MAXIMUM LIKELIHOOD ESTIMATION OF AN UNKNOWN CHANGE-POINT IN THE PARAMETERS OF A MULTIVARIATE GAUSSIAN SERIES WITH APPLICATIONS TO ENVIRONMENTAL MONITORING

By

PENGYU LIU

A dissertation/thesis submitted in partial fulfillment of
the requirements for the degree of

DOCTOR OF PHILOSOPHY

WASHINGTON STATE UNIVERSITY
Department of Mathematics

MAY 2010

To the Faculty of Washington State University:

The members of the Committee appointed to examine the dissertation/thesis of PENGYU LIU find it satisfactory and recommend that it be accepted.

V. Krishna Jandhyala, Ph.D., Chair

Stergios B. Fotopoulos, Ph.D.

Nairanjana Dasgupta, Ph.D.

ACKNOWLEDGMENT

I would like to thank all people who have helped and inspired me during my doctoral study.

I am especially thankful to my advisor, Professor Jandhyala, whose encouragement, guidance and support from the initial to the final level enabled me to develop an understanding of the subject. His insight to change-point problem is second to none. Besides, he sets an example of a world-class researcher for his rigor and passion on research.

Professor Fotopoulos and Professor Dasgupta deserve special thanks as my thesis committee members and advisors. I would like to thank them for their encouragement, insightful comments. I especially thank Professor Dasgupta for helping me to find a position at Axio Research and apply my expertise to the real-world problems.

My sincere thanks go to Axio Research, especially CEO Lee Hooks and my supervisor Alice Birnbaum, who granted me so much convenience so that I was able to finish my thesis on time.

I thank my friends: Qinan Bao, Yun Wang, Junjian Sun, Limin Yang, Jimmy Burk, Lois Kwon, Xiaohui Huang, Jing You, Chris Gast, Ya Zhang, Juntao Yu, and all that had helped and supported me during my study.

Last but not the least, I would like to thank my family: my parents Juxing Liu and Jiaju Gu and my sister Pengcheng Liu, for supporting me financially and spiritually throughout my life.

Pengyu Liu

MAXIMUM LIKELIHOOD ESTIMATION OF AN UNKNOWN CHANGE-POINT IN THE PARAMETERS OF A MULTIVARIATE GAUSSIAN SERIES WITH APPLICATIONS TO ENVIRONMENTAL MONITORING

Abstract

by Pengyu Liu, Ph.D.
Washington State University
May 2010

Chair: V. Krishna Jandhyala

The computable expressions for the asymptotic distribution of the change-point maximum likelihood estimator (mle) were derived when a change occurred in the mean and covariance matrix at an unknown point of a sequence of independently distributed multivariate Gaussian series. The derivation was based on ladder heights of Gaussian random walks hitting the half-line. We then demonstrated that change in a single parameter or change-point analysis in a univariate series can be derived as special cases. A simulation study was carried out to investigate the robustness of the asymptotic distribution to departure from normality, the sample size, location of change-point and amount of change under the multivariate and univariate case. The comparison of the asymptotic mle with Cobb's conditional MLE and Bayesian estimation method using non-informative prior and conjugate prior was also carried out in the simulation study. The asymptotic distribution of the change-point mle was

used to compute the confidence interval of the change-point of the stream flows at Northern Quebec Labrador Region and zonal annual mean temperature deviations.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Dedication

This dissertation/thesis is dedicated to my mother and father

who provided both emotional and financial support

1 INTRODUCTION

Classic change-point methods involve two fundamental inferential problems, detection and estimation. Under the maximum likelihood based approach, the detection part is addressed through likelihood ratio statistics and their asymptotic sampling distributions. The estimation part started with the point estimate of the change-point from the detection part. Even though asymptotic distributions of change detection statistics are non-standard, much progress has been made in this regard, at least for the case of detecting a single unknown change-point in a time series. The specific scenarios include changes in the parameters of univariate and multivariate exponential families, multiple linear regression models, autoregressive models, and even long range dependent time series models. Chapter 2 gave a comprehensive review about the change-point analysis using maximum likelihood method and other alternative methods. The maximum likelihood ratio statistics for change-point detection is also derived for the estimation problems in our study.

Tackling the estimation problem, we derive in this study the computable expressions for the distribution of the change-point mle when a change occurs in the mean and/or variance/covariance of a univariate or multivariate Gaussian series. The derived asymptotic distribution is quite elegant and can be computed in a simple and straightforward manner. For the Gaussian case, Fotopoulos et al. (2009) demonstrates that the second suggested approximation in Jandhyala and Fotopoulos (1999) is the exact solution to the estimation of change-point mle. In Chapter 3, the asymptotic distribution was derived for the change-point mle for change in mean only in multivariate Gaussian series. Chapter 4 derived the case for change in both mean and covariance. As the estimation requires computing the distribution of a linear combination of non-central chi-square random variables, Chapter 4 also discussed this issue for presenting the algorithmic procedure for estimating change-point mle.

It should be noted that the parameters of the distribution before and after the change-point are assumed known. However, this should not pose difficulties, since Hinkley (1970) has shown that the asymptotic distribution of the change-point mle when the parameters are unknown is equivalent to that when the parameters are known. From a practical point of view this asymptotic equivalence result is extremely important. In practice, apart from the change-point being unknown, the parameters before and after the change-point also invariably remain unknown. The problem of deriving the distribution of the change-point mle when the parameters are unknown is the one that practitioners would be most interested as opposed to the distribution of the change-point mle for the case when the parameters are known. There is no apriori reason to believe that the distributions of the change-point mle for the known and unknown cases be asymptotically equivalent. It is in this sense that the asymptotic equivalence result of Hinkley (1972) plays a key role for practitioners. One only needs to examine whether this asymptotic property holds well for reasonable sample sizes, and for this we carried out a simulation study in Chapter 5, where the asymptotic distributions are computed under different combination of sample size, location of change-point, dimension of the observations, and the choice of estimating parameters before and after the change-point.

Since the solution derived in the paper assumes Gaussianity, we also explored the robustness of this computable expression when the series deviates from Gaussianity. If the derived result is indeed robust to such departures, then it can be applied more widely than merely Gaussian processes. While a simulation study covering a wide class of non-Gaussian families of distributions may be of interest for practitioners, in this study, a limited robustness study is pursued by performing large scale simulations wherein the error terms follows the univariate or multivariate t-distribution. The degrees of freedom were changed from being small to large, so that we are able to observe how the asymptotic distributions behave as the underlying distribution approaches the Gaussianity. The simulation for univariate and multivariate case are carried out in Chapter 5.

Hinkley's (1972) approach to deriving distribution of the change-point mle is perceived as the unconditional approach in the literature. Against this, Cobb (1978) proposed a conditional approach to the distribution of the change-point mle, wherein the distribution of the mle is derived by conditioning upon sufficient information on either side of the unknown change-point. It is relevant to compare the conditional and unconditional distributions in terms of their performance, including robustness properties. Thus Cobb's conditional distribution is also included in the simulation study. As pointed out by Cobb (1978), since the conditional distribution of the change-point mle can also be interpreted as the Bayesian posterior for the change-point under a uniform prior on the unknown change-point, the comparisons between the two distributions have a broader appeal than what might appear at first glance. The simulation study for Cobb's method in Chapter 5 also includes the cases for known and unknown parameters.

In Chapter 6, we apply the methodology derived in Chapter 3 and 4 to multivariate analysis of hydrological and the climatology data. The hydrological data, previously analyzed with Bayesian method using conjugate priors by Perreault et al. (2000), represents the average spring stream flows of six rivers during 1957-1995 in the northern Québec Labrador region. The multivariate change-point analysis shows that a significant increase in mean stream flow has occurred 1984. The climatology data, which was provided by Angell (2009), represents the mean annual air temperature for surface and upper layers (850 – 300, 300 – 100 and 100 – 50 mb) from 1958 to 2008 at north and south polar. The analysis showed that at the south polar, a cooling effect has occurred at the lower stratosphere at 1981, where the change is in both mean and covariance matrix, and a warming effect at the surface temperature at 1976, where only change in mean has happened. At the north polar, a cooling effect at the lower stratosphere and a warming effect at the surface occurred at 1988, and the change has happened in mean only.

2 LITERATURE REVIEW

Maximum likelihood estimation of an unknown change-point first begins with obtaining the mle as a point estimate. Interval estimates of any desired level, which are preferred over point estimates can be constructed around the mle provided distribution theory for the mle is available. Hawkins (1977) and James et Al. (1987) studied change-point detection in the series following independent univariate normal distribution with possible change of mean. Kim and Siegnumd, D. (1989) and Chu and White (1992) developed the detection for change in simple linear regression for slope and intercept. Worsley (1988) and Henderson (1990) used likelihood ratio test for the change in hazard ratio. Worsley and Srivastava (1986) tested change in mean in multivariate normal series. As for real life applications of change-point detectioin, see Braun and Müller (1998) for application of change point methods in DNA segmentation and bioinformatics; Fearnhead (2006), Ruggieri et al (2009) for applications in geology; Perreault et al (2000a, 2000b) for applications in hydrology; Jarušková (1996) for applications in meteorology; Fealy and Sweeney (2005), DeGaetano (2006) for applications in climatology; Kaplan and Shishkin (2000), Lebarbier (2005) for applications in signal processing; Andrews (1993), Hansen (2000) for applications in econometrics; and Lai (1995), Wu et al. (2005), Zou et al. (2009) for applications in statistical process control. However, distribution theory for a change-point mle can be analytically intractable, particularly when no smoothness conditions are assumed regarding the amount of change. Convincing arguments have not yet been made in the literature regarding the appropriateness of imposing smoothness conditions on the amount of change. As a consequence of its intractability, only a few computationally useful results for the distribution of the change-point mle under abrupt change have been developed. For univariate models, the distribution theory and computational procedures have been derived by Jandhyala and Fotopoulos (1999) for change in mean only of a normal distribution, Jandhyala et al (2002) for change in variance only of a normal distribution and Jandhyala and Fotopoulos (2007) for estimating simultaneous change in both mean and variance of the univariate normal distribution. Earlier Jandhyala et al (1999) computed asymptotic distribution of the change-point mle for Weibull models and applied it to estimate change in minimum temperatures at Uppsala, Sweden. For multivariate models, Jandhyala et al (2008) derived the estimation for change in mean vector only of a multivariate normal distribution. However, distribution theory for change-point MLEs in the other parameters (covariance matrix only, or both mean and covariance matrix) of multivariate models, Gaussian or otherwise, has not yet been derived in the literature. Similarly, the methodology has not been developed for estimating changes in the parameters of regression models, and thus one cannot yet handle changes in polynomial trends under the MLE approach. Note that, as mentioned previously, the detection part has been developed for all these situations and it is the distribution theory of the change-point MLE that poses greater analytical difficulties. In this sense, this project makes an important progress by considering the problem of estimating change in both the mean vector and the covariance matrix of a multivariate normal distribution by the MLE method, and then by applying it to the analysis of zonal temperature deviations from surface to lower stratosphere layer.

In contrast, advances in the Bayesian approach to change-point methodology have been occurring at a faster pace. Ever since Markov Chain Monte Carlo (MCMC) methods were seen as a tool for overcoming the computational complexities in Bayesian analysis, there has been rapid progress in the overall development of this important methodological tool, and advances in Bayesian change-point analysis have not lagged behind. The main advantage of the Bayesian approach to the change-problem is that both detection and estimation parts of the problem are solved simultaneously once posterior distribution of the unknown change-point is made available, mainly because all inferences about the unknown change-point are made from the posterior distribution. Consequently, with recent advances in the methodology, the Bayesian approach to change-point analysis is able to provide inferential methods ranging from simple to complex situations, some of which include change in mean and/or variance of the univariate normal distribution (Perreault et al 1999, Perreault et al 2000a, 2000b), change in the mean vector of a multivariate normal distribution (Perreault et al 2000), change in mean and/or covariance of a multivariate Guassian series (Son and Kim 2005), single change in the parameters of a multiple linear regression model (Seidou et al 2007), nonlinear change (Schleip et al. 2009), and also the more complex case of estimating multiple change-points (Barry and Hartigan 1993, Fearnhead, 2005, 2006; Seidou and Ouarda 2007). Carlin et al.

(1992) proposed hierarchical Bayesian change-points model using the Gibbs sampler with application to changing regressions, Poisson process and Markov chains.

Clearly, developments in the mle methodology under abrupt changes lag behind its Bayesian counterpart. As tools for statistical modeling and analysis, it is desirable that both methods be available for practitioners. As such, data analysis will benefit from having a choice of competing methods for any given scenario and there is no need to curtail advances in either of the two approaches. It is entirely possible that one of the methods may be more suitable for the analysis of a particular data series, and seen from this perspective, it is difficult to argue against further advancements in the mle methodology for change-point analysis.

Asymptotic distribution theory for the change-point mle in the abrupt case was first initiated by Hinkley (1970, 1971, 1972). While Hinkley (1970) derived the asymptotic theory for the change-point mle in a general set-up, the distribution was not in a computable form primarily due to the technical difficulties in nature. It turned out that Hinkley (1970) computed the distribution for change in the mean of a normal distribution only through certain approximations. While Hu and Rukhin (1995) provided a lower bound for the probability of the mle being in error of capturing the true change-point, Jandhyala and Fotopoulos (1999) and Fotopoulos and Jandhyala (2001) derived upper and lower bounds and also suggested two approximations for the asymptotic distribution of the change-point mle. Similarly, Borovkov (1999) also provided only upper and lower bounds for the distribution of the change-point mle. Computable expressions for the asymptotic distribution of the change-point mle was derived in for multivariate Gaussian series with change in mean vector only by Jandhyala et al. (2008) and Fotopoulos et al. (2009), and in exponential distributions by Fotopoulos et al. (2001) . Thus, despite the attempts of various authors, the problem of deriving computable expressions for the asymptotic distribution of the change-point mle remained unsolved to the multivariate Gaussian series with the change occurred in both mean and covariance matrix.

2.1 Change-point Detection for Mean and/or Covariance

In this project, our main goal is one of advancing the mle method for the estimation of an unknown change-point in the mean vector and/or covariance matrix of a sequence of multivariate normal observations. In Section 2.1, the procedures for the change-point detection for change in both mean and covariance/variance were presented following Jandhyala and Fotopoulos (1999). The detection results would be applied in the change-point estimation and the applications.

2.1.1 Change in both mean and covariance/variance

Let $Y_1, Y_2, \dots, Y_n$ be a sequence of time series valued independent random vectors such that $\mathrm{Y_i} \in \mathbb{R}^{\mathrm{d}}, \mathrm{i} = 1, \dots, \mathrm{n}$. Furthermore, for each $\mathrm{i} = 1, 2, \dots, \mathrm{n}$, let Y_i follow the multivariate Gaussian distribution with mean vector μ and variance-covariance matrix Σ. Without loss of generality, we let the parameter set to be (μ, Σ), and the corresponding multivariate Gaussian density function to be $f(\cdot; \mu, \Sigma)$. Then, under the classical change point model in which the mean vector μ changes from an initial value μ_{0,τ_n} to a subsequent value μ_{1,τ_n}, and the covariance matrix Σ changes from Σ_{0,τ_n} to Σ_{1,τ_n},at some unknown change-point $\tau_n \in \{1,2,\dots,n-1\}$. For the purposes of this section, it will be assumed that the parameters μ_{0,τ_n}, μ_{1,τ_n}, Σ_{0,τ_n} and Σ_{1,τ_n} are all unknown. Under the change point model, one has

$$Y_i \sim \begin{cases} f\left(\cdot;\mu_{0,\tau_n},\Sigma_{0,\tau_n}\right), & i=1, \quad \dots,\tau_n \\ f\left(\cdot;\mu_{1,\tau_n},\Sigma_{1,\tau_n}\right), & i=\tau_n+1,\dots,n \end{cases} \tag{2.1}$$

where $\tau_n \in \{1,\dots,n-1\}$.

On the other hand, when there is no change occurs in the model, a single parameter set is applicable throughout the sampling period such that under the no change model one has

$$Y_i \sim f\left(\cdot;\mu_{0,n},\Sigma_{0,n}\right), \quad i=1, \quad \dots,n \tag{2.2}$$

One is confronted with having to decide whether the given data set can be modeled by the no change model (2.2) , or by the change point model (2.1) with a change occurring in the mean vector and covariance matrix at an unknown change point τ_n. Thus, the statistical problem is one of carrying out a test of the following hypotheses:

H0: The data conforms to no change model (2.2)

Against H1: The data conforms to change point model (2.1)

While a number of authors have addressed the above hypothesis testing problem at the univariate level under various approaches, the likelihood ratio approach to the multivariate version of the problem has been adequately addressed by Csörgő and Horváth (1997) , Chen and Gupta (2000), and also earlier by Worsley and Srivastava (1986). While the derivation of the statistic is fairly straight forward, its asymptotic distribution is nonstandard and requires careful analytical arguments. Below, we outline the details of deriving the likelihood ratio statistic and then state its asymptotic distribution. Under multivariate Gaussianity, given $\tau_n = t$, where $t \in \{1, \dots, n-1\}$, the likelihood function under the change-point model (2.1) is

$$L_1(t) = f\left(Y_1, \dots, Y_t; \mu_{0,t}, \Sigma_{0,t}\right) f\left(Y_{t+1}, \dots, Y_n; \mu_{1,t}, \Sigma_{1,t}\right) \tag{2.3}$$

$$= \prod_{i=1}^{t} f(Y_i; \mu_{0,t}, \Sigma_{0,\mathrm{t}}) \prod_{i=t+1}^{n} f(Y_i; \mu_{1,t}, \Sigma_{1,\mathrm{t}})$$

$$= \prod_{i=1}^{t} \frac{1}{(2\pi)^{\frac{d}{2}} \left|\Sigma_{0,\mathrm{t}}\right|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\left(Y_i - \mu_{0,t}\right)^T \Sigma_{0,\mathrm{t}}^{-1}\left(Y_i - \mu_{0,t}\right)\right)$$

$$\cdot \prod_{i=t+1}^{n} \frac{1}{(2\pi)^{\frac{d}{2}} \left|\Sigma_{1,\mathrm{t}}\right|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\left(Y_i - \mu_{1,t}\right)^T \Sigma_{1,\mathrm{t}}^{-1}\left(Y_i - \mu_{1,t}\right)\right)$$

The parameter estimates under the change-point model (2.1) are

$$\hat{\mu}_{0,t} = \frac{1}{t}\sum_{i=1}^{t} Y_i\,, \qquad \hat{\mu}_{1,t} = \frac{1}{n-t}\sum_{i=t+1}^{n} Y_i \tag{2.4}$$

$$\hat{\Sigma}_{0,t} = \frac{1}{t}\sum_{i=1}^{t}\left(Y_i - \hat{\mu}_{0,t}\right)\left(Y_i - \hat{\mu}_{0,t}\right)^T$$

$$\hat{\Sigma}_{1,t} = \frac{1}{n-t}\sum_{i=t+1}^{n}\left(Y_i - \hat{\mu}_{1,t}\right)\left(Y_i - \hat{\mu}_{1,t}\right)^T$$

So the estimate for the log-likelihood function is

$$\log \hat{L}_1(t) = -\frac{t}{2}\log\left|\hat{\Sigma}_{0,t}\right| - \frac{1}{2}\sum_{i=1}^{t}\left(Y_i - \hat{\mu}_{0,t}\right)^T \hat{\Sigma}_{0,t}^{-1}\left(Y_i - \hat{\mu}_{0,t}\right) \tag{2.5}$$

$$- \frac{td}{2}\log 2\pi - \frac{n-t}{2}\log\left|\hat{\Sigma}_{1,t}\right|$$

$$- \frac{1}{2}\sum_{i=t+1}^{n}\left(Y_i - \hat{\mu}_{1,t}\right)^T \hat{\Sigma}_{1,t}^{-1}\left(Y_i - \hat{\mu}_{1,t}\right) - \frac{(n-t)d}{2}\log 2\pi$$

$$= -\frac{t}{2}\log\left|\hat{\Sigma}_{0,t}\right| - \frac{n-t}{2}\log\left|\hat{\Sigma}_{1,t}\right|$$

$$- \frac{1}{2}\sum_{i=1}^{t} tr\left[\hat{\Sigma}_{0,t}^{-1}\left(Y_i - \hat{\mu}_{0,t}\right)\left(Y_i - \hat{\mu}_{0,t}\right)^T\right]$$

$$- \frac{1}{2}\sum_{i=t+1}^{n} tr\left[\hat{\Sigma}_{1,t}^{-1}\left(Y_i - \hat{\mu}_{1,t}\right)\left(Y_i - \hat{\mu}_{1,t}\right)^T\right] - \frac{nd}{2}\log 2\pi$$

$$= -\frac{t}{2}\log\left|\hat{\Sigma}_{0,t}\right| - \frac{n-t}{2}\log\left|\hat{\Sigma}_{1,t}\right| - \frac{nd}{2}\log 2\pi - \frac{1}{2}tr\left\{\hat{\Sigma}_{0,t}^{-1}\mathrm{t}\hat{\Sigma}_{0,\mathrm{t}}\right\}$$

$$- \frac{1}{2}tr\left\{\hat{\Sigma}_{1,t}^{-1}(n-t)\hat{\Sigma}_{1,\mathrm{t}}\right\}$$

$$= -\frac{t}{2}\log\left|\hat{\Sigma}_{0,\mathrm{t}}\right| - \frac{n-t}{2}\log\left|\hat{\Sigma}_{1,\mathrm{t}}\right| - \frac{nd}{2}\log 2\pi - \frac{1}{2}t - \frac{1}{2}(n-t)$$

$$= -\frac{t}{2}\log\left|\hat{\Sigma}_{0,\mathrm{t}}\right| - \frac{n-t}{2}\log\left|\hat{\Sigma}_{1,\mathrm{t}}\right| - \frac{nd}{2}\log 2\pi - \frac{n}{2}$$

Similarly, the likelihood function under model (2.2) is

$$L_0(t) = f\left(Y_1, \dots, Y_n; \mu_{0,n}, \Sigma_{0,\mathrm{n}}\right)$$

$$= \prod_{i=1}^{n} f\left(Y_i; \mu_{0,n}, \Sigma_{0,\mathrm{n}}\right)$$

$$= \prod_{i=1}^{n} \frac{1}{(2\pi)^{\frac{d}{2}}\left|\Sigma_{0,\mathrm{n}}\right|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\left(Y_i - \mu_{0,n}\right)^T \Sigma_{0,\mathrm{n}}^{-1}\left(Y_i - \mu_{0,n}\right)\right)$$

The parameter estimates are

$$\hat{\mu}_{0,n} = \frac{1}{n}\sum_{i=1}^{n} Y_i, \quad \hat{\Sigma}_{0,n} = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{\mu}_{0,n}\right)\left(Y_i - \hat{\mu}_{0,n}\right)^T \qquad (2.6)$$

So the estimate for the log-likelihood function is

$$\log \hat{L}_0(t) = -\frac{n}{2}\log\left|\hat{\Sigma}_{0,n}\right| - \frac{1}{2}\sum_{i=1}^{n}\left(Y_i - \hat{\mu}_{0,n}\right)^T \hat{\Sigma}_{0,n}^{-1}\left(Y_i - \hat{\mu}_{0,n}\right) \qquad (2.7)$$

$$- \frac{nd}{2}\log 2\pi$$

$$= -\frac{n}{2}\log\left|\hat{\Sigma}_{0,n}\right| - \frac{1}{2}\sum_{i=1}^{n} tr\left[\hat{\Sigma}_{0,n}^{-1}\left(Y_i - \hat{\mu}_{0,n}\right)\left(Y_i - \hat{\mu}_{0,n}\right)^T\right]$$

$$-\frac{nd}{2}\log 2\pi$$

$$= -\frac{n}{2}\log\left|\hat{\Sigma}_{0,n}\right| - \frac{nd}{2}\log 2\pi$$

$$-\frac{1}{2} tr\left\{\hat{\Sigma}_{0,n}^{-1}\sum_{i=1}^{n}\left(Y_i - \hat{\mu}_{0,n}\right)\left(Y_i - \hat{\mu}_{0,n}\right)^T\right\}$$

$$= -\frac{n}{2}\log\left|\hat{\Sigma}_{0,n}\right| - \frac{nd}{2}\log 2\pi - \frac{1}{2} tr\left\{\hat{\Sigma}_{0,n}^{-1} n\hat{\Sigma}_0\right\}$$

$$= -\frac{n}{2}\log\left|\hat{\Sigma}_{0,n}\right| - \frac{nd}{2}\log 2\pi - \frac{n}{2}$$

The log likelihood ratio for a given t can be obtained from (2.5) and (2.7) as follows

$$\log\frac{\hat{L}_1(t)}{\hat{L}_0(t)} = \log\hat{L}_1(t) - \log\hat{L}_0(t)$$
$$= -\frac{t}{2}\log\left|\hat{\Sigma}_{0,t}\right| - \frac{n-t}{2}\log\left|\hat{\Sigma}_{1,t}\right| - \frac{nd}{2}\log 2\pi - \frac{n}{2}$$
$$-\left(-\frac{n}{2}\log\left|\hat{\Sigma}_{0,n}\right| - \frac{nd}{2}\log 2\pi - \frac{n}{2}\right)$$
$$= -\frac{t}{2}\log\left|\hat{\Sigma}_{0,t}\right| - \frac{n-t}{2}\log\left|\hat{\Sigma}_{1,t}\right| + \frac{n}{2}\log\left|\hat{\Sigma}_{0,n}\right|$$

Denote

$$U_{n,t} = 2\log\frac{\hat{L}_1(t)}{\hat{L}_0(t)} = -t\log\left|\hat{\Sigma}_{0,t}\right| - (n-t)\log\left|\hat{\Sigma}_{1,t}\right| + n\log\left|\hat{\Sigma}_{0,n}\right| \quad (2.8)$$

Univariate case can be regarded as a special case of multivariate change-point problem with dimension $d = 1$. Thus $U_{n,t}$ can be directly derived from (2.8) as

$$U_{n,t} = 2\log\frac{\hat{L}_1(t)}{\hat{L}_0(t)} = -t\log\hat{\sigma}_{0,t}^2 - (n-t)\log\hat{\sigma}_{1,t}^2 + n\log\hat{\sigma}_{0,n}^2 \tag{2.9}$$

where $\hat{\sigma}_{0,t}^2 = \frac{1}{t}\sum_{i=1}^{t}\left(Y_i - \hat{\mu}_{0,t}\right)^2, \hat{\sigma}_{1,t}^2 = \frac{1}{n-t}\sum_{i=t+1}^{n}\left(Y_i - \hat{\mu}_{1,t}\right)^2$

$$\hat{\sigma}_{0,n}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{\mu}_{0,t}\right)^2$$

Then, by letting $U_n = \max_{1\leq t\leq n-1} U_{n,t}$, twice the log likelihood ratio statistic is given by

$$U_n = \max_{1\leq t\leq n-1}\left\{U_{n,t}\right\} \tag{2.10}$$

From Csörgő and Horváth (1997), the asymptotic distribution of the above log-likelihood ratio statistic is based upon

$$W_n = (2\log\log nU_n)^2 - \left(2\log\log n + \frac{p}{2}\log\log\log n - \log\Gamma\left(\frac{p}{2}\right)\right) \tag{2.11}$$

where p is the number of parameters. Under the case of change in mean and covariance matrix, the number of parameters equals the sum of the dimension of the mean vector d, and the unique number of parameters in the covariance matrix $d(d+1)/2$. That is to say

$$p = d + \frac{d(d+1)}{2} = \frac{d(d+3)}{2} \qquad (2.12)$$

The limiting distribution of W_n assumes the familiar Gumbel type of the extreme value distribution given by

$$\lim_{n\to\infty} P(W_n \leq t) = \exp(-2e^{-t}) \qquad (2.13)$$

where $-\infty < t < \infty$

For a given data set, if the computed value of W_n equals w, then the approximate P-value associated with testing H0 against H1 is seen to be

$$\lim_{n\to\infty} P(|W_n| > |w|) = \lim_{n\to\infty} [P(W_n > |w|) + P(W_n < -|w|)] \qquad (2.14)$$

$$= 1 - \exp\left(-2e^{-|w|}\right) + \exp\left(-2e^{|w|}\right)$$

2.1.2 Change in mean only

Let $Y_1, Y_2, \dots, Y_n$ be a sequence of time series valued independent random vectors such that $Y_i \in \mathbb{R}^d, i = 1, \dots, n$. Furthermore, for each $i = 1, 2, \dots, n$, let Y_i follow the multivariate Gaussian distribution with mean vector μ and variance-covariance matrix Σ. Under the classical change point model in which the covariance matrix Σ remains stationary throughout the sampling period, the mean vector μ changes from an initial value μ_0 to a subsequent value μ_1 at some unknown change point $\tau_n \in \{1,2, \dots, n-1\}$. It is still assumed that parameters μ_0, μ_1 and Σ are all unknown. Thus, under the change point model, one has

$$Y_i \sim \begin{cases} f\left(\cdot; \mu_{0,\tau_n}, \Sigma_{\tau_n}\right), & i = 1, \quad \dots, \tau_n \\ f\left(\cdot; \mu_{1,\tau_n}, \Sigma_{\tau_n}\right), & i = \tau_n + 1, \dots, n \end{cases} \tag{2.15}$$

where $\tau_n \in \{1, \dots, n-1\}$

On the other hand, when there is no change point in the model, a single parameter set is applicable throughout the sampling period such that under the no change model one has

$$Y_i \sim f\left(\cdot; \mu_{0,n}, \Sigma_n\right), \qquad i = 1, \quad \dots, n \tag{2.16}$$

One is confronted with having to decide whether the given data set can be modeled by the no change model (2.16) , or by the change point model (2.15) with a change occurring in the mean vector at an unknown change point τ_n. Thus, the statistical problem is one of carrying out a test of the following hypotheses:

H0: The data conforms to no change model (2.16)

Against H1: The data conforms to change point model (2.15)

Using the likelihood function presented in (2.3), given $\tau_n = t$, where $t \in \{1, \dots, n-1\}$, we have the log-likelihood function for (2.15) to be

$$\log L_1(t) = -\frac{t}{2}\log|\Sigma_t| - \frac{1}{2}\sum_{i=1}^{t}\left(Y_i - \mu_{0,t}\right)^T \Sigma_t^{-1}\left(Y_i - \mu_{0,t}\right) - \frac{td}{2}\log 2\pi$$

$$-\frac{n-t}{2}\log|\Sigma_t| - \frac{1}{2}\sum_{i=t+1}^{n}\left(Y_i - \mu_{1,t}\right)^T \Sigma_t^{-1}\left(Y_i - \mu_{1,t}\right)$$

$$-\frac{(n-t)d}{2}\log 2\pi$$

The estimates for the parameters under model (2.15) are

$$\hat{\mu}_{0,t} = \frac{1}{t}\sum_{i=1}^{t} Y_i\,, \qquad \hat{\mu}_{1,t} = \frac{1}{n-t}\sum_{i=t+1}^{n} Y_i$$

$$\hat{\Sigma}_t = \frac{1}{n}\left\{\sum_{i=1}^{t}\left(Y_i - \hat{\mu}_{0,t}\right)\left(Y_i - \hat{\mu}_{0,t}\right)^T + \sum_{i=t+1}^{n}\left(Y_i - \hat{\mu}_{1,t}\right)\left(Y_i - \hat{\mu}_{1,t}\right)^T\right\}$$

The estimate for the log-likelihood function for model (2.15) is

$$\log \hat{L}_1(t) = -\frac{t}{2}\log\left|\hat{\Sigma}_{\mathrm{t}}\right| - \frac{1}{2}\sum_{i=1}^{t}\left(Y_i - \hat{\mu}_{0,t}\right)^T \hat{\Sigma}_t^{-1}\left(Y_i - \hat{\mu}_{0,t}\right) - \frac{td}{2}\log 2\pi \tag{2.17}$$

$$- \frac{n-t}{2}\log\left|\hat{\Sigma}_{\mathrm{t}}\right| - \frac{1}{2}\sum_{i=t+1}^{n}\left(Y_i - \hat{\mu}_{1,t}\right)^T \hat{\Sigma}_t^{-1}\left(Y_i - \hat{\mu}_{1,t}\right)$$

$$- \frac{(n-t)d}{2}\log 2\pi$$

$$= -\frac{n}{2}\log\left|\hat{\Sigma}_{\mathrm{t}}\right| - \frac{nd}{2}\log 2\pi - \frac{1}{2}\sum_{i=1}^{t} tr\left[\hat{\Sigma}_t^{-1}\left(Y_i - \hat{\mu}_{0,t}\right)\left(Y_i - \hat{\mu}_{0,t}\right)^T\right]$$

$$- \frac{1}{2}\sum_{i=t+1}^{n} tr\left[\hat{\Sigma}_t^{-1}\left(Y_i - \hat{\mu}_{1,t}\right)\left(Y_i - \hat{\mu}_{1,t}\right)^T\right]$$

$$= -\frac{n}{2}\log\left|\hat{\Sigma}_{\mathrm{t}}\right| - \frac{nd}{2}\log 2\pi$$

$$- \frac{1}{2} tr\left\{\hat{\Sigma}_t^{-1}\left[\sum_{i=1}^{t}\left(Y_i - \hat{\mu}_{0,t}\right)\left(Y_i - \hat{\mu}_{0,t}\right)^T\right.\right.$$

$$\left.\left. + \sum_{i=t+1}^{n}\hat{\Sigma}_t^{-1}\left(Y_i - \hat{\mu}_{1,t}\right)\left(Y_i - \hat{\mu}_{1,t}\right)^T\right]\right\}$$

$$= -\frac{n}{2}\log\left|\hat{\Sigma}_{\mathrm{t}}\right| - \frac{nd}{2}\log 2\pi - \frac{1}{2} tr\left\{\hat{\Sigma}_t^{-1} n\hat{\Sigma}_{\mathrm{t}}\right\}$$

$$= -\frac{n}{2}\log\left|\hat{\Sigma}_{\mathrm{t}}\right| - \frac{nd}{2}\log 2\pi - \frac{n}{2}$$

Similarly, under multivariate Gaussianity, the likelihood function under model (2.16) can be directly adapted from (2.7) as

$$\log \hat{L}_0(t) = -\frac{n}{2}\log\left|\hat{\Sigma}_{\mathrm{n}}\right| - \frac{nd}{2}\log 2\pi - \frac{n}{2} \tag{2.18}$$

where $\quad \hat{\mu}_{0,n} = \frac{1}{n}\sum_{i=1}^{n} Y_i, \quad \hat{\Sigma}_n = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{\mu}_{0,n}\right)\left(Y_i - \hat{\mu}_{0,n}\right)^T$

The log likelihood ratio for a given t from (2.17) and (2.18) is

$$\log\frac{\hat{L}_1(t)}{\hat{L}_0(t)} = \log\hat{L}_1(t) - \log\hat{L}_0(t)$$
$$= \left(-\frac{n}{2}\log\left|\hat{\Sigma}_{\mathrm{t}}\right| - \frac{nd}{2}\log 2\pi - \frac{n}{2}\right) - \left(-\frac{n}{2}\log\left|\hat{\Sigma}_n\right| - \frac{nd}{2}\log 2\pi - \frac{n}{2}\right)$$
$$= -\frac{n}{2}\log\left|\hat{\Sigma}_t\right| + \frac{n}{2}\log\left|\hat{\Sigma}_n\right|$$

Then as in Section 2.1.1, twice the log likelihood ratio statistic is given by

$$U_n = \max_{1\le t\le n-1}\left\{U_{n,t}\right\} \tag{2.19}$$

where $\quad U_{n,t} = 2\log\frac{\hat{L}_1(t)}{\hat{L}_0(t)} = -\mathrm{n}\log\left|\hat{\Sigma}_{\mathrm{t}}\right| + \mathrm{n}\log\left|\hat{\Sigma}_{\mathrm{n}}\right|$

The p-value of the change-point detection follows (2.11), (2.13) and (2.14). Under the change in mean only, the number of parameters, p, equals the dimension of the mean vector d.

2.1.3 *Change in covariance/variance only*

Let $Y_1, Y_2, \dots, Y_n$ be a sequence of time series valued independent random vectors such that $\mathrm{Y_i} \in \mathbb{R}^{\mathrm{d}}, \mathrm{i} = 1, \dots, \mathrm{n}$. Furthermore, for each $\mathrm{i} = 1, 2, \dots, \mathrm{n}$, let Y_i follow the multivariate Gaussian distribution with mean vector μ and variance-covariance matrix Σ. Under the classical change point model in which the mean vector μ remains stationary throughout the sampling period, and the covariance matrix Σ changes from an initial value Σ_0 to a subsequent value Σ_1 at some unknown change-point $\tau_n \in \{1,2, \dots, n-1\}$. It is still assumed that parameters μ, Σ_0 and Σ_1 are all unknown. Thus, under the change point model, one has

$$Y_i \sim \begin{cases} f\left(\cdot; \mu_{\tau_n}, \Sigma_{0,\tau_n}\right), & i = 1, \quad \dots, \tau_n \\ f\left(\cdot; \mu_{\tau_n}, \Sigma_{1,\tau_n}\right), & i = \tau_n + 1, \dots, n \end{cases} \tag{2.20}$$

where $\tau_n \in \{1, \dots, n-1\}$.

On the other hand, when there is no change point in the model, a single parameter set is applicable throughout the sampling period such that under the no change model one has

$$Y_i \sim f\left(\cdot; \mu_n, \Sigma_{0,n}\right), \qquad i = 1, \quad \dots, n \tag{2.21}$$

One is confronted with having to decide whether the given data set can be modeled by the no change model (2.21), or by the change point model (2.20) with a change occurring in the mean vector at an unknown change-point τ_n. Thus, the statistical problem is one of carrying out a test of the following hypotheses:

H0: The data conforms to no change model (2.21)

Against H1: The data conforms to change point model (2.20)

Using the likelihood function presented in (2.3), we have the log-likelihood function for (2.20) to be

$$\log L_1(t) = -\frac{t}{2}\log|\Sigma_0| - \frac{1}{2}\sum_{i=1}^{t}(Y_i-\mu)^T\Sigma_0^{-1}(Y_i-\mu) - \frac{td}{2}\log 2\pi$$
$$-\frac{n-t}{2}\log|\Sigma_1| - \frac{1}{2}\sum_{i=t+1}^{n}(Y_i-\mu)^T\Sigma_1^{-1}(Y_i-\mu)$$
$$-\frac{(n-t)d}{2}\log 2\pi$$

Let the estimate for the parameters under model (2.20) be:

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

$$\hat{\Sigma}_{0,t} = \frac{1}{t}\sum_{i=1}^{t}(Y_i-\hat{\mu}_n)(Y_i-\hat{\mu}_n)^T$$

$$\hat{\Sigma}_{1,t} = \frac{1}{n-t}\sum_{i=t+1}^{n}(Y_i-\hat{\mu}_n)(Y_i-\hat{\mu}_n)^T$$

The estimated log-likelihood function for model (2.20) is

$$\log \hat{L}_1(t) = -\frac{t}{2}\log\left|\hat{\Sigma}_{0,\mathrm{t}}\right| - \frac{1}{2}\sum_{i=1}^{t}(Y_i - \hat{\mu}_n)^T \hat{\Sigma}_{0,\mathrm{t}}^{-1}(Y_i - \hat{\mu}_n) \tag{2.22}$$

$$- \frac{td}{2}\log 2\pi - \frac{n-t}{2}\log\left|\hat{\Sigma}_{1,\mathrm{t}}\right|$$

$$- \frac{1}{2}\sum_{i=t+1}^{n}(Y_i - \hat{\mu}_n)^T \hat{\Sigma}_{1,\mathrm{t}}^{-1}(Y_i - \hat{\mu}_n) - \frac{(n-t)d}{2}\log 2\pi$$

$$= -\frac{t}{2}\log\left|\hat{\Sigma}_{0,t}\right| - \frac{n-t}{2}\log\left|\hat{\Sigma}_{1,t}\right| - \frac{nd}{2}\log 2\pi$$

$$- \frac{1}{2}\sum_{i=1}^{t} tr\left[\hat{\Sigma}_{0,t}^{-1}(Y_i - \hat{\mu}_n)(Y_i - \hat{\mu}_n)^T\right]$$

$$- \frac{1}{2}\sum_{i=t+1}^{n} tr\left[\hat{\Sigma}_{1,t}^{-1}(Y_i - \hat{\mu}_n)(Y_i - \hat{\mu}_n)^T\right]$$

$$= -\frac{t}{2}\log\left|\hat{\Sigma}_{0,t}\right| - \frac{n-t}{2}\log\left|\hat{\Sigma}_{1,t}\right| - \frac{nd}{2}\log 2\pi - \frac{1}{2}tr\left\{\hat{\Sigma}_{0,t}^{-1}\mathrm{t}\hat{\Sigma}_{0,t}\right\}$$

$$- \frac{1}{2}tr\left\{\hat{\Sigma}_{1,t}^{-1}(n-t)\hat{\Sigma}_{1,t}\right\}$$

$$= -\frac{t}{2}\log\left|\hat{\Sigma}_{0,t}\right| - \frac{n-t}{2}\log\left|\hat{\Sigma}_{1,t}\right| - \frac{nd}{2}\log 2\pi - \frac{n}{2}$$

Similarly, under multivariate Gaussianity, the likelihood function under no-change model (2.21) can be directly adapted from (2.7) as

$$\mathrm{Log}\,\hat{L}_0(t) = -\frac{n}{2}\log\left|\hat{\Sigma}_{0,\mathrm{n}}\right| - \frac{nd}{2}\log 2\pi - \frac{n}{2} \tag{2.23}$$

where $$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i\,, \qquad \hat{\Sigma}_{0,n} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{\mu}_n)(Y_i - \hat{\mu}_n)^T$$

Then log likelihood ratio for a given t is

$$\begin{aligned}
&\log\frac{\hat{L}_1(t)}{\hat{L}_0(t)} = \log\hat{L}_1(t) - \log\hat{L}_0(t) \\
&= \left(-\frac{t}{2}\log\left|\hat{\Sigma}_{0,t}\right| - \frac{n-t}{2}\log\left|\hat{\Sigma}_{1,t}\right| - \frac{nd}{2}\log 2\pi - \frac{n}{2}\right) \\
&\qquad\qquad - \left(-\frac{n}{2}log\left|\hat{\Sigma}_{0,n}\right| - \frac{nd}{2}log\ 2\pi - \frac{n}{2}\right) \\
&= -\frac{t}{2}\log\left|\hat{\Sigma}_{0,t}\right| - \frac{n-t}{2}\log\left|\hat{\Sigma}_{1,t}\right| + \frac{n}{2}\log\left|\hat{\Sigma}_{0,\mathrm{n}}\right|
\end{aligned}$$

As in Section 2.1.1, twice the log likelihood ratio statistic is given by

$$U_n = \max_{1\le t\le n-1}\{U_{n,t}\} \tag{2.24}$$

where $$U_{n,t} = 2\log\frac{\hat{L}_1(t)}{\hat{L}_0(t)} = -t\log\left|\hat{\Sigma}_{0,t}\right| - (n-t)\log\left|\hat{\Sigma}_{1,t}\right| + n\log\left|\hat{\Sigma}_{0,n}\right|$$

The p-value of the change-point detection follows (2.11), (2.13) and (2.14). Under the change in covariance only, the number of parameters, p, equals the unique number of parameters in a covariance matrix, $d(d+1)/2$.

2.2 Change-point Estimation Setup

In change-point estimation, our interest here is to pursue the asymptotic distribution of the maximum likelihood estimator $\hat{\tau}_n$ when the parameters are unknown. Now suppose that $\tilde{\tau}_n$ is the mle of τ_n when the parameters are known. Hinkley (1972) has shown that asymptotic distributions of both $\hat{\tau}_n$ and $\tilde{\tau}_n$ are equivalent. Hence in the sequel, we shall first pursue the asymptotic distribution of $\tilde{\tau}_n$ only.

In deriving the asymptotic distribution of $\tilde{\tau}_n$, we first note that we can begin with the basic methodology that Jandhyala and Fotopoulos (1999) derived for the univariate situation and adapt it for the multivariate problems. While it is true that Hinkley (1970, 1971) was the first to initiate the study of the distribution theory for the mle of a change point in a sequence of independent observations, the distribution theory was not detailed enough from a computational point of view. In this regard, Jandhyala and Fotopoulos (1999, 2001), Jandhyala et al (2002, 2006), and Fotopoulos and Jandhyala (2001) studied the distributional aspects of Hinkley's (1972) mle, mainly to make the distribution of the mle computationally more tractable. Importantly, by deriving alternative expressions for the distribution of the mle they developed an algorithmic approach for computing the lower and upper bounds and also good approximations. Thus far, their algorithmic approach has been applied to compute the asymptotic distribution of the change-point to univariate datasets, and change in mean only for multivariate dataset. Here we shall adapt their algorithm to compute the asymptotic distribution of the change-point mle for multivariate datasets with change in both mean and covariance matrix.

Assume that $Y_1, Y_2, \ldots, Y_{\tau_n}$ are i.i.d. with common multivariate Gaussian density function $f_0(\cdot)$, and $Y_{\tau_n+1}, \ldots, Y_n$ are i.i.d. with common multivariate Gaussian density function $f_1(\cdot)$, where all the means and covariance matrices are known. Following Hinkley (1972), $\tilde{\tau}_n$ the mle of τ_n may be expressed as:

$$\tilde{\tau}_n = \arg \max_{1 \le j \le n-1} \sum_{i=1}^{j} a(Y_i) \tag{2.25}$$

where $a(Y_i) = \log f_0(Y_i)/f_1(Y_i)$.

For the purposes of establishing distribution theory, it is convenient to work with $\xi_n = \tilde{\tau}_n - \tau_n \in \{-\tau_n + 1, \ldots, n - \tau_n - 1\}$ instead of $\tilde{\tau}_n$. Then, it turns out that

$$\xi_n = \arg \max_{-\tau_n + 1 \le j \le n - \tau_n - 1} \sum_{i=1}^{j+\tau_n} a(Y_i) \tag{2.26}$$

where the maximizer is a result of the following two-sided random walk $\Gamma_n(\cdot)$ on $\mathbb{Z}$ such that:

$$\Gamma_n(j; \tau_n) \tag{2.27}$$

$$= \begin{cases} \sum_{i=1}^{j} a(Y_i^*) = \sum_{i=1}^{j} X_i^* = S_j^* \ , & j \in \{1, \ldots, n - \tau_n - 1\} \\ 0 \ , & j = 0 \\ -\sum_{i=1}^{-j} a(Y_i^o) = \sum_{i=1}^{-j} X_i^o = S_{-j}^o \ , & j \in \{-1, \ldots, -\tau_n - 1\} \end{cases}$$

In deriving the asymptotic theory, both τ_n and $n - \tau_n$ tend to infinity so that we will have enough information on both sides. Denote $\tilde{\tau}_\infty$ to be the maximum likelihood estimate of τ_n based on the sample $\{Y_1, \ldots, Y_n\}$ with $n \to \infty$. Note that the distribution of $\xi_n = \tilde{\tau}_n - \tau_n$ depends on the mean and covariance matrix before and after the change-point, as well as τ_n and $n - \tau_n$, while ξ_∞ depends only on the mean and covariance matrix after the change-point. In practice, ξ_n is rather inadequate since its distribution always depends on the unknown change-point τ_n. On the other hand, Fotopoulos and Jandhyala (2001) showed that ξ_∞ is a proper random variable and $\xi_n \to \xi_\infty$ almost surely, which implies that the distribution of ξ_n can be well-approximated by ξ_∞ for moderately large sample sizes.

Let $\{Y^o, Y_i^o : i \in \mathbb{N}\}$ be a sequence of i.i.d. random vectors such that Y^o is distributed according to $f_0(\cdot)$ and $\{Y^*, Y_i^* : i \in \mathbb{N}\}$ is another sequence of i.i.d. random vectors such that Y^* is distributed according to $f_1(\cdot)$. Furthermore the two sequences are independent of each other. It follows that the sequences $\{X^o, X_i^o : i \in \mathbb{N}\}$ (before the change) and $\{X^*, X_i^* : i \in \mathbb{N}\}$ (after the change) defined in (2.27) are independent.

The immediate goal is to establish the explicit functional relationship of X^o with Y^o and that of X^* with Y^*, respectively. These relationships will enable us in identifying the distributions of both X^o and X^*, a step that is fundamental in the algorithmic procedure of Jandhyala and Fotopoulos (1999). In Chapter 3 and 4, the estimation of change-point mle for possible combinations of change in mean and/or covariance/variance is explored.

3 INFERENCE FOR CHANGE-POINT IN THE MEAN ONLY OF A GAUSSIAN SERIES

In classical change-point literature, model (2.15) is known as the abrupt change model for change in mean only. On the basis of the detection statistics in Chapter 2, we begin this chapter by assuming that there is a change in the mean of observations $Y_1, Y_2, \ldots, Y_n$ at some unknown point, with the underlying assumption that the variance remained constant throughout the sampling period. In this chapter, the asymptotic distribution of the change-point mle is derived for both multivariate and univariate cases when the parameters for mean and variance/covariance are unknown.

3.1 Multivariate Case

Let $Y_1, Y_2, \ldots, Y_n$ be a sequence of time series valued independent random vectors such that $\mathrm{Y_i} \in \mathbb{R}^{\mathrm{d}}, \mathrm{i} = 1, \ldots, \mathrm{n}$. The mean vector of the series changes from μ_0 to μ_1 at some unknown point τ_n such that $\mu_0 \neq \mu_1$, and the covariance matrix Σ remained constant. Both the mean vector and covariance matrix are unknown.

As discussed in Section 2.2, the immediate goal is to establish the explicit functional relationship of X^o with Y^o and that of X^* with Y^*, respectively for the two-sided random walk defined in (2.27) . These relationships will enable us in identifying the distributions of both X^o and X^*, a step that is fundamental in the algorithmic procedure of Jandhyala and Fotopoulos (1999).

First, let the symmetric matrix Σ admit the usual orthogonal decomposition given by $\Sigma = Q\Lambda Q^T$ where Q is an orthogonal matrix, and Λ is a real diagonal matrix with positive entries $\lambda_1, \ldots, \lambda_d$.

It follows that Y^o, the random variable before the change-point, admits the representation

$$Y^o =^D \mu_0 + \Sigma^{1/2} Z \tag{3.1}$$

or

$$Y^o - \mu_0 =^D \Sigma^{1/2} Z \tag{3.2}$$

where Z is the standard multivariate normal random variable. Consequently, the random variable X^o may be expressed as

$$X^o = -a(Y^o) = -\ln \frac{f(Y^o;\mu_0,\Sigma)}{f(Y^o;\mu_1,\Sigma)}$$

$$= -\ln \frac{\frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(Y^o-\mu_0)^T\Sigma^{-1}(Y^o-\mu_0)\right)}{\frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(Y^o-\mu_1)^T\Sigma^{-1}(Y^o-\mu_1)\right)}$$

$$= \frac{1}{2}(Y^o-\mu_0)^T\Sigma^{-1}(Y^o-\mu_0)$$

$$-\frac{1}{2}(Y^o-\mu_0+\mu_0-\mu_1)^T\Sigma^{-1}(Y^o-\mu_0+\mu_0-\mu_1)$$

$$=^D \frac{1}{2}\left(\Sigma^{1/2}Z\right)^T\Sigma^{-1}\left(\Sigma^{1/2}Z\right)$$

$$-\frac{1}{2}\left(\Sigma^{1/2}Z+\mu_0-\mu_1\right)^T\Sigma^{-1}\left(\Sigma^{1/2}Z+\mu_0-\mu_1\right)$$

$$= \frac{1}{2}\left(\Sigma^{1/2}Z\right)^T\Sigma^{-1}\left(\Sigma^{1/2}Z\right) - \frac{1}{2}\left(\Sigma^{1/2}Z\right)^T\Sigma^{-1}\left(\Sigma^{1/2}Z\right)$$

$$-\frac{1}{2}\left(\Sigma^{1/2}Z\right)^T\Sigma^{-1}\,(\mu_0 - \mu_1)$$

$$-\frac{1}{2}(\mu_0 - \mu_1)^T\Sigma^{-1}\,(\Sigma^{1/2}Z)\; - \frac{1}{2}(\mu_0 - \mu_1)^T\Sigma^{-1}(\mu_0$$

$$- \mu_1)$$

$$= (\mu_1 - \mu_0)^T\Sigma^{-1}\,(\Sigma^{1/2}Z)\; - \frac{1}{2}(\mu_1 - \mu_0)^T\Sigma^{-1}(\mu_1 - \mu_0)$$

Note as we can decompose $\Sigma = Q\Lambda Q^T$, then

$$\Sigma^{-1} = (Q\Lambda Q^T)^{-1} = Q\Lambda^{-1}Q^T = Q\mathrm{diag}\{\lambda_1^{-1}, \dots, \lambda_\mathrm{d}^{-1}\}Q^T$$

$$= Q\mathrm{diag}\left\{\lambda_1^{-1/2}, \dots, \lambda_\mathrm{d}^{-1/2}\right\}\mathrm{diag}\left\{\lambda_1^{-1/2}, \dots, \lambda_\mathrm{d}^{-1/2}\right\}Q^T$$

$$= Q\mathrm{diag}\left\{\lambda_1^{-1/2}, \dots, \lambda_\mathrm{d}^{-1/2}\right\}\mathrm{Q}^\mathrm{T}\mathrm{Q}\mathrm{diag}\left\{\lambda_1^{-1/2}, \dots, \lambda_\mathrm{d}^{-1/2}\right\}Q^T$$

$$:= Q\Lambda^{-1/2}\mathrm{Q}^\mathrm{T}\mathrm{Q}\Lambda^{-1/2}Q^T$$

$$:= \Sigma^{-1/2}\Sigma^{-1/2}$$

where $\Sigma^{-1/2}$ is also a symmetric matrix.

After the simplification, one obtains

$$X^o =^D (\mu_1 - \mu_0)^T\Sigma^{-1/2}\Sigma^{-1/2}\,\Sigma^{1/2}Z\; - \frac{1}{2}(\mu_1 - \mu_0)^T\Sigma^{-1/2}\Sigma^{-1/2}(\mu_1$$

$$- \mu_0)$$

$$= (\mu_1 - \mu_0)^T\Sigma^{-1/2}Z\; - \frac{1}{2}\left[\Sigma^{-1/2}(\mu_1 - \mu_0)\right]^T\left[\Sigma^{-1/2}(\mu_1 - \mu_0)\right]$$

$$= \left[\Sigma^{-1/2}(\mu_1 - \mu_0)\right]^T Z\; - \frac{1}{2}\left[\Sigma^{-1/2}(\mu_1 - \mu_0)\right]^T\left[\Sigma^{-1/2}(\mu_1 - \mu_0)\right]$$

Now, upon letting $\eta = \Sigma^{-1/2}\ (\mu_1 - \mu_0)$, we have

$$X^o =^D -\frac{1}{2}\eta^T\eta + \eta^T Z \tag{3.3}$$

Clearly, it follows from (3.3) that $X^o \sim N(-\frac{1}{2}\eta^T\eta, \eta^T\eta)$.

For purposes of finding the distribution of $X^* = a(Y^*)$, we can follow similar derivation for the distribution of X^o.

First note that Y^*, the random variable after the change-point admits the representation

$$Y^* =^D \mu_1 + \Sigma^{1/2} Z \tag{3.4}$$

or

$$Y^* - \mu_1 =^D \Sigma^{1/2} Z \tag{3.5}$$

where Z is the standard multivariate normal random variable. Consequently, the random variable X^* may be expressed as

$$X^* = a(Y^*) = \ln\frac{f(Y^*;\mu_0,\Sigma)}{f(Y^*;\mu_1,\Sigma)}$$

$$= \ln\frac{\frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(Y^*-\mu_0)^T\Sigma^{-1}(Y^*-\mu_0)\right)}{\frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(Y^*-\mu_1)^T\Sigma^{-1}(Y^*-\mu_1)\right)}$$

$$= -\frac{1}{2}(Y^* - \mu_1 + \mu_1 - \mu_0)^T \Sigma^{-1} (Y^* - \mu_1 + \mu_1 - \mu_0)$$

$$+ \frac{1}{2}(Y^* - \mu_1)^T \Sigma^{-1} (Y^* - \mu_1)$$

$$=^D - \frac{1}{2}\left(\Sigma^{1/2} Z + \mu_1 - \mu_0\right)^T \Sigma^{-1} \left(\Sigma^{1/2} Z + \mu_1 - \mu_0\right)$$

$$+ \frac{1}{2}\left(\Sigma^{1/2} Z\right)^T \Sigma^{-1} \left(\Sigma^{1/2} Z\right)$$

$$= -\frac{1}{2}\left(\Sigma^{1/2} Z\right)^T \Sigma^{-1} \left(\Sigma^{1/2} Z\right) - \frac{1}{2}\left(\Sigma^{1/2} Z\right)^T \Sigma^{-1} (\mu_1 - \mu_0)$$

$$- \frac{1}{2}(\mu_1 - \mu_0)^T \Sigma^{-1} \left(\Sigma^{\frac{1}{2}} Z\right)$$

$$- \frac{1}{2}(\mu_1 - \mu_0)^T \Sigma^{-1} (\mu_1 - \mu_0) + \frac{1}{2}\left(\Sigma^{1/2} Z\right)^T \Sigma^{-1} \left(\Sigma^{1/2} Z\right)$$

$$= -(\mu_1 - \mu_0)^T \Sigma^{-1} (\Sigma^{1/2} Z) - \frac{1}{2}(\mu_1 - \mu_0)^T \Sigma^{-1} (\mu_1 - \mu_0)$$

Note as we can decompose $\Sigma = \Sigma^{-1/2} \Sigma^{-1/2}$, then

$$X^* =^D -(\mu_1 - \mu_0)^T \Sigma^{-1/2} \Sigma^{-1/2} \Sigma^{1/2} Z - \frac{1}{2}(\mu_1 - \mu_0)^T \Sigma^{-1/2} \Sigma^{-1/2} (\mu_1 - \mu_0)$$

$$= -(\mu_1 - \mu_0)^T \Sigma^{-1/2} Z - \frac{1}{2}\left[\Sigma^{-1/2}(\mu_1 - \mu_0)\right]^T \left[\Sigma^{-1/2}(\mu_1 - \mu_0)\right]$$

$$= -\left[\Sigma^{-1/2}(\mu_1 - \mu_0)\right]^T Z - \frac{1}{2}\left[\Sigma^{-1/2}(\mu_1 - \mu_0)\right]^T \left[\Sigma^{-1/2}(\mu_1 - \mu_0)\right]$$

As $\eta = \Sigma^{-1/2} (\mu_1 - \mu_0)$, we have

$$X^* =^D -\frac{1}{2}\eta^T\eta - \eta^T Z \tag{3.6}$$

Clearly, it follows from (3.6) that $X^* \sim N(-\frac{1}{2}\eta^T\eta, \eta^T\eta)$, too.

Thus, both X^o and X^* have identical univariate normal distributions. This result coincides with the distributions found by Jandhyala and Fotopoulos (1999) for the situation of finding the distribution of the change point MLE in the mean of univariate normal observations. Note that the random walks S^o and S^* that are defined in (2.27) are independent of each other, and both have negative means. Thus both walks eventually drift to $-\infty$.

One can apply the asymptotic distribution of the change point MLE for estimating change in the mean of univariate normal observations as derived and computed by Jandhyala and Fotopoulos (1999) to the multivariate change point MLE. It may be further noted that the vector valued definition of $\eta = \Sigma^{-1/2}(\mu_1 - \mu_0)$ when specialized to the univariate case agrees with the corresponding definition of Jandhyala and Fotopoulos (1999), after adjusting for the slight change in the definition of η.

The asymptotic distribution of the change point MLE for the univariate case as presented by Jandhyala and Fotopoulos (1999, Table 1) was computed on the basis of $\delta = \frac{1}{2}\eta$. We can use the same table for the multivariate case simply by defining $\delta = \frac{1}{2}\sqrt{\eta^T\eta}$. We can rewrite $\eta^T\eta = 4\delta^2$.

Let X_1^o and X_1^* represent the initial random variables associated with these two independent random walks S^o and S^* respectively. All assumptions stated in Jandhyala and Fotopoulos (1999) are satisfied.

Assumption 1: $-\infty \leq E(X_1^O) = E(X_1^*) = -\frac{1}{2}\delta^2 < 0$

Assumption 2: the moment generating functions of X_1^O are $\phi(s) = \exp\left(-\frac{1}{2}\delta^2 s + \frac{1}{2}\delta^2 s^2\right)$, which is convergent for $0 \leq Re(s) < 1$

Assumption 3: For $s \in \mathbb{R}$, $\frac{d}{ds}\phi(s) = \exp\left(-\frac{1}{2}\delta^2 s + \frac{1}{2}\delta^2 s^2\right)\left(-\frac{1}{2}\delta^2 + \delta^2 s\right)$. Thus when $s = \frac{1}{2}$, $\frac{d}{ds}\phi(s) = 0$ and $\frac{d^2}{ds^2}\phi\left(\frac{1}{2}\right) > 0$. It can be verified that $\phi\left(\frac{1}{2}\right) < \phi(0) = \phi(1) = 1$. So $\phi(s)$ attains a unique minimum on $[0,1]$. According to Fotopoulos, Jandhyala and Khapalova (2010), this assumption is true automatically, which coincides with the computation above. In the future, this assumption will be assumed to be true without proving.

The algorithmic procedure that was proposed in Jandhyala and Fotopoulos (1999) can be applied as follows:

Let $\tau_0 = \inf\{j \geq 1: S_j \leq 0\}$ be the weak descending ladder epoch, and let $\sigma_x = \inf\{j \geq 0: S_j > x\}$, for $x \geq 0$, where σ_0 denotes the strict ascending ladder epoch. Let $M_j = \max\{S_k: 0 \leq k \leq j\}$ be the maximum of the first j partial sums, and Let $M = \max\{S_j: j = 0, 1,2, ..\}$ be the overall maximum. For $x \geq 0$, the followings are defined:

(i) $G_j(x) = \Pr(\mathrm{M_j} \leq \mathrm{x}) = \Pr(\sigma_x > 0)$, for $j \geq 0$.

(ii) $G(x) = \Pr(\mathrm{M} \leq \mathrm{x})$

(iii) $u_j(x) = \Pr(\tau_0 > j, S_j \in (0, x])$ for $j \geq 0$ and $x \geq 0$.

Define $V_0 = 0$ and let $V_j = G_j(0) = \Pr(\sigma_0 > j)$, for $j \geq 1$. Note that $u_0(x) = \Pr(\tau_0 > 0, S_j \in (0, x])$. Thus $u_0(x) = 0$ for $x > 0$, and $u_0(0) = 1$. Then let $q_j = u_j(\infty) = \Pr(\tau_0 > j)$ for $j \geq 1$.

It is well known from Spitzer's identity that $V_\infty = e^{-B(1)}$, where $B(s) = \sum (s^j b_j)/n$ and $b_j = \Pr(S_j > 0)$, for $j \geq 1$. Therefore,

$$G^o(0) = V_\infty^o = e^{-B^o(1)}, G^*(0) = V_\infty^* = e^{-B^*(1)} \tag{3.7}$$

Jandhyala and Fotopoulos (1999) derived the following iterative procedure for the sequence of probabilities $\{q_j, j \geq 0\}$ as

$$q_0 = 1, \qquad jq_j = \sum_{k=0}^{j-1} b_{j-k} q_k \quad (j \geq 1) \tag{3.8}$$

Under this setup, Jandhyala and Fotopoulos (1999) proved that the probability distribution function for the maximum likelihood estimator $\hat{\tau}_\infty$ of the change-point τ is

$$\Pr(\hat{\xi}_{\infty} = i) \tag{3.9}$$

$$= \begin{cases} e^{-B^*(1)}\left[q_i^* - \int_{0^+}^{\infty}\{1 - G^o(x)\}du_i^*(x)\right], & i > 0 \\ e^{-B^*(1)-B^o(1)}, & i = 0 \\ e^{-B^o(1)}\left[q_{-i}^o - \int_{0^+}^{\infty}\{1 - G^*(x)\}du_{-i}^0(x)\right], & i < 0 \end{cases}$$

Jandhyala and Fotopoulos (1999) showed that the distribution function $G(x)$ of M satisfied exponential form, and derived computable inequalities for the probability distribution of $\hat{\xi}_{\infty}$.

Let $\tilde{u}_i(\lambda) = \int e^{-\lambda x} du_i(x)$ be the Laplace transformation of $u_i(x)$ for $x > 0$, and let $\tilde{u}_0(\lambda) = 1$. From Spitzer's identity, the iterative procedure for computing $\tilde{u}_i(\lambda)$ was derived as follows:

$$j\tilde{u}_j(\lambda) = \sum_{k=0}^{j-1} \tilde{b}_{j-k}(\lambda)\tilde{u}_k(\lambda) \quad (j \geq 1) \tag{3.10}$$

where $\tilde{b}_j(\lambda) = E\{e^{-\lambda S_j} I(S_j > 0)\}$.

The distribution of $\hat{\xi}_{\infty}$ can be evaluated by

$$\Pr(\hat{\xi}_{\infty} = i) = \begin{cases} e^{-B^*(1)}\left[q_i^* - \left(1 - e^{-B^o(1)}\right)\tilde{u}_i^*(\vartheta^o)\right], & i > 0 \\ e^{-B^*(1)-B^o(1)}, & i = 0 \\ e^{-B^o(1)}\left[q_i^o - \left(1 - e^{-B^*(1)}\right)\tilde{u}_i^o(\vartheta^*)\right], & i < 0 \end{cases} \tag{3.11}$$

The algorithmic procedure can be follows as

Step S0: Let $Y^o \sim N(\mu_0, \Sigma)$, $Y^* \sim N(\mu_1, \Sigma)$.

Stop S1: As derived above, $X^o \sim N(-2\delta^2, 4\delta^2)$ and $X^* \sim N(-2\delta^2, 4\delta^2)$. Since X^* and X^o are identically distributed, so their partial sums S_{-j}^o, where $j \in$

$\{-1, \ldots, -\tau_n - 1\}$ and S_j^*, where $j \in \{1, \ldots, n - \tau_n - 1\}$, are also identically distributed, whose distribution are both $N(-2|j|\delta^2, 4|j|\delta^2)$. $|j| \in \{1,2, \ldots\}$.

Step S2: Compute $\{b_j^o\}$ and $\{b_j^*\}$ for $j = 1, 2, \ldots$, where $b_j^o = \Pr(\mathrm{S_j^o} > 0)$ and $b_j^* = \Pr(\mathrm{S_j^*} > 0)$. They can be both computed by the cumulative distribution function of normal distribution $N(-2j\delta^2, 4j\delta^2)$. In statistical software R, b_j^o and b_j^* can be computed by the function $\mathrm{pnorm}\big(0, \mathrm{mean} = -2j\delta^2, \mathrm{sd} = 2\delta\sqrt{j}, \mathrm{lower.tail} = \mathrm{FALSE}\big)$.

Step S3: Compute $B^o(1)$ and $B^*(1)$ as $B^o(1) = \sum b_j^o / j$ and $B^*(1) = \sum b_j^* / j$.

Step S4: Compute both $\{\tilde{b}_j^o(\vartheta^*)\}$ and $\{\tilde{b}_j^*(\vartheta^o)\}$ as $E\left(e^{-\vartheta^* S_j^o}\mathrm{I}(\mathrm{S_j^o} > 0)\right)$ and $E\left(e^{-\vartheta^o S_j^*}\mathrm{I}(\mathrm{S_j^*} > 0)\right)$ for $j = 1, 2, \ldots$, respectively. Since both ϑ^o and $\vartheta^*$are 1, and both $\mathrm{S_j^o}$ and $\mathrm{S_j^*}$ follow the univariate normal distribution with mean $-2j\delta^2$ and variance $4j\delta^2$, $E\left(e^{-\vartheta^* S_j^o}\mathrm{I}(\mathrm{S_j^o} > 0)\right)$ can be computed as $\int_0^\infty e^{-S} f_N(s; -2j\delta^2, 4j\delta^2) ds$, where $f_N(s; -2j\delta^2, 4j\delta^2)$ is the probability density function for the univariate normal distribution. The integration can be computed using the integration function in R as $\mathrm{integrate}(f, \ \mathrm{lower} = 0, \mathrm{upper} = Inf)$ where integrand f can be defined as $f = \mathrm{function}(s)\{x = \exp(-s)\,\mathrm{dnorm}\,\big(s, \mathrm{mean} = -2j\delta^2, \mathrm{sd} = 2\delta\sqrt{j}\big)\}$.

Step S5: Implement the iterative procedures for $\{q_j^o\}$, $\{\tilde{u}_j^o(\vartheta^*)\}$ and $\{q_j^*\}$, $\{\tilde{u}_j^*(\vartheta^o)\}$ as follows:

$$q_0^o = 1,\ jq_j^o = \sum_{k=0}^{j-1} b_{j-k}^o q_k^o \ \ ;\ \tilde{u}_j^o(\vartheta^*) = 1,\ j\tilde{u}_j^o(\vartheta^*) = \sum_{k=0}^{j-1} \tilde{b}_{j-k}^o \tilde{u}_k^o$$

$$q_0^* = 1,\ jq_j^* = \sum_{k=0}^{j-1} b_{j-k}^* q_k^* \ \ ;\ \tilde{u}_j^*(\vartheta^o) = 1,\ j\tilde{u}_j^*(\vartheta^o) = \sum_{k=0}^{j-1} \tilde{b}_{j-k}^* \tilde{u}_k^*$$

Step S6: Estimate $\Pr(\hat{\xi}_n = i)$ by (3.11) as follows

$$\Pr(\hat{\xi}_\infty = i) = \begin{cases} e^{-B^*(1)}\left[q_i^* - \left(1 - e^{-B^o(1)}\right)\tilde{u}_i^*(\vartheta^o)\right], & i > 0 \\ e^{-B^*(1)-B^o(1)}, & i = 0 \\ e^{-B^o(1)}\left[q_i^o - \left(1 - e^{-B^*(1)}\right)\tilde{u}_i^o(\vartheta^*)\right], & i < 0 \end{cases}$$

Table 3.1 presents the computed distribution of $\hat{\xi}_n$ using the algorithmic procedure S0 – S6 for various values of δ. Under the assumption that the covariance matrix keeps constant before and after change, and that τ_n and $n - \tau_n$ both approach ∞, the distribution of $\hat{\xi}_\infty$ is symmetric.

Table 3.1. Asymptotic probabilities $\Pr(\hat{\xi}_\infty = \pm k)$, where $k = 0, 1, 2, \dots$ for the maximum likelihood estimate of the change-point in the case of normal distribution.

k	$\Pr(\hat{\xi}_\infty = \pm k)$			
	$\delta = 0.5$	$\delta = 1$	$\delta = 1.5$	$\delta = 2$
0	0.2802	0.6409	0.8568	0.9531
1	0.1181	0.1152	0.0599	0.0220
2	0.0689	0.0385	0.0097	0.0014
3	0.0454	0.0156	0.0020	0.0001
4	0.0318	0.0069	0.0005	0.0000
5	0.0231	0.0033	0.0001	
6	0.0173	0.0016	0.0000	
7	0.0132	0.0008		
8	0.0102	0.0004		
9	0.0080	0.0002		
10	0.0064	0.0001		
15	0.0022	0.0000		
20	0.0008			
25	0.0003			
sum	1.0213	1.0063	1.0011	1.0001
sd	5.1358	1.2436	0.5057	0.2399

3.2 Univariate Case

The univariate change-point problem can be regarded as a special case of multivariate change-point problem where the dimensionality decreases to 1. Therefore, the estimation for univariate change-point follows the method of that of the multivariate case in section 3.1. Let $Y_1, Y_2, \ldots, Y_n$ be a sequence of time series valued independent random vectors such that $\mathrm{Y_i} \in \mathbb{R}, \mathrm{i} = 1, \ldots, \mathrm{n}$. The mean of the series changes from μ_0 to μ_1 at some unknown point τ_n such that $\mu_0 \neq \mu_1$, and the variance σ^2 remained constant. Both the mean and variance are unknown. The asymptotic distribution of the maximum likelihood estimator of the change-point $\hat{\tau}_n$ when assuming μ_0, μ_1 and σ^2 are unknown is equivalent to $\tilde{\tau}_n$ when assuming μ_0, μ_1 and σ^2 are known.

Under the set-up for the multivariate change-point estimation in Section 2.1, it follows that $Y_1, Y_2, \ldots, Y_{\tau_n}$ are i.i.d. with common Gaussian density function $f(\cdot; \mu_0, \sigma^2)$, and $Y_{\tau_n+1}, \ldots, Y_n$ are i.i.d. with common Gaussian density function $f(\cdot; \mu_1, \sigma^2)$, wherein we assume that μ_0, μ_1 and σ^2 are known. The two-sided random walk is construction as in (2.27) . $\{Y^o, Y_i^o : i \in \mathbb{N}\}$ is a sequence of i.i.d. random variables such that Y^o is distributed according to $f(\cdot; \mu_0, \sigma^2)$ and $\{Y^*, Y_i^* : i \in \mathbb{N}\}$ is another sequence of i.i.d. random variables such that Y^* is distributed according to $f(\cdot; \mu_1, \sigma^2)$. Furthermore the two sequences are independent of each other. It follows that the sequences $\{X^o, X_i^o : i \in \mathbb{N}\}$ (before the change) and $\{X^*, X_i^* : i \in \mathbb{N}\}$

(after the change) are also independent. The functional relationship of X^o with Y^o and that of X^* with Y^* can be identified similarly as in section section 3.1.

It follows that Y^o, the random variable before the change-point, admits the representation

$$Y^o =^D \mu_0 + \sigma Z \tag{3.12}$$

or

$$Y^o - \mu_0 =^D \sigma Z \tag{3.13}$$

where Z is the standard univariate normal random variable. Consequently, the random variable X^o may be expressed as

$$X^o = -a(Y^o) = -\ln\frac{f(Y^o;\mu_0,\sigma^2)}{f(Y^o;\mu_1,\sigma^2)}$$

$$= -\ln\frac{\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(Y^o-\mu_0)^2}{2\sigma^2}\right)}{\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(Y^o-\mu_1)^2}{2\sigma^2}\right)}$$

$$= \frac{(Y^o-\mu_0)^2}{2\sigma^2} - \frac{(Y^o-\mu_1)^2}{2\sigma^2}$$

$$= \frac{(Y^o-\mu_0)^2}{2\sigma^2} - \frac{(Y^o-\mu_0+\mu_0-\mu_1)^2}{2\sigma^2}$$

$$=^D \frac{(\sigma Z)^2}{2\sigma^2} - \frac{(\sigma Z+\mu_0-\mu_1)^2}{2\sigma^2}$$

$$= -\frac{2\sigma Z(\mu_0-\mu_1)+(\mu_0-\mu_1)^2}{2\sigma^2}$$

$$= \frac{(\mu_1-\mu_0)}{\sigma}Z - \frac{(\mu_1-\mu_0)^2}{2\sigma^2}$$

Let $\eta = \frac{(\mu_1-\mu_0)}{\sigma}$, we have

$$X^o =^D -\frac{1}{2}\eta^2 + \eta Z \tag{3.14}$$

Clearly, it follows that $X^o \sim N(-\frac{1}{2}\eta^2, \eta^2)$.

For the purposes of finding the distribution of $X^* = a(Y^*)$, we can follow the similar derivation.

First note that Y^*, the random variable after the change-point admits the representation

$$Y^* =^D \mu_1 + \sigma Z \tag{3.15}$$

or

$$Y^* - \mu_1 =^D \sigma Z \tag{3.16}$$

where Z is the standard multivariate normal random variable. Consequently, the random variable X^* may be expressed as

$$X^* = a(Y^*) = \ln \frac{f(Y^*; \mu_0, \sigma^2)}{f(Y^*; \mu_1, \sigma^2)}$$

$$= \ln \frac{\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(Y^* - \mu_0)^2}{2\sigma^2}\right)}{\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(Y^* - \mu_1)^2}{2\sigma^2}\right)}$$

$$= -\frac{(Y^* - \mu_0)^2}{2\sigma^2} + \frac{(Y^* - \mu_1)^2}{2\sigma^2}$$

$$= -\frac{(Y^* - \mu_1 + \mu_1 - \mu_0)^2}{2\sigma^2} + \frac{(Y^* - \mu_1)^2}{2\sigma^2}$$

$$=^D - \frac{(\sigma Z + \mu_1 - \mu_0)^2}{2\sigma^2} + \frac{(\sigma Z)^2}{2\sigma^2}$$

$$= -\frac{2\sigma Z(\mu_1 - \mu_0) + (\mu_1 - \mu_0)^2}{2\sigma^2}$$

$$= -\frac{(\mu_1 - \mu_0)}{\sigma} Z - \frac{(\mu_1 - \mu_0)^2}{2\sigma^2}$$

As $\eta = \frac{(\mu_1 - \mu_0)}{\sigma}$, we have

$$X^* =^D - \frac{1}{2}\eta^2 - \eta Z \tag{3.17}$$

Clearly, X^* follows the normal distribution $N(-\frac{1}{2}\eta^2, \eta^2)$, too.

Let $\delta = \frac{1}{2}\eta$ as in section 3.1, then both X^o and X^* follow the normal distribution $N(-2\delta^2, 4\delta^2)$, which is identical to the distribution of X^o and X^* under the multivariate case, and to the derivation of Jandhyala and Fotopoulos (1999). Therefore, the method for univariate change-point estimation can exactly follow the algorithmic procedure for multivariate change-point analysis when assuming only the mean changes at some unknown point of time. Table 3.1 in Section 3.1 can also be applied to the univariate case.

4 INFERENCE FOR CHANGE-POINT IN MEAN AND COVARIANCE OF A GAUSSIAN SERIES

Chapter 3 discussed the inference for the abrupt change model (2.1) when change occurred in mean only. Under this case, the partial sums of the two-sided random walk follow normal distribution. In this chapter, the inference for change-point in both mean and covariance is discussed. The asymptotic distribution of the change-point is still formulated as a maximizer of a two-sided random walk defined in (2.27) . Due to the complexity of the parameter change, the partial sums involves a linear combination of noncentral chi-squared distribution.

4.1 MLE of a Change-point in Mean and Covariance of a Multivariate Gaussian Series

In section 4.1.1, the asymptotic distribution of the change-point MLE is first derived for multivariate Gaussian series with change in both mean and covariance, which turns out to be a linear combination of independent noncentral chi-square distribution. Then the detailed method about how to compute a linear combination of chi-square distribution is discussed in section 4.1.2. Finally the algorithmic procedure and some results are presented in section 4.1.3.

4.1.1 *Asymptotic distribution of change-point MLE*

Let $Y_1, Y_2, \dots, Y_n$ be a sequence of time series valued independent random vectors such that $Y_i \in \mathbb{R}^{\mathrm{d}}, i = 1, \dots, n$. The mean vector of the series changes from μ_0 to μ_1, and the covariance matrix changes from Σ_0 to Σ_1 at some unknown point τ_n such that $\mu_0 \neq \mu_1$ and $\Sigma_0 \neq \Sigma_1$. Both the mean vector and covariance matrices are unknown. If a change-point is indeed detected, one would like to estimate confidence interval of the change-point and see how accurate the detected change-point is. The change-point mle has been modeled as a two-sided random walk in (2.27) , where $\{Y^o, Y_i^o : i \in \mathbb{N}\}$ is a sequence of i.i.d. random vectors such that Y^o is distributed according to $f(\cdot; \mu_0, \Sigma_0)$ and $\{Y^*, Y_i^* : i \in \mathbb{N}\}$ is another sequence of i.i.d. random vectors such that Y^* is distributed according to $f(\cdot; \mu_1, \Sigma_1)$. Furthermore the two sequences are independent of each other. It follows that the sequences $\{X^o, X_i^o : i \in \mathbb{N}\}$ and $\{X^*, X_i^* : i \in \mathbb{N}\}$ defined in (2.27) are independent.

As Σ_0 is a positive definite symmetric matrix, it can be decomposed to $\Sigma_0 = Q \Lambda Q^T$, where Q is an orthogonal matrix, and Λ is a real diagonal matrix with positive entries $\lambda_1, \dots, \lambda_d$. As discussed in Chapter 3, the covariance matrix Σ_0 and Σ_1 can be decomposed as

$$\Sigma_0 = \Sigma_0^{1/2} \Sigma_0^{1/2}, \qquad \Sigma_0^{-1} = \Sigma_0^{-1/2} \Sigma_0^{-1/2} \tag{4.1}$$

$$\Sigma_1 = \Sigma_1^{1/2} \Sigma_1^{1/2}, \qquad \Sigma_1^{-1} = \Sigma_1^{-1/2} \Sigma_1^{-1/2}$$

where $\Sigma_0^{1/2}$, $\Sigma_0^{-1/2}$, $\Sigma_1^{1/2}$, and $\Sigma_1^{-1/2}$ are all positive definite symmetric matrices.

Before the change-point, the random variable Y^o admits the presentation

$$Y^o =^D \mu_0 + \Sigma_0^{1/2} Z \tag{4.2}$$

or

$$Y^o - \mu_0 =^D \Sigma_0^{1/2} Z \tag{4.3}$$

where Z is the standard multivariate normal random variable. Consequently, the random variable X^o may be expressed as

$$X^o = -\ln \frac{f(Y^o; \mu_0, \Sigma_0)}{f(Y^o; \mu_1, \Sigma_1)}$$

$$= -\ln \frac{\frac{1}{(2\pi)^{d/2}|\Sigma_0|^{1/2}} \exp\left(-\frac{1}{2}(Y^o - \mu_0)^T \Sigma_0^{-1} (Y^o - \mu_0)\right)}{\frac{1}{(2\pi)^{d/2}|\Sigma_1|^{1/2}} \exp\left(-\frac{1}{2}(Y^o - \mu_1)^T \Sigma_1^{-1} (Y^o - \mu_1)\right)}$$

$$= \ln\left(|\Sigma_0|^{1/2} |\Sigma_1|^{-1/2}\right) + \frac{1}{2}(Y^o - \mu_0)^T \Sigma_0^{-1} (Y^o - \mu_0)$$

$$- \frac{1}{2}(Y^o - \mu_0 + \mu_0 - \mu_1)^T \Sigma_1^{-1} (Y^o - \mu_0 + \mu_0 - \mu_1)$$

$$=^D \ln\left|\Sigma_0^{1/2} \Sigma_1^{-1/2}\right| + \frac{1}{2}\left(\Sigma_0^{1/2} Z\right)^T \Sigma_0^{-1} \left(\Sigma_0^{1/2} Z\right)$$

$$- \frac{1}{2}\left(\Sigma_0^{1/2} Z + \mu_0 - \mu_1\right)^T \Sigma_1^{-1} \left(\Sigma_0^{1/2} Z + \mu_0 - \mu_1\right)$$

$$= \ln\left|\Sigma_0^{1/2}\Sigma_1^{-1/2}\right| + \frac{1}{2}Z^T\Sigma_0^{1/2}\Sigma_0^{-1}\Sigma_0^{1/2}Z - \frac{1}{2}\left(\Sigma_0^{1/2}Z\right)^T\Sigma_1^{-1}\left(\Sigma_0^{1/2}Z\right)$$

$$-\frac{1}{2}(\mu_0 - \mu_1)^T\Sigma_1^{-1}\left(\Sigma_0^{1/2}Z\right)$$

$$-\frac{1}{2}\left(\Sigma_0^{1/2}Z\right)^T\Sigma_1^{-1}(\mu_0 - \mu_1)$$

$$-\frac{1}{2}(\mu_0 - \mu_1)^T\Sigma_1^{-1}(\mu_0 - \mu_1)$$

$$= \ln\left|\Sigma_0^{1/2}\Sigma_1^{-1/2}\right| + \frac{1}{2}Z^TZ - \frac{1}{2}Z^T\Sigma_0^{1/2}\Sigma_1^{-1}\Sigma_0^{1/2}Z$$

$$-\frac{1}{2}(\mu_0 - \mu_1)^T\Sigma_1^{-1}\Sigma_0^{1/2}Z - \frac{1}{2}Z^T\Sigma_0^{1/2}\Sigma_1^{-1}(\mu_0 - \mu_1)$$

$$-\frac{1}{2}(\mu_0 - \mu_1)^T\Sigma_1^{-1}(\mu_0 - \mu_1)$$

$$= \ln\left|\Sigma_0^{1/2}\Sigma_1^{-1/2}\right| - \frac{1}{2}(\mu_0 - \mu_1)^T\Sigma_1^{-1}(\mu_0 - \mu_1) + \frac{1}{2}Z^TZ$$

$$-\frac{1}{2}Z^T\Sigma_0^{1/2}\Sigma_1^{-1/2}\Sigma_1^{-1/2}\Sigma_0^{1/2}Z - (\mu_0 - \mu_1)^T\Sigma_1^{-1}\Sigma_0^{1/2}Z$$

$$= \ln\left|\Sigma_0^{1/2}\Sigma_1^{-1/2}\right| - \frac{1}{2}(\mu_1 - \mu_0)^T\Sigma_1^{-1}(\mu_1 - \mu_0)$$

$$+\frac{1}{2}Z^T\left(I - \Sigma_0^{1/2}\Sigma_1^{-1/2}\Sigma_1^{-1/2}\Sigma_0^{1/2}\right)Z$$

$$+ (\mu_1 - \mu_0)^T\Sigma_1^{-1}\Sigma_0^{1/2}Z$$

Let

$$K = \Sigma_0^{1/2}\Sigma_1^{-1/2} \tag{4.4}$$

$$\eta = \Sigma_0^{-1/2}(\mu_1 - \mu_0) \tag{4.5}$$

then $K^T = \left(\Sigma_0^{1/2}\Sigma_1^{-1/2}\right)^T = \Sigma_1^{-1/2}\Sigma_0^{1/2}$, and $\eta^T = \left(\Sigma_0^{-1/2}(\mu_1 - \mu_0)\right)^T = (\mu_1 - \mu_0)^T\Sigma_0^{-1/2}$. η can be regarded as the standardized amount of change in the mean vector, and K can be regarded as the amount of change in the covariance matrix.

Continuing the above derivation for X^o as follows

$$
\begin{aligned}
X^o =^D \ln\left|\Sigma_0^{1/2}\Sigma_1^{-1/2}\right| &- \frac{1}{2}(\mu_1 - \mu_0)^T\Sigma_1^{-1}(\mu_1 - \mu_0) \\
&+ \frac{1}{2}Z^T\left(I - \Sigma_0^{1/2}\Sigma_1^{-1/2}\Sigma_1^{-1/2}\Sigma_0^{1/2}\right)Z \\
&+ (\mu_1 - \mu_0)^T\Sigma_1^{-1}\Sigma_0^{1/2}Z \\
= \ln|K| &- \frac{1}{2}(\mu_1 - \mu_0)^T\Sigma_0^{-1/2}\Sigma_0^{1/2}\Sigma_1^{-1/2}\Sigma_1^{-1/2}\Sigma_0^{1/2}\Sigma_0^{-1/2}(\mu_1 - \mu_0) \\
&+ \frac{1}{2}Z^T\left(I - \Sigma_0^{1/2}\Sigma_1^{-1/2}\Sigma_1^{-1/2}\Sigma_0^{1/2}\right)Z \\
&+ (\mu_1 - \mu_0)^T\Sigma_0^{-1/2}\Sigma_0^{1/2}\Sigma_1^{-1/2}\Sigma_1^{-1/2}\Sigma_0^{1/2}Z \\
= \ln|K| &- \frac{1}{2}\eta^T KK^T\eta + \frac{1}{2}Z^T(I - KK^T)Z + \eta^T KK^T Z
\end{aligned}
$$

In order to investigate the distribution of the random variable X^o, we can make use of the quadratic form of the multivariate normal random variable. On the right-hand-side, we already have a quadratic term and a linear term in the standard multivariate normal variable Z. Notice if $a = (I - KK^T)^{-1}KK^T\eta$, then

$$\frac{1}{2}(Z+a)^T(I-KK^T)(Z+a)$$

$$=\frac{1}{2}Z^T(I-KK^T)Z+a^T(I-KK^T)Z+\frac{1}{2}a^T(I-KK^T)a$$

$$=\frac{1}{2}Z^T(I-KK^T)Z+((I-KK^T)^{-1}KK^T\eta)^T(I-KK^T)Z$$

$$+\frac{1}{2}((I-KK^T)^{-1}KK^T\eta)^T(I$$

$$-KK^T)((I-KK^T)^{-1}KK^T\eta)$$

$$=\frac{1}{2}Z^T(I-KK^T)Z+(\eta^T KK^T(I-KK^T)^{-1})(I-KK^T)Z$$

$$+\frac{1}{2}(\eta^T KK^T(I-KK^T)^{-1})(I$$

$$-KK^T)((I-KK^T)^{-1}KK^T\eta)$$

$$=\frac{1}{2}Z^T(I-KK^T)Z+\eta^T KK^T Z+\frac{1}{2}\eta^T KK^T(I-KK^T)^{-1}KK^T\eta$$

Thus

$$X^{\mathrm{o}} =^D \ln|K| -\frac{1}{2}\eta^T KK^T\eta+\frac{1}{2}Z^T(I-KK^T)Z+\eta^T KK^T Z$$

$$=\ln|K| -\frac{1}{2}\eta^T KK^T\eta+\frac{1}{2}(Z+a)^T(I-KK^T)(Z+a)$$

$$-\frac{1}{2}\eta^T KK^T(I-KK^T)^{-1}KK^T\eta$$

$$= \ln|K| \ -\frac{1}{2}\eta^T KK^T\eta$$

$$+\frac{1}{2}[Z + (I - KK^T)^{-1}KK^T\eta]^T(I - KK^T)[Z$$

$$+ (I - KK^T)^{-1}KK^T\eta]$$

$$-\frac{1}{2}\eta^T KK^T(I - KK^T)^{-1}KK^T\eta$$

Because KK^T is also a positive definite matrix, it can also be decomposed as

$$KK^T = \Theta\Psi\Theta^T \tag{4.6}$$

where $\Psi = \text{diag}\{\psi_1, \psi_2, \ldots, \psi_d\}, \qquad \Theta^{-1} = \Theta^{\mathrm{T}}$

Θ is an orthogonal matrix, and $\psi_1, \psi_2, \ldots, \psi_d$ are eigenvalues of KK^T. Thus $|K| = |KK^T|^{1/2} = (\psi_1\psi_2 \ldots \psi_d)^{1/2}$.

Let

$$\Theta^{\mathrm{T}}\eta := \omega = [\omega_1, \omega_2, \ldots, \omega_d]^T \tag{4.7}$$

then the distribution of X^o is

$$X^o =^D \ln(\psi_1\psi_2 \ldots \psi_d)^{\frac{1}{2}} - \frac{1}{2}\eta^T\Theta\Psi\Theta^T\eta$$

$$+\frac{1}{2}[Z + (I - \Theta\Psi\Theta^T)^{-1}\Theta\Psi\Theta^T\eta]^T(I - \Theta\Psi\Theta^T)[Z$$

$$+ (I - \Theta\Psi\Theta^T)^{-1}\Theta\Psi\Theta^T\eta]$$

$$-\frac{1}{2}\eta^T\Theta\Psi\Theta^T(I - \Theta\Psi\Theta^T)^{-1}\Theta\Psi\Theta^T\eta$$

$$= \ln(\psi_1\psi_2 \ldots \psi_d)^{1/2} \ -\frac{1}{2}\omega^T\Psi\omega$$

$$+\frac{1}{2}\{Z + [\Theta(\mathrm{I}-\Psi)\Theta^T]^{-1}\Theta\Psi\omega\}^T[\Theta(\mathrm{I}-\Psi)\Theta^T]^{-1}\{Z$$

$$+ [\Theta(\mathrm{I}-\Psi)\Theta^T]^{-1}\Theta\Psi\omega\}$$

$$-\frac{1}{2}\omega^T\Psi\Theta^T[\Theta(\mathrm{I}-\Psi)\Theta^T]^{-1}\Theta\Psi\omega$$

$$= \ln(\psi_1\psi_2 \ldots \psi_d)^{1/2} \ -\frac{1}{2}\omega^T\Psi\omega$$

$$+\frac{1}{2}\{Z + \Theta(\mathrm{I}-\Psi)^{-1}\Theta^T\Theta\Psi\omega\}^T\Theta(\mathrm{I}-\Psi)\Theta^T\{Z$$

$$+ \Theta(\mathrm{I}-\Psi)^{-1}\Theta^T\Theta\Psi\omega\}$$

$$-\frac{1}{2}\omega^T\Psi\Theta^T\Theta(\mathrm{I}-\Psi)^{-1}\Theta^T\Theta\Psi\omega$$

$$= \ln(\psi_1\psi_2 \ldots \psi_d)^{1/2} \ -\frac{1}{2}\omega^T\Psi\omega$$

$$+\frac{1}{2}\{Z + \Theta(\mathrm{I}-\Psi)^{-1}\Psi\omega\}^T\Theta(\mathrm{I}-\Psi)\Theta^T\{Z$$

$$+ \Theta(\mathrm{I}-\Psi)^{-1}\Psi\omega\} - \frac{1}{2}\omega^T\Psi(\mathrm{I}-\Psi)^{-1}\Psi\omega$$

$$= \ln(\psi_1\psi_2 \ldots \psi_d)^{1/2} \ -\frac{1}{2}\omega^T\Psi\omega$$

$$+\frac{1}{2}\{\Theta^T Z + \Theta^T\Theta(\mathrm{I}-\Psi)^{-1}\Psi\omega\}^T(\mathrm{I}-\Psi)\{\Theta^T Z$$

$$+ \Theta^T\Theta(\mathrm{I}-\Psi)^{-1}\Psi\omega\} - \frac{1}{2}\omega^T\Psi(\mathrm{I}-\Psi)^{-1}\Psi\omega$$

$$= \ln(\psi_1\psi_2 \dots \psi_d)^{1/2} - \frac{1}{2}\omega^T\Psi\omega$$

$$+ \frac{1}{2}\{\Theta^T Z + (\mathrm{I} - \Psi)^{-1}\Psi\omega\}^T(\mathrm{I} - \Psi)\{\Theta^T Z$$

$$+ (\mathrm{I} - \Psi)^{-1}\Psi\omega\} - \frac{1}{2}\omega^T\Psi(\mathrm{I} - \Psi)^{-1}\Psi\omega$$

As Θ^T is an orthogonal matrix, Z follows standard multivariate normal distribution, then $\Theta^T Z$ also follows standard normal distribution. Thus

$$X^o =^D \ln(\psi_1\psi_2 \dots \psi_d)^{1/2} - \frac{1}{2}\sum_{s=1}^{d}\psi_s\omega_s^2 \tag{4.8}$$

$$+ \frac{1}{2}\sum_{s=1}^{d}(1 - \psi_s)\left(z_s + \frac{\psi_s\omega_s}{1 - \psi_s}\right)^2 - \frac{1}{2}\sum_{s=1}^{d}\frac{\psi_s^2\omega_s^2}{1 - \psi_s}$$

$$= \ln(\psi_1\psi_2 \dots \psi_d)^{1/2} - \frac{1}{2}\sum_{s=1}^{d}\left(\psi_s + \frac{\psi_s^2}{1 - \psi_s}\right)\omega_s^2$$

$$+ \frac{1}{2}\sum_{s=1}^{d}(1 - \psi_s)\left(z_s + \frac{\psi_s\omega_s}{1 - \psi_s}\right)^2$$

$$= \ln(\psi_1\psi_2 \dots \psi_d)^{1/2} - \frac{1}{2}\sum_{s=1}^{d}\frac{\psi_s\omega_s^2}{1 - \psi_s}$$

$$+ \sum_{i=1}^{d}\frac{1 - \psi_i}{2}\left(z_i + \frac{\psi_i\omega_i}{1 - \psi_i}\right)^2$$

That is to say, the distribution of X^o is the same as

$$C^o + \sum_{s=1}^{d} a_s^o \chi^2_{1,\sigma_s^{o2}} \tag{4.9}$$

which is the sum of a constant term C^o and a linear combination of noncentral chi-square random variables with the degree of freedom 1, the noncentral parameter σ_s^{o2} and the coefficient a_s^o, where

$$C^o = \frac{1}{2}\ln(\psi_1\psi_2 \ldots \psi_d) - \frac{1}{2}\sum_{s=1}^{d} \frac{\psi_s\omega_s^2}{1-\psi_s} \tag{4.10}$$

$$\sigma_s^{o\,2} = \left(\frac{\psi_s\omega_s}{1-\psi_s}\right)^2 \tag{4.11}$$

$$a_s^o = \frac{1}{2}(1-\psi_s) \tag{4.12}$$

For the two-sided random walk for $\xi_n = \hat{\tau}_n - \tau_n$ as defined in (2.27) , in the partial sum $S_{-j}^o = \sum_{i=1}^{-j} X_i^o$ where $j \in \{-1, \ldots, -\tau_n - 1\}$, all the X_i^o have independent and identical distribution as defined in (4.9), which can be rewritten as

$$C^o + \sum_{s=1}^{d} a_{i;s}^o \chi_{i;1,\sigma_s^{o\,2}}^2 \tag{4.13}$$

wher $a_{i;s}^o = a_s^o$, $\chi_{i;1,\sigma_s^{o\,2}}^2 = \chi_{1,\sigma_s^{o\,2}}^2$

e

The $-j^{th}$ partial sum of X_i^o , S_{-j}^o where $j \in \{-1, \ldots, -\tau_n - 1\}$, has the same distribution as

$$\sum_{i=1}^{-j}\left(C^o + \sum_{s=1}^{d} a_{i;s}^o \chi_{i;1,\sigma_s^{o\,2}}^2\right) \tag{4.14}$$

$$= -jC^o + \sum_{i=1}^{-j}\left(\sum_{s=1}^{d} a_{i;s}^o \chi_{i;1,\sigma_s^{o\,2}}^2\right)$$

which is the sum of a constant term $-jC^o$ and jd terms of noncentral chi-square random variables with the degree of freedom being 1. In the linear combination of chi-square random terms for S_{-j}^o where $j \in \{-1, \ldots, -\tau_n - 1\}$, $-j$ terms have the noncentral parameter being σ_s^{o2}, and the coefficient for chi-square random variables being a_s^o for $s = 1, 2, \ldots, d$.

After the change-point τ_n, the observations Y^* follow multivariate normal distribution $N(\mu_1, \Sigma_1)$. We can write

$$Y^* =^D \mu_1 + \Sigma_1^{1/2} Z \tag{4.15}$$

or

$$Y^* - \mu_1 =^D \Sigma_1^{1/2} Z \tag{4.16}$$

where Z is the standard multivariate normal random variable. Consequently, the random variable X^* may be expressed as

$$X^* = \ln \frac{f(Y^*; \mu_0, \Sigma_0)}{f(Y^*; \mu_1, \Sigma_1)}$$

$$= \ln \frac{\frac{1}{(2\pi)^{d/2}|\Sigma_0|^{1/2}} \exp\left(-\frac{1}{2}(Y^* - \mu_0)^T \Sigma_0^{-1} (Y^* - \mu_0)\right)}{\frac{1}{(2\pi)^{d/2}|\Sigma_1|^{1/2}} \exp\left(-\frac{1}{2}(Y^* - \mu_1)^T \Sigma_1^{-1} (Y^* - \mu_1)\right)}$$

$$= -\ln\left(|\Sigma_0|^{1/2} |\Sigma_1|^{-1/2}\right)$$

$$-\frac{1}{2}(Y_i - \mu_1 + \mu_1 - \mu_0)^T \Sigma_0^{-1} (Y_i - \mu_1 + \mu_1 - \mu_0)$$

$$+\frac{1}{2}(Y_i - \mu_1)^T \Sigma_1^{-1} (Y_i - \mu_1)$$

$$=^D - \ln\left|\Sigma_0^{1/2} \Sigma_1^{-1/2}\right| - \frac{1}{2}\left(\Sigma_1^{1/2} Z + \mu_1 - \mu_0\right)^T \Sigma_0^{-1} \left(\Sigma_1^{1/2} Z + \mu_1 - \mu_0\right)$$

$$+\frac{1}{2}\left(\Sigma_1^{1/2} Z\right)^T \Sigma_1^{-1} \left(\Sigma_1^{1/2} Z\right)$$

$$= -\ln\left|\Sigma_0^{1/2}\Sigma_1^{-1/2}\right| - \frac{1}{2}\left(\Sigma_1^{1/2}Z\right)^T\Sigma_0^{-1}\left(\Sigma_1^{1/2}Z\right)$$

$$-\frac{1}{2}\left(\Sigma_1^{1/2}Z\right)^T\Sigma_0^{-1}(\mu_1-\mu_0) - \frac{1}{2}(\mu_1-\mu_0)^T\Sigma_0^{-1}\left(\Sigma_1^{1/2}Z\right)$$

$$-\frac{1}{2}(\mu_1-\mu_0)^T\Sigma_0^{-1}(\mu_1-\mu_0) + \frac{1}{2}Z^T\Sigma_1^{1/2}\Sigma_1^{-1}\Sigma_1^{1/2}Z$$

$$= -\ln\left|\Sigma_0^{1/2}\Sigma_1^{-1/2}\right| - \frac{1}{2}Z^T\Sigma_1^{1/2}\Sigma_0^{-1/2}\Sigma_0^{-1/2}\Sigma_1^{1/2}Z$$

$$-(\mu_1-\mu_0)^T\Sigma_1^{-1/2}\Sigma_1^{1/2}\Sigma_0^{-1/2}\Sigma_0^{-1/2}\Sigma_1^{1/2}Z$$

$$-\frac{1}{2}(\mu_1-\mu_0)^T\Sigma_1^{-1/2}\Sigma_1^{1/2}\Sigma_0^{-1/2}\Sigma_0^{-1/2}\Sigma_1^{1/2}\Sigma_1^{-1/2}(\mu_1$$

$$-\mu_0) + \frac{1}{2}Z^TZ$$

The same parameterization is used for X^* as for X^o as in (4.4). As $K = \Sigma_0^{1/2}\Sigma_1^{-1/2}$, then $K^T = \left(\Sigma_0^{1/2}\Sigma_1^{-1/2}\right)^T = \Sigma_1^{-1/2}\Sigma_0^{1/2}$, $K^{-1} = \left(\Sigma_0^{1/2}\Sigma_1^{-1/2}\right)^{-1} = \Sigma_1^{1/2}\Sigma_0^{-1/2}$, $(K^{-1})^T = \Sigma_0^{-1/2}\Sigma_1^{1/2}$.

Let

$$\eta^* = \Sigma_1^{-1/2}(\mu_1-\mu_0) \tag{4.17}$$

then $\eta^{*T} = \left(\Sigma_1^{-1/2}(\mu_1-\mu_0)\right)^T = (\mu_1-\mu_0)^T\Sigma_1^{-1/2}$. The distribution of X^* can be derived as follows

$$X^* =^D \ln|K^{-1}| - \frac{1}{2}Z^TK^{-1}(K^{-1})^TZ - \eta^{*T}K^{-1}(K^{-1})^TZ$$

$$-\frac{1}{2}\eta^TK^{-1}(K^{-1})^T\eta + \frac{1}{2}Z^TZ$$

$$= \ln|K^{-1}| - \frac{1}{2}\eta^T K^{-1}(K^{-1})^T\eta + \frac{1}{2}Z^T[I - K^{-1}(K^{-1})^T]Z$$
$$- \eta^{*T}K^{-1}(K^{-1})^T Z$$

All the terms have exactly the same structure as X^o if the terms K and η for X^o are substituted by the terms K^{-1} and η^* for X^*. So we can complete a quadratic term for the distribution of X^* the same way as what has been done for X^o as follows:

$$X^* =^D \ln|K^{-1}| - \frac{1}{2}\eta^{*T}K^{-1}(K^{-1})^T\eta^*$$
$$+ \frac{1}{2}[Z - (I - K^{-1}(K^{-1})^T)^{-1}K^{-1}(K^{-1})^T\eta^*]^T(I$$
$$- K^{-1}(K^{-1})^T)[Z - (I - K^{-1}(K^{-1})^T)^{-1}K^{-1}(K^{-1})^T\eta^*]$$
$$- \frac{1}{2}\eta^{*T}K^{-1}(K^{-1})^T(I - K^{-1}(K^{-1})^T)^{-1}K^{-1}(K^{-1})^T\eta^*$$

In deriving the distribution for X^o, KK^T is decomposed as $KK^T = \Theta\Psi\Theta^T$ in (4.6), where Θ is an orthogonal matrix, $\Psi = diag\{\psi_1, \psi_2, \ldots, \psi_d\}$, and $\psi_1, \psi_2, \ldots, \psi_d$ are eigenvalues of KK^T. Thus

$$K^{-1}(K^{-1})^T = (KK^T)^{-1} = (\Theta\Psi\Theta^T)^{-1} = \Theta\Psi^{-1}\Theta^T \quad (4.18)$$

where $\Psi^{-1} = diag\{\psi_1^{-1}, \psi_2^{-1}, \ldots, \psi_d^{-1}\}$, and $\Theta^{-1} = \Theta^{\mathrm{T}}$

$\psi_1^{-1}, \psi_2^{-1}, \ldots, \psi_d^{-1}$ are eigenvalues of $K^{-1}(K^{-1})^T$.

Let

$$\Theta^{\mathrm{T}}\eta^* := \omega^* = [\omega_1^*, \omega_2^*, \ldots, \omega_d^*]^T \quad (4.19)$$

then the distribution of X^* is

$$X^* =^D \ln(\psi_1\psi_2 \dots \psi_d)^{-1/2} - \frac{1}{2}\omega^{*T}\Psi^{-1}\omega^* \qquad (4.20)$$

$$+\frac{1}{2}[\Theta^{\mathrm{T}}Z - (I-\Psi^{-1})^{-1}\Psi^{-1}\omega^*]^T(I-\Psi^{-1})[\Theta^{\mathrm{T}}Z$$

$$-(I-\Psi^{-1})^{-1}\Psi^{-1}\omega^*]$$

$$-\frac{1}{2}\omega^{*T}\Psi^{-1}(I-\Psi^{-1})^{-1}\Psi^{-1}\omega^*$$

$$= \ln(\psi_1\psi_2 \dots \psi_d)^{-1/2} - \frac{1}{2}\sum\nolimits_{s=1}^{d}\psi_s^{-1}\omega_s^{*2}$$

$$+\frac{1}{2}\sum\nolimits_{s=1}^{d}(1-\psi_s^{-1})\left(z_s - \frac{\psi_s^{-1}\omega_s^*}{1-\psi_s^{-1}}\right)^2$$

$$-\frac{1}{2}\sum\nolimits_{s=1}^{d}\frac{\psi_s^{-2}\omega_s^{*2}}{1-\psi_s^{-1}}$$

$$= \ln(\psi_1\psi_2 \dots \psi_d)^{-1/2} - \frac{1}{2}\sum\nolimits_{s=1}^{d}\frac{\psi_s^{-1}\omega_s^{*2}}{1-\psi_s^{-1}}$$

$$+\sum\nolimits_{s=1}^{d}\frac{1-\psi_s^{-1}}{2}\left(z_s - \frac{\psi_s^{-1}\omega_s^*}{1-\psi_s^{-1}}\right)^2$$

That is to say, the distribution of X^* is the same as

$$C^* + \sum\nolimits_{s=1}^{d} a_s^*\chi^2_{1,\sigma_s^{*2}} \qquad (4.21)$$

which is the sum of the constant term C^* and a linear combination of noncentral chi-square random variables with, the degree of freedom 1, the noncentral parameter σ_s^{*2} and the coefficient a_s^*, where

$$C^* = -\frac{1}{2}\ln(\psi_1\psi_2 \dots \psi_d) - \frac{1}{2}\sum\nolimits_{s=1}^{d}\frac{\psi_s^{-1}\omega_s^{*2}}{1-\psi_s^{-1}} \qquad (4.22)$$

$$\sigma_s^{*2} = \left(\frac{\psi_s^{-1}\omega_i^*}{1-\psi_s^{-1}}\right)^2 \tag{4.23}$$

$$a_s^* = \frac{1}{2}(1-\psi_s^{-1}) \tag{4.24}$$

For the two-sided random walk for $\xi_n = \hat{\tau}_n - \tau_n$ as defined in (2.27) , in the partial sum $S_j^* = \sum_{i=1}^{j} X_i$ where $j \in \{1, \ldots, n-\tau_n-1\}$, all the X_i^* have independent and identical distribution as defined in (4.21). The j^{th} partial sum of X_i^*, S_j^*, has the same distribution as

$$\sum\nolimits_{i=1}^{j}\left(C^* + \sum\nolimits_{s=1}^{d} a_{i;s}^*\chi_{i;1,\sigma_s^{o\,2}}^2\right) = jC^* + \sum\nolimits_{i=1}^{j}\left(\sum\nolimits_{s=1}^{d} a_{i;s}^*\chi_{i;1,\sigma_s^{*2}}^2\right) \tag{4.25}$$

which is the sum of a constant term jC^* and jd terms of noncentral chi-square random variables with the degree of freedom being 1. In the linear combination of chi-squared random terms for S_j^*, j terms have the noncentral parameter being σ_s^{*2}, and the coefficient for chi-square random variable being a_s^* for $s = 1, 2, \ldots, d$.

Both linear combinations of chi-square distribution have a constant term and $|jd|$ terms of independent non-central chi-squared random variables with 1 degree of freedom. The distribution change-point can be derived by the maximizer of the two-sided random walk defined in (2.27) . As the derivation involves computing a linear combination of noncentral chi-square distribution, it is not as straightforward as the case that was discussed in Chapter 3, where the parameter change is confined to mean only. In the following sections, the computation for the linear combination of chi-square distribution will be discussed before presenting the algorithmic procedure for the distribution of change-point mle.

4.1.2 Distribution of linear combination of chi-square distribution

From above derivation, the distribution theory of the change-point estimate can be established by the two-sided random walk defined by X^o and X^*. Under the univariate setup, according to Fotopoulos and Jandhyala (2000), each step follows normal distribution if assuming change only in mean value, or chi-square distribution if change in variance is also assumed. As the dimensionality of the observation increases, it seems necessary to find out a way to compute the distribution of a linear combination of non-central chi-square distribution. As we would like to generalized our method and derive a fast algorithm for the estimation method without the loss of accuracy, the method about how to deal with the linear combination is critical. In the literature, several methods were proposed to achieve a balance of speed and accuracy. Imhof(1961) gave exact and approximate method for computing the distribution of the form

$$Q = \sum_{s=1}^{d} a_s \chi^2_{h_s,\sigma_s^2} \tag{4.26}$$

Gil-Pelaez's (1951) numerical inversion of the characteristic function was used to obtain the exact form of the cumulative distribution function as

$$F(q) = \frac{1}{2} - \frac{1}{\pi}\int_0^{\infty} \frac{\sin\theta(u)}{u\rho(u)}\,du \tag{4.27}$$

where

$$\theta(u) = \frac{1}{2}\sum_{s=1}^{d}[\, h_s \tan^{-1}(a_s u) + \sigma_s^2 a_s u(1 + a_s^2 u^2)^{-1}] - \frac{1}{2}qu \tag{4.28}$$

$$\rho(u) = \prod_{s=1}^{d} (1 + a_s^2 u^2)^{\frac{1}{4}h_s} \exp\left\{\frac{1}{2}\sum_{s=1}^{d} \frac{(\sigma_s a_s u)^2}{(1 + a_s^2 u^2)}\right\} \tag{4.29}$$

As the integrand of the improper integral satisfies $\lim_{u\to\infty} \frac{\sin\,\theta(u)}{u\rho\,(u)} = 0$, the numerical integration for approximation was carried on a finite range $[0, U]$ in Imhof (1961), where U is determined by the accuracy requirement of the approximation. Imhof (1961) showed two sources of errors: the error of integration from using numerical integration and the error of truncation from using the finite-range integral. Imhof (1961) also showed that if the upper integration limit is U, the upper bound of the truncation error T_U can be determined by

$$T_U = \left\{\pi k U^k \prod_{s=1}^{d} |a_s|^{\frac{1}{2}h_s} \exp\left[\frac{1}{2}\sum_{s=1}^{d} \frac{(\sigma_s a_s U)^2}{(1 + a_s^2 U^2)}\right]\right\}^{-1} \tag{4.30}$$

where

$$k = \frac{1}{2}\sum_{s=1}^{d} h_s$$

Thus if T_U is set as the accuracy, the corresponding U guarantees that the truncation error will not exceed the accuracy requirement. The trapezoidal rule and Simpson's rule were proposed by Imhof (1961) to compute the truncated integral. The integration method is computing intensive, as the length of the integration interval was determined by trial-and-error method, until a desired accuracy was achieved. Koerts and Abrahamse (1969) provided the FORTRAN program for Imhof's (1961) method. Farebrother (1990) gave the Pascal translation of Koerts and Abrahames's (1969) Fortran procedure with minor modification in implementation.

Davies (1973, 1980) also followed Gil-Pelaez's (1951) inversion formula of the characteristic function. Fourier cosine series summation formula was used to find a bound on the integration error when the numerical integration by trapezoidal rule was applied.

The distribution function $\Pr\left(\sum_{s=1}^{d} a_s \chi^2_{h_s,\sigma_s^2} \leq q\right)$ was computed by Davies (1980) as

$$\frac{1}{2} - \sum_{l=0}^{U} \left\{ \exp\left[-2b_l^2 \sum_{s=1}^{d} a_s^2 \sigma_s^2 / (1 + 4b_l^2 a_s^2)\right] \prod_{s=1}^{d} (1 + 4b_l^2 a_s^2)^{-h_s/4} \sin\left\{ \sum_{s=1}^{d} [h_s \arctan(2b_l a_s)/2 + b_l a_s \sigma_s^2 / (1 + 4b_l^2 a_s^2)] - b_l q \right\} / \left[\pi\left(l + \frac{1}{2}\right)\right] \right\}$$

where $b_l = \left(l + \frac{1}{2}\right)\Delta$, Δ is the length of subinterval for trapezoidal rule

U is the truncation limit that keeps the truncation error

$$\sum_{l=U+1}^{\infty} \mathrm{Im}[\Phi(b_l) \exp(-\mathrm{i} b_l q)] / \left[\pi\left(l + \frac{1}{2}\right)\right]$$

less that desired tolerance of error.

In the summation for numerical integration, Davies' formula might contain terms with large magnitudes with different signs, which might cumulate significant round-off error, although the author claims that it is not a problem in practice.

Lu (2006) proposed two truncation bounds to control the truncation error. The bounds might be more efficient under certain situations, but cannot be solved analytically. Iterative method such as Newton's method is required, which add the complexity of the calculation. Our concern is the accuracy of the results. If the computation time is not unreasonably long, we would prefer simpler method. What is more, the method is not applicable if the sum of the degrees of freedom of the chi-square random variables are no greater than 2; while we need to calculate the distribution function under this scenario. For above reasons, we will not adopt their methods, even though the new bounds may be more efficient.

Kuonen (1999) proposed saddlepoint approximation to the survival function, and claimed that the method was fast, accurate and easy to program. The method started with the cumulant generating function for (4.26)

$$\kappa(\zeta) = -\frac{1}{2}\sum_{s=1}^{d} h_s \log(1 - 2\zeta \mathrm{a_s}) + \sum_{s=1}^{d} \frac{\sigma_s^2 a_s}{1 - 2\zeta a_s}$$

assuming $1 - 2\zeta a_s$ to be positive. The saddlepoint $\hat{\zeta}$ of $\kappa(\zeta)$ is computed by solving the equation

$$\kappa'(\zeta) = q$$

The corresponding approximation to the cumulative density function,

$\Pr\left(\sum_{s=1}^{d} a_s \chi^2_{h_s,\sigma_s^2}\right) \leq q$, is

$$\Phi\left[w + \frac{1}{w}\log\left(\frac{v}{w}\right)\right]$$

where $$w = sign(\hat{\zeta})\left[2\left(\hat{\zeta}q - \kappa(\hat{\zeta})\right)\right]^{1/2}, v = \hat{\zeta}\left[\kappa''^{(\hat{\zeta})}\right]^{1/2}$$

By surveying the existing methods, the methods for computing the linear combination of chi-square distribution can be categorized into (a) Inversion of the characteristic function, and (b) saddlepoint approximation. Method (a) is the most studied of the two methods. The exact form for the distribution function has been proposed by Imhof. The studies following the inversion method focused on how to truncate the improper integral into a proper one and achieve the desired accuracy at the same time, and on which numerical integration method should be adopted for the proper one. The Simpson's method or trapezoidal rule can achieve desired accuracy, if the integration interval is divided into subintervals that are small enough. As high accuracy is desired, it would be very time consuming to determine the length of the subintervals and then estimate the integral by summation. In this aspect, method (b) seems to have a greater advantage on computation time. However, method (b) assumes $1 - 2\zeta a_s > 0$, which means that we might not be able to find the saddlepoint for certain coefficient a_s .

Kuonen (2003) surveyed the numerical integration in the statistical software package R. He pointed out that the *integrate* function in R implements one-dimensional adaptive 15-point Gauss-Kronrod quadrature, and a 128-point Gauss-Lagrange rule. A set of 71 well-designed test examples were tested using this function, and it delivered very accurate results most of the time.

The following table is a comparison of results using Imhof's estimation, Davies' method, Imhof's exact formula using the numerical integration in R statistical package, and Kuonen's saddlepoint approximation. All the methods were implemented by R program. The test examples are taken from Imhof (1961). As the estimation for change-point required high accuracy, the tolerance of error is chosen to be 10^{-6}.

Table 4.1. Probability of linear combination of chi-squared distribution using (ii) Imhof's (1961) estimation; (iii) Davies' (1973) method; (iv) Imhof's (1961) exact formula using R integration; (v) Saddlepoint approximation by Kuonen (1999). (i) is the true values from Imhof (1961).

	x	(i)	(ii)	(iii)	(iv)	(v)
$Q_1 = 0.6\chi_1^2 + 0.3\chi_1^2 + 0.1\chi_1^2$	0.1	0.0542	0.0542	0.0542	0.0542	0.0551
	0.7	0.4936	0.4936	0.4936	0.4936	0.5004
	2	0.8760	0.8760	0.8760	0.8760	0.8783
$Q_2 = 0.6\chi_2^2 + 0.3\chi_2^2 + 0.1\chi_2^2$	0.2	0.0064	0.0065	0.0065	0.0064	0.0065
	2	0.6001	0.6002	0.6002	0.6002	-
	6	0.9839	0.9839	0.9839	0.9839	0.9838
$Q_3 = 0.6\chi_6^2 + 0.3\chi_4^2 + 0.1\chi_2^2$	1	0.0027	0.0027	0.0027	0.0027	0.0027
	5	0.5647	0.5647	0.5647	0.5647	-
	12	0.9912	0.9912	0.9912	0.9912	0.9912
$Q_4 = 0.6\chi_2^2 + 0.3\chi_4^2 + 0.1\chi_4^2$	1	0.0334	0.0334	0.0334	0.0334	0.0336
	3	0.5802	0.5804	0.5804	0.5804	-
	8	0.9913	0.9913	0.9913	0.9913	0.9913
$Q_5 = 0.7\chi_{6;6}^2 + 0.3\chi_{2;2}^2$	2	0.0061	0.0061	0.0061	0.0061	-
	10	0.5913	0.5913	0.5913	0.5913	-
	20	0.9779	0.9779	0.9779	0.9779	-
$Q_6 = 0.7\chi_{1;6}^2 + 0.3\chi_{1;2}^2$	1	0.0451	0.0451	0.0451	0.0452	-
	6	0.5924	0.5924	0.5924	0.5924	-
	15	0.9777	0.9777	0.9777	0.9777	-
$1/3Q_3 + 2/3Q_4$	1.5	0.0109	0.0109	0.0109	0.0109	0.0110
	4	0.6547	0.6547	0.6547	0.6547	0.6571
	7	0.9846	0.9846	0.9846	0.9846	0.9850

From Table 4.1 we can see that Imhof's (1961) estimation, Davies' (1973) method and Imhof's (1961) exact formula using integration function in R are almost equivalent. Kuonen's (1999) saddlepoint estimation computes the fastest. However, from the above table, it is not as accurate as other methods. If the saddlepoint cannot be found, this method should not be used for the computation for change-point estimation. Computing the distribution of a linear combination of chi-squared distribution is an intermediate step toward the change-point inference, and high accuracy is required. The accuracy is more important than efficiency. Therefore, the Imhof's (1961) exact formula for the distribution function using inversion of characteristic function method will be adopted. The integration function in R will be chosen to evaluate the improper integral to achieve the accuracy and to avoid the complexity of determining the truncation bound.

4.1.3 Algorithmic procedure to compute the change-point mle

As discussed in Section 4.1.1, for change in mean and covariance case, we still modeled the mle as a two-sided random walk defined in (2.29). The algorithmic procedure derived in Jandhyala and Fotopoulos (1999) still applies to this case. This section will state the detailed algorithmic steps that compute the change-point mle assuming change in both the mean vector and the covariance matrix. The derivation for the distribution of the random walk in Section 4.1.1 and the distribution of the linear combination of chi-square distribution in Section 4.1.2 will be applied for the algorithm. The following is the detailed steps.

Step S0: Assume $Y^o \sim N(\mu_0, \Sigma_0)$, $Y^* \sim N(\mu_1, \Sigma_1)$. Compute the parameters derived in 4.1.1 for the distribution for the two-sided random walk. The following parameters are set in Section 4.1.1. Here they are listed for summary purpose only.

$$K = \Sigma_0^{1/2}\Sigma_1^{-1/2}.\ KK^T = \Theta\Psi\Theta^T \text{ where } \Psi = diag\{\psi_1, \psi_2, \dots, \psi_d\},\ \Theta^{-1} = \Theta^{\mathrm{T}}$$

$$\eta = \Sigma_0^{-\frac{1}{2}}(\mu_1 - \mu_0), \qquad \omega = [\omega_1, \omega_2, \dots, \omega_d]^T \Theta^{\mathrm{T}}\eta$$

$$C^o = \frac{1}{2} ln(\psi_1\psi_2 \dots \psi_d) - \frac{1}{2}\sum_{s=1}^{d} \frac{\psi_s \omega_s^2}{1-\psi_s}, \sigma_s^{o2} = \left(\frac{\psi_s \omega_s}{1-\psi_s}\right)^2, a_s^o = \frac{1}{2}(1-\psi_s)$$

$$\eta^* = \Sigma_1^{-\frac{1}{2}}(\mu_1 - \mu_0), \qquad \omega^* = [\omega_1^*, \omega_2^*, \dots, \omega_d^*]^T = \Theta^{\mathrm{T}}\eta^*$$

$$C^* = -\frac{1}{2}\ln(\psi_1\psi_2 \dots \psi_d) - \frac{1}{2}\sum\nolimits_{s=1}^{d} \frac{\psi_s^{-1}\omega_s^{*2}}{1-\psi_s^{-1}}, \sigma_s^{*2} = \left(\frac{\psi_s^{-1}\omega_s^*}{1-\psi_s^{-1}}\right)^2, a_s^*$$

$$= \frac{1}{2}(1-\psi_s^{-1})$$

Step S1: As derived in (4.14) , the partial sum S_{-j}^{o}, where $j \in \{-1, \dots, -\tau_n - 1\}$, has the same distribution as $-jC^o + \sum_{i=1}^{-j}\left(\sum_{s=1}^{d} a_{i;s}^{o}\chi_{i;1,\sigma_s^{o2}}^{2}\right)$ where $a_{i;s}^{o} = a_s^o$ and $\chi_{i;1,\sigma_s^{o2}}^{2} = \chi_{1,\sigma_s^{o2}}^{2}$. The random variable is comprised of a constant term and $-jd$ terms of weighted sum of noncentral chi-square random variables, where $-j$ terms have the noncentral parameter being σ_s^{o2}, and the coefficient for chi-square random variables being a_s^o for $s = 1, 2, \dots, d$. Similarly, as derived in (4.25), S_j^*, where $j \in \{1, \dots, n - \tau_n - 1\}$, has the same distribution as the random variable $jC^* + \sum_{i=1}^{j}\left(\sum_{s=1}^{d} a_{i;s}^{*}\chi_{i;1,\sigma_s^{*2}}^{2}\right)$, where $a_{i;s}^{*} = a_s^*$ and $\chi_{i;1,\sigma_s^{*2}}^{2} = \chi_{1,\sigma_s^{*2}}^{2}$. The random variable is comprised of a constant term and jd terms of weighted sum of noncentral chi-square random variables, where j terms have the noncentral parameter being σ_s^{*2}, and the coefficient for chi-square random variables being a_s^* for $s = 1, 2, \dots, d$.

Step S2: Compute $\{b_j^o\}$ and $\{b_j^*\}$ for $j = 1, 2, \dots$, where

$$b_j^o = \Pr\left(S_j^o > 0\right) = \Pr\left(\sum_{i=1}^{j}\left(\sum_{s=1}^{d} a_{i;s}^{o}\chi_{i;1,\sigma_s^{o2}}^{2}\right) > -jC^o\right)$$

$$b_j^* = \Pr\left(S_j^* > 0\right) = \Pr\left(\sum_{i=1}^{j}\left(\sum_{s=1}^{d} a_{i;s}^{*}\chi_{i;1,\sigma_s^{*2}}^{2}\right) > -jC^*\right)$$

The probability will be computed using Imhof (1961)'s exact formula and the integration function in R which was discussed in section 4.1.2.

Step S3: Compute $B^o(1)$ and $B^*(1)$ as $B^o(1) = \sum b_j^o / j$ and $B^*(1) = \sum b_j^* / j$.

Step S4: Compute both $\{\tilde{b}_j^o(\vartheta^*) = \tilde{b}_j^o(1)\}$ and $\{\tilde{b}_j^*(\vartheta^o) = \tilde{b}_j^*(1)\}$ as $E\left(e^{-S_j^o}\mathrm{I}\left(S_j^o > 0\right)\right)$ and $E\left(e^{-S_j^*}\mathrm{I}\left(S_j^* > 0\right)\right)$, respectively.

$$E\left(e^{-S_j^o}\mathrm{I}\left(\mathrm{S_j^o}>0\right)\right)=\int_0^{\infty}e^{-S^o}f_{S^o}(s^o)ds^o=\lim_{U\to\infty}\int_0^{U}e^{-S^o}f_{S^o}(s^o)ds^o$$

In Section 4.1.2, the probability density function of the linear combination of chi-squared distribution has not been discussed. However, by integration by part, the computation can be converted to cumulative distribution function problem as follows

$$E\left(e^{-S_j^o}\mathrm{I}\left(\mathrm{S_j^o}>0\right)\right)=\int_0^{\infty}e^{-S^o}f_{S^o}(s^o)ds^o=\lim_{U\to\infty}\int_0^{U}e^{-S^o}f_{S^o}(s^o)ds^o$$

$$=\lim_{U\to\infty}\int_0^{U}e^{-S^o}dF_{S^o}(s^o)$$

$$=\lim_{U\to\infty}\left[e^{-s^o}F_{S^o}(s^o)\Big|_0^U-\int_0^{U}F_{S^o}(s^o)de^{-S^o}\right]$$

$$=\lim_{U\to\infty}\left[e^{-U}F_{S^o}(U)-e^{0}F_{S^o}(0)-\int_0^{U}F_{S^o}(s^o)de^{-S^o}\right]$$

$$=0-F_{S^o}(0)-\int_0^{\infty}-e^{-S^o}F_{S^o}(s^o)ds^o$$

$$=-F_{S^o}(0)+\int_0^{\infty}e^{-S^o}F_{S^o}(s^o)ds^o$$

$F_{S^o}(\cdot)$ can be determined by Imhof's exact formula, and the integrate $\int_0^{\infty}e^{-S^o}F_{S^o}(s^o)ds^o$ will be computed by R's integration function. $E\left(e^{-S_j^*}\mathrm{I}\left(\mathrm{S_j^*}>0\right)\right)$ will be computed using the same method.

Step S5: Implement the iterative procedures for $\{q_j^o\}$, $\{\tilde{u}_j^o(\vartheta^*)\}$ and $\{q_j^*\}$, $\{\tilde{u}_j^*(\vartheta^o)\}$ as follows:

$$q_0^o=1,\ jq_j^o=\sum_{k=0}^{j-1}b_{j-k}^o q_k^o\ ;\ \tilde{u}_j^o(\vartheta^*)=1,\ j\tilde{u}_j^o(\vartheta^*)=\sum_{k=0}^{j-1}\tilde{b}_{j-k}^o\tilde{u}_k^o$$

$$q_0^* = 1,\ jq_j^* = \sum_{k=0}^{j-1} b_{j-k}^* q_k^* \ \ ;\ \tilde{u}_j^*(\vartheta^o) = 1,\ j\tilde{u}_j^*(\vartheta^o) = \sum_{k=0}^{j-1} \tilde{b}_{j-k}^* \tilde{u}_k^*$$

Step S6: Estimate $\Pr(\hat{\xi}_n = i)$ by (3.11) as follows

$$\Pr(\hat{\xi}_\infty = i) = \begin{cases} e^{-B^*(1)}\left[q_i^* - \left(1 - e^{-B^o(1)}\right)\tilde{u}_i^*(\vartheta^o)\right], & i > 0 \\ e^{-B^*(1)-B^o(1)}, & i = 0 \\ e^{-B^o(1)}\left[q_i^o - \left(1 - e^{-B^*(1)}\right)\tilde{u}_i^o(\vartheta^*)\right], & i < 0 \end{cases}$$

4.2 Special Cases

After the change-point analysis for change in both mean and covariance in multivariate series solved, the univariate change-point analysis can be regarded as the multivariate series with dimension equals 1, and the change in covariance case can be derived from the case for both mean and covariance by setting the mean before and after change-point being the same. In this section, the special cases are discussed and the change-point estimation is derived.

4.2.1 Mean and Variance of a Univariate Gaussian Series

The univariate change-point problem that assumes change occurs in both mean and variance can be regarded as a special case of the multivariate change-point problem. As the dimension decreases to 1, X^o and X^* contain only the single term of noncentral chi-square random variable. Let $Y_1, Y_2, \dots, Y_n$ be a sequence of time series valued independent random vectors such that $\mathrm{Y_i} \in \mathbb{R}, \mathrm{i} = 1, \dots, \mathrm{n}$. The mean of the series changes from μ_0 to μ_1 at some unknown point τ_n such that $\mu_0 \neq \mu_1$, and the variance changes from σ_0^2 to σ_1^2 at τ_n such that $\sigma_0^2 \neq \sigma_1^2$. Both the mean and variance are unknown. The asymptotic distribution of the maximum likelihood estimator of the change-point $\hat{\tau}_n$ when assuming μ_0, μ_1, σ_0^2 and σ_1^2 are unknown is equivalent to $\tilde{\tau}_n$ when assuming these parameters are known. $\xi_n = \tilde{\tau}_n - \tau_n \in \{-\tau_n + 1, \dots, n - \tau_n - 1\}$ is the maximizer of the two-sided random walk specified in (2.27) .

In the case of change in both mean and variance, $\{Y^o, Y_i^o : i \in \mathbb{N}\}$ is a sequence of i.i.d. random variables such that Y^o is distributed according to $f(\cdot; \mu_0, \sigma_0^2)$ and $\{Y^*, Y_i^* : i \in \mathbb{N}\}$ is another sequence of i.i.d. random variables such that Y^* is distributed according to $f(\cdot; \mu_1, \sigma_1^2)$. Furthermore the two sequences are independent of each other. It follows that the sequences $\{X^o, X_i^o : i \in \mathbb{N}\}$ and $\{X^*, X_i^* : i \in \mathbb{N}\}$ are also independent.

It follows that Y^o, the random variable before the change-point admits the representation

$$Y^o =^D \mu_0 + \sigma_0 \mathrm{Z}$$

or

$$Y^o - \mu_0 =^D \sigma_0 Z$$

where Z is the standard univariate normal vector. Consequently, the random variable X^o may be expressed as

$$X^o = -a(Y^o) = -\ln\frac{f(Y^o;\mu_0,\sigma_0^2)}{f(Y^o;\mu_1,\sigma_1^2)}$$

$$= -\ln\frac{\frac{1}{\sqrt{2\pi}\sigma_0}\exp\left(-\frac{(Y^o-\mu_0)^2}{2\sigma_0^2}\right)}{\frac{1}{\sqrt{2\pi}\sigma_1}\exp\left(-\frac{(Y^o-\mu_1)^2}{2\sigma_1^2}\right)}$$

$$= \ln\frac{\sigma_0}{\sigma_1} + \frac{(Y^o-\mu_0)^2}{2\sigma_0^2} - \frac{(Y^o-\mu_1)^2}{2\sigma_1^2}$$

$$= \ln\frac{\sigma_0}{\sigma_1} + \frac{(Y^o-\mu_0)^2}{2\sigma_0^2} - \frac{(Y^o-\mu_0+\mu_0-\mu_1)^2}{2\sigma_1^2}$$

$$=^D \ln\frac{\sigma_0}{\sigma_1} + \frac{\sigma_1^2 Z^2}{2\sigma_1^2} - \frac{(\sigma_0 Z+\mu_0-\mu_1)^2}{2\sigma_1^2}$$

$$= \ln\frac{\sigma_0}{\sigma_1} + \frac{(\sigma_1^2-\sigma_0^2)Z^2 - 2\sigma_0 Z(\mu_0-\mu_1) - (\mu_0-\mu_1)^2}{2\sigma_1^2}$$

$$= \ln\frac{\sigma_0}{\sigma_1}$$
$$+ \frac{(\sigma_1^2-\sigma_0^2)\left\{Z^2 - \frac{2\sigma_0 Z(\mu_0-\mu_1)}{(\sigma_1^2-\sigma_0^2)} + \left[\frac{\sigma_0(\mu_0-\mu_1)}{(\sigma_1^2-\sigma_0^2)}\right]^2 - \left[\frac{\sigma_0(\mu_0-\mu_1)}{(\sigma_1^2-\sigma_0^2)}\right]^2\right.}{2\sigma_1^2}$$
$$- \frac{(\mu_0-\mu_1)^2}{2\sigma_1^2}$$

$$= \ln\frac{\sigma_0}{\sigma_1} + \frac{(\sigma_1^2-\sigma_0^2)\left\{\left[Z - \frac{\sigma_0(\mu_0-\mu_1)}{(\sigma_1^2-\sigma_0^2)}\right]^2 - \left[\frac{\sigma_0(\mu_0-\mu_1)}{(\sigma_1^2-\sigma_0^2)}\right]^2\right\}}{2\sigma_1^2}$$
$$- \frac{(\mu_0-\mu_1)^2}{2\sigma_1^2}$$

$$= \ln\frac{\sigma_0}{\sigma_1} - \frac{(\mu_0-\mu_1)^2}{2\sigma_1^2} - \frac{\sigma_0^2(\mu_0-\mu_1)^2}{2\sigma_1^2(\sigma_1^2-\sigma_0^2)}$$
$$+ \frac{(\sigma_1^2-\sigma_0^2)}{2\sigma_1^2}\left[Z - \frac{\sigma_0(\mu_0-\mu_1)}{(\sigma_1^2-\sigma_0^2)}\right]^2$$

$$= \ln\frac{\sigma_0}{\sigma_1} - \frac{(\mu_1-\mu_0)^2}{2(\sigma_1^2-\sigma_0^2)} + \frac{(\sigma_1^2-\sigma_0^2)}{2\sigma_1^2}\left[Z + \frac{\sigma_0(\mu_1-\mu_0)}{(\sigma_1^2-\sigma_0^2)}\right]^2$$

$$= \ln\frac{\sigma_0}{\sigma_1} - \frac{\left(\frac{\mu_1-\mu_0}{\sigma_0}\right)^2}{\frac{2(\sigma_1^2-\sigma_0^2)}{\sigma_0^2}} + \frac{(\sigma_1^2-\sigma_0^2)}{2\sigma_1^2}\left[Z + \frac{\left(\frac{\mu_1-\mu_0}{\sigma_0}\right)}{\frac{(\sigma_1^2-\sigma_0^2)}{\sigma_0^2}}\right]^2$$

Let

$$K = \frac{\sigma_0}{\sigma_1}, \qquad \eta = \frac{\mu_1-\mu_0}{\sigma_0}$$

then

$$X^o =^D \ln K - \frac{\eta^2}{2\left(\frac{1}{K^2}-1\right)} + \frac{(1-K^2)}{2}\left[Z + \frac{\eta}{\left(\frac{1}{K^2}-1\right)}\right]^2$$

$$= \ln K - \frac{K^2\eta^2}{2(1-K^2)} + \frac{(1-K^2)}{2}\left[Z + \frac{K^2\eta}{(1-K^2)}\right]^2$$

The distribution of X^o coincides with the result in (4.8) if the dimension is decreased to 1. Therefore, we can directly derive the distribution for X^* from (4.20) by reducing the dimension to 1 as follows

$$X^* =^D - \ln K - \frac{K^{-2}\eta^{*2}}{2(1-K^{-2})} + \frac{(1-K^{-2})}{2}\left[Z - \frac{K^{-2}\eta^*}{(1-K^{-2})}\right]^2$$

where $\eta^* = \frac{\mu_1-\mu_0}{\sigma_1}$

It is obvious that the distribution of X^o and X^* are simpler in univariate case, because both involve only one term of noncentral chi-square random variable. However, the partial sums S_j^o and S_j^* for the random walk, which are defined as the sums of first j terms of X^o or X^* respectively, are still comprised of a constant term and j terms of linear combination of noncentral chi-square random variables. The computation for the distribution for the change-point mle will exactly follow Step S1 – S6 of the algorithmic procedure specified in Section 4.1.3 with the following parameterization.

$$K = \frac{\sigma_0}{\sigma_1}, \qquad \eta = \frac{\mu_1 - \mu_0}{\sigma_0}$$

$$C^o = \ln K - \frac{K^2\eta^2}{2(1-K^2)}, \sigma_i^{o2} = \left(\frac{K^2\eta}{1-K^2}\right)^2, a_i^o = \frac{1-K^2}{2}$$

$$\eta^* = \frac{\mu_1 - \mu_0}{\sigma_1}$$

$$C^* = -\ln K - \frac{K^{-2}\eta^{*2}}{2(1-K^{-2})}, \sigma_i^{*2} = \left(\frac{K^{-2}\eta^*}{1-K^{-2}}\right)^2, a_i^* = \frac{1-K^{-2}}{2}$$

4.2.2 Covariance Only of a Multivariate Gaussian Series

Another special case is the case when the change occurs only to covariance matrix. Let $Y_1, Y_2, \dots, Y_n$ be a sequence of time series valued independent random vectors such that $\mathrm{Y_i} \in \mathbb{R}^\mathrm{d}, \mathrm{i} = 1, \dots, \mathrm{n}$. The mean vector μ of the series keeps constant, and the covariance matrix changes from Σ_0 to Σ_1 at some unknown point τ_n such that $\Sigma_0 \neq \Sigma_1$. Both the mean vector and covariance matrices are unknown. The distribution of X^o and X^* for the two sided random walk can be adapted from Section

4.1.1 by letting both μ_0 and μ_1 equal μ. Under this setup, both η and η^* that were defined in (4.5) and (4.17) are zero. As ω and ω^* depend on η and η^*, respectively, both of them are also zero under the case of change in covariance only. We follow the distribution of X^o in (4.8) and set $\omega = 0$, we have

$$X^o =^D \ln(\psi_1\psi_2 \dots \psi_d)^{1/2} - \frac{1}{2}\sum_{s=1}^{d} \frac{\psi_s \omega_s^2}{1-\psi_s} + \sum_{s=1}^{d} \frac{1-\psi_s}{2}\left(z_s + \frac{\psi_s \omega_s}{1-\psi_s}\right)^2$$

$$= \ln(\psi_1\psi_2 \dots \psi_d)^{1/2} + \sum_{s=1}^{d} \frac{1-\psi_s}{2} {z_s}^2$$

We follow the distribution of X^* in (4.20) and set $\omega^* = 0$, we have

$$X^* =^D \ln(\psi_1\psi_2 \dots \psi_d)^{-1/2} - \frac{1}{2}\sum_{s=1}^{d} \frac{\psi_s^{-1} \omega_s^{*2}}{1-\psi_s^{-1}} + \sum_{s=1}^{d} \frac{1-\psi_s^{-1}}{2}\left(z_s - \frac{\psi_s^{-1} \omega_s^*}{1-\psi_s^{-1}}\right)^2$$

$$= \ln(\psi_1\psi_2 \dots \psi_d)^{-1/2} + \sum_{s=1}^{d} \frac{1-\psi_s^{-1}}{2} {z_s}^2$$

The distributions of X^o and X^* are comprised of the linear combination of d terms of central chi-square random variables and a constant term. The partial sums S_j^o and S_j^* for the random walk follow the same way. The Imhof's (1961) formula also applies to the linear combination of central chi-square distribution. Thus the algorithmic procedure in Section 4.1.3 can be followed using the following parameterization:

$K = \Sigma_0^{1/2}\Sigma_1^{-1/2}$. $KK^T = \Theta\Psi\Theta^T$ where $\Psi = diag\{\psi_1, \psi_2, \dots, \psi_d\}$, $\Theta^{-1} = \Theta^{\mathrm{T}}$

$$C^o = \frac{1}{2} ln(\psi_1\psi_2 \dots \psi_d), \sigma_s^{o2} = 0, a_s^o = \frac{1}{2}(1 - \psi_s)$$

$$C^* = -\frac{1}{2}\ln(\psi_1\psi_2 \dots \psi_d), \sigma_s^{*2} = 0, a_s^* = \frac{1}{2}(1 - \psi_s^{-1})$$

4.3 Bayesian Method for Estimating Change-point in Mean and/or Covariance of a Multivariate Gaussian Series

Ever since Markov Chain Monte Carlo (MCMC) methods were seen as a tool for overcoming the computational complexities in Bayesian analysis, there has been rapid progress in the overall development of this important methodological tool, and advances in Bayesian change-point analysis have not lagged behind. The main advantage of the Bayesian approach to the change-problem is that both detection and estimation parts of the problem are solved simultaneously once posterior distribution of the unknown change-point is made available, mainly because all inferences about the unknown change-point are made from the posterior distribution. Consequently, with recent advances in the methodology, the Bayesian approach to change-point analysis is able to provide inferential methods ranging from simple to complex situations, some of which include change in mean and/or variance of the univariate normal distribution (Perreault et al 1999, Perreault et al 2000a, 2000b), Change in the mean vector of a multivariate normal distribution (Perreault et al 2000), single change in the parameters of a multiple linear regression model (Seidou et al 2007), and also the more complex case of estimating multiple change-points (Fearnhead, 2005, 2006; Seidou and Ouarda 2007). In this chapter, the Bayesian change-point analysis using two types of prior information will be studied. One used the conjugate priors, multivariate normal distribution for mean vectors and Wishart distribution for covariance matrices as prior distributions. The other one uses Jeffery's non-informative prior. The results from Bayesian change-point analysis will be implemented and compared with mle method.

4.3.1 Conjugate Prior

Perreault et al. (2000) performed Bayesian analysis for multivariate change-point problem where there assumed that there was change in mean vector only. The prior distribution of the mean vector and covariance matrix followed Multivariate Normal and Wishart distribution respectively, and the prior distribution for the change-point was assumed to be uniform over all possible candidates. For the multivariate observations with change in both mean and covariance, the same assumptions can be made. The posterior distribution can be derived as follows.

Let $Y_1, Y_2, \ldots, Y_n$ be a sequence of time series valued independent random vectors such that $\mathrm{Y_i} \in \mathbb{R}^{\mathrm{d}}, \mathrm{i} = 1, \ldots, \mathrm{n}$. Furthermore, for each $\mathrm{i} = 1, 2, \ldots, \mathrm{n}$, let Y_i follow the multivariate Gaussian distribution with mean vector μ and variance-covariance matrix Σ. Assume at time τ_n, the mean vector and covariance matrix of the observations change from μ_0 to μ_1 and Σ_0 to Σ_1, respectively. Then we have

$$Y_i \sim f(x) \sim \begin{cases} N(\mu_0, \Sigma_0), & i = 1, 2, \ldots, \tau_n \\ N(\mu_1, \Sigma_1), & i = \tau_n + 1, \ldots, n' \end{cases} \quad (4.31)$$

Assume $\mathrm{P_0} = \Sigma_0^{-1}$ and $\mathrm{P_1} = \Sigma_1^{-1}$, then the prior distributions of parameters are

$$\mu_0 \sim N(\Phi_0, (\lambda_0 P_0)^{-1}), \mu_1 \sim N(\Phi_1, (\lambda_1 P_1)^{-1}) \quad (4.32)$$

$$P_0 \sim \mathrm{Wishart}(a_0, B_0), P_1 \sim \mathrm{Wishart}(a_1, B_1)$$

$$\tau_n \sim \mathrm{Unif}(1, n-1)$$

The probability density function of Wishart distribution is

$$f_w(P|a,B) = \frac{|\mathrm{B}|^{\frac{a}{2}}|\mathrm{P}|^{(a-d-1)/2}\exp[-\frac{1}{2}tr(BP)]}{2^{\frac{ad}{2}}\Gamma_{\mathrm{d}}(\frac{\mathrm{a}}{2})} \tag{4.33}$$

where a is the degree of freedom of the Wishart distribution satisfying $\mathrm{a} > d-1$, and Γ_{d} is the multivariate gamma function defined as $\Gamma_{\mathrm{d}}\left(\frac{\mathrm{a}}{2}\right) = \pi^{\frac{\mathrm{d}(\mathrm{d}-1)}{4}}\prod_{\mathrm{j}=1}^{\mathrm{d}}\Gamma(\frac{\mathrm{a}}{2}+\frac{1-\mathrm{j}}{2})$. The mean of the Wishart random variable is aB^{-1}.

The prior distribution function of the change-point is $P(\tau_n) = \frac{1}{n-1}$.

Assuming all the parameters, μ_0, μ_1, P_0, P_1 and τ_n are independent from each other, the joint distribution of the parameters can be computed as

$$P(\mu_0, \mu_1, P_0, P_1) = P(\mu_0|P_0)P(P_0)P(\mu_1|P_1)P(P_1)$$

$$=^D f_N(\mu_0|\Phi_0, (\lambda_0 P_0)^{-1})f_W(P_0|a_0, B_0)$$

$$f_N(\mu_1|\Phi_1, (\lambda_1 P_1)^{-1})f_W(P_1|a_1, B_1)$$

$$\sim NWNW(\mu_0, \mu_1, \Sigma_0, \Sigma_1;\ \Phi_0, \lambda_0, a_0, B_0, \Phi_1, \lambda_1, a_1, B_1)$$

As τ_n is independent of μ_0, μ_1, P_0, P_1,

$$P(\mu_0, \mu_1, P_0, P_1, \tau_n)\sim$$

$$NWNW(\mu_0, \mu_1, \Sigma_0, \Sigma_1;\ \Phi_0, \lambda_0, a_0, B_0, \Phi_1, \lambda_1, a_1, B_1)P(\tau_n)$$

So

$$P(\mu_0, \mu_1, P_0, P_1|\tau_n)\sim$$

$$NWNW(\mu_0, \mu_1, P_0, P_1|\Phi_0, \lambda_0, a_0, B_0, \Phi_1, \lambda_1, a_1, B_1)$$

Let τ_n be fixed,

$$P(\mu_0, \mu_1, P_0, P_1|\tau_n, Y_1, Y_2, \dots, Y_n) \tag{4.34}$$

$$= \frac{P(Y_1, Y_2, \dots, Y_n | \mu_0, \mu_1, P_0, P_1, \tau_n) P(\mu_0, \mu_1, P_0, P_1 | \tau_n)}{P(Y_1, Y_2, \dots, Y_n | \tau_n)}$$

$$\propto P(Y_1, Y_2, \dots, Y_n | \mu_0, \mu_1, P_0, P_1, \tau_n) P(\mu_0, \mu_1, P_0, P_1 | \tau_n)$$

$$\propto P(Y_1, Y_2, \dots, Y_n | \mu_0, \mu_1, P_0, P_1, \tau_n)$$

$$NWNW(\mu_0, \mu_1, P_0, P_1 | \Phi_0, \lambda_0, a_0, B_0, \Phi_1, \lambda_1, a_1, B_1)$$

In the following steps, the two terms $P(Y_1, Y_2, \dots, Y_n | \mu_0, \mu_1, P_0, P_1, \tau_n)$ and $NWNW(\mu_0, \mu_1, P_0, P_1 | \Phi_0, \lambda_0, a_0, B_0, \Phi_1, \lambda_1, a_1, B_1)$ are evaluated respectively to derive the joint posterior distribution of (μ_0, μ_1, P_0, P_1).

The likelihood function of the observations $Y_1, Y_2, \dots, Y_n$ is

$$P(Y_1, Y_2, \dots, Y_n | \mu_0, \mu_1, P_0, P_1, \tau_n)$$

$$= \prod_{i=1}^{\tau_n} N(Y_i | \mu_0, P_0) \prod_{i=\tau_n+1}^{n} N(Y_i | \mu_1, P_1)$$

$$= (2\pi)^{-nd/2} |P_0|^{\tau_n/2} \exp\left\{-\frac{1}{2} \sum_{i=1}^{\tau} (Y_i - \mu_0)^T P_0 (Y_i - \mu_0)\right\} |P_1|^{(n-\tau_n)/2} \exp\left\{-\frac{1}{2} \sum_{i=\tau+1}^{n} (Y_i - \mu_1)^T P_1 (Y_i - \mu_1)\right\}$$

$$= (2\pi)^{-nd/2} |P_0|^{\tau_n/2} |P_1|^{(n-\tau_n)/2} \exp\left\{-\frac{1}{2} tr\left(P_0 \sum_{i=1}^{\tau} (Y_i - \mu_0)(Y_i - \mu_0)^T\right)\right\} \exp\left\{-\frac{1}{2} tr\left(P_1 \sum_{i=1}^{\tau} (Y_i - \mu_1)(Y_i - \mu_1)^T\right)\right\}$$

Let

$$\bar{Y}_{\tau_n} = \frac{1}{\tau_n}\sum_{i=1}^{\tau_n} Y_i\,, \qquad \bar{Y}_{n-\tau_n} = \frac{1}{n-\tau_n}\sum_{i=\tau_n+1}^{n} Y_i \tag{4.35}$$

$$S_{\tau_n} = \frac{1}{\tau_n}\sum_{i=1}^{\tau_n}\left(Y_i - \bar{Y}_{\tau_n}\right)\left(Y_i - \bar{Y}_{\tau_n}\right)^T, \tag{4.36}$$

$$S_{n-\tau_n} = \frac{1}{n-\tau_n}\sum_{i=\tau_n+1}^{n}\left(Y_i - \bar{Y}_{n-\tau_n}\right)\left(Y_i - \bar{Y}_{n-\tau_n}\right)^T$$

The likelihood function can be written as

$$P(Y_1, Y_2, \dots, Y_n | \mu_0, \mu_1, P_0, P_1, \tau_n) \tag{4.37}$$

$$= (2\pi)^{-nd/2}|P_0|^{\tau_n/2}|P_1|^{(n-\tau_n)/2} \exp\left\{-\frac{1}{2} tr\left[P_0\left(\tau_n S_{\tau_n} + \tau_n\left(\bar{Y}_{\tau_n} - \mu_0\right)\left(\bar{Y}_{\tau_n} - \mu_0\right)^T\right)\right]\right\} \exp\left\{-\frac{1}{2} tr\left[P_1\left((n - \tau_n)S_{n-\tau_n} + (n-\tau_n)\left(\bar{Y}_{n-\tau_n} - \mu_1\right)\left(\bar{Y}_{n-\tau_n} - \mu_1\right)^T\right)\right]\right\}$$

$$= (2\pi)^{-nd/2}|P_0|^{\tau_n/2}|P_1|^{(n-\tau_n)/2} \exp\left\{-\frac{1}{2} tr\left[P_0\left(\tau_n S_{\tau_n} + \tau_n\left(\bar{Y}_{\tau_n} - \mu_0\right)\left(\bar{Y}_{\tau_n} - \mu_0\right)^T\right)\right]\right\} \exp\left\{-\frac{1}{2} tr\left[P_1\left((n - \tau_n)S_{n-\tau_n} + (n-\tau_n)\left(\bar{Y}_{n-\tau_n} - \mu_1\right)\left(\bar{Y}_{n-\tau_n} - \mu_1\right)^T\right)\right]\right\}$$

$$= (2\pi)^{-nd/2}|P_0|^{\tau_n/2}|P_1|^{(n-\tau_n)/2} \exp\left\{-\frac{\tau_n}{2} tr\left[P_0\left(S_{\tau_n} + \left(\bar{Y}_{\tau_n} - \mu_0\right)\left(\bar{Y}_{\tau_n} - \mu_0\right)^T\right)\right]\right\} \exp\left\{-\frac{n-\tau_n}{2} tr\left[P_1\left(S_{n-\tau_n} + \left(\bar{Y}_{n-\tau_n} - \mu_1\right)\left(\bar{Y}_{n-\tau_n} - \mu_1\right)^T\right)\right]\right\}$$

Now we begin to derive for $NWNW(\mu_0, \mu_1, P_0, P_1 | \Phi_0, \lambda_0, a_0, B_0, \Phi_1, \lambda_1, a_1, B_1)$.

$$NWNW(\mu_0, \mu_1, P_0, P_1 | \Phi_0, \lambda_0, a_0, B_0, \Phi_1, \lambda_1, a_1, B_1) \tag{4.38}$$

$$= f_N(\mu_0 | \Phi_0, (\lambda_0 P_0)^{-1}) f_W(P_0 | a_0, B_0)$$

$$f_N(\mu_1 | \Phi_1, (\lambda_1 P_1)^{-1}) f_W(P_1 | a_1, B_1)$$

$$= (2\pi)^{-d/2} |\lambda_0 P_0|^{1/2} \exp\left\{-\frac{1}{2}(\mu_0 - \Phi_0)^T \lambda_0 P_0 (\mu_0 - \Phi_0)\right\}$$

$$\cdot (2\pi)^{-d/2} |\lambda_1 P_1|^{1/2} \exp\left\{-\frac{1}{2}(\mu_1 - \Phi_1)^T \lambda_1 P_1 (\mu_1 - \Phi_1)\right\}$$

$$\cdot \frac{|B_0|^{\frac{a_0}{2}} |P_0|^{\frac{a_0 - d - 1}{2}}}{2^{\frac{a_0 d}{2}} \Gamma_d(\frac{a_0}{2})} \exp\left\{-\frac{1}{2} tr(B_0 P_0)\right\}$$

$$\cdot \frac{|B_1|^{\frac{a_1}{2}} |P_1|^{\frac{a_1 - d - 1}{2}}}{2^{\frac{a_1 d}{2}} \Gamma_d(\frac{a_1}{2})} \exp\left\{-\frac{1}{2} tr(B_1 P_1)\right\}$$

$$= (2\pi)^{-d} |\lambda_0 \lambda_1|^{\frac{d}{2}} |P_0|^{\frac{a_0 - d}{2}} \, |P_1|^{\frac{a_1 - d}{2}} \frac{|B_0|^{\frac{a_0}{2}}}{2^{\frac{a_0 d}{2}} \Gamma_d\left(\frac{a_0}{2}\right)} \frac{|B_1|^{\frac{a_1}{2}}}{2^{\frac{a_1 d}{2}} \Gamma_d\left(\frac{a_1}{2}\right)}$$

$$\cdot \exp\left\{-tr\frac{1}{2}[\lambda_0 P_0 (\mu_0 - \Phi_0)(\mu_0 - \Phi_0)^T + P_0 B_0]\right.$$

$$\left. - tr\frac{1}{2}[\lambda_1 P_1 (\mu_1 - \Phi_1)(\mu_1 - \Phi_1)^T + P_1 B_1]\right\}$$

Combine the results in (4.37) and (4.38) , the posterior distribution of (μ_0, μ_1, P_0, P_1) in (4.34) can be computed as

$$P(\mu_0, \mu_1, P_0, P_1 | \tau_n, Y_1, Y_2, \dots, Y_n) \propto \tag{4.39}$$

$$P(Y_1, Y_2, \dots, Y_n | \mu_0, \mu_1, P_0, P_1, \tau_n)$$

$$NWNW(\mu_0, \mu_1, P_0, P_1 | \Phi_0, \lambda_0, a_0, B_0, \Phi_1, \lambda_1, a_1, B_1)$$

$$= (2\pi)^{-\frac{nd}{2}}|P_0|^{\frac{\tau_n}{2}}|P_1|^{\frac{n-\tau_n}{2}} \exp\left\{-\frac{1}{2} tr \left[P_0 \left(\tau_n S_{\tau_n} + \tau_n\left(\bar{Y}_{\tau_n} - \mu_0\right)\left(\bar{Y}_{\tau_n} - \mu_0\right)^T\right)\right]\right\} \exp\left\{-\frac{1}{2} tr \left[P_1 \left((n - \tau_n)S_{n-\tau_n} + (n-\tau_n)\left(\bar{Y}_{n-\tau_n} - \mu_1\right)\left(\bar{Y}_{n-\tau_n} - \mu_1\right)^T\right)\right]\right\}$$

$$\cdot (2\pi)^{-d}|\lambda_0\lambda_1|^{\frac{d}{2}}|P_0|^{\frac{a_0-d}{2}} |P_1|^{\frac{a_1-d}{2}} \frac{|B_0|^{\frac{a_0}{2}}}{2^{\frac{a_0 d}{2}}\Gamma_d\left(\frac{a_0}{2}\right)} \frac{|B_1|^{\frac{a_1}{2}}}{2^{\frac{a_1 d}{2}}\Gamma_d\left(\frac{a_1}{2}\right)}$$

$$\cdot \exp\left\{-tr\frac{1}{2}[\lambda_0 P_0(\mu_0 - \Phi_0)(\mu_0 - \Phi_0)^T + P_0 B_0] - tr\frac{1}{2}[\lambda_1 P_1(\mu_1 - \Phi_1)(\mu_1 - \Phi_1)^T + P_1 B_1]\right\}$$

The exponent term in (4.39) involving the parameters before the change-point is

$$\tau_n P_0 \left(S_{\tau_n} + \left(\bar{Y}_{\tau_n} - \mu_0\right)\left(\bar{Y}_{\tau_n} - \mu_0\right)^T\right) + \lambda_0 P_0(\mu_0 - \Phi_0)(\mu_0 - \Phi_0)^T + P_0 B_0$$

$$= P_0 \left[\tau_n \left(S_{\tau_n} + \left(\mu_0 - \bar{Y}_{\tau_n}\right)\left(\mu_0 - \bar{Y}_{\tau_n}\right)^T\right) + \lambda_0(\mu_0 - \Phi_0)(\mu_0 - \Phi_0)^T + B_0\right]$$

$$= P_0 \left[(\tau_n + \lambda_0)\mu_0\mu_0^T - \mu_0\left(\lambda_0\Phi_0 + \tau_n\bar{Y}_{\tau_n}\right)^T - \left(\lambda_0\Phi_0 + \tau_n\bar{Y}_{\tau_n}\right)\mu_0^T + \lambda_0\Phi_0\Phi_0^T + \tau_n\bar{Y}_{\tau_n}\bar{Y}_{\tau_n}^T + \tau_n S_{\tau_n} + B_0\right]$$

$$= (\tau_n + \lambda_0)P_0 \Bigg[\mu_0\mu_0^T - \mu_0 \frac{(\lambda_0\Phi_0 + \tau_n\bar{Y}_\tau)^T}{\tau_n + \lambda_0} - \frac{\left(\lambda_0\Phi_0 + \tau_n\bar{Y}_{\tau_n}\right)}{\tau_n + \lambda_0}\mu_0^T$$

$$+ \left(\frac{\lambda_0\Phi_0 + \tau_n\bar{Y}_{\tau_n}}{\tau_n + \lambda_0}\right)\left(\frac{\lambda_0\Phi_0 + \tau_n\bar{Y}_{\tau_n}}{\tau_n + \lambda_0}\right)^T$$

$$- \left(\frac{\lambda_0\Phi_0 + \tau_n\bar{Y}_{\tau_n}}{\tau_n + \lambda_0}\right)\left(\frac{\lambda_0\Phi_0 + \tau_n\bar{Y}_{\tau_n}}{\tau_n + \lambda_0}\right)^T$$

$$+ \frac{\lambda_0\Phi_0\Phi_0^T + \tau_n\bar{Y}_{\tau_n}\bar{Y}_{\tau_n}^T + \tau_n S_{\tau_n} + B_0}{\tau_n + \lambda_0}\Bigg]$$

$$= (\tau_n + \lambda_0)P_0 \Bigg[\left(\mu_0 - \frac{\lambda_0\Phi_0 + \tau_n\bar{Y}_{\tau_n}}{\tau_n + \lambda_0}\right)\left(\mu_0 - \frac{\lambda_0\Phi_0 + \tau_n\bar{Y}_{\tau_n}}{\tau_n + \lambda_0}\right)^T$$

$$- \left(\frac{\lambda_0\Phi_0 + \tau_n\bar{Y}_{\tau_n}}{\tau_n + \lambda_0}\right)\left(\frac{\lambda_0\Phi_0 + \tau_n\bar{Y}_{\tau_n}}{\tau_n + \lambda_0}\right)^T$$

$$+ \frac{\lambda_0\Phi_0\Phi_0^T + \tau_n\bar{Y}_{\tau_n}\bar{Y}_{\tau_n}^T + \tau_n S_{\tau_n} + B_0}{\tau_n + \lambda_0}\Bigg]$$

$$= (\tau_n + \lambda_0)P_0 \Bigg[\left(\mu_0 - \frac{\lambda_0\Phi_0 + \tau_n\bar{Y}_{\tau_n}}{\tau_n + \lambda_0}\right)\left(\mu_0 - \frac{\lambda_0\Phi_0 + \tau_n\bar{Y}_{\tau_n}}{\tau_n + \lambda_0}\right)^T$$

$$- \frac{\lambda_0^2\Phi_0\Phi_0^T + \tau_n\lambda_0\bar{Y}_{\tau_n}\Phi_0^T + \tau_n\lambda_0\Phi_0\bar{Y}_{\tau_n}^T + \tau^2\bar{Y}_{\tau_n}\bar{Y}_{\tau_n}^T}{(\tau_n + \lambda_0)^2}$$

$$+ \frac{\lambda_0(\tau_n + \lambda_0)\Phi_0\Phi_0^T + \tau_n(\tau_n + \lambda_0)\bar{Y}_\tau\bar{Y}_\tau^T + (\tau_n + \lambda_0)(\tau_n S_\tau + B_0)}{(\tau_n + \lambda_0)^2}\Bigg]$$

$$= (\tau_n + \lambda_0)P_0\left[\left(\mu_0 - \frac{\lambda_0\Phi_0 + \tau_n\bar{Y}_{\tau_n}}{\tau_n + \lambda_0}\right)\left(\mu_0 - \frac{\lambda_0\Phi_0 + \tau_n\bar{Y}_{\tau_n}}{\tau_n + \lambda_0}\right)^T\right.$$

$$\left. + \frac{\tau_n\lambda_0\Phi_0\Phi_0^T - \tau_n\lambda_0\bar{Y}_{\tau_n}\Phi_0^T - \tau_n\lambda_0\Phi_0\bar{Y}_{\tau_n}^T + \tau_n\lambda_0\bar{Y}_{\tau_n}\bar{Y}_{\tau_n}^T + (\tau_n + \lambda_0)(\tau_n S_{\tau_n} + B_0)}{(\tau_n + \lambda_0)^2}\right]$$

$$= (\tau_n + \lambda_0)P_0\left[\left(\mu_0 - \frac{\lambda_0\Phi_0 + \tau_n\bar{Y}_{\tau_n}}{\tau_n + \lambda_0}\right)\left(\mu_0 - \frac{\lambda_0\Phi_0 + \tau_n\bar{Y}_{\tau_n}}{\tau_n + \lambda_0}\right)^T\right.$$

$$\left. + \frac{\tau_n\lambda_0(\bar{Y}_{\tau_n} - \Phi_0)(\bar{Y}_{\tau_n} - \Phi_0)^T}{(\tau_n + \lambda_0)^2} + \frac{\tau_n S_{\tau_n} + B_0}{\tau_n + \lambda_0}\right]$$

$$= P_0\left[(\tau_n + \lambda_0)\left(\mu_0 - \frac{\lambda_0\Phi_0 + \tau_n\bar{Y}_{\tau_n}}{\tau_n + \lambda_0}\right)\left(\mu_0 - \frac{\lambda_0\Phi_0 + \tau_n\bar{Y}_{\tau_n}}{\tau_n + \lambda_0}\right)^T\right.$$

$$\left. + \frac{\tau_n\lambda_0(\bar{Y}_{\tau_n} - \Phi_0)(\bar{Y}_{\tau_n} - \Phi_0)^T}{\tau_n + \lambda_0} + \tau_n S_{\tau_n} + B_0\right]$$

Let

$$\lambda_0^{'} = \lambda_0 + \tau_\mathrm{n}, \Phi_0^{'} = \frac{\lambda_0\Phi_0 + \tau_n\bar{Y}_\tau}{\lambda_0 + \tau_n}, \tag{4.40}$$

$$B_0^{'} = \frac{\tau_n\lambda_0(\bar{Y}_{\tau_n} - \Phi_0)(\bar{Y}_{\tau_n} - \Phi_0)^T}{(\tau_n + \lambda_0)} + \tau_n S_{\tau_n} + B_0$$

The above expression can be simplified as

$$P_0[\lambda_0^{'}(\mu_0 - \Phi_0^{'})(\mu_0 - \Phi_0^{'})^T + B_0^{'}] \tag{4.41}$$

Similarly for the exponent terms in (4.39) involving the parameters after the change-point, similar parameterization can be assumed as

$$\lambda_1^{'} = \lambda_1 + n - \tau_n, \Phi_1^{'} = \frac{\lambda_1 \Phi_1 + (n - \tau_n)\bar{Y}_{n-\tau_n}}{\lambda_1 + n - \tau_n}, \tag{4.42}$$

$$B_1^{'} = \frac{(n - \tau_n)\lambda_1 \left(\bar{Y}_{n-\tau_n} - \Phi_1\right)\left(\bar{Y}_{n-\tau_n} - \Phi_1\right)^T}{2(n - \tau_n + \lambda_1)} + \frac{(n - \tau_n)S_{n-\tau_n}}{2}$$

$$+ B_1$$

Also let $a_0^{'} = a_0 + \tau_n$ and $a_1^{'} = a_1 + n - \tau_n$

Then the posterior distribution of (μ_0, μ_1, P_0, P_1) can be reduced to

$$P(\mu_0, \mu_1, P_0, P_1 | \tau, Y_1, Y_2, \dots, Y_n) \tag{4.43}$$

$$\propto |P_0|^{\frac{a_0^{'} - d}{2}} |P_1|^{\frac{a_1^{'} - d}{2}} \exp\left\{-tr\frac{1}{2}[\lambda_0^{'} P_0 (\mu_0 - \Phi_0^{'})(\mu_0 - \Phi_0^{'})^T + P_0 B_0^{'}]\right.$$

$$\left. - tr\frac{1}{2}[\lambda_1^{'} P_1 (\mu_1 - \Phi_1^{'})(\mu_1 - \Phi_1^{'})^T + P_1 B_1^{'}]\right\}$$

$$\propto NWNW(\mu_0, \mu_1, P_0, P_1 | \Phi_0^{'}, \lambda_0^{'}, a_0^{'}, B_0^{'}, \Phi_1^{'}, \lambda_1^{'}, a_1^{'}, B_1^{'})$$

That is to say, the prior and posterior distributions of the joint distribution of (μ_0, μ_1, P_0, P_1) are both NWNW distributions, thus the distribution for the normalizing constant can be calculated as

$$P(Y_1, Y_2, \dots, Y_n | \tau_n)$$

$$= \frac{P(Y_1, Y_2, \dots, Y_n | \mu_0, \mu_1, P_0, P_1, \tau_n) P(\mu_0, \mu_1, P_0, P_1 | \tau_n)}{P(\mu_0, \mu_1, P_0, P_1 | Y_1, Y_2, \dots, Y_n, \tau_n)}$$

$$= \frac{P(Y_1, Y_2, \dots, Y_n | \mu_0, \mu_1, P_0, P_1, \tau_n) \mathrm{NWNW}(\mu_0, \mu_1, P_0, P_1 | \Phi_0, \lambda_0, a_0, B_0, \Phi_1, \lambda_1, a_1, B_1)}{\mathrm{NWNW}(\mu_0, \mu_1, P_0, P_1 | \Phi_0^{'}, \lambda_0^{'}, a_0^{'}, B_0^{'}, \Phi_1^{'}, \lambda_1^{'}, a_1^{'}, B_1^{'})}$$

The distribution function can be simplified by keeping the factors that do not contain (μ_0, μ_1, P_0, P_1).

$$P(Y_1, Y_2, \dots, Y_n|\tau_n) \propto \frac{|\lambda_0\lambda_1|^{\frac{d}{2}} \frac{|B_0|^{a_0}}{2^{\frac{a_0 d}{2}}\Gamma_d\left(\frac{a_0}{2}\right)} \frac{|B_1|^{a_1}}{2^{\frac{a_1 d}{2}}\Gamma_d\left(\frac{a_1}{2}\right)}}{|\lambda_0^{'}\lambda_1^{'}|^{\frac{d}{2}} \frac{|B_0^{'}|^{a_0^{'}}}{2^{\frac{a_0^{'} d}{2}}\Gamma_d\left(\frac{a_0^{'}}{2}\right)} \frac{|B_1^{'}|^{a_1^{'}}}{2^{\frac{a_1^{'} d}{2}}\Gamma_d\left(\frac{a_1^{'}}{2}\right)}} \tag{4.44}$$

$$\propto \left|\frac{\lambda_0\lambda_1}{\lambda_0^{'}\lambda_1^{'}}\right|^{\frac{d}{2}} \frac{|B_0|^{a_0}|B_1|^{a_1}}{|B_0^{'}|^{a_0^{'}}|B_1^{'}|^{a_1^{'}}} \prod_{k=1}^{d} \frac{\Gamma\left[\frac{a_0^{'}+1-k}{2}\right]\Gamma\left[\frac{a_1^{'}+1-k}{2}\right]}{\Gamma\left[\frac{a_0+1-k}{2}\right]\Gamma\left[\frac{a_1+1-k}{2}\right]}$$

$$\propto |\lambda_0^{'}\lambda_1^{'}|^{-\frac{d}{2}}|B_0^{'}|^{-a_0^{'}}|B_1^{'}|^{-a_1^{'}} \prod_{k=1}^{d} \Gamma\left[\frac{a_0^{'}+1-k}{2}\right]\Gamma\left[\frac{a_1^{'}+1-k}{2}\right]$$

By Bayes' Theorem, the posterior distribution of the change point τ is

$$P(\tau_n|Y_1, Y_2, \dots, Y_n) = \frac{P(Y_1, Y_2, \dots, Y_n|\tau_n)P(\tau_n)}{P(Y_1, Y_2, \dots, Y_n)} \tag{4.45}$$

As we assumed that τ_n is uniformly distributed, and the marginal distribution of $P(Y_1, Y_2, \dots, Y_n)$ is constant, so the posterior distribution for the change-point τ_n can be computed following

$$P(\tau_n|Y_1, Y_2, \dots, Y_n) \tag{4.46}$$

$$\propto P(Y_1, Y_2, \dots, Y_n|\tau_n)$$

$$\propto |\lambda_0^{'}\lambda_1^{'}|^{-\frac{d}{2}}|B_0^{'}|^{-a_0^{'}}|B_1^{'}|^{-a_1^{'}} \prod_{k=1}^{d} \Gamma\left[\frac{a_0^{'}+1-k}{2}\right]\Gamma\left[\frac{a_1^{'}+1-k}{2}\right]$$

for $\tau_n = 1, 2, 3, \dots, n-1.$

4.3.2 Non-informative Prior

The Bayesian method discussed in section 4.3.1 used conjugate priors for the change-point models, that is to say, multivariate normal for mean vectors, and Wishart for covariance matrix. Although Perrault et al (2000) compared the results using different parameters for the prior distributions; there is no rule about how to determine the numbers. The selection of parameters depends heavily on experience and is subjective. Son and Kim (2005) proposed to use non-informative priors for parameters for change-point analysis. In the analysis, the Jeffreys prior, which assumes that the prior distribution is proportional to the square root of the determinant of the Fisher information matrix, is applied for the joint distribution of $(\mu_0, \mu_1, \Sigma_0, \Sigma_1)$.

$$\pi(\mu_0, \mu_1, \Sigma_0, \Sigma_1) = c\{|\Sigma_0||\Sigma_1|\}^{-(d+1)/2} \tag{4.47}$$

where c is the normalizing constant whose value is unknown.

The prior distribution for the change-point τ_n is still set as uniform on $\{1, 2, \dots, n-1\}$.

Son and Kim (2005) derived the marginal density of the full sample assuming change occurs for both mean and covariance as

$$f(\tau_n, Y_1, Y_2, \dots, Y_n) \tag{4.48}$$

$$= c \cdot \left\{ \frac{\prod_{i=1}^{d} \Gamma\left(\frac{\tau_n - i}{2}\right) \Gamma\left(\frac{n - \tau_n - i}{2}\right)}{\tau_n^{d/2} (n - \tau_n)^{d/2} \pi^{d(n-d-1)/2} \left|S_{\tau_n}\right|^{(\tau_n - 1)/2} \left|S_{n-\tau_n}\right|^{(n-\tau_n-1)/2}} \right\}$$

where the definition for S_{τ_n} and $S_{n-\tau_n}$ follows (4.36) .

The posterior distribution of the change-point can be computed by Bayes theorem as

$$f(\tau_n|\, Y_1, Y_2, \dots, Y_n) = \frac{f(\tau_n, Y_1, Y_2, \dots, Y_n)}{\sum_{j=1}^{n-1} f(j, Y_1, Y_2, \dots, Y_n)} \tag{4.49}$$

4.4 Conditional MLE Method for Estimating Change-point in Mean and/or Covariance of a Multivariate Gaussian Series

MLE and Bayesian methods represent two major methods that derive the distribution for the change-point estimate. Although the frequency's view and Bayesian view are quite different, they are not completely contrary to each other. Cobb (1978) proposed the third approach from a conditional frequentist's view. His motivation came from the fact that the mle is not a sufficient statistics, so he derived conditional distribution of the change-point mle by conditioning upon sufficient observations around the true change-point. According to Cobb (1978), this is equivalent to the Bayesian posterior with uniform prior for the unknown change-point.

Suppose $Y_1, Y_2, \dots, Y_n$ are observations with detected change in their parameters at $\hat{\tau}_n$ using maximum likelihood method. Cobb's (1978) conditional solution only considered D observations on either side of $\hat{\tau}_n$. Cobb's (1978) conditional probabilities for the change-point conditional on the observations $\{Y_{\hat{\tau}_n - D}, \dots, Y_{\hat{\tau}_n + D}\}$ was computed by

$$\Pr\left(\tau_n - \hat{\tau}_n = d \mid Y_{\hat{\tau}_n - D}, \dots, Y_{\hat{\tau}_n + D,}\right) \tag{4.50}$$

$$\cong \frac{\Pr(Y_1, Y_2, \dots, Y_n; \tau = \hat{\tau}_n + d)}{\sum_{j=-D}^{D} \Pr(Y_1, Y_2, \dots, Y_n; \tau = \hat{\tau}_n + j)}$$

Cobb's (1978) solution conditioned on the event that $|\tau_n - \hat{\tau}_n| \leq D$, thus the choice for D need guarantee that the event occurs with arbitrarily high probability.

In maximum likelihood method, Cobb (1978) pointed out that the maximum likelihood estimate $\hat{\tau}_n$ does not provide the shape of the likelihood function in proximity of $\hat{\tau}_n$, thus the observations that close to $\hat{\tau}_n$, $Y_{\hat{\tau}_n-D}, \dots, Y_{\hat{\tau}_n+D}$, are used as ancillary information. To better determine the log likelihood functions around $\hat{\tau}_n$, they were translated so that the log likelihood function at $\hat{\tau}_n$ becomes the origin of the log likelihood functions, which leads to the transformed likelihood function as

$$L^*_{\hat{\tau}_n}(0) = 0 \tag{4.51}$$

$$L^*_{\hat{\tau}_n}(d) = \begin{cases} \sum_{j=1}^{d} \log \dfrac{f_0(\hat{\tau}_n + d)}{f_1(\hat{\tau}_n + d)}, & d > 0 \\ -\sum_{j=d+1}^{0} \log \dfrac{f_0(\hat{\tau}_n + d)}{f_1(\hat{\tau}_n + d)}, & d \le 0 \end{cases}$$

The fact that $\hat{\tau}_n$ is the estimate requires that the log likelihood function before transformation is larger than others, which means $L^*_{\hat{\tau}_n}(d) < 0$. Denote $\alpha(Y, \hat{\tau}_n, D)$ and $\beta(Y, \hat{\tau}_n, D)$ be the probabilities that $L^*_{\hat{\tau}_n}(d)$ would increase to positive from right- and left-hand walk, then the probability that the change-point would fall into the range $\{Y_{\hat{\tau}_n-D}, \dots, Y_{\hat{\tau}_n+D}\}$ is $1 - \epsilon(Y, D) = (1 - \alpha(Y, \hat{\tau}_n, D))(1 - \beta(Y, \hat{\tau}_n, D))$. Cobb (1978) proved that $\alpha(Y, \hat{\tau}_n, D) \le \prod_{d=1}^{D}\{f_1(Y_{\hat{\tau}_n+d})/f_2(Y_{\hat{\tau}_n+d})\}$, and $\beta(Y, \hat{\tau}_n, D) \le \prod_{d=-D+1}^{0}\{f_2(Y_{\hat{\tau}_n+d})/f_1(Y_{\hat{\tau}_n+d})\}$. Therefore, as the data given we were able to have arbitrarily small $\epsilon(Y, D)$ by expanding the range of observations that would be considered in the conditional MLE.

Therefore, in implementing the Cobb's method, $\epsilon(Y, D) = 1 - (1 - \alpha(Y, \hat{\tau}_n, D))(1 - \beta(Y, \hat{\tau}_n, D))$ was computed with increasing D such that $\epsilon(Y, D)$ was less than a pre-specified threshold, which was 0.00009 throughout the study. As the range of observations were determined, the probabilities of the change-point occurred at each time point within the range were computed as in (4.50).

5 SIMULATION STUDIES TO ASSESS ROBUSTNESS

In this chapter we carry out a simulation study for assessing the robustness of the asymptotic distribution developed in Chapter 4 for departures from normality and closeness to finite samples. The simulation study for multivariate observations will be discussed in section 5.1, and the case for univariate observations will be discussed in section 5.2. In both sections, the setup for simulations, numerical results and exploratory figures will be presented.

5.1 Simulation Setup

In our simulation, a sample of d-dimensional observations $Y_1, Y_2, \dots, Y_T$ with a change-point τ would be generated. Before the change, the mean vector and covariance matrix were μ_o and Σ_0 (variance is σ_0^2 if $d = 1$). After the change, the mean vector and covariance matrix were μ_1 and Σ_1 (variance is σ_1^2 if $d = 1$). According to Chapter 4, the parameters that determined the asymptotic distribution of the change-point mle were $\eta = \Sigma_0^{-1/2}(\mu_1 - \mu_0)$ and $K = \Sigma_0^{1/2}\Sigma_1^{-1/2}$ for multivariate case, and $\eta = (\mu_1 - \mu_0)/\sigma_0$ and $K = \sigma_0/\sigma_1$ for univariate case. In order to conform with the parameter setup of Jandhyala and Fotopoulos (1999), the parameter $\delta = \frac{1}{2}\eta$ would be used in the simulation study. In the algorithmic procedures for estimation of change-point mle, the linear combination of chi-square distribution was uniquely determined by η and KK^T. Therefore, in the simulation study, the values of $\delta = \frac{1}{2}\eta$ and $\det(KK^T)$ were fixed, and then the mean and covariance matrix before and after the change-point were set accordingly. Without loss of generality, the mean before the change was set to $\mu_o = [0, 0, \dots, 0]^T$, and the covariance (variance) before the change was set to $\Sigma_0 = \text{diag}\{1,1,\dots 1\}$. The mean and covariance (variance) after the change was computed so that they satisfied the pre-specified values of δ and $\det(KK^T)$. In the following study, the δ was chosen be to 1.5 or 2, and $\det(KK^T)$ was chosen to be 1, 1.1, 1.6, where 1 corresponds to the case with change in mean only.

As the observations were generated, the change-point estimation methods that were mentioned in Chapter 3 and 4, maximum likelihood, Cobb's method, and Bayesian method with non-informative priors and Normal-Wishart conjugate priors were applied. The mean and square root of the mean squared error (MSE) for the change-point estimates were computed to compare the differences between methods. The above procedure was repeated 100,000 times to eliminate the random errors that were introduced during random sample generation. For maximum likelihood method, the change-point was detected for each repetition so that the mean and mean squared error for the change-point mle was computed from the sample estimation. For the Cobb's and Bayesian methods, the mean and MSE were computed for each repetition, and the reported values were the average of the results.

Hinkley (1972) had proved that the distribution of change-point mle when the parameters were unknown were equivalent to the case when the parameters were unknown. Hence, in change-point estimation, if the change occurred at $\hat{\tau}_n$, then the estimated mean and covariance before and after the change-point, $\hat{\mu}_0, \hat{\mu}_1, \hat{\Sigma}_0, \hat{\Sigma}_1$, were regarded as the true parameters for the observations. In our simulation study, we would investigate the differences of using the true parameters and the estimated ones. Therefore, when applying maximum likelihood estimation and Cobb's method, four cases were applied: (i) "kk" – use the known mean and covariance/variance as the true parameters; (ii) "ke" – use the know mean and the estimated covariance/variance as the true parameters; (iii) "ek" – use the estimated mean and the known covariance/variance as the true parameters; (iv) "ee" – use the estimated mean and covariance/variance as the the parameters.

As we discussed the asymptotic distributions for change-point mle, all the observations were assumed to follow independent normal distributions. In order to investigate how much the estimation would be affected by the deviation from normality, we also let the observations follow multivariate or univariate t-distribution with an increasing degree of freedom, 5, 10, 20, to see how close the estimations were to the normal case, which could be regarded as a t-distribution of infinity degree of freedom.

In the j^{th} iteration of the simulation study, we generate the j^{th} sample of T observations, $Y_{j,1}, Y_{j,2}, \dots, Y_{j,T}$ where $Y_{j,i} \in \mathbb{R}^d$ with a change in parameters at time τ. Before the change-point, the mean and covariance are τ_0 and Σ_0 respectively, after the change-point, the corresponding parameters are τ_1 and Σ_1.

The maximum likelihood change-point detection is performed on the j^{th} sample, $Y_{j,1}, Y_{j,2}, \dots, Y_{j,T}$. In computing the log-likelihood function for the sample when change occurs at observation t, where $t = d + 1, \dots, T - d$, the mean and covariance matrix are required. These parameters are pre-specified in the simulation; however, in real-world applications, it is impossible to know the true values of mean and covariance matrix. The parameters used in the likelihood functions are the maximum likelihood estimates of the parameters. We would like to know whether we can obtain similar results using simulation and theoretical distributions in Chapter 4. We would also like to know if the closeness can be obtained for real life applications when the parameters are estimated, and how the parameter estimation affects the change-point estimation. Therefore when the mean and covariance matrix are used, we would use the following 4 cases:

1. 'kk' - known mean and covariance matrix:

$$\mu_{j,0}^{kk} = \mu_0, \qquad \Sigma_{j,o}^{kk} = \Sigma_0, \qquad \mu_{j,1}^{kk} = \mu_1, \qquad \Sigma_{j,1}^{kk} = \Sigma_1$$

2. 'ke' - known mean and estimated covariance matrix:

$$\mu_{j,0}^{ke} = \mu_0, \qquad \Sigma_{j,o}^{ke} = \frac{1}{t}\sum_{i=1}^{t}\left(Y_{j,i} - \mu_0\right)\left(Y_{j,i} - \mu_0\right)^T$$

$$\mu_{j,1}^{ke} = \mu_1, \qquad \Sigma_{j,1}^{ke} = \frac{1}{T-t}\sum_{i=1}^{T-t}\left(Y_{j,i} - \mu_1\right)\left(Y_{j,i} - \mu_1\right)^T$$

3. 'ek' - estimated mean and known covariance matrix:

$$\mu_{j,0}^{ek} = \frac{1}{t}\sum_{i=1}^{t} Y_{j,i}\,, \qquad \Sigma_{j,0}^{ek} = \Sigma_0$$

$$\mu_{j,1}^{ek} = \frac{1}{T-t}\sum_{i=1}^{T-t} Y_{j,i}\,, \qquad \Sigma_{j,1}^{ek} = \Sigma_1$$

4. 'ee' - estimated mean and covariance matrix:

$$\mu_{j,0}^{ee} = \frac{1}{t}\sum_{i=1}^{t} Y_{j,i}\,, \qquad \Sigma_{j,0}^{ee} = \frac{1}{t}\sum_{i=1}^{t}\left(Y_{j,i} - \mu_0^{ee}\right)\left(Y_{j,i} - \mu_0^{ee}\right)^T$$

$$\mu_{j,1}^{ee} = \frac{1}{T-t}\sum_{i=1}^{T-t} Y_{j,i}\,, \qquad \Sigma_{j,1}^{ee} = \frac{1}{T-t}\sum_{i=1}^{T-t}\left(Y_{j,i} - \mu_1^{ee}\right)\left(Y_{j,i} - \mu_1^{ee}\right)^T$$

Thus the likelihood functions for above 4 estimations are computed as

$$L_j^{est}(t) = \sum\nolimits_{i=1}^{t} \log f\left(Y_{j,i}; \mu_{j,0}^{est}, \Sigma_{j,0}^{est}\right) + \sum\nolimits_{i=t+1}^{T} \log f\left(Y_{j,i}; \mu_{j,1}^{est}, \Sigma_{j,1}^{est}\right)$$

where $est \in \{\mathrm{kk}, \mathrm{ke}, \mathrm{ek}, \mathrm{ee}\}, \qquad t \in \{d+1, \dots, T-d\}$

The detected change-point under the 4 cases is

$$\hat{\tau}_j^{est} = \arg \max_{\mathrm{d+1 \le t \le T-d}} \{L_j^{est}(t)\}$$

where $est \in \{\mathrm{kk}, \mathrm{ke}, \mathrm{ek}, \mathrm{ee}\}$

After a significant change-point is detected, estimation using Cobb's method is applied for the j^{th} sample. In Cobb's method, we need to determine the range of samples that provides the ancillary information, which is determined as $\{\hat{\tau}_j^{est} - D_j^{est}, \dots, \hat{\tau}_j^{est} + D_j^{est}\}$ where D_j^{est} is the minimum d_j^{est} that satisfies

$$\epsilon\left(Y_j, d_j^{est}\right) = 1 - \left(1 - \alpha(Y_j, \hat{\tau}_j^{est}, d_j^{est})\right)\left(1 - \beta(Y_j, \hat{\tau}_j^{est}, d_j^{est})\right) \le \epsilon$$

where
$$\alpha\left(Y_j, \hat{\tau}_j^{est}, d_j^{est}\right) = \prod_{d=1}^{d_j^{est}} \left\{ f(Y_{\hat{\tau}_j^{est}+d}; \mu_{j,0}^{est}, \Sigma_{j,0}^{est}) / f(Y_{\hat{\tau}_j^{est}+d}; \mu_{j,1}^{est}, \Sigma_{j,1}^{est}) \right\}$$

$$\beta\left(Y_j, \hat{\tau}_j^{est}, d_j^{est}\right) = \prod_{d=-d_j^{est}+1}^{0} \left\{ f(Y_{\hat{\tau}_j^{est}+d}; \mu_{j,1}^{est}, \Sigma_{j,1}^{est}) / f(Y_{\hat{\tau}_j^{est}+d}; \mu_{j,0}^{est}, \Sigma_{j,0}^{est}) \right\}$$

ϵ is the pre-specified threshold, and $est \in \{\mathrm{kk}, \mathrm{ke}, \mathrm{ek}, \mathrm{ee}\}$.

Then the probabilities using Cobb's method is

$$\Pr(\tau_j^{est} - \hat{\tau}_j^{est} = d) = \frac{\exp[L_j^{est}\left(\hat{\tau}_j^{est} + d\right)]}{\sum_{i=-D_j^{est}}^{D_j^{est}} \exp[L_j^{est}\left(\hat{\tau}_j^{est} + i\right)]}$$

where $d \in \{-D_j^{est}, \dots, D_j^{est}\}$ and $est \in \{\mathrm{kk}, \mathrm{ke}, \mathrm{ek}, \mathrm{ee}\}$

Then the computation for the Bayesian probabilities directly follows (4.47) - (4.49) for non-informative prior, and follows (4.40), (4.42) and (4.46) for conjugate prior.

The above procedures are repeated for $N = 100{,}000$ times. The bias and mean square error for the maximum likelihood method are computed using the sample mean and mean square error. That is to say,

$$\mathrm{Bias}(\hat{\tau}^{ml.est}) = \frac{1}{N} \sum_{j=1}^{N} \hat{\tau}_j^{est} - \tau$$

$$\mathrm{MSE}(\hat{\tau}^{ml.est}) = \frac{1}{N} \sum_{j=1}^{N} \left(\hat{\tau}_j^{est} - \tau\right)^2$$

Where $est \in \{\mathrm{kk}, \mathrm{ke}, \mathrm{ek}, \mathrm{ee}\}$

For the other methods, Cobb's conditional probabilities and Bayesian method, the probability of a change-point at each observation is computed, thus we are able to compute the Bias and MSE as

$$\text{Bias}(\hat{\tau}) = \frac{1}{N}\sum_{j=1}^{N}\sum_{t=1}^{T-1} t p_{j,t} - \tau$$

$$\text{MSE}(\hat{\tau}) = \frac{1}{N}\sum_{j=1}^{N}\sum_{t=1}^{T-1} (t-\tau)^2 p_{j,t}$$

where $p_{j,t}$ is the probability that the change-point occurs at the observation t for the j^{th} sample. In Cobb's method, $p_{j,t} = 0$ if $t \notin\in \{-D_j^{est} + \hat{\tau}_j^{cobb.est}, \ldots, D_j^{est} + \hat{\tau}_j^{cobb.est}\}$

Lastly, we let $d = 1\ or\ 2$ to investigate the effect of dimensionality to the change-point estimation. We chose the combination of $T/\tau = 100/50, 100/30, 50/25$ to investigate the effect of sample size and position of the change-point. That corresponded to the following cases: (i) 100 observations with change occurred at the 50th; (ii) 100 observations with change occurred at the 30th; (iii) 50 observations with change occurred at the 25th.

Section 5.2 would perform the simulation for bivariate observations, and 5.3 for univariate observations. The tables of mean and square root of MSE were presented in tables figures. The conclusions would be drawn in Section 5.4.

5.2 Multivariate Simulations

As specified in Section 5.1, a sample of T observations with change-point τ would be generated 100,000 times. Before the change-point, the mean was $\mu_0 = [0 \quad 0]^T$. After the change-point, $\delta = 1.5$ corresponded to $\mu_1 = \left[3/\sqrt{2} \quad 3/\sqrt{2}\right]^T$, and $\delta = 2$ corresponded to $\mu_1 = [2\sqrt{2} \quad 2\sqrt{2}]^T$. Before the change-point, the covariance matrix was $\Sigma_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. After the change-point, $\det(KK^T) = 1.1$ corresponded to $\Sigma_1 = \begin{bmatrix} 1 & 0.3 \\ 0.3 & 1 \end{bmatrix}$, and $\det(KK^T) = 1.6$ corresponded to $\Sigma_1 = \begin{bmatrix} 1 & 0.6 \\ 0.6 & 1 \end{bmatrix}$. The square roots of the MSE were presented in Tables 5.1 – 5.6, and the biases of the change-point mle were presented in Tables 5.7 – 5.12.

Table 5.1 Square root of mean squared error of the change-point mle when $T/\tau = 100/50$ and $\delta = 1.5$ for bivariate series.

K2	df	Theory	MLE				Cobb				Bayesian	
			kk	ke	ek	ee	kk	ke	ek	ee	Non-Info.	Conj.
1	5		0.9388	0.9366	1.0240	0.9894	0.9988	1.1293	1.0511	1.1965	1.2640	1.2483
1	10		0.6711	0.6779	0.7030	0.7080	0.7555	0.8191	0.7928	0.8572	0.8857	0.8731
1	20		0.5780	0.5821	0.6021	0.6069	0.6749	0.7119	0.7063	0.7416	0.7632	0.7527
1	Inf	0.5057	0.5061	0.5109	0.5255	0.5333	0.6146	0.6294	0.6420	0.6542	0.6710	0.6620
1.1	5		1.0912	1.2630	1.2672	1.7049	1.1524	1.4573	1.2628	1.6658	2.2159	1.3959
1.1	10		0.7617	0.8156	0.8045	0.8823	0.8616	0.9870	0.9120	1.0575	1.1134	0.9999
1.1	20		0.6627	0.6973	0.6968	0.7463	0.7759	0.8533	0.8187	0.9062	0.9421	0.8693
1.1	Inf	0.5842	0.5815	0.6084	0.6080	0.6396	0.7086	0.7519	0.7444	0.7872	0.8167	0.7694
1.6	5		1.2085	1.3068	1.4383	1.7290	1.2595	1.5213	1.4306	1.7334	2.2245	1.4438
1.6	10		0.8012	0.8545	0.8499	0.9256	0.9089	1.0438	0.9636	1.1209	1.1743	1.0479
1.6	20		0.7033	0.7424	0.7419	0.7885	0.8249	0.9107	0.8717	0.9657	1.0027	0.9212
1.6	Inf	0.6279	0.6277	0.6554	0.6567	0.6895	0.7636	0.8096	0.8058	0.8514	0.8797	0.8216

Table 5.2 Square root of mean squared error of the change-point mle when $T/\tau = 100/50$ and $\delta = 2$ for bivariate series.

K2	df	Theory	MLE				Cobb				Bayesian	
			kk	ke	ek	ee	kk	ke	ek	ee	Non-Info.	Conj.
1	5		0.5288	0.5278	0.5618	0.5442	0.5510	0.6173	0.5678	0.6365	0.6538	0.6453
1	10		0.3509	0.3544	0.3624	0.3656	0.3902	0.4238	0.4034	0.4358	0.4445	0.4389
1	20		0.2924	0.2968	0.2993	0.3041	0.3385	0.3601	0.3487	0.3693	0.3750	0.3702
1	Inf	0.2398	0.2397	0.2444	0.2481	0.2517	0.2952	0.3052	0.3051	0.3130	0.3178	0.3136
1.1	5		0.6156	0.6886	0.6645	0.7994	0.6388	0.7941	0.6688	0.8511	0.9841	0.7956
1.1	10		0.4067	0.4337	0.4196	0.4506	0.4546	0.5225	0.4701	0.5410	0.5580	0.5322
1.1	20		0.3418	0.3641	0.3502	0.3753	0.3980	0.4442	0.4111	0.4577	0.4684	0.4517
1.1	Inf	0.2873	0.2895	0.3052	0.2975	0.3162	0.3538	0.3813	0.3660	0.3936	0.4013	0.3889
1.6	5		0.6854	0.7272	0.7661	0.8289	0.7010	0.8514	0.7587	0.9180	1.0508	0.8484
1.6	10		0.4423	0.4693	0.4558	0.4863	0.4964	0.5699	0.5151	0.5927	0.6102	0.5803
1.6	20		0.3765	0.3972	0.3857	0.4089	0.4392	0.4887	0.4543	0.5047	0.5171	0.4968
1.6	Inf	0.3216	0.3238	0.3401	0.3323	0.3526	0.3964	0.4258	0.4107	0.4396	0.4481	0.4332

Table 5.3 Square root of mean squared error of the change-point mle when $T/\tau = 100/30$ and $\delta = 1.5$ for bivariate series.

K2	df	Theory	MLE				Cobb				Bayesian	
			kk	ke	ek	ee	kk	ke	ek	ee	Non-Info.	Conj.
1	5		0.9492	0.9442	1.0796	1.0541	1.0065	1.1312	1.0843	1.2171	1.3207	1.3086
1	10		0.6667	0.6732	0.7034	0.7133	0.7504	0.8163	0.7934	0.8609	0.8956	0.8812
1	20		0.5782	0.5847	0.6067	0.6168	0.6739	0.7125	0.7089	0.7473	0.7737	0.7618
1	Inf	0.5057	0.5057	0.5116	0.5313	0.5377	0.6138	0.6292	0.6448	0.6568	0.6770	0.6670
1.1	5		1.1048	1.4299	1.3423	2.7625	1.1544	1.5983	1.2983	1.8653	3.4017	4.0741
1.1	10		0.7616	0.8438	0.8140	0.9526	0.8594	1.0091	0.9165	1.0955	1.1854	2.1246
1.1	20		0.6628	0.7181	0.6995	0.7960	0.7744	0.8684	0.8197	0.9290	0.9807	1.6063
1.1	Inf	0.5842	0.5839	0.6191	0.6162	0.6623	0.7097	0.7606	0.7485	0.8045	0.8455	1.2701
1.6	5		1.2205	1.4770	1.7063	2.7452	1.2543	1.6705	1.6252	1.9364	3.4062	3.9952
1.6	10		0.8066	0.8875	0.8571	1.0123	0.9112	1.0676	0.9703	1.1589	1.2509	2.1124
1.6	20		0.7067	0.7573	0.7458	0.8322	0.8270	0.9222	0.8755	0.9848	1.0374	1.6171
1.6	Inf	0.6279	0.6280	0.6652	0.6596	0.7114	0.7639	0.8158	0.8057	0.8644	0.9034	1.2958

Table 5.4 Square root of mean squared error of the change-point mle when $T/\tau = 100/30$ and $\delta = 2$ for bivariate series.

K2	df	Theory	MLE				Cobb				Bayesian	
			kk	ke	ek	ee	kk	ke	ek	ee	Non-Info.	Conj.
1	5		0.5325	0.5312	0.5719	0.5498	0.5579	0.6201	0.5773	0.6429	0.6662	0.6570
1	10		0.3535	0.3571	0.3636	0.3682	0.3920	0.4259	0.4050	0.4388	0.4480	0.4420
1	20		0.2923	0.2969	0.3028	0.3061	0.3387	0.3605	0.3511	0.3709	0.3781	0.3730
1	Inf	0.2398	0.2397	0.2442	0.2489	0.2540	0.2953	0.3052	0.3068	0.3148	0.3199	0.3154
1.1	5		0.6289	0.7853	0.7124	1.2784	0.6384	0.8822	0.6745	0.9857	1.5121	1.7881
1.1	10		0.4099	0.4405	0.4238	0.4671	0.4562	0.5279	0.4729	0.5546	0.5786	0.7616
1.1	20		0.3436	0.3673	0.3532	0.3828	0.3989	0.4482	0.4130	0.4649	0.4812	0.5821
1.1	Inf	0.2873	0.2893	0.3066	0.2990	0.3203	0.3544	0.3843	0.3680	0.3980	0.4093	0.4711
1.6	5		0.6854	0.7272	0.7661	0.8289	0.7010	0.8514	0.7587	0.9180	1.0508	1.8565
1.6	10		0.4423	0.4693	0.4558	0.4863	0.4964	0.5699	0.5151	0.5927	0.6102	0.8121
1.6	20		0.3765	0.3972	0.3857	0.4089	0.4392	0.4887	0.4543	0.5047	0.5171	0.6253
1.6	Inf	0.3216	0.3244	0.3409	0.3352	0.3577	0.3969	0.4268	0.4121	0.4422	0.4546	0.5132

Table 5.5 Square root of mean squared error of the change-point mle when $T/\tau = 50/25$ and $\delta = 1.5$ for bivariate series.

K2	df	Theory	MLE				Cobb				Bayesian	
			kk	ke	ek	ee	kk	ke	ek	ee	Non-Info.	Conj.
1	5		0.9382	0.9338	1.1072	1.0864	0.9979	1.1340	1.1307	1.2832	1.4983	1.4058
1	10		0.6622	0.6770	0.7264	0.7499	0.7475	0.8323	0.8238	0.9166	0.9883	0.9446
1	20		0.5741	0.5873	0.6246	0.6439	0.6710	0.7274	0.7364	0.7948	0.8447	0.8113
1	Inf	0.5057	0.5014	0.5153	0.5436	0.5620	0.6108	0.6436	0.6686	0.6982	0.7351	0.7106
1.1	5		1.0760	1.3205	1.3277	2.7627	1.1420	1.5186	1.3353	1.8096	3.4087	1.3708
1.1	10		0.7516	0.8748	0.8408	1.6060	0.8526	1.0582	0.9572	1.2248	1.8857	1.0305
1.1	20		0.6594	0.7409	0.7233	1.2694	0.7718	0.9117	0.8586	1.0449	1.4520	0.9108
1.1	Inf	0.5842	0.5805	0.6472	0.6336	0.9930	0.7064	0.8027	0.7831	0.9196	1.1392	0.8115
1.6	5		1.1577	1.3607	1.4447	2.7119	1.2203	1.5775	1.4471	1.8698	3.3626	1.4069
1.6	10		0.7977	0.9168	0.8879	1.6210	0.9036	1.1176	1.0136	1.2913	1.9197	1.0740
1.6	20		0.6971	0.7855	0.7693	1.2914	0.8205	0.9708	0.9177	1.1125	1.5082	0.9575
1.6	Inf	0.6279	0.6213	0.6869	0.6798	1.0266	0.7585	0.8574	0.8466	0.9894	1.2032	0.8604

Table 5.6 Square root of mean squared error of the change-point mle when $T/\tau = 50/25$ and $\delta = 2$ for bivariate series.

K2	df	Theory	MLE				Cobb				Bayesian	
			kk	ke	ek	ee	kk	ke	ek	ee	Non-Info.	Conj.
1	5		0.5252	0.5226	0.5610	0.5530	0.5510	0.6227	0.5813	0.6644	0.7188	0.6882
1	10		0.3470	0.3562	0.3658	0.3776	0.3873	0.4346	0.4104	0.4586	0.4775	0.4618
1	20		0.2897	0.2982	0.3064	0.3168	0.3355	0.3689	0.3566	0.3888	0.4021	0.3891
1	Inf	0.2398	0.2390	0.2458	0.2507	0.2603	0.2932	0.3138	0.3116	0.3303	0.3408	0.3301
1.1	5		0.6056	0.7177	0.6837	1.1000	0.6366	0.8324	0.6982	0.9473	1.3275	0.8185
1.1	10		0.4038	0.4627	0.4280	0.5754	0.4514	0.5628	0.4824	0.6120	0.6939	0.5702
1.1	20		0.3405	0.3803	0.3602	0.4725	0.3954	0.4740	0.4224	0.5120	0.5610	0.4849
1.1	Inf	0.2873	0.2867	0.3179	0.3042	0.3720	0.3517	0.4080	0.3767	0.4385	0.4648	0.4183
1.6	5		0.6650	0.7581	0.7799	1.1525	0.6971	0.8927	0.7871	1.0314	1.4060	0.8701
1.6	10		0.4390	0.4992	0.4666	0.6176	0.4935	0.6124	0.5288	0.6715	0.7585	0.6049
1.6	20		0.3758	0.4171	0.3964	0.5094	0.4374	0.5225	0.4691	0.5672	0.6187	0.5210
1.6	Inf	0.3216	0.3211	0.3526	0.3414	0.4037	0.3945	0.4535	0.4241	0.4899	0.5199	0.4637

Table 5.7 Bias of the change-point mle when $T/\tau = 100/50$ and $\delta = 1.5$ for bivariate series.

K2	df	Theory	MLE				Cobb				Bayesian	
			kk	ke	ek	ee	kk	ke	ek	ee	Non-Info.	Conj.
1	5		-0.0037	-0.0024	-0.0026	-0.0027	-0.0315	-0.0177	-0.0329	-0.0209	-0.0035	-0.0034
1	10		-0.0006	-0.0012	-0.0001	-0.0005	-0.0068	-0.0088	-0.0076	-0.0104	-0.0104	-0.0009
1	20		-0.0011	-0.0016	-0.0011	-0.0012	-0.0054	-0.0066	-0.0066	-0.0078	-0.0011	-0.0011
1	Inf	0.0000	0.0001	0.0001	0.0006	0.0006	-0.0036	-0.0043	-0.0042	-0.0049	-0.0005	-0.0005
1.1	5		0.0353	0.0237	0.0291	0.0143	-0.0113	-0.0400	-0.0215	-0.0908	-0.0490	-0.0283
1.1	10		0.0267	0.0224	0.0269	0.0209	0.0023	-0.0180	-0.0021	-0.0259	-0.0175	-0.0106
1.1	20		0.0229	0.0205	0.0209	0.0194	-0.0013	-0.0117	-0.0059	-0.0168	-0.0105	-0.0061
1.1	Inf	0.0190	0.0211	0.0198	0.0200	0.0206	-0.0023	-0.0070	-0.0064	-0.0100	-0.0061	-0.0047
1.6	5		0.1088	0.0540	0.1096	0.0413	0.0515	-0.0538	0.0455	-0.1115	-0.0854	-0.0433
1.6	10		0.0619	0.0454	0.0606	0.0424	0.0226	-0.0238	0.0157	-0.0358	-0.0331	-0.0182
1.6	20		0.0485	0.0418	0.0450	0.0379	0.0084	-0.0153	-0.0004	-0.0247	-0.0221	-0.0121
1.6	Inf	0.0351	0.0354	0.0348	0.0327	0.0320	-0.0037	-0.0114	-0.0119	-0.0183	-0.0166	-0.0095

Table 5.8 Bias of the change-point mle when $T/\tau = 100/50$ and $\delta = 2$ for bivariate series.

K2	df	Theory	MLE				Cobb				Bayesian	
			kk	ke	ek	ee	kk	ke	ek	ee	Non-Info.	Conj.
1	5		-0.0023	-0.0011	-0.0009	0.0000	-0.0230	-0.0048	-0.0216	-0.0047	-0.0009	-0.0009
1	10		-0.0004	-0.0006	-0.0002	-0.0004	-0.0013	-0.0019	-0.0014	-0.0021	-0.0005	-0.0005
1	20		0.0006	0.0004	0.0007	0.0009	-0.0004	-0.0008	-0.0003	-0.0008	0.0002	0.0002
1	Inf	0.0000	0.0006	0.0005	0.0005	0.0004	0.0000	-0.0002	0.0001	-0.0001	0.0004	0.0004
1.1	5		0.0141	0.0109	0.0144	0.0094	-0.0128	-0.0131	-0.0125	-0.0219	-0.0152	-0.0098
1.1	10		0.0102	0.0082	0.0099	0.0081	0.0033	-0.0048	0.0022	-0.0064	-0.0050	-0.0032
1.1	20		0.0085	0.0073	0.0087	0.0079	0.0017	-0.0027	0.0005	-0.0037	-0.0029	-0.0018
1.1	Inf	0.0055	0.0062	0.0057	0.0062	0.0055	-0.0001	-0.0022	-0.0009	-0.0030	-0.0024	-0.0016
1.6	5		0.0445	0.0288	0.0440	0.0254	0.0143	-0.0174	0.0127	-0.0286	-0.0258	-0.0143
1.6	10		0.0245	0.0189	0.0238	0.0182	0.0117	-0.0064	0.0095	-0.0097	-0.0098	-0.0055
1.6	20		0.0177	0.0157	0.0173	0.0152	0.0052	-0.0046	0.0030	-0.0069	-0.0069	-0.0041
1.6	Inf	0.0114	0.0121	0.0114	0.0118	0.0115	-0.0001	-0.0043	-0.0023	-0.0057	-0.0060	-0.0098

Table 5.9 Bias of the change-point mle when $T/\tau = 100/30$ and $\delta = 1.5$ for bivariate series.

K2	df	Theory	MLE				Cobb				Bayesian	
			kk	ke	ek	ee	kk	ke	ek	ee	Non-Info.	Conj.
1	5		-0.0006	-0.0001	0.0216	0.0200	-0.0193	-0.0144	0.0009	0.0093	0.0197	0.0191
1	10		-0.0006	-0.0004	0.0089	0.0081	-0.0067	-0.0082	0.0039	0.0041	0.0082	0.0077
1	20		0.0009	0.0004	0.0068	0.0077	-0.0032	-0.0042	0.0053	0.0055	0.0077	0.0074
1	Inf	0.0000	-0.0002	0.0004	0.0058	0.0061	-0.0026	-0.0031	0.0050	0.0047	0.0056	0.0053
1.1	5		0.0389	0.0727	0.0609	0.1847	-0.0001	0.0193	0.0189	-0.0193	0.1566	1.5087
1.1	10		0.0269	0.0284	0.0364	0.0351	0.0022	-0.0042	0.0108	0.0002	-0.0017	0.6533
1.1	20		0.0258	0.0240	0.0330	0.0289	0.0021	-0.0010	0.0084	0.0030	-0.0003	0.4640
1.1	Inf	0.0190	0.0209	0.0198	0.0255	0.0233	-0.0027	0.0006	0.0032	0.0035	0.0000	0.3464
1.6	5		0.1112	0.1029	0.1542	0.2056	0.0602	0.0048	0.0955	-0.0440	0.1017	1.4537
1.6	10		0.0613	0.0498	0.0672	0.0554	0.0214	-0.0104	0.0259	-0.0106	-0.0232	0.6522
1.6	20		0.0495	0.0422	0.0540	0.0463	0.0097	-0.0048	0.0122	-0.0046	-0.0161	0.4727
1.6	Inf	0.0351	0.0366	0.0365	0.0399	0.0389	-0.0028	-0.0023	-0.0014	-0.0021	-0.0117	0.3599

Table 5.10 Bias of the change-point mle when $T/\tau = 100/30$ and $\delta = 2$ for bivariate series.

K2	df	Theory	MLE				Cobb				Bayesian	
			kk	ke	ek	ee	kk	ke	ek	ee	Non-Info.	Conj.
1	5		-0.0007	-0.0007	0.0047	0.0049	-0.0136	-0.0039	-0.0070	0.0039	0.0053	0.0049
1	10		-0.0022	-0.0016	0.0014	0.0017	-0.0022	-0.0024	0.0015	0.0018	0.0021	0.0019
1	20		-0.0004	-0.0008	0.0013	0.0010	-0.0010	-0.0012	0.0017	0.0018	0.0017	0.0016
1	Inf	0.0000	-0.0007	-0.0004	0.0009	0.0006	-0.0006	-0.0008	0.0017	0.0016	0.0013	0.0012
1.1	5		0.0141	0.0331	0.0205	0.0600	-0.0043	0.0158	0.0023	0.0100	0.0506	0.3844
1.1	10		0.0082	0.0106	0.0115	0.0119	0.0024	0.0024	0.0057	0.0037	0.0040	0.1409
1.1	20		0.0071	0.0077	0.0092	0.0091	0.0011	0.0023	0.0036	0.0034	0.0031	0.1003
1.1	Inf	0.0055	0.0046	0.0052	0.0063	0.0059	-0.0008	0.0017	0.0013	0.0024	0.0020	0.0749
1.6	5		0.0445	0.0288	0.0440	0.0254	0.0143	-0.0174	0.0127	-0.0286	-0.0258	0.4111
1.6	10		0.0245	0.0189	0.0238	0.0182	0.0117	-0.0064	0.0095	-0.0097	-0.0098	0.1568
1.6	20		0.0177	0.0157	0.0173	0.0152	0.0052	-0.0046	0.0030	-0.0069	-0.0069	0.1131
1.6	Inf	0.0114	0.0108	0.0103	0.0124	0.0117	-0.0008	-0.0001	0.0006	0.0002	-0.0021	0.0855

Table 5.11 Bias of the change-point mle when $T/\tau = 50/25$ and $\delta = 1.5$ for bivariate series.

K2	df	Theory	MLE				Cobb				Bayesian	
			kk	ke	ek	ee	kk	ke	ek	ee	Non-Info.	Conj.
1	5		0.0001	-0.0007	0.0017	0.0002	-0.0101	-0.0250	-0.0158	-0.0338	0.0003	0.0004
1	10		0.0022	0.0019	0.0013	0.0023	-0.0046	-0.0071	-0.0074	0.8402	0.0004	0.0006
1	20		0.0004	-0.0008	0.0001	-0.0003	-0.0041	-0.0056	-0.0069	-0.0093	-0.0001	-0.0001
1	Inf	0.0000	0.0002	0.0004	-0.0011	0.0002	-0.0027	-0.0039	-0.0056	-0.0069	-0.0005	-0.0006
1.1	5		0.0421	0.0272	0.0285	-0.0105	0.0138	-0.0410	-0.0135	-0.4484	-0.0948	-0.0258
1.1	10		0.0301	0.0268	0.0254	0.0154	0.0051	-0.0186	-0.0077	-0.1355	-0.0386	-0.0132
1.1	20		0.0237	0.0223	0.0210	0.0173	0.0005	-0.0146	-0.0107	-0.0796	-0.0293	-0.0105
1.1	Inf	0.0190	0.0198	0.0190	0.0160	0.0113	-0.0029	-0.0112	-0.0136	-0.0431	-0.0219	-0.0094
1.6	5		0.1108	0.0505	0.1066	-0.0001	0.0688	-0.0603	0.0416	-0.4640	-0.1808	-0.0448
1.6	10		0.0636	0.0485	0.0562	0.0317	0.0238	-0.0280	0.0032	-0.1531	-0.0805	-0.0257
1.6	20		0.0477	0.0405	0.0391	0.0293	0.0091	-0.0222	-0.0118	-0.0944	-0.0605	-0.0215
1.6	Inf	0.0351	0.0352	0.0343	0.0278	0.0219	-0.0029	-0.0179	-0.0230	-0.0567	-0.0461	-0.0189

Table 5.12 Bias of mean squared error of the change-point mle when $T/\tau = 50/25$ and $\delta = 2$ for bivariate series.

K2	df	Theory	MLE				Cobb				Bayesian	
			kk	ke	ek	ee	kk	ke	ek	ee	Non-Info.	Conj.
1.00	5		-0.0002	0.0001	0.0006	-0.0002	-0.0029	-0.0047	-0.0034	-0.0051	0.0001	0.0003
1.00	10		0.0005	0.0000	-0.0002	0.0000	-0.0007	-0.0013	-0.0017	-0.0021	0.0002	0.0001
1.00	20		-0.0006	0.0005	-0.0009	-0.0002	-0.0007	-0.0007	-0.0015	-0.0016	-0.0002	-0.0002
1.00	Inf	0.0000	-0.0002	0.0002	-0.0004	0.0000	-0.0002	-0.0002	-0.0006	-0.0008	-0.0001	-0.0001
1.10	5		0.0170	0.0106	0.0128	0.0061	0.0079	-0.0141	0.0020	-0.0658	-0.0276	-0.0109
1.10	10		0.0105	0.0092	0.0090	0.0079	0.0038	-0.0064	0.0006	-0.0160	-0.0108	-0.0053
1.10	20		0.0070	0.0067	0.0059	0.0043	0.0012	-0.0054	-0.0017	-0.0117	-0.0090	-0.0048
1.10	Inf	0.0055	0.0058	0.0051	0.0046	0.0040	-0.0003	-0.0046	-0.0028	-0.0080	-0.0070	-0.0045
1.56	5		0.0463	0.0291	0.0425	0.0156	0.0324	-0.0206	0.0233	-0.0849	-0.0564	-0.0193
1.56	10		0.0233	0.0179	0.0211	0.0159	0.0119	-0.0112	0.0059	-0.0249	-0.0246	-0.0091
1.56	20		0.0169	0.0136	0.0144	0.0109	0.0048	-0.0096	-0.0011	-0.0189	-0.0199	-0.0088
1.56	Inf	0.0114	0.0120	0.0102	0.0091	0.0074	-0.0006	-0.0087	-0.0059	-0.0145	-0.0164	0.0012

The mse varied with the methods that were applied, size of the change, sample size, position of the change-point, dimensionality and deviation from Normality. The mean from different methods did not vary very much. Therefore, the value of the mean squared error determined how accurately we can detect the change-point. However, the tables listed above did not provide an obvious relationship between the mse and various factors that could potentially affect it. In the rest of the section, the figures of square root of mse were plotted with certain factors fixed.

Figures 5.1-5.3 were the figures of square root of mse versus $\det(KK^T)$ under the parameter estimations using 'kk', 'ke', 'ek', 'ee' for MLE and Cobb's method. The figures for $\delta = 1.5$ and $\delta = 2$ were plotted side by side. Figures 5.1 - 5.3 represented the combination of sample size and change-point position, $T/\tau = 100/50,\ 100/30,\ 50/25$, respectively. It could be seen that the mse under MLE method were close to each other most of the time, except that when the sample size was too small (50/25), the change in mean was small, and there was change in covariance. The mse under Cobb's method were close to each using the parameter estimation method 'kk', 'ke', 'ek', 'ee'. They were systematically larger than the mse using MLE method. They were more affected by the size of the mean change, and were less vulnerable to the sample size and position of the change-point.

Figures 5.4-5.6 could be used to investigate the effect of sample size. The mse under different combinations of T/τ was plotted in the same figures. Figure 5.4 presented the mse using MLE method, including the 4 parameter estimation options. Under the case where both mean and covariance changed, if the mean and covariance were estimated, the amount of change was small, and the sample size was small, the MLE method tended to be bigger than the theoretical values. Otherwise, whether the parameters were estimated or not did not affect the mse of change-point mle very much. Figure 5.5 presented the mse using Cobb's method with the 4 parameter estimation options. This method behaved similar to the MLE method, except that the combined effect of small sample size, small change in mean and change in covariance was smaller than the MLE method. Figure 5.6 presented the mse using Bayesian methods with a non-informative prior and conjugate prior. When the change in mean was small, the Bayesian method with non-informative prior was more affected by the small sample size, while the Bayesian method with conjugate prior was more affected by the skewness of the position of the change-point.

Figures 5.7-5.10 compared the mse under MLE, Cobb's and Bayesian method. Separate figures were generated for 'kk', 'ke', 'ek' and 'ee' for MLE and Cobb. It could be observed that when the change in mean was big enough, all the methods obtained similar results. When the amount of change was small, MLE and Cobb's methods gave smaller mse than Bayesian method using non-informative or conjugate prior.

Figures 5.11-5.13 investigated the effect of departure from normality under each method. Figure 5.11 was for MLE method. When the degree of freedom was greater than 5, even if the series followed t-distribution, the mse were still close to each other. MLE method was quite resistant to the departure from the normality assumption. Figure 5.12 was for Cobb's method, and Figure 5.13 was for Bayesian method. Although the mse also decreased as df increased, the lines for mse under different degrees of freedom were not as close to each other as the MLE method. Thus, Cobb's method and Bayesian method were also resistant to departure from Normality, but they did not perform as good as the MLE method.

Figures 5.14-5.16 compared the behavior of each method under departure from Normality with degree of freedom equaled 5, 10, and 20 respectively. When df=5, the Bayesian method with non-informative prior overestimated the mse the most. It became closer to the theoretical values when the amount of change in mean increased. The MLE and Cobb's method produced very close mse, which were smaller than the mse of Bayesian method; however, when the parameters before and after change were estimated and the change in mean was small, the Bayesian method with conjugate prior produced smaller mse.

From the observations of the figures, it was concluded that the MLE method were more resistant to departure from assumptions than Cobb's and Bayesian method. Although Cobb's method produced very close results to MLE method, the mse was still slightly bigger than the mse produced by MLE. Bayesian method produced similar mse only when the amount of change was large. Otherwise, it was more sensitive to the sample size, position of change-point, and the departure from normality.

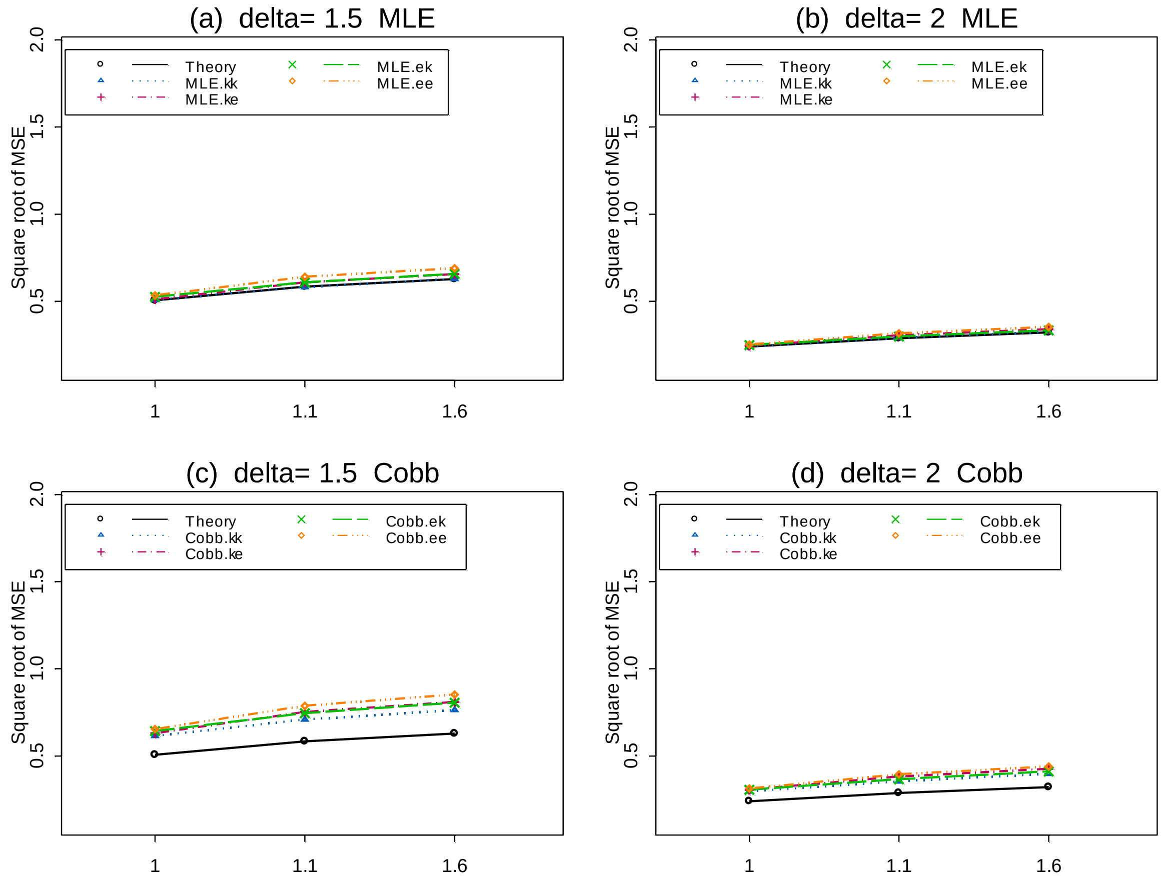

Figure 5.1 Comparison of the kk, ke, ek, and ee estimation methods for MLE and Cobb's method when $T/\tau = 100/50$ for bivariate series.

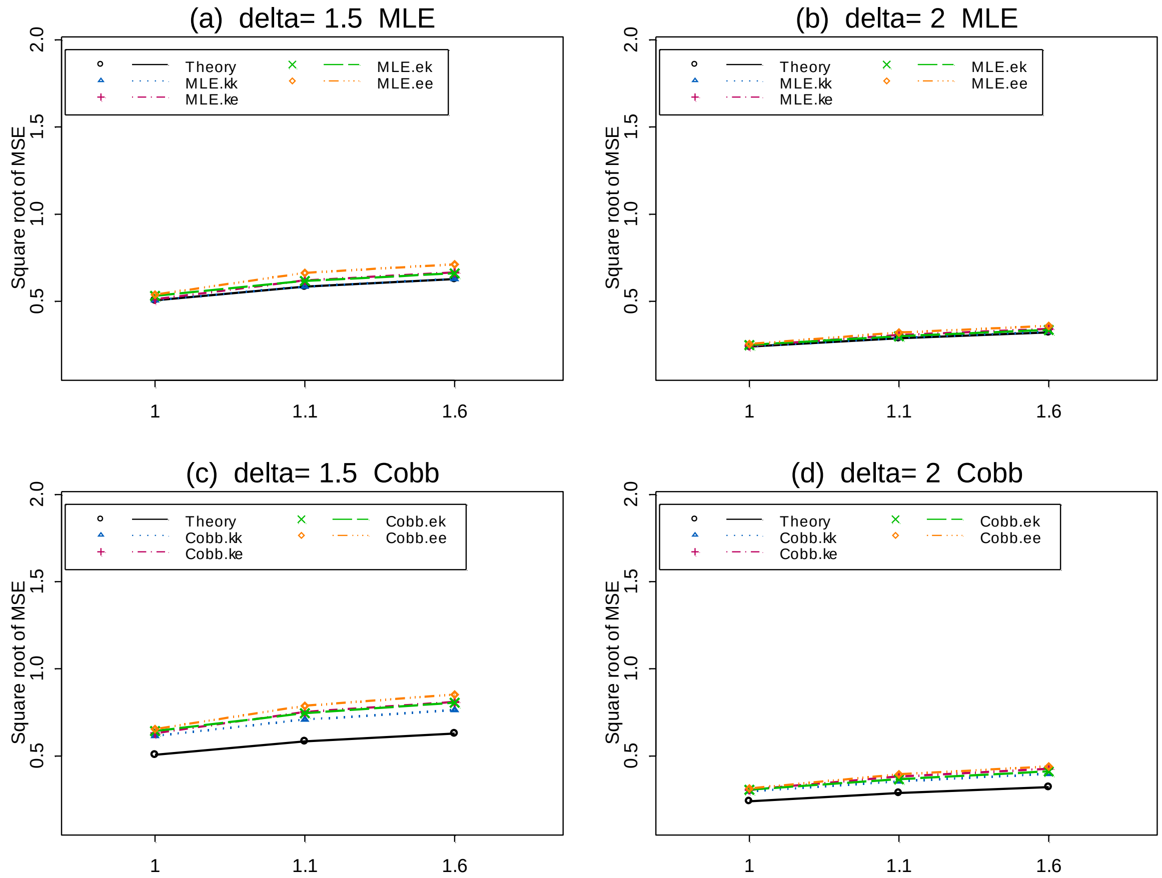

Figure 5.2 Comparison of the kk, ke, ek, and ee estimation method for MLE and Cobb's method when $T/\tau = 100/30$ for bivariate series.

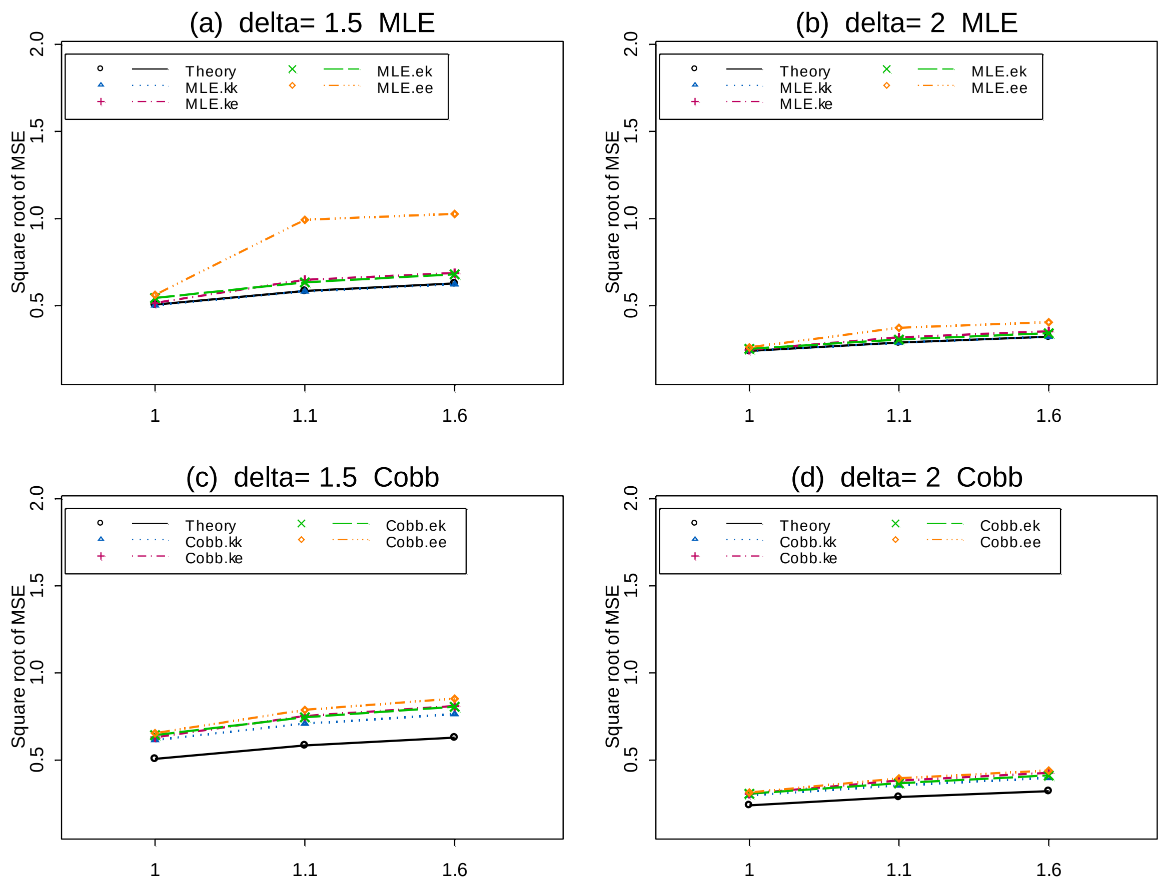

Figure 5.3 Comparison of the kk, ke, ek, and ee estimation method for MLE and Cobb's method when $T/\tau = 50/25$ for bivariate series.

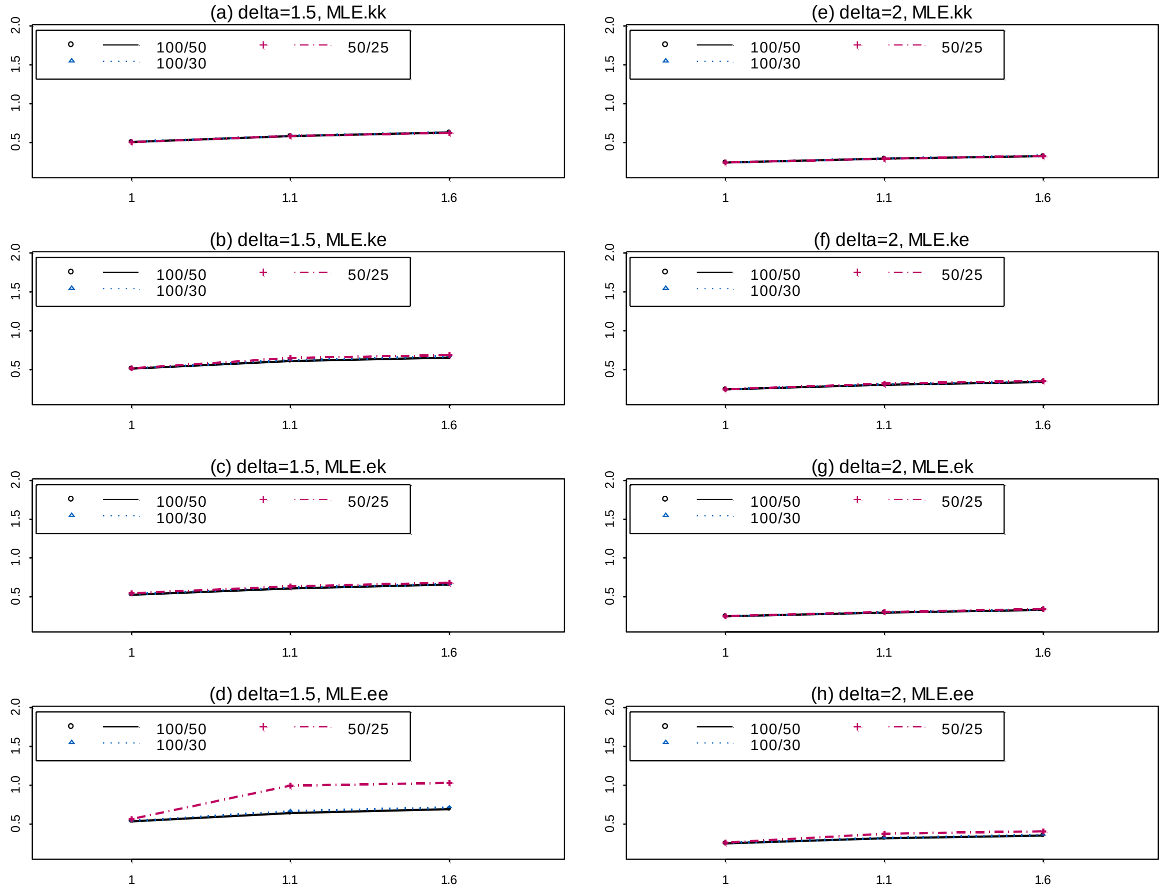

Figure 5.4 The effect of sample size and change-point position to the MLE estimation method for bivariate series.

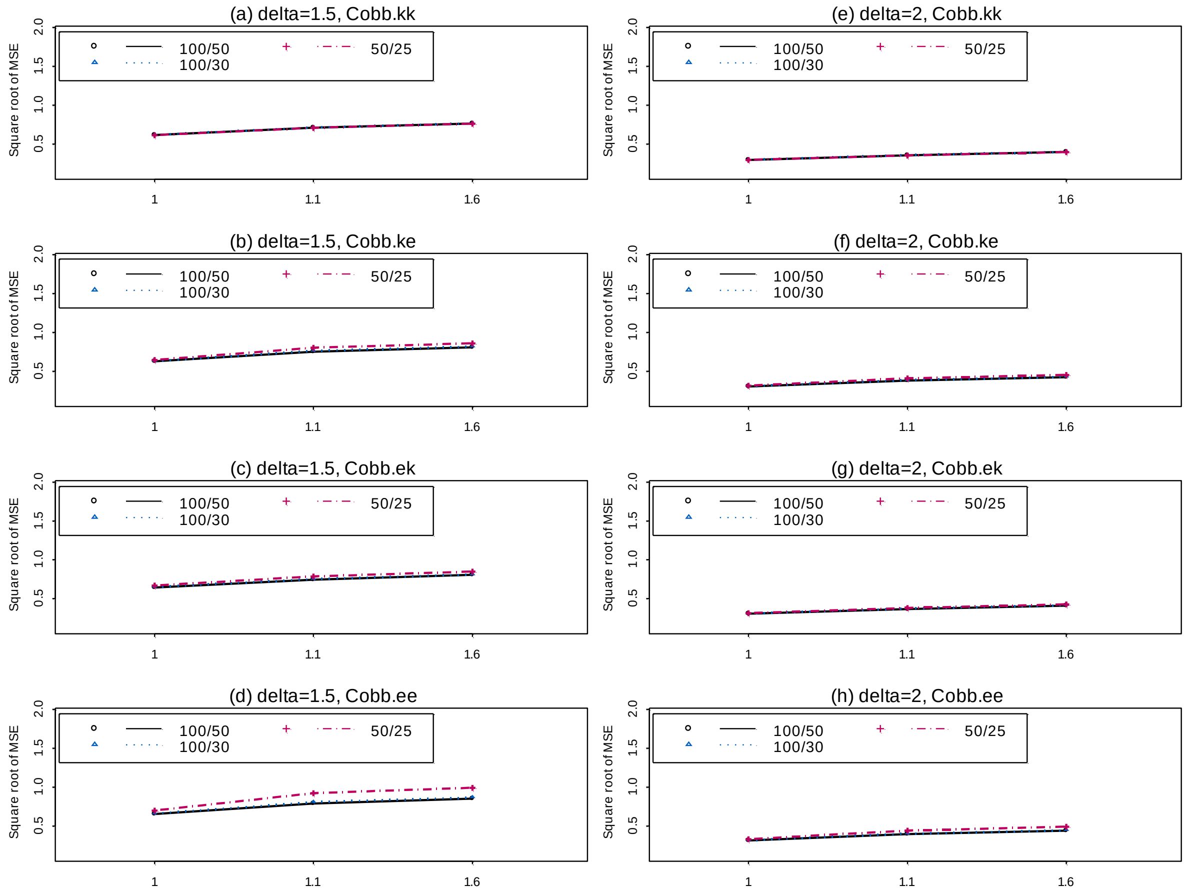

Figure 5.5 The effect of sample size and change-point position to the Cobb's estimation method for bivariate series.

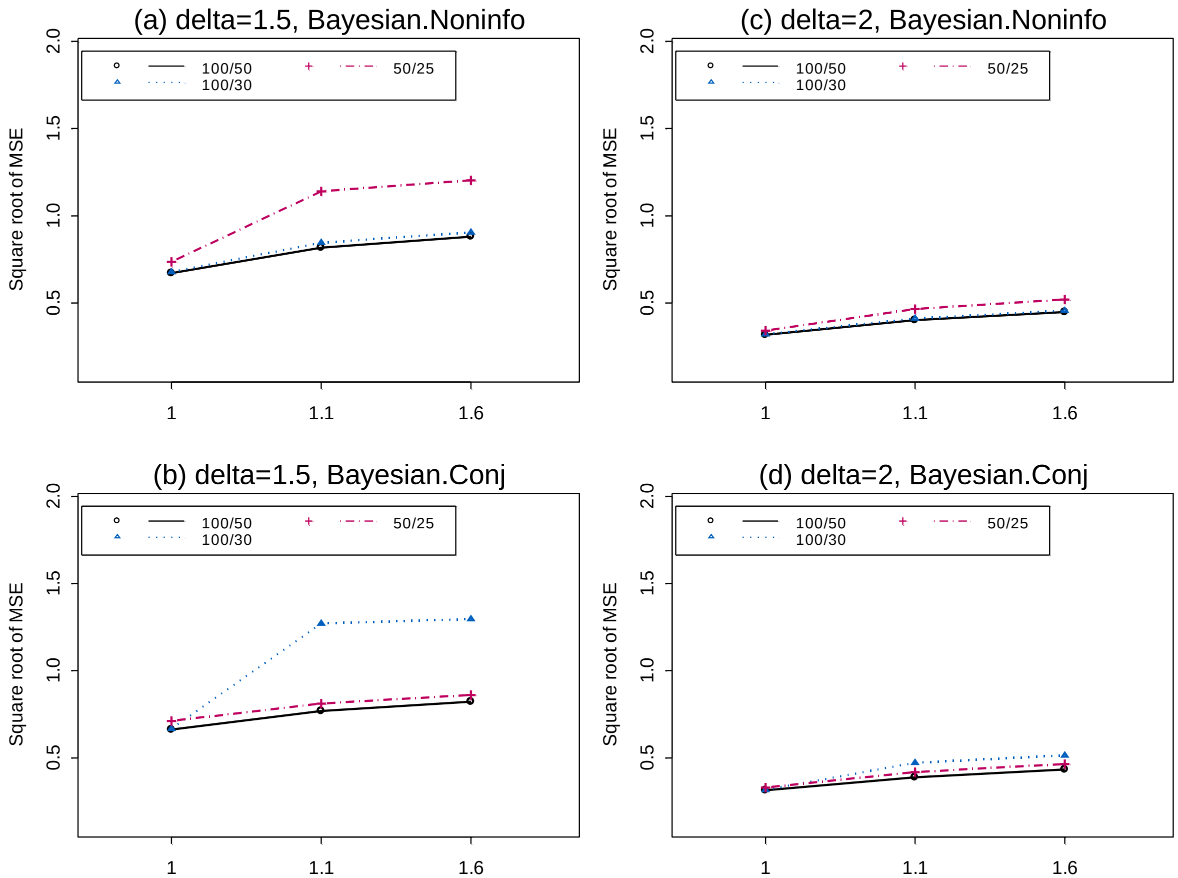

Figure 5.6 The effect of sample size and change-point position to the Bayesian's estimation method for bivariate series.

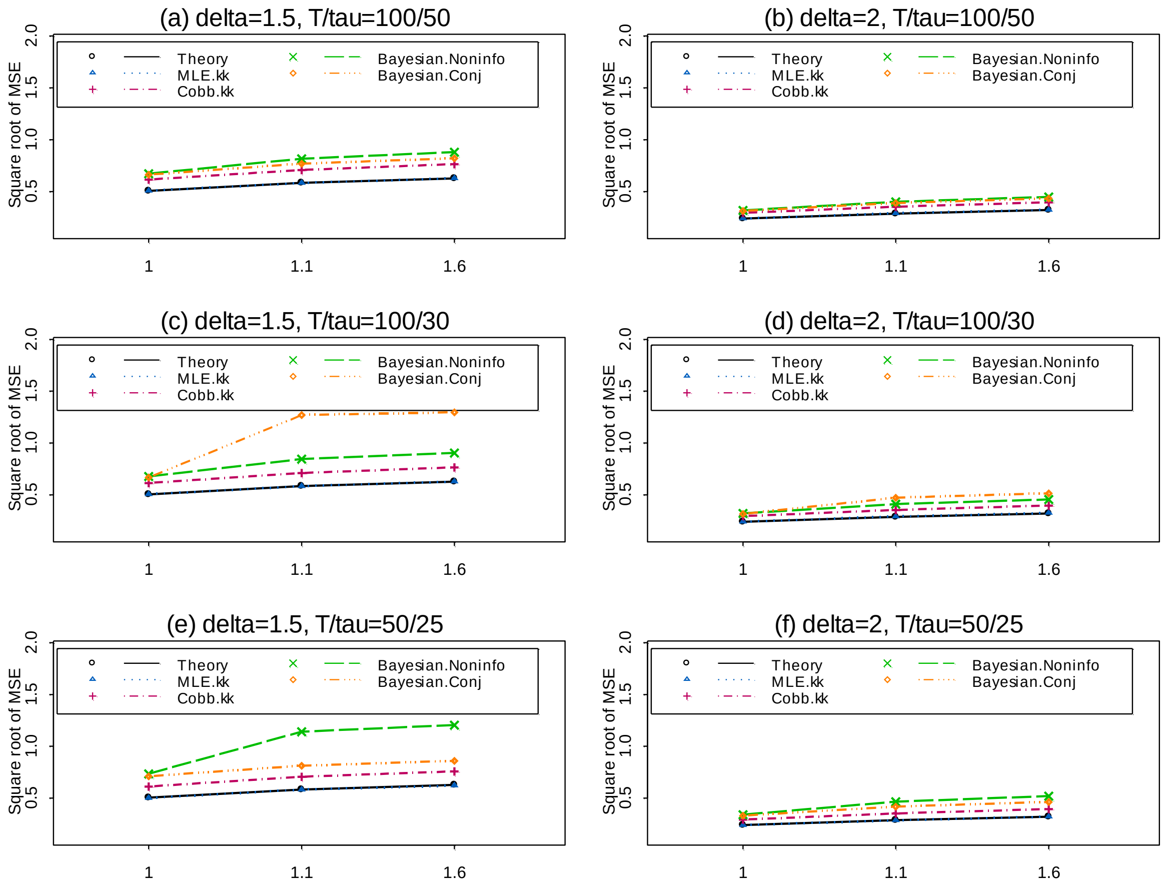

Figure 5.7 Comparison of estimation methods when the MLE and Cobb used ‘kk’ for parameter estimates for bivariate series.

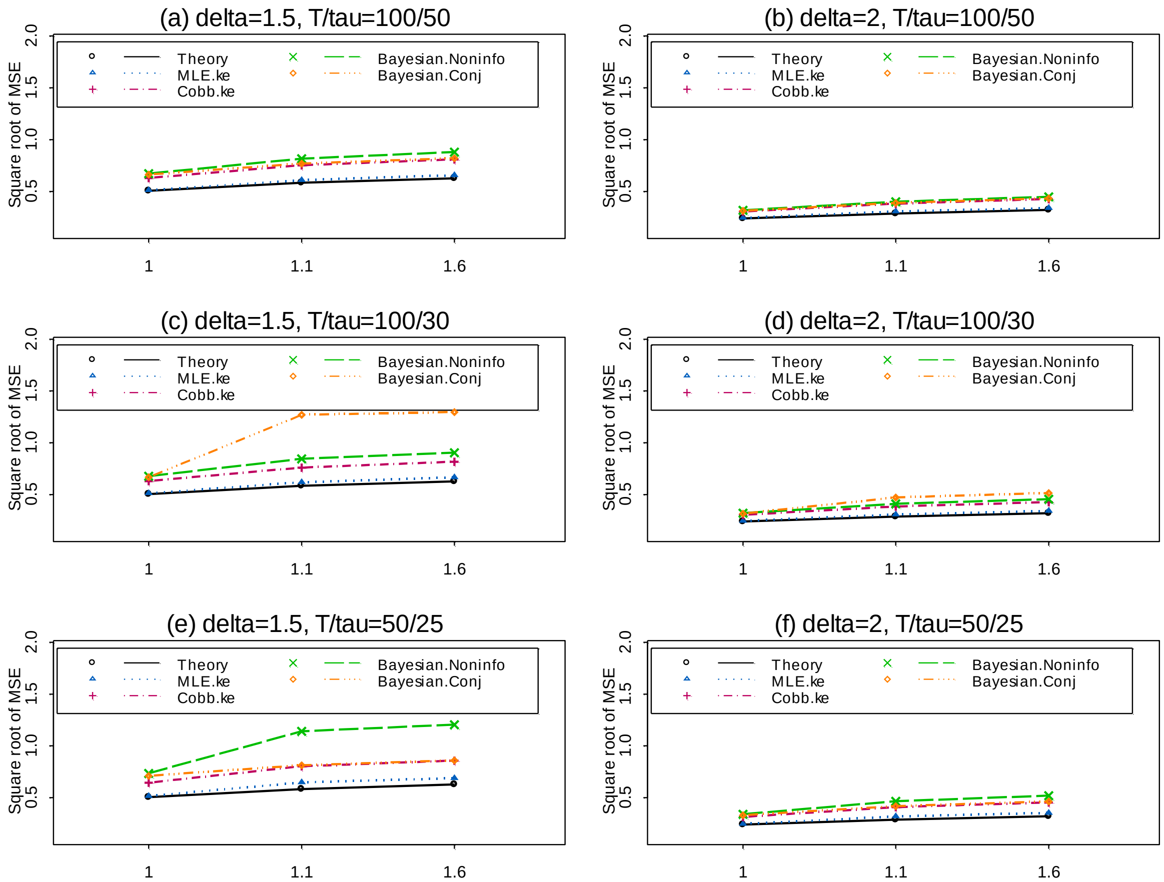

Figure 5.8 Comparison of estimation methods when the MLE and Cobb used ‘ke’ for parameter estimates for bivariate series.

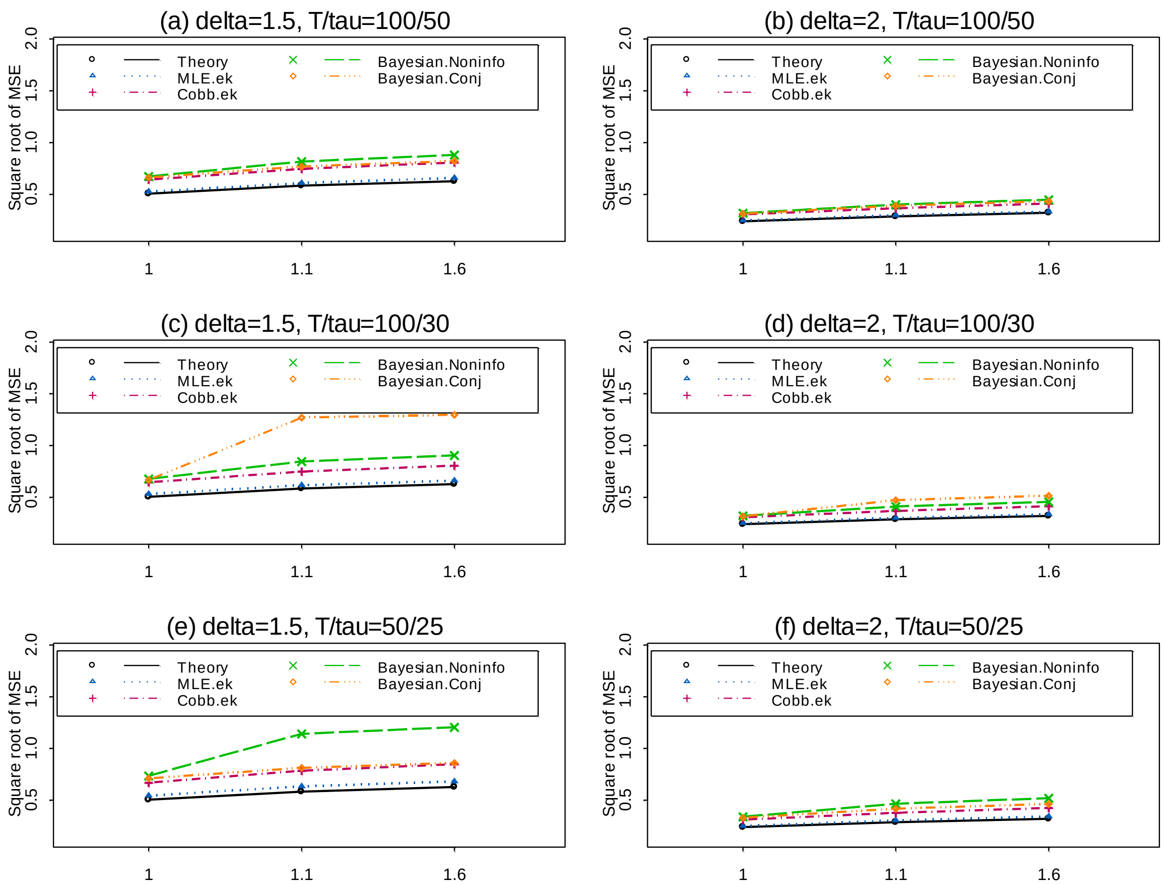

Figure 5.9 Comparison of estimation methods when the MLE and Cobb used ‘ek’ for parameter estimates for bivariate series.

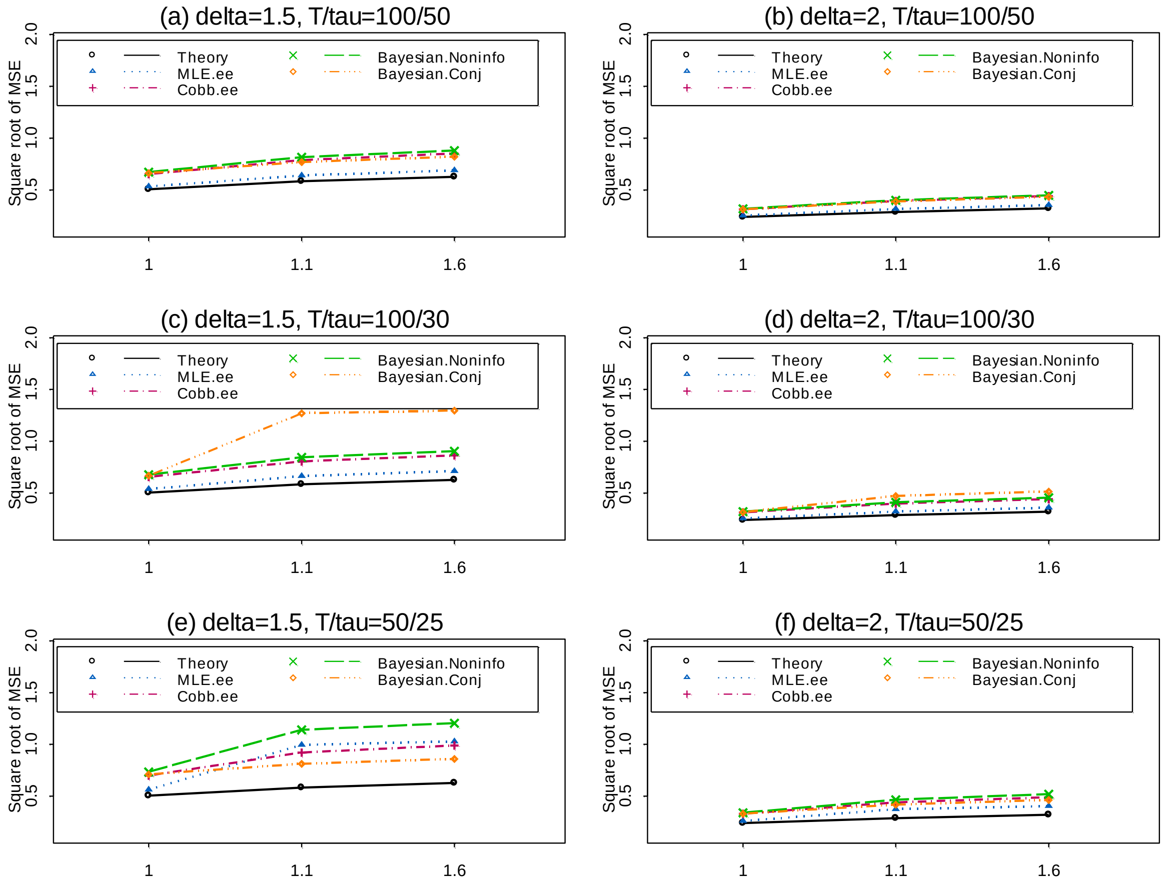

Figure 5.10 Comparison of estimation methods when the MLE and Cobb used ‘ee’ for parameter estimates for bivariate series.

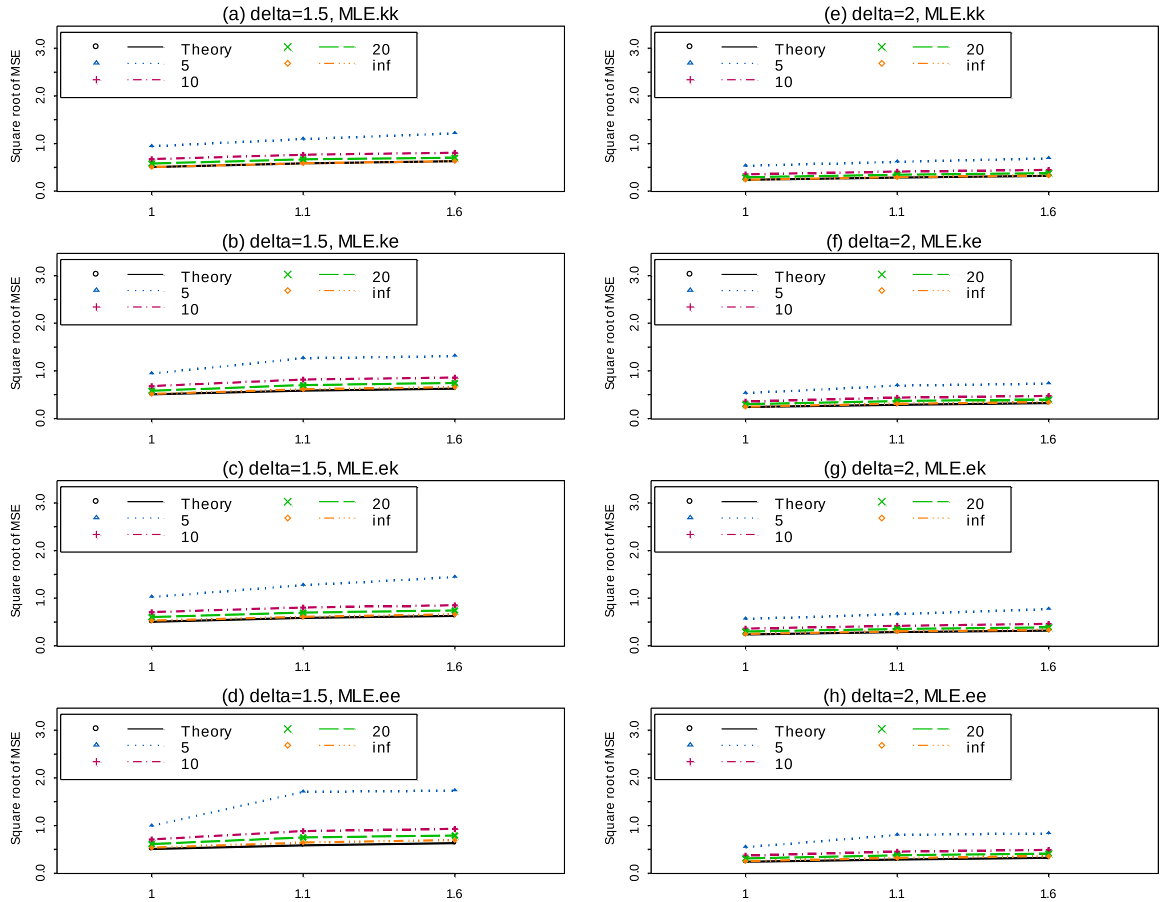

Figure 5.11 Effect of the degrees of freedom when the series follow multivariate t-distribution using MLE method for bivariate series.

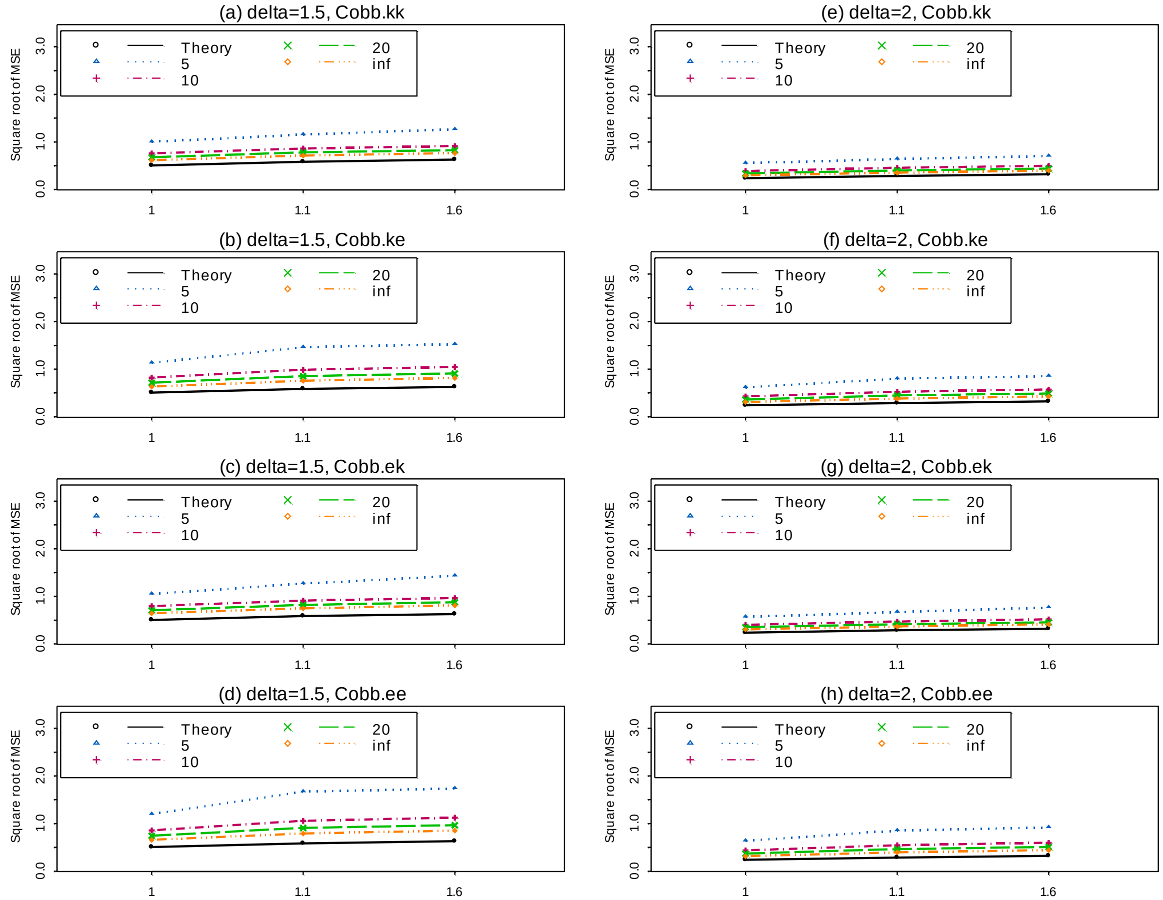

Figure 5.12 Effect of the degrees of freedom when the series follow multivariate t-distribution using Cobb's method for bivariate series.

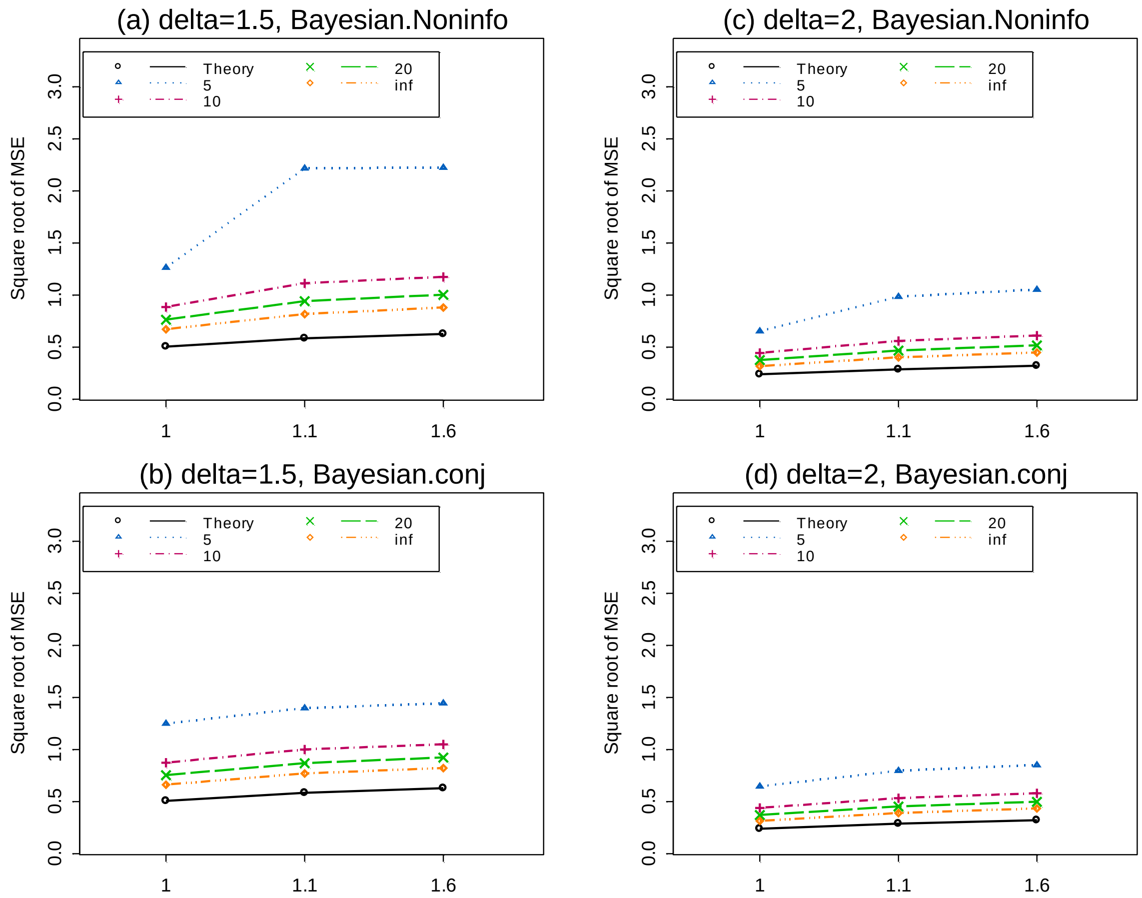

Figure 5.13 Effect of the degrees of freedom when the series follow multivariate t-distribution using Bayesian method for bivariate series.

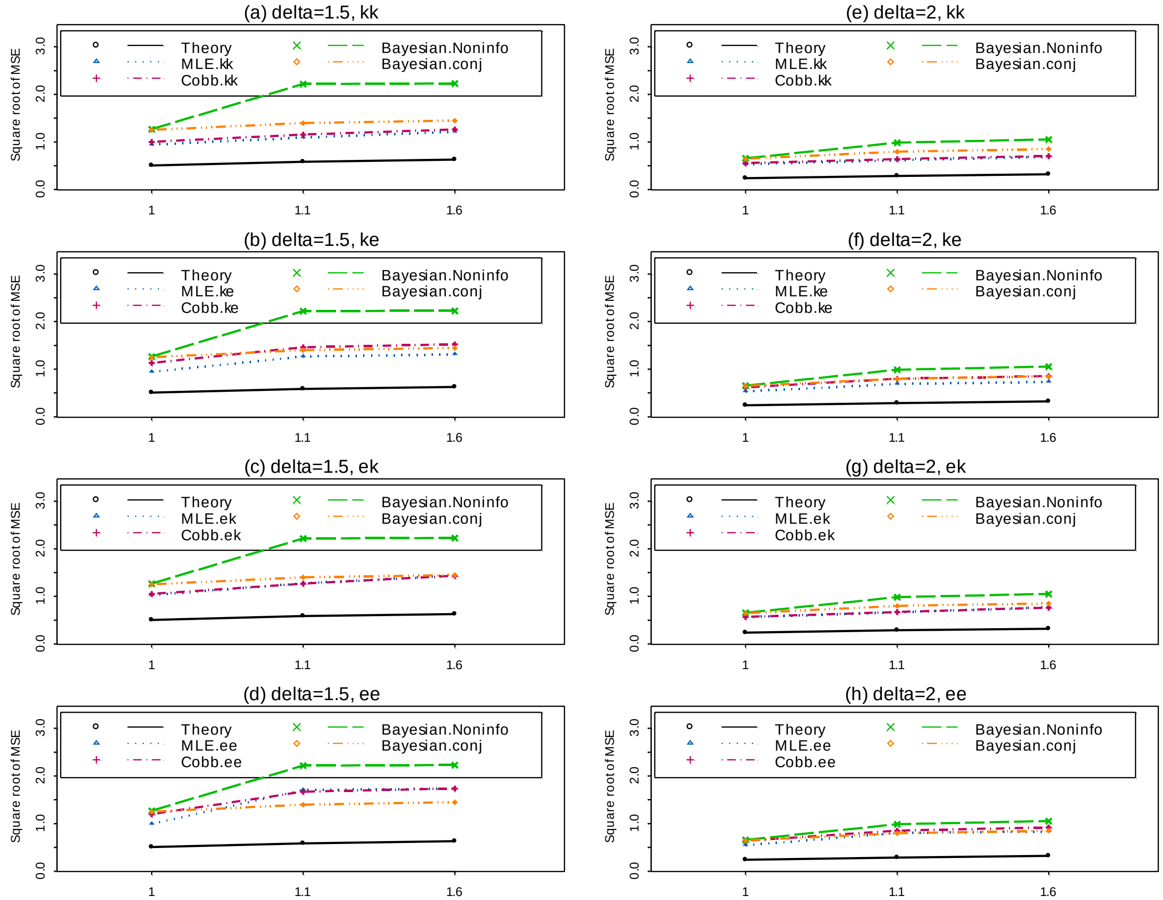

Figure 5.14 Comparison of estimation methods when the series follow multivariate t-distribution with df=5 for bivariate series.

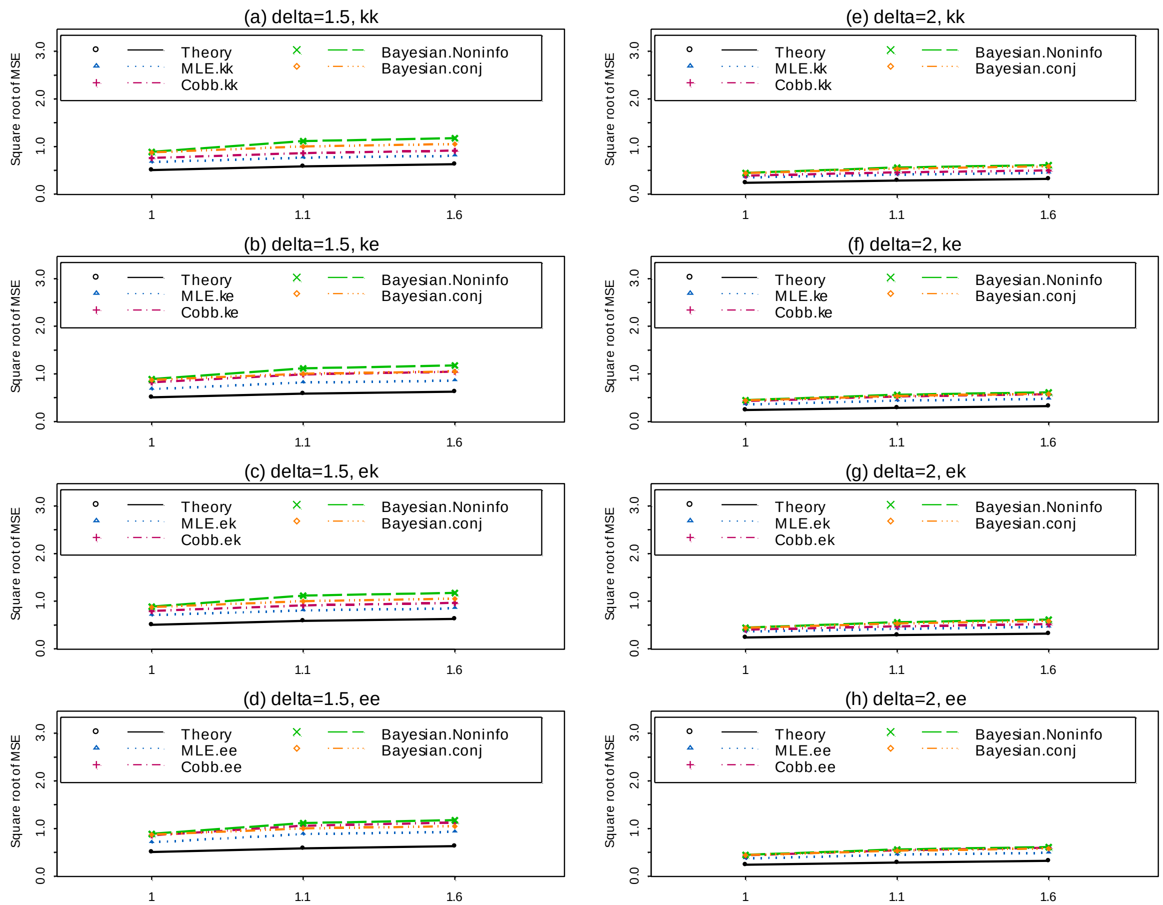

Figure 5.15 Comparison of estimation methods when the series follow multivariate t-distribution with df=10 for bivariate series.

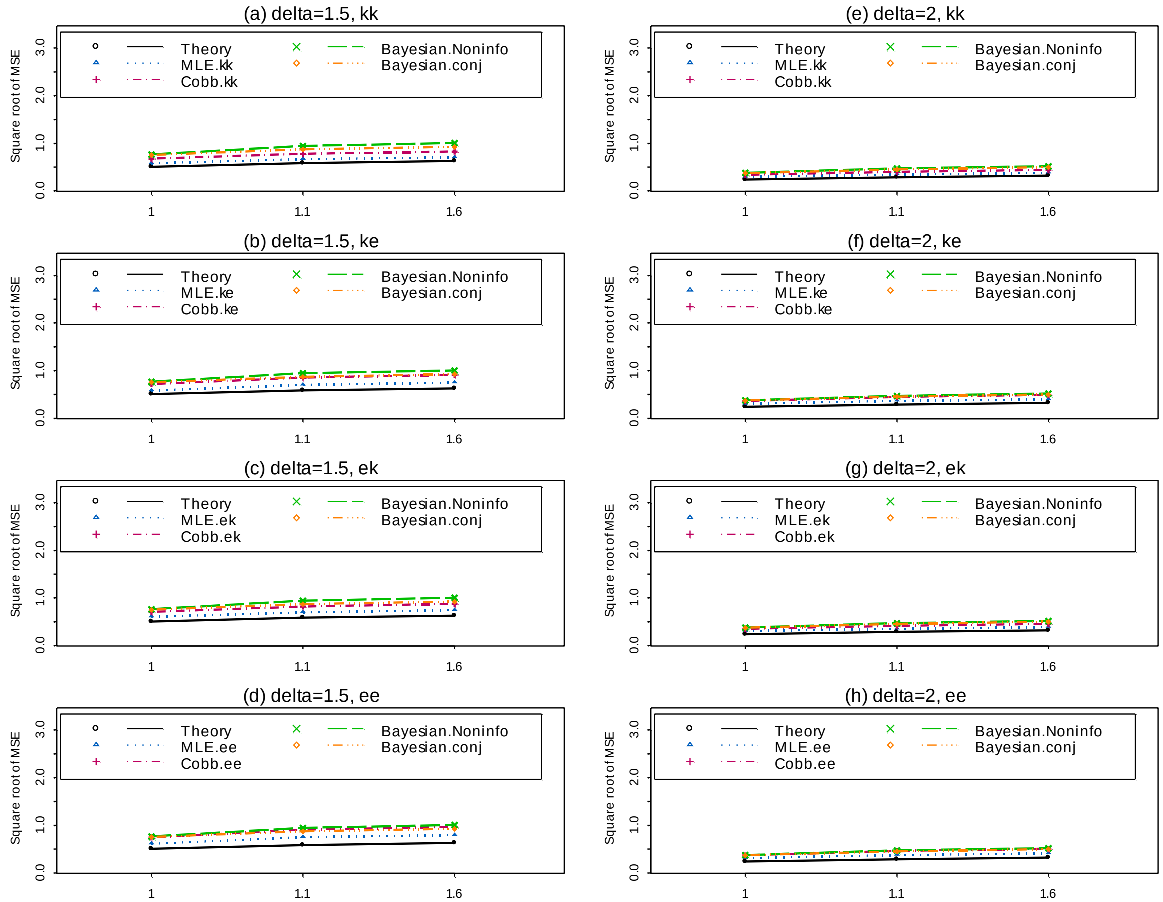

Figure 5.16 Comparison of estimation methods when the series follow multivariate t-distribution with df=20 for bivariate series.

5.3 Univariate Simulations

Under the univariate simulation, the setup followed the multivariate simulations. The same number of repetition, values for T/τ, δ, $\det(KK^T)$, and the departure from normality were used as in the multivariate case. As specified in Section 5.1, a sample of T observations with change-point τ would be generated 100,000 times. Before the change-point, the mean was $\mu_0 = 0$. After the change-point, $\delta = 1.5$ corresponded to $\mu_1 = 3$, and $\delta = 2$ corresponded to $\mu_1 = 4$. Before the change-point, the variance was $\Sigma_0 = 1$. After the change-point, $\det(KK^T) = 1.1$ corresponded to $\sigma_1^2 = 0.91$, and $\det(KK^T) = 1.6$ corresponded to $\sigma_1^2 = 0.64$. The square roots of the MSE were presented in Tables 5.13 – 5.18, and the biases of the change-point mle were presented in Tables 5.19 – 5.24. The same set of figures was produced as in Section 5.2.

Table 5.13 Square root of mean squared error of the change-point mle when $T/\tau = 100/50$ and $\delta = 1.5$ for univariate series.

K2	df	Theory	MLE				Cobb				Bayesian
			kk	ke	ek	ee	kk	ke	ek	ee	Non-Info.
1	5		0.9495	0.9495	0.9910	0.9910	1.0138	1.1571	1.0612	1.2250	1.2648
1	10		0.6583	0.6583	0.6855	0.6855	0.7519	0.8184	0.7850	0.8542	0.8663
1	20		0.5767	0.5767	0.5974	0.5974	0.6783	0.7134	0.7086	0.7426	0.7517
1	Inf	0.5057	0.5025	0.5025	0.5233	0.5233	0.6176	0.6292	0.6454	0.6538	0.6608
1.1	5		0.9019	1.0165	0.9469	1.1452	0.9615	1.2073	1.0127	1.3281	1.3787
1.1	10		0.6257	0.6637	0.6527	0.6936	0.7149	0.8173	0.7453	0.8562	0.8725
1.1	20		0.5473	0.5607	0.5681	0.5887	0.6433	0.7000	0.6707	0.7349	0.7469
1.1	Inf	0.4803	0.4758	0.4885	0.4922	0.5103	0.5844	0.6151	0.6101	0.6406	0.6505
1.6	5		0.8041	0.8616	0.9387	0.9554	0.8478	1.0205	0.9614	1.1181	1.1538
1.6	10		0.5243	0.5535	0.5429	0.5749	0.5934	0.6769	0.6175	0.7039	0.7177
1.6	20		0.4551	0.4644	0.4655	0.4769	0.5318	0.5766	0.5500	0.5967	0.6077
1.6	Inf	0.3874	0.3858	0.3983	0.3968	0.4118	0.4771	0.5039	0.4954	0.5222	0.5285

Table 5.14 Square root of mean squared error of the change-point mle when $T/\tau = 100/50$ and $\delta = 2$ for univariate series.

K2	df	Theory	MLE				Cobb				Bayesian
			kk	ke	ek	ee	kk	ke	ek	ee	Non-Info.
1	5		0.5241	0.5241	0.5402	0.5402	0.5501	0.6200	0.5708	0.6434	0.6576
1	10		0.3468	0.3468	0.3534	0.3534	0.3876	0.4198	0.3972	0.4296	0.4348
1	20		0.2941	0.2941	0.3006	0.3006	0.3401	0.3587	0.3489	0.3662	0.3700
1	Inf	0.2398	0.2408	0.2408	0.2459	0.2459	0.2959	0.3027	0.3047	0.3095	0.3124
1.1	5		0.5024	0.5725	0.5213	0.6089	0.5274	0.6623	0.5484	0.7040	0.7212
1.1	10		0.3299	0.3433	0.3341	0.3538	0.3668	0.4170	0.3750	0.4289	0.4363
1.1	20		0.2781	0.2868	0.2836	0.2925	0.3200	0.3517	0.3283	0.3597	0.3641
1.1	Inf	0.2222	0.2238	0.2324	0.2294	0.2376	0.2766	0.2956	0.2848	0.3027	0.3063
1.6	5		0.4389	0.4872	0.4739	0.5272	0.4584	0.5582	0.4862	0.5987	0.6055
1.6	10		0.2706	0.2828	0.2744	0.2895	0.2998	0.3393	0.3057	0.3472	0.3538
1.6	20		0.2204	0.2281	0.2249	0.2331	0.2545	0.2815	0.2609	0.2872	0.2905
1.6	Inf	0.1694	0.1746	0.1823	0.1798	0.1872	0.2150	0.2316	0.2214	0.2369	0.2395

Table 5.15 Square root of mean squared error of the change-point mle when $T/\tau = 100/30$ and $\delta = 1.5$ for univariate series.

K2	df	Theory	MLE				Cobb				Bayesian
			kk	ke	ek	ee	kk	ke	ek	ee	Non-Info.
1	5		0.9594	0.9594	1.0564	1.0564	1.0227	1.1568	1.0825	1.2500	1.3331
1	10		0.6638	0.6638	0.7028	0.7028	0.7549	0.8213	0.8019	0.8707	0.8843
1	20		0.5775	0.5775	0.6061	0.6061	0.6788	0.7155	0.7191	0.7548	0.7644
1	Inf	0.5057	0.5031	0.5031	0.5219	0.5219	0.6170	0.6284	0.6494	0.6569	0.6641
1.1	5		0.9112	1.1456	1.0415	1.4881	0.9709	1.3273	1.0612	1.6307	1.8291
1.1	10		0.6355	0.6739	0.6699	0.7424	0.7187	0.8288	0.7630	0.9035	0.9318
1.1	20		0.5500	0.5749	0.5745	0.5965	0.6441	0.7127	0.6817	0.7489	0.7695
1.1	Inf	0.4803	0.4740	0.4892	0.4928	0.5111	0.5837	0.6205	0.6141	0.6508	0.6627
1.6	5		0.8480	1.0073	1.2564	1.3167	0.8917	1.1533	1.1696	1.4506	1.5377
1.6	10		0.5311	0.5571	0.5621	0.6098	0.5982	0.6862	0.6360	0.7449	0.7684
1.6	20		0.4533	0.4681	0.4702	0.4909	0.5308	0.5863	0.5588	0.6164	0.6299
1.6	Inf	0.3874	0.3861	0.3982	0.3968	0.4138	0.4759	0.5085	0.4989	0.5303	0.5392

Table 5.16 Square root of mean squared error of the change-point mle when $T/\tau = 100/30$ and $\delta = 2$ for univariate series.

K2	df	Theory	MLE				Cobb				Bayesian
			kk	ke	ek	ee	kk	ke	ek	ee	Non-Info.
1	5		0.5326	0.5326	0.5452	0.5452	0.5583	0.6235	0.5748	0.6492	0.6655
1	10		0.3530	0.3530	0.3633	0.3633	0.3927	0.4244	0.4052	0.4371	0.4428
1	20		0.2895	0.2895	0.2985	0.2985	0.3378	0.3568	0.3500	0.3680	0.3719
1	Inf	0.2398	0.2402	0.2402	0.2451	0.2451	0.2942	0.3010	0.3048	0.3098	0.3127
1.1	5		0.5084	0.6633	0.5287	0.7997	0.5320	0.7480	0.5560	0.8856	0.9292
1.1	10		0.3343	0.3540	0.3420	0.3703	0.3710	0.4282	0.3828	0.4459	0.4584
1.1	20		0.2735	0.2877	0.2807	0.2940	0.3179	0.3560	0.3291	0.3687	0.3755
1.1	Inf	0.2222	0.2249	0.2343	0.2286	0.2416	0.2751	0.2986	0.2850	0.3086	0.3133
1.6	5		0.4642	0.5886	0.5973	0.7294	0.4819	0.6546	0.6156	0.7956	0.8164
1.6	10		0.2750	0.2900	0.2822	0.3018	0.3023	0.3502	0.3124	0.3639	0.3738
1.6	20		0.2191	0.2310	0.2258	0.2395	0.2534	0.2853	0.2626	0.2945	0.3006
1.6	Inf	0.1694	0.1737	0.1810	0.1782	0.1867	0.2142	0.2351	0.2220	0.2420	0.2461

Table 5.17 Square root of mean squared error of the change-point mle when $T/\tau = 50/25$ and $\delta = 1.5$ for univariate series.

K2	df	Theory	MLE				Cobb				Bayesian
			kk	ke	ek	ee	kk	ke	ek	ee	Non-Info.
1	5		0.9553	0.9553	1.0679	1.0679	1.0160	1.1635	1.1197	1.3119	1.4652
1	10		0.6654	0.6654	0.7165	0.7165	0.7556	0.8385	0.8280	0.9205	0.9604
1	20		0.5809	0.5809	0.6208	0.6208	0.6821	0.7328	0.7422	0.7917	0.8180
1	Inf	0.5057	0.5030	0.5030	0.5360	0.5360	0.6171	0.6411	0.6702	0.6885	0.7092
1.1	5		0.9064	1.0459	1.0222	1.3418	0.9640	1.2592	1.0646	1.4846	1.6589
1.1	10		0.6330	0.6876	0.6822	0.7826	0.7177	0.8607	0.7847	0.9654	1.0296
1.1	20		0.5510	0.5852	0.5884	0.6588	0.6470	0.7418	0.7017	0.8126	0.8538
1.1	Inf	0.4803	0.4749	0.4998	0.5057	0.5446	0.5842	0.6434	0.6326	0.7018	0.7324
1.6	5		0.8246	0.8812	0.9862	1.0931	0.8545	1.0574	0.9719	1.2221	1.3357
1.6	10		0.5261	0.5692	0.5639	0.6287	0.5962	0.7134	0.6464	0.7839	0.8276
1.6	20		0.4508	0.4764	0.4779	0.5256	0.5307	0.6099	0.5719	0.6589	0.6870
1.6	Inf	0.3874	0.3865	0.4064	0.4087	0.4333	0.4765	0.5277	0.5113	0.5658	0.5882

Table 5.18 Square root of mean squared error of the change-point mle when $T/\tau = 50/25$ and $\delta = 2$ for univariate series.

K2	df	Theory	MLE				Cobb				Bayesian
			kk	ke	ek	ee	kk	ke	ek	ee	Non-Info.
1	5		0.5263	0.5263	0.5478	0.5478	0.5532	0.6352	0.5765	0.6788	0.7264
1	10		0.3528	0.3528	0.3690	0.3690	0.3929	0.4345	0.4144	0.4555	0.4682
1	20		0.2931	0.2931	0.3070	0.3070	0.3407	0.3672	0.3589	0.3838	0.3926
1	Inf	0.2398	0.2401	0.2401	0.2505	0.2505	0.2954	0.3097	0.3132	0.3238	0.3310
1.1	5		0.5053	0.5777	0.5303	0.6731	0.5291	0.6897	0.5550	0.7772	0.8297
1.1	10		0.3344	0.3680	0.3450	0.3927	0.3705	0.4536	0.3890	0.4819	0.5008
1.1	20		0.2766	0.2985	0.2885	0.3118	0.3203	0.3759	0.3376	0.3950	0.4085
1.1	Inf	0.2222	0.2247	0.2385	0.2340	0.2552	0.2764	0.3141	0.2926	0.3301	0.3406
1.6	5		0.4451	0.4834	0.4950	0.5539	0.4574	0.5724	0.4961	0.6397	0.6771
1.6	10		0.2752	0.2984	0.2822	0.3167	0.3038	0.3668	0.3153	0.3884	0.4027
1.6	20		0.2217	0.2389	0.2292	0.2526	0.2556	0.3029	0.2682	0.3164	0.3264
1.6	Inf	0.1694	0.1758	0.1867	0.1821	0.1963	0.2154	0.2477	0.2273	0.2598	0.2674

Table 5.19 Bias of the change-point mle when $T/\tau = 100/50$ and $\delta = 1.5$ for univariate series.

K2	df	Theory	MLE				Cobb				Bayesian
			kk	ke	ek	ee	kk	ke	ek	ee	Non-Info.
1	5		-0.0034	-0.0034	-0.0038	-0.0038	-0.0038	-0.0033	-0.0059	-0.0060	-0.0043
1	10		-0.0012	-0.0012	-0.0001	-0.0001	-0.0001	0.0001	0.0005	0.0008	0.0008
1	20		0.0001	0.0001	0.0017	0.0017	0.0015	0.0016	0.0019	0.0019	0.0019
1	Inf	0.0000	0.0028	0.0028	0.0037	0.0037	0.0002	0.0002	0.0004	0.0006	0.0006
1.1	5		0.0058	-0.0001	0.0087	0.0027	0.0089	0.0093	0.0113	0.0104	0.0139
1.1	10		-0.0048	-0.0051	-0.0021	-0.0052	0.0016	0.0022	0.0034	0.0034	0.0041
1.1	20		-0.0035	-0.0048	-0.0022	-0.0041	0.0020	0.0021	0.0029	0.0031	0.0032
1.1	Inf	-0.0049	-0.0025	-0.0028	-0.0024	-0.0022	0.0003	0.0007	0.0009	0.0016	0.0017
1.6	5		0.0232	-0.0008	0.0370	0.0059	0.0373	0.0368	0.0504	0.0468	0.0504
1.6	10		-0.0109	-0.0184	-0.0096	-0.0152	0.0057	0.0082	0.0090	0.0128	0.0134
1.6	20		-0.0139	-0.0174	-0.0135	-0.0172	0.0024	0.0036	0.0045	0.0058	0.0064
1.6	Inf	-0.0161	-0.0145	-0.0156	-0.0153	-0.0155	0.0005	0.0012	0.0020	0.0027	0.0032

Table 5.20 Bias of the change-point mle when $T/\tau = 100/50$ and $\delta = 2$ for univariate series.

K2	df	Theory	MLE				Cobb				Bayesian
			kk	ke	ek	ee	kk	ke	ek	ee	Non-Info.
1	5		-0.0001	-0.0001	-0.0003	-0.0003	-0.0009	-0.0005	-0.0009	-0.0006	-0.0009
1	10		0.0013	0.0013	0.0013	0.0013	0.0002	0.0002	0.0003	0.0003	0.0003
1	20		-0.0002	-0.0002	0.0000	0.0000	0.0000	0.0001	-0.0001	0.0000	0.0000
1	Inf	0.0000	-0.0005	-0.0005	-0.0001	-0.0001	0.0004	0.0004	0.0003	0.0003	0.0003
1.1	5		0.0017	0.0004	0.0027	0.0006	0.0025	0.0039	0.0032	0.0053	0.0060
1.1	10		0.0005	-0.0014	0.0004	-0.0002	0.0007	0.0006	0.0009	0.0017	0.0016
1.1	20		-0.0015	-0.0019	-0.0013	-0.0013	0.0000	0.0004	0.0000	0.0005	0.0006
1.1	Inf	-0.0013	-0.0017	-0.0010	-0.0013	-0.0010	0.0003	0.0008	0.0004	0.0007	0.0008
1.6	5		0.0069	0.0017	0.0103	0.0037	0.0101	0.0148	0.0134	0.0185	0.0194
1.6	10		-0.0020	-0.0044	-0.0020	-0.0039	0.0011	0.0025	0.0018	0.0034	0.0041
1.6	20		-0.0042	-0.0050	-0.0040	-0.0046	-0.0003	0.0008	0.0002	0.0013	0.0015
1.6	Inf	-0.0036	-0.0038	-0.0038	-0.0034	-0.0033	0.0001	0.0008	0.0005	0.0011	0.0013

Table 5.21 Bias of the change-point mle when $T/\tau = 100/30$ and $\delta = 1.5$ for univariate series.

K2	df	Theory	MLE				Cobb				Bayesian
			kk	ke	ek	ee	kk	ke	ek	ee	Non-Info.
1	5		0.0020	0.0020	0.0162	0.0162	0.0017	-0.0016	0.0182	0.0267	0.0282
1	10		-0.0010	-0.0010	0.0086	0.0086	-0.0020	-0.0023	0.0102	0.0124	0.0098
1	20		-0.0044	-0.0044	0.0029	0.0029	-0.0034	-0.0035	0.0075	0.0085	0.0060
1	Inf	0.0000	0.0001	0.0001	0.0070	0.0070	-0.0003	-0.0003	0.0096	0.0097	0.0078
1.1	5		0.0102	0.0301	0.0294	0.0697	0.0143	0.0504	0.0355	0.0990	0.1179
1.1	10		-0.0034	-0.0064	0.0057	0.0038	0.0002	0.0089	0.0124	0.0224	0.0258
1.1	20		-0.0080	-0.0068	-0.0010	-0.0026	-0.0028	0.0048	0.0079	0.0130	0.0150
1.1	Inf	-0.0049	-0.0051	-0.0042	0.0020	-0.0007	-0.0001	0.0062	0.0094	0.0130	0.0141
1.6	5		0.0294	0.0246	0.0690	0.0632	0.0436	0.0719	0.0813	0.1218	0.1328
1.6	10		-0.0108	-0.0173	-0.0033	-0.0104	0.0049	0.0141	0.0161	0.0262	0.0301
1.6	20		-0.0171	-0.0201	-0.0126	-0.0175	-0.0014	0.0063	0.0076	0.0135	0.0157
1.6	Inf	-0.0161	-0.0154	-0.0158	-0.0111	-0.0120	0.0004	0.0062	0.0083	0.0119	0.0133

Table 5.22 Bias of the change-point mle when $T/\tau = 100/30$ and $\delta = 2$ for univariate series.

K2	df	Theory	MLE				Cobb				Bayesian
			kk	ke	ek	ee	kk	ke	ek	ee	Non-Info.
1	5		0.0022	0.0022	0.0081	0.0081	0.0017	0.0012	0.0084	0.0104	0.0099
1	10		-0.0005	-0.0005	0.0022	0.0022	-0.0005	-0.0007	0.0029	0.0036	0.0029
1	20		-0.0026	-0.0026	-0.0007	-0.0007	-0.0021	-0.0020	0.0007	0.0011	0.0006
1	Inf	0.0000	0.0005	0.0005	0.0011	0.0011	0.0003	0.0003	0.0027	0.0028	0.0024
1.1	5		0.0046	0.0182	0.0109	0.0301	0.0054	0.0280	0.0125	0.0429	0.0499
1.1	10		-0.0021	-0.0011	0.0009	0.0012	-0.0002	0.0049	0.0032	0.0078	0.0099
1.1	20		-0.0034	-0.0028	-0.0015	-0.0018	-0.0020	0.0024	0.0008	0.0041	0.0055
1.1	Inf	-0.0013	-0.0008	-0.0008	0.0000	0.0000	0.0003	0.0037	0.0026	0.0051	0.0061
1.6	5		0.0109	0.0185	0.0207	0.0315	0.0140	0.0358	0.0252	0.0510	0.0563
1.6	10		-0.0039	-0.0040	-0.0017	-0.0026	-0.0001	0.0058	0.0030	0.0083	0.0104
1.6	20		-0.0053	-0.0056	-0.0040	-0.0047	-0.0015	0.0026	0.0007	0.0042	0.0056
1.6	Inf	-0.0036	-0.0033	-0.0029	-0.0026	-0.0024	0.0002	0.0035	0.0020	0.0046	0.0056

Table 5.23 Bias of the change-point mle when $T/\tau = 50/25$ and $\delta = 1.5$ for univariate series.

K2	df	Theory	MLE				Cobb				Bayesian
			kk	ke	ek	ee	kk	ke	ek	ee	Non-Info.
1	5		-0.0055	-0.0055	-0.0031	-0.0031	-0.0065	-0.0309	-0.0072	-0.0318	-0.0042
1	10		-0.0008	-0.0008	-0.0008	-0.0008	-0.0006	-0.0006	-0.0012	-0.0013	-0.0010
1	20		-0.0030	-0.0030	-0.0015	-0.0015	-0.0017	-0.0015	-0.0007	-0.0005	-0.0006
1	Inf	0.0000	0.0023	0.0023	0.0018	0.0018	0.0020	0.0020	0.0016	0.0018	0.0018
1.1	5		0.0031	-0.0069	0.0114	-0.0028	0.0064	0.0010	0.0114	-0.0094	0.0076
1.1	10		-0.0041	-0.0066	-0.0040	-0.0062	0.0010	0.0010	0.0023	0.0018	0.0032
1.1	20		-0.0078	-0.0079	-0.0048	-0.0066	-0.0014	-0.0010	0.0009	0.0001	0.0014
1.1	Inf	-0.0049	-0.0029	-0.0039	-0.0038	-0.0037	0.0019	0.0015	0.0025	0.0030	0.0030
1.6	5		0.0247	-0.0131	0.0469	0.0051	0.0370	0.0244	0.0507	0.0373	0.0477
1.6	10		-0.0122	-0.0208	-0.0102	-0.0186	0.0043	0.0062	0.0102	0.0125	0.0150
1.6	20		-0.0161	-0.0210	-0.0149	-0.0173	-0.0004	0.0014	0.0046	0.0062	0.0078
1.6	Inf	-0.0161	-0.0159	-0.0171	-0.0160	-0.0165	0.0014	0.0019	0.0044	0.0052	0.0063

Table 5.24 Bias of the change-point mle when $T/\tau = 50/25$ and $\delta = 2$ for univariate series.

K2	df	Theory	MLE				Cobb				Bayesian
			kk	ke	ek	ee	kk	ke	ek	ee	Non-Info.
1	5		-0.0012	-0.0012	-0.0002	-0.0002	-0.0011	-0.0035	-0.0007	-0.0047	-0.0017
1	10		0.0005	0.0005	-0.0008	-0.0008	-0.0002	-0.0002	-0.0011	-0.0011	-0.0011
1	20		-0.0010	-0.0010	-0.0007	-0.0007	-0.0009	-0.0010	-0.0006	-0.0007	-0.0008
1	Inf	0.0000	-0.0001	-0.0001	0.0006	0.0006	0.0005	0.0006	0.0006	0.0005	0.0005
1.1	5		0.0017	-0.0024	0.0033	-0.0006	0.0029	0.0006	0.0042	0.0019	0.0042
1.1	10		-0.0005	-0.0019	-0.0014	-0.0022	0.0003	0.0005	-0.0001	0.0004	0.0009
1.1	20		-0.0023	-0.0029	-0.0020	-0.0024	-0.0009	-0.0008	-0.0004	0.0000	0.0002
1.1	Inf	-0.0013	-0.0015	-0.0010	-0.0009	-0.0002	0.0005	0.0009	0.0007	0.0012	0.0012
1.6	5		0.0079	-0.0020	0.0123	0.0027	0.0106	0.0112	0.0142	0.0176	0.0196
1.6	10		-0.0031	-0.0057	-0.0027	-0.0043	0.0006	0.0028	0.0015	0.0046	0.0053
1.6	20		-0.0043	-0.0057	-0.0046	-0.0057	-0.0009	0.0006	-0.0001	0.0017	0.0023
1.6	Inf	-0.0036	-0.0036	-0.0040	-0.0038	-0.0040	0.0002	0.0015	0.0010	0.0022	0.0026

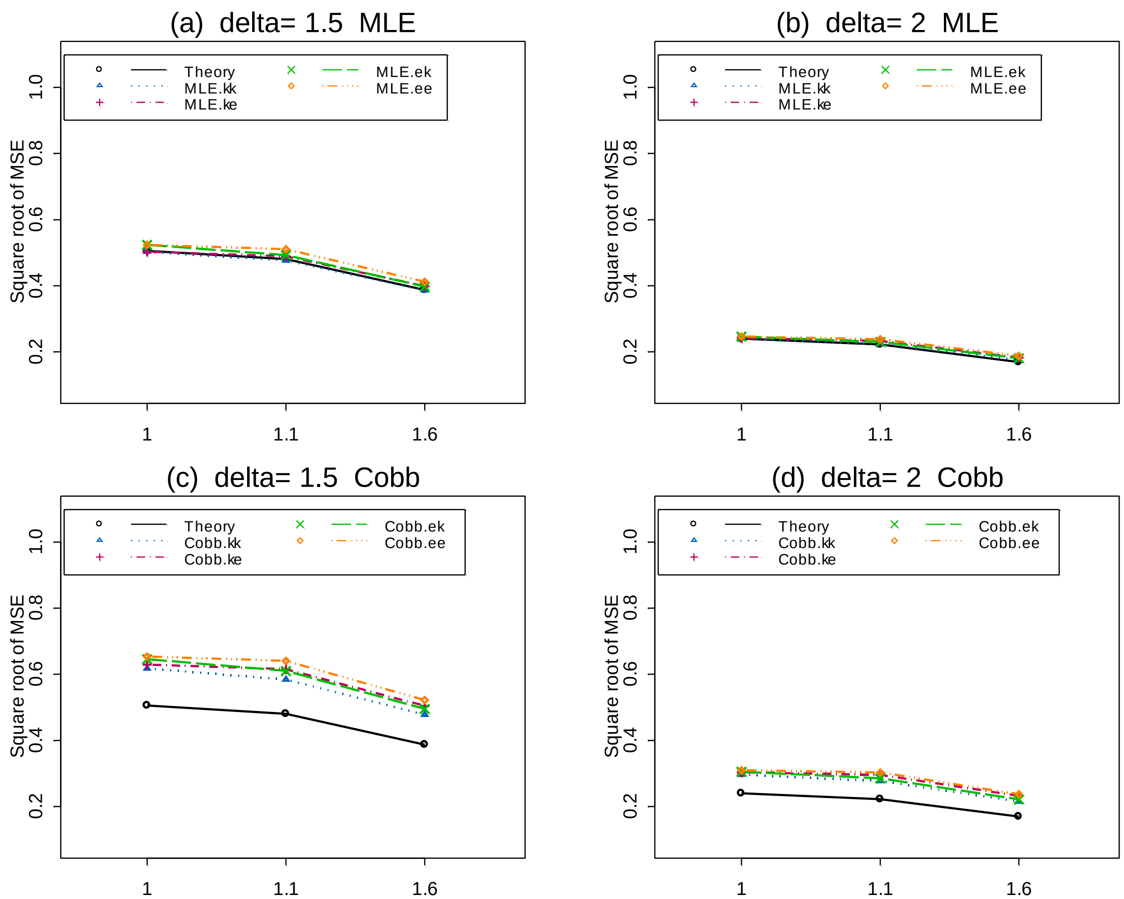

Figure 5.17 Comparison of the kk, ke, ek, and ee estimation method for MLE and Cobb's method when $T/\tau = 100/50$ for univariate series.

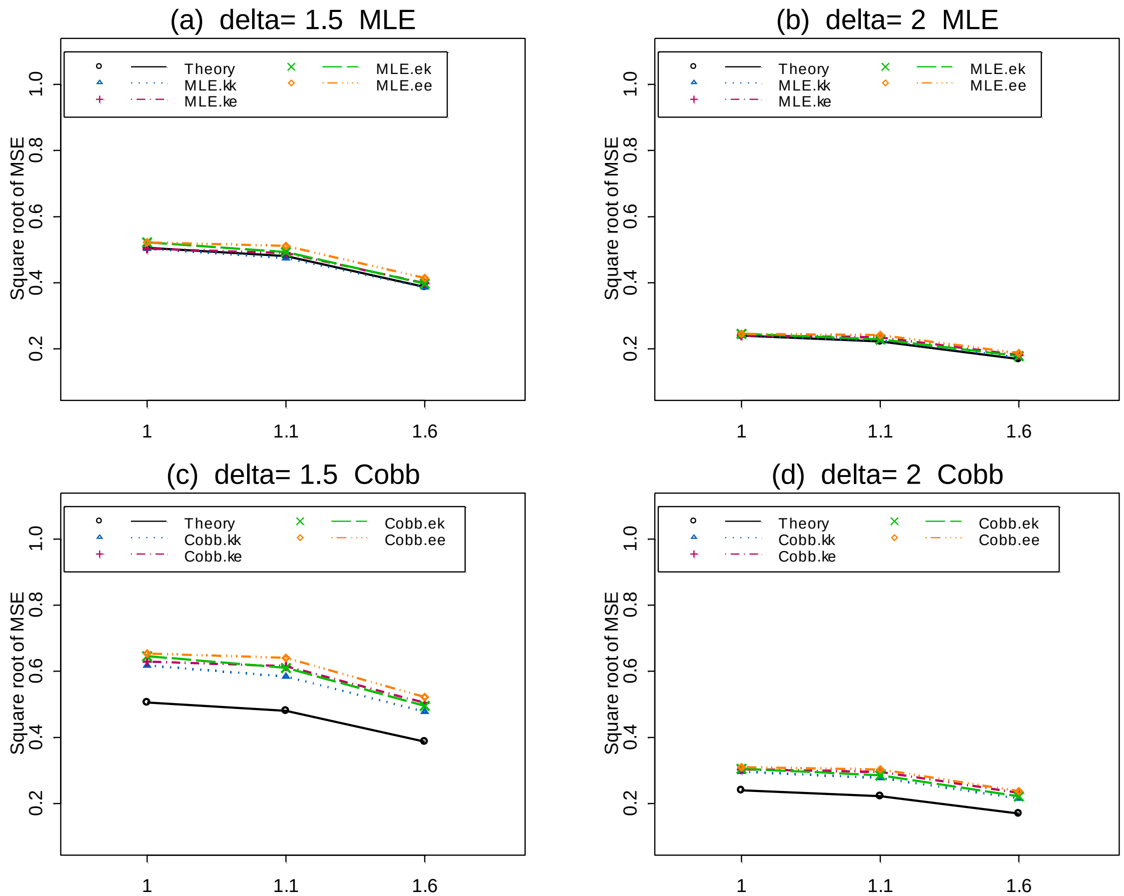

Figure 5.18 Comparison of the kk, ke, ek, and ee estimation method for MLE and Cobb's method when $T/\tau = 100/30$ for univariate series.

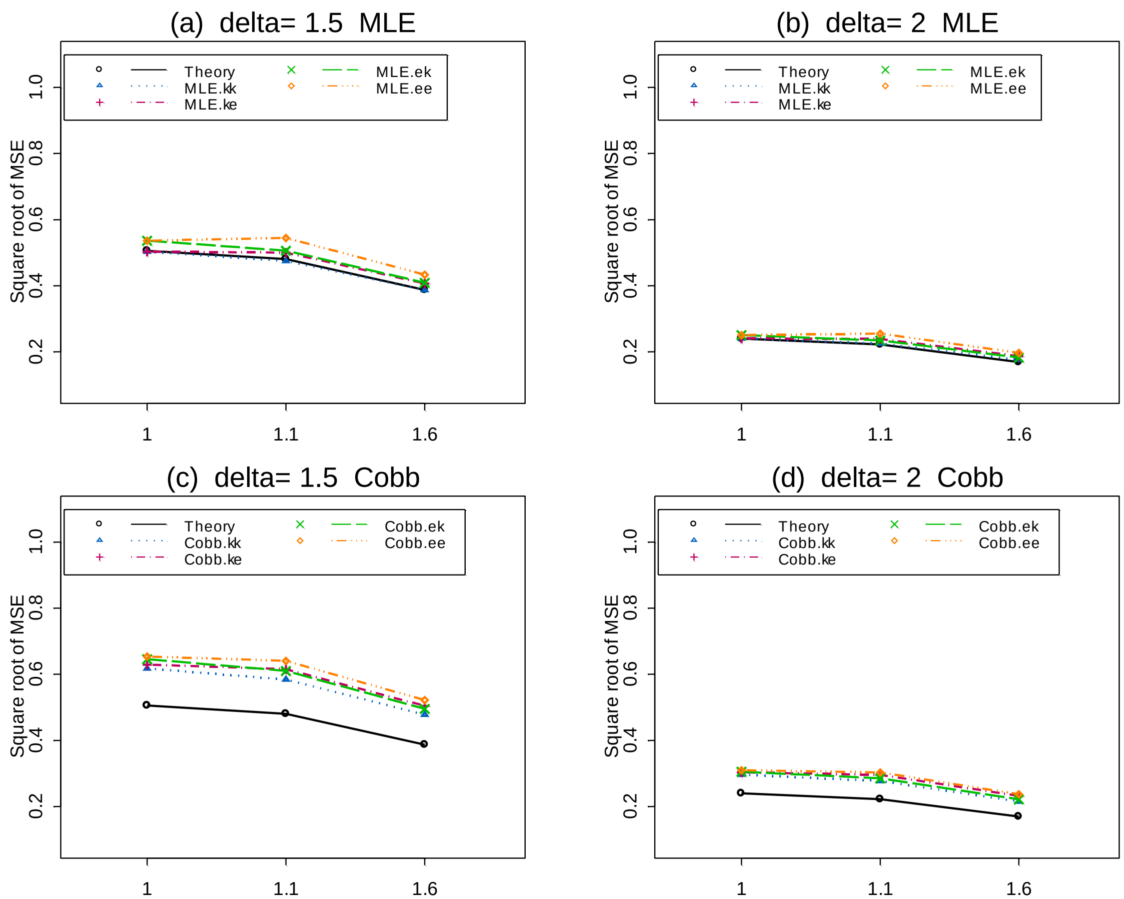

Figure 5.19 Comparison of the kk, ke, ek, and ee estimation method for MLE and Cobb's method when $T/\tau = 50/25$ for univariate series.

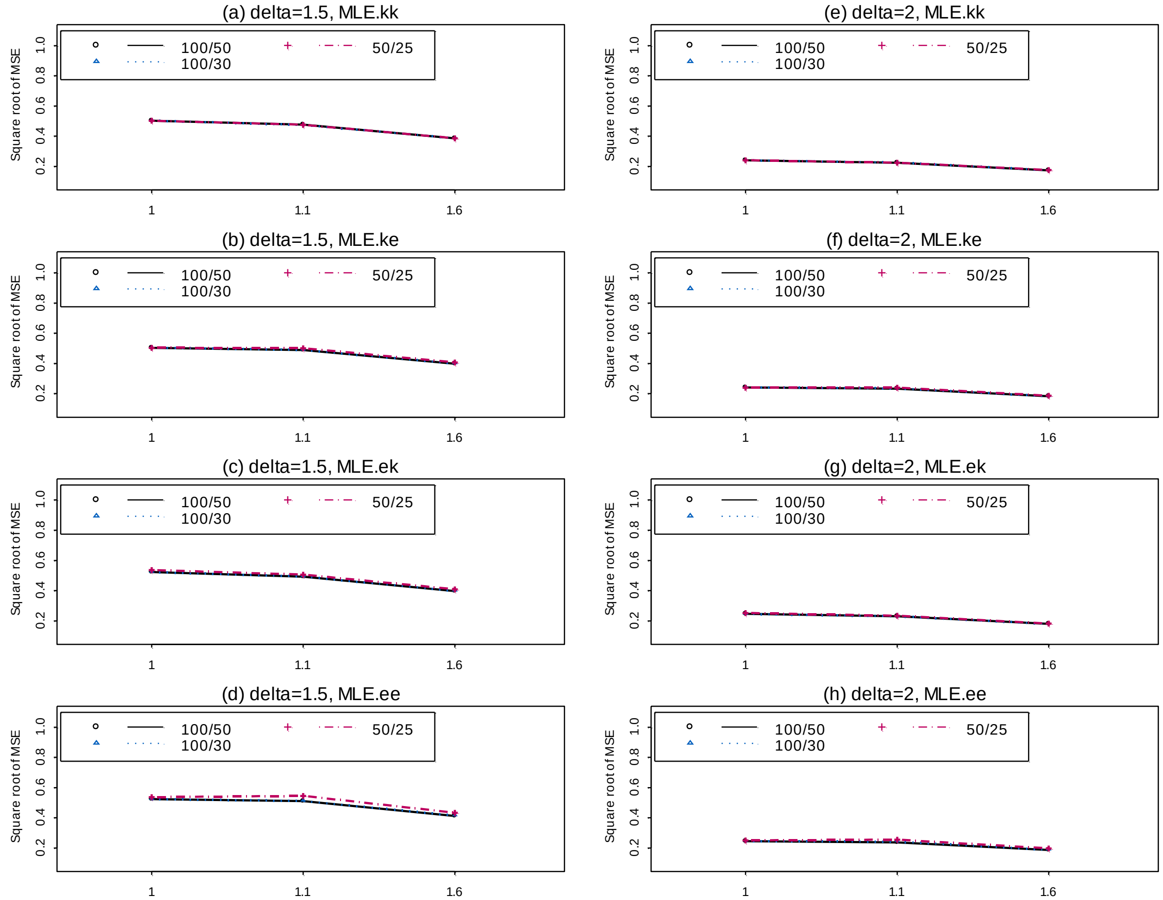

Figure 5.20 The effect of sample size and change-point position to the MLE estimation method for univariate series.

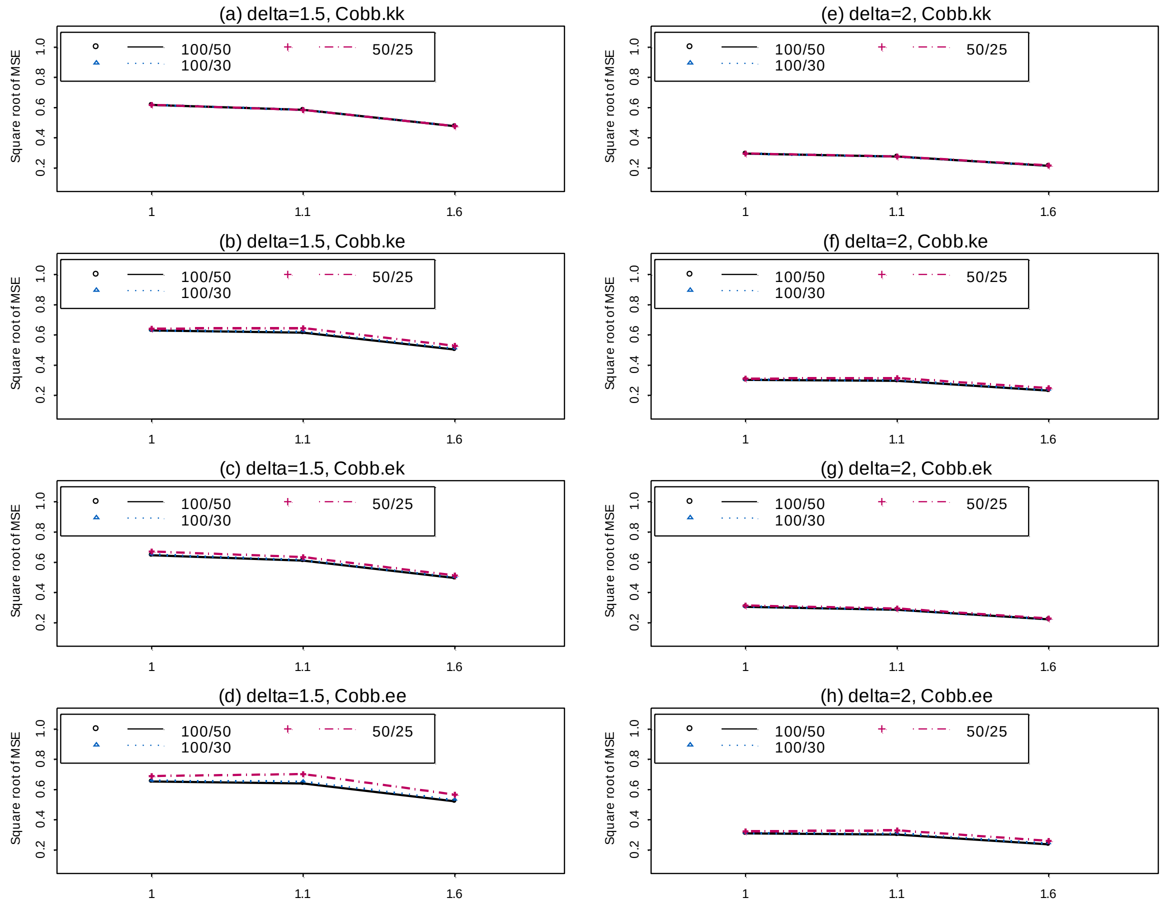

Figure 5.21 The effect of sample size and change-point position to the Cobb's estimation method for univariate series.

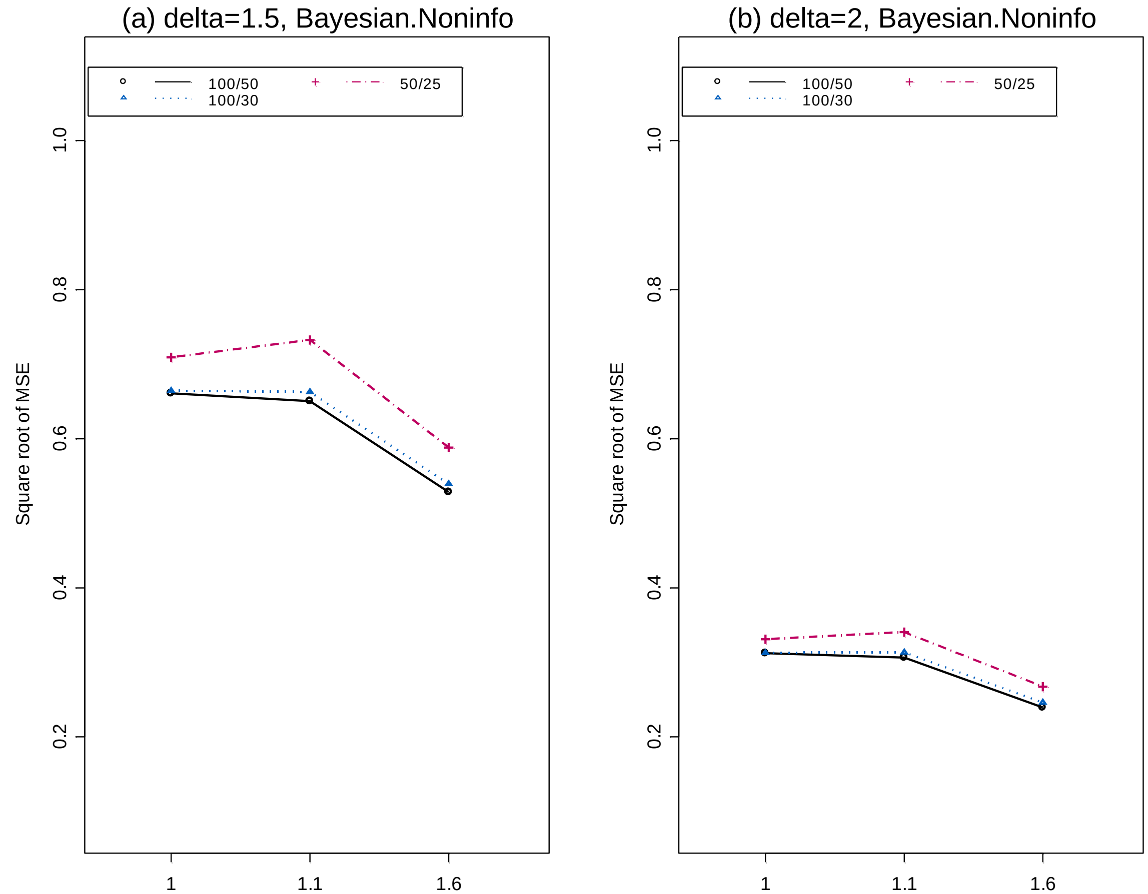

Figure 5.22 The effect of sample size and change-point position to the Bayesian's estimation method for univariate series.

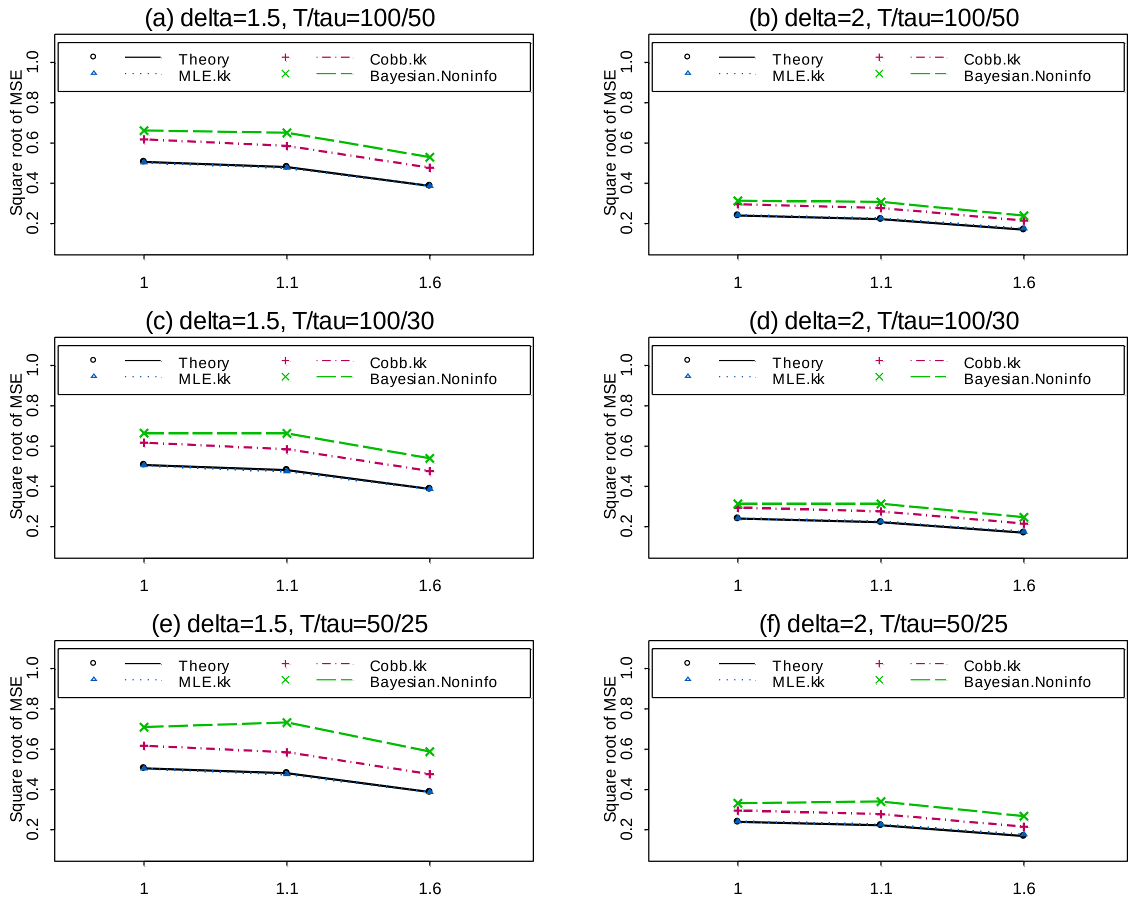

Figure 5.23 Comparison of estimation methods when the MLE and Cobb used ‘kk’ for parameter estimates for univariate series.

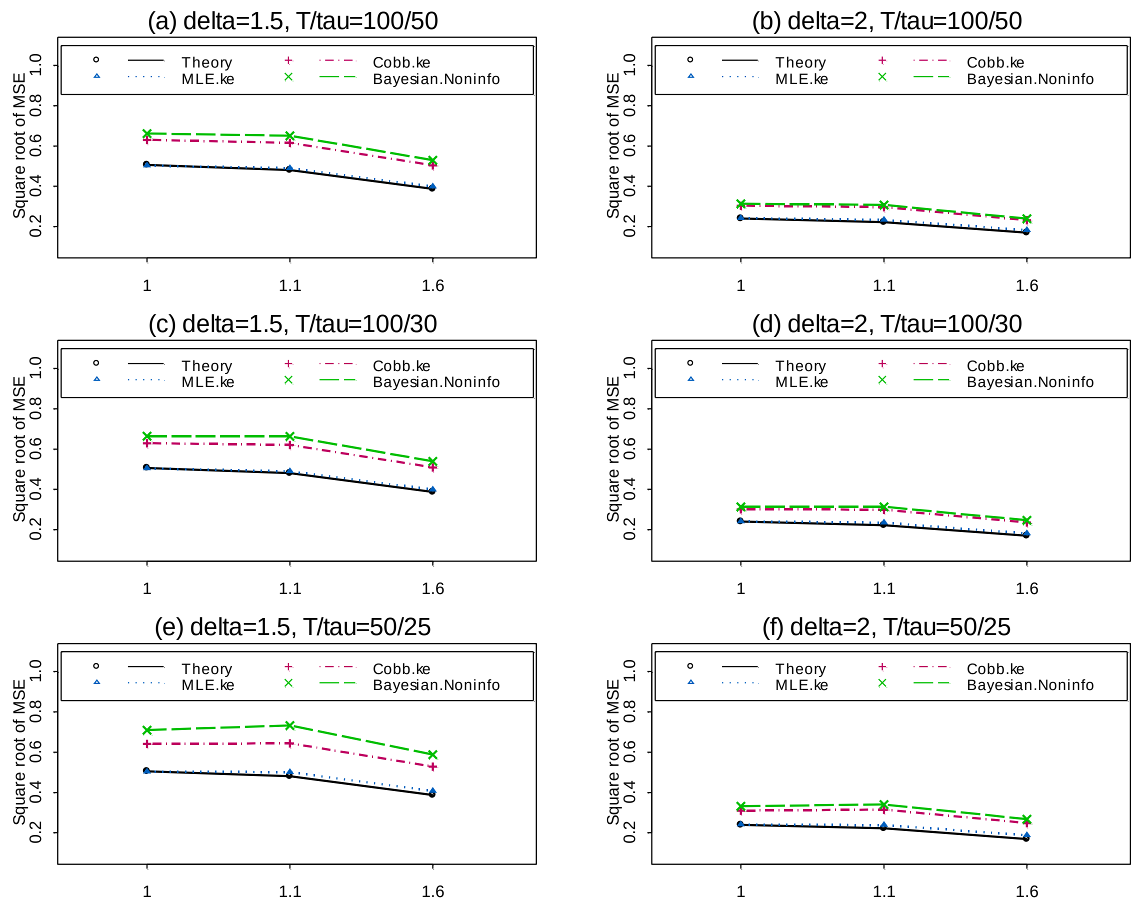

Figure 5.24 Comparison of estimation methods when the MLE and Cobb used ‘ke’ for parameter estimates for univariate series.

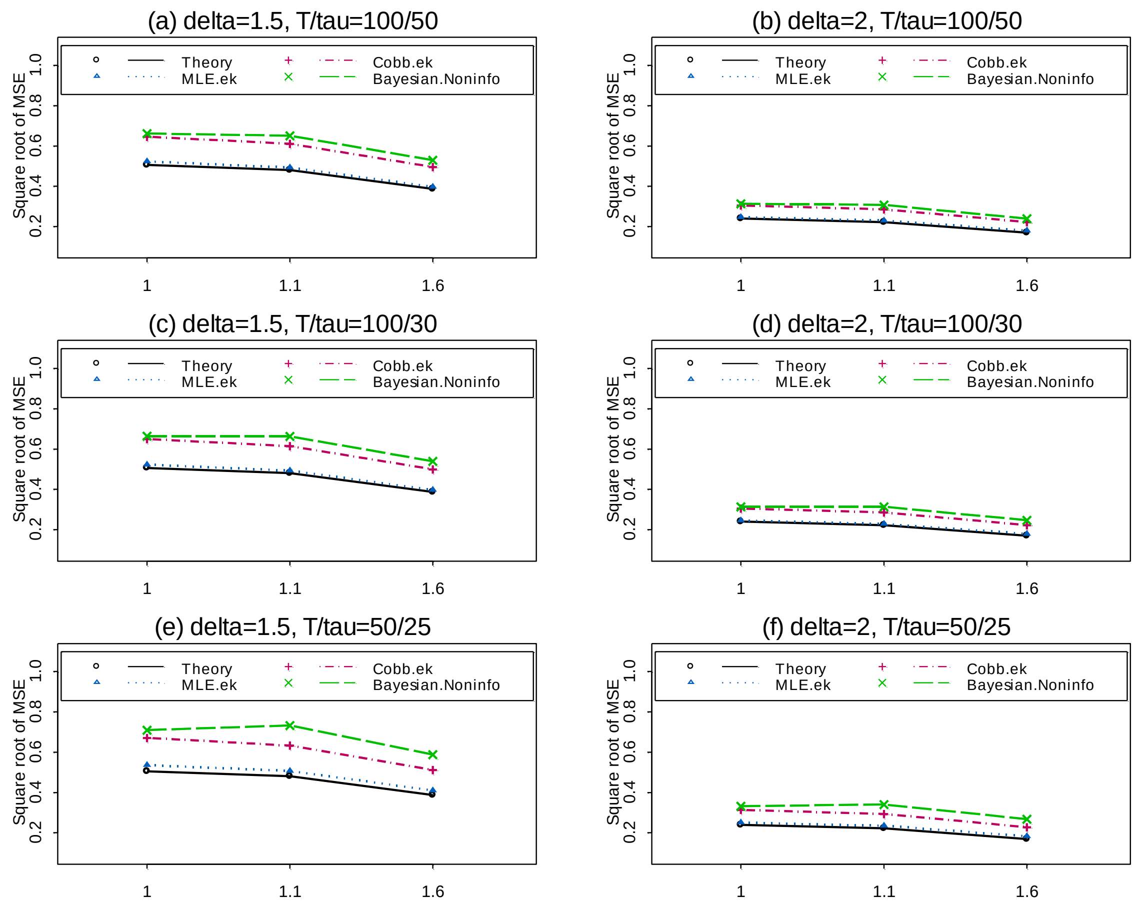

Figure 5.25 Comparison of estimation methods when the MLE and Cobb used 'ek' for parameter estimates for univariate series.

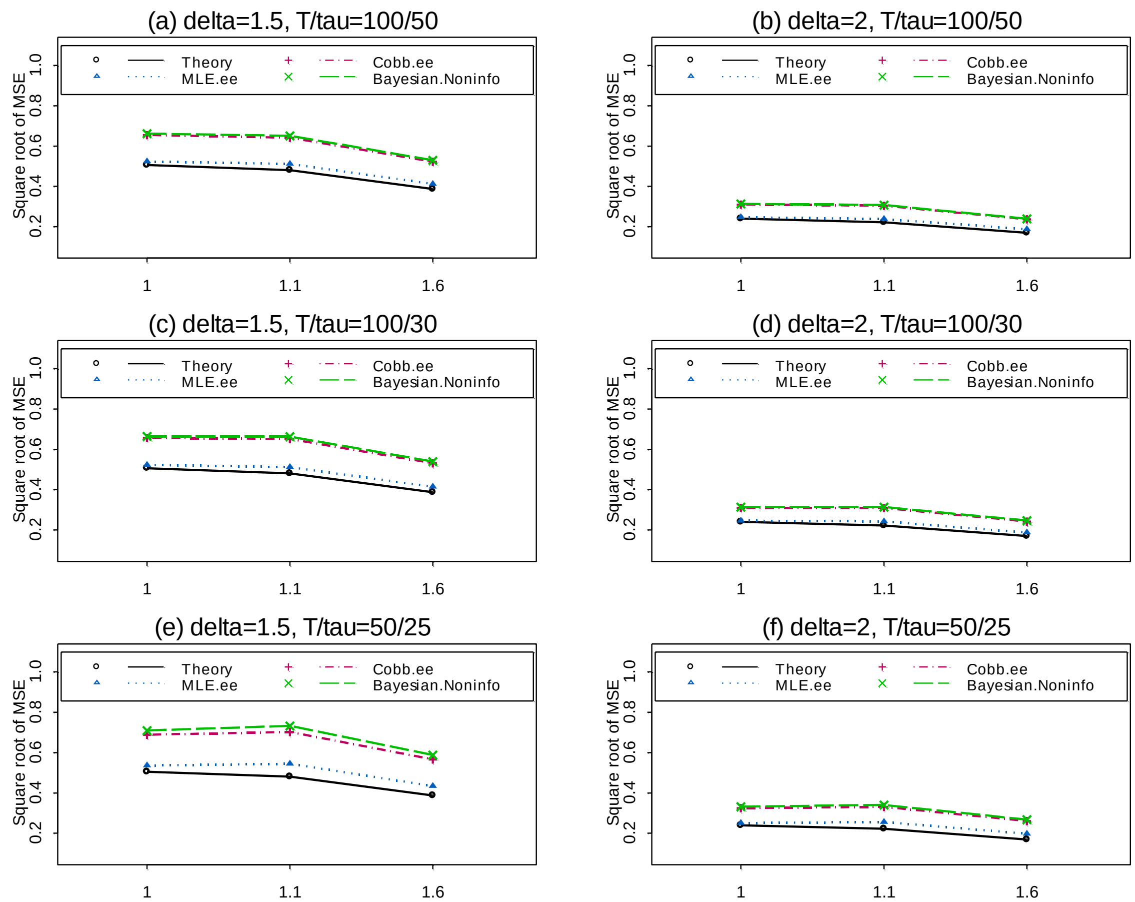

Figure 5.26 Comparison of estimation methods when the MLE and Cobb used ‘ee’ for parameter estimates for univariate series.

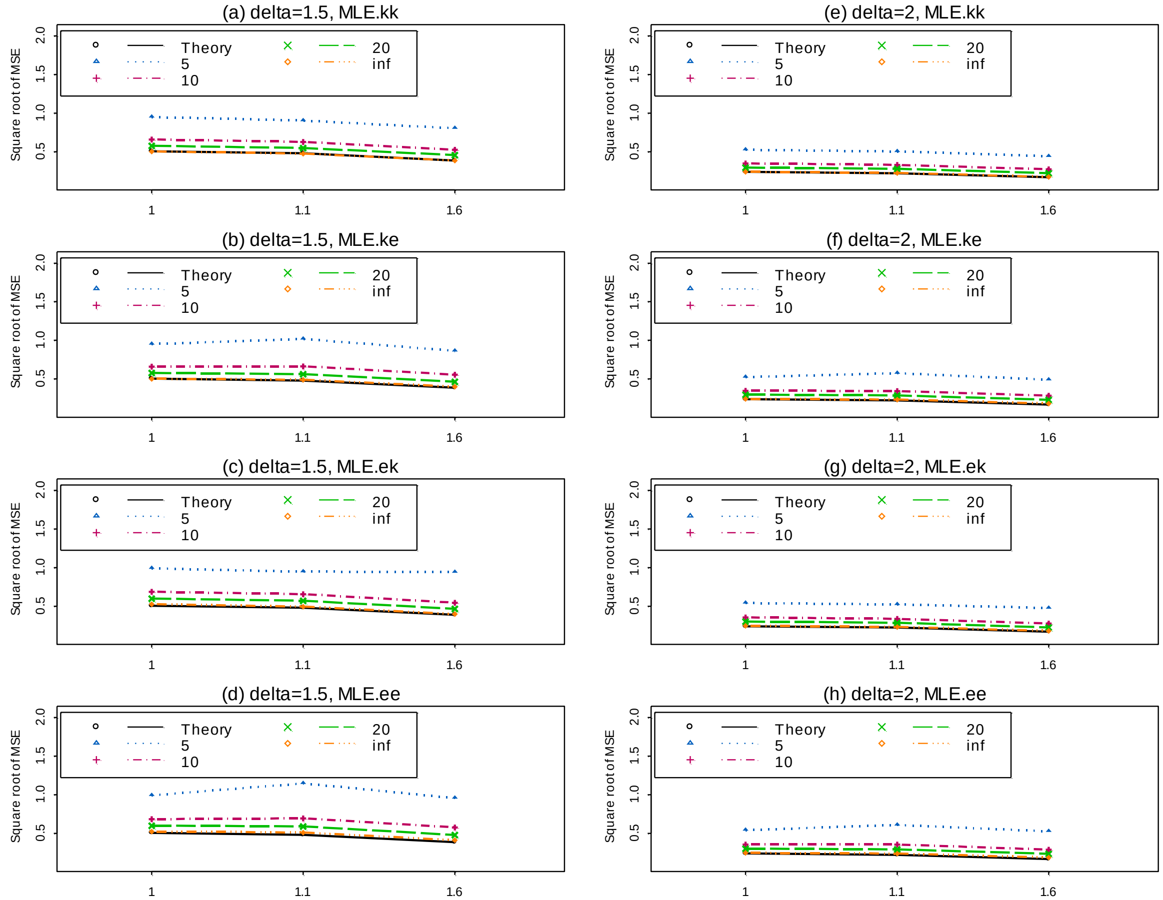

Figure 5.27 Effect of the degrees of freedom when the series follow univariate t-distribution using MLE method for univariate series.

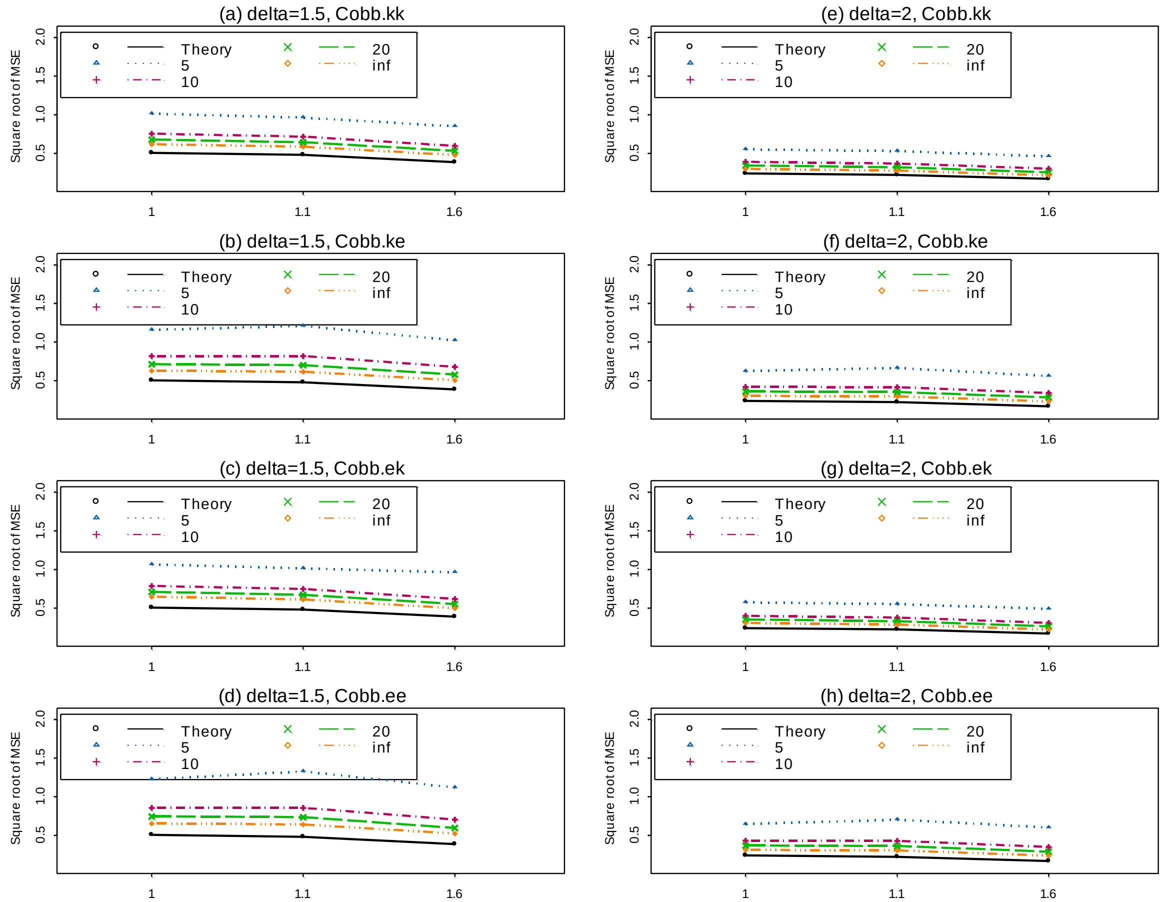

Figure 5.28 Effect of the degrees of freedom when the series follow univariate t-distribution using Cobb's method for univariate series.

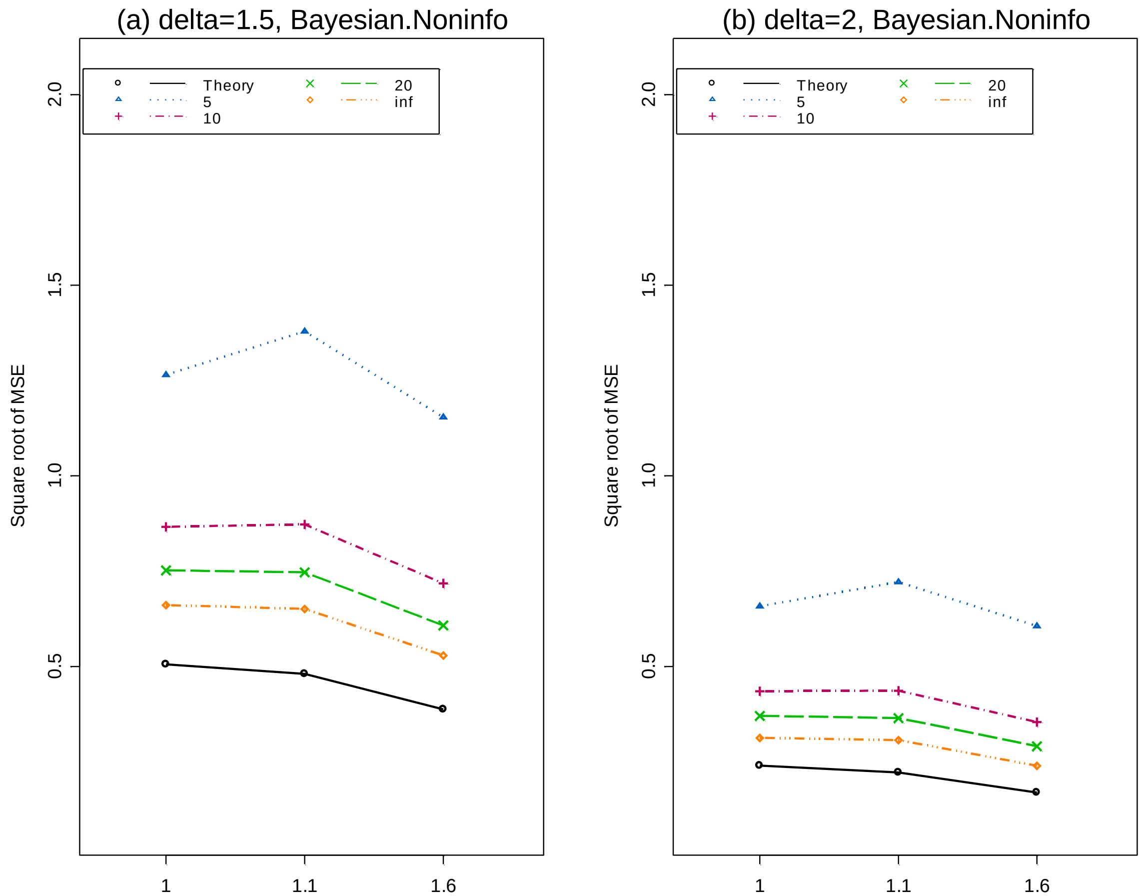

Figure 5.29 Effect of the degrees of freedom when the series follow univariate t-distribution using Bayesian method for univariate series.

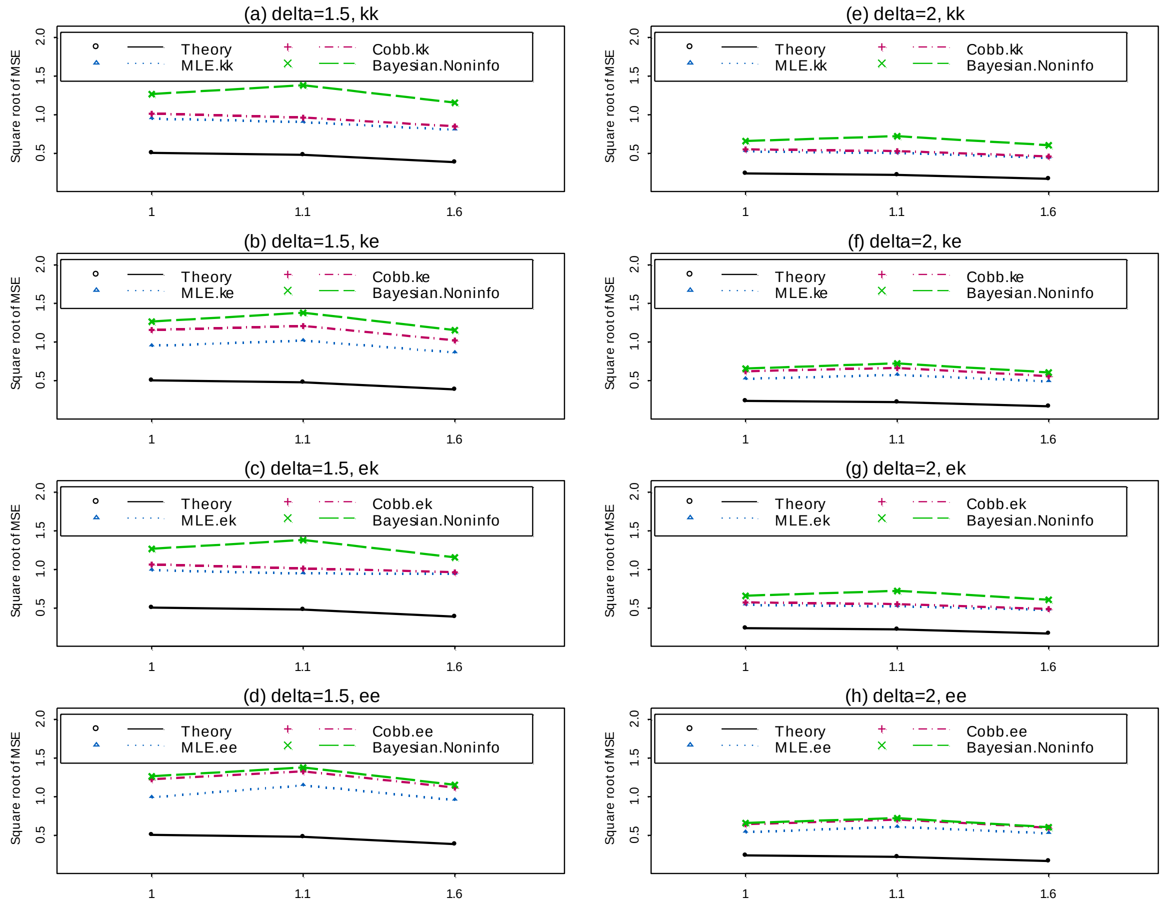

Figure 5.30 Comparison of estimation methods when the series follow univariate t-distribution with df=5 for univariate series.

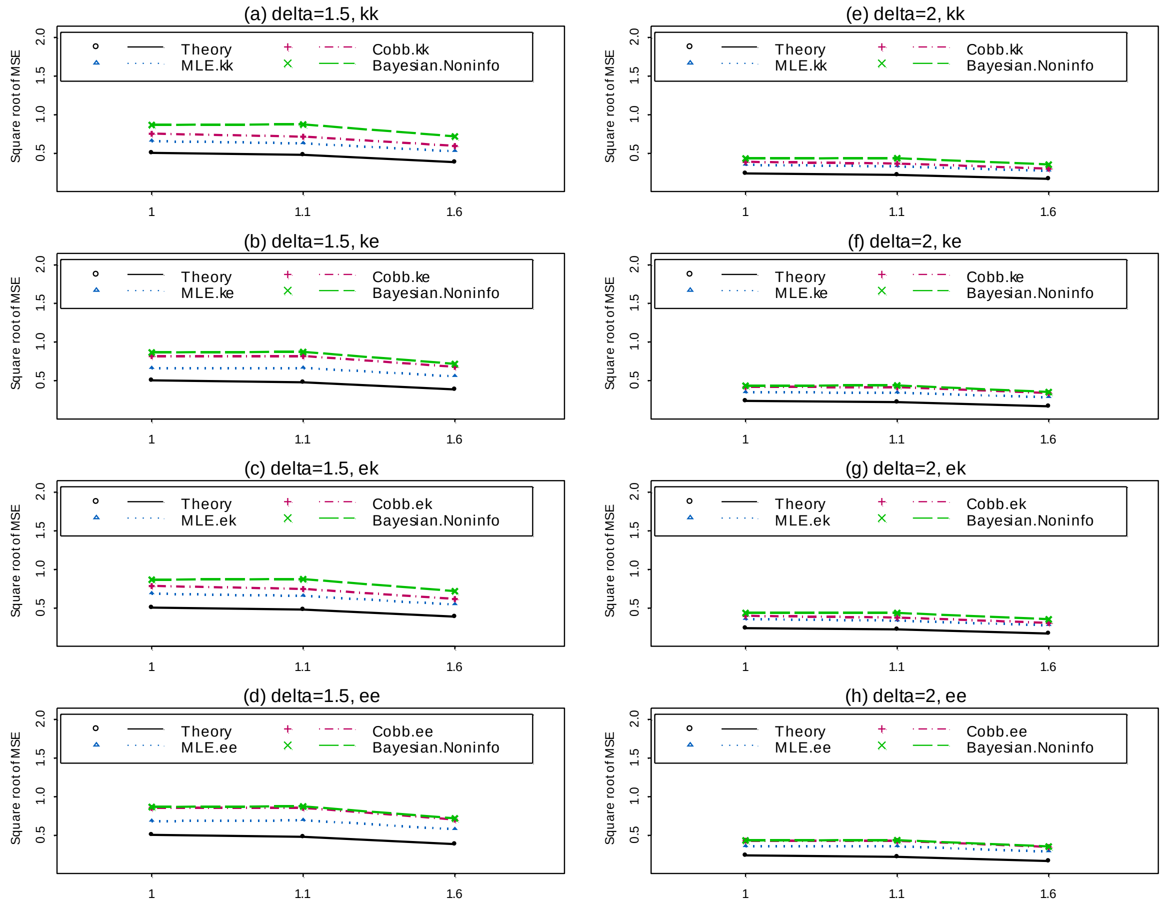

Figure 5.31 Comparison of estimation methods when the series follow univariate t-distribution with df=10 for univariate series.

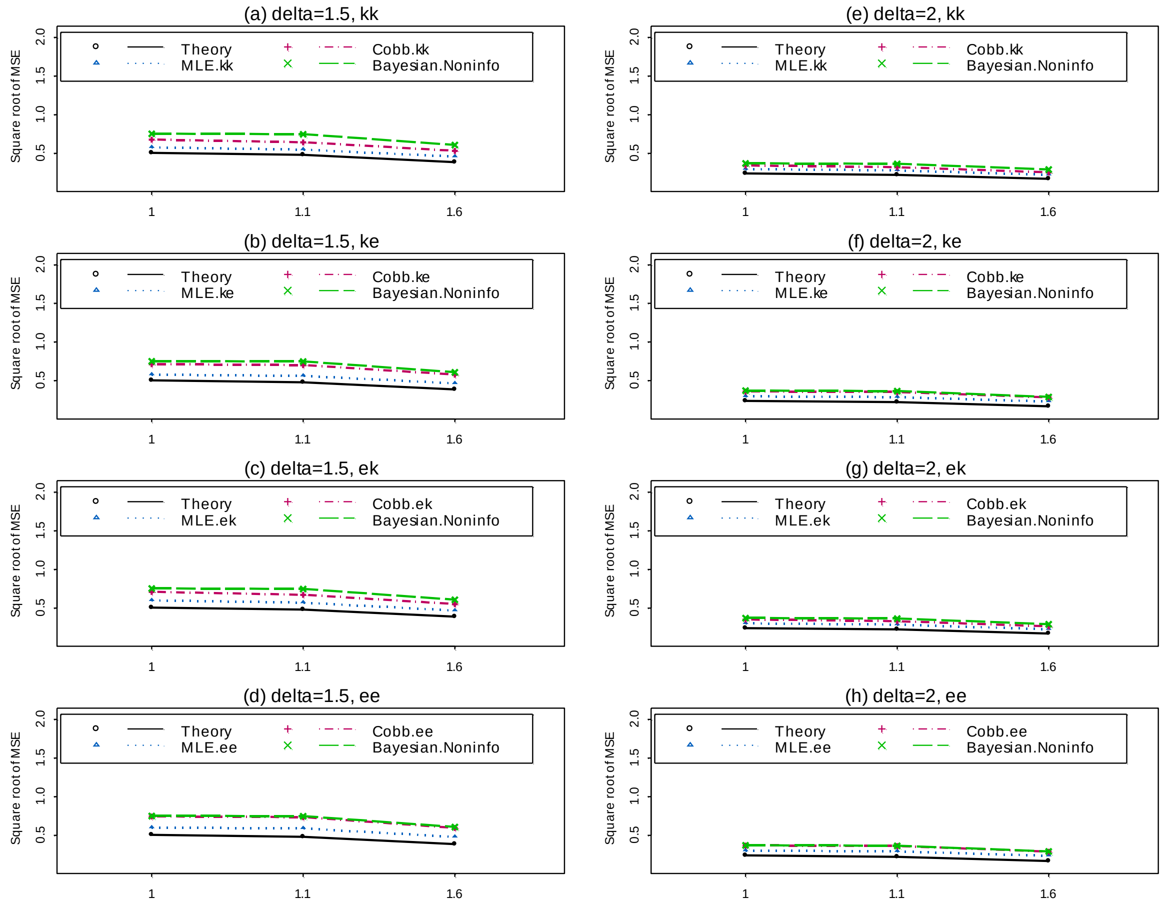

Figure 5.32 Comparison of estimation methods when the series follow univariate t-distribution with df=20 for univariate series.

Under the univariate case, the behavior of the mse was very similar to the multivariate case. But in general, all the methods produced closer results to the theoretical values. Both MLE and Cobb's method were not greatly affected by how the parameters before and after the change-point were estimated. The effect of the sample size and the location of the change-point could be negligible except using the Bayesian method. The MLE method produced the smallest mse, and the Bayesian method produced the largest, and the results using Cobb's method were in between. The mse produced by the MLE method was very close to the theoretical mse.

If the series followed t-distribution, all the methods produced larger mse when the degrees of freedom decreased. It could be seen from Figures 5.27 – 5.32 that unless the degree of freedom was 5, all the methods provided fairly close mse to the results when the normality assumption was satisfied. Among all the methods, the MLE consistently produced the smallest mse.

6 APPLICATION TO ENVIRONMENTAL MONITORING

6.1 River Stream Flows in the Northern Québec Labrador Region

The Québec-Labrador peninsula is located in the eastern part of Canada. It is surrounded by the United States in the south, James Bay and Hudson Bay in the west, the Strait of Hudson and Ungava Bay in the north, and the Atlantic Ocean in the east. Over half of the region is covered by forest. There are a large number of lakes in this region and several rivers run through it. This is one reason why the Québec-Labrador peninsula is of great importance from a hydrological point of view.

The climate in the region varies from being moderate to arctic as one travels from south to north. The eastern part of the peninsula has a marine climate because of the Atlantic Ocean. The southern part has a continental climate where the winters are long and chilly and the summers are warm and humid. The northern part has an arctic climate not only due to its high altitude but also due to the surrounding waters of James Bay and Hudson Bay. Consequently, the winters in this region are extremely cold and long, and the summers are short and cool.

The waters of James Bay and Hudson Bay make the area very moist. During fall and winter, the prevailing westerly winds pick up moisture from the James Bay and the Hudson Bay and dump it on the eastern part of the peninsula. The western part of the peninsula has many great glaciers of the world. Melted water from these glaciers drains north into the Ungava Bay, south to the gulf of Saint Lawrence River, east to the Atlantic Ocean, and west into the James and Hudson Bays. This region has more running water per unit area than any other place in the world, not only because of the abundance of melted snow and precipitation, but because it is so cold here most of the time that there is less water evaporated from the land (MacCutcheon, 1991).

All of this running water forms into a number of streams and rivers, and stream flow in them enters a natural recession when the air temperature drops below zero during the winter-period (December through April). It is in this winter-period that snowfall accumulates and forms a spectacular snow-cover in the region. When the winter snow-cover melts in the spring, spring floods occur and these spring floods account for most of the discharge into the rivers of the region.

Québec has one of the most extensive hydroelectric developments in North America. Nearly 96% of all of Québec's power is hydroelectric, and over 45% of all hydroelectric power in Canada is produced in the Province of Québec alone. Seen from this perspective, it is extremely important to monitor stream flows of various rivers of the region and look for signs of any appreciable changes; lest the economic consequences can be staggering.

Recently, Perreault et al (2000) considered data on average stream flows in spring (January to June runoff) for six rivers, namely, Romaine, Churchill Falls, Manicougan, Outardes, Sainte-Marguerite, and À la Baleine, that flow in the Northern Québec Labrador region. The collected data on the average stream flow from these rivers was expressed in $1/(km^2 \times s)$ and spanned through the years 1957-1995 except for the À la Baleine river, for which data was available only between 1963-1995. We present below in Figure 6.1 and also Appendix 1 the stream flow data on the six rivers.

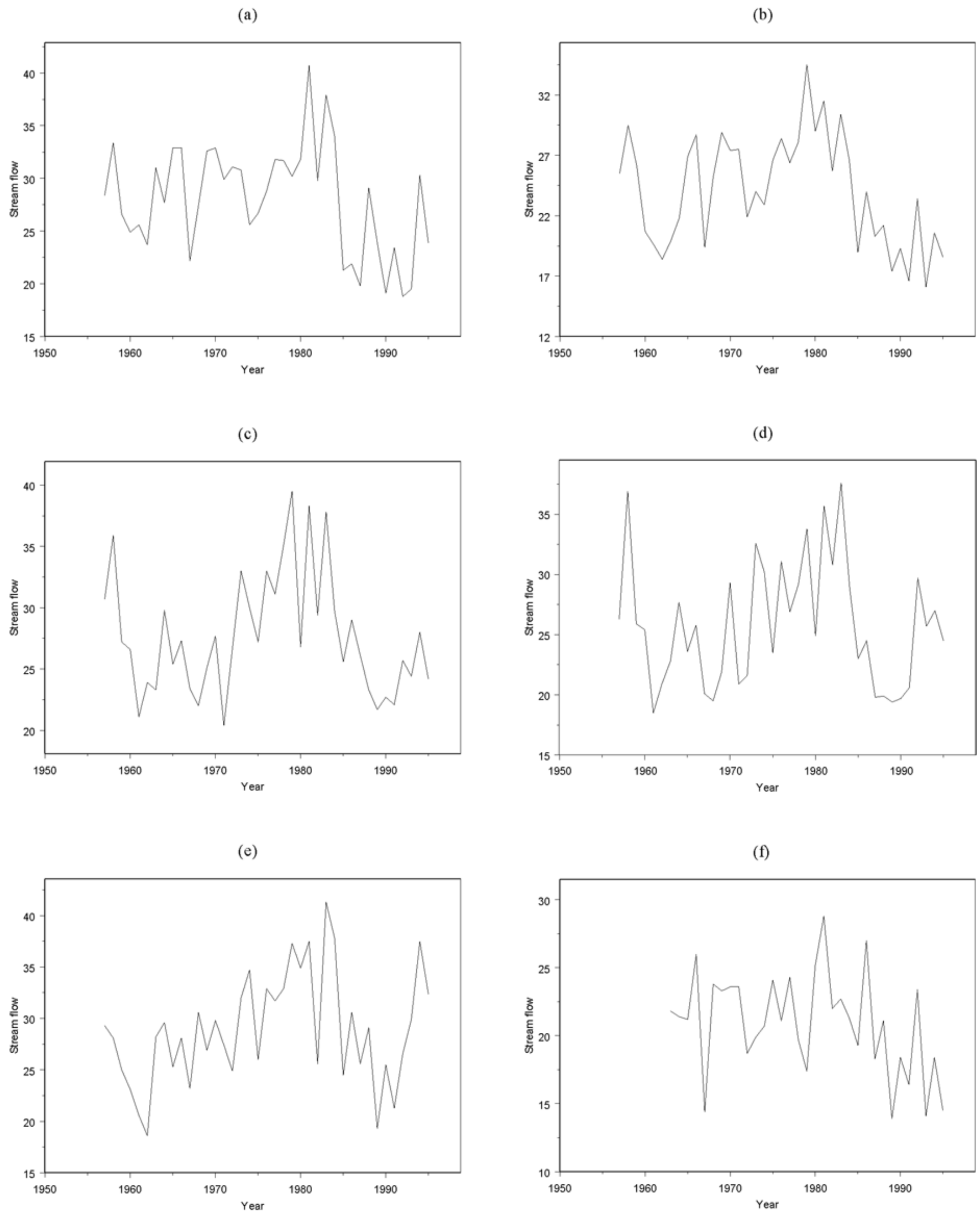

Figure 6.1. Average Spring flows of six rivers: (a) Romaine, (b) Churchill Falls, (c) Manicougan, (d) Outardes, (e) Sainte-Marguerite, (f) À la Baleine during 1957-1995 from the Northern Québec Labrador region.

The geographic locations of these six rivers are diverse in that they represent the region's varied climatic conditions quite well. Rivers Romaine, Manicouagan, Outardes, and Sainte-Marguerite are located south of Quebec. Their waters flow into the Saint Lawrence river, where it has a continental climate. The Churchill Falls is a group of waterfalls located in the Labrador region to the east side of the peninsula, where it has a marine climate. At a staggering 5948 MW of hydroelectric power, this series of rapids constitutes the second largest hydroelectric power generating capacity in North America. The river À la Baleine, otherwise known as the Great Wale River is located north of Québec and has an arctic climate. Thus, the six rivers could be viewed to represent the region's river system reasonably well. Hydroelectric power generation in the region critically depends on steady flows in the region's river system, year after year. Hence it is vital that stream flows in the river system are monitored from various aspects in order to understand the steadiness of flows or any departures thereof.

Our main goal is one of advancing the mle method for the estimation of an unknown change-point in the mean vector of a sequence of multivariate normal observations. The need for multivariate change-point methods is quite compelling. While analyzing data on six rivers from the Québec Labrador region, Perreault et al (2000) correctly argued that univariate formulation of each river independently would fail to take into account significant spatial correlations one can expect among the rivers that flow in the same region. Moreover, when modeled in a multivariate framework, any change-point identified in the river flows could be viewed as a global change-point for the whole region. Such region-wide conclusions would be inappropriate on the basis of individual univariate analyses only.

In our effort to extend the mle method for estimation of change in the mean of a multivariate normal distribution, we have shown in Chapter 3 that the multivariate problem can be directly translated into an equivalent univariate problem. Such a simplistic solution is not enjoyed by the Bayesian approach (see Perreault et al 2000). Thus while there are complexities that are difficult to overcome under the mle approach, it seems there is potential for it to yield delightful solutions as well. Of course, the greater aim of this chapter is to ensure that this simplistic solution under the mle is accessible to hydrologists, and to this extent we carry out a detailed application of the methodology to river flows from the Northern Québec Labrador region.

Due to the problem of missing data for the À la Baleine river between the years 1957-1962, Perreault et al (2000) utilized data only for the common period 1963-1965 in their main analysis, and used data from the five rivers for the period 1957-1962 as a prior sample to estimate hyperparameters of their priors. In our analysis, we pursue the application of the mle method for the data on river flows presented in Appendix 1 under three different cases: (i) full data for the first five rivers between the years 1957-1995; (ii) data for all six rivers between the years 1963-1995; and (iii) data for all six rivers between the years 1957-1995, in which we treat the years 1957-1962 for À la Baleine river as missing. Where appropriate we pursue the application under case (i) with greater emphasis. It will be evident in the subsequent sections that the three analyses yield quite similar results.

Before we go ahead with the technical formulation of the model, we need to address issues about the distribution to be considered, serial correlations in the data, and the behavior of spatial covariances over time. The same issues were also relevant for Perreault et al (2000). Adopting their assumptions, we initially formulate the river flows by the multivariate Gaussian family in which spatial covariance is stationary, and the observations are serially uncorrelated and hence independent over time. We shall revisit the assumptions of Gaussianity and independence and check for their validity through residual analysis of the fitted model. The assumption of stationarity of spatial covariance seems less of a concern for this data and hence we shall not pursue validation of this aspect in our residual analysis.

The rest of the section is organized as follows: first, we formulate the multivariate change-point model and then carry out the likelihood ratio test for change-point detection under the three cases of data choices, second, residual analysis will be performed for the fitted model under case (i), and lastly, asymptotic distribution of the change-point mle will be developed.

6.1.1 *Multivariate change-point model setup*

We begin by first formulating the change-point model for the river stream flow data in Appendix 1. The problem formulation will follow the problem setup in Chapter 2. Accordingly, let $Y_1, Y_2, \ldots, Y_{39}$ be a sequence of time-series valued independent random vectors such that $Y_i \in \mathbb{R}$, $i = 1, \ldots, 39$. Furthermore, for each $i = 1, \ldots, 39$ let Y_i follow the multivariate Gaussian distribution with mean vector μ and variance-covariance matrix Σ. Then, under the classical change-point model in which the covariance matrix Σ remains stationary throughout the sampling period, the mean vector μ changes from an initial value μ_0 to a subsequent value μ_1 at some unknown change-point $\tau_{39} \in \{1,2, \ldots, 38\}$. Following the assumption of our study, it will be still assumed that parameters μ_0, μ_1 and Σ are all unknown. Thus, under the change point model, one has

$$Y_i \sim \begin{cases} f(\cdot; \mu_0, \Sigma), & i = 1, \quad \ldots, \tau_{39} \\ f(\cdot; \mu_1, \Sigma), & i = \tau_{39} + 1, \ldots, 39 \end{cases} \tag{6.1}$$

where $\tau_{39} \in \{1,2, \ldots, 38\}$.

On the other hand, when there is no change point in the model, a single parameter set is applicable throughout the sampling period such that under the no change model one has

$$Y_i \sim f(\cdot; \mu_0, \Sigma), \qquad i = 1, \ldots, 39 \tag{6.2}$$

One is confronted with having to decide whether the given river flow data set can be modeled by the no change model (6.2) , or by the change point model (6.1) with a change occurring in the mean vector at an unknown change point τ_{39}. Thus, the statistical problem is one of carrying out a test of the following hypotheses:

H0: The data conforms to no change model (6.2)

Against H1: The data conforms to change point model (6.1).

6.1.2 *Detection of an unknown Change-Point in River Stream Flows*

We shall now apply the above detection methodology to cases (i) – (iii) as elaborated in Chapter 1. While applying the methodology to case (i), we have computed the likelihood ratio $U_{n,t}$ for $t = 3, \dots, 37$. This allowed us to have minimum number of observations at either end to compute mles $\hat{\mu}_{0,t}$, $\hat{\mu}_{1,t}$ and $\hat{\Sigma}_t$ for each value of t. Similar approach was used for cases (ii) and (iii) also. Also, while analyzing the data under case (iii) we first replaced each of the six missing data values for À la Baleine river by the average of the data for the years 1963-1995 for the same river.

Figure 6.2 is a plot of $U_{39,t}$ for $t = 3, \dots, 37$ under case (i). The computed values of the statistic W and the corresponding P-values for the three cases were: (i) $w_1 = 5.99, P - value_1 = 0.0050$; (ii) $w_2 = 6.39, P - value_2 = 0.0033$; (iii) $w_3 = 6.31, P - value_3 = 0.0036$. Clearly all three cases provide strong evidence to conclude that there is an unknown point of time in the river flow data subsequent to which the mean vector has changed significantly.

The mle $\hat{\tau}_{39}$ of the unknown change-point τ_{39} is obtained as the value of *t* at which $U_{39,t}$ is maximized. The mle under case (iii) is $\hat{\tau}_{39} = 28$, which implies that mean vector for the six rivers considered in the data has changed significantly subsequent to the year 1984. Cases (i) and (ii) also provide 1984 as the year of change. It is important to note that the mle coincides with the Bayesian posterior mode obtained by Perreault et al (2000). However, it is not sufficient to merely obtain the above point estimate for estimating the time of change. It is much more preferable to have confidence interval estimates at any desired level.

Figure 6.2: Twice log-likelihood ratio for a given change-point for the six rivers from the Northern Québec Labrador region..

Before moving on, we shall first perform residual analysis on the basis of residuals from the fitted model under case (iii), mainly to investigate the appropriateness of the assumptions of multivariate Gaussianity and independence over time. The results of appropriateness analysis for cases (i) and (ii) are similar and will not be further discussed here.

First, we need mles of all model parameters in order to fit the model and get the residuals. On the basis of the mle being $\hat{\tau}_{39} = 28$, the computed mles $\hat{\mu}_{0,28}$, $\hat{\mu}_{1,28}$ and $\hat{\Sigma}_{28}$, of the model parameters were

$$\hat{\mu}_{0,28} = (30.11, 25.77, 28.84, 26.87, 29.40, 22.05)$$

$$\hat{\mu}_{1,28} = (22.82, 19.68, 24.80, 23.07, 27.48, 18.62)$$

and

$$\hat{\Sigma}_{28} = \begin{pmatrix} 15.48 & 7.93 & 8.19 & 8.68 & 12.26 & 5.45 \\ 7.93 & 12.70 & 10.21 & 9.84 & 11.32 & 7.06 \\ 8.19 & 10.21 & 20.14 & 19.10 & 15.96 & 3.28 \\ 9.68 & 9.84 & 19.10 & 23.03 & 16.86 & 4.07 \\ 12.26 & 11.32 & 15.96 & 16.86 & 27.47 & 5.55 \\ 5.45 & 7.06 & 3.28 & 4.07 & 5.55 & 11.60 \end{pmatrix}$$

Apart from the above point estimates, standard errors for $\hat{\mu}_{0,28}$ and $\hat{\mu}_{1,28}$ would also be of interest. Based upon variance estimates represented by the diagonal elements in $\hat{\Sigma}_{28}$ above, and corresponding sample size of 28, standard errors for components in $\hat{\mu}_{0,28}$ were $se_0 = (0.743 \quad 0.673 \quad 0.848 \quad 0.907 \quad 0.990 \quad 0.644)$. Similarly, with sample size 11, the standard errors for $\hat{\mu}_{1,28}$ were $se_0 = (1.186 \quad 1.074 \quad 1.353 \quad 1.447 \quad 1.580 \quad 1.027)$

The above parameter estimates yield the fitted model, which in itself yields the corresponding residuals for the fitted multivariate model. As part of residual analysis, we first applied the standard Shapiro-Wilk test for univariate normality of residuals from each river and we found the corresponding P-values to be 0.5144, 0.9616, 0.3780, 0.4241, 0.7063, and 0.8980, respectively. Thus, residuals from all six rivers confirm the assumption of Gaussianity in a univariate way. However, it is well known that univariate Gaussianity of each series may not necessarily imply that the residual data in vector form would be multivariate Gaussian. Thus, in order to test for multivariate Gaussianity of the residuals, we applied Mardia's skewness and kurtosis tests (Mardia, 1970), as well as the test proposed by Henze and Zirkler (1990). The P-values for the three tests were 0.0294, 0.7505 and 0.9463, respectively. Thus, except for the marginal evidence based on skewness test, there is no evidence otherwise in the multivariate tests that multivariate Gaussianity assumption is in violation.

Next, we utilize the residuals for investigating the assumption of independence over time. If the model under independence is truly a good fit, then both autocorrelations and partial autocorrelations of different lags for residuals from each of the six rivers should show no appreciable significances. Moreover, independence at the multivariate level should mean that forward as well as backward cross-correlations of different lags for residuals from any pair of rivers should also show no significances. First, we present in Figures 6.3a-e, autocorrelation plots up to the first ten lags together with the corresponding 5% significance curves, for residuals of each of the six rivers.

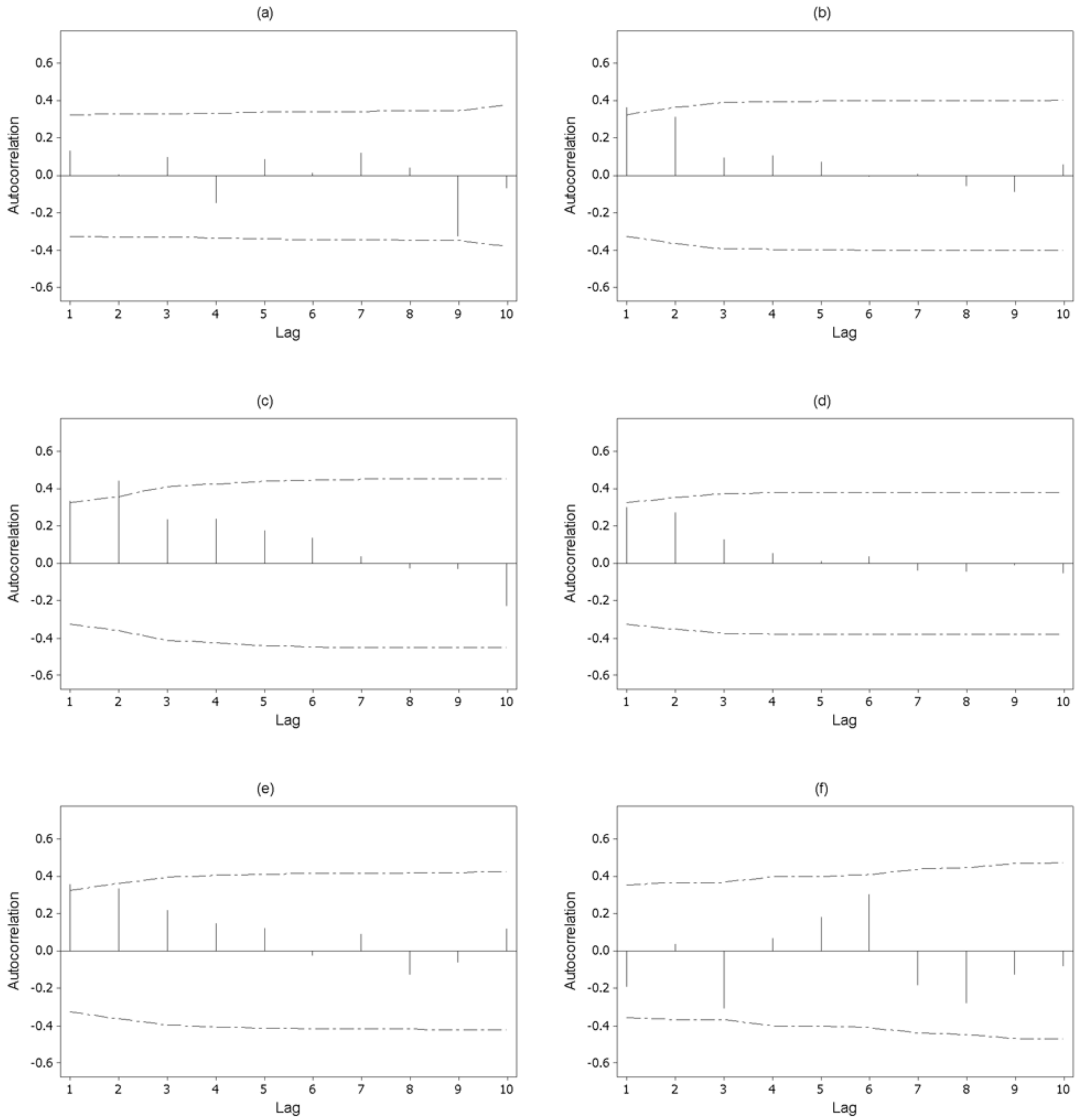

Figure 6.3: Plot of auto correlations for residuals from six rivers, (a) Romaine, (b) Churchill Falls, (c) Manicougan, (d) Outardes, (e) Sainte-Marguerite, (f) À la Baleine.

Except for the autocorrelation plot for the Manicouagan river (Figure 6.3c) in which the auto correlation at lag 2 is significant, the remaining autocorrelation plots obey independence quite well. While Figure 6.3(a), 6.3(d) and 6.3(f) show no significance at all, the significances at lag 1 in Figure 6.3(b) and 6.3(e) seem too marginal to be of any real concern. While we do not display the partial autocorrelation plots, we found that their behavior was uniformly better than the autocorrelation plots for all the five rivers and hence they show even less of a concern. Next, we plotted the cross-correlation plots for all pairs of rivers (both forward and backward) and noted that they were all essentially similar to Figure 6.3(a)-(f). Thus, the auto correlation, partial auto correlation, and cross-correlation plots put together do not show any strong evidence that would be indicative of violation in the assumption of time independence among the vector valued observations. Overall, the residual analysis firmly validates multivariate Gaussianity and does not indicate violation of independence over time in a manner that it would pose real concerns.

6.1.3 *Asymptotic Distribution of the Change Point MLE for River Stream Flows*

In section 6.1.3, we found the MLEs of the model parameters under case (iii) to be $\hat{\tau}_{39} = 28$. We shall now obtain confidence interval estimate of τ_{39} through the asymptotic distribution of $\hat{\tau}_{39}$ for which, we can assume that $\hat{\mu}_{0,28}$, $\hat{\mu}_{1,28}$ and $\hat{\Sigma}_{28}$ given above are true values rather than estimates. As discussed in Chapter 3, this is possible because the asymptotic distributions of $\hat{\tau}_{39}$ and $\tilde{\tau}_{39}$ are identical. This allowed us to compute $\eta = \hat{\Sigma}_{28}{}^{-1/2}(\hat{\mu}_{1,28} - \hat{\mu}_{0,28})$ and hence $\delta = \frac{1}{2}\sqrt{\eta^T \eta}$. The value of δ under case (i) was $\delta_1 = 1.22$. The corresponding values under cases (ii) and (iii) were $\delta_2 = 1.34$ and $\delta_3 = 1.23$, respectively. Adapting the algorithmic procedure in section 3.1, we computed the asymptotic distribution of $\hat{\tau}_{39} - \tau_{39}$ for δ_1, δ_2 and δ_3. Since the asymptotic distribution of $\hat{\tau}_{39} - \tau_{39}$ is symmetric around zero, we present probabilities only for nonnegative integers of $\hat{\tau}_{39} - \tau_{39}$. We also find it convenient to present in the same table cumulative probabilities of the form $\Pr(|\hat{\tau}_{39} - \tau_{39}| \leq i)$.

On the basis of the distributions for $\hat{\tau}_{39} - \tau_{39}$, the standard deviation of the asymptotic distribution of the change point MLE under case (i) was found to be 0.80 years, it was 0.65 years under case (ii), and 0.79 years under case (iii). Moreover, Table 6.1 has allowed us to compute confidence interval estimates for τ_{39} of any desired level. For example, a 93% confidence interval estimate for the true change point τ_{39} under cases (i) and (iii) is {27, 28, 29}, and a 97% confidence interval is {26, 27, 28, 29, 30}. The same interval estimates expressed in years are {1983, 1984, 1985}, and {1982, 1983, 1984, 1985, 1986}, respectively. Under case (ii), these intervals can be seen to have confidence levels of 95% and 98%, respectively. For comparison purposes, we note that the standard deviation for the change point estimate under the MLE approach is smaller than the reported value of 2.33 years under the Bayesian approach (Perreaul et al. 2000a). Further, the 90% posterior credibility interval under the Bayesian approach reported by (Perreaul et al. 2000a) was {1982, 1983, 1984, 1985}. While the MLE approach seems to do better than the Bayesian approach in this case, one should not conclude that this would be the case with other data sets. There is scope for uncertainty in the asymptotic distribution of the MLE due to the fact that the mean vectors before and after the change point were assumed to be known, whereas, in reality they are estimated from data. Even though the equivalence result of (Hinkley 1972) justifies this consideration for large samples, one should be prepared for some uncertainty because of the limited nature of the sample size in the river flow data.

Table 6.1. Asymptotic distribution of $\hat{\xi}_\infty$ under case (i), (ii) and (iii) for the change-point mle of the six rivers from the Northern Québec Labrador region..

i	$\Pr(\hat{\xi}_\infty = \pm i)$			$\Pr(\lvert\hat{\xi}_\infty\rvert \le i)$		
	$\delta_1 = 1.22$	$\delta_2 = 1.34$	$\delta_3 = 1.23$	$\delta_1 = 1.22$	$\delta_2 = 1.34$	$\delta_3 = 1.23$
0	0.7543	0.8036	0.7588	0.7543	0.8036	0.7588
1	0.0892	0.0760	0.0881	0.9328	0.9555	0.9350
2	0.0223	0.0159	0.0217	0.9773	0.9872	0.9784
3	0.0068	0.0041	0.0066	0.9910	0.9954	0.9915
4	0.0023	0.0012	0.0022	0.9957	0.9978	0.9960
5	0.0008	0.0004	0.0008	0.9974	0.9986	0.9975
6	0.0003	0.0001	0.0003	0.9981	0.9988	0.9981
7	0.0001	0.0000	0.0001	0.9983	0.9989	0.9984

6.2 Change-point Analysis of Zonal Temperature Deviations

6.2.1 Dataset description

In our example, the dataset is from Angell's (2009) study on global temperature deviation derived from radiosonde records. The dataset contains mean annual and seasonal air temperature for surface and upper layers (850 – 300, 300 – 100 and 100 – 50 mb) from 1958 to 2008, where the 850 – 300 mb layer represented the troposphere, 300 – 100 represented the tropopause, and 100 – 50 mb layer represented the lower stratosphere. The data was obtained from 63 globally distributed radiosonde stations. Angell's (2009) illustration for the atmosphere layers are as in Figure 6.4.

The upper-air temperature was obtained from the difference in height between constant-pressure layers at each individual station. Angell (2009) obtained the pressure-height data before 1980 from published values in *Monthly Climatic Data for the World*. Between 1980 and 1990, Angell (2009) obtained the data from *the Climatic Data for the World* and *the Global Telecommunications System (GTS) Network* which was available at the National Meteorological Center. Between 1990 and 1995, Angell (2009) obtained the data from GTS only. Since 1995, Angell (2009) has obtained the data from National Center for Atmospheric Research files. The data are evaluated as deviations from the mean based on the interval 1958-1977.

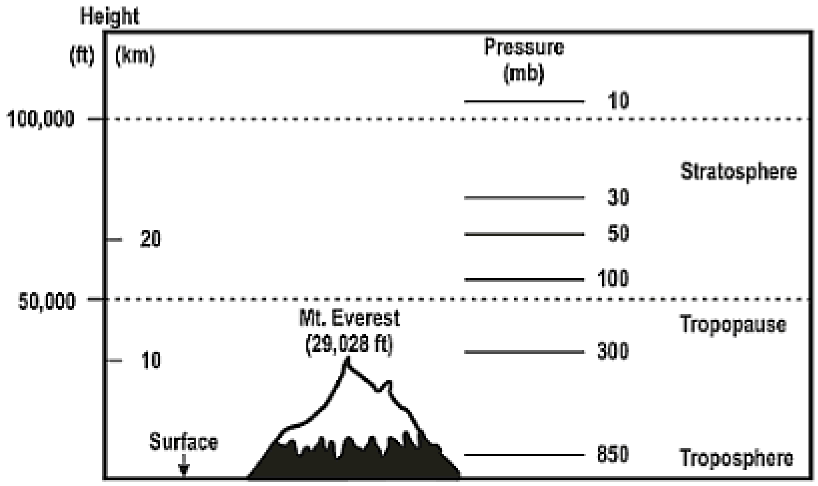

Figure 6.4. Layers of atmosphere for Angell's (2009) radiosonde temperature data.

All the data have been presented as the deviations from 1958-1977 mean temperatures. Then the deviations from all the stations were averaged with equal weights to obtain annual and seasonal mean temperature deviations. Currently the data are available for South Polar ($60° S - 90° S$), and North Polar ($60° N - 90° N$). The dataset for South Polar temperature deviation is in Appendix II, and the dataset for North Polar is available in Appendix III. Figures 6.5 and 6.6 are the time series plot for the annual mean temperature deviations for South and North Polar respectively.

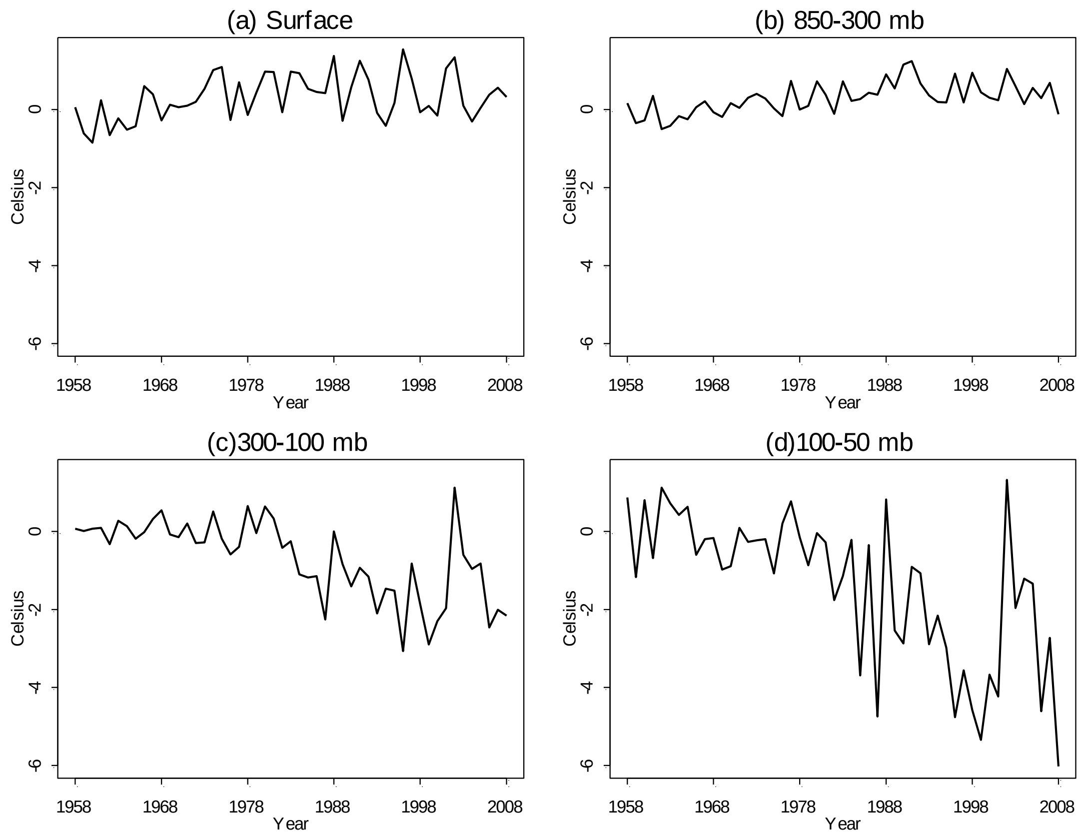

Figure 6.5. South Polar annual mean temperature deviations during 1958 – 2008.

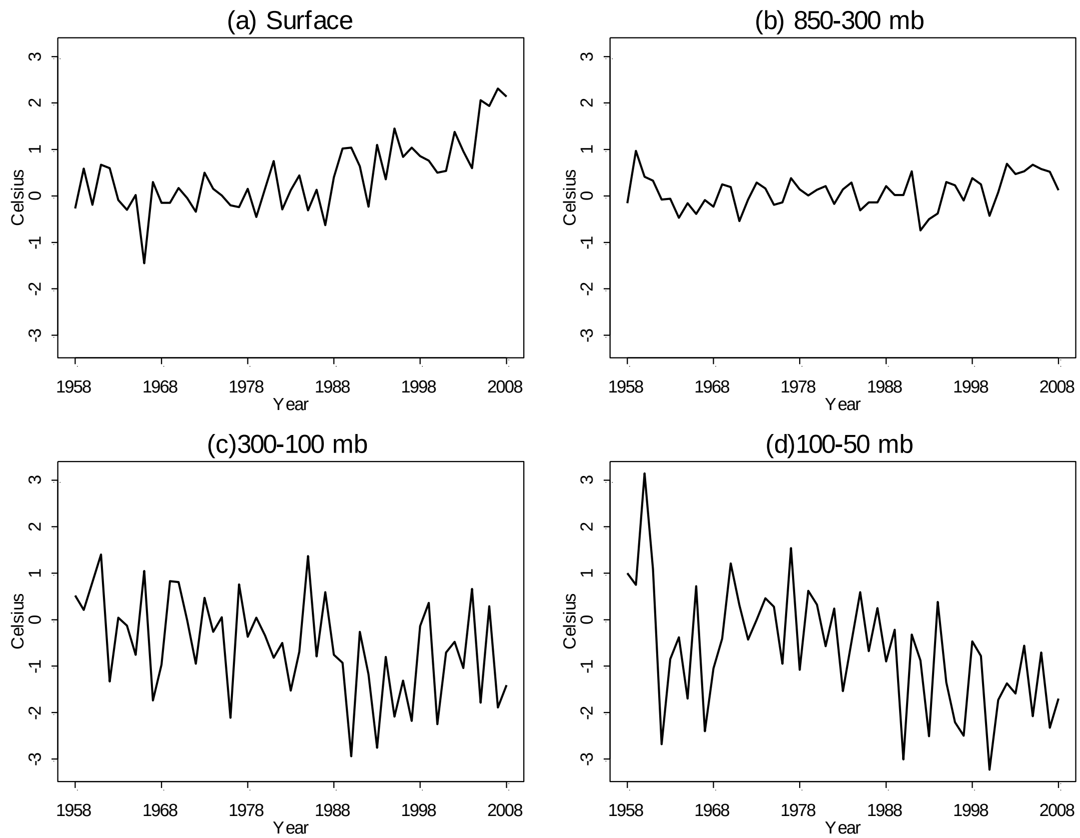

Figure 6.6. North Polar annual mean temperature deviations during 1958 – 2008.

From the time series plots in Figures 6.5 and 6.6, it was obvious that some parameter change had occurred in most of the series; however, in the past, there had been little formal statistical analysis for the change-point, and we were not able to tell exactly when the change had occurred. In the following sections, the change-point analysis using mle methods were applied to the data. The change-point was detected using maximum likelihood method, and the mean and confidence interval about the detection were computed using the method specified in Chapter 4.

6.2.2 *Change-point Analysis at South Polar*

We begin by first formulating the change-point model for the South Polar temperature deviation data in Appendix II. The problem formulation will follow the problem setup in Chapter 2. Accordingly, let $Y_1, Y_2, \dots, Y_{51}$ be a sequence of time-series valued independent random vectors. Furthermore, for each $i = 1, \dots, 51$ let Y_i follow the Gaussian distribution with mean vector μ and variance-covariance matrix Σ (variance σ^2 for univariate case). Assuming the parameters for the observations, μ and Σ (or σ^2), keep constant, the no-change model can is

$$M_0\colon\ Y_i \sim N(\mu, \Sigma), \qquad i = 1, \qquad \dots, 51 \tag{6.3}$$

If a change in the parameters occurs at some unknown point of time, we have 3 possible models: (i) change in mean only, that is to say, the mean changed from μ_0 to μ_1; (ii) change in covariance (or variance) only, that is to say, the covariance matrix (or variance) changed from $\Sigma_0\,(or\ \sigma_0^2)$ to $\Sigma_1\,(or\ \sigma_1^2)$; (iii) change in both mean and covariance (or variance), that is to say, the mean changed from μ_0 to μ_1, and the covariance matrix (or variance) changed from $\Sigma_0\,(or\ \sigma_0^2)$ to $\Sigma_1\,(or\ \sigma_1^2)$. The models with change-point under the three cases can be written as

$$M_1\colon\ Y_i \sim \begin{cases} N(\mu_0, \Sigma), & i = 1, \qquad \dots, \tau_n \\ N(\mu_1, \Sigma), & i = \tau_n + 1, \dots, 51 \end{cases} \tag{6.4}$$

$$M_2\colon\ Y_i \sim \begin{cases} N(\mu, \Sigma_0), & i = 1, \qquad \dots, \tau_n \\ N(\mu, \Sigma_1), & i = \tau_n + 1, \dots, 51 \end{cases}$$

$$M_3\colon\ Y_i \sim \begin{cases} N(\mu_0, \Sigma_0), & i = 1, \qquad \dots, \tau_n \\ N(\mu_1, \Sigma_1), & i = \tau_n + 1, \dots, 51 \end{cases}$$

The change-point detection was to test the following hypothesis:

H_0: The data conforms to no change model M_0 (6.5)

Against H_1: The data conforms to change point model M_i

where $i = 1, 2, 3$.

6.2.2.1 Change-point Detection

Before the change-point analysis, the position and the nature of the change-point remain unknown for us, that is to say, we do not know whether the change happens at only one layer or multiple layers, and whether the change of parameters involves mean only, covariance (variance) only or both. Therefore, an exhaustive detection method will be applied on the dataset. We will use the maximum likelihood change-point detection method for univariate and multivariate on all the possible combinations of layers that were specified in Chapter 2 to test the 3 hypotheses in (6.5). The detected change-points and p-values of all cases are listed in Table 6.2. We will start from the 4-dimensional multivariate change-point detection for all layers. If no change is detected, it means that no change of parameters occurred in the dataset. If a statistically significant change-point is detected, then we will perform the change-point detection on all the combinations of 3-dimensional data. If the detection is not significant for a combination for 3 layers, then no change occurred in any of the layers, and they will be excluded in the detection procedure in the next step, which makes the detection more focused on the data with possible occurrences of change-point. The same detection will be applied to all possible combinations of the temperature deviations of 2 layers, excluding the ones that do not show significant change in 3-dimensional data. This procedure will continue until all the univariate data are detected. Then a conclusion can be made about which layer or combination of layers has a change of parameters at an unknown point of time. The combination of layers will be used together for inference on the change-point to gain the maximum power.

In determining the nature of the change-point, similar exhaustive steps will be applied. First, whether there is a change in both mean and/or covariance (variance) will be detected. When a significant change is detected, it means that a change in mean, covariance (variance) or both has occurred in the dataset. Then the change-point detection in mean only will be performed. If the change in mean is significant, then the residual of the data will be obtained by adjusting for the mean before and after the change-point, and the detection on change in covariance (variance) only will be applied on the residuals. If the change in mean is not significant, then no adjustment is required, and the detection on change in covariance (variance) only will be performed directly on the data. Table 6.2 shows the detection results following the above procedure. For convenience, the different layers of the atmosphere are numbered as follows: 1 represents the surface layer, 2 represents the 850 – 300 mb layer, 3 represents the 300 – 100 mb layer, and 4 represents the 100 – 50 mb layer.

Table 6.2. Change-point detection of South Polar annual mean temperature deviations during 1958 – 2008 for mean and/or covariance (variance), mean only and covariance (variance) only.

	Mean and Covariance/Variance M_3		Mean Only M_1		Covariance/Variance Only M_2	
	$\hat{\tau}_n$	p-value	$\hat{\tau}_n$	p-value	$\hat{\tau}_n$	p-value
1, 2, 3, 4	25	< 0.0001	27	0.0002	26	0.0001
1, 2, 3	25	< 0.0001	25	0.0002	4	0.0019
1, 2, 4	27	< 0.0001	27	0.0003	28	0.0034
1, 3, 4	24	< 0.0001	26	0.0009	24	0.0013
2, 3, 4	25	< 0.0001	25	0.0002	24	0.0015
1, 2	14	0.0049	19	0.0156	16	0.3983
1, 3	26	0.0001	26	0.0007	29	0.0108
1, 4	24	0.0001	27	0.0030	28	0.0114
2, 3	25	< 0.0001	25	0.0002	18	0.0159
2, 4	27	< 0.0001	27	0.0004	27	0.0166
3, 4	24	< 0.0001	26	0.0012	24	0.0033
1	8	0.0424	8	0.0291	16	0.5386
2	19	0.0116	19	0.0076	32	0.8357
3	26	0.0001	26	0.0006	29	0.0041
4	27	0.0003	27	0.0019	28	0.0097

From the table for detection, we found that all the 4-dimensional and 3-dimensional observations showed significant changes in mean and covariance. In 2-dimensional change-point detection, layers 1 and 2 only showed significant change in mean only, and all the other combinations showed significant change in both mean and covariance. That is to say, the layers 1 and 2 possibly have changes in mean only, but not in both mean and variance. The layers 3 and 4 possibly have changes in mean and/or variance. As the detection for univariate data was obtained, it verified the conclusion drawn from the detection on 2-4 dimensional data: the data for layers 3 and 4 had significant change in mean and variance, the data for layers 1 and 2 had significant change in mean only. The detected change-points were close for layers 3 and 4, and were far apart for layers 1 and 2. In the inference, in order the get the maximum power, we will perform bivariate change in mean and covariance for layers 3 and 4. Univariate change-point analysis will performed on layers 1 and 2 respectively.

6.2.2.2 Bivariate change-point analysis for layer 3 and 4

As in Table 6.2, the bivariate temperature deviations for layer 3 (300 – 100 mb) and layer 4 (100 – 50 mb) were detected to have a change in mean both mean and covariance. Twice the likelihood ratio statistics was plotted in Figure 6.8.

The statistics is maximized at the 24th observation with the p-value < 0.0001. The corresponding year of change is 1981. The change-point model followed M_3:

$$Y_i \sim \begin{cases} N(\mu_{0,24}, \Sigma_{0,24}), & i = 1, \quad \dots, 24 \\ N(\mu_{1,24}, \Sigma_{1,24}), & i = 25, \dots, 51 \end{cases} \tag{6.6}$$

Where the maximum likelihood estimate of the parameters is

$$\hat{\mu}_{0,24} = [0.0525 \quad -0.0913]^T, \qquad \hat{\mu}_{1,24} = [-1.3556 \quad -2.5626]^T \tag{6.7}$$

$$\hat{\Sigma}_{0,24} = \begin{bmatrix} 0.1069 & -0.0147 \\ -0.01475 & 0.4329 \end{bmatrix}, \qquad \hat{\Sigma}_{1,24} = \begin{bmatrix} 0.8351 & 1.4090 \\ 1.4090 & 3.4279 \end{bmatrix}$$

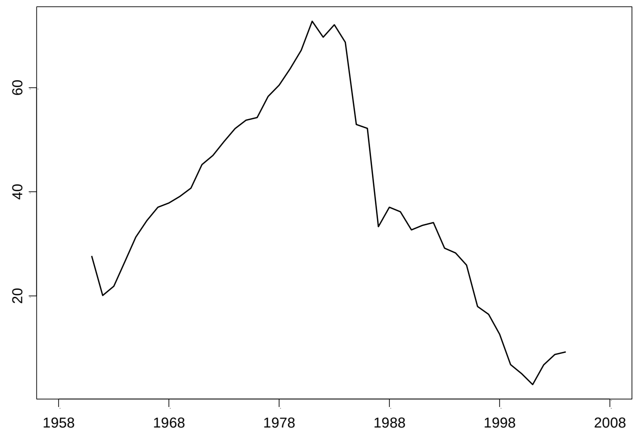

Figure 6.8. Twice the log likelihood ratio statistics for South Polar annual mean temperature deviations during 1958 – 2008 at layer 3 (850 – 300 mb) and layer 4 (100 – 50 mb).

In the last section, the significant change-points were detected in both bivariate observations. In order to determine the accuracy of the detection, the confidence interval for the change-point detection will be calculated using the multivariate maximum likelihood estimation method in Sections 3.1 and 4.1. In this section, the assumptions of maximum likelihood estimation will be tested: the bivariate observations follow multivariate normal distribution and all the observations are independent over time. If the assumptions are not satisfied, the maximum likelihood change-point analysis should not be applied on the data.

As the parameters for both sets of observations changed, each observation was adjusted by its own mean and covariance so that the residual is obtained. After the adjustment, the residuals all satisfied $\mu = 0$ and $\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. The task then is to test whether the residuals follow independent standard bivariate normal distribution.

In test the multivariate normality, the standard Shapiro-Wilk test for univariate normality of residuals for each component was applied first. The p-values are 0.4155 for layer 3 and 0.1555 for layer 4. The residuals from both layers confirm the assumption of Normality in a univariate way. However, it is well known that univariate Normality of each series may not necessarily imply that the residual data in vector form would be multivariate Normal. Thus, in order to test for multivariate Normality of the residuals, we applied Mardia’s skewness and kurtosis tests (Mardia, 1970), as well as the test proposed by Henze and Zirkler (1990). The P-values for the three tests were 0.8053, 0.0778, and 0.1140, respectively. There is no evidence otherwise in the multivariate tests that multivariate Normality assumption is in violation for the bivariate series for layers 3 and 4.

Next, we utilize the residuals for investigating the assumption of independence over time. If the model under independence is truly a good fit, then both autocorrelations and partial autocorrelations of different lags for residuals from each of the bivariate series should show no appreciable significances. Moreover, independence at the multivariate level should mean that forward as well as backward cross-correlations of different lags for residuals from any pair of rivers should also show no significances. First, we present in Figures 6.9, the autocorrelation and partial correlation plots up to the first ten lags together with the corresponding 95% significance curves, for residuals of the bivariate data for layers 3 and 4. The cross correlations of are in Table 6.3. None of the correlations are significant. Therefore, the bivariate series comprised of layers 3 and 4 temperature deviations satisfied the normality and independence assumptions.

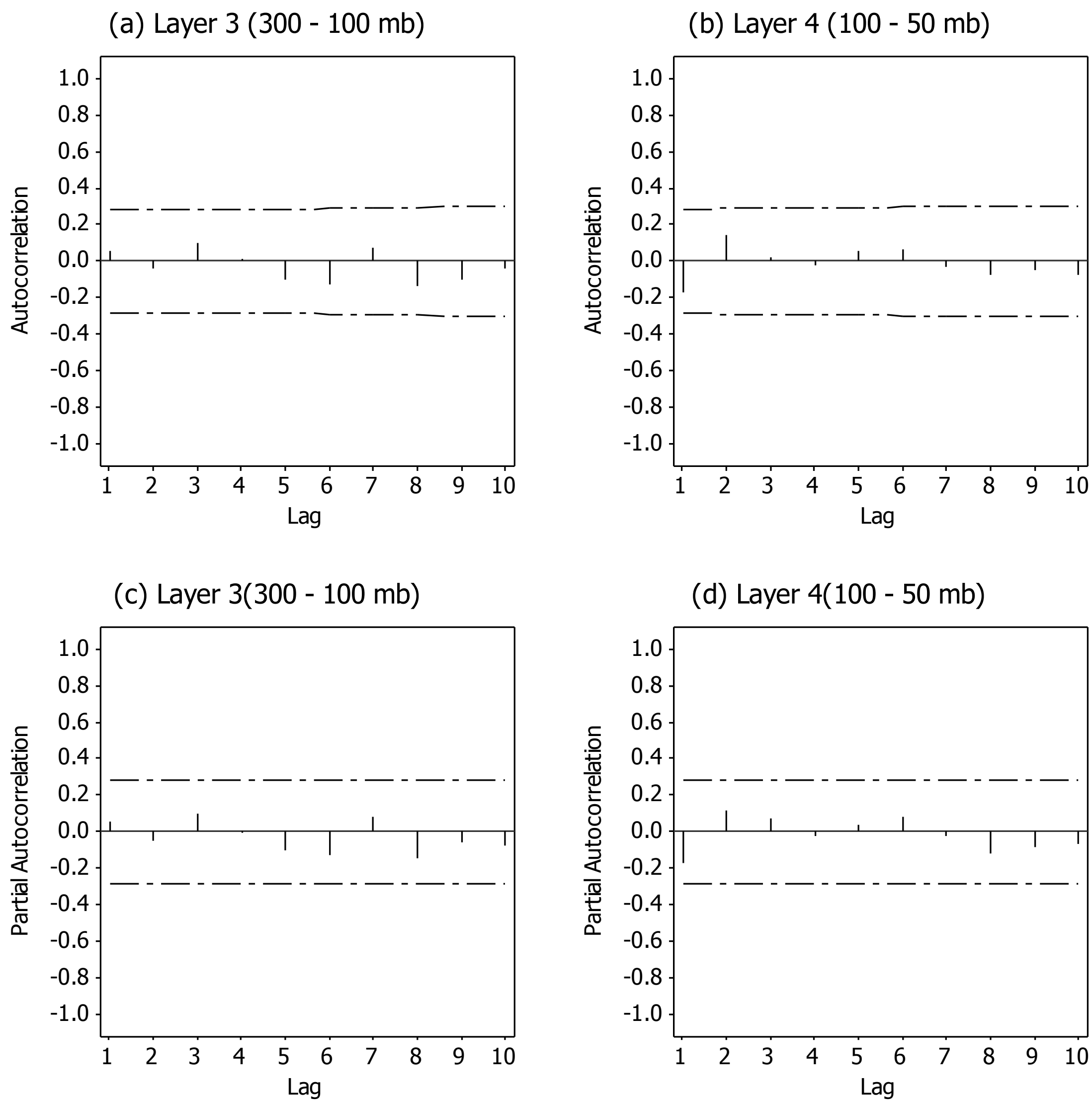

Figure 6.9. Autocorrelation and partial autocorrelation plots of residuals with 95% significant limits for South Polar annual mean temperature deviations during 1958 – 2008 at layer 3 (300 – 100 mb) and layer 4 (100 – 50 mb). (a) and (b) are correlations, and (c) and (d) are partial correlations.

Table 6.3 Cross correlations of the residuals at layers 3 and 4 for South Polar annual mean temperature deviations during 1958 – 2008.

Lag	Cross correlation
-5	0.053
-4	-0.084
-3	0.121
-2	-0.056
-1	0.105
0	-0.006
1	-0.063
2	-0.031
3	0.305
4	0.117
5	-0.046

We shall now obtain confidence interval estimate of τ_{51} through the asymptotic distribution of $\hat{\tau}_{51}$ for which, we can assume that $\hat{\mu}_{0,24}$, $\hat{\mu}_{1,24}$, $\hat{\Sigma}_{0,24}$ and $\hat{\Sigma}_{1,24}$ given above are true values rather than estimates. As discussed in Chapter 4, this is possible because the asymptotic distributions of $\hat{\tau}_{51}$ and $\tilde{\tau}_{51}$ are identical. This allowed us to compute all the parameters that are necessary for the algorithmic procedure specified in Section 4.1.3.

$$K = \hat{\Sigma}_{0,24}^{1/2}\hat{\Sigma}_{1,24}^{-1/2} = \begin{bmatrix} 0.6235 & -0.2098 \\ -0.4276 & 0.5114 \end{bmatrix}$$

$$KK^T = \Theta\Psi\Theta^T \text{ where } \Psi = diag\{0.8124, 0.0646\},\ \Theta = \begin{bmatrix} -0.7016 & -0.7126 \\ 0.7126 & -0.7016 \end{bmatrix}$$

$$\eta = \hat{\Sigma}_{0,24}^{-\frac{1}{2}}\left(\hat{\mu}_{1,24} - \hat{\mu}_{0,24}\right) = [-4.4881 \quad -3.8593]^T$$

$$\omega = \Theta^{\mathrm{T}}\eta = [0.3987 \quad 5.9058]^T$$

$$C^o = \frac{1}{2} ln(\psi_1\psi_2 \dots \psi_d) - \frac{1}{2}\sum_{i=1}^{d} \frac{\psi_i\omega_i^2}{1-\psi_i} = -3.0225$$

$$\sigma_i^{o2} = \left(\frac{\psi_i\omega_i}{1-\psi_i}\right)^2 = (2.9816 \quad 0.1664)$$

$$a_i^o = \frac{1}{2}(1-\psi_i) = (0.0938 \quad 0.4677)$$

$$\eta^* = \Sigma_1^{-\frac{1}{2}}(\mu_1 - \mu_0) = [-1.1479 \quad -1.0321]^T$$

$$\omega^* = \Theta^{\mathrm{T}}\eta^* = [0.0699 \quad 1.5421]^T$$

$$C^* = -\frac{1}{2}\ln(\psi_1\psi_2 \dots \psi_d) - \frac{1}{2}\sum_{i=1}^{d} \frac{\psi_i^{-1}\omega_i^{*2}}{1-\psi_i^{-1}} = 2.7577$$

$$\sigma_i^{*2} = \left(\frac{\psi_i^{-1}\omega_i^*}{1-\psi_i^{-1}}\right)^2 = (0.1390 \quad 2.7179)$$

$$a_i^* = \frac{1}{2}(1-\psi_i^{-1}) = (-0.1155 \quad -7.2380)$$

Among these parameters, the parameter $\delta = \frac{\sqrt{\eta^T\eta}}{2} = 2.9597$ and $|KK^T| = \frac{|\hat{\Sigma}_{0,24}|}{|\hat{\Sigma}_{1,24}|} = 0.0523$ uniquely determined the size of the change in mean and covariance matrix.

As all the parameters in Step S0 of the algorithm in Section 4.1.3, the asymptotic distribution of $\hat{\xi}_\infty$ was computed in Table 6.4, and the cumulative probabilities were computed in Table 6.5. For comparison purpose, the distribution computed by Cobb's conditional mle, Bayesian method using conjugate priors and non-informative priors were also presented in Tables 6.4 and 6.5.

Table 6.4 Computed probabilities for $\hat{\xi}_\infty$ using Maximum Likelihood, Cobb's conditional mle, and Bayesian methods using conjugate and non-informative priors for South Polar annual mean temperature deviations during 1958 – 2008.

N	Year	i	$P(\hat{\xi}_\infty = i)$			
			ML	Cobb*	Bayesian (Conjugate Prior)	Bayesian (Non-informative Prior)
17	1974	-7	0.0000		0.0000	0.0000
18	1975	-6	0.0000		0.0000	0.0000
19	1976	-5	0.0000		0.0000	0.0001
20	1977	-4	0.0002	0.0000	0.0001	0.0004
21	1978	-3	0.0008	0.0005	0.0005	0.0011
22	1979	-2	0.0041	0.0030	0.0033	0.0049
23	1980	-1	0.0266	0.0394	0.0261	0.0289
24	1981	0	0.8276	0.7442	0.4859	0.4344
25	1982	1	0.1150	0.0785	0.1055	0.1064
26	1983	2	0.0210	0.1283	0.3267	0.3505
27	1984	3	0.0044	0.0060	0.0519	0.0733
28	1985	4	0.0010	0.0000	0.0000	0.0000
29	1986	5	0.0002		0.0000	0.0000
30	1987	6	0.0001		0.0000	0.0000
31	1988	7	0.0000		0.0000	0.0000
Mean			24.1377	24.3061	24.8797	24.9834
Variance			0.3235	0.6084	1.9258	1.2706
SD			0.5688	0.7800	1.3877	1.1272

*Note: With the tolerance of error 0.0001, the number of observations for Cobb's conditional method is 4 before and after the detected change-point mle.

Table 6.5 Computed cumulative probabilities for $\hat{\xi}_\infty$ using Maximum Likelihood, Cobb's conditional mle, and Bayesian methods using conjugate and non-informative priors for South Polar annual mean temperature deviations during 1958 – 2008.

N	Year	i	$P\left(\left\|\hat{\xi}_\infty\right\| \leq i\right)$			
			ML	Cobb	Bayesian (Conjugate Prior)	Bayesian (Non-informative Prior)
24	1981	0	0.8276	0.7442	0.4859	0.4344
23 – 25	1980 – 1982	1	0.9692	0.8621	0.6175	0.5697
22 – 26	1979 – 1983	2	0.9943	0.9934	0.9475	0.9251
21 – 27	1978 – 1984	3	0.9995	0.9999	0.9999	0.9995
20 – 28	1977 – 1985	4	1.0007	0.9999	1.0000	0.9999
19 – 29	1976 – 1986	5	1.0009	0.9999	1.0000	1.0000
18 – 30	1975 – 1987	6	1.0010	0.9999	1.0000	1.0000
17 – 31	1974 – 1988	7	1.0010	0.9999	1.0000	1.0000

The 95% confidence interval is {1980, 1981, 1982} when using ML method, {1979, 1980, 1981, 1982, 1983} when using Cobb's conditional mle method, and {1978, 1979, 1980, 1981, 1982, 1983, 1984} when using Bayesian methods. For comparison purposes, we note that the standard deviation for the change point estimate under the MLE approach is smaller under the other methods. While the MLE approach seems to do better than the Bayesian approach in this case, one should not conclude that this would be the case with other data sets. There is scope for uncertainty in the asymptotic distribution of the MLE due to the fact that the mean vectors before and after the change point were assumed to be known, whereas, in reality they are estimated from data. Even though the equivalence result of (Hinkley 1972) justifies this consideration for large samples, one should be prepared for some uncertainty because of the limited nature of the sample size in the temperature deviation data.

6.2.2.3 Univariate change-point analysis for layer 1

As in Table 6.2, the univariate temperature deviations for layer 1 (surface) was detected to have a change in mean only. Twice the likelihood ratio statistics was plotted in Figure 6.10. The statistics is maximized at the 8th observation with the p-value 0.0291. The corresponding year of change is 1965. The change-point model for this dataset is

$$Y_i \sim \begin{cases} N\left(\mu_{0,8}, \sigma_8^2\right), & i = 1, \quad \dots, 8 \\ N\left(\mu_{1,8}, \sigma_8^2\right), & i = 9, \dots, 51 \end{cases}$$

Where the maximum likelihood estimator of the parameters are

$$\hat{\mu}_{0,8} = -0.375, \hat{\mu}_{1,8} = 0.4347$$

$$\hat{\sigma}_8^2 = 0.2419, \hat{\sigma} = 0.4919$$

The tests for normality and independence assumptions were applied on the residuals. The p-value for the standard Shapiro-Wilk test is 0.2329. The normality assumption was not violated. The autocorrelation and partial autocorrelations functions of the residuals for the first 10 lags were plotted in Figure 6.11. The lag 2 autocorrelation and partial autocorrelation were marginally significant. We regarded that the assumption of independence was not in violation.

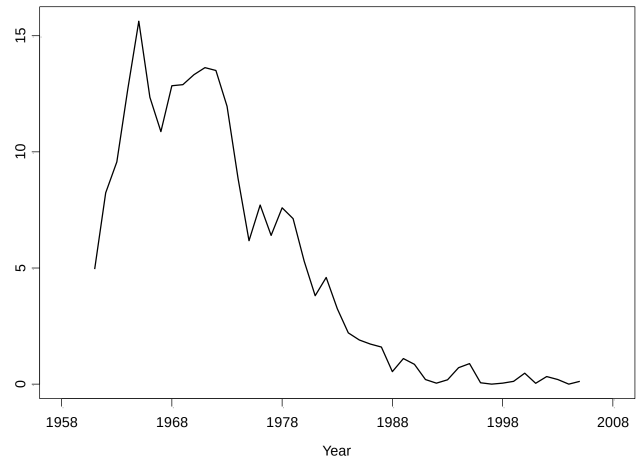

Figure 6.10. Twice the log likelihood ratio statistics for South Polar annual mean temperature deviations during 1958 – 2008 at layer 1 (surface).

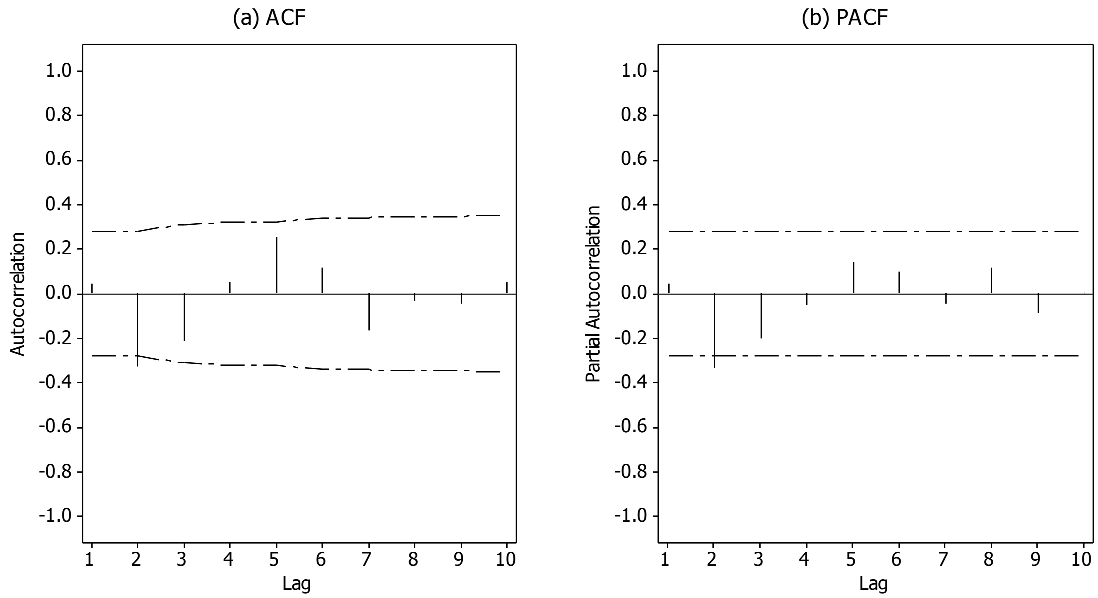

Figure 6.11. Autocorrelation and partial autocorrelation plots of residuals with 95% significant limits for South Polar annual mean temperature deviations during 1958 – 2008 at layer 1 (surface).

We shall now obtain confidence interval estimate of τ_{51} through the asymptotic distribution of $\hat{\tau}_{51}$ for which, we can assume that $\hat{\mu}_{0,8}$, $\hat{\mu}_{1,8}$, and $\hat{\sigma}_8^2$ given above are true values rather than estimates. As discussed in Chapter 3, this is possible because the asymptotic distributions of $\hat{\tau}_{51}$ and $\tilde{\tau}_{51}$ are identical. This allowed us to compute all the parameters that are necessary for the algorithmic procedure specified in Section 3.1.3. The parameter for the estimation is $\eta = \left(\hat{\mu}_{1,8} - \hat{\mu}_{0,8}\right)/\hat{\sigma}_8 = 1.6461$. The probabilities are shown in Table 6.6. When changed occurred in mean only, the asymptotic distribution for the change-point mle, $\hat{\tau}_{51}$ is symmetric about τ_{51}. The change-point mle $\hat{\xi}_\infty$ had mean equal to 0, and standard deviation equal to 1.8674. The 97% confidence interval for $\hat{\xi}_\infty$ is $\{-4, -3, -2, -1, 0, 1, 2, 3, 4\}$. As $\hat{\tau}_{51} = 8$, the 97% confidence interval for $\hat{\tau}_{51}$ included the list the years $\{1961, 1962, 1963, 1964, 1965, 1966, 1967, 1968, 1969\}$.

Table 6.6 Computed probabilities and cumulative probabilities for $\hat{\xi}_\infty$ at South Polar during 1958 – 2008 at layer 1 (surface).

i	$P(\lvert\hat{\xi}_\infty\rvert = i)$	$P(\lvert\hat{\xi}_\infty\rvert \le i)$
0	0.5272	0.5272
1	0.1288	0.7848
2	0.0533	0.8913
3	0.0261	0.9434
4	0.0139	0.9713
5	0.0078	0.9869
6	0.0046	0.9960
7	0.0027	1.0014
8	0.0017	1.0047
9	0.0010	1.0068
10	0.0006	1.0081
11	0.0004	1.0089
12	0.0003	1.0094
13	0.0002	1.0098
14	0.0001	1.0100
15	0.0001	1.0101
16	0.0000	1.0102
Mean	0	
Variance	3.4872	
SD	1.8674	

6.2.2.4 Univariate change-point analysis for layer 2

As in Table 6.2, the univariate temperature deviations for layer 2 (850 – 300 mb) was detected to have a change in mean only. Twice the likelihood ratio statistics was plotted in Figure 6.12. The statistics is maximized at the 19th observation with the p-value 0.0076. The corresponding year of change is 1965. The change-point model for this dataset is

$$Y_i \sim \begin{cases} N\left(\mu_{0,19}, \sigma_{19}^2\right), & i = 1, \quad \dots, 19 \\ N\left(\mu_{1,19}, \sigma_{19}^2\right), & i = 20, \dots, 51 \end{cases}$$

Where the maximum likelihood estimator of the parameters are

$$\hat{\mu}_{0,19} = -0.021, \hat{\mu}_{1,19} = 0.4769$$

$$\hat{\sigma}_{19}^2 = 0.1035, \hat{\sigma}_{19} = 0.3217$$

The tests for normality and independence assumptions were applied on the residuals. The p-value for the standard Shapiro-Wilk test is 0.5611. The normality assumption was not violated. The autocorrelation and partial autocorrelations functions of the residuals for the first 10 lags were plotted in Figure 6.13. None of the correlations were significant. Therefore, the independence assumption was not violated, either.

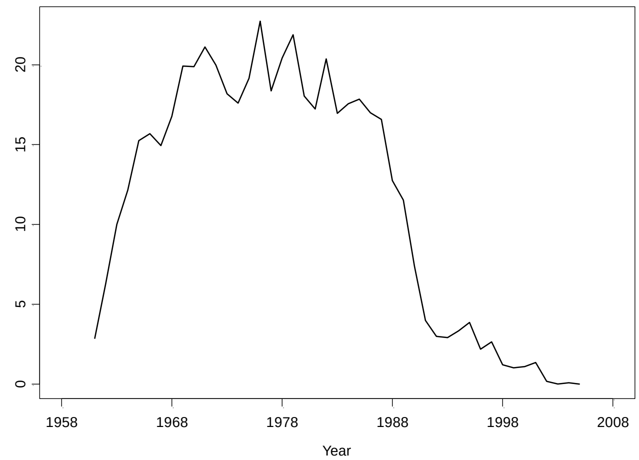

Figure 6.12. Twice the log likelihood ratio statistics for South Polar annual mean temperature deviations during 1958 – 2008 at layer 2 (850 – 300 mb).

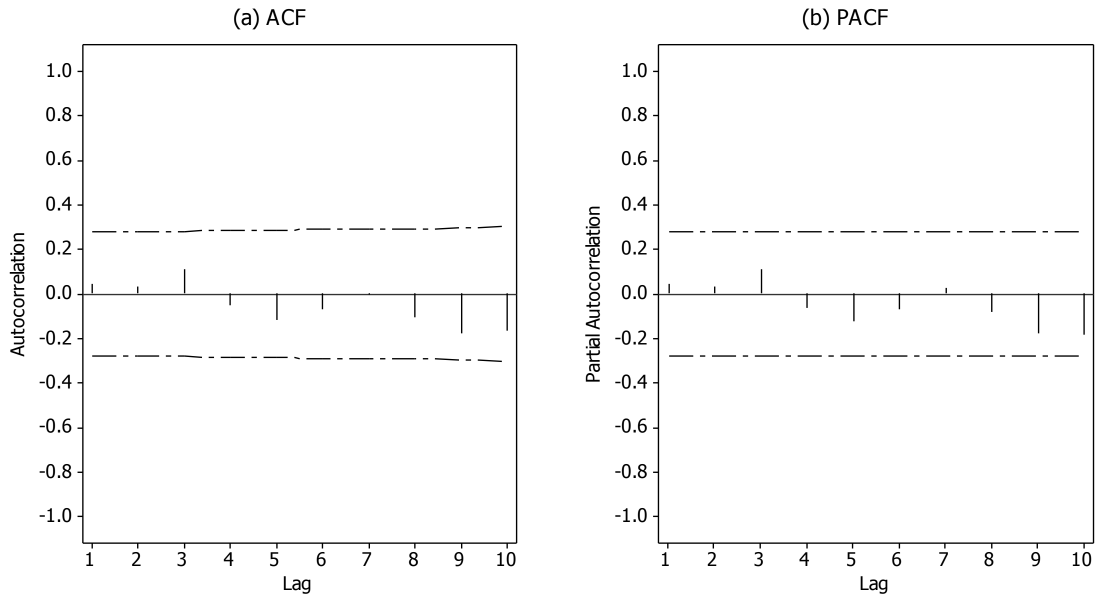

Figure 6.13. Autocorrelation and partial autocorrelation plots of residuals with 95% significant limits for South Polar annual mean temperature deviations during 1958 – 2008 at layer 2 (850 – 100 mb).

We shall now obtain confidence interval estimate of τ_{51} through the asymptotic distribution of $\hat{\xi}_\infty$ for which, we can assume that $\hat{\mu}_{0,19}$, $\hat{\mu}_{1,19}$, and $\hat{\sigma}_{19}^2$ given above are true values rather than estimates. As discussed in Chapter 3, this is possible because the asymptotic distributions of $\hat{\tau}_{51}$ and $\tilde{\tau}_{51}$ are identical. This allowed us to compute all the parameters that are necessary for the algorithmic procedure specified in Section 3.1.3. The parameter for the estimation is $\eta = (\hat{\mu}_{1,19} - \hat{\mu}_{0,19})/\hat{\sigma}_{19} = 1.5496$. The probabilities are shown in Table 6.7. When the change occurred in mean only, the asymptotic distribution for the change-point mle, $\hat{\tau}_{51}$ is symmetric about τ_{51}. The change-point mle $\hat{\xi}_\infty$ had mean equal to 0, and standard deviation equal to 2.1144. The 96% confidence interval for $\hat{\xi}_\infty$ is $\{-4, -3, -2, -1, 0, 1, 2, 3, 4\}$. As $\hat{\tau}_{51} = 19$, the 96% confidence interval for $\hat{\tau}_{51}$ included the list the years $\{1972, 1973, 1974, 1975, 1976, 1977, 1978, 1979, 1980\}$.

Table 6.7 Computed probabilities and cumulative probabilities for $\hat{\xi}_\infty$ at South Polar during 1958 – 2008 at layer 2 (850 – 100 mb).

i	$P(\lvert\hat{\xi}_\infty\rvert = i)$	$P(\lvert\hat{\xi}_\infty\rvert \leq i)$
0	0.4930	0.4930
1	0.1307	0.7545
2	0.0571	0.8686
3	0.0293	0.9273
4	0.0164	0.9601
5	0.0096	0.9793
6	0.0058	0.9910
7	0.0036	0.9983
8	0.0023	1.0029
9	0.0015	1.0059
10	0.0010	1.0078
11	0.0006	1.0091
12	0.0004	1.0100
13	0.0003	1.0105
14	0.0002	1.0109
15	0.0001	1.0112
16	0.0001	1.0114
17	0.0001	1.0115
18	0.0000	1.0116
Mean	0	
Variance	4.4708	
SD	2.1144	

6.2.3 *Change-point Analysis at North Polar*

For the data of temperature deviation at the North Polar, the same analysis procedures applied as in Section 6.2.2. The no-change model M_0 and the change-point models $\mathrm{M}_1 - \mathrm{M}_3$ were set up the same way as in Section 6.2.2.

$$M_1:\ Y_i \sim \begin{cases} N(\mu_0, \Sigma), & i = 1, \quad \dots, \tau_n \\ N(\mu_1, \Sigma), & i = \tau_n + 1, \dots, 51 \end{cases}$$

$$M_2:\ Y_i \sim \begin{cases} N(\mu, \Sigma_0), & i = 1, \quad \dots, \tau_n \\ N(\mu, \Sigma_1), & i = \tau_n + 1, \dots, 51 \end{cases}$$

$$M_3:\ Y_i \sim \begin{cases} N(\mu_0, \Sigma_0), & i = 1, \quad \dots, \tau_n \\ N(\mu_1, \Sigma_1), & i = \tau_n + 1, \dots, 51 \end{cases}$$

The hypothesis test for the change-point was

H_0: The data conforms to no change model M_0

Against H_1: The data conforms to change point model M_i

where $i = 1, 2, 3$.

First, the exhaustive change-point detection was applied to all single and combinations of series for the 4 layers of atmosphere: layer 1 for surface, layer 2 for 850 – 300 mb, layer 3 for 300 – 100 mb and layer 4 for 100 – 50 mb. All the detection results are available in Table 6.8. As the change-point detection was applied from the high-dimensional data down to the univariate data, we found that significant change in both mean and covariance matrix was detected for the 4-dimensional data. As we detected the 3-dimensional data, the series comprised of the temperature deviation from layers 1, 2 and 3 had change in mean only, but not in covariance matrix. Therefore, the temperature deviations from these 3 layers possibly had change in mean, but no change occurred in the variances. In the change-point detection in 2-dimensional data, significant change in mean only were detected in all series except the combination of layers 3 and 4, which showed significant change in both mean and covariance matrix, although the change in covariance was marginal. As we proceeded to the univariate cases, the temperature deviations for layer 3 did not have parameter change, and those of layers 1, 3 and 4 had change in mean only. The p-value of layer 3 for the change in mean only showed marginal change. Therefore we only considered that layers 1 and 4 had real significant change in mean only. The change-point analysis would be applied on the bivariate series comprised of the temperature deviation of layers 1 and 4.

Table 6.8. Change-point detection of North Polar annual mean temperature deviations during 1958 – 2008 for mean and/or covariance (variance), mean only and covariance (variance) only. 1 = Surface, 2 = 850 – 300 mb, 3= 300 – 100 mb, 4=100 – 50 mb.

	Mean and Covariance/Variance		Mean Only		Covariance/Variance Only	
	$\hat{\tau}_n$	p-value	$\hat{\tau}_n$	p-value	$\hat{\tau}_n$	p-value
1, 2, 3, 4	44	<0.0001	37	0.0015	5	0.0008
1, 2, 3	31	0.0001	31	0.0016	47	0.1199
1, 2, 4	31	0.0001	31	0.0015	43	0.0634
1, 3, 4	44	<0.0001	37	0.0015	44	0.0131
2, 3, 4	44	0.0001	38	0.0078	10	0.0229
1, 2	31	0.0007	31	0.0013	47	0.2807
1, 3	47	0.0010	31	0.0018	47	0.3055
1, 4	47	0.0007	31	0.0016	43	0.1937
2, 3	32	0.0297	30	0.0397	33	0.9792
2, 4	37	0.0034	37	0.0075	5	0.2457
3, 4	38	0.0056	38	0.0419	12	0.0425
1	31	0.0015	31	0.0011	47	0.1468
2	44	0.0421	44	0.0406	4	0.3762
3	30	0.1215	30	0.0637	32	0.7495
4	32	0.0390	32	0.0222	10	0.1043

As in Table 6.8, the bivariate temperature deviations for layer 1 (surface) and layer 4 (100 – 50 mb) was detected to have a change in mean only. The detected change-point under univariate analysis was close to each other. Therefore, the temperature deviations for layer 1 and 4 were analyzed using bivariate change-point analysis for mean only. Twice of the log likelihood function was shown in Figure 6.14. The statistics is maximized at the 31th observation with the p-value 0.0016. The corresponding year of change is 1988. The change-point model for this dataset is

$$Y_i \sim \begin{cases} f(\cdot; \mu_{0,31}, \Sigma_{31}), & i = 1, \quad \dots, 31 \\ f(\cdot; \mu_{1,31}, \Sigma_{31}), & i = 32, \dots, 51 \end{cases} \tag{6.8}$$

Where the maximum likelihood estimator of the parameters are

$$\hat{\mu}_{0,31} = [0.0013 \quad -0.1152]^T, \qquad \hat{\mu}_{1,31} = [1.0655 \quad -1.4585]^T \tag{6.9}$$

$$\hat{\Sigma}_{31} = \begin{bmatrix} 0.2739 & -0.1535 \\ -0.1535 & 1.2028 \end{bmatrix}$$

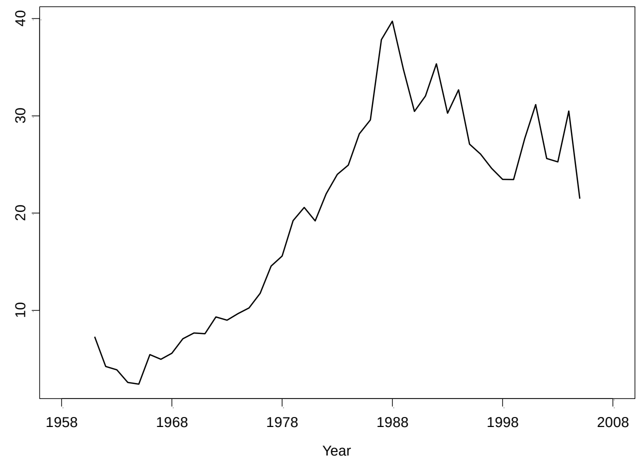

Figure 6.14. Twice log likelihood ratio statistics for North Polar annual mean temperature deviations during 1958 – 2008 at layer 1 (surface) and layer 4 (100 – 50 mb).

In change-point mle the estimated parameters were regarded as the true parameters of the observations. The assumptions for change-point mle shall be verified before the change-point estimation. The residuals are obtained by adjusting each observation by their mean and covariance matrix depending whether it was before or after the change-point estimate $\hat{\tau}_n = 31$.

The p-values of Shapiro-Wilk test were 0.1871 and 0.9399 for layer 1 and 4 respectively. The p-values of the multivariate normality test, Mardia's skewness and kurtosis tests, and Henze- Zirkler test, are 0.9767, 0.8181, and 0.6089, respectively. The normality assumption was not violated. The autocorrelation and partial autocorrelation were plotted in Figure 6.15. The cross correlation for lags -5 to 5 were shown in Table 6.9. The autocorrelations and partial autocorrelations for the temperature deviations at layer 1 and 4 were not significant. Even though the cross correlations showed marginal significant at lags 1, 4 and 5, the layers 1 and 4 are quite separated. Therefore, we still assume that the assumption of independence was not violated.

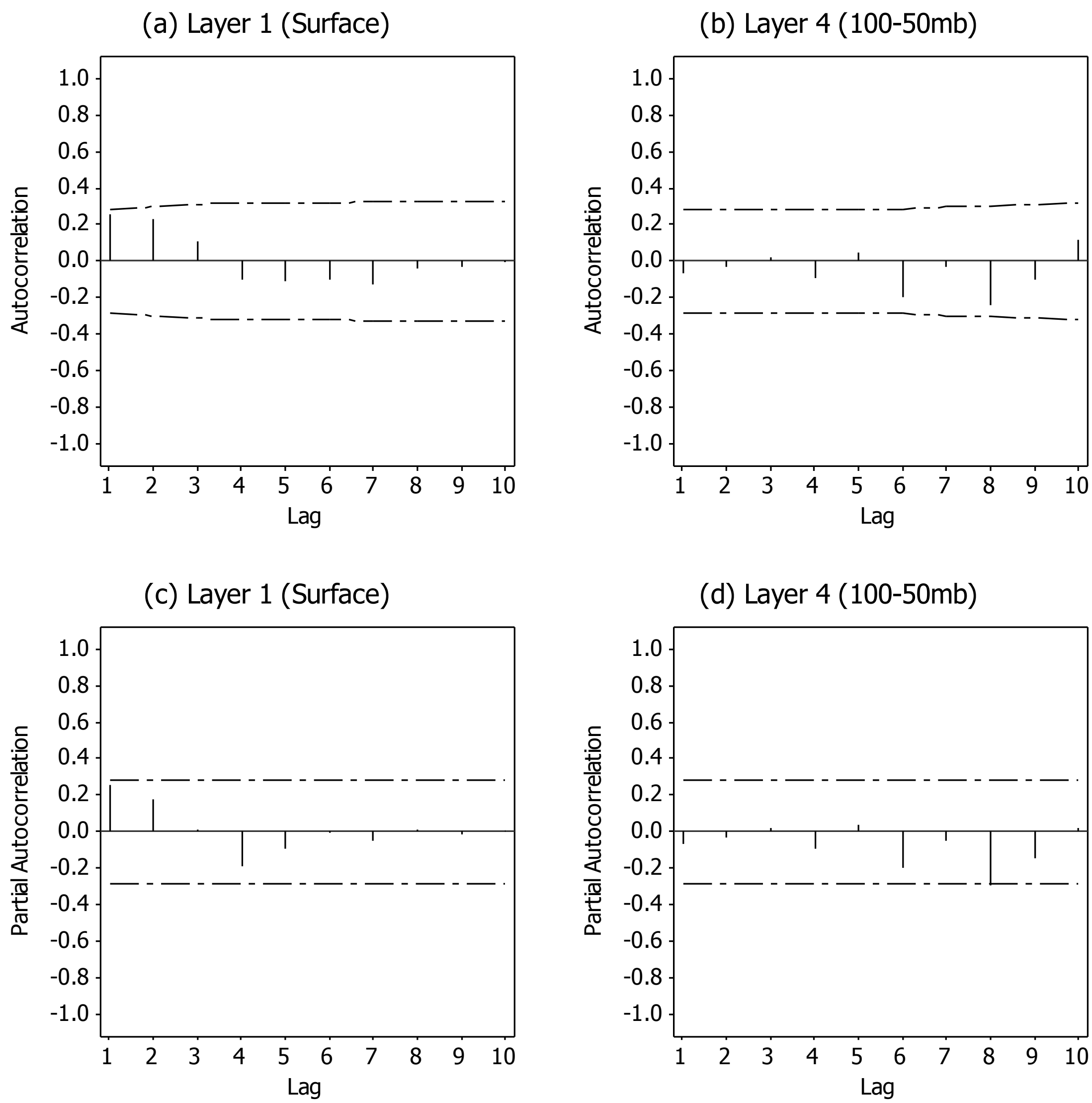

Figure 6.15. Autocorrelation and partial autocorrelation plots of residuals with 95% significant limits for North Polar annual mean temperature deviations during 1958 – 2008 at layer 1 (surface) and layer 4 (100 – 50 mb). (a) and (b) are autocorrelations, and (c) and (d) are partial autocorrelations.

Table 6.9 Cross correlations for the residuals at layers 1 and 4 for North Polar annul mean temperature deviations during 1958 – 2008.

Lag	Cross Correlation
-5	-0.333
-4	0.363
-3	0.061
-2	0.076
-1	0.334
0	0
1	0.228
2	-0.152
3	-0.036
4	-0.009
5	-0.068

When only change in mean was detected in the bivariate data, the algorithmic procedure in Section 3.1 should be followed. The parameters for the procedure were calculated as follows

$$\eta = \hat{\Sigma}_{31}^{-\frac{1}{2}} \left(\hat{\mu}_{1,31} - \hat{\mu}_{0,31}\right) = [1.8704 \quad -1.0660]^T$$

$$\delta = \frac{1}{2}\sqrt{\eta^T \eta} = 1.0764$$

The asymptotic distribution of $\hat{\xi}_\infty$ can be computed as in Tables 6.10 and 6.11. When the change occurred in mean only, the asymptotic distribution for $\hat{\tau}_{51}$ is symmetric about τ_{51}, hence the mean for $\hat{\xi}_\infty$ is 0. The standard deviation for $\hat{\xi}_\infty$ was 1.1304. The 96% confidence interval for $\hat{\tau}_{51}$ included 2 years before and after the 31st observation, which was {1986, 1987, 1988, 1989, 1990}. The 98% confidence interval included 3 years before and after the 31st observation, which was {1985, 1986, 1987, 1988, 1989, 1990, 1991}.

Table 6.10 Computed probabilities of $\hat{\xi}_\infty$ using Maximum Likelihood, Cobb's conditional mle, and Bayesian methods using conjugate and non-informative priors for North Polar annual mean temperature deviations at layers 1 and 4 .

N	Year	i	$P(\hat{\xi}_\infty = i)$			
			ML	Cobb*	Bayesian (Conjugate Prior)	Bayesian (Non-informative Prior)
21	1978	-10	0.0000		0.0000	0.0000
22	1979	-9	0.0001		0.0000	0.0000
23	1980	-8	0.0002	0.0000	0.0000	0.0001
24	1981	-7	0.0004	0.0000	0.0000	0.0000
25	1982	-6	0.0009	0.0000	0.0001	0.0002
26	1983	-5	0.0021	0.0000	0.0001	0.0002
27	1984	-4	0.0049	0.0000	0.0001	0.0002
28	1985	-3	0.0120	0.0006	0.0017	0.0023
29	1986	-2	0.0325	0.0027	0.0043	0.0054
30	1987	-1	0.1073	0.3177	0.2748	0.2414
31	1988	0	0.6838	0.4705	0.3959	0.3382
32	1989	1	0.1073	0.1245	0.1203	0.1138
33	1990	2	0.0325	0.0048	0.0064	0.0077
34	1991	3	0.0120	0.0045	0.0072	0.0087
35	1992	4	0.0049	0.0612	0.1156	0.1174
36	1993	5	0.0021	0.0026	0.0105	0.0133
37	1994	6	0.0009	0.0106	0.0528	0.0625
38	1995	7	0.0004	0.0003	0.0031	0.0047
39	1996	8	0.0002	0.0000	0.0011	0.0020
40	1997	9	0.0001		0.0003	0.0006
41	1998	10	0.0000		0.0002	0.0005
Mean			31	31.1467	31.8041	33.1043
Variance			1.1304	1.9415	5.8473	20.3538
SD			1.0632	1.3934	2.4181	4.5115

*Note: With the tolerance of error 0.0001, the number of observations for Cobb's conditional method is 8 before and after the detected change-point mle.

Table 6.11 Computed cumulative probabilities for $\hat{\xi}_\infty$ using Maximum Likelihood, Cobb's conditional mle, and Bayesian methods using conjugate and non-informative priors for North Polar annual mean temperature deviations at layers 1 (surface) and layer 4 (100 – 50 mb) during 1958 – 2008.

N	Year	i	$P(\lvert\hat{\xi}_\infty\rvert \leq i)$			
			ML	Cobb	Bayesian (Conjugate Prior)	Bayesian (Non-informative Prior)
31	1988	0	0.6838	0.4705	0.3959	0.3382
30 – 32	1987 – 1989	1	0.8984	0.9127	0.7910	0.6934
29 – 33	1986 – 1990	2	0.9634	0.9202	0.8017	0.7065
28 – 34	1985 – 1991	3	0.9874	0.9253	0.8106	0.7175
27 – 35	1984 – 1992	4	0.9972	0.9865	0.9263	0.8351
26 – 36	1983 – 1993	5	1.0014	0.9891	0.9369	0.8486
25 – 37	1982 – 1994	6	1.0032	0.9997	0.9898	0.9113
24 – 38	1981 – 1995	7	1.0040	1.0000	0.9929	0.9160
23 – 39	1980 – 1996	8	1.0044	1.0000	0.9940	0.9181
22 – 40	1979 – 1997	9	1.0046	1.0000	0.9943	0.9187
21 – 41	1978 – 1998	10	1.0046	1.0000	0.9945	0.9192

6.2.4 Discussion about Polar Temperature Deviations

From the change-point detection and estimation for North and South polar temperature deviations at the 4 layers of the atmosphere, we found that at the south polar, a cooling effect and increased variations occurred around 1981 at the lower stratosphere. At the surface and lower troposphere layer, there was a slight temperature increment. At the north polar, there was a cooling effect at the lower stratosphere, and an increased temperature at the surface that happened at around 1988.

The cooling of the lower stratosphere temperature had been discovered since 1980 according to Angell (1986), Randel and Wu (1999), Compagnucci et al (2000), Ramaswamy et al. (2001), Schleip et al (2009). Angell (1986) analyzed the same dataset as in our study. In Angell's (1986) study, the data was divided into subintervals as 1960 –85, 1965 – 85, 1970 – 85, 75 – 85, and linear regression was applied to each subintervals. He found significant cooling effect in both South and North Hemisphere at the tropopause (300 – 100 mb) layer and the lower stratosphere layer (100 – 50 mb), and he concluded that the cooling effect was more pronounced in South Hemisphere than North. However, the piecewise linear regression methods could not tell when the change had occurred. Randel and Wu (1999) noticed strong cooling of lower stratosphere since approximately 1985, which was maximized in spring (October – December), and the cooling of the Arctic lower stratosphere occurred in 1990s. The conclusion was drawn by observing the fitted curve to the monthly time series data. No formal test was applied to the change-point. They studied on Radiosonde, NCEP reanalysis and satellite data, and found good overall agreement under comparison.

Compagnucci et al (2000) studied the lower stratosphere temperature derived from soundings that were made the Microwave Sounding Unit (MSU), which was regarded as satellite data. But the data was only available since 1979. The principle component method was applied on the monthly time series data. They also discovered the cooling of the lower stratosphere, which was largest over Antarctica. Ramaswamy (2001) found consistent cooling of lower stratosphere over 1979 – 1994 using different data source, including radiosonde, satellite and rocketsonde. He observed substantial cooling in the lower stratosphere during winter/spring time at Antarctica since about early 1980s, and at Arctic since the 1990s. The radiosonde records prior to 1980 showed little cooling effect. The findings were also obtained by observing various time series plots. Schleip et al (2009) applied Bayesian analysis on radiosonde data for global annual mean lower stratosphere temperature anomalies using linear model. The rate of cooling during 1979 – 2004 was detected to be much greater than the period during 1958 – 1978. In this sense, there was a general agreement that there the cooling effect occurred to south polar around early 1980s, and to the north polar around 1990s. The cooling was more prominent at the south polar than at the north polar. Although the both the radiosonde and satellite data contained uncertainty, they produced similar results. As the record of radiosonde data was longer that the satellite data, thus it was adopted in our study for a more powerful analysis of the data.

The surface of both south and north polar, and the troposphere (850 – 300 mb) layer of the south polar, showed a warming effect over the years. This had been recognized by several studies, too. Angell's (1986) study on 1960 – 1985 radiosonde data showed that the surface and the troposphere had warmed. Randel and Wu (1999) observed the warming of upper troposphere (500 – 300 mb) during midwinter at most individual stations at south polar, which matched our detection of change in mean at layer 2 (850 – 300 mb) at south polar only. Angell's (1999) study on the radiosonde data during 1958 – 1998 detected the warming of surface on both northern and southern hemisphere. The surface of the northern hemisphere had warmed more than southern hemisphere, but in the troposphere layer (850 – 300 mb) the warming was greater in southern hemisphere than northern one. Comiso (2003) studied the satellite thermal infrared data on surface temperature at arctic, and discovered sustained warmings from 1988. Both the monthly data and annual mean was studied, and they produced very similar results. Karcher et al (2003) studied the water temperature of the Atlantic Ocean in the central Arctic during 1979 – 1999, and found warming since 1991. Schleip et al (2009) detected high change point probability around 1985 and 1995 for the warming from the surface up to the tropopause layer using the Bayesian approach.

Many studies regarded ozone depletion since 1980s as the major factor for the cooling of the lower stratosphere (Angell 1986, Randel and Wu 1999, Ramaswamy et al 2001, Steinbrecht et al 2003, Cagnazzo et al 2006). Forster et al. (2007) explained that the cooling was due to the decrease absorption of longwave radiation from the reduced ozone level. Ramaswamy's (2001) survey also pointed out that some studies had shown that the cooling of the lower stratosphere at Southern Hemisphere showed more obvious cooling effect than the Northern Hemisphere. Solomon et al (2007) pointed out that the depth and frequency of the ozone depletion in the Arctic was far less than that in the Antarctic. That explained why the cooling at the South polar was more prominent than at the North polar.

Besideds ozone depletion, Ramaswamy (2001) also stated that the green house gases, carbon dioxide(CO_2), not only warmed the surface, but also affected the temperature of the lower stratosphere. The CO_2 enhanced the thermal emission from above layers, and retained the heat close to the surface, which caused the warming of the lower layer, and cooling at the upper layers.

From the review above, our detection well matches the findings. In the past, people generally compare time series plots over time, or applied linear regression on short-term data. The findings were about the change in the mean temperatures over the years. Our change-point analysis provided strict change-point estimation and confidence intervals. The insight to the position of the change-point can help climatologist to investigate exactly what have caused the change in the temperature, and what people can do to stop the trend of global warming. Our change-point analysis of the layer 3 (300 – 100 mb) and 4 (100 – 50 mb) also detected significant change in the variance and covariance, which was not mentioned in the literatures. The reasons why the temperature fluctuated more than before, and why the covariance also changed, could be a question to climatologists for further investigations.

BIBLIOGRAPHY

Andrews, D. W. K. and Ploberger, W., (1994), Optimal tests when a nuisance parameter is present only under the alternative, *Econometrica*, 62, 1383-1414.

Angell, J. K. (1986), Annual and seasonal global temperature changes in the troposphere and low stratosphere, 1960-1985, *Monthly Weather Review*, 114, 1922.

Angell, J. K. (1999), Comparison of surface and tropospheric temperature trends estimated from a 63-Station Radiosonde Network, 1958–1998, *Geophys. Res. Lett.*, 26(17), 2761–2764.

Angell, J.K. (2009), Global, hemispheric, and zonal temperature deviations derived from radiosonde records, *In Trends Online: A Compendium of Data on Global Change. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee, U.S.A*. doi: 10.3334/CDIAC/cli.005.

Barry, D., and Hartigan, J. A. (1993), A Bayesian Analysis for Change Point Problems, *Journal of the American Statistical Association*, 88, 421, 309-319.

Braun, J. V. and Muller, H. G. (1998), Statistical methods for DNA sequence segmentation, *Statist. Sci.* 13 142-162.

Cagnazzo, C., Claud, Chantal. and Hare, S., (2006), Aspects of stratospheric long-term changes induced by ozone depletion, *Climate Dynamics*, 27, 101-111.

Chen, J., and A. K. Gupta (2000), *Parametric change point analysis*, Birkhäuser, New York.

Chu, J. C., and White, H. (1992), A Direct Test for Changing Trend, *Journal of Business & Economic Statistic,* 10, 3, 289-299.

Cobb, G (1978), The problem of the Nile: conditional solution to a changepoint problem, *Biometrika*, 65, 243-251.

Csörgő, M., and L. Horváth (1997), *Limit Theorems in Change-Point Analysis*, Wiley, Chichester, New York.

Davies, R. B. (1973), Numerical inversion of a characteristic function, *Biometrika*, 60, 415-417.

Davies, R. B. (1980), Algorithm AS 155: the distribution of a linear combination of χ^2 random variables, *Appl. Statist.*, 29, 323-333.

Degaetano, A. T., (2006), Attributes of several methods for detecting discontinuities in temperature series: Prospects for a hybrid homogenization procedure, *J. Climate*, 9, 1646-1660.

Hansen, B. E., (2000), Testing for structural change in conditional models, *J. Econometr.*, 97, 93-115.

Hinkley, D. V. (1970), Inference about the change-point in a sequence of random variables, *Biometrika*, 57, 1-17.

Hinkley, D. V. (1971), Inference about the change-point from cumulative sum tests, *Biometrika*, 58, 509-523.

Hinkley, D. V. (1972), Time order classification, *Biometrika*, 59, 509-523.

Worsley, K. J., and M. S. Srivastava (1986), Likelihood ratio tests for a change in the multivariate mean, *J. Amer. Statist. Assoc*., 81, 199-204.

Fealy, R. and Sweeney, J., (2005), Detection of a possible change point in atmospheric variability in the North Atlantic and its effect on Scandinavian glacier mass balance, *Int. J. Climatol*. 25 1819-1833.

Farebrother, R. W. (1990), The Distribution of a Quadratic Form in Normal Variables, *Applied Statistics,* 39, 294-309.

Fearnhead, P. (2005), Exact Bayesian curve fitting and signal segmentation, *IEEE Trans. Signal Process*., 53, 2160-2166.

Fearnhead, P. (2006), Exact and efficient Bayesian inference for multiple change-point problems, *Stat. Comput*., 16, 203-213.

Forster, P. M., G. Bodeker, R. Schofield, S. Solomon, and D. Thompson (2007), Effects of ozone cooling in the tropical lower stratosphere and upper troposphere, *Geophys. Res. Lett.*, 34, L23813.

Fotopoulos, S. B., and V. K. Jandhyala (2001), Maximum Likelihood Estimation of a Change Point for Exponentially Distributed Random Variables, *Stat. Prob. Letrs.*, 51, 423-429.

Fotopoulos, S. B., and V. K. Jandhyala (2007), On Hinkley's estimator: Inference about the change-point, *Stat. Prob. Letrs.*, 77, 1449-1458.

Fotopoulos, S. B., V. K. Jandhyala, and E. Khapalova (2009), Exact Asymptotic Distribution of Change-point MLE for Change in the Mean of Gaussian Sequences, *Annals of Applied Statistics,* in press.

Gil-Pelaez, J (1951), Note on the inversion theorem, *Biometrika*, 38, 481-482.

Hawkins D. M. (1977), Testing a Sequence of Observations for a Shift in Location, *Journal of the American Statistical Association*, 72, 357, 180-186.

Henderson, R (1990), A Problem with the Likelihood Ratio Test for a Change-Point Hazard Rate Model, *Biometrika*, 77, 4, 835-843.

Henze, N., and B. Zirkler (1990), A class of invariant and consistent tests for multivariate normality, *Com. Stat. Theo. Meth*., 19, 3595-3617.

Hinkley, D. V. (1970), Inference about the change-point in a sequence of random variables, *Biometrika*, 57, 1-17.

Hinkley, D. V. (1971), Inference about the change-point from cumulative sum tests, *Biometrika*, 58, 509-523.

Hinkley, D. V. (1972), Time order classification, *Biometrika*, 59, 509-523.

Imhof, J.P. (1961), Computing the distribution of a quadratic form in normal variables, *Biometrika*, 48, 419-426

James, B., James K. L. and Siegnumd, D. (1987), Tests for a Change-Point, *Biometrika,* 74, 1, 71-83.

Jandhyala, V. K., and S. B. Fotopoulos (1999), Capturing the distributional behavior of the maximum likelihood estimator of a change-point, *Biometrika*, 86, 129-140.

Jandhyala, V. K., S. B. Fotopoulos, and N. Evaggelopoulos (1999), Change-point Methods for Weibull Models with Applications to Detection of Trends in Extreme Temperatures. *Environmetrics*, 10, 547-564.

Jandhyala, V. K., and S. B. Fotopoulos (2001), Rate of convergence of the maximum likelihood estimate of a change-point, *Sankhyă Ser.A,* 63, 277-285.

Jandhyala, V. K., S. B. Fotopoulos, and D. M. Hawkins (2002), Detection and Estimation of Abrupt Changes in the Variability of the Process, *Comp. Stat. Data Anal.*, 40, 1-19.

Jandhyala, V. K., S. B. Fotopoulos, and D. M. Hawkins (2006), On the inconsistency of the change-point estimator for the NE family, *Metrika*, 63, 309-315.

Jarušková, D. (1996), Change-point measurement in meteorological measurement, *Mon. Wea. Rev*., 124, 1535-1543.

Jarušková, D. (1997), Some problems with application of change-point detection methods to environmental data. *Environmetrics,* 8, 469–483.

Jarušková, D., and M. Rencová (2007), Analysis of annual maximal and minimal temperatures for some European cities by change point methods, *Environmetrics*.

Kaplan, A. Y. and Shishkin, S. L., (2000), Application of the change-point analysis to the investigation of the brain's electrical activity, *In B. E. Brodsky and B. S. Darkhovsky, editors, Non-Parametric Statistical Diagnosis : Problems and Methods*, 333–388.

Kim, H-J., and Siegnumd, D. (1989), The Likelihood Ratio Test for a Change-Point in Simple Linear Regression, *Biometrika*, 76, 3, 409-423.

Koerts, J. and Abrahamse, A. P. J. (1969) *On the theory and application of the general linear model.* Rotterdam: Rotterdam University Press.

Kuonen, D. (1999), Saddlepoint Approximations for Distributions of Quadratic Forms in Normal Variables, *Biometrika*, 86, 929-935

Kuonen, D. (2003), Numerical integration in S-Plus or R: A survey, *Journal of Statistical Software*, 8, 1–14.

Lai, T. L. (1995), Sequential change-point detection in quality control and dynamical systems, J. Roy. *Statist. Soc. B.*, 57 613-658.

Lebarbier, L. (2005), Detecting multiple change-points in the mean of Gaussian process by model selection, *Sign. Proc.*, 85, 717-736.

Lu, Z. (2006), The numerical evaluation of the probability density function of a quadratic form in normal variables, *Computational Statistics and Data Analysis*, 51.

Mardia, K.V. (1970), Measures of multivariate skewness and kurtosis with applications, Biometrika, 57, 519-530.

McCutcheon, S. (1991), *Electric Rivers: The Story of the James Bay* Project, Black Rose Books Ltd.

Perreault, L., M. Haché M. Slivitzky, B. Bobée (2000), Detection of changes in precipitation and runoff over eastern Canada and U.S. using Bayesian approach, *Stoch. Environ. Res. Risk Asses*., 13, 201-216.

Perreault, L., É. Parent, J. Bernier, B. Bobée, and M. Slivitzky (2000), Retrospective multivariate Bayesian change-point analysis: A simultaneous single change in the mean of several hydrological sequences, *Stoch. Environ. Res. Risk Asses*., 14, 243-261.

Perreault, L., J. Bernier, B. Bobée, and É. Parent (2000a), Bayesian change-point analysis in hydrometeorological time series 1, Part 1, *J. Hydrol*., 235, 221-241.

Perreault, L., J. Bernier, B. Bobée, and É. Parent (2000a), Bayesian change-point analysis in hydrometeorological time series 2, Part 2, *J. Hydrol*., 235, 242-263.

Ramaswamy, V. Chanin, M.-L. Angell, J. Barnett, J. Gaffen, D.; Gelman, M. Keckhut, P. Koshelkov, Y. Labitzke, K. Lin, J.-J. R, (2001), Stratospheric Temperature Trends: Observations and Model Simulations, Rev. Geophys., 39, 71-122.

Randel, W. J. Wu, F., (1999), Cooling of the Arctic and Antarctic Polar Stratospheres due to Ozone Depletion, *Journal of Climate*, 12, 1467-1479

Ruggieri, E., Herbert, T., Lawrance, K. T. and Lawrance, C. E. (2009). Change point method for detecting regime shifts in paleoclimatic time series: Application to δ18O time series of the Plio-Pleistocene, Paleoceanography, 24, PA1204.

Schleip, C., Menzel, A. and Dose, V., (2009), Bayesian analysis of changes in Radiosonde Atmospheric Temperature, *Int. J. of Climatology*, 29, 629-641.

Seidou, O., J. J. Asselin, and T. B. M. J. Ouarda (2007), Bayesian multivariate linear regression with application to changepoint models in hydrometeorological variables, *Wat. Res. Res*..

Seidou, O., and T. B. M. J. Ouarda (2007), Recursion-based multiple changepoint detection in multivariate linear regression and application to river streamflows, *Wat. Res. Res*..

Solomon, S., Portmann, R. W., Thompson, D. W. J. (2007), Contrasts between Antarctic and Arctic ozone depletion, *Proceedings of the National Academy of Sciences*, 104, 445-449.

Spitzer, F. (1976), *Principles of Random Walks*. New York: Springer-Verlag.

Worsley, K. J., and M. S. Srivastava (1986), Likelihood ratio tests for a change in the multivariate mean, *J. Amer. Statist. Assoc.*, 81, 199-204.

Worsley J. (1988), Exact Percentage Points of the Likelihood-Ratio Test for a Change-Point Hazard-Rate Model, *Biometrics*, 44, 1, 259-263.

Wu, Q-Z, Cheng, H-Y. and Jeng, B-S. (2005), Motion detection via change-point detection for cumulative histograms of ratio images, *Pattrn. Recog. Lett.*, 26, 555-563.

Zou, C., Qiu, P. and Hawkins, D. (2009), Nonparametric control chart for monitoring profiles using change point formulation and adaptive smoothing, *Statist. Sini.*, 19, 1337—1357.

APPENDIX

A. Average Spring stream flows during 1957-1995 in the Northern Québec Labrador region

Year	Romaine	Churchill Falls	Manicou-agan	Outardes	Sainte-Marguerite	À la Baleine
195	28.4	25.5	30.7	26.3	29.3	
195	33.4	29.5	35.9	36.9	28.1	
195	26.6	26.2	27.2	25.9	25.0	
196	24.9	20.7	26.6	25.4	23.1	
196	25.6	19.6	21.1	18.5	20.6	
196	23.7	18.4	23.9	20.9	18.6	
196	31.0	19.9	23.3	22.8	28.2	21.8
196	27.7	21.8	29.8	27.7	29.6	21.4
196	32.9	26.9	25.4	23.6	25.3	21.2
196	32.9	28.7	27.3	25.8	28.1	26.0
196	22.2	19.4	23.4	20.1	23.2	14.4
196	27.4	25.2	22.0	19.5	30.6	23.8
196	32.6	28.9	25.1	21.9	26.9	23.3
197	32.9	27.4	27.7	29.3	29.8	23.6
197	29.9	27.5	20.4	20.9	27.4	23.6
197	31.1	21.9	26.7	21.6	24.9	18.7
197	30.8	24.0	33.0	32.6	32.0	19.9
197	25.6	22.9	30.0	30.2	34.7	20.7
197	26.7	26.6	27.2	23.5	26.0	24.1
197	28.8	28.4	33.0	31.1	32.9	21.1
197	31.8	26.4	31.1	26.9	31.7	24.3
197	31.7	28.1	35.1	29.2	32.9	19.7
197	30.2	34.5	39.5	33.8	37.3	17.4
198	31.8	29.0	26.8	24.9	34.9	25.2
198	40.7	31.5	38.3	35.7	37.5	28.8
198	29.8	25.7	29.4	30.8	25.6	22.0
198	37.9	30.4	37.8	37.6	41.3	22.7
198	34.0	26.6	29.7	28.9	37.8	21.3
198	21.3	19.0	25.6	23.0	24.5	19.3
198	21.9	24.0	29.0	24.5	30.6	27.0
198	19.8	20.3	26.1	19.8	25.6	18.3
198	29.1	21.2	23.3	19.9	29.1	21.1
198	23.9	17.4	21.7	19.4	19.3	13.9
199	19.1	19.3	22.7	19.7	25.5	18.4
199	23.4	16.6	22.1	20.6	21.3	16.4
199	18.8	23.4	25.7	29.7	26.6	23.4
199	19.5	16.1	24.4	25.7	29.9	14.1
199	30.3	20.6	28.0	27.0	37.5	18.4

199	23.9	18.6	24.2	24.5	32.4	14.5

B. Annual mean temperature deviation for South Polar

Year	Surface	850 – 300 mb	300 – 100 mb	100 – 50 mb
1958	0.06	0.16	0.07	0.87
1959	-0.61	-0.35	0.01	-1.17
1960	-0.85	-0.28	0.07	0.8
1961	0.24	0.35	0.09	-0.68
1962	-0.66	-0.5	-0.33	1.12
1963	-0.23	-0.42	0.27	0.72
1964	-0.52	-0.17	0.13	0.42
1965	-0.43	-0.25	-0.19	0.63
1966	0.6	0.06	-0.02	-0.6
1967	0.39	0.21	0.32	-0.2
1968	-0.28	-0.07	0.54	-0.17
1969	0.12	-0.19	-0.08	-0.98
1970	0.06	0.16	-0.15	-0.89
1971	0.1	0.04	0.2	0.09
1972	0.2	0.3	-0.3	-0.27
1973	0.53	0.4	-0.28	-0.23
1974	1.01	0.28	0.51	-0.2
1975	1.09	0.03	-0.19	-1.08
1976	-0.27	-0.17	-0.59	0.2
1977	0.7	0.73	-0.4	0.77
1978	-0.14	0	0.65	-0.15
1979	0.43	0.09	-0.04	-0.87
1980	0.97	0.72	0.64	-0.04
1981	0.96	0.38	0.33	-0.28
1982	-0.07	-0.11	-0.42	-1.76
1983	0.97	0.72	-0.25	-1.14
1984	0.93	0.22	-1.1	-0.22
1985	0.53	0.27	-1.18	-3.69
1986	0.45	0.43	-1.15	-0.35
1987	0.42	0.38	-2.26	-4.74
1988	1.37	0.9	0	0.82
1989	-0.29	0.54	-0.84	-2.54
1990	0.57	1.15	-1.41	-2.87
1991	1.25	1.24	-0.93	-0.91
1992	0.76	0.67	-1.16	-1.07
1993	-0.09	0.36	-2.1	-2.89
1994	-0.42	0.19	-1.47	-2.16
1995	0.17	0.18	-1.52	-2.98
1996	1.54	0.92	-3.07	-4.76
1997	0.79	0.18	-0.82	-3.56
1998	-0.07	0.94	-1.86	-4.58

1999	0.09	0.44	-2.9	-5.34
2000	-0.16	0.3	-2.3	-3.67
2001	1.05	0.24	-1.97	-4.23
2002	1.34	1.04	1.12	1.32
2003	0.1	0.6	-0.6	-1.96
2004	-0.31	0.14	-0.96	-1.21
2005	0.04	0.55	-0.82	-1.34
2006	0.38	0.29	-2.46	-4.61
2007	0.56	0.68	-2.01	-2.73
2008	0.32	-0.12	-2.16	-6.02

C. Annual mean temperature deviation for North Polar

Year	Surface	850 – 300 mb	300 – 100 mb	100 – 50 mb
1958	-0.27	-0.16	0.52	1
1959	0.59	0.97	0.21	0.75
1960	-0.19	0.41	0.81	3.15
1961	0.67	0.33	1.4	1.09
1962	0.6	-0.08	-1.33	-2.68
1963	-0.09	-0.06	0.04	-0.85
1964	-0.3	-0.47	-0.13	-0.38
1965	0.02	-0.16	-0.76	-1.7
1966	-1.45	-0.39	1.05	0.72
1967	0.3	-0.09	-1.74	-2.4
1968	-0.15	-0.23	-0.97	-1.05
1969	-0.15	0.25	0.83	-0.41
1970	0.17	0.19	0.81	1.21
1971	-0.05	-0.54	-0.02	0.31
1972	-0.34	-0.08	-0.95	-0.43
1973	0.5	0.29	0.47	0
1974	0.15	0.16	-0.26	0.46
1975	0.01	-0.19	0.05	0.28
1976	-0.2	-0.14	-2.11	-0.95
1977	-0.24	0.38	0.76	1.54
1978	0.15	0.14	-0.37	-1.08
1979	-0.45	0.01	0.04	0.62
1980	0.15	0.13	-0.34	0.32
1981	0.75	0.21	-0.82	-0.57
1982	-0.29	-0.17	-0.5	0.24
1983	0.12	0.14	-1.53	-1.54
1984	0.44	0.29	-0.69	-0.48
1985	-0.31	-0.31	1.37	0.59
1986	0.13	-0.14	-0.79	-0.68
1987	-0.63	-0.14	0.59	0.25
1988	0.4	0.21	-0.76	-0.9
1989	1.02	0.02	-0.93	-0.22
1990	1.04	0.02	-2.94	-3.01
1991	0.64	0.53	-0.26	-0.32
1992	-0.23	-0.74	-1.18	-0.88
1993	1.1	-0.5	-2.76	-2.51
1994	0.36	-0.38	-0.8	0.38
1995	1.45	0.3	-2.09	-1.35
1996	0.84	0.23	-1.31	-2.21
1997	1.04	-0.1	-2.18	-2.5
1998	0.86	0.38	-0.14	-0.47

1999	0.76	0.25	0.36	-0.78
2000	0.5	-0.43	-2.25	-3.23
2001	0.54	0.08	-0.71	-1.73
2002	1.38	0.69	-0.48	-1.37
2003	0.96	0.47	-1.04	-1.59
2004	0.6	0.53	0.66	-0.56
2005	2.06	0.67	-1.79	-2.08
2006	1.94	0.58	0.29	-0.71
2007	2.31	0.52	-1.89	-2.33
2008	2.14	0.12	-1.41	-1.7