 0.6$, and various σ_{Ab}. Simulated power (Power) and average stopping time (AVST). The level of significance is $\alpha = 0.05$, $d = 1$, and $C(\alpha) = 2.24$.

n_0	τ	σ_{Ab}	POWER	AVST
30	1	0.6	0.041	29.67
		0.7	0.263	27.67
		0.8	0.608	23.34
		1.0	0.948	15.11
		1.2	0.997	10.97
60	31	0.6	0.046	59.24
		0.7	0.195	57.66
		0.8	0.496	54.36
		1.0	0.919	45.59
		1.2	0.994	40.27

Table 2.7: Test for simultaneously monitoring all variance components σ^2, σ_w^2, and σ_b^2 of a nested random effects model with $\mu = 0$, $\sigma_0 = 0.2$, $\sigma_{0w} = 0.3$, and $\sigma_{0b} = 0.6$. Only σ is changing. Simulated power (Power) and average stopping time (AVST).The level of significance is $\alpha = 0.05$, $d = 3$ and $C(\alpha) = 2.632$.

n_0	τ	σ_A	POWER	AVST
30	1	0.20	0.055	29.66
		0.22	0.420	26.80
		0.25	0.975	15.76
		0.27	0.999	11.17
		0.30	1.000	7.60
60	31	0.20	0.054	59.30
		0.22	0.290	57.47
		0.25	0.933	47.76
		0.27	0.996	42.68
		0.30	1.000	38.59

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 2.8: Test for simultaneously monitoring all variance components σ^2, σ_w^2, and σ_b^2 of a nested random effects model with $\mu = 0$, $\sigma_0 = 0.2$, $\sigma_{0w} = 0.3$, and $\sigma_{0b} = 0.6$. Only σ_w is changing. Simulated power (Power) and average stopping time (AVST). The level of significance is $\alpha = 0.05$, $d = 3$ and $C(\alpha) = 2.632$.

n_0	τ	σ_{Aw}	POWER	AVST
30	1	0.3	0.055	29.66
		0.4	0.598	23.61
		0.5	0.958	14.52
		0.6	0.997	9.59
		0.7	1.000	7.00
60	31	0.3	0.054	59.30
		0.4	0.467	53.03
		0.5	0.909	46.28
		0.6	0.995	40.70
		0.7	1.000	37.73

Table 2.9: Test for simultaneously monitoring all variance components σ^2, σ_w^2, and σ_b^2 of a nested random effects model with $\mu = 0$, $\sigma_0 = 0.2$, $\sigma_{0w} = 0.3$, and $\sigma_{0b} = 0.6$. Only σ_b is changing. Simulated power (Power) and average stopping time (AVST). The level of significance is $\alpha = 0.05$, $d = 3$ and $C(\alpha) = 2.632$.

n_0	τ	σ_{Ab}	POWER	AVST
30	1	0.6	0.055	29.66
		0.7	0.205	28.43
		0.8	0.525	24.87
		0.9	0.793	20.25
		1.0	0.929	16.24
60	31	0.6	0.054	59.30
		0.7	0.150	58.36
		0.8	0.405	55.85
		0.9	0.700	51.75
		1.0	0.884	47.51

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 2.10: Test for simultaneously monitoring all four parameters of a nested random effects model with $\mu_0 = 0$, $\sigma_0 = 0.2$, $\sigma_{0w} = 0.3$, and $\sigma_{0b} = 0.6$. Only the mean is changing. Simulated power (Power) and average stopping time (AVST). The level of significance is $\alpha = 0.05$, $d = 4$ and $C(\alpha) = 2.727$.

n_0	τ	μ_A	POWER	AVST
30	1	0.0	0.056	29.64
		0.1	0.104	29.43
		0.2	0.257	28.45
		0.3	0.540	26.29
		0.4	0.816	22.92
		0.5	0.962	19.33
		0.6	0.995	16.29
		0.7	1.000	13.88
		0.8	1.000	11.96
60	31	0.0	0.057	59.27
		0.1	0.091	59.05
		0.2	0.185	58.38
		0.3	0.380	56.86
		0.4	0.648	54.41
		0.5	0.871	51.23
		0.6	0.971	48.11
		0.7	0.996	45.45
		0.8	1.000	43.22

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

2.5 Consistency of the test

In this section we will derive the asymptotic distribution of the test statistic (2.14) under the alternative hypothesis, H_A. To simplify discussion we assume that the observations come from the exponential family of distributions (Serfling (1980)), that is, the canonical form of the log-likelihood is

$$\log f(y;\theta,\eta) = T_1(y)\theta^t + T_2(y)\eta^t + S(y) - A(\theta,\eta)\,, \tag{2.15}$$

where $T_1(\cdot)$ and $T_2(\cdot)$ are vector valued functions of the data, $S(\cdot)$ is a real valued function of the data and $A(\theta,\eta)$ is a function of the parameters only. The superscript t denotes a vector or matrix transpose.

The following regularity conditions will be needed for the results of this section

C9. Vectors $\nabla_\theta A(\theta_0,\eta)$ and $\nabla_\eta A(\theta_0,\eta)$ exist, are continuous and have unique inverses that are Lipschitz continuous of order one in each argument.

C10. Matrix $\nabla^2_{\xi^2} A(\theta_0,\eta)$ exists, is positive definite and Lipschitz continuous of order one in each argument.

C11. $\frac{\partial^3}{\partial\xi_i\partial\xi_j\partial\xi_k}A(\xi)$, $1 \le i,j,k \le d+p$, exist and are bounded in a neighborhood of ξ.

C12. $E_\xi\|T_i(Y)\|^{2+\delta} < \infty$, for $i = 1, 2$, and some $\delta > 0$.

The following Theorem states that, under H_A, after the change, the score vector will drift away from a process with mean zero. In all our examples the drift is proportional to the number of observations after change. Here θ_A and η^* are the true values of θ and η respectively, after change.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Theorem 2.2 Under H_A, if C9-C12 are satisfied in an open neighborhood of the interval connecting (θ_0, η) and (θ_A, η^*), then for τ fixed, and any $k \geq \tau$, as $k \longrightarrow \infty$,

$$V_k = \sum_{i=1}^{k} \Big\{ \big[T_1(Y_i) - ET_1(Y_i)\big] - \big[T_2(Y_i) - ET_2(Y_i)\big] I_{22}^{-1}(\theta_0, \eta_{\tau k}) I_{21}(\theta_0, \eta_{\tau k}) \Big\}$$

$$+ k d_{\tau k} + O(\log\log k) \ , \tag{2.16}$$

where $\eta_{\tau k}$ is the solution of equation

$$\nabla_\eta A(\theta_0, \eta) = \frac{\tau - 1}{k} ET_2(Y_1) + \frac{(k - \tau + 1)}{k} ET_2(Y_\tau) \ , \tag{2.17}$$

and the drift component $d_{\tau k}$ is

$$d_{\tau k} = \frac{\tau - 1}{k} \nabla_\theta A(\theta_0, \eta) + \frac{(k - \tau + 1)}{k} \nabla_\theta A(\theta_A, \eta^*) - \nabla_\theta A(\theta_0, \eta_{\tau k}) \ . \tag{2.18}$$

Proof. Let $\eta_{\tau k}$ be the solution of the non-random equation (2.17). First we will prove that

$$\limsup_{k \longrightarrow \infty} \| \hat{\eta}_k - \eta_{\tau k} \| = O\big(\sqrt{k^{-1} \log\log k}\big) \quad a.s. \tag{2.19}$$

We can write

$$\nabla_\eta A(\theta_0, \hat{\eta}_k) - \nabla_\eta A(\theta_0, \eta_{\tau k}) = \frac{1}{k} \sum_{i=1}^{\tau - 1} \big[T_2(Y_i) - ET_2(Y_1)\big]$$
$$+ \frac{k - \tau + 1}{k} \frac{1}{k - \tau + 1} \sum_{i=\tau}^{k} \big[T_2(Y_i) - ET_2(Y_\tau)\big] \ ,$$

and by the LIL we have

$$\limsup_{k \longrightarrow \infty} \left\| \frac{1}{k - \tau + 1} \sum_{i=\tau}^{k} \big[T_2(Y_i) - ET_2(Y_\tau)\big] \right\| = O\left(\sqrt{\frac{\log\log(k - \tau + 1)}{(k - \tau + 1)}}\right) \quad a.s.$$

Then, from the last two relations we obtain

$$\limsup_{k \longrightarrow \infty} \left\| \nabla_\eta A(\theta_0, \hat{\eta}_k) - \nabla_\eta A(\theta_0, \eta_{\tau k}) \right\| = O\left(\sqrt{\frac{\log\log k}{k}}\right) \quad a.s.$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

By C9, $\nabla_\eta A(\theta_0;\cdot)$ has a unique inverse which is Lipschitz continuous of order one so from the above relation (2.19) follows immediately. We can write

$$\begin{aligned}
V_k(\theta_0,\hat{\eta}_k) &= \sum_{i=1}^{k} \nabla_\theta \log f(Y_i,\theta_0,\hat{\eta}_k) \\
&= \sum_{i=1}^{\tau-1} \left[T_1(Y_i) - \nabla_\theta A(\theta_0,\hat{\eta}_k)\right] + \sum_{i=\tau}^{k} \left[T_1(Y_i) - \nabla_\theta A(\theta_0,\hat{\eta}_k)\right] \\
&= \sum_{i=1}^{\tau-1} \left[T_1(Y_i) - \nabla_\theta A(\theta_0,\eta)\right] + \sum_{i=\tau}^{k} \left[T_1(Y_i) - \nabla_\theta A(\theta_A,\eta^*)\right] \\
&\quad + (\tau-1)\nabla_\theta A(\theta_0,\eta) + (k-\tau+1)\nabla_\theta A(\theta_A,\eta^*) - k\nabla_\theta A(\theta_0,\hat{\eta}_k)\,,
\end{aligned}$$

which can be written as

$$\begin{aligned}
V_k &= \sum_{i=1}^{\tau-1} \left[T_1(Y_i) - \nabla_\theta A(\theta_0,\eta)\right] + \sum_{i=\tau}^{k} \left[T_1(Y_i) - \nabla_\theta A(\theta_A,\eta^*)\right] \\
&\quad + (\tau-1)\nabla_\theta A(\theta_0,\eta) + (k-\tau+1)\nabla_\theta A(\theta_A,\eta^*) - k\nabla_\theta A(\theta_0,\eta_{\tau k}) \\
&\quad + k\left[\nabla_\theta A(\theta_0,\eta_{\tau k}) - \nabla_\theta A(\theta_0,\hat{\eta}_k)\right]\,. \qquad (2.20)
\end{aligned}$$

By C10, C11 and (2.19) the last term can be written as

$$\begin{aligned}
\nabla_\theta A(\theta_0,\eta_{\tau k}) - \nabla_\theta A(\theta_0,\hat{\eta}_k) &= (\eta_{\tau k} - \hat{\eta}_k)\nabla^2_{\theta\eta} A(\theta_0,\eta_{\tau k}) + O(\|\hat{\eta}_k - \eta_{\tau k}\|^2) \\
&= (\hat{\eta}_k - \eta_{\tau k}) I_{21}(\theta_0,\eta_{\tau k}) + O\left(k^{-1}\log\log k\right) \quad a.s.
\end{aligned} \qquad (2.21)$$

In order to obtain an expression for $\hat{\eta}_k - \eta_{\tau k}$ we use the following three-term Taylor expansion

$$\nabla_\eta A(\theta_0,\hat{\eta}_k) - \nabla_\eta A(\theta_0,\eta_{\tau k}) = (\hat{\eta}_k - \eta_{\tau k})\nabla^2_{\eta^2} A(\theta_0,\eta_{\tau k}) + O\left(k^{-1}\log\log k\right)\,,$$

and by the definition of $\hat{\eta}_k$ we have

$$\nabla_\eta A(\theta_0,\hat{\eta}_k) = \frac{1}{k}\sum_{i=1}^{k} T_2(Y_i).$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Now by using C11 we obtain

$$\hat{\eta}_k - \eta_{\tau k} = -\left\{\frac{1}{k}\sum_{i=1}^{k} T_2(Y_i) - \nabla_\eta A(\theta_0, \eta_{\tau k})\right\} I_{22}^{-1}(\theta_0, \eta_{\tau k}) + O\left(\sqrt{\frac{\log\log k}{k}}\right) . \tag{2.22}$$

From (2.21) and (2.22) we have

$$\nabla_\theta A(\theta_0, \eta_{\tau k}) - \nabla_\theta A(\theta_0, \hat{\eta}_k) = -\left\{\frac{1}{k}\sum_{i=1}^{k} T_2(Y_i) - \nabla_\eta A(\theta_0, \eta_{\tau k})\right\}$$
$$\times I_{22}^{-1}(\theta_0, \eta_{\tau k}) I_{21}(\theta_0, \eta_{\tau k}) + O\left(k^{-1}\log\log k\right) ,$$

and combine this result with (2.20) we get

$$\begin{aligned} V_k &= \sum_{i=1}^{\tau-1}\left[T_1(Y_i) - \nabla_\theta A(\theta_0, \eta)\right] + \sum_{i=\tau}^{k}\left[T_1(Y_i) - \nabla_\theta A(\theta_A, \eta^*)\right] \\ &+ \left[(\tau-1)\nabla_\theta A(\theta_0, \eta) + (k-\tau+1)\nabla_\theta A(\theta_A, \eta^*) - k\nabla_\theta A(\theta_0, \eta_{\tau k})\right] \\ &- \left\{\sum_{i=1}^{k} T_2(Y_i) - k\nabla_\eta A(\theta_0, \eta_{\tau k})\right\} I_{22}^{-1}(\theta_0, \eta_{\tau k}) I_{21}(\theta_0, \eta_{\tau k}) + O(\log\log k) . \end{aligned}$$

By using (2.17) and (2.18) we obtain

$$\begin{aligned} V_k &= \sum_{i=1}^{k}\left\{\left[T_1(Y_i) - ET_1(Y_i)\right] - \left[T_2(Y_i) - ET_2(Y_i)\right] I_{22}^{-1}(\theta_0, \eta_{\tau k}) I_{21}(\theta_0, \eta_{\tau k})\right\} \\ &+ kd_{\tau k} + O(\log\log k) . \end{aligned}$$

If we denote

$$Z_i = \left[T_1(Y_i) - ET_1(Y_i)\right] - \left[T_2(Y_i) - ET_2(Y_i)\right] I_{22}^{-1}(\theta_0, \eta_{\tau k}) I_{21}(\theta_0, \eta_{\tau k}) ,$$

then we can write

$$V_k = \sum_{i=1}^{k} Z_i + kd_{\tau k} + O(\log\log k) ,$$

Z_i are independent random vectors with mean zero and finite covariance structure.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

□

Remark 2.5 *The test (2.14) is based on* $V_{k,j}(\theta_0, \hat{\eta}_{k,j})$. *By Theorem 2.5.1* $V_{k,j}$ *can be represented as*

$$V_{k,j} = \sum_{i=j}^{k} Z_i + (k - j + 1)d_{\tau k}(j) + O(\log\log k) ,$$

where the drift $d_{\tau k}(j)$ *is defined as follows*

(i) *if* $\tau \le j < k$, *and* η_{1k} *is the solution of the non-random equation* $\nabla_\eta A(\theta_0, \eta) = ET_2(Y_\tau)$, *then*

$$d_{\tau k}(j) = \nabla_\theta A(\theta_A, \eta^*) - \nabla_\theta A(\theta_0, \eta_{1k}) .$$

(ii) *if* $j < \tau \le k$, *and* $\eta_{\tau k}(j)$ *is the solution of the non-random equation*

$$\nabla_\eta A(\theta_0, \eta) = \frac{\tau - j}{k - j + 1}\, ET_2(Y_1) + \frac{k - \tau + 1}{k - j + 1}\, ET_2(Y_\tau),$$

then

$$d_{\tau k}(j) = \frac{(\tau - j)\nabla_\theta A(\theta_0, \eta) + (k - \tau + 1)\nabla_\theta A(\theta_A, \eta^*)}{(k - j + 1)} - \nabla_\theta A(\theta_0, \eta_{\tau k}(j)) .$$

Example 2.1 We will compute now the drift in the test (2.14) in the case of monitoring the mean of a normal distribution (section 2.3.1). The test is based on

$$T_k = \frac{1}{\sqrt{n_0}} \max_{1 \le j < k} \frac{\sum_{i=j}^{k} Y_i}{\sqrt{\frac{1}{k}\sum_{i=1}^{k} Y_i^2}} .$$

It can be easily seen that under the alternative H_A, $\sigma^2_{\tau k} = \sigma^2 + \frac{k-\tau+1}{k}\mu_A^2$ and the drift in $V_{k,j}$ is

$$\begin{cases} d_{\tau k}(j) = (k - j + 1)\mu_A , & \text{if } \tau \le j \le k, \\ d_{\tau k}(j) = (k - \tau + 1)\mu_A , & \text{if } 2 \le j \le \tau \le k. \end{cases}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Then the maximum drift is obtained for any $j \leq \tau$ and for $j = \tau$ the drift in the test statistic is

$$D_{\tau k} = \frac{(k-\tau+1)\mu_A}{\sqrt{n_0\left(\sigma^2 + \frac{(k-\tau+1)}{k}\mu_A^2\right)}} .$$

Example 2.2 In the case of monitoring the variance of a normal distribution the test is based on (section 2.3.1)

$$T_k = \frac{1}{\sqrt{n_0}} \max_{1 \leq j < k} \sum_{i=j}^{k} \frac{\left(Y_i^2 - \sigma_0^2 - \hat{\mu}_{k-j}^2\right)}{\sigma_0^2\sqrt{2}} .$$

In this case $\mu_{\tau k} = \mu_{1k} = \mu$ and under the alternative, H_A, the drift in $V_{k,j}$ is

$$\begin{cases} d_{\tau k}(j) = (k-j+1)(\sigma_A^2 - \sigma_0^2), & \text{if } \tau \leq j \leq k , \\ d_{\tau k}(j) = (k-\tau+1)(\sigma_A^2 - \sigma_0^2), & \text{if } 2 \leq j \leq \tau \leq k . \end{cases}$$

Then the maximum drift is obtained for any $j \leq \tau$, and for $j = \tau$ the drift in the test statistic is

$$D_{\tau k} = \frac{(k-\tau+1)(\sigma_A^2 - \sigma_0^2)}{\sqrt{2n_0\sigma_0^2}} .$$

In the same way one can compute the drift when monitoring the parameters of a nested random effects.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Chapter 3

Monitoring Parameter Change in AR(p) Models

3.1 Preliminaries

Traditional Statistical Process Control (SPC) techniques are based on the assumption that process data are independent. As Wetherill (1977) has pointed out, observations from modern industrial processes are often autocorrelated and the process itself can behave like an autoregressive process. Such behavior must be taken into account when setting up monitoring procedures. So, for practical purposes the methods available for independent observations needed to be extended to the case of non-independent observations.

Johnson and Bagshaw (1975) obtained the limit processes for partial sums of observations from ARMA processes and explored the effect of ARMA noise on CUSUM statistics. Bagshaw and Johnson (1975) examined the effect of ARMA noise on the run length distribution for CUSUMs. Their method is based on the first passage distribution of a Wiener process moving between

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

reflecting and absorbing barriers, adjusting the variance parameter to allow for serial correlation, but their approximations can be inadequate in many cases (c.f., Yashchin 1993). Tang and MacNeill (1993) contains theoretical results and simulations on the effect of correlation.

Starting with Brown *et al.* (1975), residuals became one of the most important tools in change-point analysis for testing the constancy of parameters of a process over time. Kulperger (1985), Bai (1993), and Horváth (1993) investigated various asymptotic properties of sums of residuals. Boldin (1982) and Bai (1994b) obtained the weak convergence of empirical processes of residuals in stationary ARMA processes. Alwan and Roberts (1988) and Montgomery and Friedman (1989) discuss an alternative approach which consists of fitting a time series model to the $\{Y_i\}$ series when it is autocorrelated and then applying a control chart to the series of one-step-ahead prediction errors. In the above papers various tests for detecting a change in the parameters of a process have been suggested. Our simulations have been showed that a test based on residuals can be powerful in detecting a change in the mean of the process but is not too sensitive when we test for change in the coefficients of the process.

Several extensions of CUSUM and GLR schemes to handle autocorrelated observations have appeared. In principle Page's likelihood ratio CUSUM scheme (Page 1955) can be easily extended to non-independent observations, simply by replacing $f(Y_i)$ by $f(Y_i|Y_1, ..., Y_{i-1})$. However, according to Basseville and Nikiforov (1993), practical implementation of the GLR algorithm is not always possible because of computational difficulties.

Likelihood ratio methodology for testing for changes in the parameters of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

an autoregressive process was developed by Picard (1985), and later by Davis *et al.* (1995), and Csörgő and Horváth (1997).

As an alternative to likelihood ratio tests, a Bayes-type method was introduced by Chernoff and Zacks (1964) who applied it to the problem of one-sided changes at unknown times in the mean of a sequence of independent normal random variables. Lurie and Neerchal (1999) extended this method to the problem of testing for a change in autoregressive parameters for a general stationary AR process. In the presence of nuisance parameters the Bayes-type test statistics can be expressed in terms of certain Brownian integrals. Their simulation studies have shown that neither method is powerful.

In this chapter we focus on the change-point problems occurring in autocorrelated data. Truncated CUSUM-type sequential tests are proposed to detect an abrupt change in the parameters of a sequence of autocorrelated observations. The tests are based on large sample approximations of the components of efficient score vector. At the end of this chapter the empirical power of the proposed tests is analyzed in a simulation study.

To set up the problem let $Y_{-p+1}, Y_{-p+2}, \ldots, Y_0, Y_1, Y_2, \ldots,$ be consecutive observations from the model

$$Y_i - \mu = \phi_1(Y_{i-1} - \mu) + \ldots + \phi_p(Y_{i-p} - \mu) + \varepsilon_i, \quad i \geq -p+1\,, \tag{3.1}$$

where $\mu, \phi_j,\ j = 1, \ldots, p$, are constants and $\{\varepsilon_i\}$ is a sequence of random variables.

The assumptions on innovations ε_i vary in the literature. In the most simple case they are assumed to satisfy

$$\{\varepsilon_i\} \quad \text{is an i.i.d.} \quad N(0, \sigma^2) \quad \text{sequence}\,. \tag{3.2}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

For simplicity of exposition we shall work under this condition on the $\{\varepsilon_i\}$ sequence, but note that this is not a crucial assumption and our results will be valid if $\{\varepsilon_i\}$ were a martingale difference sequence, or some other sequence as long as the results of Eberlein (1986) are valid for the strong approximation of the corresponding sequence of partial sums by a Brownian motions. In the more general case the likelihood function will be replaced by a quasi-likelihood, and some moment conditions are specified.

We shall assume that the process is stationary, that is, the characteristic polynomial $\boldsymbol{\phi}(z) = 1 - \phi_1 z - \ldots - \phi_p z^p$ satisfies

$$\boldsymbol{\phi}(z) \neq 0 \text{ for all } |z| \leq 1 , \tag{3.3}$$

(c.f. Brockwell and Davis 1991).

The stochastic behavior of the the sequence is determined by the vector of parameters $(\mu, \sigma^2, \phi_1, ..., \phi_p)$. All the components of this vector may be of interest or we can deal with nuisance parameters. Here we consider the problem of detecting change in parameter θ from the initially given θ_0 value, where θ can be $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)$, or σ^2, or μ, or (μ, σ^2), with the other parameters unknown, that is, nuisance parameters.

We denote $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)^t$ the $p \times 1$ vector of the coefficients and assume that the covariance matrix

$$\boldsymbol{\Gamma} = \begin{pmatrix} c_0 & c_1 & \ldots & c_{p-1} \\ c_1 & c_0 & \ldots & c_{p-2} \\ \vdots & \vdots & & \vdots \\ c_{p-1} & c_{p-2} & \ldots & c_0 \end{pmatrix} , \tag{3.4}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

is invertible, where $c_k = Cov(Y_i, Y_{i+k})$. Let

$$\boldsymbol{X}_k = \begin{pmatrix} Y_0 - \mu & Y_{-1} - \mu & \dots & Y_{-p+1} - \mu \\ Y_1 - \mu & Y_0 - \mu & \dots & Y_{-p+2} - \mu \\ \vdots & \vdots & & \vdots \\ Y_{k-1} - \mu & Y_{k-2} - \mu & \dots & Y_{k-p} - \mu \end{pmatrix},$$

be the design matrix at stage k. It is well known that $(1/k)\mathbf{X}_k^t\mathbf{X}_k \xrightarrow{a.s.} \mathbf{\Gamma}$, hence it is invertible for k large enough.

We assume that the observations $Y_{-p+1}, Y_{-p+2}, \dots, Y_0, Y_1, \dots,$ come from the model (3.1). Under the assumption (3.2), for each $k \geq 2$, the joint distribution function of $(Y_1, \dots, Y_k)$ conditionally on $Y_{-p+1}, Y_{-p+2}, \dots, Y_0$ is

$$\begin{aligned} f(Y_1, \dots, Y_k) &= \prod_{i=1}^{k} f(Y_i | Y_{i-1}, \dots, Y_{i-p}) \\ &= (2\pi\sigma^2)^{-\frac{k}{2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{k} \left[(Y_i - \mu) - \sum_{j=1}^{p} \phi_j (Y_{i-j} - \mu) \right]^2 \right\}, \end{aligned} \tag{3.5}$$

and the log-likelihood function is given by

$$\begin{aligned} l_k(\mu, \sigma^2, \phi_1, \dots, \phi_p) &= -\frac{k}{2} \log(2\pi) - \frac{k}{2} \log(\sigma^2) \\ &\quad - \frac{1}{2\sigma^2} \sum_{i=1}^{k} \left[(Y_i - \mu) - \sum_{j=1}^{p} \phi_j (Y_{i-j} - \mu) \right]^2. \end{aligned} \tag{3.6}$$

Using the relationship $\varepsilon_i = (Y_i - \mu) - \sum_{j=1}^{p} \phi_j (Y_{i-j} - \mu)$, the components of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the efficient score vector are

$$\frac{\partial l_k}{\partial \mu} = \frac{(1-\phi_1-\ldots-\phi_p)}{\sigma^2}\sum_{i=1}^{k}\left[(Y_i-\mu)-\sum_{j=1}^{p}\phi_j(Y_{i-j}-\mu)\right]$$

$$= \frac{(1-\phi_1-\ldots-\phi_p)}{\sigma^2}\sum_{i=1}^{k}\varepsilon_i \,. \tag{3.7}$$

$$\frac{\partial l_k}{\partial \sigma^2} = -\frac{k}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{k}\left[(Y_i-\mu)-\sum_{j=1}^{p}\phi_j(Y_{i-j}-\mu)\right]^2$$

$$= \frac{1}{2\sigma^4}\sum_{i=1}^{k}\left(\varepsilon_i^2-\sigma^2\right) . \tag{3.8}$$

$$\frac{\partial l_k}{\partial \phi_s} = \frac{1}{\sigma^2}\sum_{i=1}^{k}(Y_{i-s}-\mu)\left[(Y_i-\mu)-\sum_{j=1}^{p}\phi_j(Y_{i-j}-\mu)\right]$$

$$= \frac{1}{\sigma^2}\sum_{i=1}^{k}\left(Y_{i-s}-\mu\right)\varepsilon_i \,, \quad \text{for all } \ s=1,\ldots,p\,. \tag{3.9}$$

The $(p+2)\times(p+2)$ information matrix is given by

$$\boldsymbol{I}(\mu,\sigma^2,\phi_1,\ldots,\phi_p) = \begin{pmatrix} \frac{(1-\phi_1-\ldots-\phi_p)^2}{\sigma^2} & 0 & \mathbf{0} \\ 0 & \frac{1}{2\sigma^4} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{1}{\sigma^2}\mathbf{\Gamma} \end{pmatrix},$$

where $\mathbf{\Gamma}$ is the covariance matrix given in (3.4).

The following estimators will be used

$$\hat{\mu}_k = \frac{1}{k}\sum_{i=1}^{k} Y_i \,, \tag{3.10}$$

$$\hat{\sigma}_k^2 = \frac{1}{k}\sum_{i=1}^{k}\left[(Y_i-\mu)-\sum_{j=1}^{p}\phi_j(Y_{i-j}-\mu)\right]^2 , \tag{3.11}$$

and

$$\hat{\boldsymbol{\phi}}_k = (\mathbf{X}_k^t\mathbf{X}_\mathbf{k})^{-1}\mathbf{X}_k^t\mathbf{Z}_k \,, \tag{3.12}$$

where $\mathbf{Z}_k = (Y_1-\mu,\ldots,Y_k-\mu)^t$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

We are testing whether parameter $\theta = \theta_0$ changes along the sequence, so only the nuisance parameters have to be estimated by their restricted m.l.e's, where value θ_0 is used in the formulas. The following lemmas on the maximum likelihood estimators give rates that allow the strong approximation of the statistics process by a Brownian motion. Invariance principles, strong approximations found in the literature (Phillips and Solo (1992), Konev and Pergamenshchikov (1997), and references therein) are not sufficient for our purposes, so although the limits are not new, the improved rates are. As our methods are based on large sample approximations, for large k, $Y_{-p+1}, Y_{-p+2}, \ldots, Y_0$, can be replaced by any random variable or constant without changing the limit.

Lemma 3.1 Under the hypothesis $\phi = \phi_0$ or $\sigma^2 = \sigma_0^2$ and conditions (3.2) and (3.3)

$$|\hat{\mu}_k - \mu| = O\left(\sqrt{k^{-1} \log\log k}\right) \qquad a.s. \tag{3.13}$$

Proof. The m.l.e. of μ is

$$\begin{aligned}
\bar{\mu}_k &= \frac{1}{k(1-\phi_{01} - \ldots - \phi_{0p})} \sum_{i=1}^{k} \left(Y_i - \phi_{01} Y_{i-1} - \ldots - \phi_{0p} Y_{i-p}\right) \\
&= \frac{1}{k(1-\phi_{01} - \ldots - \phi_{0p})} \sum_{i=1}^{k} \left[(Y_i - \mu) - \phi_{01}(Y_{i-1} - \mu) - \ldots - \phi_{0p}(Y_{i-p} - \mu)\right] \\
&+ \frac{k\mu}{k(1-\phi_{01} - \ldots - \phi_{0p})}(1 - \phi_{01} - \ldots - \phi_{0p}) \\
&= \mu + \frac{1}{(1-\phi_{01} - \ldots - \phi_{0p})} \frac{1}{k} \sum_{i=1}^{k} \varepsilon_i \,.
\end{aligned}$$

By the strong invariance principles for the innovation sequence, under our conditions, we get $k^{-1} \sum_{i=1}^{k} \varepsilon_i \overset{a.s.}{=} O\left(\sqrt{k^{-1} \log\log k}\right)$, and this implies

$$|\bar{\mu}_k - \mu| = O\left(\sqrt{k^{-1} \log\log k}\right) \qquad a.s.$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Note that

$$\bar{\mu}_k = \frac{1}{k(1-\phi_{01}-\ldots-\phi_{0p})}\left\{\sum_{i=1}^{k} Y_i - \phi_{01}\sum_{i=1}^{k} Y_{i-1} - \ldots - \phi_{0p}\sum_{i=1}^{k} Y_{i-p}\right\}$$
$$= \hat{\mu}_k + O\big(\sqrt{k^{-1}\log\log k}\big) \quad a.s.$$

Now putting together the last two relationships we obtain (3.13).

□

Lemma 3.2 Under the hypothesis $\mu = \mu_0$ or $\sigma^2 = \sigma_0^2$ and conditions (3.2) and (3.3)

$$\|\hat{\phi}_k - \phi\| = O\big(\sqrt{k^{-1}\log\log k}\big) \qquad a.s. \tag{3.14}$$

Proof. From $\hat{\phi}_k = (\mathbf{X}_k^t\mathbf{X}_\mathbf{k})^{-1}\mathbf{X}_k^t\mathbf{Z}_k$ using the description of the model in the form $\mathbf{Z}_k = \mathbf{X}_k\phi + \varepsilon$ we get $\hat{\phi}_k - \phi = (X_k^t X_k)^{-1}X_k^t\varepsilon$. By Lemma A.2 each component of the vector $\left(\frac{1}{k}\mathbf{X}_k^t\varepsilon\right)_{p\times 1}$ is of order $O(\sqrt{k^{-1}\log\log k})$ a.s. As $(1/k)\mathbf{X}_k^t\mathbf{X}_k \xrightarrow{a.s.} \mathbf{\Gamma}$, we have

$$\|\hat{\phi}_k - \phi\| = \left\|\left(\frac{1}{k}\mathbf{X}_k^t\mathbf{X}_\mathbf{k}\right)^{-1}\left(\frac{1}{k}\mathbf{X}_k^t\varepsilon\right)\right\| \stackrel{a.s.}{=} O\left(\sqrt{\frac{\log\log k}{k}}\right).$$

Using Lemma 3.1 we obtain that replacing μ by $\hat{\mu}_k$ (given by (3.10)) in $\mathbf{X}_k^t\varepsilon$ the error committed is $(\hat{\mu}_k - \mu)\sum_{i=1}^k \varepsilon_i \stackrel{a.s.}{=} O(\log\log k)$, which is negligible after the standardization. Similarly, it is easy to check by calculations, that $(1/k)\mathbf{X}_k^t\mathbf{X}_k\big|_{\mu=\hat{\mu}_k} \xrightarrow{a.s.} \mathbf{\Gamma}$.

□

Lemma 3.3 Under the hypothesis $\mu = \mu_0$ or $\phi = \phi_0$ and conditions (3.2) and (3.3)

$$|\hat{\sigma}_k^2 - \sigma^2| = O\big(\sqrt{k^{-1}\log\log k}\big) \qquad a.s. \tag{3.15}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Proof. The m.l.e. of σ^2 is

$$\hat{\sigma}_k^2 = \frac{1}{k}\sum_{i=1}^{k}\Big[(Y_i-\mu)-\phi_1(Y_{i-1}-\mu)-\ldots-\phi_p(Y_{i-p}-\mu)\Big]^2 ,$$

where either μ or ϕ has to be replaced by its m.l.e. Consider the first case.

$$\begin{aligned}\hat{\sigma}_k^2 &= \frac{1}{k}\sum_{i=1}^{k}\Big[(Y_i-\mu)-\sum_{j=1}^{p}\phi_j(Y_{i-j}-\mu)+(\mu-\hat{\mu}_k)(1-\sum_{j=1}^{p}\phi_j)\Big]^2\\ &= \frac{1}{k}\sum_{i=1}^{k}\varepsilon_i^2+(\mu-\hat{\mu}_k)^2(1-\sum_{j=1}^{p}\phi_j)^2+2(\mu-\hat{\mu}_k)(1-\sum_{j=1}^{p}\phi_j)\frac{1}{k}\sum_{i=1}^{k}\varepsilon_i .\end{aligned} \tag{3.16}$$

Under condition (3.2) the invariance principle holds for the first term and we have $|\sigma^2-k^{-1}\sum_{i=1}^{k}\varepsilon_i^2| \overset{a.s.}{=} O\big(\sqrt{k^{-1}\log\log k}\big)$. The second term is of order $O\big(k^{-1}\log\log k\big)$ by Lemma 3.1, while the invariance principle for the ε_i sequence and Lemma 3.1 make the last term of order $O\big(k^{-1}\log\log k\big)$. Putting these together the lemma is proved in this case.

Consider now the second case. If ϕ is replaced by $\hat{\phi}_k$ the variance σ^2 is estimated by

$$\begin{aligned}\hat{\sigma}_k^2 &= \frac{1}{k}\sum_{i=1}^{k}\Big\{\Big[(Y_i-\mu)-\sum_{j=1}^{p}\phi_j(Y_{i-j}-\mu)\Big]-\Big[\sum_{j=1}^{p}(\hat{\phi}_{kj}-\phi_j)(Y_{i-j}-\mu)\Big]\Big\}^2\\ &= \frac{1}{k}\sum_{i=1}^{k}\varepsilon_i^2-2\frac{1}{k}\sum_{i=1}^{k}\varepsilon_i\Big[\sum_{j=1}^{p}(\hat{\phi}_{kj}-\phi_j)(Y_{i-j}-\mu)\Big]\\ &\quad+\frac{1}{k}\sum_{i=1}^{k}\Big[\sum_{j=1}^{p}(\hat{\phi}_{kj}-\phi_j)(Y_{i-j}-\mu)\Big]^2\end{aligned} \tag{3.17}$$

The first term is as in (3.16) and we have $|\sigma^2-k^{-1}\sum_{i=1}^{k}\varepsilon_i^2| \overset{a.s.}{=} O\big(\sqrt{k^{-1}\log\log k}\big)$. For the second term

$$(\hat{\phi}_{k1}-\phi_1)\frac{1}{k}\sum_{i=1}^{k}\varepsilon_i(Y_{i-1}-\mu)+\ldots+(\hat{\phi}_{kp}-\phi_p)\frac{1}{k}\sum_{i=1}^{k}\varepsilon_i(Y_{i-p}-\mu) ,$$

we use Lemma 3.2 and the invariance principle for each sum of the form $\sum_{i=1}^{k}\varepsilon_i(Y_{i-s}-\mu)$, (see Lemma A.2), and we obtain a rate of $O\big(k^{-1}\log\log k\big)$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

For the last term in (3.17) we apply the Cauchy-Schwarz inequality and obtain

$$\frac{1}{k}\sum_{i=1}^{k}\left[\sum_{j=1}^{p}(\hat{\phi}_{kj}-\phi_j)(Y_{i-j}-\mu)\right]^2 \leq \frac{1}{k}\sum_{i=1}^{k}\left[p\sum_{j=1}^{p}(\hat{\phi}_{kj}-\phi_j)^2(Y_{i-j}-\mu)^2\right]$$
$$= p\sum_{j=1}^{p}(\hat{\phi}_{kj}-\phi_j)^2\left[\frac{1}{k}\sum_{i=1}^{k}(Y_{i-j}-\mu)^2\right].$$

Now, as $k^{-1}\sum_{i=1}^{k}(Y_{i-j}-\mu)^2$ converges almost surely to the diagonal element of the matrix $\mathbf{\Gamma}$, by using Lemma 3.2 we obtain again a rate of $O(k^{-1}\log\log k)$ a.s.

□

3.2 Monitoring the mean of an AR(p) process

In this section we consider the problem of detecting a change in the mean μ of an AR(p) model. σ^2 and ϕ are nuisance parameters. The hypothesis of interest is

$$H_0 : \mu = \mu_0\ ,\ \ \sigma^2,\ \ \phi \text{ unknown, for all } \ i \geq -p+1,$$

against the alternative

$$H_A : \begin{cases} \mu = \mu_0\ ,\ \ \sigma^2,\ \ \phi \text{ unknown, for } \ -p+1 \leq i < \tau, \\ \mu = \mu_A > \mu_0\ ,\ \ \sigma^2,\ \ \phi \text{ unknown, for } \ i \geq \tau\ , \end{cases}$$

where τ is the unknown change-point.

The test will be based on the first component of the efficient score vector given by (3.7). Under H_0 defined above, the efficient score vector is

$$\frac{\partial l_k}{\partial \mu}(\mu_0, \hat{\sigma}_k^2, \hat{\phi}_k) = \frac{(1-\hat{\phi}_{k1}-\ldots-\hat{\phi}_{kp})}{\hat{\sigma}_k^2}\sum_{i=1}^{k}\left[(Y_i-\mu_0)-\sum_{j=1}^{p}\hat{\phi}_{kj}(Y_{i-j}-\mu_0)\right].$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

For each $k \geq 2$ define

$$\hat{\varepsilon}_{ki} = (Y_i - \mu_0) - \sum_{j=1}^{p} \hat{\phi}_{kj}(Y_{i-j} - \mu_0) \,. \tag{3.18}$$

Then the standardized efficient score vector is given by

$$W_k(\mu_0, \hat{\sigma}_k^2, \hat{\phi}_k) = \frac{1}{\hat{\sigma}_k} \sum_{i=1}^{k} \hat{\varepsilon}_{ki} \,, \tag{3.19}$$

where $\hat{\boldsymbol{\phi}}_k$ is given by (3.12). In the following lemma we assume, without a loss of generality, the existence of a Brownian motion $W(\cdot)$ used in the approximations.

Lemma 3.4 Under the hypothesis $H_0 : \mu = \mu_0$, σ^2, $\boldsymbol{\phi}$ unknown, and conditions (3.2) and (3.3), there exists a Brownian motion $W(\cdot)$, such that

$$|\, W_k(\mu_0, \hat{\sigma}_k^2, \hat{\phi}_k) - W(k) \,| \overset{a.s.}{=} o(k^{1/\nu}) \tag{3.20}$$

for some $\nu > 2$.

Proof. The standardized efficient score vector can be written as

$$\begin{aligned} W_k(\mu_0, \hat{\sigma}_k^2, \hat{\phi}_k) &= \frac{1}{\hat{\sigma}_k} \sum_{i=1}^{k} \hat{\varepsilon}_{ki} = \frac{1}{\hat{\sigma}_k} \sum_{i=1}^{k} \left[(Y_i - \mu_0) - \sum_{j=1}^{p} \hat{\phi}_{kj}(Y_{i-j} - \mu_0) \right] \\ &= \frac{1}{\hat{\sigma}_k} \sum_{i=1}^{k} \left[(Y_i - \mu_0) - \sum_{j=1}^{p} \phi_j (Y_{i-j} - \mu_0) + \sum_{j=1}^{p} (\phi_j - \hat{\phi}_{kj})(Y_{i-j} - \mu_0) \right] \\ &= \frac{1}{\hat{\sigma}_k} \sum_{i=1}^{k} \varepsilon_i + \frac{1}{\hat{\sigma}_k} \sum_{j=1}^{p} \left[(\phi_j - \hat{\phi}_{kj}) \sum_{i=1}^{k} (Y_{i-j} - \mu_0) \right] \\ &= \frac{1}{\sigma} \sum_{i=1}^{k} \varepsilon_i + \frac{\sigma - \hat{\sigma}_k}{\hat{\sigma}_k \sigma} \sum_{i=1}^{k} \varepsilon_i + \frac{\sigma}{\hat{\sigma}_k} \frac{1}{\sigma} \sum_{j=1}^{p} \left[k(\phi_j - \hat{\phi}_{kj}) \frac{1}{k} \sum_{i=1}^{k} (Y_{i-j} - \mu_0) \right] \end{aligned} \tag{3.21}$$

By the invariance principle there exists a Brownian motion $W(\cdot)$, such that

$$\left| \frac{1}{\sigma} \sum_{i=1}^{k} \varepsilon_i - W(k) \right| \overset{a.s.}{=} o(k^{1/\nu}) \,,$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

for some $\nu > 2$. Replacing $\frac{1}{\hat{\sigma}_k}$ by $\frac{1}{\sigma}$ causes an error given by the second term of (3.21). By Lemma 3.3 and the invariance principle for the sequence $\sum_{i=1}^{k} \varepsilon_i$, the second term is of order $O(\sqrt{k^{-1} \log\log k}) \cdot O(\sqrt{k \log\log k}) = O(\log\log k)$ a.s. Multiplier $\frac{\sigma}{\hat{\sigma}_k} \overset{a.s.}{\to} 1$, and the sum of the error term is negligible by using Lemma 3.2 and recalling (see Lemma A.1) that the assumptions on the sequence $\{\varepsilon_i\}$ entails the invariance principle for $\sum_{i=1}^{k}(Y_i - \mu_0)$. Hence the error term is $O(\log\log k)$ a.s.

□

Remark 3.1 *Based on the above lemma it is easy to see that, as $n \to \infty$,*

$$\max_{1\le l<k\le n} \frac{1}{\hat{\sigma}_k \sqrt{n}} \sum_{i=l}^{k} \hat{\varepsilon}_{ki} = \max_{1\le l<k\le n} \frac{1}{\sqrt{n}} \left\{ W_k(\mu_0, \hat{\sigma}_k^2, \hat{\phi}_k) - W_l(\mu_0, \hat{\sigma}_k^2, \hat{\phi}_k) \right\}$$

$$\xrightarrow{\mathcal{D}} \sup_{0\le u<v\le 1} \left\{ W(v) - W(u) \right\},$$

which implies

$$\max_{1\le l<k\le n} \frac{1}{\hat{\sigma}_k \sqrt{n}} \sum_{i=l}^{k} \hat{\varepsilon}_{ki} \xrightarrow{\mathcal{D}} \sup_{0\le t\le 1} |W(t)| .$$

Now we are able to define a truncated CUSUM-type sequential test for detecting a change in the mean μ_0 as follows.

TEST 1. *Stop and conclude that H_0 is not supported by the data at the first k when*

$$T_\mu(k) = \max_{1\le l<k} \frac{1}{\hat{\sigma}_k \sqrt{n_0}} \sum_{i=l}^{k} \hat{\varepsilon}_{ki} \ge C(\alpha) . \tag{3.22}$$

Do not reject H_0 if it is not rejected by $k = n_0$.

Here n_0 is the truncation point, α is the level of significance and $\hat{\sigma}_k^2$ is the least square estimator of σ^2 given in (3.9)(with μ replaced by μ_0). The critical value $C(\alpha)$ is taken from the well known distribution of $\sup_{0\le t\le 1} |W(t)|$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

3.3 Monitoring the variance of an AR(p) process

In this section we consider the problem of detecting a change in the variance σ^2, while the mean μ and ϕ are nuisance parameters. We test

$$H_0 : \sigma^2 = \sigma_0^2 \ , \quad \mu \quad \phi \ \text{unknown}, \ \text{for all} \quad i \geq -p+1,$$

against the alternative

$$H_A : \begin{cases} \sigma^2 = \sigma_0^2 \ , \quad \mu, \quad \phi \ \text{unknown}, \ \text{for} \quad -p+1 \leq i < \tau, \\ \sigma^2 = \sigma_A^2 > \sigma_0^2 \ , \quad \mu, \quad \phi \ \text{unknown}, \ \text{for} \quad i \geq \tau \ , \end{cases}$$

where τ is the unknown change-point.

The test will be based on the second component of the efficient score vector given by (3.8), that is

$$\frac{\partial l_k}{\partial \sigma^2}(\sigma_0^2, \hat{\mu}_k, \hat{\phi}_k) = -\frac{k}{2\sigma_0^2} + \frac{1}{2\sigma_0^4} \sum_{i=1}^{k} \left[(Y_i - \hat{\mu}_k) - \sum_{j=1}^{p} \hat{\phi}_{kj}(Y_{i-j} - \hat{\mu}_k) \right]^2 .$$

By standardization we obtain

$$W_k(\sigma_0^2, \hat{\mu}_k, \hat{\phi}_k) = \frac{1}{\sqrt{2}\sigma_0^2} \sum_{i=1}^{k} \left(\hat{\varepsilon}_{ki}^2 - \sigma_0^2 \right) , \tag{3.23}$$

where now $\hat{\varepsilon}_{ki}$ are defined by

$$\hat{\varepsilon}_{ki} = (Y_i - \hat{\mu}_k) - \sum_{j=1}^{p} \hat{\phi}_{kj}(Y_{i-j} - \hat{\mu}_k) . \tag{3.24}$$

Here $\hat{\mu}_k$ and $\hat{\phi}_k$ are given by (3.10) and (3.12) respectively. In the following lemma we show that the standardized component of the efficient score vector can be approximated by a Brownian process $W(\cdot)$.

Lemma 3.5 Under the hypothesis $H_0 : \sigma^2 = \sigma_0^2$, μ, ϕ unknown, and conditions (3.2), (3.3), there exists a Brownian motion $W(\cdot)$, such that

$$|W_k(\sigma_0^2, \hat{\mu}_k, \hat{\phi}_k) - W(k)| \overset{a.s.}{=} o(k^{1/\nu}) \tag{3.25}$$

for some $\nu > 2$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Proof. Relationship (3.23) can be rewritten as

$$W_k = \frac{1}{\sqrt{2}\sigma_0^2}\sum_{i=1}^{k}\left(\hat{\varepsilon}_{ki}^2 - \sigma_0^2\right) = \frac{1}{\sqrt{2}\sigma_0^2}\sum_{i=1}^{k}\left(\varepsilon_i^2 - \sigma_0^2\right) + \frac{1}{\sqrt{2}\sigma_0^2}\sum_{i=1}^{k}\left(\hat{\varepsilon}_{ki}^2 - \varepsilon_i^2\right).$$

Under our conditions the invariance principle in the claim of the lemma holds for the first sum. Neglecting the coefficient $1/\sqrt{2}\sigma_0^2$, in the error term, we add $\pm\sum_{i=1}^{k}\left[(Y_i-\mu)-\sum_{j=1}^{p}\hat{\phi}_{kj}(Y_{i-j}-\mu)\right]^2$ and obtain

$$\sum_{i=1}^{k}\left(\hat{\varepsilon}_{ki}^2-\varepsilon_i^2\right) = \sum_{i=1}^{k}\left[(Y_i-\mu)-\sum_{j=1}^{p}\hat{\phi}_{kj}(Y_{i-j}-\mu)\right]^2 - \sum_{i=1}^{k}\varepsilon_i^2 +$$

$$+\sum_{i=1}^{k}\left[(Y_i-\hat{\mu}_k)-\sum_{j=1}^{p}\hat{\phi}_{kj}(Y_{i-j}-\hat{\mu}_k)\right]^2 - \sum_{i=1}^{k}\left[(Y_i-\mu)-\sum_{j=1}^{p}\hat{\phi}_{kj}(Y_{i-j}-\mu)\right]^2.$$

The difference between the first two terms can be written as

$$\sum_{i=1}^{k}\left[\varepsilon_i+\sum_{j=1}^{p}(\phi_j-\hat{\phi}_{kj})(Y_{i-j}-\mu)\right]^2 - \sum_{i=1}^{k}\varepsilon_i^2 =$$
$$=\sum_{i=1}^{k}\left[\sum_{j=1}^{p}(\phi_j-\hat{\phi}_{kj})(Y_{i-j}-\mu)\right]^2 + 2\sum_{i=1}^{k}\varepsilon_i\left[\sum_{j=1}^{p}(\phi_j-\hat{\phi}_{kj})(Y_{i-j}-\mu)\right]$$
$$=\sum_{i=1}^{k}\left[\sum_{j=1}^{p}(\phi_j-\hat{\phi}_{kj})(Y_{i-j}-\mu)\right]^2 + 2\sum_{j=1}^{p}(\phi_j-\hat{\phi}_{kj})\left[\sum_{i=1}^{k}\varepsilon_i(Y_{i-j}-\mu)\right],$$

and now, as in the last part of the proof of Lemma 3.3, one can see that the difference is of order $O(\log\log k)$.

The remaining terms

$$\sum_{i=1}^{k}\left[(Y_i-\hat{\mu}_k)-\sum_{j=1}^{p}\hat{\phi}_{kj}(Y_{i-j}-\hat{\mu}_k)\right]^2 - \sum_{i=1}^{k}\left[(Y_i-\mu)-\sum_{j=1}^{p}\hat{\phi}_{kj}(Y_{i-j}-\mu)\right]^2,$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

can be expanded as

$$\sum_{i=1}^{k}\left[(Y_i-\hat{\mu}_k)-\sum_{j=1}^{p}\phi_j(Y_{i-j}-\hat{\mu}_k)-\sum_{j=1}^{p}(\hat{\phi}_{kj}-\phi_j)(Y_{i-j}-\mu+\mu-\hat{\mu}_k)\right]^2$$

$$-\sum_{i=1}^{k}\left[\varepsilon_i-\sum_{j=1}^{p}(\hat{\phi}_{kj}-\phi_j)(Y_{i-j}-\mu)\right]^2$$

$$=\sum_{i=1}^{k}\left[(Y_i-\hat{\mu}_k)-\sum_{j=1}^{p}\phi_j(Y_{i-j}-\hat{\mu}_k)\right]^2-\sum_{i=1}^{k}\varepsilon_i^2$$

$$+\sum_{i=1}^{k}\left[\sum_{j=1}^{p}(\hat{\phi}_{kj}-\phi_j)(Y_{i-j}-\mu)+(\mu-\hat{\mu}_k)\sum_{j=1}^{p}(\hat{\phi}_{kj}-\phi_j)\right]^2$$

$$-\sum_{i=1}^{k}\left[\sum_{j=1}^{p}(\hat{\phi}_{kj}-\phi_j)(Y_{i-j}-\mu)\right]^2$$

$$-2\sum_{i=1}^{k}\left\{\left[(Y_i-\hat{\mu}_k)-\sum_{j=1}^{p}\phi_j(Y_{i-j}-\hat{\mu}_k)\right]\cdot\sum_{j=1}^{p}(\hat{\phi}_{kj}-\phi_j)(Y_{i-j}-\mu)\right\}$$

$$-2\sum_{i=1}^{k}\left\{\left[(Y_i-\hat{\mu}_k)-\sum_{j=1}^{p}\phi_j(Y_{i-j}-\hat{\mu}_k)\right](\hat{\mu}_k-\mu)\sum_{j=1}^{p}(\hat{\phi}_{kj}-\phi_j)\right\}$$

$$+2\sum_{i=1}^{k}\varepsilon_i\left[\sum_{j=1}^{p}(\hat{\phi}_{kj}-\phi_j)(Y_{i-j}-\mu)\right].$$

In Lemma 3.3 was proved that the difference between the first two terms is of order $O(\log\log k)$ a.s. As in Lemma 3.3 one can prove that the third and the fourth terms are of order $O(\log\log k)$ a.s. By Lemma A.2 the invariance principle holds for $\sum_{i=1}^{k}\varepsilon_i(Y_{i-j}-\mu)$, $j=1,\ldots,p$, from which one can see that the mixed terms with coefficient 2 are $O(\log\log k)$ a.s. by Lemmas 3.1 and 3.2. and the Lemma is proved.

□

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Based on the above lemma and reasoning as in Remark 3.1 we can define the following test.

TEST 2. *Stop and conclude that H_0 is not supported by the data at the first k when*

$$T_\sigma(k) = \max_{1 \le l < k} \frac{1}{\sigma_0^2 \sqrt{2n_0}} \sum_{i=l}^{k} \left(\hat{\varepsilon}_{ki}^2 - \sigma_0^2\right) \ge C(\alpha) \,. \tag{3.26}$$

Do not reject H_0 if it is not rejected by $k = n_0$.

Again, n_0 is the truncation point and α is the level of significance. Note that, here $\hat{\varepsilon}_{ki}$ are given by (3.24).

Remark 3.2 *If we want to monitor both μ and σ^2 simultaneously we can monitor both statistics $T_\mu(k)$(with $\hat{\sigma}_k^2$ replaced by σ_0^2, as this is assumed to be known) and $T_\sigma(k)$. Note that in this case μ is also known, and (3.18) and (3.24) are the same. Now we test*

$$H_0 : \mu = \mu_0,\ \sigma^2 = \sigma_0^2\ ,\quad \phi \ \textit{unknown},\ \textit{for all}\quad i \ge -p+1,$$

against the alternative

$$H_A : \begin{cases} \mu = \mu_0,\ \sigma^2 = \sigma_0^2\ ,\quad \phi \ \textit{unknown},\ \textit{for}\quad -p+1 \le i < \tau, \\ \mu = \mu_A,\ \sigma^2 = \sigma_A^2\ ,\quad \phi \ \textit{unknown},\ \textit{for}\quad i \ge \tau\ , \end{cases}$$

where τ is the unknown change-point. The test is defined as follows.

TEST 3. *Stop and conclude that H_0 is not supported by the data at the first k when*

$$\max\left\{T_\mu(k)\ ,\ T_\sigma(k)\right\} \ge C(\alpha^*) \,. \tag{3.27}$$

Do not reject H_0 if it is not rejected by $k = n_0$.

Here we monitor two parameters, so each test statistic is monitored with a level of significance of α^* such that the overall level of significance is $\alpha = 1 - (1 - \alpha^*)^2$ (see Remark 2.2). For example, if we want the overall level of significance to be $\alpha = 0.05$, we have to choose $\alpha^* = 0.0253$, and in this case the critical value is $C(\alpha^*) = 2.48$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

3.4 Monitoring the coefficients of an AR(p) process

In this section we shall consider the problem of testing for a change in the coefficients ϕ_i 's of an AR(p) process defined by (3.1). Now the mean μ and the variance σ^2 are nuisance parameters. The hypothesis of interest

$$H_0 : \phi = \phi_0, \ \ \mu, \ \sigma^2 \text{ unknown}, \ \text{ for all } \ \ i \geq -p+1,$$

is tested against the alternative

$$H_A : \begin{cases} \phi = \phi_0, \ \mu, \ \sigma^2 \text{ unknown}, \ \text{ for } \ -p+1 \leq i < \tau, \\ \phi = \phi_A, \ \mu, \ \sigma^2 \text{ unknown}, \ \text{ for } \ i \geq \tau \ , \end{cases}$$

where τ is the unknown change-point, and the new value ϕ_A is also unknown.

In this case the initial value ϕ_0 is assumed to be known and at each step $k \geq 2$, we shall estimate the nuisance parameters μ and σ^2 by

$$\hat{\mu}_k = \frac{1}{k}\sum_{i=1}^{k} Y_i \quad \text{and} \quad \hat{\sigma}_k^2 = \frac{1}{k}\sum_{i=1}^{k} \hat{\varepsilon}_{ki}^2 \ ,$$

respectively, where the residuals at step k, $\hat{\varepsilon}_{ki}$, are now given by

$$\hat{\varepsilon}_{ki} = \left(Y_i - \hat{\mu}_k\right) - \sum_{j=1}^{p} \phi_{0j}\left(Y_{i-j} - \hat{\mu}_k\right) . \tag{3.28}$$

For each $s = 1, 2, \ldots, p$, the component corresponding to ϕ_s is given by

$$V^{(s)} = \frac{1}{\sigma^2}\sum_{i=1}^{k}(Y_{i-s} - \mu)\varepsilon_i \ .$$

When we replace the unknown parameters by their estimators this becomes

$$\hat{V}^{(s)} = \frac{1}{\hat{\sigma}_k^2}\sum_{i=1}^{k}(Y_{i-s} - \hat{\mu}_k)\hat{\varepsilon}_{ki} \ ,$$

and standardizing this we have

$$W_k(\phi_0, \hat{\mu}_k, \hat{\sigma}_k^2) = \left(\frac{1}{\hat{\sigma}_k^2}\Gamma(\phi_0, \hat{\mu}_k, \hat{\sigma}_k^2)\right)^{-1/2} \nabla_\phi l(\phi_0, \hat{\mu}_k, \hat{\sigma}_k^2) \ .$$

The test is based on the following lemma.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Lemma 3.6 Under the hypothesis $H_0 : \phi = \phi_0$, μ, σ^2 unknown, and conditions (3.2) and (3.3), there exists a process $W(\cdot) = (W^{(1)}, \ldots, W^{(p)})$, with independent Brownian motion components, such that

$$\|W_k(\phi_0, \hat{\mu}_k, \hat{\sigma}_k^2) - W(k)\| \overset{a.s.}{=} o(k^{1/\nu}) \tag{3.29}$$

for some $\nu > 2$.

Proof. Let $W_k = W_k(\phi_0, \mu, \sigma^2)$ be the standardized efficient score vector evaluated at the true values of the parameters, and for each $r = 1, 2, \ldots, p$, let $W_k^{(r)}$ be the r^{th} component of W_k. Then $W_k^{(r)}$ can be written as

$$W_k^{(r)} = \gamma_{r1} V^{(1)} + \gamma_{r2} V^{(2)} + \ldots + \gamma_{rp} V^{(p)} , \tag{3.30}$$

where γ_{rj} are the components of the matrix $\left(\sigma^{-2}\mathbf{\Gamma}\right)^{-1/2}$. By using the definition of $V^{(s)}$ we can write

$$W_k^{(r)} = \sum_{t=1}^{k} \varepsilon_t \left\{ \sum_{s=1}^{p} \gamma_{rs} \frac{(Y_{t-s} - \mu)}{\sigma^2} \right\} = \sum_{t=1}^{k} \varepsilon_t X_{t-1}^{(r)} ,$$

where $X_t^{(r)} = \sum_{s=1}^{p} \gamma_{rs} \sigma^{-2} (Y_{t+1-s} - \mu)$. As in the proof of Lemma A.1, for each $t \in \mathbb{Z}$, $X_t^{(r)}$ can be written as $X_t^{(r)} = \sum_{i \geq 0} \tilde{\alpha}_i^{(r)} \varepsilon_{t-i}$, where the constants $\tilde{\alpha}_i^{(r)}$ satisfy the inequality $|\tilde{\alpha}_i^{(r)}| \leq \tilde{M}^{(r)} \alpha^i$, for some $\alpha \in (0, 1)$ and some $\tilde{M}^{(r)} \in (0, \infty)$. Arguing as in the proof of Lemma A.2 one can prove that Eberlein's theorem applies for each component $W_k^{(r)}$, $r = 1, 2, \ldots, p$. As the components are uncorrelated there exists a process $W(\cdot) = (W^{(1)}, \ldots, W^{(p)})$, with independent Brownian motion components, such that

$$\|W_k - W(k)\| \overset{a.s.}{=} o(k^{1/\nu}) ,$$

for some $\nu > 2$.

Now it is enough to prove that $\|W_k(\phi_0, \hat{\mu}_k, \hat{\sigma}_k^2) - W_k\| \overset{a.s.}{=} O(\log \log k)$, and this follows from (3.30) if we prove that $|\hat{V}^{(s)} - V^{(s)}| \overset{a.s.}{=} O(\log \log k)$, for

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

each $s = 1, 2, \ldots, p$. The difference $\hat{V}^{(s)} - V^{(s)}$ is written as

$$\hat{V}^{(s)} - V^{(s)} = \left(\frac{1}{\hat{\sigma}_k^2} - \frac{1}{\sigma^2}\right) \sum_{i=1}^{k} (Y_{i-s} - \mu)\varepsilon_i + \frac{\sigma^2}{\hat{\sigma}_k^2}\frac{1}{\sigma^2}(\mu - \hat{\mu}_k) \sum_{i=1}^{k} \varepsilon_i$$
$$+ \frac{\sigma^2}{\hat{\sigma}_k^2}\frac{1}{\sigma^2} \sum_{i=1}^{k} (Y_{i-s} - \hat{\mu}_k)\big(\hat{\varepsilon}_{ki} - \varepsilon_i\big) .$$

The rate is obtained by application of the invariance principle for $\sum_{i=1}^{k}(Y_{i-s} - \mu)\varepsilon_i$, (see Lemma A.2), and $\sum_{i=1}^{k} \varepsilon_i$ in the first and the second term of the above relationship.

For the sum in the last term we write

$$\sum_{i=1}^{k} (Y_{i-s} - \hat{\mu}_k)\big(\hat{\varepsilon}_{ki} - \varepsilon_i\big) =$$
$$= \sum_{i=1}^{k} (Y_{i-s} - \hat{\mu}_k)\left[(Y_i - \hat{\mu}_k) - (Y_i - \mu) - \sum_{j=1}^{p} \phi_{0j}(Y_{i-j} - \hat{\mu}_k) + \sum_{j=1}^{p} \phi_{0j}(Y_{i-j} - \mu)\right]$$
$$= \sum_{i=1}^{k} \left\{ (Y_{i-s} - \hat{\mu}_k)\big[(\mu - \hat{\mu}_k)(1 - \sum_{j=1}^{p} \phi_{0j})\big] \right\}$$
$$= \sum_{i=1}^{k} \left\{ \big[(Y_{i-s} - \mu) + (\mu - \hat{\mu}_k)\big]\big[(\mu - \hat{\mu}_k)(1 - \sum_{j=1}^{p} \phi_{0j})\big] \right\}$$
$$= k(\mu - \hat{\mu}_k)^2\big(1 - \sum_{j=1}^{p} \phi_{0j}\big) + (\mu - \hat{\mu}_k)\big(1 - \sum_{j=1}^{p} \phi_{0j}\big) \sum_{i=1}^{k} (Y_{i-s} - \mu) .$$

Now, by Lemma A.1, from the invariance principle for $\sum_{i=1}^{k}(Y_{i-s} - \mu)$ and Lemmas 3.1 and 3.2 we get rate $O(\log\log k)$ a.s. again and the proof is completed.

□

Based on the above lemma we can define a procedure useful in testing for a change in the coefficients ϕ_i 's. Denote $\hat{\boldsymbol{V}} = (\hat{V}^{(1)}, \ldots, \hat{V}^{(p)})^t$. Then the vector $\boldsymbol{T} = \big(\sigma^{-2}\boldsymbol{\Gamma}\big)^{-1/2}\hat{\boldsymbol{V}}$ has p independent components and we can monitor each component with a level of significance $\alpha^* = 1 - (1 - \alpha)^{1/p}$ and the overall level of significance will be α.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Remark 3.3 *Note that the matrix $\sigma^{-2}\mathbf{\Gamma}$ does not depend on μ or σ^2, but only on the vector of the coefficients ϕ which is assumed to be known in this case. This method can be applied for any $p \geq 1$ but for $p \geq 3$ the matrix $\left(\sigma^{-2}\mathbf{\Gamma}\right)^{-1/2}$ is complicated to compute symbolically. In any particular problem ϕ_i's are known and $\sigma^{-2}\mathbf{\Gamma}$ is a numerical matrix so $\left(\sigma^{-2}\mathbf{\Gamma}\right)^{-1/2}$ can be easily computed. For simplicity we shall consider in the next sections only the cases of $p = 1$ and $p = 2$ respectively.*

3.4.1 AR(1) process

Consider the model (3.1) with $p = 1$ and $\phi_1 = \phi$, that is

$$Y_i - \mu = \phi(Y_{i-1} - \mu) + \varepsilon_i, \quad i \geq 0\,. \tag{3.31}$$

In this case the stationarity condition (3.3) is equivalent to $|\phi| < 1$. From (3.9) the component of the efficient score vector corresponding to ϕ is given by

$$V^{(1)} = \frac{1}{\sigma^2}\sum_{i=1}^{k}(Y_{i-1} - \mu)\varepsilon_i\,,$$

and the information matrix is given by

$$\boldsymbol{I}(\phi, \mu, \sigma^2) = \begin{pmatrix} \frac{1}{1-\phi^2} & 0 & 0 \\ 0 & \frac{(1-\phi)^2}{\sigma^2} & 0 \\ 0 & 0 & \frac{1}{2\sigma^4} \end{pmatrix}, \qquad \text{and} \qquad \mathbf{\Gamma} = \frac{\sigma^2}{1-\phi^2}\,.$$

Based on Lemma 3.6, reasoning as in Remark 3.1, one can prove that, as $n \to \infty$

$$\max_{1 \leq l < k \leq n} \frac{\sqrt{1-\phi^2}}{\hat{\sigma}_k^2\sqrt{n}}\sum_{i=l}^{k}(Y_{i-1} - \hat{\mu}_k)\hat{\varepsilon}_{ki} \xrightarrow{\mathcal{D}} \sup_{0 \leq t \leq 1}|W(t)|\,,$$

where $\{W(t)\}_{0 \leq t \leq 1}$ is a standard Wiener process. The test for monitoring the coefficient ϕ is defined as follows.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

TEST 4. *Stop and conclude that H_0 is not supported by the data at the first k when*

$$T_\phi(k) = \max_{1 \le l < k} \frac{\sqrt{1-\phi_0^2}}{\hat{\sigma}_k^2 \sqrt{n_0}} \sum_{i=1}^{k} (Y_{i-1} - \hat{\mu}_k)\hat{\varepsilon}_{ki} \ge C(\alpha) \; . \tag{3.32}$$

Do not reject H_0 if it is not rejected by $k = n_0$, where n_0 is the truncation point and α is the level of significance of the test.

3.4.2 AR(2) process

Now consider the model (3.1) with $p = 2$, that is

$$Y_i - \mu = \phi_1(Y_{i-1} - \mu) + \phi_2(Y_{i-2} - \mu) + \varepsilon_i, \quad i \ge -1 \; . \tag{3.33}$$

When the observations come from an AR(2) process the stationarity condition implies that the parameters ϕ_1 and ϕ_2 must lie in the triangular region

$$\begin{cases} \phi_1 + \phi_2 < 1 \\ \phi_2 - \phi_1 < 1 \\ -1 < \phi_2 < 1 \end{cases}$$

The components of the efficient score vector corresponding to ϕ_1 and ϕ_2 are given by

$$V^{(1)} = \frac{\partial l_k}{\partial \phi_1} = \frac{1}{\sigma^2} \sum_{i=1}^{k} (Y_{i-1} - \mu)\varepsilon_i \; , \quad \text{and} \quad V^{(2)} = \frac{\partial l_k}{\partial \phi_2} = \frac{1}{\sigma^2} \sum_{i=1}^{k} (Y_{i-2} - \mu)\varepsilon_i \; ,$$

respectively. The information matrix is given by

$$\mathbf{I}(\phi_1, \phi_2, \mu, \sigma^2) = \begin{pmatrix} \frac{1-\phi_2}{(1+\phi_2)[(1-\phi_2)^2-\phi_1^2]} & \frac{\phi_1}{(1+\phi_2)[(1-\phi_2)^2-\phi_1^2]} & 0 & 0 \\ \frac{\phi_1}{(1+\phi_2)[(1-\phi_2)^2-\phi_1^2]} & \frac{1-\phi_2}{(1+\phi_2)[(1-\phi_2)^2-\phi_1^2]} & 0 & 0 \\ 0 & 0 & \frac{(1-\phi_1-\phi_2)^2}{\sigma^2} & 0 \\ 0 & 0 & 0 & \frac{1}{4\sigma^4} \end{pmatrix} .$$

In this case

$$\left(\frac{1}{\sigma^2}\mathbf{\Gamma}\right)^{-1/2} = \begin{pmatrix} \gamma_1 & \gamma_2 \\ \gamma_2 & \gamma_1 \end{pmatrix} ,$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

where

$$\gamma_1 = \frac{\sqrt{1+\phi_2}}{2}\left\{\sqrt{1-\phi_1-\phi_2}+\sqrt{1+\phi_1-\phi_2}\right\},$$

$$\gamma_2 = \frac{\sqrt{1+\phi_2}}{2}\left\{\sqrt{1-\phi_1-\phi_2}-\sqrt{1+\phi_1-\phi_2}\right\}.$$

We shall denote

$$\hat{V}_{kl}^{(1)}(\hat{\mu}_k,\hat{\sigma}_k^2) = \frac{1}{\hat{\sigma}_k^2}\sum_{i=l}^{k}(Y_{i-1}-\hat{\mu}_k)\hat{\varepsilon}_{ki},$$

$$\hat{V}_{kl}^{(2)}(\hat{\mu}_k,\hat{\sigma}_k^2) = \frac{1}{\hat{\sigma}_k^2}\sum_{i=l}^{k}(Y_{i-2}-\hat{\mu}_k)\hat{\varepsilon}_{ki},$$

for any $l < k$, where $\hat{\varepsilon}_{ki}$ are given in (3.28). Then for each $k \geq 2$ we define

$$T^{(1)}(k) = \max_{1\leq l<k}\left\{\gamma_1\hat{V}_{kl}^{(1)}+\gamma_2\hat{V}_{kl}^{(2)}\right\},$$

$$T^{(2)}(k) = \max_{1\leq l<k}\left\{\gamma_2\hat{V}_{kl}^{(1)}+\gamma_1\hat{V}_{kl}^{(2)}\right\}.$$

The two statistics are independent, and by Lemma 3.6, as $n \to \infty$,

$$\max_{2\leq k\leq n}\frac{1}{\sqrt{n}}T^{(i)}(k) \xrightarrow{\mathcal{D}} \sup_{0\leq t\leq 1}|W(t)| \quad \text{for} \quad i=1,2.$$

Now we can test

$$H_0: \phi_1=\phi_{10},\ \phi_2=\phi_{20},\ \ \mu\ ,\ \sigma^2 \ \text{unknown, for all} \ \ i\geq -p+1,$$

against the alternative

$$H_A: \begin{cases}\phi_1=\phi_{10},\ \phi_2=\phi_{20},\ \ \mu\ ,\ \sigma^2\ \text{unknown, for} \ \ -p+1\leq i<\tau,\\ \phi_1=\phi_{1A},\ \phi_2=\phi_{2A},\ \ \mu\ ,\ \sigma^2\ \text{unknown, for} \ \ i\geq\tau\ ,\end{cases}$$

by using the following procedure.

TEST 5. *Stop and conclude that H_0 is not supported by the data at the first k when*

$$T(k) = \frac{1}{\sqrt{n_0}}\max\left\{T^{(1)}(k),T^{(2)}(k)\right\} \geq C(\alpha^*)\,. \tag{3.34}$$

Fail to reject H_0 if it is not rejected by $k=n_0$, where n_0 is the truncation point and $\alpha^* = 1-(1-\alpha)^{1/2}$, where α is the level of significance of the test.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Remark 3.4 *In the case of independent observations Gombay (2003) developed two other tests based on $k^{-1/2}W_k$ and $n^{-1/2}W_k$ statistics, respectively. Lemmas 3.1-3.6 allow us to extend these two tests to the case of an AR(p) process. Our simulations (unreported here) showed that these tests outperform the CUSUM-type test only if the change is at the beginning, i.e. $\tau = 1$. As we are interested in the more realistic case $\tau > 1$, we have chosen to present only the CUSUM-type test.*

3.5 Simulation Studies

To evaluate the power of the truncated sequential tests proposed in the previous sections we have carried out some Monte Carlo experiments for each case discussed before.

3.5.1 The case of AR(1) process

First we consider the autoregressive model AR(1), that is

$$Y_i - \mu = \phi(Y_{i-1} - \mu) + \varepsilon_i, \quad i \geq 0 ,$$

where $|\phi| < 1$, and the errors ε_i are independent identically distributed normal random variables with mean zero and variance one. In all tables n_0 represents the truncation point, τ is the change-point, and the level of significance is $\alpha = 0.05$. Each table presents two different situations, $n_0 = 100$ with a change at $\tau = 50$, and $n_0 = 200$ with a change at $\tau = 100$, respectively. Each scenario in these simulations is based on 5,000 replicates.

a) Monitoring the mean μ . When testing for a change in the mean of the process we test the hypothesis

$$H_0 : \mu = \mu_0, \;\; \sigma^2 \text{ and } \phi \text{ unknown for all } i \geq 1 ,$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

against the alternative

$$H_A : \begin{cases} \mu = \mu_0, \ \ \sigma^2 \text{ and } \phi \ \text{ unknown for all } \ 1 \le i < \tau \ , \\ \mu = \mu_A > \mu_0, \ \ \sigma^2 \text{ and } \phi \ \text{ unknown for all } \ i \ge \tau \ . \end{cases}$$

The in-control value of the mean is $\mu_0 = 0$, and μ_A was varied over the set $\{0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$. Here σ^2 and ϕ are nuisance parameters. In each case the variance was $\sigma^2 = 1$. We present the empirical power of Test 1 for three different values of the coefficient, $\phi = 0.1$, $\phi = 0.5$, and $\phi = -0.5$. The test statistic is $T_\mu(k)$ defined in (3.22).

These simulations support that Test 1 is consistent against the change point alternative, that is, as the truncation point n_0 grows to infinity, the power of the test goes to unity for any change point alternative. From Table 3.1 we also note that, for fixed truncation point and fixed change point, the power of the test decreases as the coefficient ϕ increases from -1 to 1. This is easier seen in Figure 3.1 and it happens because the drift D is proportional with $(1 - \phi)$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 3.1: Test for monitoring the mean of an AR(1) model $Y_i - \mu_0 = \phi(Y_{i-1} - \mu_0) + \varepsilon_i$. Simulated power (Power) and average stopping time (AVST). $\sigma^2 = 1$ and the in-control mean is $\mu_0 = 0$. The level of significance is $\alpha = 0.05$.

			$\phi = 0.1$		$\phi = 0.5$		$\phi = -0.5$	
n_0	τ	μ_A	POWER	AVST	POWER	AVST	POWER	AVST
100	50	0.0	0.031	99.54	0.025	99.72	0.033	99.46
		0.1	0.070	99.06	0.038	99.56	0.134	98.23
		0.2	0.156	98.18	0.063	99.34	0.393	95.12
		0.3	0.290	96.70	0.095	99.00	0.725	89.83
		0.4	0.463	94.63	0.136	98.66	0.935	83.87
		0.5	0.652	92.15	0.186	98.24	0.992	78.95
		0.6	0.809	89.38	0.248	97.72	0.999	75.48
		0.7	0.916	86.83	0.313	97.18	1.000	73.07
		0.8	0.969	84.63	0.388	96.61	1.000	71.35
		0.9	0.988	82.90	0.455	96.06	1.000	70.11
		1.0	0.996	81.63	0.518	95.56	1.000	69.20
200	100	0.0	0.037	198.59	0.033	198.86	0.036	198.45
		0.1	0.118	196.40	0.061	198.03	0.243	192.67
		0.2	0.305	191.52	0.120	196.74	0.700	177.87
		0.3	0.578	183.42	0.204	194.71	0.968	159.41
		0.4	0.828	173.12	0.325	191.83	0.999	146.67
		0.5	0.957	163.28	0.455	188.44	1.000	138.80
		0.6	0.994	155.51	0.600	184.30	1.000	133.63
		0.7	1.000	149.91	0.739	179.93	1.000	130.04
		0.8	1.000	145.73	0.847	175.57	1.000	127.42
		0.9	1.000	142.65	0.917	171.61	1.000	125.44
		1.0	1.000	140.40	0.962	168.16	1.000	123.92

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

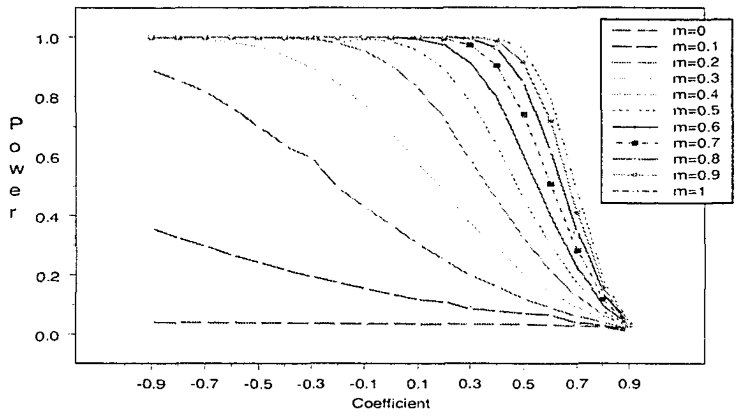

Figure 3.1: Power vs Coefficient when testing change in the mean of an AR(1) model. The truncation point is $n_0 = 200$ and the change point is $\tau = 100$. The in-control value is μ_0 and the drift after the change is $m = \mu_A - \mu_0$.

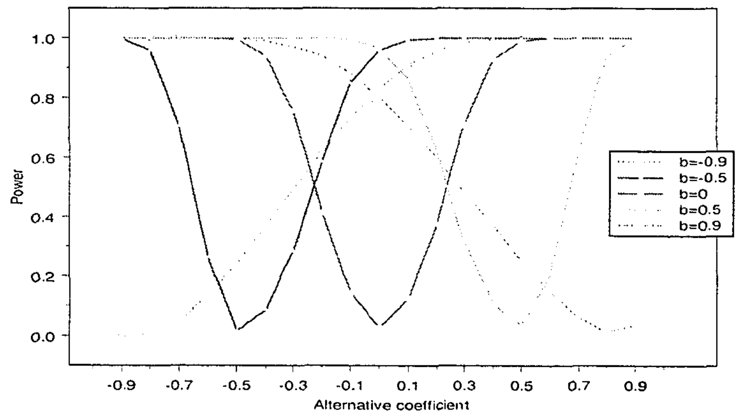

Figure 3.2: Power when testing for change in the coefficient for different AR(1) models ($b = \phi$). The truncation point is $n_0 = 200$ and the change point is $\tau = 100$. The initial coefficient value is -0.9, -0.5, 0, 0.5, 0.9.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

b) Monitoring the variance σ^2. We consider the same AR(1) model but now we shall test for a change in the variance σ^2, while the mean μ and the coefficient ϕ are nuisance parameters. The hypothesis of interest is

$$H_0 : \sigma^2 = \sigma_0^2 \ , \quad \mu \text{ and } \phi \ \text{ unknown for all } \ i \geq 1 \ ,$$

against the alternative

$$H_A : \begin{cases} \sigma^2 = \sigma_0^2 \ , & \mu \text{ and } \phi \ \text{ unknown for all } \ 1 \leq i < \tau \ , \\ \sigma^2 = \sigma_A^2 > \sigma_0^2 \ , & \mu \text{ and } \phi \ \text{ unknown for all } \ i \geq \tau \ . \end{cases}$$

The in control value of the variance is $\sigma_0 = 1$, and σ_A was varied over the set $\{1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0\}$. In Table 3.2 we present the power of Test 2 for two coefficient values, $\phi = 0.1$, and $\phi = 0.5$. In each case the mean of the process was $\mu = 0$. The test statistic is $T_\sigma(k)$ defined in (3.26).

These simulations support that Test 2 is consistent against the change point alternative. From Table 3.2 (and other simulations studies unreported here) we noted that, for fixed truncation point and fixed change point, the power of the test remains almost the same for any ϕ between -1 and 1.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 3.2: Test for monitoring σ of an AR(1) model $Y_i - \mu = \phi(Y_{i-1} - \mu) + \varepsilon_i$. Simulated power (Power) and average stopping time (AVST). $\mu = 0$ and the in-control value is $\sigma_0 = 1$. The level of significance is $\alpha = 0.05$.

			$\phi = 0.1$		$\phi = 0.5$	
n_0	τ	σ_A	POWER	AVST	POWER	AVST
100	50	1.0	0.046	98.77	0.046	98.85
		1.1	0.171	96.75	0.170	96.74
		1.2	0.433	91.91	0.433	91.11
		1.3	0.712	85.25	0.715	85.26
		1.4	0.897	78.45	0.899	78.47
		1.5	0.970	72.86	0.969	72.81
		1.6	0.990	68.77	0.990	68.77
		1.7	0.998	65.71	0.998	65.67
		1.8	1.000	63.36	1.000	63.35
200	100	1.0	0.043	198.03	0.043	198.00
		1.1	0.260	190.94	0.260	190.84
		1.2	0.674	173.95	0.673	154.30
		1.3	0.935	154.41	0.935	154.30
		1.4	0.993	140.42	0.994	140.26
		1.5	1.000	131.37	1.000	131.27
		1.6	1.000	125.34	1.000	125.25
		1.7	1.000	121.14	1.000	121.03
		1.8	1.000	117.95	1.000	117.83

c) Monitoring the coefficient ϕ. When monitoring the coefficient ϕ of the AR(1) model the hypotheses of interest are

$$H_0 : \phi = \phi_0 \ , \quad \mu \text{ and } \sigma^2 \text{ unknown for all } \ i \geq 1 \ ,$$

against the alternative

$$H_A : \begin{cases} \phi = \phi_0 \ , \quad \mu \text{ and } \sigma^2 \text{ unknown for all } \ 1 \leq i < \tau \ , \\ \phi = \phi_A > \phi_0 \ , \quad \mu \text{ and } \sigma^2 \text{ unknown for all } \ i \geq \tau \ . \end{cases}$$

Here μ and σ^2 are nuisance parameters. The test statistics is $T_\phi(k)$ defined in (3.32). In each case the mean was $\mu = 0$ and the variance was

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$\sigma^2 = 1$. We present three different situations. In Table 3.3, the in control value of the coefficient is $\phi_0 = 0.1$, and ϕ_A was varied over the set $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$. In Table 3.4, the in control value of the coefficient is $\phi_0 = -0.5$, and ϕ_A was varied over the set {-0.5, -0.4, -0.3, -0.2, -0.1, 0.0, 0.1, 0.2, 0.3, 0.4, 0.5}, and in Table 3.5, the in control value of the coefficient is $\phi_0 = 0.5$, and ϕ_A was varied over the set $\{0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$.

Again, these simulations support that Test 4 is consistent against the change-point alternative. From these tables we also note that, for fixed truncation point and fixed change point, the power of the test increases as the coefficient ϕ increases from -1 to 1. this can easily be seen in Figure 3.2 which presents the power of the test when the in-control value of the coefficient is $-0.9, -0.5, 0, 0.5, 0.9$.

Based on our simulation results (some unreported here) we can observe that the power of the tests depends on the coefficients ϕ_i's of the autoregressive process. For an AR(1) process we recommend to choose a truncation point $n_0 \geq 100$ when $|\phi| < 0.5$, and a truncation point $n_0 \geq 200$ when $0.5 \leq |\phi| < 0.7$. For a coefficient $0.7 \leq |\phi| \leq 0.9$ the truncation point around 1000 will provide a good power. This is in accordance with the recommendation of Tang and MacNeill(1993) in a study of effect of serial correlation on tests for parameter change in an AR(1) model.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 3.3: Test for monitoring the coefficient ϕ_0 of the AR(1) model $Y_i - \mu = \phi_0(Y_{i-1} - \mu) + \varepsilon_i$. Simulated power (Power) and average stopping time (AVST). $\phi_0 = 0.1$, $\mu = 0$, $\sigma^2 = 1$ and various ϕ_A. The level of significance is $\alpha = 0.05$ and $C(\alpha) = 2.24$.

n_0	τ	ϕ_A	POWER	AVST
100	50	0.1	0.025	99.54
		0.2	0.076	98.81
		0.3	0.207	97.19
		0.4	0.436	93.63
		0.5	0.692	88.08
		0.6	0.876	81.39
		0.7	0.967	75.05
		0.8	0.993	69.57
		0.9	0.999	65.12
		1.0	1.000	59.25
200	100	0.1	0.033	198.59
		0.2	0.135	195.87
		0.3	0.407	187.73
		0.4	0.754	173.04
		0.5	0.949	156.20
		0.6	0.995	142.07
		0.7	1.000	131.34
		0.8	1.000	123.35
		0.9	1.000	117.25
		1.0	1.000	112.71

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 3.4: Test for monitoring the coefficient ϕ_0 of the AR(1) model $Y_i - \mu = \phi_0(Y_{i-1} - \mu) + \varepsilon_i$. Simulated power (Power) and average stopping time (AVST). $\phi_0 = -0.5$, $\mu = 0$, $\sigma^2 = 1$ and various ϕ_A. The level of significance is $\alpha = 0.05$ and $C(\alpha) = 2.24$.

n_0	τ	ϕ_A	POWER	AVST
100	50	-0.5	0.010	99.89
		-0.4	0.039	99.60
		-0.3	0.111	99.03
		-0.2	0.232	97.88
		-0.1	0.418	95.86
		0.0	0.623	93.03
		0.1	0.798	89.51
		0.2	0.906	85.70
		0.3	0.966	82.05
		0.4	0.989	78.67
		0.5	0.998	75.47
200	100	-0.5	0.021	199.56
		-0.4	0.090	197.89
		-0.3	0.285	193.60
		-0.2	0.588	185.57
		-0.1	0.853	174.71
		0.0	0.959	164.23
		0.1	0.994	155.41
		0.2	0.999	148.17
		0.3	1.000	142.02
		0.4	1.000	136.63
		0.5	1.000	131.64

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 3.5: Test for monitoring the coefficient ϕ_0 of the AR(1) model $Y_i - \mu = \phi_0(Y_{i-1} - \mu) + \varepsilon_i$. Simulated power (Power) and average stopping time (AVST). $\phi_0 = 0.5$, $\mu = 0$, $\sigma^2 = 1$ and various ϕ_A. The level of significance is $\alpha = 0.05$ and $C(\alpha) = 2.24$.

n_0	τ	ϕ_A	POWER	AVST
100	50	0.5	0.032	99.31
		0.6	0.121	97.92
		0.7	0.366	93.65
		0.8	0.711	85.24
		0.9	0.927	75.13
		1.0	0.993	66.70
200	100	0.5	0.037	198.30
		0.6	0.210	192.79
		0.7	0.638	175.37
		0.8	0.943	150.20
		0.9	0.997	130.36
		1.0	1.000	118.44

3.5.2 The case of AR(2) process

Lai (1995) extended some classical sequential change-detection schemes to the AR(p) model. In this part of the simulation study we shall consider the problem of monitoring the coefficients of an AR(2) model and our test will be compared to Lai's algorithm. As in Lai's paper we consider an AR(2) model

$$Y_i - \mu = \phi_1(Y_{i-1} - \mu) + \phi_2(Y_{i-2} - \mu) + \varepsilon_i ,$$

with ε_i i.i.d. standard normal and bivariate normal (Y_0, Y_{-1}), such that $EY_0 = EY_{-1} = 0$, $Var(Y_0) = Var(Y_{-1}) = 1.0227$ and $Cov(Y_0, Y_{-1}) = -0.1136$.

We shall monitor a change in the coefficients ϕ_1 and ϕ_2. The in-control values are $\phi_{10} = -0.1$ and $\phi_{20} = 0.1$ and a change will occur only in ϕ_{10}. Note that this is known in Lai's algorithm while our test is monitoring both coefficients. Also note that $\sigma^2 = 1$ is a nuisance parameter in both methods,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

while the mean $\mu = 0$ is assumed to be known with Lai's method but it is a nuisance parameter in our test.

In Table 3.6 the CusumT columns give the empirical power and average stopping time of our test. The truncation point is $n_0 = 1000$ almost equal to the theoretical average stopping time of Lai's algorithm (i.e. 1006). The level of significance is $\alpha = 0.05$, the dimension is $d = 2$, so $\alpha^* = 0.0253$ and the critical value is $C = 2.48$.

The procedure in Lai (1995) is set up so that the average stopping time under the no-change hypothesis is around 1006. We analyzed this algorithm as follows. The columns *Total* and *Before* τ below the *POWER* column represent the proportion of stops (power) and the proportion of stops before the change-point, respectively. The columns *Total* and *Before* τ below the *AVST* column represent the average stopping time and the average stopping time when the algorithm stops before the change-point, respectively. The algorithm was stopped after 6,500 observations.

In Table 3.6 we present two different situations. In the first part of the table the change point is at $\tau = 70$ as in Lai's paper. In this case Lai's algorithm has a lower AVST and it only stops 4% of the time before the change-point which is an acceptable percent of false alarms if we consider a level of significance of 0.05. However, in real life we do not know the change point. In the second part of the table the change-point is at $\tau = 500$. Now the CUSUM test still has good power, while Lai's algorithm stops more then 35% of the time before the change point, i.e. too many false alarms. This behavior is common to all open ended monitoring schemes, where the expected stopping time under the null hypothesis controls the process parameters. We have a truncated-type algorithm, where the overall probability of type I error is under control. We pay for this advantage with increased delay in detection.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 3.6: Monitoring the coefficients ϕ_1 and ϕ_2 of the AR(2) model $Y_i - \mu = \phi_1(Y_{i-1}-\mu)+\phi_2(Y_{i-2}-\mu)+\varepsilon_i$. Simulated power (Power) and average stopping time (AVST). The in-control parameters are $\phi_{10} = -0.1$ and $\phi_{20} = 0.1$. Only ϕ_{10} is changing to ϕ_{1A}.

		POWER			AVST		
τ	ϕ_{1A}	Total	Before τ	CusumT	Total	Before τ	CusumT
70	-0.10	0.998	0.043	0.048	1134.69	42.48	990.40
	0.00	1.000	0.041	0.835	558.40	43.62	737.30
	0.10	1.000	0.040	1.000	261.21	43.70	422.60
	0.20	1.000	0.039	1.000	162.08	44.11	297.23
	0.30	1.000	0.041	1.000	122.91	40.03	230.31
	0.40	1.000	0.039	1.000	103.72	40.51	186.25
500	-0.10	1.000	0.358	0.042	1137.54	240.57	992.10
	0.00	1.000	0.365	0.477	744.43	238.36	934.23
	0.10	1.000	0.371	0.962	543.95	238.66	798.69
	0.20	1.000	0.370	1.000	470.93	239.93	700.20
	0.30	1.000	0.364	1.000	444.42	242.27	638.39
	0.40	1.000	0.360	1.000	432.17	244.31	601.25

3.6 Consistency of the test statistics

To understand the process under the alternative hypothesis of change we shall describe in detail the asymptotic behavior of the standardized efficient score vector in the case of an AR(1) model. In the case of a higher order process the conclusions are similar but will not be written out in detail because of unavoidable notational complexities.

We will consider only the cases of monitoring the mean and the coefficients of the process. In the case of monitoring the variance the conclusion is similar and much simpler to obtain so it will be omitted.

From Lemmas 3.7 and 3.8 we can see the consistency of Tests 1 and 4, respectively, provided $\frac{n_0-\tau}{\sqrt{n_0}} \rightarrow \infty$, as $n_0 \rightarrow \infty$, where τ is the unknown fixed time of change, and n_0 is the truncation point of the test. Consider first the case of monitoring the mean of the process.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Lemma 3.7 Assume conditions (3.1), (3.2), and (3.3) hold with $\mu = \mu_A$ for the sequence $(Y_k)_{k\in\mathbb{Z}}$. Then there exists a Brownian motion $W(\cdot)$, such that

$$\left|W_k(\mu_0, \hat{\sigma}_k^2, \hat{\phi}_k) - kD - \gamma W(k)\right| \overset{a.s.}{=} o(k^{1/\nu}) \tag{3.35}$$

for some $\nu > 2$, where γ and D are constants specified in the proof. The sign of drift D is the same as that of the difference $\mu_A - \mu_0$.

Proof. First, we describe the behavior of the restricted m.l.e.'s if $\mu = \mu_A$. In the AR(1) model

$$\hat{\phi}_k = \frac{\sum_{i=1}^k (Y_i - \mu_0)(Y_{i-1} - \mu_0)}{\sum_{i=1}^k (Y_{i-1} - \mu_0)^2} .$$

In the denominator we have

$$\begin{aligned}
\frac{1}{k}\sum_{i=1}^k (Y_{i-1} - \mu_0)^2 &= \frac{1}{k}\sum_{i=1}^k \left(Y_{i-1} - \mu_A + \mu_A - \mu_0\right)^2 \\
&= \frac{1}{k}\sum_{i=1}^k \left[(Y_{i-1} - \mu_A)^2 + (\mu_A - \mu_0)^2 + 2(Y_{i-1} - \mu_A)(\mu_A - \mu_0)\right] \\
&= \frac{\sigma^2}{1-\phi^2} + (\mu_A - \mu_0)^2 + O\left(\sqrt{k^{-1}\log\log k}\right) \qquad a.s.
\end{aligned}$$

using the invariance principle for $\sum_{i=1}^k (Y_{i-1} - \mu_A)$ (Lemma A.1). Similarly, the numerator can be approximated as

$$\begin{aligned}
&\frac{1}{k}\sum_{i=1}^k (Y_i - \mu_0)(Y_{i-1} - \mu_0) = \\
&= \frac{1}{k}\sum_{i=1}^k \left[(Y_{i-1} - \mu_A)(Y_i - \mu_A) + (Y_i - \mu_A)(\mu_A - \mu_0) + (Y_{i-1} - \mu_A)(\mu_A - \mu_0)\right] \\
&\quad + (\mu_A - \mu_0)^2 \\
&= \phi\frac{\sigma^2}{1-\phi^2} + (\mu_A - \mu_0)^2 + O\left(\sqrt{k^{-1}\log\log k}\right) \qquad a.s.
\end{aligned}$$

Then, with

$$\phi_{0A} = \frac{\phi\frac{\sigma^2}{1-\phi^2} + (\mu_A - \mu_0)^2}{\frac{\sigma^2}{1-\phi^2} + (\mu_A - \mu_0)^2} \quad \text{we have } \left|\hat{\phi}_k - \phi_{0A}\right| = O\left(\sqrt{k^{-1}\log\log k}\right) \ a.s.$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

For the variance estimator, as $\varepsilon_i = (Y_i - \mu_A) - \phi(Y_{i-1} - \mu_A)$, using the same principles, we have

$$\hat{\sigma}_k^2 = \frac{1}{k}\sum_{i=1}^{k}\left[(Y_i - \mu_0) - \hat{\phi}_k(Y_{i-1} - \mu_0)\right]^2$$
$$= \frac{1}{k}\sum_{i=1}^{k}\left[(Y_i - \mu_A) - \phi(Y_{i-1} - \mu_A) + (\phi - \hat{\phi}_k)(Y_{i-1} - \mu_A) + (1 - \hat{\phi}_k)(\mu_A - \mu_0)\right]^2$$
$$= \frac{1}{k}\sum_{i=1}^{k}\left[\varepsilon_i + (\phi - \hat{\phi}_k)(Y_{i-1} - \mu_A) + (1 - \hat{\phi}_k)(\mu_A - \mu_0)\right]^2$$
$$= \frac{1}{k}\sum_{i=1}^{k}\varepsilon_i^2 + \frac{(\phi - \hat{\phi}_k)^2}{k}\sum_{i=1}^{k}(Y_{i-1} - \mu_A)^2 + (\mu_A - \mu_0)^2(1 - \hat{\phi}_k)^2$$
$$+ 2\frac{(\phi - \hat{\phi}_k)}{k}\sum_{i=1}^{k}\varepsilon_i(Y_{i-1} - \mu_A) + 2\frac{(1 - \hat{\phi}_k)(\mu_A - \mu_0)}{k}\sum_{i=1}^{k}\varepsilon_i$$
$$+ 2(\mu_A - \mu_0)\frac{(\phi - \hat{\phi}_k)(1 - \hat{\phi}_k)}{k}\sum_{i=1}^{k}(Y_{i-1} - \mu_A)$$
$$= \sigma^2 + \frac{\sigma^2}{1 - \phi^2}(\phi - \phi_{0A})^2 + (\mu_A - \mu_0)^2(1 - \phi_{0A})^2 + O\left(\sqrt{k^{-1}\log\log k}\right) \qquad a.s.$$
$$= \sigma_{0A}^2 + O\left(\sqrt{k^{-1}\log\log k}\right) \qquad a.s.$$

where $\sigma_{0A}^2 = \sigma^2 + \frac{\sigma^2}{1-\phi^2}(\phi - \phi_{0A})^2 + (\mu_A - \mu_0)^2(1 - \phi_{0A})^2$.
The standardized efficient score vector is approximated as

$$W_k(\mu_0, \hat{\sigma}_k, \hat{\phi}_k) = \frac{1}{\hat{\sigma}_k}\sum_{i=1}^{k}\hat{\varepsilon}_{ki} = \frac{1}{\hat{\sigma}_k}\sum_{i=1}^{k}\left[(Y_i - \mu_0) - \hat{\phi}_k(Y_{i-1} - \mu_0)\right]$$
$$= \frac{1}{\hat{\sigma}_k}\sum_{i=1}^{k}\left[(Y_i - \mu_A) - \phi(Y_{i-1} - \mu_A) + (\phi - \hat{\phi}_k)(Y_{i-1} - \mu_A) + (1 - \hat{\phi}_k)(\mu_A - \mu_0)\right]$$
$$= \frac{1}{\hat{\sigma}_k}\sum_{i=1}^{k}\left[\varepsilon_i + (\phi - \phi_{0A})(Y_{i-1} - \mu_A)\right] + (\phi_{0A} - \hat{\phi}_k)\frac{1}{\hat{\sigma}_k}\sum_{i=1}^{k}(Y_{i-1} - \mu_A)$$
$$+ \frac{1}{\hat{\sigma}_k}\left[k(1 - \phi_{0A})(\mu_A - \mu_0) + k(\phi_{0A} - \hat{\phi}_k)(\mu_A - \mu_0)\right].$$

As in Lemma A.1 the first sum can be approximated by $\gamma W(k)$ with an $a.s.$ error of order $o(k^{1/\nu})$, where $\gamma^2 = \sigma_{0A}^{-2}[\sigma^2 + (\phi - \phi_{0A})^2\frac{\sigma^2}{1-\phi^2}]$, and $W(\cdot)$ is a

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

standard Brownian motion. The dominant term in the error is $k(\phi_{0A} - \hat{\phi}_k) \overset{a.s.}{=} O(\sqrt{k \log\log k})$, so we get for the drift of the process

$$D = \frac{(1-\phi_{0A})}{\sigma_{0A}}(\mu_A - \mu_0) + O\big(\sqrt{k^{-1}\log\log k}\big), \quad a.s.$$

per observation. Note that

$$1 - \phi_{0A} = \frac{(1-\phi)\sigma^2}{\sigma^2 + (1-\phi^2)(\mu_A - \mu_0)^2},$$

so the drift is proportional to $(1-\phi)$ and this explains the results obtained in the first part of our simulation study.

□

Remark 3.5 *In Lemma 3.7 we assumed that the change point is $\tau = 1$. When the change point is $\tau \geq 2$, in the same way one can prove that the drift is*

$$D = \frac{(1-\phi_{0A})}{\sigma_{0A}}(\mu_A - \mu_0) + O\big(\sqrt{(k-\tau)^{-1}\log\log(k-\tau)}\big) \quad a.s.$$

per observation after the change. So the drift in the standardized efficient score vector will be

$$(k-\tau)D = (k-\tau)\frac{(1-\phi_{0A})}{\sigma_{0A}}(\mu_A - \mu_0) + O\big(\sqrt{(k-\tau)\log\log(k-\tau)}\big) \quad a.s.$$

In the case of monitoring the coefficient ϕ the consistency of the test statistic follows from the next lemma.

Lemma 3.8 Assume conditions (3.1), (3.2), and (3.3) hold with $\phi = \phi_A$ for the sequence $(Y_k)_{k\in\mathbb{Z}}$. Then there exists a Brownian motion $W(\cdot)$, such that

$$\big|W_k(\phi_0, \hat{\mu}_k, \hat{\sigma}_k^2) - kD - \gamma W(k)\big| \overset{a.s.}{=} o(k^{1/\nu}) \tag{3.36}$$

for some $\nu > 2$, where γ and D are constants specified in the proof. The sign of drift D is the same as that of the difference $\phi_A - \phi_0$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Proof. As in Lemma 3.1 we can prove that $|\hat{\mu}_k - \mu| = O\big(\sqrt{k^{-1}\log\log k}\big)$. Note that under the alternative the errors are given by $\varepsilon_i = (Y_i - \mu) - \phi_A(Y_{i-1} - \mu)$, and we estimate them by $\hat{\varepsilon}_{ki} = (Y_i - \hat{\mu}_k) - \phi_0(Y_{i-1} - \hat{\mu}_k)$. For the variance estimator we can write

$$\begin{aligned}
\hat{\sigma}_k^2 &= \frac{1}{k}\sum_{i=1}^{k}\hat{\varepsilon}_{ki}^2 = \frac{1}{k}\sum_{i=1}^{k}\Big[(Y_i - \hat{\mu}_k) - \phi_0(Y_{i-1} - \hat{\mu}_k)\Big]^2 \\
&= \frac{1}{k}\sum_{i=1}^{k}\Big[(Y_i - \mu) + (\mu - \hat{\mu}_k) - \phi_A(Y_{i-1} - \mu) + \phi_A(Y_{i-1} - \mu) \\
&\quad - \phi_0(Y_{i-1} - \mu) + \phi_0(Y_{i-1} - \mu) - \phi_0(Y_{i-1} - \hat{\mu}_k)\Big]^2,
\end{aligned}$$

which gives

$$\begin{aligned}
\hat{\sigma}_k^2 &= \frac{1}{k}\sum_{i=1}^{k}\Big[\varepsilon_i + (\phi_A - \phi_0)(Y_{i-1} - \mu) + (1 - \phi_0)(\mu - \hat{\mu}_k)\Big]^2 \\
&= \frac{1}{k}\sum_{i=1}^{k}\varepsilon_i^2 + \frac{(\phi_A - \phi_0)^2}{k}\sum_{i=1}^{k}(Y_{i-1} - \mu)^2 + 2\frac{(\phi_A - \phi_0)}{k}\sum_{i=1}^{k}\varepsilon_i(Y_{i-1} - \mu) \\
&\quad + 2(1 - \phi_0)\left[\frac{(\mu - \hat{\mu}_k)}{k}\sum_{i=1}^{k}\varepsilon_i + (\phi_A - \phi_0)\frac{(\mu - \hat{\mu}_k)}{k}\sum_{i=1}^{k}(Y_{i-1} - \mu)\right] \\
&\quad + (1 - \phi_0)^2(\mu - \hat{\mu}_k)^2,
\end{aligned}$$

and using the same principles, we obtain

$$\begin{aligned}
\hat{\sigma}_k^2 &= \sigma^2 + (\phi_A - \phi_0)^2\frac{\sigma^2}{1 - \phi_A^2} + O\big(\sqrt{k^{-1}\log\log k}\big) \qquad a.s. \\
&= \sigma_{0A}^2 + O\big(\sqrt{k^{-1}\log\log k}\big) \qquad a.s.
\end{aligned}$$

where $\sigma_{0A}^2 = \sigma^2 + (\phi_A - \phi_0)^2\frac{\sigma^2}{1-\phi_A^2}$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The standardized efficient score vector is approximated as

$$
\begin{aligned}
W_k(\phi_0, \hat{\mu}_k, \hat{\sigma}_k^2) &= \frac{\sqrt{1-\phi_0^2}}{\hat{\sigma}_k^2} \sum_{i=1}^{k} (Y_{i-1} - \hat{\mu}_k)\hat{\varepsilon}_{ki} \\
&= \frac{\sqrt{1-\phi_0^2}}{\hat{\sigma}_k^2} \sum_{i=1}^{k} \left\{ \left[\varepsilon_i + (\phi_A - \phi_0)(Y_{i-1} - \mu) + (1-\phi_0)(\mu - \hat{\mu}_k) \right] \right. \\
&\quad \times \left. \left[(Y_{i-1} - \mu) + (\mu - \hat{\mu}_k) \right] \right\} \\
&= \frac{\sqrt{1-\phi_0^2}}{\hat{\sigma}_k^2} \sum_{i=1}^{k} \varepsilon_i (Y_{i-1} - \mu) + (\phi_A - \phi_0) \frac{\sqrt{1-\phi_0^2}}{\hat{\sigma}_k^2} \sum_{i=1}^{k} (Y_{i-1} - \mu)^2 \\
&\quad + \frac{\sqrt{1-\phi_0^2}}{\hat{\sigma}_k^2} (\mu - \hat{\mu}_k) \sum_{i=1}^{k} \varepsilon_i + k(\mu - \hat{\mu}_k)^2 (1-\phi_0) \frac{\sqrt{1-\phi_0^2}}{\hat{\sigma}_k^2} \\
&\quad + (1 - 2\phi_0 + \phi_A) \frac{\sqrt{1-\phi_0^2}}{\hat{\sigma}_k^2} (\mu - \hat{\mu}_k) \sum_{i=1}^{k} (Y_{i-1} - \mu).
\end{aligned}
$$

The first sum can be approximated by $\gamma W(k)$ with an *a.s.* error of order $o(k^{1/\nu})$ for some $\nu > 2$, where $\gamma^2 = \sigma_{0A}^{-4}(1-\phi_0^2)^{-1}[\sigma^2\phi_A(1-\phi_0)^2]$, and $W(\cdot)$ is a standard Brownian motion. The second term can be approximated by $k(\phi_A - \phi_0)\frac{\sigma^2\sqrt{1-\phi_0^2}}{\sigma_{0A}^2(1-\phi_A^2)}$ with an error of order $O\big(\sqrt{k \log\log k}\big)$. The last three terms produce an error of order $O\big(\log\log k\big)$. So, we get for the drift of the process

$$
D = (\phi_A - \phi_0)\frac{\sigma^2\sqrt{1-\phi_0^2}}{\sigma_{0A}^2(1-\phi_A^2)} + O\big(\sqrt{k^{-1}\log\log k}\big) \quad a.s.
$$

per observation. Replacing σ_{0A}^2 by its formulae the drift becomes

$$
D = \frac{(\phi_A - \phi_0)\sqrt{1-\phi_0^2}}{(1-\phi_A^2) + (\phi_A - \phi_0)^2} + O\big(\sqrt{k^{-1}\log\log k}\big) \quad a.s.
$$

per observation and the result of the lemma follows. Note that the drift is proportional with $(\phi_A - \phi_0)$ and this explains the results obtained in our simulation study.

□

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Remark 3.6 *In the above lemma we assumed that the change point is* $\tau = 1$. *When the change point is* $\tau \geq 2$, *in the same way one can prove that the drift is*

$$D = \frac{(\phi_A - \phi_0)\sqrt{1-\phi_0^2}}{(1-\phi_A^2) + (\phi_A - \phi_0)^2} + O\big(\sqrt{(k-\tau)^{-1}\log\log(k-\tau)}\big) \quad a.s.$$

for each observation after the change point. So the drift in the standardized efficient score vector will be

$$(k-\tau)D = (k-\tau)\frac{(\phi_A - \phi_0)\sqrt{1-\phi_0^2}}{(1-\phi_A^2) + (\phi_A - \phi_0)^2} + O\big(\sqrt{(k-\tau)\log\log(k-\tau)}\big) \quad a.s.$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Chapter 4

Sequential ANOVA

4.1 Preliminaries

This chapter is concerned with the problem of sequential comparison of three or more groups. The problem is frequently faced in clinical trials. Some work has already been done in the case of comparison of three groups. For more details we refer to Siegmund (1993) and Betensky (1996). The methods proposed are dealing with the simplest situation where the responses are independent, normally distributed random variables with a common known variance. The sequential F-test proposed by Siegmund (1980) deals with the case of more then three groups but the analytic approximations will be more complicated.

Here we compute the critical values for the tests of Gombay (2003a) and show how these tests can be used for this propose. These tests are compared to the sequential F-test in a simulation study. In the last section we shall show how the CUSUM test defined in the second chapter can be used for a sequential change-point ANOVA problem.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

We shall assume that observations are made sequentially on vectors $\mathbf{Y_k} = (Y_{1k}, Y_{2k}, \ldots, Y_{dk})^t$, $k \geq 1$, where d denotes the number of groups (treatments), and Y_{ik} is the k^{th} observation from group (treatment) i. The observations are assumed to be independently and normally distributed with common variance, i.e., for each $i = 1, 2, \ldots, d$,

$$Y_{ik} \ \text{iid} \ N(\mu_i, \sigma^2), \quad \text{for all} \quad k \geq 1 . \tag{4.1}$$

Considering σ^2 as a nuisance parameter, we are interested in testing

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_d, \quad \text{against} \quad H_A : \textit{not} \ H_0 . \tag{4.2}$$

4.2 Sequential F-test

The sequential F-test is proposed by Siegmund (1980) and it is based on the log likelihood ratio statistics. With the above notation let $\bar{Y}_{i.} = k^{-1} \sum_{j=1}^{k} Y_{ij}$ and $\bar{Y}_{..} = (kd)^{-1} \sum_{i=1}^{d} \sum_{j=1}^{k} Y_{ij}$ be the group sample means and the overall sample mean, respectively, based on k observations. The log likelihood ratio statistics for testing (4.2) is

$$L_k = \frac{kd}{2} \log \left\{ 1 + \frac{k \sum_{i=1}^{d} (\bar{Y}_{i.} - \bar{Y}_{..})^2}{\sum_{i=1}^{d} \sum_{j=1}^{k} (Y_{ij} - \bar{Y}_{i.})^2} \right\} . \tag{4.3}$$

It will be convenient to use the following parametrization. Define $\mu = d^{-1} \sum_{i=1}^{d} \mu_i$ and $\alpha_i = \mu - \mu_i$. Then $E(Y_{ik}) = \mu + \alpha_i$, with $\sum_{i=1}^{d} \alpha_i = 0$. Now the hypothesis of interest becomes

$$H_0 : \sum_{i=1}^{d} \alpha_i^2 = 0 , \quad \text{versus} \quad H_A : \sum_{i=1}^{d} \alpha_i^2 > 0 . \tag{4.4}$$

The test is defined as follows.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Sequential F-test. *Given integers* $k_0 < n_0$ *and constants* $0 < c \le a$, *stop sampling at* $\min(T, n_0)$, *where*

$$T = \inf\left\{k : k \ge k_0, L_k > a\right\}, \tag{4.5}$$

and reject H_0 *if either* $T \le n_0$ *or* $T > n_0$ *and* $L_{n_0} > c$.

According to Siegmund (1980) the power of the test depends on the parameters $\alpha_1, \ldots, \alpha_d$, μ, and σ^2 only through the value of $\delta = \sigma^{-1}(\sum_{i=1}^{d} \alpha_i^2)^{1/2}$. The power function of the test is defined by

$$P_\delta(T \le n_0) + P_\delta(T > n_0, L_{n_0} > c).$$

An exact value of the power function is very hard to obtain for $d \ge 3$. To obtain a level of significance α, the constants c and a are chosen based on the following approximation given in Siegmund (1980)

$$\alpha \approx P_0(L_{k_0} \ge a) + P_0(c < L_{n_0} \le a) + P_0(k_0 < T \le n_0). \tag{4.6}$$

The first two terms on the right-hand side of (4.6) may be obtained directly from tables of F distribution and the last term is approximated as follows

$$P_0(k_0 < T \le n_0) \approx 2\exp\{-a\}\left(\frac{a}{d}\right)^{\frac{d-1}{2}}\left\{\Gamma\left(\frac{d-1}{2}\right)\right\}^{-1}$$
$$\times \int_{l_1}^{l_2} x^{d-2} v_d(x)\sqrt{1+\frac{x^2}{d}}\left\{\log\left(1+\frac{x^2}{d}\right)\right\}^{\frac{1-d}{2}} dx,$$

where $\Gamma(\cdot)$ is the well known gamma function, $l_1 = \sqrt{d\{\exp(\frac{2a}{n_0 d}) - 1\}}$, and $l_2 = \sqrt{d\{\exp(\frac{2a}{k_0 d}) - 1\}}$. Because of computational difficulties the function $v_d(x)$ will be approximated by $\exp\{-0.583x(1 + x^2/d)^{-1}\}$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

4.3 Tests based on Rao's statistic

In Gombay (2002) two sequential tests are proposed to test the hypotheses

$$H_0 : \theta = \theta_0, \quad \eta \ \textit{unknown}, \quad \text{against} \quad H_A : \theta \neq \theta_0, \quad \eta \ \textit{unknown} \tag{4.7}$$

where $\theta \in \mathbb{R}^d$, $d \geq 1$, is the parameter of interest, and $\eta \in \mathbb{R}^p$, $p \geq 0$, is the nuisance parameter. It is assumed that the independent observations $Y_1, Y_2, \ldots,$ come sequentially from a distribution with density function $f(y; \theta, \eta)$. The efficient score vector is defined as

$$V_k(\xi) = \frac{1}{\sqrt{k}} \sum_{i=1}^{k} \nabla_\xi \log f(Y_i, \xi) \ , \tag{4.8}$$

where for brevity we denote $\xi = (\theta, \eta)$, and ∇_ξ denotes the vector of partial derivatives. Rao's statistic can be defined as

$$R_k(\xi) = V_k(\xi) \boldsymbol{I}^{-1}(\xi) V_k^t(\xi) \ ,$$

where $\boldsymbol{I}(\xi) = -E_\xi \big(\partial^2 / \partial\xi_i \partial\xi_j \log f(Y; \xi) \big)$ is the $(d+p) \times (d+p)$ information matrix. As before we shall partition this matrix based on the partition of parameter vector $\xi = (\theta, \eta)$ as

$$\boldsymbol{I} = \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix} .$$

The inverse of $\boldsymbol{I}$ will also be partitioned and denoted by

$$\boldsymbol{I}^{-1} = \begin{pmatrix} I^{11} & I^{12} \\ I^{21} & I^{22} \end{pmatrix} .$$

To test the hypotheses defined in (4.7), the nuisance parameter will be replaced with its maximum likelihood estimator under H_0, that is, by the solution of the equation

$$\sum_{i=1}^{k} \nabla_\eta \log f(Y_i; \theta_0, \eta) = 0 \ .$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Then the efficient score vector becomes

$$V_k(\theta, \hat{\eta}_k) = \frac{1}{\sqrt{k}} \sum_{i=1}^{k} \nabla_\theta \log f(Y_i; \theta_0, \hat{\eta}_k) \ ,$$

and Rao's statistics can be written as

$$R_k(\theta_0, \hat{\eta}_k) = V_k(\theta_0, \hat{\eta}_k) I^{11}(\theta_0, \hat{\eta}_k) V_k^t(\theta_0, \hat{\eta}_k) \ . \tag{4.9}$$

The following theorem and its corollary gives the asymptotics under H_0 for Rao's statistics process. For simplicity we shall assume that the observations come from a normal distribution.

Theorem 4.1 **(Gombay (2002))**. Under H_0, there exist independent Wiener processes $W_i(t)$, $i = 1, \ldots, d$, such that for $\zeta \leq 1/2 - 1/(2+\gamma)$

$$\sup_{1 \leq t < \infty} |V_{[nt]}(\theta_0, \hat{\eta}_{[nt]}) I^{11}(\theta_0, \eta) V^t_{[nt]}(\theta_0, \hat{\eta}_{[nt]}) - R^{(d)}(nt)| = O\big(n^{-\zeta} \sqrt{\log \log n}\big) \ \ a.s.$$

where $R^{(d)}(t) = t^{-1} \sum_{i=1}^{d} W_i^2(t)$ and $\gamma > 0$.

Note that, when η is replaced by $\hat{\eta}_k$ in $I^{11}(\theta_0, \eta)$ the asymptotic limit remains the same.

Corollary 4.1 **(Gombay (2002))**. Under conditions of Theorem 4.1

$$\text{(i)} \lim_{n \to \infty} P\Big\{ \max_{1<k\leq n} \big[V_k(\theta_0, \hat{\eta}_k) I^{11}(\theta_0, \hat{\eta}_k) V_k^t(\theta_0, \hat{\eta}_k)\big]^{1/2} \leq \frac{t + b_d(\log n)}{a(\log n)} \Big\}$$
$$= \exp(-e^{-t}) \ \ a.s. \ ,$$

where $a(t) = (2 \log t)^{1/2}$, $b_d(t) = 2 \log t + (d/2) \log \log t - \log \Gamma(d/2)$, and $\Gamma(t) = \int_0^\infty y^{t-1} e^{-y} dy, \ \ 1 \leq t < \infty$.

(ii) As $n \longrightarrow \infty$,

$$\max_{1<k\leq n} \big[(k/n) V_k(\theta_0, \hat{\eta}_k) I^{11}(\theta_0, \hat{\eta}_k) V_k^t(\theta_0, \hat{\eta}_k)\big]^{1/2} \xrightarrow{\mathcal{D}} \sup_{0 \leq t \leq 1} \Big(\sum_{i=1}^{d} W_i^2(t) \Big)^{1/2} .$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Based on Corollary 4.1 the following truncated sequential tests are defined.

TEST 1. *Stop and conclude that H_0 is not supported by the data at the first k when*

$$T_1(k) = \sqrt{\frac{k}{n_0} R_k(\theta_0, \hat{\eta}_k)} > C_1(\alpha, d) \; .$$

Fail to reject H_0 if it is not rejected by $k = n_0$.

Here n_0 is the truncation point. In the next subsection we will show how to compute the critical values $C_1(\alpha, d)$ from the distribution of the maximum of the Bessel process.

TEST 2. *Stop and conclude that H_0 is not supported by the data at the first k when*

$$T_2(k) = \sqrt{R_k(\theta_0, \hat{\eta}_k)} > C_2(\alpha, d, n_0) \; .$$

Fail to reject H_0 if it is not rejected by $k = n_0$.

Here the critical value is given by

$$C_2(\alpha, d, n_0) = (a(\log n_0))^{-1}\Big[- \log\big(- \frac{1}{2} \log(1-\alpha)\big) + b_d(\log n_0)\Big] , \quad (4.10)$$

where $a(\cdot)$ and $b_d(\cdot)$ are defined in Corollary 4.1. A discussion about the critical values is given in the next subsection.

4.3.1 Critical values for Test 1

The critical values for Test 1 can be obtained from the distribution of the maximum of the Bessel process. Denote $\nu = d/2 - 1$. According to Borodin and Salminen (1996) for $0 \leq x \leq y$ we have the following formulae

$$P_x\big(\sup_{0 \leq s \leq t} R^{(d)}(s) \geq y \big) = 1 - \sum_{k=1}^{\infty} \frac{2x^{-\nu} J_\nu(j_{\nu,k} x/y)}{y^{-\nu} j_{\nu,k} J_{\nu+1}(j_{\nu,k})} \exp\{ - \frac{j_{\nu,k}^2 t}{2y^2} \} , \quad (4.11)$$

where $J_\nu(x)$ is the Bessel function defined by

$$J_\nu(x) = \sum_{k=0}^{\infty} \frac{(-1)^k (x/2)^{\nu+2k}}{k!(\nu+k)!} ,$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

and $0 < j_{\nu,1} < j_{\nu,2} < \ldots$ are positive zeros of $J_\nu(\cdot)$. As our process starts at $x = 0$, the critical value $C_1(\alpha, d)$ can be obtained by solving the following equation

$$1 - \alpha = \sum_{k=1}^{\infty} \frac{2x^{-\nu} J_\nu(j_{\nu,k}x/C_1)}{C_1^{-\nu} j_{\nu,k} J_{\nu+1}(j_{\nu,k})} \exp\{-\frac{j_{\nu,k}^2}{2C_1^2}\}, \tag{4.12}$$

as $x \to +0$.

When x is very small, $x \to +0$, $J_\nu(x)$ is well approximated by $x^\nu/(2^\nu \nu!)$. Using this property we can approximate $(x/C_1)^{-\nu} J_\nu(j_{\nu,k}x/C_1)$ by $j_{\nu,k}^\nu/(2^\nu \nu!)$. Replacing this in (4.12) we can obtain the critical value $C_1(\alpha, d)$ by solving

$$1 - \alpha = \sum_{k=1}^{\infty} \frac{j_{\nu,k}^{\nu-1}}{2^{\nu-1} J_{\nu+1}(j_{\nu,k}) \nu!} \exp\{-\frac{j_{\nu,k}^2}{2C_1^2}\}. \tag{4.13}$$

In Table 4.1 we give the critical values for various α and d. To obtain these critical values in our computations we have used the first 100 terms in the sum on the right-hand side of equation (4.13). Note that the convergence of the series in (4.13) is not uniform in $\nu(d)$ and it is slower for larger values of $\nu(d)$. In this case more terms should be used in the sum. More extensive tables with the left-tail probability of the maximum of the Bessel process are given in Appendix B.

4.3.2 Critical values for Test 2

Using the critical values $C_2(\alpha, d, n_0)$ defined in (4.10) the test is too conservative. A better finite approximation to the critical values can be obtained by using a result of Vostrikova (1981). According to that result the critical value, $C_2'(\alpha, d, n_0)$ can be obtained by solving

$$\alpha = \frac{(C_2')^d \exp\{-(C_2')^2/2\}}{2^{d/2}\Gamma(d/2)} \left\{ \log(n_0) \left[1 - \frac{d}{(C_2')^2} + \frac{4}{(C_2')^2}\right] + O\left(\frac{1}{(C_2')^4}\right)\right\} \tag{4.14}$$

The following table gives the critical values $C_2'(\alpha, d, n_0)$ needed later in our simulations studies.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 4.1: Critical value $C_1(\alpha, d)$ for different values of d and different levels of significance α.

	α		
d	0.10	0.05	0.01
2	2.419	2.695	3.242
3	2.751	3.023	3.562
4	3.023	3.294	3.827
5	3.260	3.530	4.059
6	3.474	3.743	4.269
7	3.669	3.938	4.461
8	3.851	4.119	4.640
10	4.183	4.450	4.968
12	4.482	4.748	5.264

Table 4.2: Critical value $C_2'(\alpha, d, n_0)$ obtained using Vostrikova's formulae (4.14), for $\alpha = 0.05$ and different values of d and n_0 used in later simulations.

	n_0			
d	50	100	200	500
2	3.490	3.540	3.585	3.633
3	3.830	3.880	3.920	3.970
4	4.105	4.155	4.200	4.249

4.4 Simulation Studies

In this section we shall study the test statistics in the cases when we compare three or four means. In each case we have carried out some Monte Carlo experiments to evaluate the empirical power and the average stopping time of the tests proposed. Each scenario in these Monte Carlo simulations is based on 5,000 replicates.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

4.4.1 Comparison of three treatments

In this case we observe vectors $\mathbf{Y}_k = (Y_{1k}, Y_{2k}, Y_{3k})^t$, $k \geq 1$, with independent normal components, i.e., for each $i = 1, 2, 3$, Y_{ik} iid $N(\mu_i, \sigma^2)$, for all $k \geq 1$. The log likelihood function based on the first k observations is given by

$$l_k = -\frac{3}{2}\log 2\pi - \frac{3}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{3}\sum_{j=1}^{k}(Y_{ij} - \mu_i)^2 .$$

Using the parametrization

$$\theta = (\theta_1, \theta_2) = \left(\frac{\mu_1 + \mu_2 - 2\mu_3}{2\sigma^2\sqrt{6}}, \frac{\mu_1 - \mu_2}{2\sigma^2\sqrt{2}}\right), \quad \eta = (\eta_1, \eta_2) = \left(\frac{\mu_3}{2\sigma^2}, \frac{-1}{2\sigma^2}\right),$$

the hypothesis of interest becomes (4.7) with $\theta_0 = (0, 0)$, $d = 2$, and $p = 2$. Under H_0 the efficient score vector is

$$V_k(0, \hat{\eta}_k) = \left(\frac{1}{3\sqrt{k}}\sum_{j=1}^{k}(Y_{1j} + Y_{2j} - 2Y_{3j}), \frac{1}{\sqrt{k}}\sum_{j=1}^{k}(Y_{1j} - Y_{2j})\right),$$

and the information matrix is given by

$$\boldsymbol{I}(\theta_0, \eta) = \begin{pmatrix} -\frac{1}{\eta_2} & 0 & -\frac{1}{\eta_2} & \frac{\eta_1}{\eta_2^2} \\ 0 & -\frac{1}{\eta_2} & 0 & 0 \\ -\frac{1}{\eta_2} & 0 & -\frac{3}{2\eta_2} & \frac{3\eta_1}{2\eta_2^2} \\ \frac{\eta_1}{\eta_2^2} & 0 & \frac{3\eta_1}{2\eta_2^2} & \frac{3\eta_2 - 3\eta_1^2}{2\eta_2^3} \end{pmatrix}.$$

The maximum likelihood estimator of σ^2 is

$$\hat{\sigma}_k^2 = \frac{-1}{2\hat{\eta}_{2k}} = \frac{1}{3k}\sum_{j=1}^{k}\left(Y_{1j}^2 + Y_{2j}^2 + Y_{3j}^2\right) - \left[\frac{1}{3k}\sum_{j=1}^{k}(Y_{1j} + Y_{2j} + Y_{3j})\right]^2 .$$

Then, in terms of the initial parameters, Rao's statistic (4.9) becomes

$$R_k = \left[\frac{\sum_{j=1}^{k}(Y_{1j} + Y_{2j} - 2Y_{3j})}{\hat{\sigma}_k\sqrt{6k}}\right]^2 + \left[\frac{\sum_{j=1}^{k}(Y_{1j} - Y_{2j})}{\hat{\sigma}_k\sqrt{2k}}\right]^2 . \tag{4.15}$$

In our simulations the truncation point is $n_0 = 50$. For the sequential F-test we have used $c = 3.38$, $a = 6.68$, and $k_0 = 7$ as in Siegmund (1980).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 4.3: Comparison of three treatments. Simulated power (Power) and average stopping time (AVST) for Test 1, Test 2, and Sequential F-test with normally distributed outcomes for various δ. Level of significance $\alpha = 0.05$.

	POWER			AVST		
δ	Test 1	Test 2	Seq F	Test 1	Test 2	Seq F
0.0	0.039	0.011	0.042	49.67	49.74	48.54
0.1	0.076	0.020	0.077	49.35	49.57	48.29
0.2	0.191	0.063	0.193	48.22	48.81	47.44
0.3	0.411	0.174	0.405	45.60	46.91	45.46
0.4	0.660	0.371	0.658	41.87	43.36	41.84
0.5	0.860	0.636	0.857	36.95	37.78	36.31
0.6	0.959	0.833	0.960	32.48	31.89	30.47
0.7	0.992	0.946	0.992	28.80	26.06	24.44
0.8	1.000	0.994	1.000	25.85	21.25	19.61

From Table 4.3 we see that Test 1 based on Rao's statistic and the sequential F-test are comparable in terms of power but the sequential F-test stops earlier for large values of δ. Test 2 which is also based on Rao's statistic is comparable with the F-test in terms of average stopping time but it is less powerful. Note that the asymptotic distribution of these two tests is the same and the only difference is due to the different stopping rules.

4.4.2 Comparison of four treatments

In the case of four treatments the log likelihood function is given by

$$l_k = -2\log(2\pi) - 2\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{4}\sum_{j=1}^{k}(Y_{ij} - \mu_i)^2 .$$

Using the parametrization

$$\theta = (\theta_1, \theta_2, \theta_3) = \left(\frac{\mu_1 - \mu_2}{2\sigma^2\sqrt{2}}, \frac{\mu_3 - \mu_4}{2\sigma^2\sqrt{2}}, \frac{(\mu_1 + \mu_2) - (\mu_3 + \mu_4)}{4\sigma^2}\right),$$

$$\eta = (\eta_1, \eta_2) = \left(\frac{\mu_3 + \mu_4}{2\sigma^2\sqrt{2}}, -\frac{1}{2\sigma^2}\right),$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the hypothesis of interest becomes (4.7) with $\theta_0 = (0, 0, 0)$, $d = 3$, and $p = 2$. Under H_0 the efficient score vector is

$$V_k = \left(\frac{\sqrt{2}}{\sqrt{k}} \sum_{j=1}^{k} (Y_{1j} - Y_{2j}), \frac{\sqrt{2}}{\sqrt{k}} \sum_{j=1}^{k} (Y_{3j} - Y_{4j}), \frac{1}{\sqrt{k}} \sum_{j=1}^{k} \left[(Y_{1j} + Y_{2j}) - (Y_{3j} + Y_{4j}) \right] \right)$$

and the information matrix is given by

$$I(\theta_0, \eta) = \begin{pmatrix} -\frac{2}{\eta_2} & 0 & 0 & 0 & 0 \\ 0 & -\frac{2}{\eta_2} & 0 & 0 & 0 \\ 0 & 0 & -\frac{4}{\eta_2} & -\frac{2\sqrt{2}}{\eta_2} & \frac{2\eta_1\sqrt{2}}{\eta_2^2} \\ 0 & 0 & -\frac{2\sqrt{2}}{\eta_2} & -\frac{4}{\eta_2} & \frac{4\eta_1}{\eta_2} \\ 0 & 0 & \frac{2\eta_1\sqrt{2}}{\eta_2^2} & \frac{4\eta_1}{\eta_2} & \frac{2\eta_2 - 4\eta_1^2}{\eta_2^3} \end{pmatrix}.$$

The maximum likelihood estimator of σ^2 is

$$\hat{\sigma}_k^2 = \frac{-1}{2\hat{\eta}_{2k}} = \frac{1}{4k} \sum_{i=1}^{4} \sum_{j=1}^{k} Y_{ij}^2 - \left(\frac{1}{4k} \sum_{i=1}^{4} \sum_{j=1}^{k} Y_{ij} \right)^2.$$

Then, in terms of the initial parameters, Rao's statistic (4.9) becomes

$$R_k = \left[\frac{\sum_{j=1}^{k} (Y_{1j} - Y_{2j})}{\hat{\sigma}_k \sqrt{2k}} \right]^2 + \left[\frac{\sum_{j=1}^{k} (Y_{3j} - Y_{4j})}{\hat{\sigma}_k \sqrt{2k}} \right]^2 + \left[\frac{\sum_{j=1}^{k} \left(Y_{1j} + Y_{2j} - Y_{3j} - Y_{4j} \right)}{2\hat{\sigma}_k \sqrt{k}} \right]^2.$$

The results of our simulations are presented in Table 4.4. The truncation point is $n_0 = 50$ in the first part of the table and $n_0 = 100$ in the second part. The values of c and a for the sequential F-test are obtained by using (4.6). In order to obtain a level of significance of $\alpha = 0.05$ we have used $k_0 = 7$, $c = 3.65$ and $a = 6.95$ when $n_0 = 50$, and $k_0 = 15$, $c = 3.60$ and $a = 6.75$ when $n_0 = 100$.

Siegmund (1980) reported a difference of $10 - 15\%$ between the approximated significance level and the empirical value in the case of $d = 3$ and $n_0 = 49$. From Table 4.4 we can see that the empirical value is larger then the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

approximated value obtained by using (4.6). It seems that the difference becomes larger as d and n_0 increase. In each part of Table 4.4 we have computed the values of c and a so that the empirical power of the F test is 0.05. We obtained $c' = 4.65$ and $a' = 7.15$ when $n_0 = 50$, and $c'' = 4.3$ and $a'' = 7.25$ when $n_0 = 100$. These values were obtained based on 10,000 replicates. The empirical power obtained by using these values is given in the brackets. As before Test 1 and the sequential F-test are comparable in terms of power. The sequential F-test stops earlier but the tests based on Rao's statistic have the advantage that the critical values can be easily computed.

Table 4.4: Comparison of four treatments. Simulated power (Power) and average stopping time (AVST) for Test 1, Test 2 and Sequential F-test, with normally distributed outcomes for various δ and n_0. Level of significance $\alpha = 0.05$.

		POWER			AVST		
n_0	δ	Test1	Test 2	Seq F	Test1	Test2	Seq F
50	0.0	0.044	0.019	0.079 (0.050)	49.68	49.50	48.88 (49.01)
	0.1	0.065	0.032	0.131 (0.071)	49.47	49.18	48.59 (48.69)
	0.2	0.153	0.071	0.253 (0.152)	48.79	48.59	47.73 (48.04)
	0.3	0.341	0.176	0.453 (0.328)	46.85	46.66	45.70 (45.91)
	0.4	0.591	0.370	0.700 (0.581)	43.73	43.11	41.53 (41.99)
	0.5	0.810	0.613	0.884 (0.798)	39.40	37.85	35.75 (36.93)
	0.6	0.940	0.818	0.969 (0.927)	35.07	31.96	29.56 (30.25)
	0.7	0.985	0.946	0.994 (0.983)	31.08	25.96	23.54 (24.34)
	0.8	0.997	0.985	0.999 (0.977)	28.00	21.50	19.05 (19.87)
100	0.0	0.047	0.022	0.087 (0.050)	99.25	98.76	97.87 (98.31)
	0.1	0.097	0.044	0.161 (0.092)	98.51	97.71	96.39 (97.43)
	0.2	0.311	0.138	0.416 (0.296)	94.04	94.15	91.42 (93.47)
	0.3	0.667	0.395	0.748 (0.646)	84.30	84.54	78.65 (82.55)
	0.4	0.915	0.749	0.952 (0.910)	71.25	67.27	61.24 (82.55)
	0.5	0.990	0.942	0.996 (0.987)	59.52	49.58	44.58 (65.45)
	0.6	1.000	0.994	1.000 (0.999)	50.98	36.86	32.79 (35.25)
	0.7	1.000	1.000	1.000 (1.000)	44.60	28.35	25.62 (27.44)
	0.8	1.000	1.000	1.000 (1.000)	39.92	22.87	21.52 (22.34)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

4.5 Sequential change-point ANOVA

We have seen in the previous section that we can find a re-parametrization so that the components of the efficient score vector are independent. This means that we can use the CUSUM test defined in the second chapter to compare two or more groups. The CUSUM test is more appropriate for an ANOVA change-point problem, that is, if we want to test if at an unknown time point $\tau > 1$ one or more of the means have changed.

As before we shall assume that observations are made sequentially on vectors $\mathbf{Y_k} = (Y_{1k}, Y_{2k}, \ldots, Y_{dk})^t$, $k \geq 1$, where d denotes the number of groups (treatments), and Y_{ik} is the k^{th} observation from group (treatment) i. The observations are assumed to be independently and normally distributed with common variance, i.e., for each $i = 1, 2, \ldots, d$,

$$Y_{ik} \text{ iid } N(\mu_{ik}, \sigma^2), \quad \text{for all} \quad k \geq 1 . \tag{4.16}$$

Considering σ^2 as a nuisance parameter, we are interested in testing

$$H_0 : \mu_{1k} = \mu_{2k} = \ldots = \mu_{dk} = \mu, \text{ unknown}, \quad \text{for all} \quad k \geq 1 ,$$

against the alternative

$$H_A : \begin{cases} \mu_{1k} = \mu_{2k} = \ldots = \mu_{dk} = \mu, \text{ unknown}, \quad \text{for all} \quad k < \tau , \\ \text{there is an } i \in \{1, 2, \ldots, d\} \text{ such that } \mu_{ik} \neq \mu, \text{ for } \quad k \geq \tau , \end{cases}$$

where τ is the unknown change point.

With the same re-parametrization and the same notations as in the previous section, for each $l \leq k$, we can define

$$W_{k,l}(\theta_0, \hat{\eta}_k) = \mathbf{\Gamma}^{-1/2}(\theta_0, \hat{\eta}_k) V_{k,l}(\theta_0, \hat{\eta}_k) ,$$

where $\mathbf{\Gamma}(\theta_0, \hat{\eta}_k) = I_{11} - I_{12} I_{22}^{-1} I_{21}$ and $V_{k,l}(\theta_0, \hat{\eta}_k)$ is the score vector based on the last $k - l$ observations. The test is defined as in (2.14).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

4.5.1 The case of three groups

When $d = 3$, by using the results of Section 4.4.1, the components of the vector $W_{k,l}(\theta_0, \hat{\eta}_k)$ are given by

$$W_{k,l}^{(1)} = \frac{1}{\hat{\sigma}_k\sqrt{6}}\sum_{j=l}^{k}(Y_{1j} + Y_{2j} - 2Y_{3j}) ,$$

$$W_{k,l}^{(2)} = \frac{1}{\hat{\sigma}_k\sqrt{2}}\sum_{j=l}^{k}(Y_{1j} - Y_{2j}) ,$$

and the test is defined as follows.

Test 3 (CUSUM, $d = 3$). *Stop and reject H_0 at the first $k \geq 2$ for which*

$$\max\left\{\frac{1}{\sqrt{n_0}}\max_{2\leq l<k} W_{k,l}^{(1)}, \frac{1}{\sqrt{n_0}}\max_{2\leq l<k} W_{k,l}^{(2)}\right\} \geq C(\alpha^*) . \tag{4.17}$$

If no such $k \leq n_0$ exists then do not reject H_0.

Here the level of significance is $\alpha = 1 - (1 - \alpha^*)^2$. For $\alpha = 0.05$ we have $\alpha^* = 0.0253$ and $C(0.0253) = 2.48$. The empirical power and the average stopping time are presented in Table 4.5. The truncation point is $n_0 = 200$ and the change point is $\tau = 100$. Before the change-point the observations come from the same distribution, i.e. $\mu_1 = \mu_2 = \mu_3 = 0$, and after the change point μ_1 is changing to $\mu_1^{(A)}$ taking on values {0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0}. The standard deviation was unchanged, $\sigma = 1$. Each scenario is based on 5,000 replications.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 4.5: Simulated power (Power) and average stopping time (AVST) for the CUSUM test with normally distributed outcomes. $\sigma^2 = 1$, $\mu_2 = \mu_3 = 0$ and for various $\mu_1^{(A)}$. Level of significance $\alpha = 0.05$.

n_0	τ	$\mu_1^{(A)}$	Power	AVST
200	100	0.0	0.041	198.59
		0.1	0.099	196.90
		0.2	0.216	194.02
		0.3	0.410	188.31
		0.4	0.645	180.76
		0.5	0.846	171.43
		0.6	0.952	162.08
		0.7	0.993	154.48
		0.8	0.999	148.19
		0.9	1.000	143.65
		1.0	1.000	139.33

Table 4.6: Simulated power (Power) and average stopping time (AVST) for the CUSUM test with normally distributed outcomes. $\sigma^2 = 1$, $\mu_2 = \mu_3 = \mu_4 = 0$ and for various $\mu_1^{(A)}$. Level of significance $\alpha = 0.05$.

n_0	τ	$\mu_1^{(A)}$	Power	AVST
200	100	0.0	0.037	198.75
		0.1	0.087	197.39
		0.2	0.180	195.24
		0.3	0.364	190.25
		0.4	0.602	183.14
		0.5	0.820	174.14
		0.6	0.946	164.86
		0.7	0.990	157.00
		0.8	0.998	150.55
		0.9	1.000	145.27
		1.0	1.000	141.14

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

4.5.2 The case of three groups

When $d = 4$, by using the results of Section 4.4.2, the components of the vector $W_{k,l}(\theta_0, \hat{\eta}_k)$ are given by

$$W_{k,l}^{(1)} = \frac{1}{\hat{\sigma}_k\sqrt{2}} \sum_{j=l}^{k} (Y_{1j} - Y_{2j}) \, ,$$

$$W_{k,l}^{(2)} = \frac{1}{\hat{\sigma}_k\sqrt{2}} \sum_{j=l}^{k} (Y_{3j} - Y_{4j}) \, ,$$

$$W_{k,l}^{(3)} = \frac{1}{2\hat{\sigma}_k} \sum_{j=l}^{k} \left(Y_{1j} + Y_{2j} - Y_{3j} - Y_{4j}\right) \, ,$$

and the test is defined as follows.

Test 4 (CUSUM, $d = 4$). *Stop and reject H_0 at the first $k \geq 2$ for which*

$$\max\left\{\frac{1}{\sqrt{n_0}} \max_{2 \leq l < k} W_{k,l}^{(1)}, \frac{1}{\sqrt{n_0}} \max_{2 \leq l < k} W_{k,l}^{(2)}, \frac{1}{\sqrt{n_0}} \max_{2 \leq l < k} W_{k,l}^{(3)}\right\} \geq C(\alpha^*) \, . \quad (4.18)$$

If no such $k \leq n_0$ exists then do not reject H_0.

Here the level of significance is $\alpha = 1 - (1 - \alpha^*)^3$, and for $\alpha = 0.05$ we have $\alpha^* = 0.0169$ and $C(0.0169) = 2.632$. The empirical power and the average stopping time are presented in Table 4.6. The truncation point is $n_0 = 200$ and the change point is $\tau = 100$. Before the change-point the observations come from the same distribution, i.e., $\mu_1 = \mu_2 = \mu_3 = \mu_4 = 0$, and after the change point μ_1 is changing to $\mu_1^{(A)}$ taking on values {0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0}. The standard deviation was unchanged, $\sigma = 1$ and each scenario is based on 5,000 replications.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Chapter 5

Concluding Remarks

In this thesis the following results were obtained

- A truncated CUSUM-type sequential test was derived. The test is based on large sample approximations of the efficient score vector. It is applicable to a large class of distributions including distributions from the non-exponential family. The main attractive features of the new test are simple accommodation of nuisance parameters and easy computations of the asymptotic critical values. The test statistic was examined under the alternative hypothesis and its consistency was demonstrated. An application where the observations come from a non-exponential distribution was provided.

- The procedure described above was extended to the case of autocorrelated observations. The test statistics are provided in the case of monitoring the parameters of an AR(p) process. In the case of AR(1) and AR(2) models the consistency of the test was proved and the empirical power was evaluated in a simulation study. As the results of Lai (1995) are the only ones we found that perform test of change in the same generality as we do, we have compared our test to his algorithm. The

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

approximations developed in this part are strong enough to allow us to extend other testing procedures available for independent observations to the case of AR(p) models.

- Some testing procedures were proposed for comparison of three or more treatments. Tests based on Rao's statistics proposed by Gombay (2003) and to the sequential F-test proposed by Siegmund (1980) are compared in an extensive simulation study. The critical values for Test 2 of Gombay (2003) are computed.

- The CUSUM test is adapted to the ANOVA change-point problem. The empirical power and the average stopping time are computed.

- Strong approximations for sums of observations from an AR (p) model are proved in the last part of the thesis (Appendix A).

The results of this thesis can be extended in the following directions

- As the efficient score vector behaves approximately as a partial sums sequence, the idea of Siegmund (1985) can be used to develop new stopping rules that can improve the average stopping time of the test. An open ended procedure might be useful in some problems.

- Procedures to estimate the change-point and the new values of the parameters following the change-point detection are of interest.

- The components of the efficient score vector might be used to develop new CUSUM and EWMA control charts. They might be implemented in the case of multiple change-point problems.

- The procedure can be extended to other time series models like MA or ARMA. An extension to the case of random coefficient autoregressive models (RCA) seems to work under certain moment conditions.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

- In the case of comparison of three treatments when the objective is to find the best treatment, methods similar to those of Betensky (1996) and Siegmund (1993) can be based on the components of the efficient score vector. By using the simple structure of the efficient score vector the approximations of the critical values might be easier to obtain.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Bibliography

Alwan, L.C. and H.V. Roberts (1988). Time Series Modeling for Statistical Process Control. *J. Business Econ. Statist.* (**6**), 87–95.

Armitage, P. (1960). *Sequential Medical Trials.* Blackwell.

Atenafu, E. and E. Gombay (2004). Sequential Tests for Change in the Parameters of Nested Random Effects Model. In: *Asymptotic Methods in Statistics: A volume in Honor of Miklós Csörgő.* Fields Institute Communications, 44, 301-314.

Bagshaw, M. and R.A. Johnson (1975). The Effect of Serial Correlation on the Performance of CUSUM Tests II. *Technometrics* (**17**), 73–80.

Bai, J. (1993). On the Partial Sums of Residuals in Autoregressive and Moving Average Models. *Journal of Time Series Analysis* (**14**), 247–260.

Bai, J. (1994a). Least Squares Estimation of a Shift in Linear Processes. *Journal of Time Series Analysis* (**15**), 453–472.

Bai, J. (1994b). Weak Convergence of Sequential Empirical Processes of Residuals in ARMA Models. *Ann. Statist.* (**22**), 2051–2061.

Bartlett, M.S. (1949). The Large Sample Theory of Sequential Tests. *Proc. Camb. Phil. Soc.* (**42**), 239–244.

Basseville, M. and I.V. Nikiforov (1993). *Detection of Abrupt Changes: Theory and Applications.* Prentice Hall. Englewood Cliffs.

Betensky, R.A. (1996). An O'Brien-Fleming Sequential Trial for Comparing Three Treatments. *Annals of Statistics* (**24**), 1765–1791.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley. New York, New York.

Boldin, M.V. (1982). Estimation of the Distribution of Noise in Autoregressive Scheme. *Theory Probab. Appl.* (**27**), 866–871.

Borodin, A.N. and P. Salminen (1996). *Handbook of Brownian Motion - Facts and Formulae.* Birkhauser.

Breslow, N. (1969). On Large Sample Sequential Analysis with Applications to Survivorship Data. *Journal of Applied Probability* (**6**), 261–274.

Brockwell, P.J. and R.A. Davis (1991). *Time Series: Theory and Methods.* Springer-Verlag. New York, New York.

Brown, R.L., J. Durbin and J.M. Evans (1975). Techniques for Testing the Constancy of Regression Relationships Over Time (with discussions). *J. Roy. Statist. Soc. Ser. B* (**37**), 149–192.

Chernoff, H. and S. Zacks (1964). Estimating the Current Mean of A Normal Distribution which Is Subjected to Changes in Time . *Ann. Math. Statist.* (**35**), 999–1018.

Cox, D.R. (1963). Large Sample Sequential Tests for Composite Hypotheses. *Sankhya* (**A25**), 5–12.

Csörgő, M. and L. Horváth (1993). *Weighted Approximations in Probability and Statistics.* Wiley. Chichester.

Csörgő, M. and L. Horváth (1997). *Limit Theorems in Change-Point Analysis.* Wiley. New York, New York.

Csörgő, M. and P. Révész (1981). *Strong Approximations in Probability and Statistics.* Academic Press. New York, New York.

Davis, R.A., D. Huang and C. Yao (1995). Testing for a Change in the Parameter Values and Order of an Autoregressive Model. *Ann. Statist.* (**23**), 282–304.

Eberlein, E. (1986). On Strong Invariance Principles under Dependence Assumptions. *Ann. Probab.* (**14**), 260–300.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Einmahl, U. (1987). Strong Invariance Principles for Partial Sums of Independent Random Vectors. *Ann. Probab.* (**15**), 1419–1440.

Ghosh, B.K. (1970). *Sequential Tests of Hypotheses.* Addison-Wesley.

Gombay, E. (1994). Testing for Change-Points with Rank and Sign Statistics. *Statistics & Probability Letters* (**20**), 49–55.

Gombay, E. (1996). The Weighted Sequential Likelihood Ratio. *Canadian J. Statist.* (**24**), 229–239.

Gombay, E. (1997). The Likelihood Ratio Under Noncontiguous Alternatives. *Canadian J. Statist.* (**25**), 417–423.

Gombay, E. (2000). Sequential Change-Detection with Likelihood Ratios. *Statistics and Probability Letters* (**49**), 195–204.

Gombay, E. (2002). Parametric Sequential Tests in the Presence of Nuisance Parameters. *Theory Stoch. Processes* (**1-2**), 107–118.

Gombay, E. (2003a). Sequential Change-Point Detection and Estimation. *Sequential Analysis* (**22**), 203–222.

Gombay, E. (2003b). Sequential Tests of Composite Hypotheses. In: *Limit Theorems of Probability and Statistics II.* I. Berkes, E. Csáki, M. Csörgő eds.. Budapest.

Gombay, E. and L. Horváth (1994). An Application of the Maximum Likelihood Test to the Change-Point Problem. *Stochastic Process. Appl.* (**50**), 161–171.

Gombay, E., G. Heo and A Hussein (2001). Nonparametric Sequential Comparison of Two Populations Using U-statistics. *Preprint.*

Horváth, L. (1993). Change in Autoregressive Processes. *Stochastic Process. Appl.* (**44**), 221–242.

Jiang, W., Tsui K.L. and W.H. Woodall (2000). A New SPC Monitoring Method: The ARMA Chart. *Technometrics* (**42**), 339–410.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Johnson, R.A. and M. Bagshaw (1975). The Effect of Serial Correlation on the Performance of CUSUM Tests. *Technometrics* (**16**), 103–112.

Komlós, J., Major P. and G. Tusnády (1975). An Approximation of Partial Sums of Independent R.V's and the Sample DF. *I.Z. Wahrsch. verw. Gebiete* (**32**), 111–131.

Konev, V. and S. Pergamenshchikov (1997). On Guaranteed Estimation of an Autoregressive Process. *Ann. of Statistics* (**25**), 2127–2163.

Kulperger, R.J. (1985). On the Residuals of Autoregressive and Polynomial Regression. *Stochastic Processes and their Applications* (**21**), 107–118.

Lai, T.L. (1995). Sequential Changepoint Detection in Quality Control and Dynamical Systems. *J. R. Statist. Soc. B* (**57**), 613–658.

Lai, T.L. (2001). Sequential Analysis: Some Classical Problems and New Challenges. *Statistica Sinica* (**11**), 303–408.

Lee, S. and S. Park (2001). The CUSUM of Squares Test for Scale Changes in Infinite Order Moving Average Processes. *Scand. J. Statist.* (**28**), 625–644.

Lehmann, E.L. (2001). Elements of Large Sample Theory. *Springer-Verlag* (**28**), 625–644.

Lerche, H.R. (1986). *Boundary Crossing of Brownian Motion.* Springer-Verlag.

Lorden, G. (1971). Procedures for Reacting to a Change in Distribution. *Ann. Math. Statist.* (**42**), 1879–1908.

Lurie, A.M. and N.K. Neerchal (1999). Bayes Type Tests for Constancy of Autoregressive Parameters. *Environmetrics* (**10**), 737–752.

Montgomery, D.C. and D.J. Friedman (1989). Statistical Process Control in a Computer-Integrated Manufacturing Environment. In: *Statistical Process Control in Automated Manufacturing.* Marcel Decker. New York. pp. 67–88.

Moustakides, G.V. (1986). Optimal Stopping Times for Detecting Changes in Distributions. *Ann. Statist.* (**14**), 1379–1387.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Nyblom, J. (1989). Likelihood Ratio Methods for Monitoring Parameters of Nested Random Effects Model. *J. Amer. Statist. Assoc.* (**84**), 223–230.

O'Brien, P.C. and T.R. Fleming (1979). A Multiple Testing Procedure for Clinical Trials. *Biometrics* (**35**), 549–556.

Page, E.S. (1955). A Test for a Change in a Parameter Occurring at an Unknown Point. *Biometrika* (**42**), 523–527.

Phillips, P. and V. Solo (1992). Asymptotics for Linear Processes. *The Ann. of Statistics* (**20**), 971–1001.

Picard, D. (1985). A Test for a Change-point in Time Series. *Adv. Appl. Probab.* (**17**), 841–867.

Pocock, S.J. (1977). Group Sequential Methods in the Design and Analysis of Clinical Trials. *Biometrika* (**64**), 191–199.

Pollak, M. and D. Siegmund (1985). A Diffusion Process and its Applications to Detecting a Change in the Drift of a Brownian Motion. *Biometrika* (**73**), 267–280.

Robbins, H. and D. Siegmund (1972). Sequential Tests Involving Two Populations. *J. Amer. Statist. Ass.* (**69**), 132–139.

Serfling, R.J. (1980). *Approximations Theorems of Mathematical Statistics.* Wiley. New York.

Siegmund, D. (1980). Sequential and F Tests and the Related Confidence Intervals. *Biometrika* (**67**), 389–402.

Siegmund, D. (1985). *Sequential Analysis: Tests and Confidence Intervals.* Springer-Verlag. New York.

Siegmund, D. (1993). A Sequential Clinical Trial for Comparing Three Treatments. *Annals of Statistics* (**21**), 464–483.

Tang, S.M. and I.B. MacNeill (1993). The Effect of Serial Correlation on Tests for Parameter Change at Unknown Time. *Ann. Statist.* (**21**), 552–575.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Telksnys, L. (1986). *Detection of Changes in Random Processes.* Optimization Software.

Vostrikova, L.J. (1981). Detection of a "Disorder" in a Wiener Process. *Theory Probab. Appl.* (**26**), 356–362.

Wald, A. (1947). *Sequential Analysis.* Wiley. New York, New York.

Wald, A. and J. Wolfowitz (1948). Optimum Character of the Sequential Probability Ratio Test. *Ann. of Math. Statis.* (**19**), 326–339.

Wetherill, G.B. (1977). *Sampling Inspection and Quality Control.* 2nd ed.. Chapman and Hall. London.

Yashchin, E. (1993). Performance of CUSUM Control Schemes for Serially Correlated Observations. *Technometrics* (**35**), 37–52.

Yashchin, E. (1995). Likelihood Ratio Methods for Monitoring Parameters of Nested Random Effects Model. *J. Amer. Statist. Assoc.* (**90**), 729–738.

Zhang, N.F. (1998). A Statistical Control Chart for Stationary Process Data. *Technometrics* (**40**), 24–38.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Appendix A

Strong Approximations for AR(p) Models

In this appendix we verify a strong invariance principle for the partial sums of an AR(p) time series model. Let $\{X_k\}_{k\in\mathbb{Z}}$ be a sequence of random variables. We denote

$$S_k(m) = X_{m+1} + \ldots + X_{m+k} \quad \text{for } m \geq 0 \text{ and } k \geq 1 \, .$$

Let $\|\cdot\|_1$ denote the L_1-norm. Our results will follow from a theorem of Eberlein (1986). As we are interested in real-valued random variables we shall state the theorem in a simpler version.

Theorem A.1 (Eberlein). Let $\{X_k\}_{k\in\mathbb{Z}}$ be a sequence of real-valued random variables such that

1. $EX_k = 0$ for all $k \in \mathbb{Z}$.
2. $\|E\big(S_k(m)|\mathcal{F}_m\big)\|_1 = O\big(k^{\frac{1}{2}-\zeta}\big)$ uniformly in $m \geq 1$ for some $\zeta \in \big(0, \frac{1}{2}\big)$, where $\mathcal{F}_m = \sigma\big(X_l : l \leq m\big)$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

3. There exists a constant σ_w^2 such that uniformly in $m \geq 1$,

$$\frac{1}{k} ES_k^2(m) - \sigma_w^2 = O(k^{-\lambda}) ,$$

as $k \longrightarrow \infty$ for some $\lambda > 0$.

4. There exists $\gamma > 0$ such that uniformly in $m \geq 1$,

$$\|E(S_k^2(m)|\mathcal{F}_m) - ES_k^2(m)\|_1 = O(k^{1-\gamma}) , \quad a.s.,$$

as $k \longrightarrow \infty$.

5. There exists a constant $M < \infty$ and $r > 2$, such that $E|X_k|^r < M$ for all $k \geq 1$.

Then, there exists a Wiener process $\{W(t) : t \geq 0\}$, such that

$$\left|\sum_{k=1}^{[t]} X_k - \sigma_w W(t)\right| = O(t^{1/\nu}) \quad a.s.,$$

as $t \longrightarrow \infty$ for some $\nu > 2$.

By verifying the assumptions made in the above theorem we get the following strong approximations for $AR(p)$ time series.

Lemma A.1 Let $\{Y_k\}_{k \in \mathbb{Z}}$ be an $AR(p)$ process defined in (3.1) and assume that (3.2) and (3.3) hold. Then, there exists a Wiener process $\{W(t) : t \geq 0\}$, such that

$$\left|\sum_{k=1}^{[t]} (Y_k - \mu) - \sigma_y W(t)\right| = O(t^{1/\nu}) \qquad a.s.,$$

as $t \longrightarrow \infty$ for some $\sigma_y > 0$ and some $\nu > 2$.

Proof. Let $X_k = Y_k - \mu$, for all $k \in \mathbb{Z}$. It is enough to prove that the sequence $\{X_k\}$ satisfies the assumptions 1-5 of Eberlein's theorem. Since $\{X_k\}$ is stationary we can find a sequence of constants $(\alpha_i)_{i \geq 0}$ such that

$$X_t = \sum_{i \geq 0} \alpha_i \varepsilon_{t-i} , \tag{A.1}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

and there are constants M and $0 < \alpha < 1$ such that

$$|\alpha_i| \leq M\alpha^i \text{ for any } \; i \geq 0\,, \tag{A.2}$$

conform Brockwell and Davis (1991, page 85).

1. It is easy to see that $EX_k = E(Y_k - \mu) = 0$, for all k, and the first condition holds.

2. As $E(\varepsilon_i) = 0$, for all i, by using relationship (A.1) we obtain

$$\begin{aligned}
E\big(S_k(m)|\mathcal{F}_m\big) &= \sum_{t=1}^{k} E\big(X_{m+t}|\mathcal{F}_m\big) \\
&= \sum_{t=1}^{k}\sum_{i\geq 0} \alpha_i E\big(\varepsilon_{m+t-i}|\mathcal{F}_m\big) \\
&= \sum_{t=1}^{k}\left\{\sum_{i=0}^{t-1} \alpha_i E\big(\varepsilon_{m+t-i}\big) + \sum_{i\geq t} \alpha_i \varepsilon_{m+t-i}\right\} \\
&= \sum_{t=1}^{k}\sum_{i\geq t} \alpha_i \varepsilon_{m+t-i}.
\end{aligned}$$

Now we can write

$$\begin{aligned}
\|E\big(S_k(m)|\mathcal{F}_m\big)\|_1 &= E\left|\sum_{t=1}^{k}\sum_{i\geq t}\alpha_i\varepsilon_{m+t-i}\right| \leq \sum_{t=1}^{k}\sum_{i\geq t}|\alpha_i|E|\varepsilon_{m+t-i}| \\
&\leq ME|\varepsilon_0|\sum_{t=1}^{k}\sum_{i\geq t}\alpha^i \leq \frac{\alpha ME|\varepsilon_0|}{1-\alpha}\sum_{t=1}^{k}\alpha^{t-1} \\
&\leq \frac{\alpha ME|\varepsilon_0|}{(1-\alpha)^2}(1-\alpha^k) = O(1), \text{ as } k \longrightarrow \infty,
\end{aligned}$$

which implies the second assumption of the theorem.

3. We can write

$$S_k^2(m) = \sum_{t=1}^{k} X_{m+t}^2 + 2\sum_{1\leq t<l\leq k} X_{m+t}X_{m+l}.$$

For each $t \geq 1$ we have

$$\begin{aligned}
EX_{m+t}^2 &= \sum_{i\geq 0}\alpha_i^2 E\big(\varepsilon_{m+t-i}^2\big) + 2\sum_{j>i\geq 0}\alpha_i\alpha_j E\big(\varepsilon_{m+t-i}\varepsilon_{m+t-j}\big) \\
&= \sigma^2\left(\sum_{i\geq 0}\alpha_i^2\right).
\end{aligned}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

In the same way, for $t < l$, we get

$$E\big(X_{m+t}X_{m+l}\big) = \sum_{i,j\geq 0} \alpha_i\alpha_j E\big(\varepsilon_{m+t-i}\varepsilon_{m+l-j}\big)$$
$$= \sigma^2 \sum_{i\geq 0} \alpha_i\alpha_{i+(l-t)} \,.$$

Now, putting together the las three relationships we obtain

$$ES_k^2(m) = k\sigma^2\Big(\sum_{i\geq 0}\alpha_i^2\Big) + 2\sigma^2 \sum_{1\leq t<l\leq k}\sum_{i\geq 0}\alpha_i\alpha_{i+(l-t)} \,. \tag{A.3}$$

This can be rewritten as

$$\frac{1}{k}ES_k^2(m) = \sigma^2\Big(\sum_{i\geq 0}\alpha_i^2\Big) + \frac{2\sigma^2}{k}\sum_{1\leq t<l\leq k}\sum_{i\geq 0}\alpha_i\alpha_{i+(l-t)}$$
$$= \sigma^2\Big(\sum_{i\geq 0}\alpha_i^2\Big) + \frac{2\sigma^2}{k}\Big\{\sum_{t=1}^{k-1}\Big(k\sum_{i\geq 0}\alpha_i\alpha_{i+t}\Big) - \sum_{t=1}^{k-1}\Big(t\sum_{i\geq 0}\alpha_i\alpha_{i+t}\Big)\Big\}$$
$$= \sigma^2\Big(\sum_{i\geq 0}\alpha_i^2\Big) + 2\sigma^2\sum_{t=1}^{k-1}\Big(\sum_{i\geq 0}\alpha_i\alpha_{i+t}\Big) - \frac{2\sigma^2}{k}\sum_{t=1}^{k-1}\Big(t\sum_{i\geq 0}\alpha_i\alpha_{i+t}\Big)$$
$$= \sigma^2\Big(\sum_{i\geq 0}\alpha_i^2\Big) + A_k + B_k.$$

By using (A.2), for any $n \geq 1$ we have

$$|A_{k+n} - A_k| = \Big|2\sigma^2\sum_{t=k}^{k+n-1}\Big(\sum_{i\geq 0}\alpha_i\alpha_{i+t}\Big)\Big| \leq 2\sigma^2\sum_{t=k}^{k+n-1}\Big(\sum_{i\geq 0}|\alpha_i\alpha_{i+t}|\Big)$$
$$\leq 2\sigma^2M^2\sum_{t=k}^{k+n-1}\Big(\alpha^t\sum_{i\geq 0}\alpha^{2i}\Big) = \frac{2\sigma^2M^2}{1-\alpha^2}\alpha^k\sum_{t=0}^{n-1}\alpha^t$$
$$= \frac{2\sigma^2M^2}{(1-\alpha^2)(1-\alpha)}\alpha^k(1-\alpha^n) \longrightarrow 0 \text{ as } k\longrightarrow\infty\,,$$

and by the Cauchy criterion the sequence $\{A_k\}$ is convergent. Let

$$A = \lim_{k\to\infty} A_k = 2\sigma^2\sum_{t\geq 1}\Big(\sum_{i\geq 0}\alpha_i\alpha_{i+t}\Big)\,.$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Then, as before we obtain

$$
\begin{aligned}
|A - A_k| &= 2\sigma^2 \Big| \sum_{t\ge k} \Big(\sum_{i\ge 0} \alpha_i \alpha_{i+t} \Big) \Big| \le 2\sigma^2 M^2 \sum_{t\ge k} \alpha^t \Big(\sum_{i\ge 0} \alpha^{2i} \Big) \\
&= \frac{2\sigma^2 M^2}{(1-\alpha^2)(1-\alpha)} \alpha^k = O(k^{-1}) \ \text{ as } k \longrightarrow \infty \, .
\end{aligned}
$$

By using the same techniques we get

$$
\begin{aligned}
|B_k| &= \frac{2\sigma^2}{k} \Big| \sum_{t=1}^{k-1} \Big(t \sum_{i\ge 0} \alpha_i \alpha_{i+t} \Big) \Big| \le \frac{2\sigma^2}{k} \sum_{t=1}^{k-1} \Big(t \sum_{i\ge 0} |\alpha_i \alpha_{i+t}| \Big) \\
&\le \frac{2\sigma^2 M^2}{k} \sum_{t=1}^{k-1} t\alpha^t \Big(\sum_{i\ge 0} \alpha^{2i} \Big) = \frac{2\sigma^2 M^2}{k(1-\alpha^2)} \sum_{t=1}^{k-1} t\alpha^t \\
&= \frac{2\alpha\sigma^2 M^2}{k(1-\alpha^2)(1-\alpha)^2} = O(k^{-1}) \ \text{ as } k \longrightarrow \infty \, .
\end{aligned}
$$

Now using the last two relationships we obtain

$$
\frac{1}{k} E S_k^2(m) - \sigma^2 \Big(\sum_{i\ge 0} \alpha_i \Big)^2 = (A_k - A) + B_k = O(k^{-1}) \, ,
$$

and the third assumption of Eberlein's theorem follows.

4. Observe that

$$
S_k(m) = \sum_{t=1}^{k} X_{m+t} = \sum_{t=1}^{k} \Big(\sum_{i\ge 0} \alpha_i \varepsilon_{m+t-i} \Big) = \sum_{j\ge 0} \gamma_j \varepsilon_{m+k-j}
$$

where $\gamma_j = \gamma_j(k)$ are defined by

$$
\begin{cases}
\gamma_j = \alpha_0 + \alpha_1 + \ldots + \alpha_j, & \text{for } j = 0, 1, \ldots, k-1, \\
\gamma_j = \alpha_j + \alpha_{j-1} + \ldots + \alpha_{j-(k-1)}, & \text{for } j \ge k \, .
\end{cases}
\tag{A.4}
$$

Note that for each $j \ge k$ we have

$$
|\gamma_j| = | \sum_{i=j-k+1}^{j} \alpha_i | \le M \sum_{i=j-k+1}^{j} \alpha^i = M\alpha^{j-k+1} \sum_{i=0}^{k-1} \alpha^i = \frac{M(1-\alpha^k)}{(1-\alpha)} \alpha^{j-k+1} \, .
\tag{A.5}
$$

Now we can write

$$
S_k^2(m) = \sum_{j\ge 0} \gamma_j^2 \varepsilon_{m+k-j}^2 + 2 \sum_{j\ge 0} \sum_{i>j} \gamma_i \gamma_j \varepsilon_{m+k-i} \varepsilon_{m+k-j} \, .
\tag{A.6}
$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

By using (A.6) we obtain

$$ES_k^2(m) = \sigma^2\Big(\sum_{j\geq 0}\gamma_j^2\Big)\,, \tag{A.7}$$

and

$$E\big(S_k^2(m)|\mathcal{F}_m\big) = \sigma^2\Big(\sum_{j=0}^{k-1}\gamma_j^2\Big) + \sum_{j\geq k}\gamma_j^2\varepsilon_{m+k-j}^2 + 2\sum_{j\geq k}\sum_{i>j}\gamma_i\gamma_j\varepsilon_{m+k-i}\varepsilon_{m+k-j}\,, \tag{A.8}$$

Now collecting (A.7) and (A.8)

$$\begin{aligned}E\big(S_k^2(m)|\mathcal{F}_m\big) - ES_k^2(m) &= \sum_{j\geq k}\gamma_j^2\big(\varepsilon_{m+k-j}^2 - \sigma^2\big) + 2\sum_{j\geq k}\sum_{i>j}\gamma_i\gamma_j\varepsilon_{m+k-i}\varepsilon_{m+k-j}\\ &= D_k(m)\,.\end{aligned}$$

We can write

$$E|D_k(m)| \leq E|\varepsilon_0^2 - \sigma^2|\sum_{j\geq k}\gamma_j^2 + 2E|\varepsilon_0\varepsilon_1|\sum_{j\geq k}\sum_{i>j}|\gamma_i\gamma_j| = A_k + B_k\,.$$

Now we shall prove that each term on the right-hand side of the above inequality is of order $O(1)$ as $k \longrightarrow \infty$. Consider the first term. By (A.5) we have

$$\begin{aligned}A_k = E|\varepsilon_0^2 - \sigma^2|\sum_{j\geq k}\gamma_j^2 &\leq \frac{M^2\alpha^2E|\varepsilon_0^2-\sigma^2|}{(1-\alpha)^2}(1-\alpha^k)^2\sum_{j\geq k}\alpha^{2(j-k)}\\ &= \frac{M^2\alpha^2E|\varepsilon_0^2-\sigma^2|}{(1-\alpha)^2(1-\alpha^2)}(1-\alpha^k)^2 = O(1) \text{ as } k\longrightarrow\infty\,.\end{aligned}$$

In the same way

$$\begin{aligned}B_k = 2E|\varepsilon_0\varepsilon_1|\sum_{j\geq k}\sum_{i>j}|\gamma_i\gamma_j| &\leq \frac{M^2\alpha^2E|\varepsilon_0\varepsilon_1|}{(1-\alpha)^2}(1-\alpha^k)^2\sum_{j\geq k}\sum_{i>j}\alpha^{i+j-2k+2}\\ &\leq \frac{M^2\alpha^3E|\varepsilon_0\varepsilon_1|}{(1-\alpha)^2}(1-\alpha^k)^2\sum_{j\geq k}\alpha^{2(j-k)}\Big(\sum_{i\geq 0}\alpha^i\Big)\\ &= \frac{M^2\alpha^3E|\varepsilon_0\varepsilon_1|}{(1-\alpha)^3}(1-\alpha^k)^2\sum_{j\geq k}\alpha^{2(j-k)}\\ &= \frac{M^2\alpha^3E|\varepsilon_0\varepsilon_1|}{(1-\alpha)^3(1-\alpha^2)}(1-\alpha^k)^2 = O(1) \text{ as } k\longrightarrow\infty\,,\end{aligned}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

and the fourth assumption is satisfied. As the last assumption is trivial to check the lemma is proved.

□

Lemma A.2 Let $\{Y_k\}_{k\in\mathbb{Z}}$ be an $AR(p)$ process defined in (3.1) and assume that (3.2) and (3.3) hold. Then, for any $s \in \{1, 2, \dots, p\}$ there exists a Wiener process $\{W(t) : t \geq 0\}$, such that

$$\left|\sum_{k=1}^{[t]} \varepsilon_k (Y_{k-s} - \mu) - \sigma_s W(t)\right| = O\big(t^{1/\nu}\big) \qquad a.s.,$$

as $t \longrightarrow \infty$ for some $\sigma_s > 0$ and some $\nu > 2$.

Proof. Let $s \in \{1, 2, \dots, p\}$ fixed and denote $X_k = \varepsilon_k(Y_{k-s} - \mu)$. Now we shall prove that the sequence $\{X_k\}$ satisfies assumptions 1-5 of Eberlein's theorem.

1. It is easy to see that $EX_k = E\big(\varepsilon_k(Y_{k-s} - \mu)\big) = E(\varepsilon_k)E(Y_{k-s} - \mu) = 0$, for all k, and the first condition holds.

2. As $E(\varepsilon_i) = 0$, for all i, by using relationship (A.1) for $(Y_{k-s} - \mu)$ we obtain

$$\begin{aligned} E\big(S_k(m)|\mathcal{F}_m\big) &= \sum_{t=1}^{k} E\big(X_{m+t}|\mathcal{F}_m\big) = \sum_{t=1}^{k} E\big(\varepsilon_{m+t}(Y_{m+t-s} - \mu)|\mathcal{F}_m\big) \\ &= \sum_{t=1}^{k}\sum_{i\geq 0} \alpha_i E\big(\varepsilon_{m+t}\varepsilon_{m+t-s-i}|\mathcal{F}_m\big) \\ &= \sum_{t=1}^{k}\left\{\sum_{i=0}^{t-s-1} \alpha_i E\big(\varepsilon_{m+t}\varepsilon_{m+t-s-i}\big) + \sum_{i\geq t-s} \alpha_i \varepsilon_{m+t-s-i} E\big(\varepsilon_{m+t}\big)\right\} \\ &= 0\,, \end{aligned}$$

and the second assumption is verified.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

3. For each $t \geq 1$ we have

$$\begin{aligned} EX_{m+t}^2 &= E\big(\varepsilon_{m+t}^2 (Y_{m+t-s}-\mu)^2\big) = E\big(\varepsilon_{m+t}^2\big) E(Y_{m+t-s}-\mu)^2 \\ &= \sigma^2 \sum_{i\geq 0} \alpha_i^2 E\big(\varepsilon_{m+t-s-i}^2\big) + 2\sigma^2 \sum_{j>i\geq 0} \alpha_i\alpha_j E\big(\varepsilon_{m+t-s-i}\varepsilon_{m+t-s-j}\big) \\ &= \sigma^4 \Big(\sum_{i\geq 0} \alpha_i^2\Big). \end{aligned}$$

In the same way, for $0 < t < l$, we get

$$\begin{aligned} E\big(X_{m+t}X_{m+l}\big) &= E\big[\varepsilon_{m+t}\varepsilon_{m+l}(Y_{m+t-s}-\mu)(Y_{m+l-s}-\mu)\big] \\ &= E\big(\varepsilon_{m+t}\big)E\big(\varepsilon_{m+l}\big)E\big[(Y_{m+t-s}-\mu)(Y_{m+l-s}-\mu)\big] \\ &= 0. \end{aligned}$$

Now using the last two relationships we obtain

$$\begin{aligned} ES_k^2(m) &= \sum_{t=1}^{k} EX_{m+t}^2 + 2 \sum_{1\leq t<l\leq k} E\big(X_{m+t}X_{m+l}\big) \\ &= k\sigma^4 \Big(\sum_{i\geq 0} \alpha_i^2\Big), \end{aligned} \tag{A.9}$$

and the third assumption follows.

4. For each $t \geq 1$ we can write

$$\begin{aligned} E\big(X_{m+t}^2|\mathcal{F}_m\big) &= E\big(\varepsilon_{m+t}^2 (Y_{m+t-s}-\mu)^2|\mathcal{F}_m\big) \\ &= \sum_{i=0}^{t-s-1} \alpha_i^2 E\big(\varepsilon_{m+t}^2\varepsilon_{m+t-s-i}^2|\mathcal{F}_m\big) + \sum_{i\geq t-s} \alpha_i^2 E\big(\varepsilon_{m+t}^2\varepsilon_{m+t-s-i}^2|\mathcal{F}_m\big) \\ &\quad + 2 \sum_{j>i\geq 0} \alpha_i\alpha_j E\big(\varepsilon_{m+t}^2\varepsilon_{m+t-s-i}\varepsilon_{m+t-s-j}|\mathcal{F}_m\big) \\ &= \sigma^4 \Big(\sum_{i=0}^{t-s-1} \alpha_i^2\Big) + \sigma^2 \sum_{i\geq t-s} \alpha_i^2\varepsilon_{m+t-s-i}^2 \\ &\quad + 2\sigma^2 \sum_{j>i\geq t-s} \alpha_i\alpha_j\varepsilon_{m+t-s-i}\varepsilon_{m+t-s-j}\,. \end{aligned}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

In a similar way, for $t < l$, we get

$$
\begin{aligned}
E\big(X_{m+t}X_{m+l}|\mathcal{F}_m\big) &= E\big[\varepsilon_{m+t}\varepsilon_{m+l}(Y_{m+t-s}-\mu)(Y_{m+l-s}-\mu)|\mathcal{F}_m\big] \\
&= \sum_{i,j\geq 0}\alpha_i\alpha_j\big[E\big(\varepsilon_{m+t}\varepsilon_{m+l}\varepsilon_{m+t-s-i}\varepsilon_{m+l-s-j}|\mathcal{F}_m\big)\big] \\
&= 0.
\end{aligned}
$$

Now, putting together the last two relationships we obtain

$$
\begin{aligned}
E\big(S_k^2(m)|\mathcal{F}_m\big) &= \sigma^4\sum_{t=1}^{k}\Big(\sum_{i=0}^{t-s-1}\alpha_i^2\Big)+\sigma^2\sum_{t=1}^{k}\Big(\sum_{i\geq t-s}\alpha_i^2\varepsilon_{m+t-s-i}^2\Big) \\
&+ 2\sigma^2\sum_{t=1}^{k}\Big(\sum_{j>i\geq t-s}\alpha_i\alpha_j\varepsilon_{m+t-s-i}\varepsilon_{m+t-s-j}\Big)
\end{aligned}
$$

From the above relation and (A.9) we get

$$
\begin{aligned}
E\big(S_k^2(m)|\mathcal{F}_m\big) - ES_k^2(m) &= \sigma^2\sum_{t=1}^{k}\Big\{\sum_{i\geq t-s}\alpha_i^2\big(\varepsilon_{m+t-s-i}^2-\sigma^2\big)\Big\} \\
&+ 2\sigma^2\sum_{t=1}^{k}\Big\{\sum_{j>i\geq t-s}\alpha_i\alpha_j\varepsilon_{m+t-s-i}\varepsilon_{m+t-s-j}\Big\} \\
&= A_k(m)+B_k(m)\ .
\end{aligned}
$$

Consider the first term

$$
\begin{aligned}
E|A_k(m)| &= \sigma^2 E\Big|\sum_{t=1}^{k}\Big\{\sum_{i\geq t-s}\alpha_i^2\big(\varepsilon_{m+t-s-i}^2-\sigma^2\big)\Big\}\Big| \\
&\leq \sigma^2\sum_{t=1}^{k}\Big\{\sum_{i\geq t-s}\alpha_i^2 E\big|\varepsilon_{m+t-s-i}^2-\sigma^2\big|\Big\} \\
&\leq \sigma^2 M^2 E|\varepsilon_0^2-\sigma^2|\sum_{t=1}^{k}\Big(\sum_{i\geq t-s}\alpha^{2i}\Big) \\
&\leq \sigma^2 M^2 E|\varepsilon_0^2-\sigma^2|\sum_{t=1}^{k}\Big(\alpha^{2(t-s)}\sum_{i\geq 0}\alpha^{2i}\Big) \\
&= \frac{\sigma^2 M^2 E|\varepsilon_0^2-\sigma^2|\alpha^{2(1-s)}}{(1-\alpha^2)}\sum_{t=1}^{k}\alpha^{2(t-1)} \\
&= \frac{\sigma^2 M^2 E|\varepsilon_0^2-\sigma^2|\alpha^{2(1-s)}}{(1-\alpha^2)^2}(1-\alpha^{2k}) = O(1),\ \text{ as } k\longrightarrow\infty\ .
\end{aligned}
$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Consider now the second term

$$
\begin{aligned}
E|B_k(m)| &= 2\sigma^2 E\Big|\sum_{t=1}^{k}\Big\{\sum_{i\geq t-s}\alpha_i\varepsilon_{m+t-s-i}\Big(\sum_{j>i}\alpha_j\varepsilon_{m+t-s-j}\Big)\Big\}\Big| \\
&\leq 2\sigma^2\sum_{t=1}^{k}\Big\{\sum_{i\geq t-s}|\alpha_i|E|\varepsilon_{m+t-s-i}|\Big(\sum_{j>i}|\alpha_j|E|\varepsilon_{m+t-s-j}|\Big)\Big\} \\
&\leq 2\sigma^2 M^2(E|\varepsilon_0|)^2\sum_{t=1}^{k}\Big\{\sum_{i\geq t-s}\alpha^i\Big(\sum_{j>i}\alpha^j\Big)\Big\} \\
&= \frac{2\sigma^2\alpha M^2(E|\varepsilon_0|)^2}{(1-\alpha)}\sum_{t=1}^{k}\Big(\sum_{i\geq t-s}\alpha^{2i}\Big) \\
&= \frac{2\sigma^2\alpha M^2(E|\varepsilon_0|)^2}{(1-\alpha)}\sum_{t=1}^{k}\alpha^{2(t-s)}\Big(\sum_{i\geq 0}\alpha^{2i}\Big) \\
&= \frac{2\sigma^2\alpha M^2(E|\varepsilon_0|)^2}{(1-\alpha)(1-\alpha^2)}\sum_{t=1}^{k}\alpha^{2(t-s)} \\
&= \frac{2\sigma^2\alpha M^2(E|\varepsilon_0|)^2\alpha^{2(1-s)}}{(1-\alpha)(1-\alpha^2)}\sum_{t=0}^{k-1}\alpha^{2t} \\
&= \frac{2\sigma^2\alpha M^2(E|\varepsilon_0|)^2\alpha^{2(1-s)}}{(1-\alpha)(1-\alpha^2)}(1-\alpha^{2k}) = O(1), \quad \text{as } k\longrightarrow\infty\,.
\end{aligned}
$$

From the last two relationships we obtain $\|A_k(m)+B_k(m)\|_1 = O(1)$ and the fourth assumption is verified. As ε_i are normally distributed and the process is stationary the last assumption follows and the lemma is proved.

□

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Appendix B

The Distribution of the Maximum of the Bessel Process

In this appendix we give more extensive tables of the distribution of the maximum of the Bessel process of different orders d. Each table presents the probability

$$P(\sup_{0 \leq t \leq 1} R^{(d)}(t) < z) ,$$

where $R^{(d)}(t)$ is defined as

$$R^{(d)}(t) = \sqrt{W_1^2(t) + \ldots + W_d^2(t)} ,$$

and $W_i(t)$, $i = 1, \ldots, d$, are independent standard Brownian processes.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table B.1: Maximum of the Bessel process probability in left-hand tail, $P(\sup_{0\leq t\leq 1} R^{(d)}(t) < z)$, for $d = 2$.

	Second decimal place of z									
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
1.5	.44190	.44937	.45680	.46421	.47158	.47891	.48620	.49346	.50067	.50784
1.6	.51496	.52205	.52908	.53607	.54300	.54989	.55673	.56351	.57024	.57692
1.7	.58354	.59011	.59662	.60307	.60947	.61580	.62208	.62830	.63446	.64055
1.8	.64659	.65256	.65847	.66432	.67010	.67582	.68148	.68708	.69261	.69808
1.9	.70348	.70882	.71409	.71930	.72445	.72953	.73455	.73950	.74439	.74921
2.0	.75397	.75867	.76330	.76787	.77238	.77682	.78120	.78552	.78977	.79396
2.1	.79809	.80216	.80617	.81012	.81401	.81783	.82160	.82531	.82896	.83255
2.2	.83608	.83956	.84298	.84634	.84965	.85290	.85610	.85924	.86233	.86536
2.3	.86834	.87127	.87415	.87698	.87975	.88248	.88515	.88778	.89036	.89289
2.4	.89537	.89780	.90019	.90254	.90484	.90709	.90930	.91147	.91359	.91568
2.5	.91772	.91972	.92168	.92359	.92548	.92732	.92912	.93089	.93261	.93431
2.6	.93596	.93759	.93917	.94072	.94224	.94373	.94518	.94660	.94799	.94935
2.7	.95068	.95198	.95325	.95449	.95570	.95689	.95804	.95917	.96028	.96135
2.8	.96241	.96343	.96444	.96542	.96637	.96730	.96821	.96910	.96997	.97081
2.9	.97164	.97244	.97323	.97399	.97473	.97546	.97617	.97686	.97753	.97818
3.0	.97882	.97944	.98005	.98064	.98121	.98177	.98231	.98284	.98336	.98386
3.1	.98435	.98482	.98528	.98573	.98617	.98659	.98701	.98741	.98780	.98818
3.2	.98855	.98891	.98926	.98959	.98992	.99024	.99055	.99085	.99115	.99143
3.3	.99171	.99198	.99224	.99249	.99273	.99297	.99320	.99342	.99364	.99385
3.4	.99406	.99425	.99445	.99463	.99481	.99499	.99516	.99532	.99548	.99563
3.5	.99578	.99593	.99607	.99620	.99633	.99646	.99658	.99670	.99682	.99693
3.6	.99704	.99714	.99724	.99734	.99744	.99753	.99762	.99770	.99778	.99786
3.7	.99794	.99802	.99809	.99816	.99822	.99829	.99835	.99841	.99847	.99853
3.8	.99858	.99864	.99869	.99874	.99878	.99883	.99887	.99892	.99896	.99900
3.9	.99903	.99907	.99911	.99914	.99917	.99921	.99924	.99927	.99930	.99932
4.0	.99935	.99937	.99940	.99942	.99945	.99947	.99949	.99951	.99953	.99955
4.1	.99957	.99958	.99960	.99962	.99963	.99965	.99966	.99968	.99969	.99970
4.2	.99971	.99973	.99974	.99975	.99976	.99977	.99978	.99979	.99980	.99980
4.3	.99981	.99982	.99983	.99984	.99984	.99985	.99986	.99986	.99987	.99987
4.4	.99988	.99988	.99989	.99989	.99990	.99990	.99991	.99991	.99992	.99992
4.5	.99992	.99993	.99993	.99993	.99994	.99994	.99994	.99994	.99995	.99995
4.6	.99995	.99995	.99996	.99996	.99996	.99996	.99996	.99997	.99997	.99997
4.7	.99997	.99997	.99997	.99997	.99998	.99998	.99998	.99998	.99998	.99998
4.8	.99998	.99998	.99998	.99998	.99999	.99999	.99999	.99999	.99999	.99999
5.0	.99999	.99999	.99999	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table B.2: Maximum of the Bessel process probability in left-hand tail, $P(\sup_{0\leq t\leq 1} R^{(d)}(t) < z)$, for $d = 3$.

	Second decimal place of z									
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
1.9	.50133	.50816	.51497	.52173	.52847	.53517	.54183	.54845	.55504	.56158
2.0	.56808	.57454	.58095	.58732	.59364	.59991	.60614	.61232	.61844	.62452
2.1	.63055	.63652	.64244	.64831	.65412	.65988	.66558	.67123	.67682	.68236
2.2	.68783	.69325	.69862	.70392	.70917	.71435	.71948	.72455	.72956	.73451
2.3	.73940	.74423	.74900	.75372	.75837	.76296	.76749	.77196	.77638	.78073
2.4	.78502	.78926	.79343	.79755	.80161	.80561	.80955	.81343	.81725	.82102
2.5	.82473	.82838	.83198	.83552	.83900	.84243	.84580	.84912	.85238	.85559
2.6	.85875	.86185	.86490	.86790	.87085	.87374	.87659	.87938	.88213	.88482
2.7	.88747	.89007	.89262	.89512	.89757	.89998	.90235	.90467	.90694	.90917
2.8	.91136	.91350	.91561	.91767	.91969	.92166	.92360	.92550	.92736	.92918
2.9	.93096	.93271	.93442	.93609	.93773	.93933	.94090	.94243	.94393	.94540
3.0	.94683	.94823	.94961	.95095	.95226	.95354	.95479	.95601	.95720	.95837
3.1	.95951	.96062	.96170	.96276	.96380	.96481	.96579	.96675	.96769	.96861
3.2	.96950	.97037	.97122	.97204	.97285	.97364	.97440	.97515	.97588	.97659
3.3	.97728	.97795	.97860	.97924	.97986	.98047	.98106	.98163	.98219	.98273
3.4	.98326	.98377	.98427	.98476	.98523	.98569	.98613	.98657	.98699	.98740
3.5	.98780	.98818	.98856	.98893	.98928	.98963	.98996	.99028	.99060	.99091
3.6	.99120	.99149	.99177	.99204	.99231	.99256	.99281	.99305	.99328	.99351
3.7	.99373	.99394	.99415	.99435	.99454	.99473	.99491	.99508	.99525	.99542
3.8	.99558	.99573	.99588	.99603	.99617	.99630	.99643	.99656	.99668	.99680
3.9	.99692	.99703	.99713	.99724	.99734	.99744	.99753	.99762	.99771	.99779
4.0	.99787	.99795	.99803	.99810	.99817	.99824	.99831	.99837	.99843	.99849
4.1	.99855	.99861	.99866	.99871	.99876	.99881	.99886	.99890	.99894	.99898
4.2	.99902	.99906	.99910	.99914	.99917	.99920	.99924	.99927	.99930	.99932
4.3	.99935	.99938	.99940	.99943	.99945	.99948	.99950	.99952	.99954	.99956
4.4	.99958	.99959	.99961	.99963	.99964	.99966	.99967	.99969	.99970	.99971
4.5	.99973	.99974	.99975	.99976	.99977	.99978	.99979	.99980	.99981	.99982
4.6	.99983	.99984	.99984	.99985	.99986	.99987	.99987	.99988	.99988	.99989
4.7	.99990	.99990	.99991	.99991	.99992	.99992	.99992	.99993	.99993	.99994
4.8	.99994	.99994	.99995	.99995	.99995	.99995	.99996	.99996	.99996	.99997
4.9	.99997	.99997	.99997	.99997	.99998	.99998	.99998	.99998	.99998	.99998
5.0	.99999	.99999	.99999	.99999	.99999	.99999	.99999	.99999	1.0000	1.0000

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table B.3: Maximum of the Bessel process probability in left-hand tail, $P(\sup_{0\leq t\leq 1} R^{(d)}(t) < z)$, for $d = 4$.

	Second decimal place of z									
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
2.0	.38912	.39596	.40280	.40964	.41648	.42331	.43014	.43697	.44378	.45059
2.1	.45738	.46416	.47092	.47767	.48439	.49110	.49778	.50444	.51108	.51769
2.2	.52427	.53082	.53735	.54384	.55029	.55672	.56310	.56945	.57576	.58204
2.3	.58827	.59446	.60061	.60672	.61278	.61880	.62477	.63070	.63657	.64240
2.4	.64818	.65391	.65959	.66522	.67080	.67632	.68179	.68721	.69258	.69789
2.5	.70314	.70835	.71349	.71858	.72361	.72859	.73351	.73838	.74318	.74793
2.6	.75263	.75726	.76184	.76636	.77082	.77523	.77957	.78386	.78809	.79227
2.7	.79639	.80045	.80445	.80840	.81229	.81612	.81989	.82361	.82728	.83089
2.8	.83444	.83794	.84138	.84477	.84811	.85139	.85462	.85779	.86091	.86398
2.9	.86700	.86997	.87288	.87575	.87856	.88133	.88404	.88671	.88933	.89190
3.0	.89443	.89691	.89934	.90172	.90406	.90636	.90861	.91082	.91299	.91511
3.1	.91719	.91923	.92123	.92318	.92510	.92698	.92882	.93062	.93239	.93411
3.2	.93580	.93746	.93908	.94066	.94221	.94372	.94521	.94666	.94807	.94946
3.3	.95081	.95213	.95343	.95469	.95592	.95713	.95831	.95946	.96058	.96168
3.4	.96275	.96379	.96481	.96580	.96677	.96772	.96864	.96954	.97042	.97127
3.5	.97211	.97292	.97371	.97448	.97524	.97597	.97668	.97738	.97805	.97871
3.6	.97935	.97998	.98059	.98118	.98175	.98231	.98286	.98339	.98390	.98440
3.7	.98489	.98536	.98583	.98627	.98671	.98713	.98754	.98794	.98833	.98870
3.8	.98907	.98942	.98977	.99010	.99043	.99074	.99105	.99134	.99163	.99191
3.9	.99218	.99244	.99270	.99294	.99318	.99341	.99364	.99385	.99406	.99427
4.0	.99447	.99466	.99484	.99502	.99520	.99537	.99553	.99569	.99584	.99599
4.1	.99613	.99627	.99640	.99653	.99665	.99678	.99689	.99701	.99711	.99722
4.2	.99732	.99742	.99752	.99761	.99770	.99778	.99786	.99794	.99802	.99810
4.3	.99817	.99824	.99830	.99837	.99843	.99849	.99855	.99860	.99866	.99871
4.4	.99876	.99881	.99886	.99890	.99894	.99898	.99902	.99906	.99910	.99914
4.5	.99917	.99920	.99924	.99927	.99930	.99932	.99935	.99938	.99940	.99943
4.6	.99945	.99947	.99950	.99952	.99954	.99956	.99957	.99959	.99961	.99963
4.7	.99964	.99966	.99967	.99969	.99970	.99971	.99972	.99974	.99975	.99976
4.8	.99977	.99978	.99979	.99980	.99981	.99982	.99982	.99983	.99984	.99985
4.9	.99985	.99986	.99987	.99987	.99988	.99988	.99989	.99989	.99990	.99990
5.0	.99991	.99991	.99992	.99992	.99992	.99993	.99993	.99993	.99994	.99994
5.1	.99994	.99995	.99995	.99995	.99995	.99996	.99996	.99996	.99996	.99996
5.2	.99997	.99997	.99997	.99997	.99997	.99997	.99997	.99998	.99998	.99998
5.3	.99998	.99998	.99998	.99998	.99998	.99998	.99999	.99999	.99999	.99999
5.5	.99999	.99999	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table B.4: Maximum of the Bessel process probability in left-hand tail, $P(\sup_{0\le t\le 1} R^{(d)}(t) < z)$, for $d = 5$.

z	Second decimal place of z									
	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
2.4	.50291	.50943	.51592	.52239	.52884	.53526	.54165	.54801	.55435	.56065
2.5	.56691	.57315	.57935	.58551	.59164	.59772	.60377	.60978	.61574	.62167
2.6	.62755	.63338	.63918	.64492	.65062	.65627	.66188	.66743	.67294	.67839
2.7	.68380	.68915	.69446	.69971	.70491	.71005	.71514	.72018	.72517	.73009
2.8	.73497	.73979	.74455	.74926	.75392	.75851	.76306	.76754	.77197	.77635
2.9	.78066	.78492	.78913	.79328	.79737	.80141	.80539	.80931	.81318	.81699
3.0	.82075	.82445	.82810	.83169	.83523	.83872	.84215	.84552	.84884	.85211
3.1	.85533	.85849	.86161	.86467	.86768	.87064	.87354	.87640	.87921	.88197
3.2	.88468	.88734	.88995	.89252	.89504	.89751	.89994	.90232	.90465	.90695
3.3	.90919	.91140	.91356	.91568	.91775	.91979	.92178	.92374	.92565	.92753
3.4	.92937	.93116	.93292	.93465	.93633	.93799	.93960	.94118	.94273	.94424
3.5	.94572	.94716	.94858	.94996	.95131	.95263	.95392	.95518	.95641	.95761
3.6	.95878	.95993	.96105	.96214	.96320	.96424	.96526	.96625	.96721	.96816
3.7	.96907	.96997	.97084	.97169	.97252	.97333	.97412	.97489	.97563	.97636
3.8	.97707	.97776	.97843	.97909	.97972	.98034	.98095	.98153	.98210	.98266
3.9	.98320	.98372	.98423	.98473	.98521	.98568	.98614	.98658	.98701	.98743
4.0	.98783	.98823	.98861	.98898	.98934	.98969	.99003	.99036	.99068	.99099
4.1	.99129	.99158	.99187	.99214	.99241	.99267	.99292	.99316	.99339	.99362
4.2	.99384	.99405	.99426	.99446	.99465	.99484	.99502	.99520	.99537	.99553
4.3	.99569	.99585	.99600	.99614	.99628	.99642	.99655	.99667	.99679	.99691
4.4	.99703	.99714	.99724	.99734	.99744	.99754	.99763	.99772	.99781	.99789
4.5	.99797	.99805	.99812	.99819	.99826	.99833	.99839	.99846	.99852	.99857
4.6	.99863	.99868	.99874	.99879	.99883	.99888	.99893	.99897	.99901	.99905
4.7	.99909	.99913	.99916	.99920	.99923	.99926	.99929	.99932	.99935	.99938
4.8	.99940	.99943	.99945	.99947	.99950	.99952	.99954	.99956	.99958	.99959
4.9	.99961	.99963	.99965	.99966	.99968	.99969	.99970	.99972	.99973	.99974
5.0	.99975	.99976	.99978	.99979	.99980	.99980	.99981	.99982	.99983	.99984
5.1	.99985	.99985	.99986	.99987	.99987	.99988	.99989	.99989	.99990	.99990
5.2	.99991	.99991	.99992	.99992	.99992	.99993	.99993	.99994	.99994	.99994
5.3	.99995	.99995	.99995	.99995	.99996	.99996	.99996	.99996	.99997	.99997
5.4	.99997	.99997	.99997	.99998	.99998	.99998	.99998	.99998	.99998	.99999
5.5	.99999	.99999	.99999	.99999	.99999	.99999	.99999	.99999	.99999	1.0000

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table B.5: Maximum of the Bessel process probability in left-hand tail, $P(\sup_{0\le t\le 1} R^{(d)}(t) < z)$, for $d = 6$.

	Second decimal place of z									
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
2.5	.43174	.43829	.44483	.45136	.45789	.46441	.47092	.47742	.48391	.49039
2.6	.49685	.50329	.50971	.51612	.52250	.52886	.53520	.54151	.54780	.55406
2.7	.56029	.56648	.57265	.57879	.58489	.59095	.59698	.60298	.60893	.61485
2.8	.62072	.62656	.63235	.63810	.64381	.64947	.65509	.66066	.66619	.67167
2.9	.67710	.68248	.68781	.69309	.69833	.70351	.70864	.71372	.71875	.72372
3.0	.72864	.73351	.73833	.74309	.74779	.75245	.75705	.76159	.76608	.77051
3.1	.77489	.77921	.78348	.78769	.79185	.79595	.80000	.80399	.80793	.81181
3.2	.81564	.81941	.82313	.82679	.83040	.83396	.83746	.84090	.84430	.84764
3.3	.85093	.85416	.85735	.86048	.86356	.86659	.86957	.87249	.87537	.87820
3.4	.88098	.88371	.88639	.88903	.89161	.89415	.89665	.89909	.90149	.90385
3.5	.90616	.90843	.91065	.91283	.91497	.91707	.91912	.92114	.92311	.92504
3.6	.92693	.92879	.93060	.93238	.93412	.93583	.93749	.93913	.94072	.94228
3.7	.94381	.94530	.94676	.94819	.94959	.95095	.95228	.95359	.95486	.95610
3.8	.95732	.95850	.95966	.96079	.96189	.96297	.96402	.96505	.96604	.96702
3.9	.96797	.96890	.96980	.97068	.97154	.97238	.97320	.97399	.97477	.97552
4.0	.97626	.97697	.97767	.97835	.97901	.97965	.98027	.98088	.98147	.98205
4.1	.98261	.98315	.98368	.98420	.98470	.98518	.98566	.98612	.98656	.98699
4.2	.98741	.98782	.98822	.98860	.98898	.98934	.98969	.99004	.99037	.99069
4.3	.99100	.99130	.99160	.99188	.99216	.99242	.99268	.99293	.99318	.99341
4.4	.99364	.99386	.99408	.99428	.99448	.99468	.99487	.99505	.99522	.99539
4.5	.99556	.99572	.99587	.99602	.99617	.99630	.99644	.99657	.99670	.99682
4.6	.99693	.99705	.99716	.99726	.99737	.99746	.99756	.99765	.99774	.99783
4.7	.99791	.99799	.99807	.99814	.99821	.99828	.99835	.99841	.99847	.99853
4.8	.99859	.99865	.99870	.99875	.99880	.99885	.99889	.99894	.99898	.99902
4.9	.99906	.99910	.99914	.99917	.99920	.99924	.99927	.99930	.99933	.99936
5.0	.99938	.99941	.99943	.99946	.99948	.99950	.99952	.99954	.99956	.99958
5.1	.99960	.99962	.99963	.99965	.99966	.99968	.99969	.99971	.99972	.99973
5.2	.99974	.99975	.99976	.99978	.99979	.99979	.99980	.99981	.99982	.99983
5.3	.99984	.99984	.99985	.99986	.99986	.99987	.99988	.99988	.99989	.99989
5.4	.99990	.99990	.99991	.99991	.99992	.99992	.99992	.99993	.99993	.99994
5.5	.99994	.99994	.99994	.99995	.99995	.99995	.99995	.99996	.99996	.99996
5.6	.99996	.99997	.99997	.99997	.99997	.99997	.99997	.99998	.99998	.99998
5.7	.99998	.99998	.99998	.99998	.99998	.99999	.99999	.99999	.99999	.99999
5.8	.99999	.99999	.99999	.99999	.99999	.99999	.99999	.99999	1.0000	1.0000

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table B.6: Maximum of the Bessel process probability in left-hand tail, $P(\sup_{0 \le t \le 1} R^{(d)}(t) < z)$, for $d = 7$.

	Second decimal place of z									
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
2.8	.50184	.50820	.51454	.52086	.52716	.53344	.53970	.54593	.55213	.55832
2.9	.56447	.57059	.57668	.58274	.58877	.59477	.60073	.60665	.61254	.61839
3.0	.62420	.62997	.63570	.64139	.64703	.65263	.65819	.66371	.66918	.67460
3.1	.67997	.68530	.69058	.69581	.70100	.70613	.71121	.71624	.72122	.72615
3.2	.73103	.73585	.74063	.74535	.75001	.75462	.75918	.76369	.76814	.77254
3.3	.77688	.78117	.78540	.78958	.79371	.79778	.80179	.80575	.80966	.81351
3.4	.81731	.82106	.82475	.82838	.83197	.83549	.83897	.84239	.84576	.84908
3.5	.85235	.85556	.85872	.86183	.86489	.86790	.87085	.87376	.87662	.87943
3.6	.88219	.88490	.88756	.89018	.89275	.89527	.89775	.90018	.90256	.90490
3.7	.90720	.90945	.91166	.91382	.91595	.91803	.92007	.92207	.92402	.92594
3.8	.92782	.92966	.93147	.93323	.93496	.93665	.93830	.93992	.94151	.94306
3.9	.94457	.94605	.94750	.94892	.95030	.95165	.95298	.95427	.95553	.95676
4.0	.95796	.95914	.96029	.96141	.96250	.96356	.96460	.96562	.96661	.96758
4.1	.96852	.96943	.97033	.97120	.97205	.97288	.97369	.97447	.97524	.97599
4.2	.97671	.97742	.97811	.97878	.97943	.98006	.98068	.98128	.98187	.98243
4.3	.98299	.98352	.98404	.98455	.98505	.98552	.98599	.98644	.98688	.98731
4.4	.98772	.98812	.98851	.98889	.98926	.98962	.98996	.99030	.99063	.99094
4.5	.99125	.99154	.99183	.99211	.99238	.99264	.99290	.99314	.99338	.99361
4.6	.99384	.99405	.99426	.99447	.99466	.99485	.99504	.99521	.99539	.99555
4.7	.99571	.99587	.99602	.99616	.99631	.99644	.99657	.99670	.99682	.99694
4.8	.99705	.99717	.99727	.99738	.99747	.99757	.99766	.99775	.99784	.99792
4.9	.99800	.99808	.99815	.99823	.99830	.99836	.99843	.99849	.99855	.99861
5.0	.99866	.99872	.99877	.99882	.99886	.99891	.99895	.99900	.99904	.99908
5.1	.99912	.99915	.99919	.99922	.99925	.99928	.99931	.99934	.99937	.99940
5.2	.99942	.99945	.99947	.99949	.99952	.99954	.99956	.99958	.99959	.99961
5.3	.99963	.99965	.99966	.99968	.99969	.99971	.99972	.99973	.99974	.99976
5.4	.99977	.99978	.99979	.99980	.99981	.99982	.99982	.99983	.99984	.99985
5.5	.99986	.99986	.99987	.99988	.99988	.99989	.99989	.99990	.99990	.99991
5.6	.99991	.99992	.99992	.99993	.99993	.99993	.99994	.99994	.99994	.99995
5.7	.99995	.99995	.99996	.99996	.99996	.99996	.99996	.99997	.99997	.99997
5.8	.99997	.99997	.99998	.99998	.99998	.99998	.99998	.99998	.99998	.99999
6.0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table B.7: Maximum of the Bessel process probability in left-hand tail, $P(\sup_{0\leq t\leq 1} R^{(d)}(t) < z)$, for $d = 8$.

	Second decimal place of z									
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
3.0	.51514	.52140	.52765	.53387	.54008	.54626	.55242	.55855	.56465	.57073
3.1	.57678	.58280	.58879	.59474	.60066	.60655	.61240	.61822	.62399	.62973
3.2	.63543	.64109	.64671	.65229	.65782	.66331	.66876	.67416	.67951	.68482
3.3	.69009	.69530	.70047	.70559	.71065	.71567	.72064	.72556	.73043	.73525
3.4	.74001	.74472	.74938	.75399	.75855	.76305	.76750	.77189	.77623	.78052
3.5	.78476	.78894	.79306	.79714	.80115	.80512	.80903	.81289	.81669	.82044
3.6	.82413	.82778	.83136	.83490	.83838	.84181	.84519	.84852	.85179	.85501
3.7	.85818	.86130	.86437	.86738	.87035	.87327	.87614	.87895	.88172	.88445
3.8	.88712	.88974	.89232	.89486	.89734	.89978	.90218	.90453	.90683	.90909
3.9	.91131	.91349	.91562	.91771	.91976	.92177	.92374	.92566	.92755	.92940
4.0	.93121	.93299	.93472	.93642	.93808	.93971	.94130	.94286	.94438	.94587
4.1	.94733	.94875	.95014	.95150	.95283	.95413	.95540	.95664	.95785	.95903
4.2	.96018	.96131	.96240	.96348	.96452	.96554	.96654	.96751	.96845	.96938
4.3	.97028	.97115	.97201	.97284	.97365	.97444	.97521	.97596	.97669	.97740
4.4	.97809	.97876	.97942	.98005	.98067	.98128	.98186	.98243	.98299	.98353
4.5	.98405	.98456	.98505	.98553	.98600	.98645	.98690	.98732	.98774	.98814
4.6	.98853	.98891	.98928	.98964	.98999	.99032	.99065	.99096	.99127	.99157
4.7	.99186	.99214	.99241	.99267	.99292	.99317	.99341	.99364	.99386	.99408
4.8	.99429	.99449	.99469	.99488	.99506	.99524	.99541	.99558	.99574	.99589
4.9	.99604	.99619	.99633	.99646	.99660	.99672	.99684	.99696	.99708	.99719
5.0	.99729	.99740	.99749	.99759	.99768	.99777	.99786	.99794	.99802	.99810
5.1	.99817	.99824	.99831	.99838	.99844	.99850	.99856	.99862	.99867	.99873
5.2	.99878	.99883	.99888	.99892	.99896	.99901	.99905	.99909	.99912	.99916
5.3	.99920	.99923	.99926	.99929	.99932	.99935	.99938	.99940	.99943	.99945
5.4	.99948	.99950	.99952	.99954	.99956	.99958	.99960	.99962	.99963	.99965
5.5	.99966	.99968	.99969	.99971	.99972	.99973	.99974	.99976	.99977	.99978
5.6	.99979	.99980	.99981	.99982	.99982	.99983	.99984	.99985	.99985	.99986
5.7	.99987	.99987	.99988	.99989	.99989	.99990	.99990	.99991	.99991	.99992
5.8	.99992	.99992	.99993	.99993	.99993	.99994	.99994	.99994	.99995	.99995
5.9	.99995	.99996	.99996	.99996	.99996	.99996	.99997	.99997	.99997	.99997
6.0	.99997	.99997	.99998	.99998	.99998	.99998	.99998	.99998	.99998	.99998
6.1	.99999	.99999	.99999	.99999	.99999	.99999	.99999	.99999	.99999	.99999
6.2	.99999	.99999	.99999	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.