Detecting Discontinuities Using Nonparametric Smoothing Techniques in Correlated Data

Christina Yap

A Dissertation Submitted to the
University of Glasgow
for the degree of
Doctor of Philosophy

Department of Statistics

March 2004

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Abstract

There is increasing interest in the detection and estimation of discontinuities in regression problems with one and two covariates, due to its wide variety of applications. Moreover, in many real life applications, we are likely to encounter a certain degree of dependence in observations that are collected over time or space. Detecting changes in dependent data in the presence of a smoothly varying trend, is a much more complicated problem that previously has not been adequately studied. Hence, the aim of this thesis is to respond to the immense need for a nonparametric discontinuity test which is capable of incorporating robust estimation of the underlying dependence structure (if unknown) into the test procedure in one and two dimensions.

By means of a difference-based method, using a local linear kernel smoothing technique, a global test of the hypothesis that an abrupt change is present in the smoothly varying mean level of a sequence of correlated data is developed in the one-dimensional setting. Accurate distributional calculations for the test statistic can be performed, using standard results on quadratic forms.

Extensive simulations are carried out to examine the performance of the test in the cases both of correlation known and unknown. For the latter, the effectiveness of the different algorithms that have been devised to incorporate the estimation of correlation, for both the equally and unequally spaced designs, is investigated. Various factors that affect the size and power of the test are also explored. In addition, a small simulation study is performed to compare the proposed test with an isotonic regression test proposed by Wu et al. (2001).

The utility of the techniques is demonstrated by applying the proposed discontinuity test to three sets of real-life data, namely the Argentina rainfall data, the global warming

data and the River Clyde data. The analysis of the results are compared to those using the isotonic regression test of Wu et al. (2001) and the Bayesian test of Thomas (2001).

Finally, the test is also extended to detect discontinuities in spatially correlated data. The same differencing principle as in the one-dimensional case is utilised here. However, the discontinuity in this context does not occur only at a point but over a smooth curve. Hence, the test has to take into account the additional element of direction. A two stage algorithm which makes use of a partitioning process to remove observations that are near the discontinuity curve is proposed. A motivating application for the approach is the analysis of radiometric data on cesium fallout in a particular area in Finland after a nuclear reactor accident in Chernobyl.

The procedures outlined for both the one and two dimensional settings are particularly useful and relatively easy to implement. Although the main focus of the work is not to identify the exact locations of the discontinuities, useful graphical tools have been employed to infer their likely locations. The dissertation closes with a summary and discussion of the results presented, and proposes potential future work in this area.

Declaration

This thesis has been composed by myself and it has not been submitted in any previous application for a degree. The work reported within was executed by myself, unless otherwise stated.

March 2004

To My Family and Ashley

Acknowledgements

The compilation of this thesis was not possible without the assistance of several people.

I am indebted to my supervisors, Professor Adrian Bowman and Professor Marian Scott, for their endless support and encouragement. They guided me with patience and wisdom, and inspired me with their enthusiasm and creativity. I am both honoured and humbled by their commitment and belief in my work.

Support from the Council of Vice Chancellors and Principals, and the University of Glasgow, is gratefully acknowledged.

It has been a privilege and pleasure to be acquainted with the members of staff and postgraduate students in the Department of Statistics. Their friendship and enthusiasm in research have made a great impact in my life.

My heartfelt gratitude goes to many of my friends in UK, Singapore and Australia, for the love they have demonstrated through their friendship, encouragement and intercessory support. In particular, I like to thank Wei Ling, Hui Min and Pui Ee.

I am very grateful to my family's unwavering support, which has always, and will always be, greatly motivating. Special appreciation is due to my sister, Marie, for her excellent insights and helpful comments in putting the finishing touches on the final draft. She has spurred me on with her love, encouragement and prayers.

And special thanks to Ashley for always being there and for his faith in me. His unfailing words of encouragement, patience and timeless help have been a constant inspiration.

Finally, thanks be to God who alone has given me the wisdom, ability, and endurance to complete this thesis, and for being my anchor. May the dedication of this work and the dedication of my life be eternally pleasing to Him.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

The last two decades have seen a burgeoning interest in the area of "change-point" analysis, equivalently known as "discontinuity" analysis. It is often assumed in regression analysis that the underlying "true" function is continuous. However, as will be illustrated later in this section, there are numerous real life applications where this assumption does not hold.

If the main focus of a statistical analysis is to estimate the underlying regression function, an increased bias will result in the estimation if one assumes that the function is continuous when it is not. Furthermore, important scientific information will be lost. Hence it is important to check the statistical homogeneity of the observed data before estimation of the parameters of statistical models, to obtain a more consistent estimate. In some instances, the main purpose might be to accurately model the underlying trend function in a setting of possible discontinuities and as such the actual locations of the discontinuities are of secondary importance. In other applications, detecting the possible presence of a discontinuity may be of primary interest and importance, as this might correspond to some physical intervention or phenomenon that the investigator is very keen to detect.

To illustrate the importance of change-point analysis, let us refer to some interesting examples that have appeared in past literature.

In the study of growth in biology, it is normally assumed that a log linear relationship

exists between the sizes of two body parts, and this relationship persists throughout the stable growth period. However, any structural shift in this relationship may be of interest as that might indicate the start of a new phase (Krishnaiah and Miao, 1988).

A recent analysis of the growth of infants by Müller and Stadmüller (1999) provides another interesting example. The given controversy here is whether the growth occurs smoothly or whether there exist discontinuities in the infant growth. The data were obtained by Lampl et al. (1992), who claimed in their findings that jump discontinuities do occur during the growth of infants, but an article by Heinrichs et al. (1995) argued that the claim was highly doubtful.

One of the most popular examples, analysed by several statisticians, is the annual volume of the Nile river flow data, which was shown to contain a discontinuity in the underlying mean function (Cobb, 1978, Müller, 1992). Independent evidence on tropical rainfall to support the possible drop in volume was mentioned by Cobb (1978).

In addition to the scientific fields of medicine and hydrology, the literature in this topic of change-point analysis is growing rapidly due to its wide range of applications in quality control, econometrics, signal and image processing and many other physical sciences. Indeed, it has been a popular topic of research for more than fifty years (see, for example, the books of Brodsky and Darkhovsky (1993, 2000) or Basseville and Nikiforov (1993)).

One motivating real-life application (yet to be published) is whether there has been any abrupt change in the dissolved oxygen level in the River Clyde, after an improved sewage treatment plant was implemented by the Scottish Environment Protection Agency, SEPA in 1985. As the data are collected over time, there might exist a certain degree of serial correlation, which will affect the discontinuity test if it is not accounted for. This points to the research question: is there a discontinuity in the River Clyde data after accounting for the possible presence of correlation? Furthermore, if we allow the trend to vary smoothly, by using nonparametric regression (the advantage of such a technique will be discussed in the next section), a test developed to answer this question will be useful in such a setting as well as in many other practical situations.

Similarly in a two-dimensional spatial setting, dependence often exists between values which are close together. One interesting application that will be analysed in this work is to

investigate if there is any abrupt variation in the deposits of cesium in an area in Finland, due possibly to different terrains. The underlying trend may be complex and using nonparametric regression to model the surface is greatly beneficial here. The research question here is: does the underlying surface of cesium deposits vary smoothly, or are there discontinuities present, perhaps due to the different terrains that are mapped?

This directs us to the main aim of this thesis: to detect discontinuities in the one-dimensional and two-dimensional settings, in the presence of a smoothly varying trend and correlation.

1.2 Different Types of Change-Point Problems

A *change-point* can be defined as a point which separates a series of observations, $x_1, x_2, \ldots, x_n$, into two groups, each of which follows a 'different' statistical model. The subscripts may denote time (equally or unequally spaced) or some other ordering of the observations. The last observation before the change or the first observation after the change can be defined as the change-point. In this work, we will refer to the change-point as the *last* observation before the change.

There are different kinds of change-point problems. According to the method of data acquirement, there exist two fundamental formulations.

The *retrospective* change-point problem (also known as the *off-line*, *posterior* or *non-sequential view*) is where the whole data set is complete at the time of analysis. It can emerge as a *hypothesis testing* problem, whereby the concern is whether a change has occurred, or as an *estimation problem*, concerning the location of the change-point if a change is present. The examples that are provided in Section 1.1 are all off-line change-point problems.

On the other hand, the *sequential* change-point problem, also known as *on-line*, *surveillance* or *turning-point detection*, is concerned with the setting where additional observations will be made available as time goes on. The objective is to detect any change in the observations as soon as possible, without too many 'false alarms'. On-line problems occur in numerous monitoring systems, for instance in areas of quality control, where successive observations of the process are monitored, and an alarm is sounded if there is a deviation from

a pre-set standard value. There are also applications in econometrics where one is interested in diagnosing the changes of an economy (also known as turning-points), and in medicine, such as the supervision of the heart rate of a patient, etc.

Various terms have been used to describe change-point analysis in its diverse fields of applications. It is most commonly referred to as "change-point" or "discontinuity", but also as "failure (fault) detection" in quality control and "structural break/change" or "intervention analysis" in time series. Other terms include "abrupt change", "level shift" or "shock detection". In this work, the author will restrict the terms to either "change-point", "discontinuity" or "abrupt change".

There can exist a *single* change-point or *multiple* change-points. There are also different kinds of changes, such as abrupt changes, gradual changes, sometimes expressed in terms of changes in model parameters. Abrupt changes, which are changes that occur very fast, if not instantaneously, with respect to the sampling period of the measurements, have been the most well-studied over the years. Abrupt changes in regression equations are also known as *switching regressions* or *two-phase regression problems.*

To illustrate some of the different change-point situations one may encounter, diagrams (1)-(3) of Figure 1.1 display an abrupt change in a constant mean, a switching regression and an abrupt change in a smoothly varying trend. Diagram (4) displays an abrupt change in slope, which is also an example of a switching regression with the addition of a continuity constraint. The bottom middle and right panels display abrupt changes in variance and correlation of a similar simulated set of data. The change-point occurs at the mid-point of the data for all these diagrams.

The underlying trend (away from the jump point) can be modelled *parametrically*, *nonparametrically* or with a *semiparametric* model. The parametric regression approach is based on the assumption that the regression function is expressed as a function with a finite known number of (unknown) parameters, whereas the nonparametric diagnosis simply assumes that the regression function is a smooth unknown function of the predictor variable, with no finite dimensional parameter space. The semiparametric diagnosis is intermediate between the two extreme positions, with a combination of a smooth unknown function and an unknown finite dimensional parameter space.

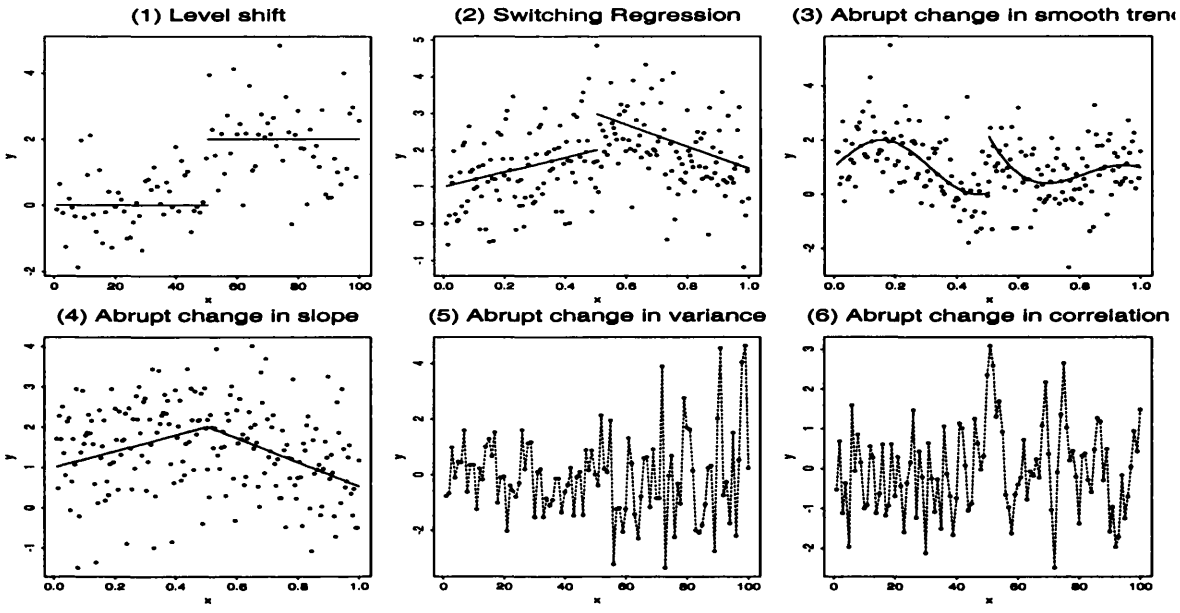

Figure 1.1. Different types of abrupt changes

The mean structures in diagrams (1) and (2), which are modelled by a constant and linear trend respectively, are the simplest forms of the parametric family. The number of parameters can be increased to give *polynomial* regression, such as quadratic or cubic regression. Though this is quite a frequent choice in modeling nonlinear curves, as Fan and Gijbels' (1996) work illustrates, one of the main shortcomings of polynomial regression is that it is still not flexible enough to model many real-life situations, thus resulting in many biases. Moreover, it might be very sensitive to the presence of outliers, which could have a great influence on remote parts of the estimated curve.

Regression is one of the most widely used statistical tools, but conventional parametric regression models are too restrictive; and other practical techniques are needed. Therefore, the rapidly growing field of nonparametric regression, whose policy is to "let the data speak for themselves", is an important and popular tool for modelling data and for inferential purposes. It is extremely flexible and enables the derivation of information from data, imposing very weak assumptions on the mean function, which cannot otherwise be explained by a parametric model. In fact, the key assumption for nonparametric regression is the presence of some degree of smoothness in the underlying trend. This meets the needs of many

analysts who deal with real-life data from a diverse range of fields, such as the environmental and medical fields, where the underlying trend may be too complex to permit the use of simple equations to accurately model it. With the same intention in mind, the author was motivated to develop a change-point algorithm that allows for a smoothly varying trend, so that it can contribute more significantly to practical real-life applications. Diagram (3) displays an example of a change-point model with an underlying smooth trend.

1.3 Scope of thesis

This thesis consists of seven chapters. In this chapter, an introduction and overview of the different change-point problems often encountered is given along with some illustrative examples. Chapter 2 provides a survey of previous literature to tackle different types of change-point problem, using both parametric and nonparametric approaches, with independent and dependent data in the one-dimensional setting. This then leads on to the main focus of the work which is using nonparametric smoothing techniques for change-point detection in an inferential context, particularly in the presence of correlated errors.

In Chapter 3, the discontinuity testing algorithm is thoroughly described and developed for the one-dimensional equally spaced case for correlated data. Different estimators of correlation and variance are proposed. An extensive simulation is carried out to evaluate the performance of the test both in the correlation known and unknown cases as we vary different factors that influence the results of the test. Comparison of the proposed nonparametric algorithm to the isotonic regression test proposed by Wu et al. (2001) is also investigated.

In Chapter 4, some modifications are made to the algorithm to extend the application to an unequally spaced setting. The structure is similar to that of Chapter 3. The test takes correlation into account by using a variogram. A new test statistic is also considered here and comparison is made with the original proposal. Simulations are carried out to see how the two test statistics perform in finite samples, and with different experimental factors, both in the correlation known and unknown cases.

Chapter 5 consists of the analysis of three real-life datasets. They are the Argentina data (Wu et al., 2001), the global warming data (Wu et al., 2001) and the River Clyde data.

The proposed nonparametric regression approach is used, and the results are compared with those using two other approaches; a Bayesian approach by Thomas (2001) and an isotonic regression test by Wu et al. (2001).

Chapter 6 extends the discontinuity test proposed in the one-dimensional case to the two-dimensional spatial setting, along with simulations representing both equally and unequally spaced data. The algorithm is then applied to investigate if there is abrupt variation in the mapping of cesium in an area in Finland (RESUME-95, 1997).

Chapter 7 concludes with an overall summary and discussion of this work, and proposes future developments.

Chapter 2

Statistical Approach to Change-Point Problems

Current literature on the general topic of "change-point detection" is too extensive to review here. Thus, in this chapter we will confine our attention to research in the one-dimensional setting that is mainly centered on abrupt changes in the mean structure, though a brief mention of the other types of change points will also be made. The list is not exhaustive. As our main interest is in a nonparametric regression approach, the whole of Section 2.2 is dedicated to this field of research. Detecting discontinuities in the two-dimensional setting will be discussed later in Chapter 6.

The organisation of this chapter is as follows. We will begin our tour by making a passing mention of sequential change-point detection and refer the reader to several cited references. The focus then turns to retrospective detection. A brief survey of the commonly used parametric approaches, namely Bayesian and maximum likelihood change-point detection, as well as nonparametric methods, will be made, starting from the independent setting and subsequently progressing to the dependent setting. The chapter ends with a section on detecting discontinuities using nonparametric regression.

2.1 History

The *sequential* change-point problem, with its connections to statistical quality control, started all the way back with Shewhart (1931), who approached *on-line* detection of a change by comparing individual means. He proposed the use of Control Charts in the manufacturing industry to decide if a manufacturing process or the quality of a manufactured product is in statistical control. Shewhart developed the use of 3-sigma control limits as action limits. An 'alarm' is sounded and the process is considered out-of-control if the quality characteristic goes beyond these limits.

Page (1954) dealt with on-line detection using the successive cumulative sums (CUSUM) instead, and thereby laid the fundamental basis for the widely used CUSUM-procedures for on-line detection of changes. Barnard (1959) extended the analysis further by modelling the underlying observations using a parametric model and basing an analysis on the estimated parameters. For more on on-line detection, see Basseville and Nikiforov (1993) or an annotated bibliography by Shaban (1980), Zacks (1983), and a recent paper of Lai (2001, Section 3) who gives a comprehensive survey of sequential change-point detection. The rest of this chapter is mainly focused on *retrospective* change-point detection, unless otherwise stated.

The *retrospective* change-point problem was formally introduced by Page (1955) who proposed the CUSUM procedure to test for a change in a parameter occurring at an unknown change-point. The problem was originally formulated to improve Shewhart's (1931) 3-sigma control chart procedures in quality control. Page's works stimulated several broad approaches to *retrospective* change-point problems.

In general, the model often considered in change-point problems is of the form

$$Y_i = f(i/n) + \varepsilon_i; \qquad i = 1, \ldots, n \tag{2.1}$$

where ε_i is usually assumed to be independent and identically distributed (i.i.d) random variables with zero means and finite variances. For convenience and without loss of generality, the total time of observations has been scaled to the interval $[0, 1]$. It is of interest to investigate whether f suffers from any abrupt jumps or sudden changes in slope (or gradual

changes in the mean level), at one or more of the points i/n where $i = 1, \ldots, n$. For the parametric approaches, a specific form of f is assumed, as illustrated in diagrams (1) and (2) of Figure 1.1. Diagram (1) can also be viewed as one where the data are independent with no additional covariate information; while diagram (2) shows that the response y varies linearly across covariate x. The nonparametric regression approach on the other hand, allows f to vary smoothly, as in diagram (3) of Figure 1.1.

There are two main aspects in change-point problems: *inference*, i.e. to test if there is a change-point; and *estimation*, to find the location of the change-point and/or its jump size, assuming that a jump is present. On the other hand, sometimes the main objective might be to estimate the underlying piecewise regression functions that are sub-divided by the change-points. This might involve an initial step of estimating the change-point locations.

A pioneering work in the *Bayesian parametric approach* to change-point detection is that proposed by Chernoff and Zacks (1964). They applied it to the problem of a one-sided likelihood ratio test for changes at unknown times in the level mean of a sequence of independent Normal random variables, using a quadratic loss function to test if there is a jump. An arbitrary prior is used here for the possible location of the change-points. Kander and Zacks (1966) generalised this one-sided test to exponential families. Gardner (1969) studied the testing problem for normal random variables when H_1 is two-sided. Smith (1975) also used Bayesian inference on the location of a change-point for data that follow a normal or binomial distribution.

Jandhyala and MacNeill (1989, 1991) also approached the change-point problem from a Bayesian framework and derived Cramér-von Mises type statistics expressed in terms of the sequence of partial sums of weighted regression residuals. The same authors, Jandhyala and MacNeill (1999), also proposed a one-sided test, using iterated partial sum sequences of regression residuals to detect change-points of segmented polynomials which are continuous at the change-points. The simplest case is the piecewise simple linear model as shown in diagram (4) of Figure 1.1. For a comprehensive survey of the Bayesian approach, see Zacks (1983) and Thomas (2001, Part I).

A key feature to highlight regarding the Bayesian approach to change-point analysis is that the unknown change-point location is considered as a *random* quantity and a probability

distribution is attributed to it. A prior probability is required for the unknown parameters in the Bayesian analysis. The improper or vague prior, such as the uniform prior, is commonly used. This implies that each position of the data has an equal probability of being the change-point location. On the contrary, the maximum likelihood method, which is a frequentist approach, considers the change-point location as an unknown but *fixed* quantity.

As an alternative to Bayesian methods, Hinkley (1970) proposed another important parametric approach, based on the *maximum likelihood test* to detect abrupt mean shifts as in diagram (1) of Figure 1.1. He based his inferences on the asymptotic sampling distribution of the maximum likelihood estimator of the change-point. The asymptotic distribution of the likelihood ratio statistic for testing the hypothesis about the change-point was derived and a large sample approximation for the p-value of the test was suggested.

Other work using maximum likelihood approaches includes that of Worsley (1983, 1986) which focused on detecting a change in the mean for a sequence of independent exponential family random variables. Siegmund (1986) developed a generalized likelihood ratio test (LRT) to avoid the case where the likelihood ratio statistic does not have a limiting distribution. Employing a method devised to solve boundary crossing problems in sequential analysis, Siegmund (1986) also obtained an approximation for the p-value of the test, which is reasonably accurate for small sample sizes. James et al. (1987) extended the results of Siegmund (1986) to the case where the variance is unknown and made comparisons of the performance of LRT with other tests. Kim and Siegmund (1989) also used likelihood ratio tests for discontinuity testing in simple linear regression models. Their alternative hypothesis can be a change simply in the intercept or a change in both the intercept and slope as in diagram (2) of Figure 1.1.

For more detailed discussions on change-point detection in parametric models for independent data, see Shaban (1980), Krishnaiah and Miao (1988), Basseville and Nikiforov (1993) and Chen and Gupta (2000).

One of the most dynamically developing areas of retrospective change-point detection for random sequences is the *nonparametric* approach. There are two very different aspects of this nonparametric approach. One is known as nonparametric because the method of detection is based on nonparametric methods using rank or score functions. This is useful when the

underlying distribution of the data is not assumed to be known. The other is referred to as nonparametric because no parametric form is assumed for the underlying trend. It is allowed to be smooth and modelled using nonparametric regression. This latter aspect will be discussed in greater detail in Section 2.2, but let us first discuss the former, which we will refer to as the *nonparametric method.*

For the nonparametric method, the goal is to address the question of whether the data are generated by one or by more probabilistic mechanisms. For instance, let Y_i be independent and identically distributed random variables such that

$$\begin{aligned} Y_1 \ldots, Y_\tau \quad & \text{has cdf } F \\ Y_{\tau+1} \ldots, Y_n \quad & \text{has cdf } G \end{aligned} \tag{2.2}$$

where $\tau \in \{2, 3, \ldots, n-1\}$ is the possible unknown change-point and F and G are unknown probability distributions. In the case where there is no change, we would expect $F = G$.

The first *retrospective* nonparametric change-point detection methods was proposed by Bhattacharya and Johnson (1968). They considered the model as expressed in Equation (2.2). Their test statistic is based on the ranks of observations, R_k, $k = 1, \ldots, n$, where $R_k = \sum_{i \leq n} I(y_i \leq y_k)$ and n is the number of observations. The difference between their test statistic and that of Chernoff and Zacks (1964) lies in the values of the data: y_i are replaced by ranks R_i in the nonparametric case.

Darkhovshky (1976) and Schechtman (1982) developed their test statistic based on the Mann Whitney statistic. Other work has also considered using generalised variants of the Kolmogorov-Smirnov test used for checking the equality of distributions (see e.g. Carlstein, 1988, Dümbgen, 1991, Darkhovsky, 1994). An alternative nonparametric test statistic is that based on the U-statistic type processes (Csörgő and Horváth, 1988).

For a survey of nonparametric methods in the setting of just one change-point, especially in testing, refer to Wolfe and Schechtman (1984). Csörgő and Horváth (1988) also provided an extensive bibliography of nonparametric methods for discontinuity testing as well as the problem of sequential detection of change in a random process. Refer also to Brodsky and Darkhovsky (1993, 2000).

Most of the earlier works considered only the detection of a single change-point. There has been very little research dedicated to detecting multiple change-points until recent years. This is a much more complex problem, especially when there is no prior knowledge of the number of changes.

Pettitt (1980) recommended an ad hoc sequential procedure involving a cumulative sum (CUSUM) method. Yao (1988) estimated the number of jumps in a sequence of independent normal sequence using the Schwarz' criterion, whilst Lomard (1987) proposed rank statistics for detecting single and multiple abrupt and smooth change-points. Schechtman and Wolfe (1985) proposed a sequential algorithm for estimating the number and the location of change-points of data from an unknown distribution, extending earlier work by Schechtman (1982). Hawkins (2001) developed an algorithm to obtain maximum likelihood estimates of the change-points of data from an exponential family distribution, by executing a multiway split instead of obtaining them sequentially.

For detection of multiple change-points from a Bayesian perspective, refer to a recent paper by Rotondi (2002) who used the *reversible jump* Markov chain Monte Carlo (RJMCMC) first proposed by Green (1995), and the references therein. The reversible jump MCMC permits the sampler to move between parameter subspaces of another dimensionality, unlike the standard MCMC which is constrained to situations of the same dimensionality. However, this technique does bring with it a substantial increase in computational load.

Most of the earlier literature has dealt only with change-point detection where the data are assumed to be independent. However, this assumption may not hold in many real-life applications. For instance, assuming independence for repeated measurements of the heart rate of a patient upon giving a treatment (which would naturally show some degree of dependency) does seem too optimistic. Hence, it would be very beneficial if a test can cope with both independent and correlated data. The call for improved statistical techniques to deal with such real-life applications has prompted many statisticians to venture into this area.

The importance of taking correlation into account was emphasized by several researchers, including Tang and MacNeill (1993) and Kim (1996), who showed that failure to account for correlation will affect the performance of various change-point test statistics in terms of the

size and power of the test. Hence it is essential to correct the test for correlation to ensure an accurate conclusion from the test.

Tang and MacNeill (1993) suggested a large sample correction factor for various test statistics defined in terms of partial sums of residuals that were considered in the independent case. This includes the test statistics proposed by Chernoff and Zacks (1964), James et al. (1987), Kim and Siegmund (1989), Jandhyala and MacNeill (1989, 1991) and many others.

Kim (1996) extended the generalised likelihood ratio test (LRT) developed by Siegmund (1986), which was mentioned earlier, to the situation where the data are correlated. He also demonstrated that if correlation of an AR(1) model is not taken into account, the true p-values will be underestimated or overestimated for positive or negative correlations respectively.

If the data follow a more specific time series model, analysis of abrupt and gradual changes are often referred to as *intervention analysis.* The central method of intervention analysis is to include more terms in the initial time series model and test for significant parameters (see Box et al. (1995) for a more in-depth discussion of this topic). However, the main drawback of intervention analysis might be that it is model specific, requiring the data to follow a particular time series structure. Hence it cannot be easily modified to a model free setting.

One of the pioneering contributions to the detection of change-points of dependent random sequences is by Box and Tiao (1965). They used a Student's t-test to detect a change in the mean value of observations, using a non-stationary moving average model, assuming a known change-point.

For sequential change-point analysis, Bagshaw and Johnson (1975) examined the effect of ARMA noise on the run length distribution for CUSUM. Yashchin (1993) proposed an alternative approach to assess the run length of a CUSUM control scheme in the presence of serial correlation. His main idea involves replacing the correlated series by an i.i.d. sequence which has similar run length characteristics. Recently, a new technique, Singular Spectrum Analysis (SSA), has also been proposed to detect change-point in time series (Moskvina, 2001).

Moving our focus back to retrospective detection, Kulperger (1985) considered polynomial regression and extended the results of MacNeill (1978) to autoregressive error processes

of order p. Picard (1985) used a maximum likelihood approach to detect a single change-point in the mean of a Gaussian autoregressive process, the order of which is known. This was also extended to detect a change in the mean and autoregressive parameters simultaneously. Bai (1994) extended these results by proposing a least squares estimate of the location of a single change-point in the mean of a linear process. This was further applied to multiple change-points by Bai and Perron (1996) for weakly dependent processes. Subsequently, Bai (1997) looked at least squares estimation of a change-point in multiple regression models, allowing for some or all the coefficients to have an abrupt change at an unknown time. Bai (1999) proposed a likelihood-ratio-type test for detecting multiple change-points in polynomial trend. A much related work includes that of Kao and Ross (1995) who used a modified CUSUM test to examine if there is a change-point in a linear regression model with serially correlated errors. Their test includes an initial estimate of the autoregressive coefficient for the errors, which is then used to transform the data.

Other researchers who have examined change-point detection in time series include Tsay (1986, 1990) who proposed some procedures for detecting outliers, level shifts, and variance changes in univariate time series.

From a Bayesian perspective, Henderson (1986) considered a Bayesian test approach for data that form a series of correlated normal variables, assuming both the error and correlation are known. In the same vein, Nagaraj (1990) obtained the likelihood ratio test statistic for a change in the mean of correlated data, under several different assumptions of known and unknown mean and variance. Thomas (2001) also suggested a Bayesian approach to retrospective detection of change-points of an AR(1) model. In Thomas (2001, Paper II), he proposed using approximate inference in the case where autocorrelation is unknown. This is computationally faster than the Markov Chain Monte Carlo techniques that he proposed in Thomas (2001, Paper III) which allows for exact inference to be carried out. The approach by Thomas (2001, Paper II) will be revisited in Chapter 5.

A test for a constant mean function when the data are dependent was also devised by Kim and Hart (1998). Under the null hypothesis, the underlying mean is a constant, while the alternative can be at least one level shift or a smoothly varying trend. Their test statistics are obtained from a Fourier series smoother that minimises an estimate of the mean integrated

squared error.

Epps (1988) focused on dependent data with abrupt changes at some known times and proposed a chi-squared statistic for testing stationarity of a Gaussian process. Lavielle and Moulines (2000) extended the approach by Yao (1988) and studied a penalised least-square estimate of an unknown number of multiple changes in a larger class of dependent processes, including strongly mixing and long-range dependent processes (see also Lavielle, 1999).

Using a nonparametric method, with Kolmogorov-Smirnov type tests for dependent data, Giraitis et al. (1996) tackled the problem in the setting where the distribution is unknown. For a detailed survey of tests developed to cope with dependent data till 1997, see Antoch et al. (1997) and the references therein.

In recent years, there has also been a growing interest in using an alternative fuzzy statistical technique for the estimation of multiple abrupt and gradual change-points of nonlinear time series; see Wu and Chen (1999) and Kumar and Wu (2001). The work by Kumar and Wu (2001) involves first detecting the location of the change-points using fuzzy logic, followed by fitting a linear regression model with time between the estimated change-points. They then used a partial F statistic to test if the slopes or the levels of the two consecutive lines are the same.

For work on change-point detections of variance in time series data, see Wiche et al. (1976), Davis (1979), Inclan and Tiao (1994) and Lee and Park (2001).

2.2 Nonparametric Regression Approach

This section is mainly focused on using nonparametric regression to estimate a smooth underlying trend which might contain discontinuities at certain locations, as illustrated in diagram (3) of Figure 1.1.

As mentioned in the earlier section, a parametric approach to the regression change-point model assumes simple linear or polynomial regression before and after a possible change-point, with the possible occurrence of a jump in the function or its first derivative. In practice, it is often difficult to find a suitable parametric method to model the underlying trend and to estimate the location and the sizes of the jumps in the regression function.

On the other hand, there has been considerable development in the use of nonparametric regression methods for change-point analysis since the 1990s. This is mainly because it is an attractive alternative when no appropriate parametric model is available. Nonparametric regression can produce a smooth fit of a regression function and its characteristics from noisy data, without any knowledge of the initial distribution of the data. Requiring weaker assumptions simply of general smoothness and differentiability as compared to parametric models, nonparametric regression enables a large class of regression functions to be considered.

In nonparametric regression analysis, it is conventionally assumed that the underlying smooth trend is continuous. However it is quite common, in a real-life setting, that the regression function may have discontinuities due to the occurrence of some external forces. From recent research (see e.g. Müller and Stadmüller, 1999), it has been highlighted that ignorance of the possible existence of discontinuities in a regression function may result in inflated bias in the regression estimate, as well as possible loss of valuable information. This points us to an *inference* problem, with a key question to tackle: is the smooth trend continuous or discontinuous?

However, besides the issue of inference, there is another scope of change-point analysis, which is the *estimation* problem of the change-point locations, and/or the underlying regression function (away from the change-points). If the focus is on the latter, the technique to estimate the discontinuous smooth regression function could be *direct* or *indirect.* The direct technique estimates the trend adapting to possible breaks, while the indirect one first estimates the location of the change-points before estimating the trend functions of each segment sub-divided by the change-points. The estimation problem has surprisingly received greater attention in the change-point literature using nonparametric regression than the inference problem, although there are some overlaps of both in certain papers.

One of the most common nonparametric smoothing techniques for change-point analysis is *kernel* estimates. This comprises of work by Müller (1992), Hall and Titterington (1992), Wu and Chu (1993), Chu (1994), Speckman (1994) and Eubank and Speckman (1994), among many others. Several authors who have used *local polynomial* methods are McDonald and Owen (1986), Müller (1993), Loader (1996), Qiu and Yandell (1998), Spokoiny

(1998), Grégoire and Hamrouni (2002a) and Bowman et al. (2003). Speckman (1995) and Cline et al. (1995) proposed fitting curves with possible discontinuities in the mean function or its derivatives using semiparametric change-point methods. Both the kernel and local polynomial approaches will be discussed in greater detail in the next section.

Besides kernel and local polynomial techniques, other forms of nonparametric smoothing which have been used for change-point estimation or discontinuous trend include *wavelets*, *splines* and *M-smoothers*.

Using a *wavelet-based* method, Wang (1995) tests for jumps and sharp cusps in the case of a continuously observed signal, with Gaussian white noise. The change-point estimator converges at rate $n^{-1}(\log n)^{1+\delta}$ where n is the number of sample points and $\delta > 0$. The convergence rate can be refined to $n^{-1}(\log n)^{1/2}$ for Normal errors. However, it was commented by Gijbels et al. (1999) that wavelets are either rough, fractal-like graphs or smooth graphs with numerous turning points and thus to obtain the rate of $n^{-1}(\log n)^{1/2}$ is in practice, very difficult. In other words, it is difficult to identify the specific location where the appropriate maximum absolute value of the empirical wavelet coefficient occurs. Other work using wavelets include Raimondo (1998), Oudshoorn (1998) and Antoniadis and Gijbels (2002).

Another established smoothing technique to retrieve discontinuous but smooth trends is *spline-based* methods. Some of the concerns encountered here are the selection of the number and position of knots within each spline, and possibly the number and location of the discontinuities. In particular, the principle behind Koo (1997) is to add additional knots near the discontinuities to cope better with jumps or cusps present in the regression function.

In the same vein, approaching the problem from a Bayesian framework, Denison et al. (1998) assumed that the underlying smooth trend can be estimated by several low order piecewise polynomials. The number and locations of the knots which link the polynomials are unknown parameters to be inferred. Their proposed algorithm was shown to approximate functions that had discontinuities or were rapidly varying. For related work using smoothing via spline-based techniques, see also Lee (2002).

Of late, an edge preserving technique based on an *M-smoother*, which directly estimates the regression function in an automatic fashion without first obtaining the discontinuities, has been employed by Chu et al. (1998) in the field of image processing. Other related work

using this technique includes that of Kauermann (2000) and Burt (2000). The latter extends the work by Chu et al. (1998) by proposing an automatic bandwidth selection criteria for the M-smoother.

For more details on the other approaches, the reader is referred to the references mentioned above and those cited within. The next section will only concentrate on the use of kernel and local polynomial regressions.

2.2.1 Using kernel and local polynomial regression

In fact, though there are slight deviations from the basic idea, the strategy of most of the kernel and local polynomial change-point techniques are similar. The most popular approach is based on finding the differences between the left and right smooths of the data at an evaluation point z. The left and right smooths are obtained by smoothing over only data at the left and right sides of z respectively. A natural estimator of the possible change-point location is where the difference between the two smooths is maximised.

Before we discuss aspects of some of the key work that has been presented in this field, it would be helpful first to summarise some of the main differing features and assumptions:

1. smoothing techniques (e.g. use of different types of kernel function and assumptions made on them; use of different degrees of local polynomials.)

2. fixed or random design of covariate x

3. assumptions of errors: a typical assumption is independent and homoscedastic errors but some have also assumed normality, heteroscedastic errors or dependent errors

In Müller (1992), the following fixed design regression model was considered.

$$y_i = g(x_i) + \varepsilon_i \qquad x_i \in [0,1], \quad 1 \leq i \leq n \tag{2.3}$$

where y_i are noisy measurements of the smooth regression function g, taken at equidistant points $x_i = i/n$ where $i = 1 \ldots, n$, and ε_i are i.i.d. errors with $E(\varepsilon_i) = 0$, $var(\varepsilon_i) = \sigma^2 < \infty$.

To illustrate the main idea of the approach by Müller (1992), we will simply focus on the detection of an abrupt change in the mean (though Muller's approach is applicable to changes in the derivatives as well) with just one change-point.

Assume that the regression function is continuous until a particular change-point τ where there is a jump of Δ and that a continuous function f exists such that

$$g(x) = f(x) + \Delta I_{[\tau,1]}(x) \quad 0 \leq x \leq 1, \tag{2.4}$$

where $I_B(\cdot)$ is the indicator function of a set B, and Δ is the jump size at the possible change-point τ. If there is no discontinuity, we would expect Δ to be 0. On the other hand, if there is a discontinuity an estimator of τ can be expressed as

$$\hat{\tau} = \arg\max_x |\hat{g}_-(x) - \hat{g}_+(x)| \tag{2.5}$$

where the left and right smooths of x are denoted as g_- and g_+ respectively.

The estimated jump size at $\hat{\tau}$ is then

$$\hat{\Delta} = \hat{g}_-(\hat{\tau}) - \hat{g}_+(\hat{\tau}) \tag{2.6}$$

To estimate the final regression function g, the segments divided by the change-point can then be estimated separately using the user's preferred smoothing techniques.

Müller (1992) uses the Gasser Müller (1979) kernel estimator as his smoother,

$$g_\pm(t) = \frac{1}{h}\sum_{i=1}^{n} y_i \int_{s_{i-1}}^{s_i} K_\pm\left(\frac{t-u}{h}\right)du$$

where $s_i = \frac{t_{i,n}+t_{i+1,n}}{2}$. K_+ and K_- are right and left one-sided smooth kernels with the support $K_+ = [-1,0]$ and $K_- = [0,1]$, $K_{-x} = K_x$ and $K_+(-1) = K_+(0) = 0$. The bandwidth h satisfies $h \to 0$, $nh \to \infty$ as $n \to \infty$, $\limsup_{n\to\infty} nb^{2(k)+1} < \infty$.

Based on asymptotic properties of Muller's (1992) one-sided kernel estimates, asymptotic

confidence intervals for both the location and size of the change-point are derived. In addition, it was shown that the rate of convergence is $O(n^{-1+\gamma})$ for arbitrary $\gamma > 0$, which is slower than $O(n^{-1})$ in the parametric case for the change-point location τ.

The method proposed by Müller (1992) assumes that there is only one discontinuity. Qiu (1994) generalized it to the situation with an unknown number of discontinuities but placed a restriction such that the jump sizes have a known lower bound.

Wu and Chu (1993) employed a similar approach based on Gasser-Muller kernel estimators to obtain change point locations on an equally spaced fixed-design nonparametric regression model as in Equation (2.3) of Müller (1992). Improvements to the performance were made by using two different bandwidths for estimating the change-point location and the jump size.

Eubank and Speckman (1994) suggested a different approach based on semiparametric estimators and weighted least squares estimators to obtain the size and location of the change point in the first derivative of an otherwise smooth function. The semiparametric method is asymptotically related to Muller's (1992) kernel estimate.

Several authors have noted the very advantageous property of the absence of edge effects when using odd local polynomial smoothing as opposed to kernel smoothing; see e.g. Hastie and Loader (1993), Fan and Gijbels (1996) and Loader (1996). The rate of convergence of the asymptotic bias between the left and right smooths is equal for the boundaries as well as for the interior points. Hence, asymptotically, odd numbered local polynomials will provide superior results for change-point inference and estimation. This will be discussed in greater detail in Chapter 3, Section 3.2.4.

Using a local polynomial smoothing technique, Loader (1996) proposed an estimate for an unknown change-point τ in the regression function g that attains a better convergence rate of $O_p(n^{-1})$ in the case where the noise is Gaussian with constant variance. He uses a similar regularly spaced model as in Equation (2.3).

The principle to obtain the change-point is similar to Müller (1992) with the exception of the smoothing techniques and the different conditions on K that Loader (1996) imposed, which led to different asymptotic performances. The use of the vital condition $K(0) > 0$ by Loader (1996) allows him to produce a superior convergence rate than Müller (1992) who

used $K(0) = 0$ for the change-point estimate.

In addition, Loader (1996) demonstrated that the change-point estimate has the same asymptotic distribution as the maximum likelihood estimates used in the parametric approach. His approach is built on ideas from likelihood based inference in a parametric model, where the limit distribution of $n(\hat{\tau} - \tau)$ is related to the location of a maximum of a two-sided random walk (see Hinkley (1970) and Kim and Siegmund (1989)).

Qiu and Yandell (1998) suggested a local polynomial jump detection algorithm to detect an unknown number of change-points in regression functions or their derivatives. To detect abrupt changes in the mean trend, the technique involves fitting a least square line in a neighbourhood of each design point and examining the derivatives of these estimates. Making use of the asymptotic property of the local linear estimates, which is approximately normal, they obtain a threshold value which can be used to decide whether a particular design point can be considered a change-point. The performance of their algorithm is compared to that of Wu and Chu (1993) and it is shown in their simulation that their algorithm, which takes into account the derivatives in discontinuity detection, performs better when the underlying regression function is steep but continuous.

Assuming that a change-point exists, Grégoire and Hamrouni (2002a) used a local linear approach to estimate the change-point location by means of the technique of differences between the two smooths as in Equation (2.6). The model is similar to Equation (2.3) except that they considered a regression model that allows for random design and heteroscedastic errors, and no assumption has been made on the noise unlike that in Loader (1996). A continuous right-sided kernel function $K_+(\cdot)$ supported on $[-1, 0]$ is used and $K_-(x) = K_+(-x)$. They also assumed that $K_-(0) > 0$ and achieved the same rate of convergence of $O(n^{-1})$ as Loader (1996) for $h \to 0$ and $nh \to \infty$. They obtained convergence of the jump size $\hat{\Delta}$, showed that $\hat{\tau}$ is consistent and derived the asymptotic normality of both the estimators of the location τ and the jump size at the change-point location, $\hat{\Delta}(\hat{\tau})$.

Other authors have proposed a two-step procedure to obtain the change-point estimator. This includes the work by Müller and Song (1997) and Gijbels et al. (1999).

To improve the asymptotic rate of convergence of the change point location τ obtained by Müller (1992) from $O(n^{-1+\gamma})$ to $O(n^{-1})$ in the fixed jump case, Müller and Song (1997)

propose a two-step procedure for the case of one jump discontinuity. A regularly spaced design similar to Müller (1992) in Equation (2.3), which only requires mild conditions on the error distribution, is studied here.

The first step involves obtaining a kernel-based pilot estimator of the change-point location, $\tilde{\tau}$ using the difference based approach as in Equation (2.6). Under suitable conditions, it was shown that $P(|\tilde{\tau}-\tau| \leq h_n)$ converges to 1 as $n \to \infty$. The second step then comprises maximising the weighted left and right differences of the mean of the data that falls within the subintervals obtained using the pilot estimator $\tilde{\tau}$ in the first step. The final change-point estimator is the location where the weighted difference is maximised. The idea behind this approach is that important observations to determine the change-point location are those that are close to it and, asymptotically, this can be found by using step functions.

Like Müller and Song (1997), Gijbels et al. (1999) employed a two-step method to estimate change-points in a smooth underlying trend, requiring only mild conditions on the error distribution. However, they also consider the detection of the locations of multiple change-points where the number is known. The first step involves using a kernel estimator and obtaining an initial pilot estimator of the change-point at the location where the estimate of the first derivative is maximized. This leads to a convergence rate of $n^{-1}(\log n)^{1/2}$, provided that the error distribution has a finite moment generating function. Gijbels et al. (1999) then combined this with a local least squares step, fitting curves that are locally constant within an interval of the pilot estimator found in the first step, to obtain the final change-point estimator $\hat{\tau}$. They showed that this improves the convergence rate to n^{-1}.

Two different bandwidths are used in the estimation procedure above, which undoubtedly do affect the performance of the estimator $\hat{\tau}$. Gijbels and Goderniaux (2002a) investigated the data-driven choices for both bandwidths using a bootstrap procedure with a suitable error criterion and showed by the use of simulations that the method performs well.

Instead of the popular difference based approach to detect discontinuities, Jose and Ismail (1997) established an alternative algorithm based on the analysis of the residuals from the nonparametric kernel regression estimate in a fixed design. They estimated the location and size of a jump in the regression function or its first derivative, as well as the discontinuous regression function. Jose and Ismail (1999) improved the algorithm to identify the number,

order, location and size of the multiple discontinuities in the function or its derivatives. It can be applied not only to a fixed design, but also to a random design. They compared their method with that of Loader (1996) based on sample mean squared error (MSE). Their method performs better in estimating the change-point and size of jump in the nonparametric regression function.

Whilst the objective of some work is to estimate the locations of the change-points and the corresponding jump sizes, there are several works that are more concerned with estimating the discontinuous regression function.

Kang et al. (2000) focused on the estimation of the entire regression function, which is assumed to be smooth, except for a finite number of points of discontinuities in a fixed design model, using an indirect technique. Their main idea is to first find estimators of the change-point location τ and the corresponding jump size Δ which possess good performance. The estimated size of the jump is then subtracted from all the observations y_i obtained after the estimated change-point. The adjusted data can then be considered smooth and be estimated by an ordinary kernel regression estimator. Subsequently, the final regression function is obtained by adding the jump back to observations after the change-point. Their global L_2 rates of convergence for the proposed regression function is demonstrated to have the same rate of convergence as the ordinary kernel regression estimators of smooth curves.

Several authors have also proposed algorithms to estimate the discontinuous smooth functions *directly*, instead of first obtaining the change-point locations. One such algorithm is suggested by McDonald and Owen (1986) who estimated the discontinuous function directly by using three smoothed estimates of the regression function, that are obtained via data on the right, center and left of a point by least squares. A split linear fit of the data is then obtained by using weighted averages of the three estimates, where the weights are decided by the goodness-of-fit values of the estimates. In the setting where there is a discontinuity, only some of these three estimates will produce good fits.

Hall and Titterington (1992) studied edge-preserving and peak-preserving by smoothing, and provided an algorithm that is easier to apply than McDonald and Owen (1986), to estimate the discontinuous regression function. Their algorithm is also based on constructing three smooths: right, central and left. Comparisons are then made between the three

smoothers, which are expected to be very different at a discontinuity point.

Approaching the problem from a similar perspective, a recent paper by Qiu (2003) suggests using local linear estimation to fit a smooth regression curve automatically, preserving any jumps present, without first estimating the change-points or assuming any knowledge of the number of change-points. The main difference of the algorithm compared to McDonald and Owen (1986) and Hall and Titterington (1992) is that the fitted value of the regression function at a particular point is obtained by either the left or right piecewise fitted line which has a lower weighted residual sum of squares.

Having reviewed past literature, it is quite apparent that considerable attention has been given to both testing and estimation of change-point using parametric models. On the other hand, discontinuity detection using nonparametric regression has grown speedily in popularity in recent years, as it gives greater flexibility to the model. However, in this research field, the main focus of most of the papers as discussed earlier, particularly those using kernel and local linear polynomial smoothing techniques, have been on estimating the locations of one or more change-points, or estimating the discontinuous but otherwise smooth function. There appears to be less attention specifically devised for inference, to first evaluate whether the underlying regression is smoothly varying or that there exists certain points where the smooth curve is discontinuous.

Some authors who have undertaken the challenge of inference in the early 1990s include Wu and Chu (1993) and Speckman (1994). Wu and Chu (1993) used the difference-based kernel smoothing approach and suggested a global test for the number of change-points in an equally spaced fixed design model. The null hypothesis is stated as $p = 0$ versus the alternative $p > 0$, where p is the number of change-points. They also give asymptotic intervals for the sizes of jumps in the regression functions, using the limiting distribution of the kernel type estimators. Speckman (1994) suggested a semiparametric change-point estimator based on kernel smoothing to detect one or more change-points in the mean or one of its derivatives.

In the last four years, Müller and Stadmüller (1999), Horváth and Kokoszka (2002) and Grégoire and Hamrouni (2002b) have also proposed tests based on asymptotic approximations.

Instead of the popular difference-based approach of the left and right smoothers, Müller and Stadmüller (1999) proposed a global test where their test statistic is based on the sums of the squared differences of the observations, formed over various span sizes in an equally spaced fixed design regression model. The test statistic can be expressed as an asymptotic simple linear model which involves the parameters of the sum of squared jump sizes and the error variance. They showed that both parameters can be estimated consistently. Their test evaluates if there is any discontinuity by examining whether the sum of the squared sizes of the jump of all change-points is close to 0 or significantly greater than 0.

Horváth and Kokoszka (2002) used local polynomial smoothing techniques, and also considered a fixed equally spaced model as in Equation (2.3). No assumption is made on the error distribution. Just as in the papers previously described, the difference-based approach is used, where local polynomials of degree v are fitted at the left and right of each evaluation point, to evaluate if the appropriate coefficients of the two polynomials are significantly different. Under the null hypothesis of the global test, the function g or its derivatives $g^{(v)}$ is continuous, versus the alternative, where the left and right smooths are not equal at the change-point τ, i.e. $g_{-}^{(v)}(\tau) \neq g_{+}^{(v)}(\tau)$. Horváth and Kokoszka (2002) also showed consistency of the change-point estimates under various conditions of the kernel functions.

A similar asymptotic test of the same nature is proposed by Grégoire and Hamrouni (2002b) using local linear regression with the difference-based approach, but both the local and global tests of whether the size of the jump is significantly greater than 0 are considered here.

Although these methods have served current research reasonably well, the dependence on asymptotic laws is limiting in real-life settings, particularly in the case of finite samples.

On the contrary, Gijbels and Goderniaux (2002b) proposed a fully data-driven bootstrap test which does not depend on asymptotic laws because it estimates the sample distribution of the test statistic using a bootstrap approach. The test comprises of a data-driven choice of bandwidth. When Gijbels and Goderniaux (2002b) compared their test via simulations to those by Müller and Stadmüller (1999) and Grégoire and Hamrouni (2002b), they demonstrated that their bootstrap test performs better.

Similarly, Bowman et al. (2003) also proposed a difference-based test that does not depend

on asymptotic approximations, but has an additional assumption of the normality of errors. A local linear smoothing approach is proposed. By expressing the test statistic as a ratio of quadratic forms, an accurate estimate of the error variance is not required, unlike most asymptotic tests, such as those of Wu and Chu (1993), Speckman (1994) and Müller and Stadmüller (1999), which require a consistent variance estimator. This might not be possible in finite sample sizes. In situations where the error variance might be underestimated, this will result in an increase in the type 1 error or the size of the test. However, the method by Bowman et al. (2003) is not fully data-driven as in Gijbels and Goderniaux (2002b) since the choice of an automatic bandwidth is not addressed here. To address this issue from another perspective, Bowman et al. (2003) proposed using a significance trace instead to examine how the significance of the test varies as the bandwidth changes. The main advantage of the test over that of Gijbels and Goderniaux (2002b) is that it is computationally less intensive.

However, these papers have only considered the independent setting. Extending the tests to allow for possible correlation will be of great utility and practical value in real-life applications. Several authors have dealt with this issue using parametric models as mentioned earlier. However, detecting abrupt changes in the presence of a smoothly varying mean is a much more difficult problem which has not been adequately studied. A closely related work is that of Wu et al. (2001) who used isotonic regression in the dependent setting. This is in fact also a form of nonparametric regression, but instead of a smooth underlying trend, it consists of piecewise step functions. A likelihood ratio test is used to examine if the trend is constant. Their alternative hypothesis supports the existence of an abrupt jump or monotonic trend. This technique will be discussed in greater detail in Chapter 5.

Nonetheless, approaching the problem by allowing for a smooth trend, in the presence of correlated errors, will broaden the practicality of discontinuity testing to a substantial magnitude. The work proposed in this thesis is a response to the pressing need for further methodological developments in this area.

The four primary research objectives of this dissertation are:

1. To extend the test proposed by Bowman et al. (2003) to a correlated setting in the one-dimensional setting, which does not rely on asymptotic results. By expressing the test statistic as a ratio of quadratic forms, an accurate estimate of the error variance is not necessary.

2. To develop robust ways of estimating correlation in both the equally and unequally-spaced one-dimensional setting.

3. To extend the test to a two-dimensional setting, incorporating a robust estimation of correlation in both the equally and unequally spaced settings.

4. Finally, to provide user friendly graphical tools to illustrate where the discontinuities might occur in both one and two dimensional settings

Chapter 3

Testing for Discontinuities in One-Dimensional Correlated Data: Equally Spaced Setting

3.1 Introduction

In many real life situations, we may have an unknown number of discontinuities in an otherwise smooth regression function. Some examples of such data-sets are the well-known Nile data (Cobb, 1978), the Mine accident data (Jarrett, 1979), the Argentina data (Wu et al., 2001), and the global warming data (Wu et al., 2001, Jones et al., 2001). As the data are collected over time, there often exists a certain degree of correlation which should be accounted for when applying a discontinuity test to detect possible change-points.

One of the primary objectives of this work is to develop a test to detect discontinuities in the one-dimensional correlated data setting. The methodology is designed to cope with two different types of design spaces: equally-spaced and unequally-spaced. This chapter develops a test that uses a local linear nonparametric kernel regression approach in the equally-spaced setting. In the next chapter, this methodology will be extended to cope with data that are irregularly spaced.

The advantages of this approach are that it not only allows for a smoothly varying mean

structure, but it also does not rely on asymptotic approximations. These are favourable properties, especially for small to moderate sample sizes, compared to other testing procedures which also use nonparametric regression for discontinuity testing, but are based on asymptotic results, such as Wu and Chu (1993), Müller and Stadmüller (1999) and Grégoire and Hamrouni (2002b). However, the test developed here does require the assumption of the normality of errors.

Chapters 3 and 4 follow the same structure. Firstly the test statistic is developed. This is then followed by a simulation study designed to investigate how different design factors affect the size and power of our test.

The simulation study consists of two main parts. Firstly, the correlation, Σ, is assumed known. Secondly, Σ is estimated using two different methods, to assess their effects on the test's performance in practice.

3.2 Methodology

3.2.1 Statistical Model

A nonparametric regression setting is considered:

$$y_i = g(x_i) + \varepsilon_i \tag{3.1}$$

where g is the nonparametric regression function, allowing the mean to be smoothly varying, and ε_i are correlated and normally distributed errors $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n) \sim N_n(0, \sigma^2\Sigma)$ for $i = 1, \ldots, n$ where Σ is the correlation matrix. Here we take our covariates x_i to be equally spaced fixed design points, that is $x_i = i/n$ over 0 to 1, without loss of generality.

There are several assumptions that have to be made about the true underlying regression function.

- The regression function g is taken to be a piecewise continuous function, C^r on the interval $[0, 1]$. This implies that g is smooth except possibly at several points $\tau_1, \ldots, \tau_d$ between 0 and 1, where g is discontinuous.

- The left and right limits $g(\tau_j-)$ and $g(\tau_j+)$ exist and $g(\tau_j-) \neq g(\tau_j+)$ at each discontinuity point τ_j, but they are identical at other points.
- g is also assumed to have at least 2 derivatives, at each $x \in [0,1]\backslash\{0, \tau_1, \ldots, \tau_d, 1\}$.
- The error variance and correlation of the data are assumed to be the same throughout the range of x.

There are two possible kinds of tests: a *local test* and a *global test.* In the case of a local test, the change-point location is assumed to be known and hence the test involves examining whether there is a discontinuity at that particular location. On the other hand, in the case of a global test, no information is given on the change-point locations. The test then involves examining more generally whether there may be one or more change-points in the underlying smooth regression function.

Bowman et al. (2003) considered a discontinuity test for independent data. We will extend their approach to detect possible discontinuities in correlated data.

3.2.2 The test statistic

Our methodology is based on a simple yet ingenious difference-based idea that was introduced earlier in Section 2.2.1. At each evaluation point, z, where we wish to test for a discontinuity, we compare two linear smooths, the left smooth and right smooth of the data, which are obtained by looking at the data to the left and right of the evaluation point respectively. At a particular evaluation point where the change-point occurs, we can expect the difference between the two smooths to be large.

The global test will first be introduced, followed by discussion on how this can be simplified to a local test if the change-point location is pre-specified.

3.2.2.1 Global test

Here, no prior knowledge of the possible change-point locations is given. Our interest lies in testing whether there are any discontinuities in the smooth regression function, g. The hypotheses can be stated as:

H_0: g is continuous over the equally-spaced interval (0,1).

H_1: g is discontinuous at one or more particular points over the equally-spaced interval (0,1).

The evaluation point, z, is taken as the midpoint of pairs of adjacent design points, i.e. $z_i = (x_i + x_{i+1})/2$, where the data are assumed to be ordered by x. The first and last 5 observations are not considered. Ignoring the possibility of discontinuities, we apply a smoother to estimate $g(z)$ at $z \in (0,1)$ as

$$\hat{g}(z_i) = \sum_{j=1}^{n} w_{ij} y_j = W_i Y \tag{3.2}$$

where $Y = (y_1, \ldots, y_n)$ is the vector of observations, w_{ij} is a weight function, and is dependent on the smoothing parameter, h. The weight function can be obtained using various smoothing techniques. This will be discussed in greater detail in Section 3.2.4.

From Equation (3.2), we can obtain the left smooth by smoothing over data that lie on the left of the evaluation point, z_i. This can be expressed as

$$\begin{aligned} \hat{g}_L(z_i) &= \sum_{j=1}^{n} w_{ij} I(x_j < z_i) Y_j \\ &= \sum_{j=1}^{n} w_{L_{ij}} y_j = W_{L_i} Y \end{aligned} \tag{3.3}$$

where I is the indicator function.

Similarly, the right smooth can be obtained via

$$\begin{aligned} \hat{g}_R(z_i) &= \sum_{j=1}^{n} w_{ij} I(x_j > z_i) Y_j \\ &= \sum_{j=1}^{n} w_{R_{ij}} y_j = W_{R_i} Y \end{aligned} \tag{3.4}$$

Note W_i,W_{L_i} and W_{R_i} are row vectors in Equations (3.2),(3.3) and (3.4) respectively.

The difference between the left and right smooths

$$\begin{aligned} r_i &= \hat{g}_L(z_i) - \hat{g}_R(z_i) \\ &= (W_{L_i} - W_{R_i})Y \end{aligned} \tag{3.5}$$

gives an estimate of the jump size. We would expect that r_i will be small if no discontinuity is present but will be large if a discontinuity is present. In computational terms, the vector of the differences of the left and right smooths, r, can be written as DY, where D is an $n \times n$ matrix and has as its ith row $(W_{L_i} - W_{R_i})$.

The idea proposed by Bowman et al. (2003) to deal with independent data was to carry out distributional calculations based on the test statistic

$$\begin{aligned} F_1(h) &= \frac{\sum_{i=1}^{n} r_i^2}{\hat{\sigma}^2}, \\ &= \frac{Y^T D^T D Y}{\hat{\sigma}^2} \end{aligned} \tag{3.6}$$

which is simply the sum of the squared differences of the left and right smooths at each evaluation point divided by an estimate of error variance, $\hat{\sigma}^2$.

An alternative form of the test statistic arises from the fact that individual comparisons $r_i = \hat{g}_L(z_i) - \hat{g}_R(z_i)$ have different variances. Those r_i in the middle of the series have smaller variances as more information is provided to create the two smooths. These are given by the diagonal elements of the variance matrix

$$Var(r) = DVar(Y)D^T = \sigma^2 DD^T$$

If u_i denotes the ith diagonal element of DD^T, then the test statistic is given as

$$\begin{aligned} F_2(h) &= \sum_{i=1}^{n} \frac{r_i^2}{\hat{\sigma}^2 u_i} \\ &= \frac{Y^T D^T \Lambda' D Y}{\hat{\sigma}^2} \end{aligned} \tag{3.7}$$

where Λ' is the inverse of the diagonal matrix whose diagonal entries are those of DD^T.

The noise variance σ^2 can be estimated in a variety of ways, all of which can be written in quadratic form, Y^TBY. The role of the quadratic form for $\hat{\sigma}^2$ in the denominator is to create a test statistic of the form $\frac{Y^TAY}{Y^TBY}$ which scales out the unknown variance of the model, and hence is independent of the true value of σ^2. This will be discussed in greater detail in Section 3.2.6.

The results when using test statistics $F_1(h)$ and $F_2(h)$ discussed in Equations (3.6) and (3.7) respectively will be slightly different. This is because the latter accounts for the fact that the variances of the individual differences between the left and right smooths are different since they depend on the amount of data to the left and right of the evaluation point, z. In contrast, $F_1(h)$ just takes a constant error variance over all the evaluation points. Hence $F_2(h)$ will be more powerful if discontinuities occur in the middle of the series.

When the data are correlated, the basic idea follows through. Though the variance estimator $\hat{\sigma}^2$ proposed in the independent case in Equation (3.6) no longer provides a valid estimate of σ^2, it does not fundamentally alter the operation of the test. This is because the main role of the denominator is to make the test statistic independent of σ^2 by constructing the test statistic as a ratio and this remains true in the case of correlated data.

The distribution of the test statistic is, however, not independent of Σ, as it is an integral part of the quadratic form calculations. One strategy of the test in the case of correlated data is simply to plug in a suitable estimate of Σ in the distributional calculations. This will be discussed later in Section 3.2.3.

Adjustment for correlation in the variances of r_i follows from

$$Var(r) = DVar(Y)D^T = \sigma^2 D\Sigma D^T \tag{3.8}$$

with diagonal vector $\sigma^2 v$. This leads to the test statistic

$$\begin{aligned} F_3(h) &= \sum_{i=1}^{n} \frac{r_i^2}{\hat{\sigma}^2 v_i} \\ &= \frac{Y^T D^T \Lambda' D Y}{\hat{\sigma}^2}, \end{aligned} \tag{3.9}$$

where Λ' is the inverse of the diagonal matrix whose diagonal entries are those of $D\Sigma D^T$. The plug-in estimate of Σ can be used in computing the test statistic. This will, however, produce very similar results as the test statistic in Equation (3.7) when the correlation matrix, Σ, is incorporated in the distributional calculations for both test statistics.

3.2.2.2 Local test

In the situation where there is prior information about the possible location of a change-point, and interest lies in whether the change is an abrupt shift or one that can be attributed to random variation, the test is a simplified version of the global test. Instead of obtaining the sum of the squared differences between the two smooths at various evaluation points, we simply look at the particular location of interest, say τ, where we wish to examine if there is a discontinuity.

The hypotheses can be stated as:

H_0: g is continuous at τ.

H_1: g has a jump discontinuity at τ.

In other words, under the null hypothesis, the difference between the left and right smooths at τ should be approximately zero. However, if a jump is present, then there will be a substantial difference between the two smooths.

Equation (3.9) of the global test can be simplified to

$$\begin{aligned} F_{local_1}(h) &= \frac{r_\tau^2}{v_\tau \hat{\sigma}^2} \\ &= \frac{Y^T d^T dY}{\hat{\sigma}^2 v_\tau} \end{aligned} \tag{3.10}$$

where w_L,w_R and d are vectors of length n, $d = (w_L - w_R)$ is evaluated at τ and $v_\tau = d\Sigma d^T$.

An alternative test statistic is

$$F_{local_2}(h) = \frac{Y^T d^T dY}{\hat{\sigma}^2} \tag{3.11}$$

with Σ taken into account in the distributional calculations for both test statistics.

3.2.3 Reference distribution of the test statistics

To assess the significance of the observed differences, we use the observed value of $F(h)$, which is of the form $F(h) = \frac{Y^T AY}{Y^T BY}$. We can take $F(h)$ as any of the test statistics mentioned above. We will reject H_0 if the observed value of the test statistic $F(h)$ is bigger than some critical value c_α, when H_0 is true, i.e.

$$P(F(h) > c_\alpha) = \alpha$$

Let the constant k be the observed value of the test statistic, then the p-value of the test can be written as

$$\begin{aligned} p &= P(F(h) > k | H_0 \text{ is true}) \\ &= P\left(\frac{Y^T AY}{Y^T BY} > F_{obs}\right) \\ &= P(Y^T(A - kB)Y > 0) \\ &= P(Y^T QY > 0) \end{aligned} \tag{3.12}$$

where A and B are positive definite matrices and $Q = A - kB$.

If, as will be discussed in Section 3.2.4, the bias term of the difference between the left and right smooths which are involved in computing $Y^T AY$ cancels, then $F(h)$ can be written as the ratio of two quadratic forms where the random variables are normal with means approximately zero. Hence

$$Y^T QY \approx \varepsilon^T Q\varepsilon \qquad \text{where } \varepsilon \sim N(0, \sigma^2\Sigma),$$

where Σ is the correlation matrix of Y. We can take the variance, σ^2 as 1 because of the original scaling form of the test statistic.

Based on well established general results about the distribution of quadratic forms (Johnson and Kotz, 1972), the probability p can be evaluated to any desired level of accuracy, in numerical form. The cumulants of $\varepsilon^T Q\varepsilon$ are calculated easily as functions of $\text{tr}(Q\Sigma)^j$. The

exact expression is given by

$$\kappa_j = 2^{j-1}(j-1)!tr(Q\Sigma)^j, \tag{3.13}$$

where $tr(\cdot)$ denotes the trace operator. From expression (3.13), we can compute the mean (κ_1) and variance (κ_2) and any higher order cumulants.

However, for the purpose of hypothesis testing, three significant figures will be sufficient for evaluating the p-values. They can be more conveniently obtained by matching the first three or four moments of the distribution of $\varepsilon^T Q\varepsilon$ to a more accessible distribution. Bowman and Azzalini (1999) suggested that the final quadratic form, $Y^T QY$, can be handled through an accurate approximation by matching moments to a shifted and scaled χ^2 distribution of the form $a\chi^2_b + c$.

The constants a, b and c can be obtained via the expressions below

$$a = |\kappa_3|/(4\kappa_2), \qquad b = (8\kappa_2^3)/\kappa_3^2, \qquad c = \kappa_1 - ab.$$

The p-value of the observed test statistic can then be estimated as $1 - q$, where q is the probability lying below the point $-c/a$ in a χ^2 distribution with b degrees of freedom.

3.2.4 Choice of smoothing techniques

The smoothers used throughout the thesis are *linear* smoothers, which can be written in the form $\hat{g}_n = SY$, where S is a suitable $n \times n$ smoothing matrix whose rows consist of the weights corresponding to each observed covariate value x and Y is the vector of observed responses. A favourable property of such linear smoothers is that the means and variances of the estimators are easily calculated.

In addition, most of the important distributional properties of nonparametric regression estimators follow from their characterisations as linear functions of the response variables y_i. For instance, in our test statistic, if the response variables are assumed to be normal, then having an estimator DY as a linear structure of the response variables allows us to write the numerator of our test statistic in quadratic from as $Y^T D^T DY$. We can then make use of

distributional results from the theory of quadratic forms of random variables to assess the significance of our test as described above.

As our test statistic is dependent on the left and right smooths of the data, which are therefore estimating at a boundary, the performance of how well the smoothers $\hat{g}$ perform at the boundary should be considered in the choice of smoothers to be used. Boundary behaviour in smoothers has been studied quite extensively in the literature (Trevor and Loader, 1993), and *local polynomial estimators* have been found to perform reasonably well in this aspect (Wand and Jones, 1995). Furthermore, the generalisation of how these estimators can be used to estimate higher dimensional surfaces is quite straightforward. Hence throughout this work, the *local polynomial estimators*, which are a form of linear smoothers since they can be expressed as weighted arithmetic means of the observed Y's are used.

A Taylor expansion for sufficiently smooth functions is given as

$$g(z) \approx g(x) + g'(x)(z-x) + g''(x)(z-x)^2\frac{1}{2!} + \cdots + g^{(p)}(x)(z-x)^p\frac{1}{p!} \tag{3.14}$$

A general local polynomial estimator of degree p of $g(z)$ proceeds by fitting a polynomial of degree p via weighted least squares, locally around x. In other words, $\hat{g}(z)$ is $\hat{\beta}_0$, where $(\beta_0, \ldots, \beta_p)$ is minimised by $(\hat{\beta}_0, \ldots, \hat{\beta}_p)$ in the expression

$$\sum\{Y_i - \beta_0 - \ldots - \beta_p(x_i - z)^p\}^2 K_h(x_i - z) \tag{3.15}$$

where $K_h(\cdot)$ is a kernel weight function defined as $K_h(\cdot) = K(\cdot/h)/h$, and $K_h(\cdot)$ is a symmetric (around zero) probability density $K(\cdot)$ and h is the bandwidth or smoothing parameter which controls the size of the local neighbourhood.

An exact expression for a general polynomial estimator can be obtained. (see Wand and Jones, 1995, pg. 119). Two special members of the local polynomial smoothers are the local mean estimators ($p = 0$) and the local linear estimators ($p = 1$).

The local mean estimators are also known as local constant fits, since they correspond to fitting local polynomials of degree zero. This is also known as the Nadaraya-Watson estimator (Nadaraya, 1964), defined as

$$\hat{g}(z) = \frac{\sum_{i=1}^{n} K_h(X_i - z)Y_i}{\sum_{i=1}^{n} K_h(X_i - z)}$$

A further well-known kernel estimator is the Gasser-Muller estimator (Gasser and Müller, 1979), defined as

$$\hat{g}(z) = \sum_{i=1}^{n} Y_i \int_{s_i-1}^{s_i} K_h(u - z)du$$

with $s_i = (X_i + X_{i+1})/2, X_0 = -\infty$ and $X_{n+1} = +\infty$, where the data are sorted in increasing order over the X variable. This is an estimator of a different type and its advantage over the Nadaraya-Watson estimator is that it does not require a normalising denominator as its weights sum to 1. This property is helpful in the derivation of the asymptotic theory of the smoother. It is also suitable to be used with both equally and unequally spaced designs.

The local linear estimator can be explicitly defined as

$$\hat{g}(z) = \frac{\sum_{i=1}^{n} w_i Y_i}{\sum_{i=1}^{n} w_i}, \qquad w_i = K_h(X_i - z)(S_{n,2} - (X_i - z)S_{n,1})$$

where $S_{n,j} = \sum_{i=1}^{n} K_h(X_i - z)(X_i - z)^j$.

In particular, the behaviour of the boundary bias of the local linear regression estimator ($p = 1$) is superior to that of the local constant estimator ($p = 0$), and is comparable to other linear estimators (Fan, 1993). In addition, the local linear smooths have a favourable design adaptive property in that the bias component does not depend on the pattern of the design points, described by the density function $f(x)$, at least asymptotically. Another favourable property of the local linear estimate is that it has no sample bias if the true regression function is linear (Ruppert and Wand, 1994), whereas the local mean estimator does not possess this finite sample property (see Fan and Gijbels, 1996, pg. 63).

Fan and Gijbels (1996) showed that for p odd, the asymptotic conditional bias and variance are of the same order, h^{p+1}, in the interior as well as the boundary, unlike that for p even, where the leading term in the asymptotic bias has a more complicated form. Hence

it is possible to extend the local linear smoothers to higher odd powered local polynomials.

Since correlation does not affect the bias but does have a considerable effect on the variance, at a point z on the boundary the asymptotic bias of the local polynomial estimator of degree p is similar to the independent case, and can be expressed as

$$E(\hat{g}(z)) - g(z) = \frac{h^{p+1}}{(p+1)!} B(K) g^{p+1}(z) + o(h^{p+1}) \tag{3.16}$$

where $B(k)$ is a function of the kernel (Wand and Jones, 1995, Section 5.4, 5.5). From Equation (3.16), we can express the biases of the left and right smooths of a local linear estimator respectively as

$$E(\hat{g}_L(z_i)) - g(z_i-) = \frac{h^2}{2} B(k) g''(z_i-) + o(h^2), \tag{3.17}$$

$$\text{and } E(\hat{g}_R(z_i)) - g(z_i+) = \frac{h^2}{2} B(k) g''(z_i+) + o(h^2), \tag{3.18}$$

Under the null hypothesis, where there is no change-point, we have $g(z_i-) = g(z_i+) = g(z_i)$. It then follows from Equations (3.17) and (3.18) that

$$\begin{aligned} E(r_i) &= E(\hat{g}_L(z_i) - \hat{g}_R(z_i)) \\ &= E(\hat{g}_L(z_i)) - E(\hat{g}_R(z_i)) \\ &= o(h^2) \end{aligned} \tag{3.19}$$

The bandwidth, $h = h_n$ is taken to be a sequence satisfying $h \to 0$ and $nh \to \infty$ as $n \to \infty$. For a more general p, where p is odd, $E(r_i) = o(h^{p+1})$.

Taking $Dy = Dg + D\varepsilon$, which can be written as $Dy = o(h^2) + D\varepsilon$, it then follows that $Y^T D^T DY = \varepsilon^T D^T D\varepsilon + o(h^2)$. Hence, we can observe that the numerator of the test statistic $Y^T D^T DY$ is approximately equivalent to $\varepsilon^T D^T D\varepsilon$. This leads to $F(h)$ being the ratio of quadratic forms in Normal random variables with means approximately zero. This is particularly attractive for inference since the distributional properties of a quadratic form for centred, zero mean normal random variables are well established (Johnson and Kotz, 1972, Chapter 29).

Other linear smoothers can also be considered for our discontinuity test if they have favourable boundary behaviour. For the choice of the kernel function, we have adopted the Gaussian probability density as its unboundedness provides finite conditional and unconditional variance of the estimators (see Simonoff, 1996, pg. 141). However, it has been noted that the performance of the resulting estimators is not influenced very much by the choice of the kernel function, both theoretically and empirically (see Fan and Gijbels, 1996, pg. 76).

The choice of the smoothing parameter is a more crucial question, which we will discuss inthe following section.

32.5 Choice of smoothing parameter

The smoothing parameter or bandwidth, h, defines the width of the kernel function and hence determines the degree of smoothing applied to the data. Using a normal probability density as the kernel function, $K_h(x-z;h) = \phi(x-z;h)$, where $\phi(x-z;h)$ is the normal density function with mean 0 and standard deviation h, a smoothing parameter value of h will imply smoothing over an effective distance of $4h$ on the covariate axis.

As we can see from the various forms of the test statistic, $F(h)$ is dependent on the value ofthe smoothing parameter h used. In the author's proposed nonparametric regression test, the choice of the smoothing parameter makes an implicit assumption on the smoothness of the underlying trend. The remaining smooth variation is attributed to the correlation of the errors.

There are numerous reasons why it will be very difficult to choose an appropriate smoothing parameter h for our discontinuity testing.

- Firstly, even in the simplest case, with independent data, the degree of smoothing is important, as it determines the nonparametric estimate of the trend function to a large extent. The most appropriate smoothing parameter depends on the roughness of the unknown underlying function $g(\cdot)$ expressed in the derivative $g''(\cdot)$. Without this knowledge, it is extremely difficult to decide how much of the irregularity of the observed data to attribute to the trend, and how much to the error variance.

- In addition, if correlation is present, the difficulty becomes greater when we have no

prior knowledge of either the parametric shape of the trend or the correlation function. Local trends in the data may be due either to movement in the underlying regression function or to correlation in the errors.

- To avoid having to use a smoothing parameter chosen by trial and error, data-driven selection methods that account for correlation have been developed to select appropriate values (see e.g. Altman (1990), Herrmann et al. (1992) and a recent survey by Opsomer et al. (2001)). Other works include that of Hart (1991, 1994) who showed that when the errors are positively correlated, the conventional cross validation approaches will tend to undersmooth even when the correlation is low. Though cross-validation provides an automatic choice of smoothing parameter, it is computationally very intensive and often suffers from high variability (Hall and Johnstone, 1992). Moreover, to complicate the issue further, these bandwidth selection methods, even after correcting for correlations, break down in the presence of discontinuities in the regression function. Consequently, the selected bandwidth is underestimated, as these automatic selection techniques perceive the mean function to be rougher than it really is. Thus applications of such automatic selection techniques are unlikely to be that straightforward.

- Oversmoothing or undersmoothing might lead to inaccurate conclusions on the presence of change-points.

In summary, it is difficult to make an appropriate global bandwidth choice when both correlation and discontinuities may be present. Without prior knowledge of the irregularity of the underlying function $g''(x)$ and the correlation of the error, the complications of this highly challenging task become even greater.

However, although the choice of the smoothing parameter is highly influential in the performance of our test, our main aim here is to use it for inference and not to estimate the underlying regression. If the latter is our intent, then techniques to optimise the properties of the estimator in their choice of smoothing parameter should be used. However, if the smoothing parameter is to be used for the purpose of inference, then its selection may not be as critical.

To avoid the specific selection of the smoothing parameter, Azzalini and Bowman (1993) proposed the idea of a significance trace, which is simply a plot of p-values against the smoothing parameters. In this approach, they avoided specifying the value of h that is to be used for the discontinuity testing. This approach is implemented here, not only because of the reasons that are mentioned above, but also due to the fact that the smoothing parameter will be random, if an automatic data driven technique is adopted, since it will be a function of the responses, Y. The weight function, W which are functions of h will also be random and hence the smooth $\hat{g}(z)$ is no longer a simple linear function of Y.

Henceforth, we will use a fixed bandwidth in our methodology. In certain situations, some knowledge of the application might be able to be used to suggest an appropriate size of smoothing parameter, for example from knowledge of the physical distance over which a disturbance in the system generating the data may have an effect.

3.2.6 Choice of variance estimator

In order to carry out inference, we require not only to estimate the underlying regression function g but also the error variance, σ^2. This latter estimate is also very important.

In this section, we will discuss some of the variance estimators that can be considered for the discontinuity test. In order for the final test statistic to be expressed in the form of a ratio of quadratic forms, so that the distributional results described above can be employed, variance estimators that can be written in quadratic form $Y^T BY$ are used. Here, B denotes a $n \times n$ matrix, consisting of constants, depending on the design space.

A natural variance estimator that can be considered is the Residual Sum of Squares (RSS) based estimator of σ^2, which defines the residuals as differences between the responses, y_i and a fitted value, $\hat{g}(x_i)$. In the nonparametric setting,

$$\begin{aligned} RSS &= \sum_{i=1}^{n} \{y_i - \hat{g}(x_i)\}^2 \\ &= y^T (I - S)^T (I - S) y \end{aligned} \tag{3.20}$$

where S is the smoothing matrix obtained using bandwidth h. The RSS based estimator is

then given by

$$\hat{\sigma}^2 = \frac{RSS}{df_{error}} \tag{3.21}$$

where the RSS is adjusted by df_{error}, the error degrees of freedom.

Analogous to the linear parametric model, df_{error} denotes the difference between the sample size n and the degrees of freedom of the smoother. Hastie and Tibshirani (1990, pg 52-55) provided various suggestions for error degrees of freedom, and one of them being $df_{error} = n - tr(2S - S^TS)$, assuming that the errors are independent. This is motivated by the expection of the RSS from a fitted nonparametric model.

$$\begin{aligned}
E(RSS) &= E\{y^T(I-S)^T(I-S)y\} \\
&= E\{(g+\varepsilon)^T(I-S)^T(I-S)(g+\varepsilon)\} \\
&= E\{g^T(I-S)^T(I-S)g + \varepsilon^T(I-S)^T(I-S)\varepsilon + 2g^T(I-S)^T(I-S)\varepsilon\} \\
&= E\{g^T(I-S)^T(I-S)g\} + E(\varepsilon^T I\varepsilon) - 2E(\varepsilon^T S\varepsilon) + E(\varepsilon^T S^T S\varepsilon) \\
&= g^T(I-S)^T(I-S)g + n\sigma^2 - 2\sigma^2 tr(S\Sigma) + \sigma^2 tr(S^T S\Sigma) \\
&= g^T(I-S)^T(I-S)g + \sigma^2\{n - tr(2S\Sigma - S^T S\Sigma)\}
\end{aligned} \tag{3.22}$$

Here, in the case where the data are independent, Σ is simply the identity matrix, and the $df_{error} = n - tr(2S - S^TS)$ is as recommended above. However, we have allowed for Σ in our notation, for later reference. The first term $g^T(I-S)^T(I-S)g$ is the sum of the squared biases over the design points, and if it is negligible, then Equation (3.21) with the defined df_{error}, will give a good estimator of σ^2. However as all nonparametric models do contain bias, $g^T(I-S)^T(I-S)g$ will be greater than 0, and so the estimator will overestimate σ^2 to some extent.

One disadvantage of the RSS estimator is that it is very dependent on the fit to the data. In particular, it also raises the question of what smoothing parameter is appropriate. In his simulation study, Bock (1999) suggested that undersmoothing yields a better estimate of σ^2 in terms of bias. However one still has to decide what degree of undersmoothing is suitable. This highlights the advantage of the next class of variance estimators, difference

based estimators, which are independent of the smoothing parameter. It was highlighted by Bock (1999) that the accuracy of these difference based estimators is comparable to the best RSS based estimators over both the equally and unequally spaced design settings.

The difference based estimator defines the residuals differently from the RSS approach. The residuals here are obtained by differencing the observed responses via various ways to remove the trend. Hence it neither requires the regression function to be estimated explicitly, nor any type of smoother or smoothing parameter. Different forms of these exist in the univariate case, two of which are investigated here, namely that of Rice (1984) (*first order differences,* $r = 1$), and Gasser et al. (1986) (*second order differences,* $r = 2$).

A simple estimator proposed by Rice (1984) for independent data is expressed as

$$\hat{\sigma}^2 = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} (y_{i+1} - y_i)^2 \tag{3.23}$$

where it is assumed that the observations (x_i, y_i) have been ordered by x_i. This estimator can also be expressed in matrix form as

$$\hat{\sigma}^2 = \frac{Y^T \Delta Y}{tr\Delta} \tag{3.24}$$

where the matrix Δ is defined as

$$\Delta = \begin{Bmatrix} 1 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \ddots & \vdots \\ 0 & -1 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 2 & -1 \\ 0 & \cdots & 0 & -1 & 1 \end{Bmatrix}$$

By considering the expected value of $(y_i - y_{i-1})^2$,

$$\begin{aligned} E(y_i - y_{i-1})^2 &= E\left[\{y_i - g(x_i)\} - \{y_{i-1} - g(x_{i-1})\} + \{g(x_i) - g(x_{i-1})\}\right]^2, \\ &= E(\varepsilon_i{}^2) - 2E(\varepsilon_i \varepsilon_{i-1}) + E(\varepsilon_{i-1}{}^2) + \{g(x_i) - g(x_{i-1})\}^2 \\ &= 2\sigma^2 - 2cov(\varepsilon_i, \varepsilon_{i-1}) + \{g(x_i) - g(x_{i-1})\}^2, \end{aligned} \tag{3.25}$$

we can see that when the data are independent, $cov(\varepsilon_i, \varepsilon_{i-1}) = 0$, so $E(y_i - y_{i-1})^2 = 2\sigma^2 + \{g(x_i) - g(x_{i-1})\}^2$ and that σ^2 is inflated due to the fluctuation in the underlying regression function g. However this effect will decrease if the amount of data increases such that the covariate values are spread more densely over the whole design space.

A second order difference estimator proposed by Gasser et al. (1986), involves creating pseudo-residuals, by taking the difference between y_i and the line which connects its two nearest neighbours.

$$\tilde{\varepsilon}_i = \frac{x_{i+1} - x_i}{x_{i+1} - x_{i-1}} y_{i-1} + \frac{x_i - x_{i-1}}{x_{i+1} - x_{i-1}} y_{i+1} - y_i \tag{3.26}$$

By expressing $\tilde{\varepsilon}_i = a_i y_{i-1} + b_i y_{i+1} - y_i$, Gasser's error variance estimator is provided as

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=2}^{n-1} \frac{1}{(a_i^2 + b_i^2 + 1)} \varepsilon_i^2 \tag{3.27}$$

This estimator is more capable of reducing the effect of the underlying function in inflating the estimate.

These estimators also have the properties of having a small bias for small sample sizes (Dette et al., 1998, pg. 754-755) and of being very simple computationally. Though they do not achieve the optimal rate of some residual based estimators, their performance can be comparable (see Hall and Marron, 1990).

Dette et al. (1998) did an extensive simulation study which showed that factors such as sample size, magnitude of residual variance, and the order of differencing of the difference based estimators do influence the performance of the different estimators. However, the most significant factor is still the underlying trend function itself. In the presence of possibly highly oscillating functions, or faced with extremely noisy data, the variance estimator of Gasser ($r = 2$) or any order greater than 2 might be preferable.

The variance estimators that have been discussed so far in this section are evaluated for errors that are independent, where $\Sigma = I$. However, in the setting where the data are correlated, these estimators will break down if the data are treated as independent. To estimate the variance of a set of correlated data, we need to apply a correction factor to the

variance estimators for independent data, as we can observe from the expected values of the estimators in Equation (3.22) and (3.25). Nevertheless, for the purpose of inference, as the test statistic is constructed in such a way to scale out the unknown variance of the model, it is not necessary to correct for correlation for the variance estimation. A good estimation of the variance in correlated data is however required in the construction of the reference band of the estimator. The latter will be discussed in Section 3.2.8.

In fact, when the variance estimator can be written in quadratic forms, $y^T D_v^T D_v y$, the general form of the expected values are

$$E(y^T D_v^T D_v y) = g^T D_v^T D_v g + \sigma^2 tr(D_v^T D_v \Sigma) \tag{3.28}$$

If $g^T D_v^T D_v g$ is negligible, then

$$\hat{\sigma}^2 = \frac{y^T D_v^T D_v y}{tr(D_v^T D_v \Sigma)} \tag{3.29}$$

From Equation (3.29), we can see that it is quite straightforward to adjust the estimators used in the independent setting to correct for the presence of correlation. For instance, in the RSS based estimator, D_v is simply $I - S$. The estimator, however, requires knowledge of the often unknown parameter, Σ. Section 3.2.7 will discuss how this can be estimated. The issue of variance estimation in the nonparametric regression setting with time series errors is also discussed in Francisco-Fernandez and Villar-Fernandez (2001) and Hall and Keilegom (2002).

In our test procedure, we have used the Rice and Gasser variance estimators, because overall, they have more favourable properties than the RSS based approach. The performance of the test will depend on how well each variance estimator performs under different settings of the model, for example how Rice's method performs under a smoother trend compared to Gasser's.

3.2.7 Adjusting the test for correlation

Another difficulty that we encounter, in addition to that of the choice of the smoothing parameter for the discontinuity test, is the estimation of the underlying correlation of the data. Since the correlation is a nuisance parameter, it is not crucial that we model it exactly. It seems reasonable in most cases to adopt a simple AR(1) model as it is convenient and should capture most of the main structure of the correlation. In our test for data which are equally spaced, we therefore use a correlation matrix Σ that follows an autoregressive AR(1) model to model the correlation. This can be expressed as

$$\varepsilon_i = \alpha\varepsilon_{i-1} + Z_i$$

where Z_i is a purely random process with $E(Z_i) = 0$ and $Var(Z_i) = \sigma_z^2$, and α is the AR(1) correlation coefficient. The series will converge if $|\alpha| < 1$ and Σ can be expressed as

$$\Sigma = \begin{bmatrix} 1 & \alpha & \alpha^2 & \cdots & \alpha^{n-1} \\ \alpha & 1 & \alpha & \cdots & \alpha^{n-2} \\ \alpha^2 & \alpha & 1 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha^{n-1} & \cdots & \cdots & \cdots & 1 \end{bmatrix}$$

To estimate this unknown correlation, we need first to remove the trend that might be present in the data. It is a common time series procedure to estimate the trend of the series by some method and then analyse the residual series. In the situation where the trend is smoothly varying, it will be most appropriate to apply nonparametric regression to remove the trend and obtain the residuals. We will refer to this technique as the *residual approach.*

However, this approach faces several challenges. The first one being the choice of the smoothing parameter which is denoted here as *h.trend*, to remove the trend. This is similar to the difficulties encountered in the choice of the smoothing parameter for the discontinuity test that has been mentioned earlier in Section 3.2.5.

It is difficult to select a suitable *h.trend* in the presence of an unknown correlation, and

this difficulty is enhanced even more in the presence of discontinuities. However, using an appropriate *h.trend* to remove the trend function is of considerable importance, as the residuals obtained as a result are used to estimate the correlation present in the data. Using too small a smoothing parameter will trace the data too closely, and correlation of the residuals will be underestimated, resulting in inflation of the size of discontinuity test. Using too large a smoothing parameter will oversmooth the data. For example, a very large bandwidth will result in almost fitting a straight line, when a more smoothly varying function might be more suitable. As the trend of the data is not totally removed, it will 'flow' into the residuals and cause the correlation to be overestimated. In addition, if a jump is present, oversmoothing will result in a set of consecutive residuals that appear to be correlated, possibly causing an overestimation of the correlation. This might result in the test having lower power.

Since the bandwidth parameter for the discontinuity test, denoted as *h.test*, makes an implicit assumption on the smoothness of the underlying trend, it would be natural to proceed using the same bandwidth to remove the trend, i.e. $h.trend = h.test$. However in view of the fact that there might be possible jumps present in the data, an alternative proposal would be using a smaller *h.trend*, e.g. $h.trend = h.test - 2/n$. (Subtraction of $2/n$ effectively reduces the number of observations over which smoothing takes place by 8, at any particular point.) The reasoning behind this is that in being slightly conservative in our choice of *h.trend* (compared to *h.test*), we hope to not only remove any underlying trend,

The residual approach

1st stage:

- Remove the trend in the data by fitting a smooth curve using a particular smoothing parameter, *h.trend*.
- Using the residuals, we obtain an estimate of the correlation matrix, Σ, by estimating the sample correlation coefficient, $\hat{\alpha}$.

2nd stage:

- Insert the estimate of Σ into the discontinuity test, with smoothing parameter, *h.test* and obtain the p-value of the test.

Figure 3.1. The residual approach

but also any considerable jump that might be present. Then the remaining variation in the residuals will be attributed to the correlation in the data to obtain a better estimate. It thus seems reasonable to use a slightly smaller bandwidth for the first step. The procedure can be summarised in Figure 3.1.

Another estimator of correlation that can be considered is based on a *moving window approach.* Unlike the *residual approach*, where we have to decide on the smoothing parameter for the trend removal, this technique avoids this extremely difficult task by assuming that the data considered within a moving window size are stationary (with constant mean and no discontinuity). However, the size of the moving window, b, has to be selected. The procedure is outlined in Figure 3.2.

The moving window approach

1st stage:

- Use a *moving window* of size b and find the sample correlation of the first b data points, $y_1, y_2, \ldots, y_b$, in that particular window.
- Shift the moving window of size b by one observation, and find the sample correlation of the next b data points, $y_2, y_3, \ldots, y_{b+1}$. Repeat this till it covers the last observation.
- The estimated sample correlation of the whole data-set is obtained by the median of the estimated sample correlations in $(n - b)$ windows.

2nd stage:

- Insert the estimate of Σ into the discontinuity test, with bandwidth, *h.test*, and obtain the p-value of the test.

Figure 3.2. The moving window approach

The sizes of the moving windows that we have considered are $b = \{n/5, n/4, n/3\}$ which are moderate in size. The advantage of the moving window procedure is that it might be more robust to the presence of jumps in the data, compared to the residual approach. However, it is not without drawbacks. Care has to be taken to select the moving window size. A small window size might result in underestimation of correlation, causing an inflation in size. On the other hand, too large a window might result in overestimation of correlation

if most of the moving windows calculate the sample correlations in the presence of trends or jumps.

In the situation where the location of a possible change-point is given, it is much easier to calculate the correlation of the data. The correlation can be estimated by considering the two segments partitioned by the potential change-point separately. Alternatively, a left and right smooth can be fitted at the potential change-point location. Any difference between the two smooths is evaluated and the data are adjusted to remove the discontinuity. A smooth curve can be used to remove any underlying trend to obtain the residuals, which are then used to estimate the correlation of the data. After obtaining an estimate of Σ, this is then used in the reference distribution calculations to evaluate if the test of discontinuity is significant.

3.2.8 Reference band

In the situation where the null hypothesis is rejected, providing sufficient evidence to suggest that there is at least one discontinuity in the data, it will be useful to have a graphical tool to indicate where the possible discontinuities might be located. To examine that, we can observe how different the left and right smooths are at different evaluation points. This can be done by plotting the left and right smooths, superimposed with a reference band (Bowman et al., 2003).

If it is appropriate to assume that the error variance follows a normal distribution, then the nonparametric estimator $\hat{g}(x)$ will also be normally distributed. Even in cases where it might not be appropriate to assume normality for the error variance, mild assumptions might still allow a normal approximation to be employed for $\hat{g}(z)$, by utilising a form of the central limit theorem.

Because of the normality of $\hat{g}(z)$, the random variable,

$$Q(z) = \frac{(\hat{g}_L(z) - \hat{g}_R(z))}{\sqrt{Var(\hat{g}_L(z) - \hat{g}_R(z))}}$$

has a standard normal distribution under H_0.

Then for each evaluation point z in the interval, and under H_0, it is expected with an approximate probability of 95% that

$$-2\sqrt{Var(\hat{g}_L(z) - \hat{g}_R(z))} \leq \hat{g}_L(z) - \hat{g}_R(z) \leq 2\sqrt{Var(\hat{g}_L(z) - \hat{g}_R(z))} \qquad (3.30)$$

where the computation of $Var(\hat{g}_L(z) - \hat{g}_R(z))$ involves using both Equations (3.8) and (3.29).

Instead of plotting the above functions, we first divide through by 2, and then center the plot on the average of the left and right smooths, $(\hat{g}_L + \hat{g}_R)/2$. The reference band is thus constructed by taking the average of the left and right smooths plus or minus one estimated standard deviation of $g_L - g_R$. Under H_0, it is expected that the left and right smooths should lie in the reference band with a probability of 0.95. It is a useful graphical follow-up to indicate where the possible change-points might be, since if both the left and right smooths leave the shaded region (reference band) then they are separated by more than two standard deviations, indicating the possible presence of a discontinuity. On the other hand, it can also help to explain why certain apparent differences between the two smooths do not lead to a significant result, through the structure of the variance.

However, the reference band is only a visual guide as an approximate graphical tool for checking discontinuities due to the pointwise nature of the bands, as it does not take into account the multiple testing problem.

As can be observed from Equation (3.30), the construction of the reference bands involves the error variance. Hence, a good estimate of the error variance in the presence of correlation is very important.

3.2.9 Change-point Significance Trace (CPST)

As the estimated change-point locations often depend on the smoothing parameter that is used for the discontinuity testing, it will be useful to see how these change over a range of smoothing parameters. A change-point locator plot is proposed to serve this purpose. This comprises of an image plot that shows the probable change-point locations, which have absolute values of the standardised differences greater than 2.5, i.e. $|st.diff| > 2.5$, over the range of smoothing parameters. The intensity of the tones of the images are determined

by the values of the standardised differences. The higher the absolute value, the darker the tone is at that particular location and bandwidth.

This plot together with the significance trace, and the plot of the size of the highest standardised difference between the left and right smooths (which corresponds to the most probable change-point location), is termed as the ***change-point significance trace, CPST***. This is very useful because the significance of the test, the possible change-point locations, and the size of abrupt change at the most probable change-point location, over various smoothing parameters, can all be visualised.

3.3 Simulation Study

In this section, we report on a simulation study to investigate how the discontinuity test performs when it is applied to a set of equally spaced data that are *correlated*. The global test is used for this section, except for Section 3.4.7 where we investigate the size and power of the local test.

The size of each test is assessed by simulating 200 sets of data from a model which follows a particular dependence structure, and applying the discontinuity test. The power simulations are carried out by using the same pattern of simulations, but adding a change-point of a particular jump size at the mid-point of the data. The *empirical size* and *power* are used to assess the performance of the discontinuity test and the effects of a variety of factors.

An extensive simulation study is carried out to investigate the effects of varying the following features of the data:

- shape of trend function
- size of correlation
- sample size
- ratio of jump to standard deviation

and the following aspects of the testing procedure:

- smoothing parameter (global bandwidth)
- estimators of error variance
- estimators of correlation

Given the complexities of the different factors that influence the test, the simulations are carried out incorporating a range of smoothing parameters. In other words, the simulations are repeated at several levels of the smoothing parameter and plots are made of the empirical size ($\alpha = 0.05$) and the power against the smoothing parameter values, *h.test*.

Firstly, we will consider the scenario where the correlation is known in Section 3.4, and then in Section 3.5, where the correlation is unknown. Although in practice we rarely know the true correlation of the observed data, the known correlation case is considered here in order to assess the behaviour of the discontinuity test when other factors are varied, and also later to assess the performance of the estimators of correlation.

In the following sections, 200 simulated sets of an AR(1) model, with a specified trend function, are generated over an equally-spaced setting between 0 and 1. The simulations are initialised using the same random number seed. In other words, the same 200 different sets of noise sequences are used throughout the study. The data conditions and parameters used in the test approaches are summarised in Table 3.1. The sample sizes of 50 and 100 are chosen to represent low and moderate values often encountered in practice. The following values for the data conditions and test approaches as shown in bold are the ones used by default unless specified otherwise. A considerable jump of 3 in the presence of noise variance, ${\sigma_z}^2$ of 1 is used for most simulations, as we hope to observe the difference in the performance of a large range of parameters, in particular from low to highly correlated data. However, the effects of other jump sizes are considered in Section 3.4.6. As the value of ${\sigma_z}^2$ is kept constant throughout the simulations, this implies that the variance of the noise, σ^2 is ${\sigma_z}^2/(1-\alpha^2)$. The effect of correlation is thus confounded with the effect of the signal to noise ratio. The global test statistic $F_3(h)$ from Equation (3.9) is used.

Data conditions	
Design space	**Equally-Spaced**
No. of simulations	**200**
Sample size	$n = 50, \mathbf{100}$
Trend function	**flat**, linear, quadratic and **sine** functions
Error	$\varepsilon \sim N(0, \sigma^2\Sigma)$ where $\varepsilon_i = \alpha\varepsilon_{i-1} + Z_i$ with $Var(Z_i) = {\sigma_z}^2 = \mathbf{1}$
Correlation	$\alpha = corr = \{-0.8, -0.6, -0.4, -0.2, \mathbf{0}, \mathbf{0.2}, \mathbf{0.4}, \mathbf{0.6}, \mathbf{0.8}\}$
Size of Jump	jump = 1, 2, **3** at midpoint of data
Test Approaches	
Smoothing parameter	$h.test = \{\mathbf{0.08}, \mathbf{0.12}, \mathbf{0.16}, \mathbf{0.2}, 0.24, 0.28\}$
Estimator of error variance	**Rice**, Gasser
Algorithm to adjust for correlation	- none (for Section 3.4); - residual approach using $h.trend = \{0.06, 0.08, 0.1, 0.12\}$ - moving window approach using $b = \{n/5, n/4, n/3\}$

Table 3.1. Simulation settings for one-dimensional equally spaced data. The parameters in bold are the ones used by default in the simulations, unless specified otherwise.

3.4 Equally spaced setting: correlation known

In this section, we examine how the test performs under different data conditions, when the correlation is assumed known. In each of the subsequent sections, only the parameter of interest is varied, leaving the rest of the data, or parameters for test approaches unchanged. Sections 3.4.1 - 3.4.6 examine in sequence the effects of trend function, of treating correlated data as independent, of correlation, sample size, variance estimator and ratio of jump to standard deviation respectively. Lastly, in Section 3.4.7, the local test statistic (provided in Equation (3.10)) is compared to the global test.

3.4.1 Effects of shape of trend function

To begin with, we will look at the effect of the shape of trend functions on the size and power of the test. This is done to assess how well the test performs when there is an underlying

trend in the data, which often occurs in time series data. Four different trend functions are considered. They are namely the

- Flat function, $g(x) = 0$, or any constant,
- Linear function, $g(x) = x$,
- Quadratic function, $g(x) = 4x^2$, and
- Smoothly varying (or sine) function $g(x) = sin(2\pi x)$.

As an illustration, the different trend functions are displayed in the top panel of Figure 3.3. Sets of 100 realisations from an AR(1) process with correlations of 0 and 0.6, and error

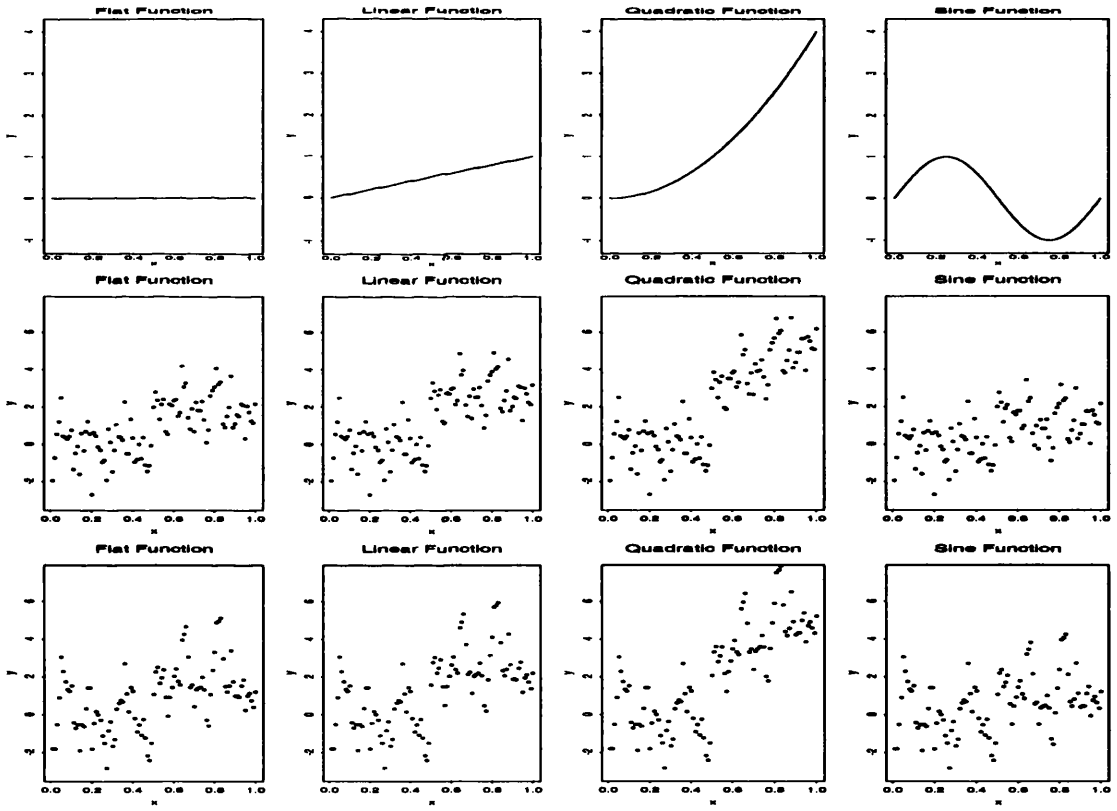

Figure 3.3. The top panel displays the four trend functions: Flat, Linear, Quadratic and Sine (from left to right plots). The middle and bottom panels display data with the various trends and with correlations of 0 and 0.6 respectively. A jump of 2 is also added at the midpoint of the data.

variance of 1, are simulated and added to the different trend functions, shown in the middle and bottom panels of the same figure respectively. A jump of size 2 at the midpoint of the data is also added. Two main observations from the figures can be made.

1. The jump is more obvious in the first three functions (flat, linear and quadratic) than in the sine function as the data appear to vary more in the latter case. Hence we would expect that the power to detect a jump in the sine function will be lower.

2. A high correlation of 0.6 seems to make the jump less obvious to the eye. The data appear to be more variable. The correlated data is also equivalent to a reduced sample size with less information. Thus, we would expect that it would be more difficult to detect a jump when the data are highly correlated. This difficulty is further enhanced when the underlying trend is more irregular.

Here, we will carry out a study of 200 simulations of data with correlation 0.2, having four different trend functions. For the power calculations, the data are shifted by a jump size of 2 added at the midpoint. (A jump of 2 is considered here to allow the reader to observe the difference in performance among the four trend functions.) A wide range of smoothing parameters, *h.test* from 0.08 to 0.28 and Rice's variance estimator are used. A point to highlight is that a smoothing parameter of 0.16 and above is effectively smoothing over quite a large range of at least 0.64 of the data. The results are shown in the Table 3.2 and are plotted in Figure 3.4.

h.test	**Size**				**Power**			
	Flat	Linear	Quad	Sine	Flat	Linear	Quad	Sine
0.08	0.050	0.050	0.050	0.055	0.170	0.170	0.165	0.170
0.12	0.030	0.030	0.025	0.040	0.340	0.340	0.340	0.325
0.16	0.025	0.025	0.030	0.110	0.525	0.525	0.535	0.375
0.20	0.030	0.030	0.035	0.245	0.645	0.645	0.655	0.330
0.24	0.040	0.040	0.050	0.385	0.710	0.710	0.720	0.295
0.28	0.040	0.040	0.065	0.505	0.765	0.765	0.770	0.275

Table 3.2. Simulation results for $n = 100$, $corr = 0.2$ and $jump = 2$ at midpoint for four different trend functions.

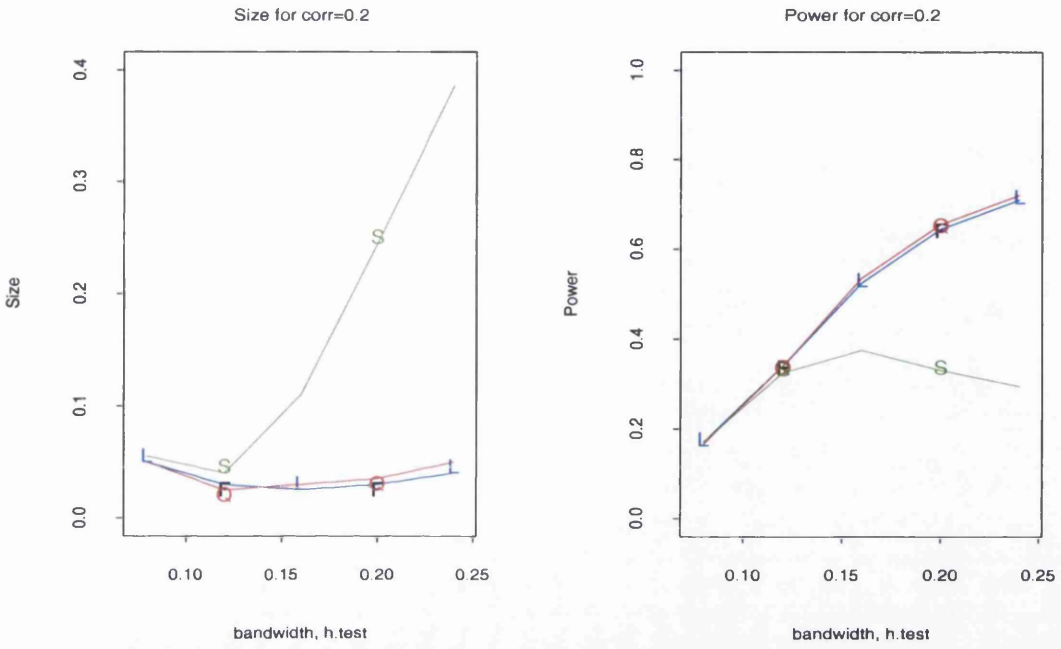

Figure 3.4. Simulation results for $n = 100$, $corr = 0.2$, $jump = 2$ at midpoint for four different trend functions. The trend functions are denoted as F: flat; L: linear; Q: quadratic; S: sine. Rice variance estimator is used.

The size and power of the test for the first three functions are quite similar, all having a size of approximately 0.05, over the whole range of *h.test*. The power increases as *h.test* increases. However, the size for the sine function is only within limits for *h.test* of 0.08 and 0.12, and power increases but levels off at higher *h.test*. The power is also much lower than for the former three trends. The highest power (with appropriate size) is 0.325, using $h.test = 0.12$.

The first three functions have very good size and power as they are very simple functions, the linear and quadratic functions being monotonically increasing trends. In the situation where a very large bandwidth is used, it is almost equivalent to fitting two straight lines, which would be most appropriate if the underlying trend is flat or linear, hence the size remains at 0.05. For the quadratic function, the size is about 0.05, at the range of *h.test* considered. It will however start to go out of limits at extremely large bandwidths.

In the case of a sine function, it is not only a non-monotonic trend, it has in fact two turning points. To trace the underlying trend, a moderate bandwidth has to be used. Using too large a bandwidth will result in false alarms, as it forces the left and right smooths to be almost linear (oversmoothing), resulting in a larger (overestimated) sum of the differences between the two smooths. Thus it tends to give a significant p-value even when the data do not have a discontinuity. As expected, the power of the sine function is lower than that of the flat function, simply because the standardised differences between the two smooths at all evaluation points are affected by the sine trend present and are thus lower.

The shape of the three functions plays an influential role in the effectiveness of the discontinuity test. If the underlying trend is a simple function (flat, linear or quadratic), power increases as *h.test* increases while size remains at an appropriate level of 0.05. However if the trend involves functions that are oscillating (eg. sine), a moderate *h.test* value should be used to obtain a suitable size. The power is lower and tends to level off at higher bandwidths. It is of interest to note that by using the same random seed, the variation in the simulations has been reduced.

Since the performances of the test on the flat, linear and quadratic trend functions are quite similar, in the following sections we will focus on the effects of the different factors only with the flat and sine trend functions over a smaller range of smoothing parameters, $h.test = \{0.08, 0.12, 0.16, 0.2\}$.

3.4.2 Effects of treating correlated data as independent

Before we look further into the performance of the test taking into account the known correlation, let us first look at how detrimental it would be to the test if we have *not* taken the correlation of the data into account, when the data are in fact correlated. In this case, Σ is set equal to I in Equations (3.9) and (3.13). Figures 3.5 and 3.6 display the effects of treating data as independent when they are correlated for sample sizes of 50 and 100 respectively, under the flat and sine trends. Correlations ranging from -0.8 to 0.8 are used, and a jump of size 3 is added at the midpoint of the data for the power calculations.

We can observe that the sizes for negatively correlated data are more conservative, staying

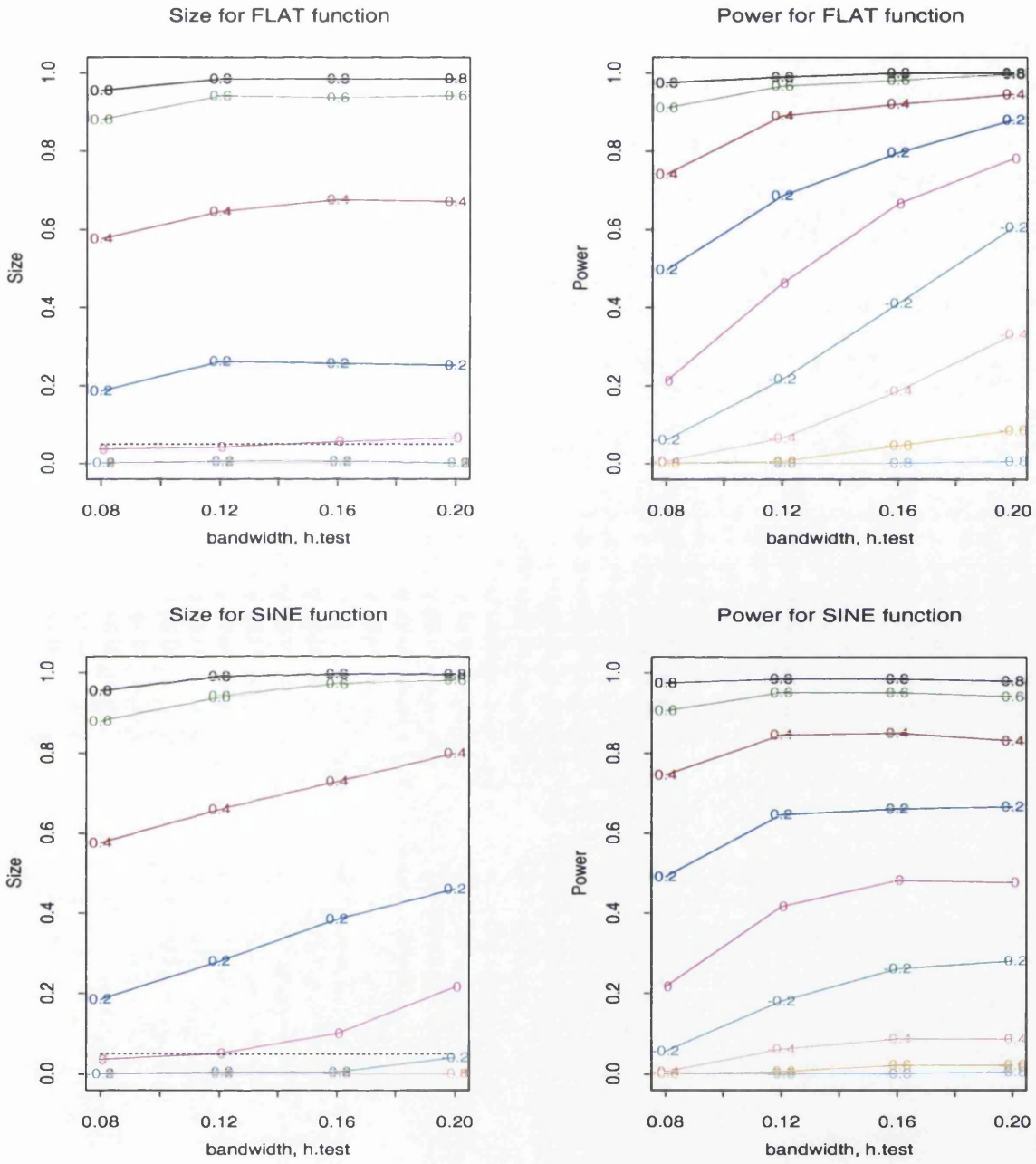

Figure 3.5. Treating correlated data as independent for $n = 50$. The top graphs show the size and power for the flat function, while the bottom graphs show the size and power for the sine function, over a range of correlations from -0.8 to 0.8. The values on the curves denote the correlation of the data.

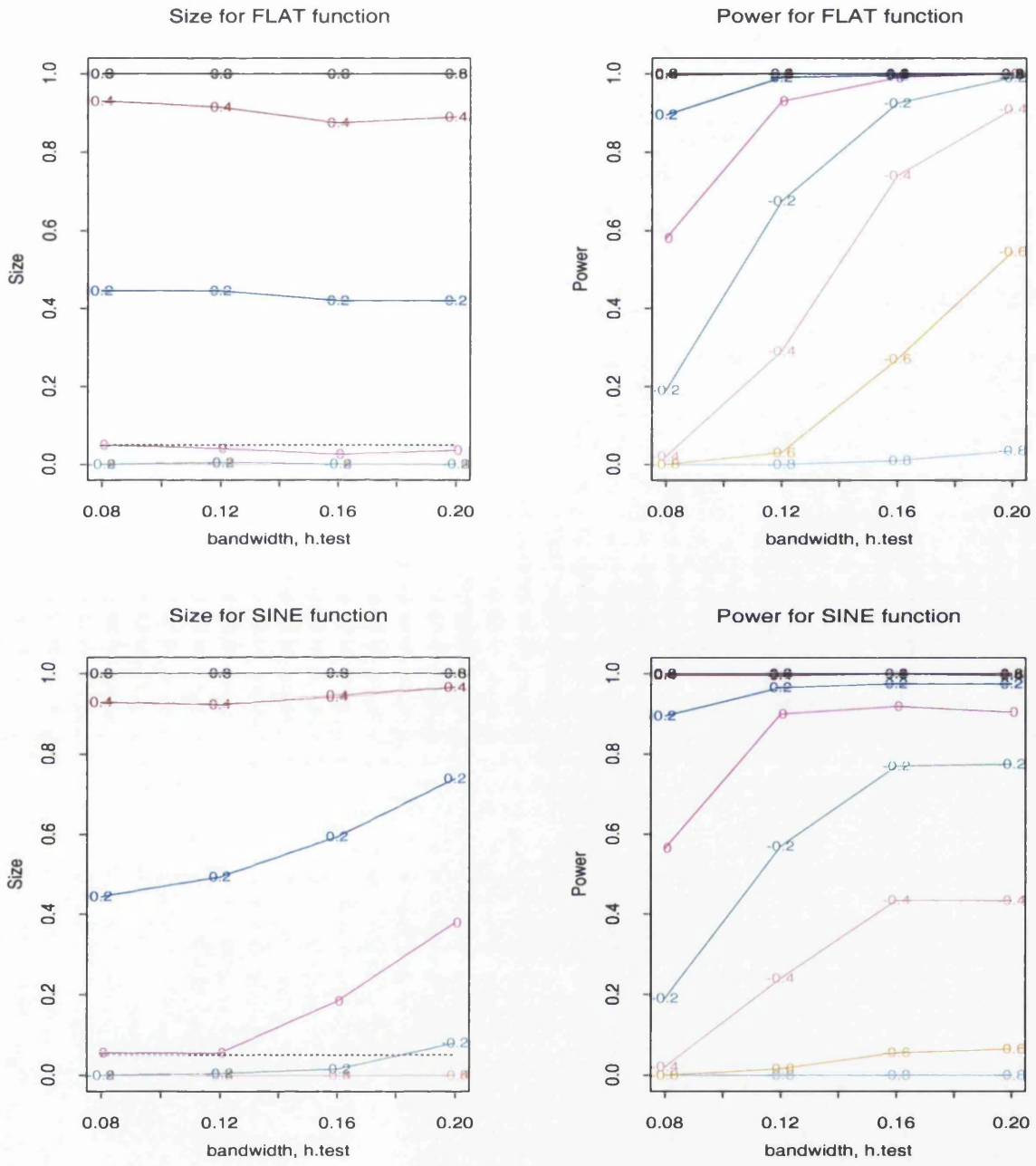

Figure 3.6. Treating correlated data as independent for $n = 100$. The top graphs show the size and power for the flat function, while the bottom graphs show the size and power for the sine function, over a range of correlations from -0.8 to 0.8. The values on the curves denote the correlation coefficients of the AR(1) errors.

almost at zero at the different bandwidths for both sample sizes of 50 and 100. On the other hand, all the sizes for positively correlated data are over 0.05, and they increase as correlation increases. The inflation of sizes for positive correlation is also greater when the sample size is larger and when the bandwidth increases.

In other words, for negative correlation, the test will be conservative and hence less powerful, having reduced size and power compared to the situation where we take the correlation into account. However in the case of positive correlation, the test will detect more false change-points, having higher size and power. (These can be compared to Figures 3.8 and 3.7 for $n = 50$ and 100 respectively in the later section, where the correlation is taken into account). Note that only the test with correlation 0 gives a good size for all smoothing parameters under a flat trend, and for low and moderate bandwidths under a sine trend.

This study thus highlights the importance of taking correlation into account if present, in order to ensure good size properties.

3.4.3 Effects of correlation

In this section, we will consider the same setting as the previous section, but the test has been correctly adjusted for the presence of correlation of an AR(1) model. A sample size of 100 is used.

Looking first at the flat function displayed at the top panel of Figure 3.7, the size of the test is approximately 0.05 for all bandwidths and correlations, except for a correlation of -0.8, which seems to be slightly conservative. There is a general trend that power increases as the correlation gets more negative. In fact, power for zero and negative correlations are approximately 1, for bandwidths of 0.12 and above. On the other hand, as positive correlation increases, power decreases. Hence the corresponding power for negatively correlated data is higher than that of their corresponding positively correlated data (eg. Power for $corr = -0.4 >$ Power for $corr = 0.4$). For data with a high positive correlation of 0.8, power is very low and increases only slightly when the bandwidth is increased.

One possible way of explaining this phenomenon is related to the better estimation of the underlying curve for negatively correlated data. This can be seen from the following

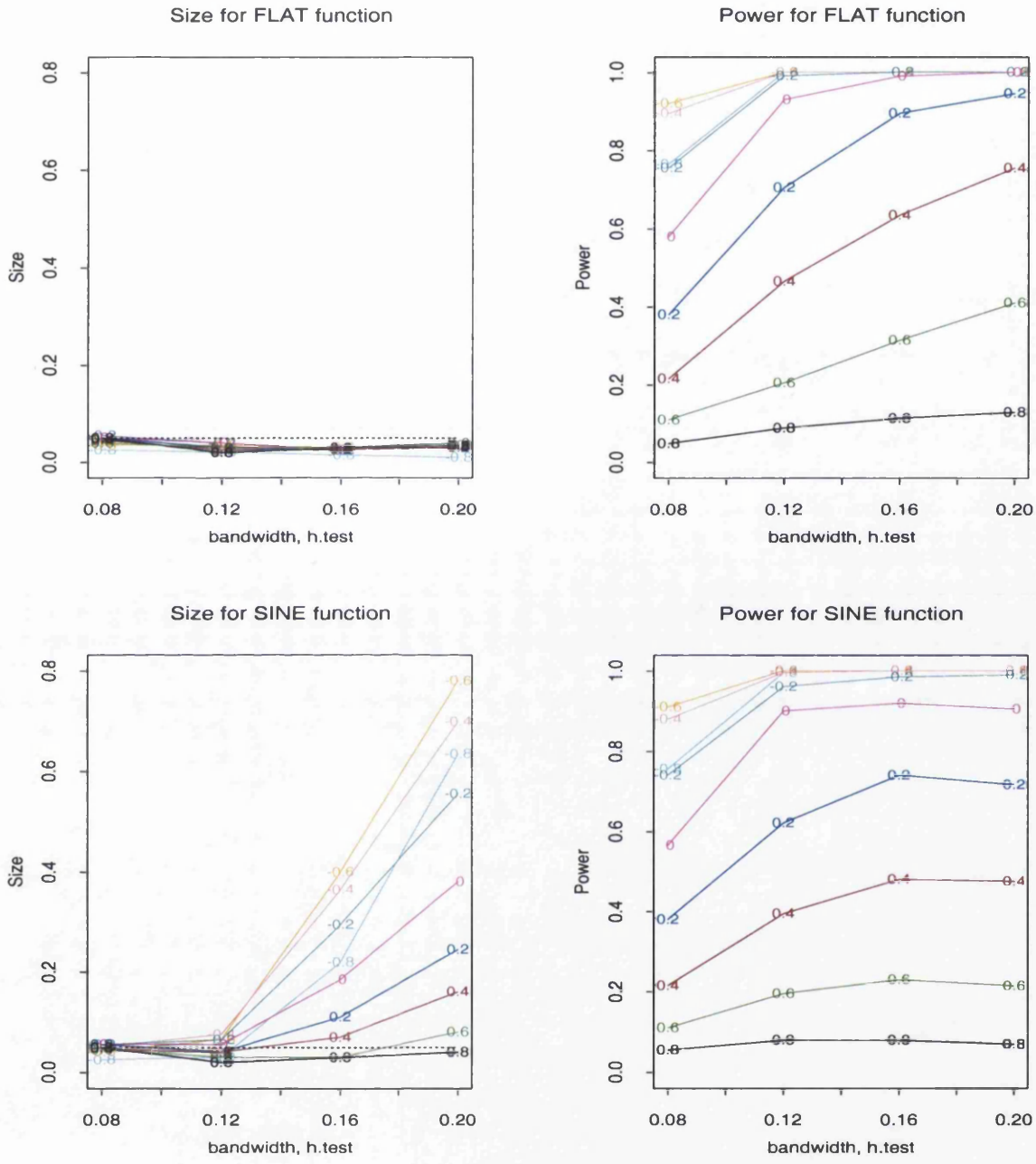

Figure 3.7. Effects of correlation and smoothing parameter for $n = 100$. The top graphs show the size and power for the flat function, while the bottom graphs show the size and power for the sine function, over a range of correlations at different *h.test*. The values on the curves denote the correlation coefficients of the AR(1) errors.

equation,

$$var(\hat{g}(x_i)) = var(\sum_i w_i y_i) = \sum {w_i}^2 \sigma^2 + 2 \sum_{i<j} w_i w_j cov(y_i, y_j) \tag{3.31}$$

In the independent case, the second term is zero, and $\sigma^2 = {\sigma_z}^2$. However, in the case where the correlation is positive, the second term is positive and the first term is increased by a factor of $1/(1-\alpha^2)$, hence the overall variance is increased. Correspondingly, the more positive the correlation is, the more variable the underlying regression will be. This leads to a decrease in power with increasing positive correlation. On the other hand, when the correlation is negative, the second term is negative, but the first term is increased by a factor of $1/(1-\alpha^2)$. Nonetheless, the overall variance is decreased. This leads to a more precise estimate of the underlying regression function and hence the power is higher for negatively correlated data, compared to positively correlated ones.

In the situation where the underlying trend is a sine function, the size of the test is approximately 0.05 for bandwidths below 0.16, but increases as bandwidth increases. The power of the test, however, shows a similar pattern for the effects of correlation and bandwidths as that for a flat function, except that it begins to level off at higher bandwidths. However, care has to be taken when using large bandwidths, as these tend to give false alarms of change-points, which we have mentioned earlier in Section 3.4.1.

3.4.4 Effects of sample size

In Section 3.4.3, we have seen the effects of correlation for a sample of $n = 100$ which are illustrated in Figure 3.7. Here, we are interested to compare the similarities or differences when a smaller sample size of 50 is used. The simulation for $n = 50$ follows exactly the same setting as Section 3.4.3 for $n = 100$, and is presented in Figure 3.8.

Comparing these two figures, the size for the flat trend is approximately 0.05 across the whole range of bandwidths considered. However, a substantial increase in power for all correlations (except 0.8) over different smoothing parameters can be observed as sample size increases. It is as expected that as the sample size increases, power increases, since more

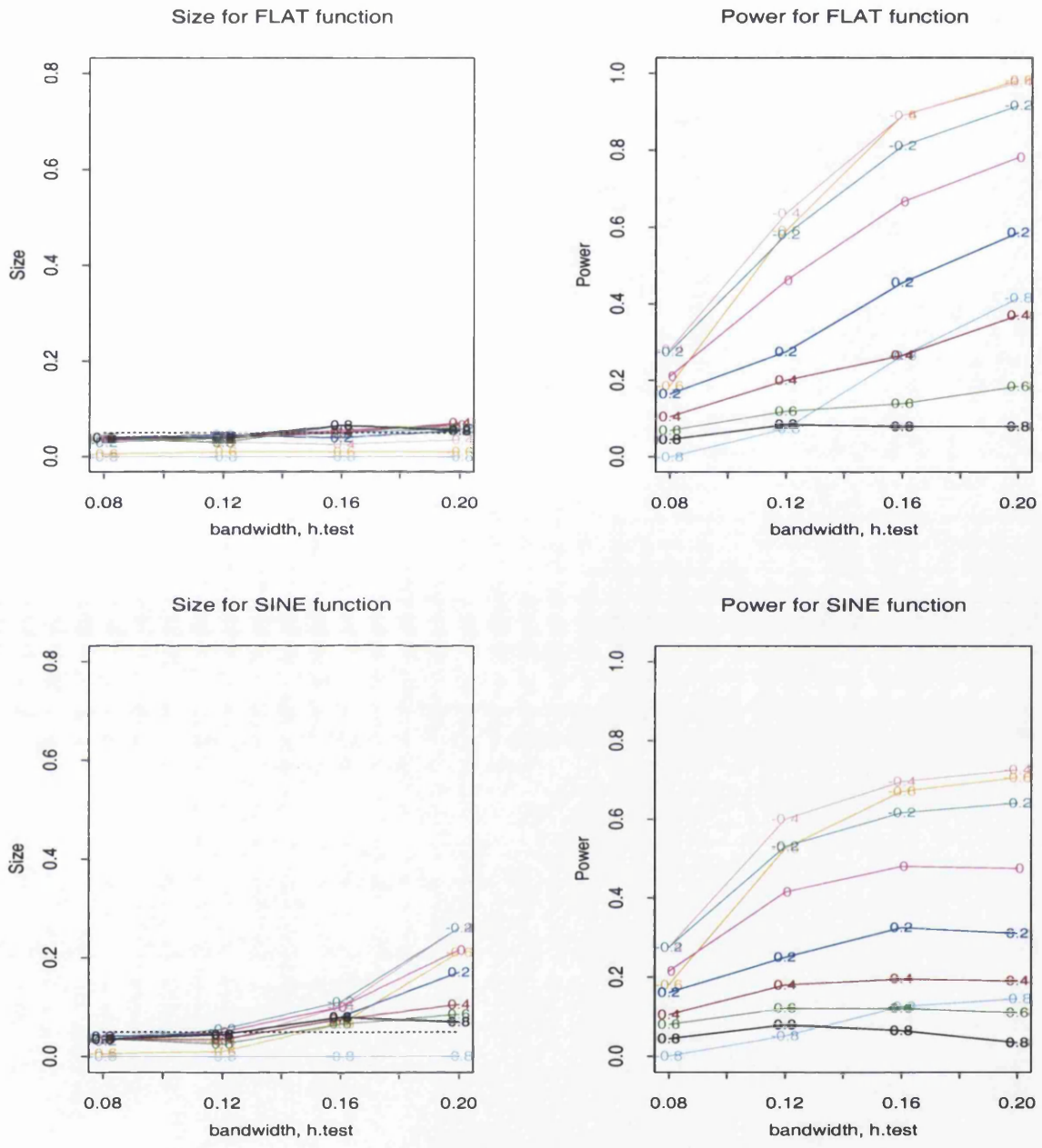

Figure 3.8. Effects of correlation and smoothing parameter for $n = 50$. The top graphs show the size and power for the flat function, while the bottom graphs show the size and power for the sine function, over a range of correlations and *h.test*. The values on the curve denote the correlation coefficients of the AR(1) errors.

data are present and the variability of the estimators decreases.

However, the increase in power, due to a larger sample size, decreases as correlation increases from -0.8 to 0.8. For a highly negative correlation of -0.8, the increase in power is exceptionally high, when n is increased from 50 to 100. On the contrary, only a slight increase is observed for correlation of 0.8, whose power is less than 0.1 even for a larger sample size of 100.

In the presence of a sine trend, the size of the test for both sample sizes is approximately 0.05 for bandwidths of 0.16 and below. However, at higher bandwidths, the size of the test gets more inflated for a larger sample of 100. This indicates that if an inappropriate bandwidth is used for the discontinuity test, there will be more false alarms of change-points for a larger sample than for a smaller one. This is probably because for a fixed bandwidth, the bias is approximately the same but the variance is smaller for the higher sample size. The mis-targetting is therefore more severe for a larger sample, giving a larger size. The power of the test, however, shows a similar pattern to that of the flat function, where it increases as sample size increases.

As negative correlations are less common in practice, the remaining simulations will concentrate on positively correlated data.

3.4.5 Effects of variance estimator

This section focuses on two different variance estimators, proposed by Rice (1984) and Gasser et al. (1986), discussed in Section 3.2.6, and investigates how they might affect the discontinuity test. Two sample sizes $n = 50$ and 100 are considered.

Figure 3.9 displays the effects of the test using Rice and Gasser variance estimators for $n = 50$. In the situation where the underlying trend is flat, the size of test is approximately 0.05. Power is slightly lower for Gasser for all correlations, compared to Rice, except for correlation 0.8, where both are similar. The difference in power in using the two variance estimators decreases when correlation increases.

In the presence of a sine trend, the size for Gasser is slightly lower than that of Rice, but in all cases, the size increases as bandwidth increases. Care has to be taken once again,

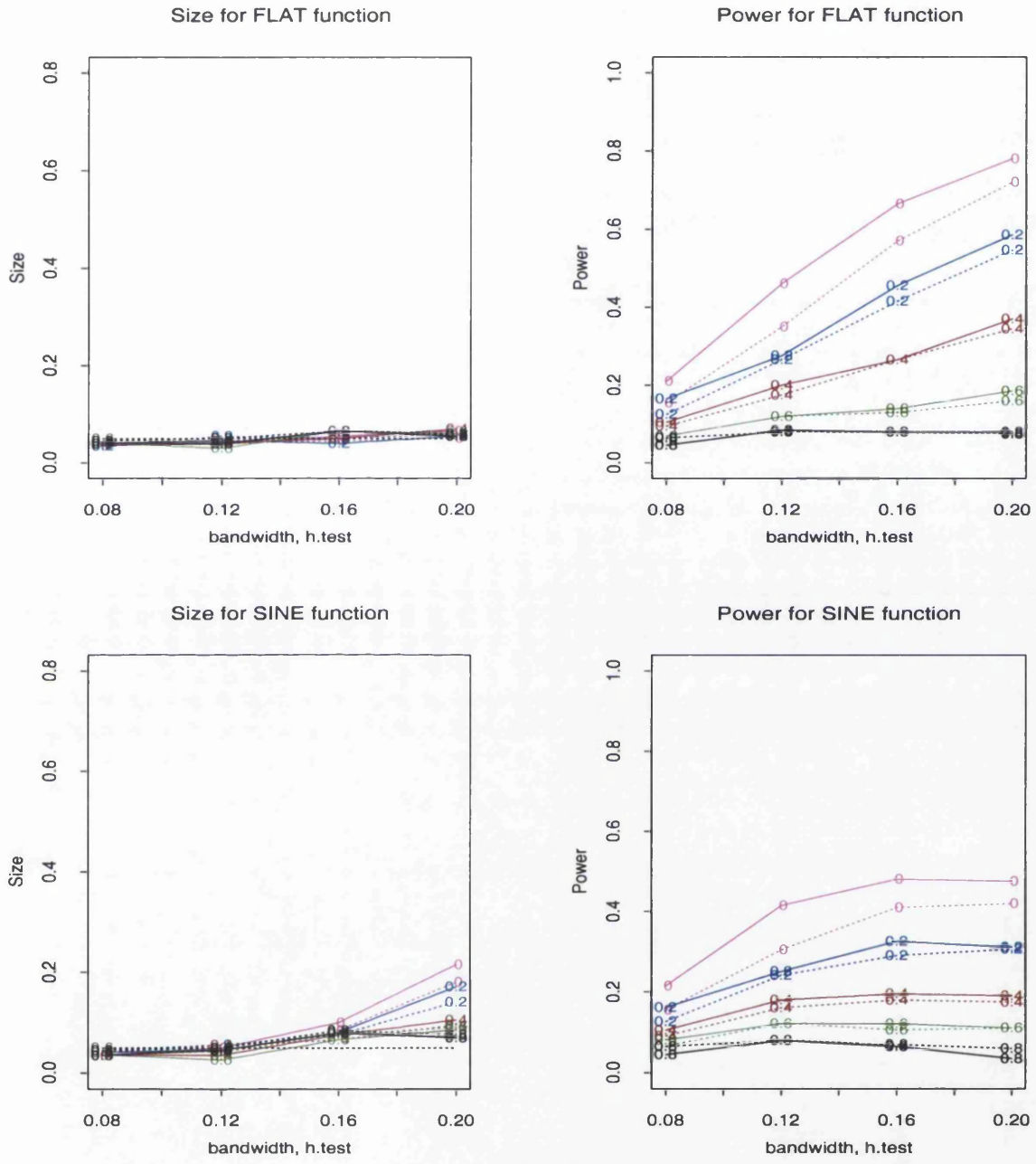

Figure 3.9. Effects of variance estimators for $n = 50$. The top and bottom panels display the results for the flat and sine trends respectively. solid: Rice, dotted: Gasser. The values on the curve denote the correlation coefficients of the AR(1) errors.

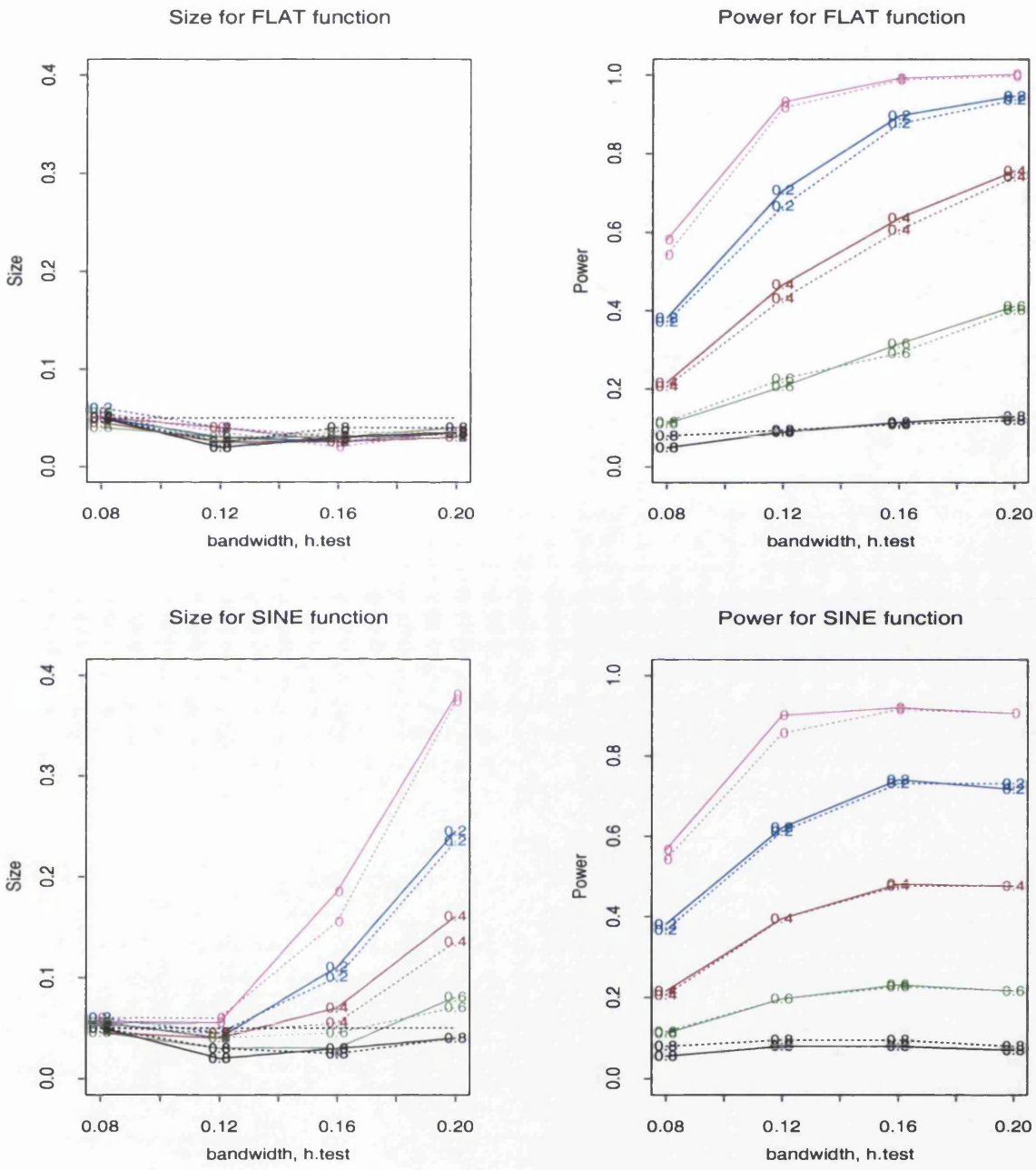

Figure 3.10. Effects of variance estimators for $n = 100$. The top and bottom panels display the results for the flat and sine trends respectively. solid: Rice, dotted: Gasser. The values on the curve denote the correlation coefficients of the AR(1) errors.

in using the appropriate bandwidth. Gasser's method allows for a larger bandwidth to be used, while maintaining a good size. However that does not improve the power compared to using Rice's approach with a slightly lower bandwidth.

Figure 3.10 displays the results for all the various correlation against bandwidth for $n = 100$. Gasser performs slightly worse than Rice for correlations 0 to 0.4, but is quite similar to Rice for higher correlations. However, the difference in power in using these two variance estimators is slightly lesser compared to that for $n = 50$.

From the simulation results, the test is more effective in detecting change-points for low to moderately correlated data under both the trend functions using the Rice variance estimator. This difference is somewhat more prominent for a smaller sample size.

The Rice variance estimator seems to be preferred here as both the flat and sine trends that we have considered are relatively smooth. If the regression function is rougher, Gasser's approach might be preferred (Dette et al., 1998).

3.4.6 Effects of ratio of jump to standard deviation

This section will investigate the effects of jump size on power. As the overall effects of this factor over all correlations are similar, we will focus on two correlations 0.2 and 0.4, which are representative of values commonly met in practice, for both $n = 50$ and 100. Jump sizes of 1,2 and 3 with $\sigma_z^2 = 1$ are used. Their size graphs have been illustrated previously in Figure 3.8 and Figure 3.7. The performance of the test is dependent not just on the jump size, but rather on the ratio of the jump size to standard deviation of the data, sometimes referred to as the signal-to-noise ratio.

As we can observe from Figure 3.11, a larger jump to noise ratio will give a higher power, as it is then easier to detect a jump. This is because a larger ratio increases the value of the test statistic, which is based on the standardised differences between the left and right smooths.

For both the flat and sine functions for sample sizes of 50 and 100, we can observe that at $h.test = 0.14$, as the signal to noise ratio increases from 2 to 3, the power increases by about 0.2 for $n = 50$, and by 0.4 for $n = 100$. This illustrates that the jump size to standard

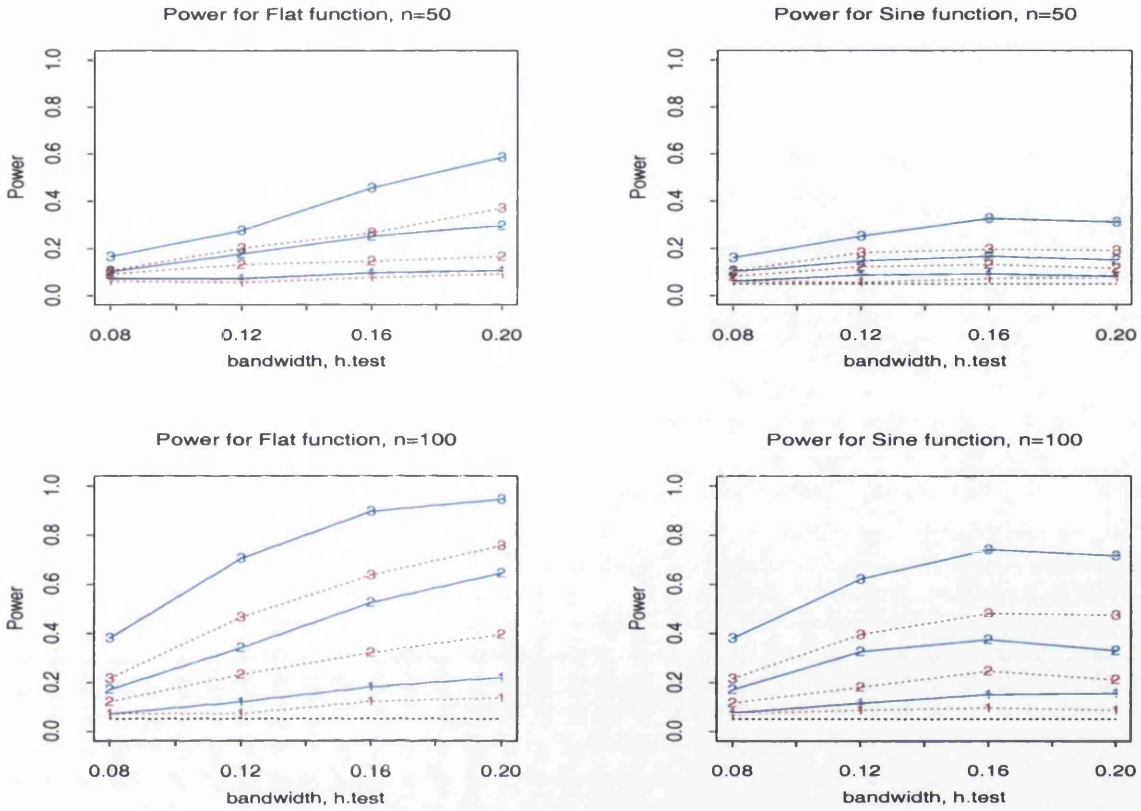

Figure 3.11. Effects of size of jump and error variance. The top panel display graphs for $n = 50$, with both flat and sine trend functions, while the bottom panel presents results for $n = 100$. Blue, solid: $corr = 0.2$, Brown, dotted: $corr = 0.4$. The values on the curve indicate the jump size.

deviation ratio does have a great influence on how effective the test is in detecting change-points and the effects of this increased ratio are greater for a larger sample size, using the same bandwidth.

3.4.7 Local test

In this section, the size and power of the local test is investigated. This is used when it is of interest to examine if a particular point in the data is the location where a discontinuity has occured.

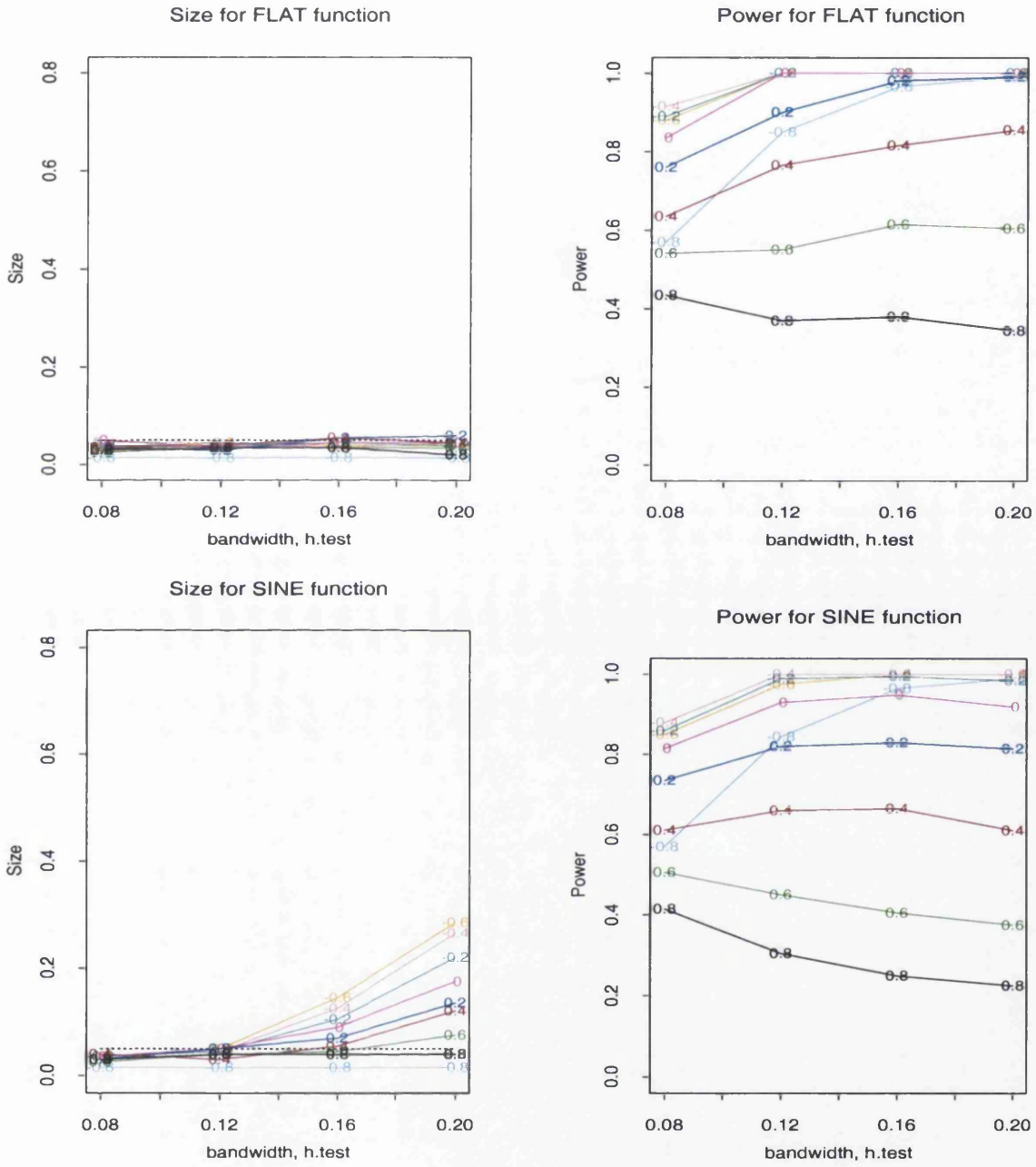

Figure 3.12. Effects of correlation and smoothing parameter using local test with $n = 50$, in the presence of flat and sine trends. The values on the curves denote both the positive and negative correlation coefficients of the AR(1) errors.

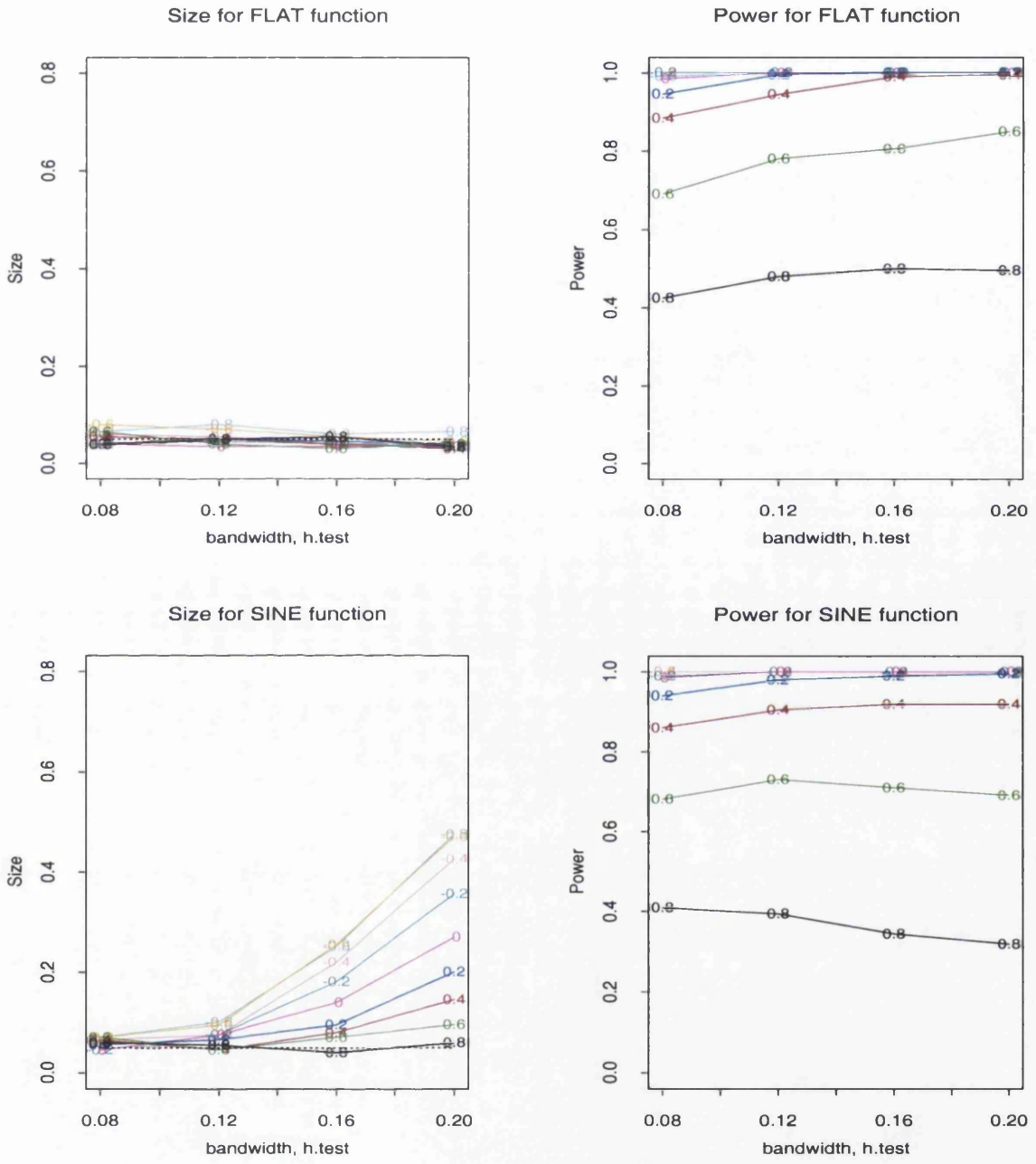

Figure 3.13. Effects of correlation and smoothing parameter using local test with $n = 100$, in the presence of flat and sine trends. The values on the curves denote both the positive and negative correlation coefficients of the AR(1) errors.

Figures 3.12 and 3.13 display the size and power for the local test for a sample size of 50 and 100 respectively, with the correlations assumed to be given. We can observe similar features, in terms of the effects of smoothing parameter, correlation and sample sizes as in the global test. However, one distinct difference observed by comparing Figures 3.8 and 3.12 for $n = 50$ and Figures 3.7 and 3.13 for $n = 100$, is that the the power of the local test is much higher than that of the global test. For instance, power for $corr = 0.4$, with $n = 100$, $h.test = 0.12$, and a sine trend, is approximately 0.4 for the global test, but 0.9 for the local test.

The result is as expected since the test statistic for the local test only involves the squared standardised differences between the left and right smooths at one location of interest, compared to the global test which involves looking at many evaluation points. Hence the latter adds more noise to the test statistic at evaluation points where no discontinuity has occurred, resulting in a decrease in the power to detect a discontinuity.

A high correlation of 0.8 which has poor power (less than 0.1) for both sample sizes in the global test, shows a substantial improvement in the local test. This implies that for highly correlated data, it will be very helpful to have prior knowledge of where possible change-points might have occurred. Another interesting feature is that when the sample size is increased from 50 to 100, the power for $corr = 0.8$ does not seem to alter much, compared to other cases with lower correlations where the power increases as expected.

3.4.8 Remarks

Of interest, it can be highlighted that, when comparing the simulation studies using a flat trend for independent data, it is observed that the proposed asymptotic tests by Müller and Stadmüller (1999) and Wu and Chu (1993) produced inflated sizes, while the discontinuity test developed here remains within limits for $corr = 0$ (refer to Figures 3.8 and 3.7). This could possibly be due to the fact that the former two tests are sensitive to the estimated error variance, but not the latter.

3.4.9 Conclusions

In summary, these simulations have shown the following features when correlation is assumed known.

1. For a flat trend, size is approximately 0.05 for all bandwidths, and power increases as bandwidth increases, and eventually levels off.

2. In the presence of a sine trend, care has to be taken in the bandwidth choice. Although power increases as bandwidth increases (and levels off eventually), oversmoothing will result in an inflation of the size of test. This inflation of size is greater if the sample size is larger.

3. The power is higher if the underlying trend is smoother. Hence the power is always higher for a flat or linear function compared to a sine function.

4. In the situation where correlated data are treated as independent, the discontinuity test will be more conservative for negative correlations, having lower size and power. It will give more false alarms (higher size) for positively correlated data.

5. It is much easier to detect change-points for negatively correlated data than for positively correlated data. Power increases as correlation increases negatively (except for $corr = -0.8$), but decreases as correlation increases positively. It is noteworthy that as the simulations have been carried out keeping ${\sigma_z}^2$ constant, the presence of correlation will increase the variance of the noise, which in turn affects the signal to noise ratio and hence the performance of the test. Thus, the sole effect of correlation on the discontinuity test is confounded with the effect of the signal to noise ratio.

6. The power of the test increases as sample size increases from $n = 50$ to $n = 100$. This difference decreases as correlation increases from -0.8 to 0.8. Hence, for highly correlated data, the increase in power is negligible.

7. As the ratio of jump size to standard deviation increases, the power of the test increases.

8. Gasser's variance estimator performs worse than Rice for the flat and sine trends, but the difference is negligible if the sample size is large.

9. The power of the local test is higher than that of the global test. There is a substantial increase in power even for highly correlated data, unlike for the global test.

3.5 Equally-spaced Setting: correlation unknown

In the previous part of the simulation study, we have assumed that the correlation of the data is known. However, in practice, this is seldom the case, and the correlation has to be estimated.

A number of problems occur when nonparametric regression is attempted in the presence of correlated errors. Indeed, in the most general setting, where no parametric shape is assumed either for the mean or the correlation function, the model is essentially unidentifiable, so that it is theoretically impossible to estimate either function separately. However, the bandwidth parameter (denoted as *h.test* here) makes an implicit assumption on the smoothness of the underlying trend. Any remaining smooth variation is attributed to the correlation.

The rest of the simulation study focuses on the performance of the test when the correlation is estimated using various procedures that were discussed earlier in Section 3.2.7, namely the *residual approach* and the *moving window approach.* Data simulated with correlations of 0.2 and 0.4 are used as they represent low and moderate values that are more commonly met in practice. Sample size of 100 and jump sizes of 2 and 3 at the midpoint of the data are used, in the presence of flat and sine underlying trends. We have used Rice's variance estimator which performs better than Gasser for these two trend functions. As we have seen earlier in our simulations, the size is greatly inflated when a large bandwidth is used in the presence of a more irregular trend, like the sine function. Hence to accommodate for this, we have narrowed down our range of *h.test* values to 0.08 to 0.16.

We will first examine how the test performs using the residual approach in Section 3.5.1 and then the moving window approach in Section 3.5.2.

3.5.1 The Residual Approach

As mentioned in Section 3.2.7, this approach to estimating Σ involves the choice of *h.trend* used to remove the trend, and possibly any jump, in the data.

The simulation study was carried out using different *h.trend* values, to illustrate how important this factor is in the performance of the test. Different *h.test* values are also used. It was suggested in Section 3.2.7 to either use the same *h.trend* value as the one for *h.test*, or a slightly smaller *h.trend* value for trend removal compared to *h.test*. The graphs are plotted over different *h.test* for each *h.trend* used to investigate how this suggestion fares in the simulations. The true size and power curves (denoted as "true", "true.j2" and "true.j3" for jumps 0, 2 and 3 respectively) are also plotted.

Figure 3.14 presents the size and power results, with jumps of 2 and 3, when estimating a correlation of 0.2. From the top left panel, which displays the size for a flat trend, it can be observed that the inflation of size occurs when too low a *h.trend* value is used to remove the trend. In order to obtain appropriate sizes, suitable *h.trend* values of 0.1 and 0.12 can be used over all *h.test* values or $h.trend = 0.08$ coupled with lower *h.test* values of 0.08 or 0.1.

The top right panel illustrates the corresponding power for jumps of 2 (dotted lines) and 3 (solid lines) using different *h.trend* values. As expected, power increases as *h.test* increases since the trend function is flat. The power of the test with $h.trend = 0.1$ or 0.12 (which control the size) is generally lower than the true power for $jump = 3$. However, its performance is much better for a jump of 2. The best *h.trend* value is 0.1 as it traces the true power quite closely for $h.test > 0.1$. Hence, for an appropriate size and good power, using $h.test = h.trend + 2/n$ as suggested in Section 3.2.7, i.e. $h.trend = 0.1$ with $h.test = 0.12$, works well. Next, we will investigate if this is also true for the sine trend.

The bottom panel of the figure presents the results for the sine trend. The same suitable *h.trend* values of 0.1 and 0.12 as before, will give sizes that are within limits with moderate *h.test*. Similarly, the size is inflated when too low *h.trend* values are used.

The corresponding power for the sine trend is lower than that of the flat function as seen earlier. The power using $h.trend = 0.1$ traces the true power quite closely and would

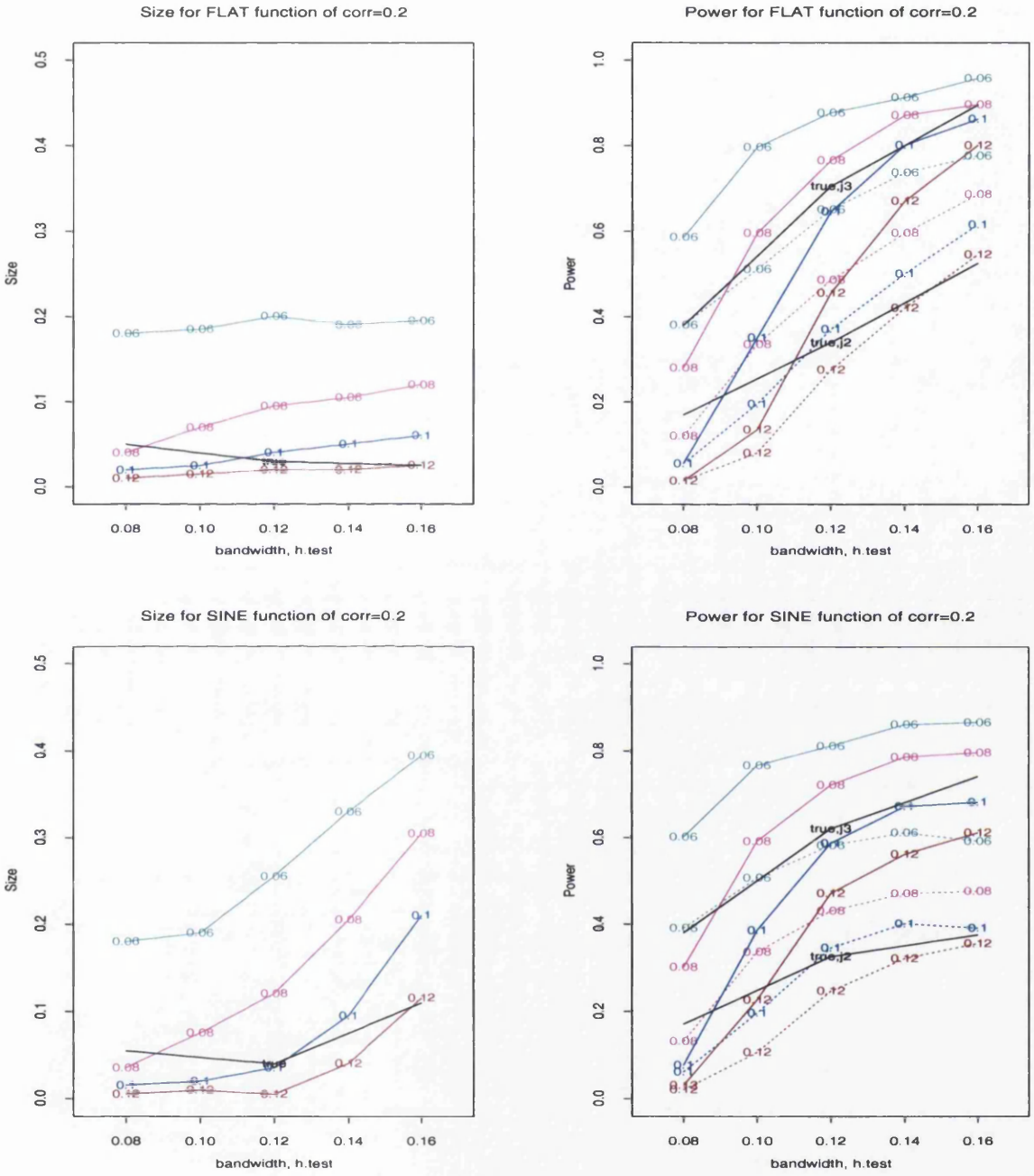

Figure 3.14. Size and Power for $corr = 0.2$ using the residual approach with $jump = 2$ and 3 (represented by dotted and solid lines respectively). The values on the curves denote the bandwidths, *h.trend*. The size and power using true Σ are represented by black solid lines.

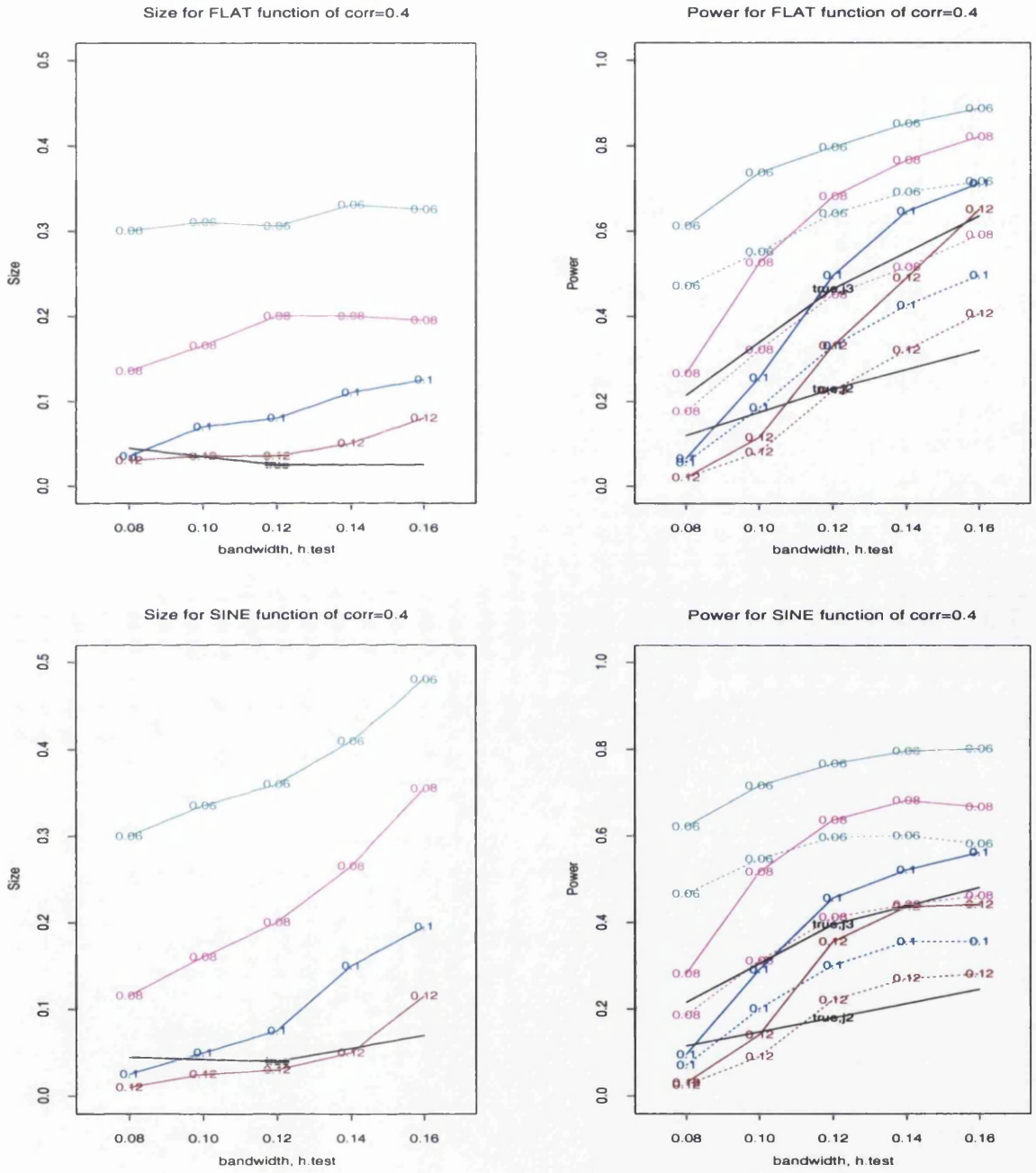

Figure 3.15. Size and Power for $corr = 0.4$ using the residual approach with $jump = 2$ and 3 (represented by dotted and solid lines respectively). The values on the curves denote the bandwidths, *h.trend*. The size and power using true Σ are represented by black solid lines.

be the recommended value in terms of both size and power considerations. The sizes for both are within limits, but using $h.trend = 0.1$ with $h.test = 0.12$ gives a better power than $h.test = h.trend = 0.12$ for both jumps.

Figure 3.15 presents the results for $corr = 0.4$. A slightly larger $h.trend$ than the ones for $corr = 0.2$ is required to give good size values. From the top left plot, a suitable $h.trend$ value is 0.12 over all $h.test$ values, or $h.trend = 0.1$ coupled with $h.test = 0.08$ to 0.12. In fact, the power using those suggested $h.trend$ values and $h.test = h.trend + 2/n$ is similar to that of the true power at that particular $h.test$ value, for both trends. It also performs better when the magnitude of the jump is small. For instance, the power curves for $corr = 0.4$, using $h.trend = 0.12$ when $jump = 3$, generally falls below the true power curve, but is comparable to the true power curve for $jump = 2$, for both trend functions.

3.5.2 The Moving Window Approach

In this section, we will investigate how the test performs when the moving window approach is applied to estimate correlation. The description of this approach was given in Section 3.2.7. Instead of the choice of $h.trend$ in the residual approach in the previous section, this technique requires the choice of the size of the moving window, b. Figure 3.16 displays the results for $corr = 0.2$ using different moving window sizes $b3$, $b4$, $b5$ (where $b3 = n/3$, $b4 = n/4$, $b5 = n/5$).

In terms of size considerations, the three different moving window sizes used are all suitable for both the flat and sine trends. However, in terms of power, $b5$ gives the best power as the curve traces the true power curve very closely for both jump sizes and trends. It appears to perform slightly better in the flat trend scenario.

As the size of the moving window increases from $b5$ to $b3$, the power decreases as a larger moving window will be less capable of estimating the correlation efficiently in the presence of jumps and/or trends, hence making the test more conservative. From the top right and bottom right, we can see that the power using $b3$ is much lower than the true curve for both jumps.

However, care has to be taken if too small a moving window is used as the size will be

higher. Though the size is within limits for both the flat and sine trends, in this case for $corr = 0.2$, it can be noted that size for $b5$ >size for $b4$ > size for $b3$.

Figure 3.17 displays the results for simulations using data with $corr = 0.4$. For the flat trend, the size is inflated when $b5$ is used but is within limits for $b3$ and $b4$. This indicates that a larger moving window is required to get a good estimate of a more highly correlated data.

In terms of both size and power considerations, $b4$ is the best moving window for a flat trend. Its performance is slightly poorer for $jump = 3$ as it lies beneath the true power curve (denoted as "true.j3") over all *h.test*, but traces the true power curve for jump 2 ("true.j2") closely. This indicates that it works better in the presence of smaller jumps. Using $b3$ gives very low power for all cases.

In the presence of a sine trend, both $b4$ and $b5$ can be used. However to keep the size within limits, a *h.test* value of less than 0.14 should be used for $b5$. The power using $b5$ performs well as it traces the true power curve very closely.

A possible explanation as to why $b5$ can be used for the sine trend, but not in the case of the flat trend, is that due to the presence of a sine trend, the estimated correlation using this size of moving window is increased slightly, hence giving a more favourable size than in the flat trend.

3.5.3 Discussion

In the first part of our simulation study in Section 3.4, the correlation is assumed to be known and the effects of various factors (such as the effects of trend functions, correlation, smoothing parameter, etc) on the discontinuity test are assessed. The summary of these effects are provided in Section 3.4.9.

The second part of the study focused on the performance of the test when the correlation is unknown and has to be estimated in different settings of different magnitudes of jump. We kept the other factors investigated earlier constant, since the effects will be similar. Two techniques, namely the residual approach, in Section 3.5.1, and the moving window approach in Section 3.5.2, are proposed to tackle this task. Both techniques require the choice of one

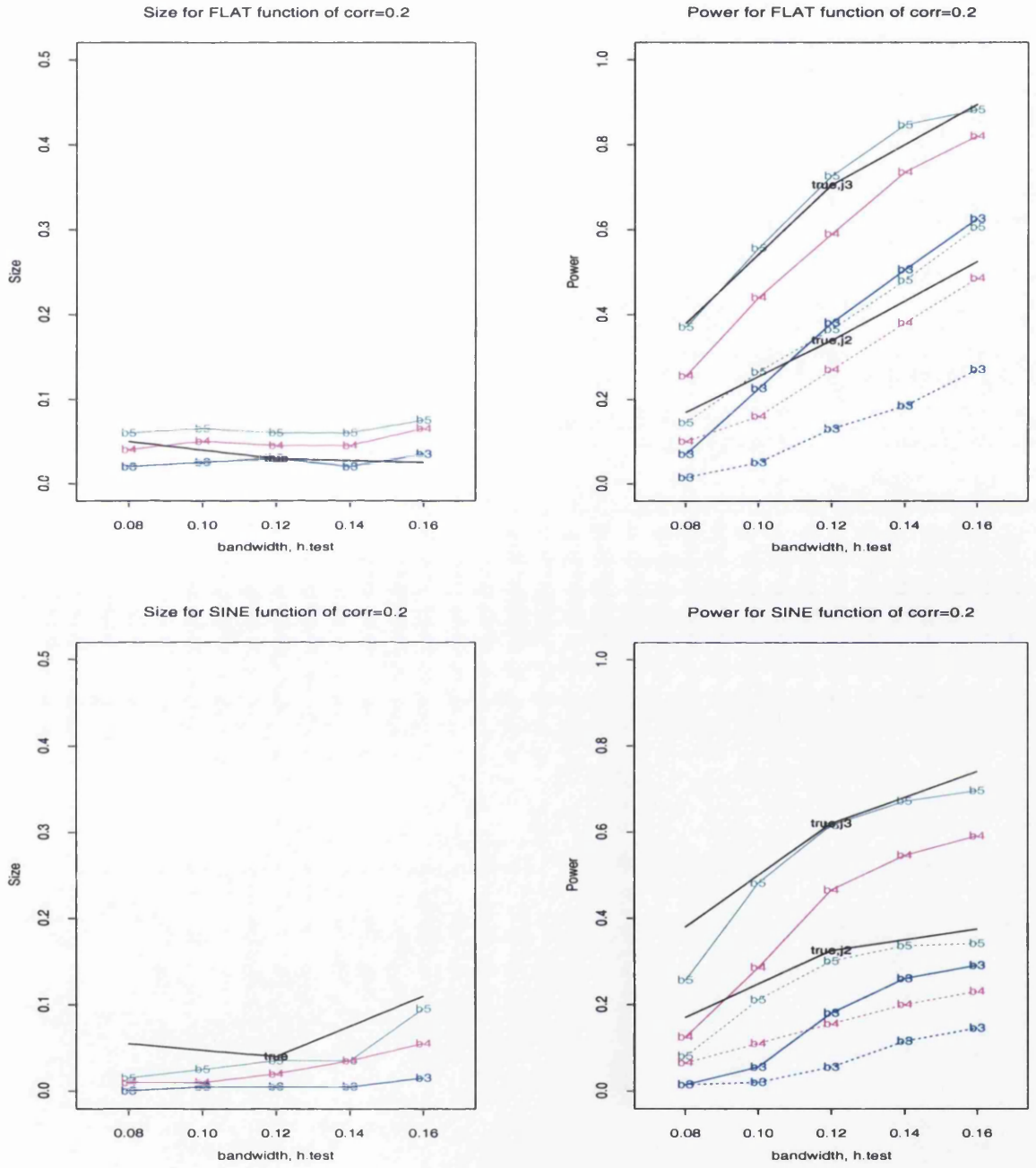

Figure 3.16. Size and Power for $corr = 0.2$ using the moving window approach with $jump = 2$ and 3 (represented by dotted and solid lines respectively). The values on the curves denote the moving window sizes, b. The size and power using true Σ are represented by black solid lines.

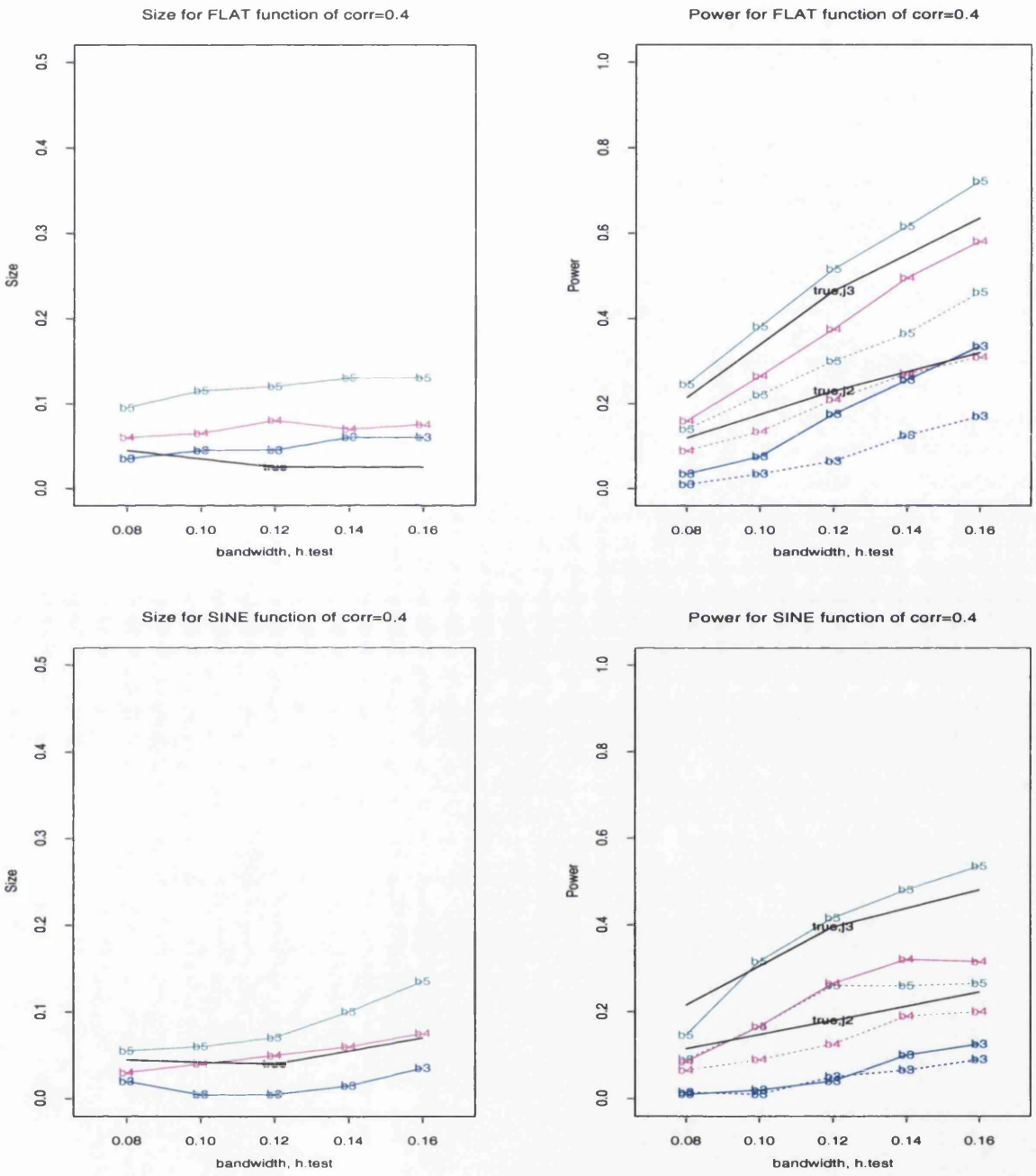

Figure 3.17. Size and Power for $corr = 0.4$ using the moving window approach with $jump = 2$ and 3 (represented by dotted and solid lines respectively). The values on the curves denote the moving window sizes, b. The size and power using true Σ are represented by black solid lines.

parameter; the bandwidth for trend removal, *h.trend*, for the residual approach and the size of the moving window, *b*, for the moving window approach. Both tests perform reasonably well when suitable values are used.

For the residual approach, it is not surprising that the test is very sensitive to the *h.trend* values used to remove the trend. When too small a *h.trend* value is used, the smoother traces the data too closely, hence the correlation is underestimated, resulting in an inflation of size. More highly correlated data will hence require a larger *h.trend* value to give well-controlled size. On the other hand, if too large a *h.trend* value is used, the test will be conservative, giving lower power, due to overestimation of the correlation of the data where the trend and/or the jump might not be efficiently removed. This gets more severe when the jump is large. One way to tackle this is to use a smaller *h.trend* value relative to the *h.test* value, to remove the jump. This suggestion is shown to achieve better performance than using $h.trend = h.test$ in the simulations.

Our simulation results show that the moving window approach is capable of coping with both trend functions and jumps considerably well. For data with low correlations, all three sizes of the moving window can be used, though it is preferable to use the smallest one, $b5$ to give a high power. However, for highly correlated data, a larger moving window $b4$ might be required to give a good estimate. On the other hand, if a trend is present, there might be a need to compromise in the increase of the size of the moving window to cope with this.

As the performance of the test is dependent on how well the two approaches estimate the underlying correlation in the presence of possible trends and jumps, let us now examine the estimated correlations under various settings that have been considered in the simulations.

Figure 3.18 gives a graphical summary of the estimated correlations 0.2 (top panel) and 0.4 (bottom panel) using boxplots for different *h.trend* values, in the presence of both trend functions and jump sizes. Each plot is divided into three segments, where each segment contains boxplots plotted over *h.trend* values from 0.06 to 0.14. The first segment is with jump 0, while the second and third are with jumps 2 and 3 respectively. Similarly, Figure 3.19 presents a graphical summary for both estimated correlations with different sizes of moving windows in various settings.

From both figures, we can see an upward trend in the estimated correlations when *h.trend*

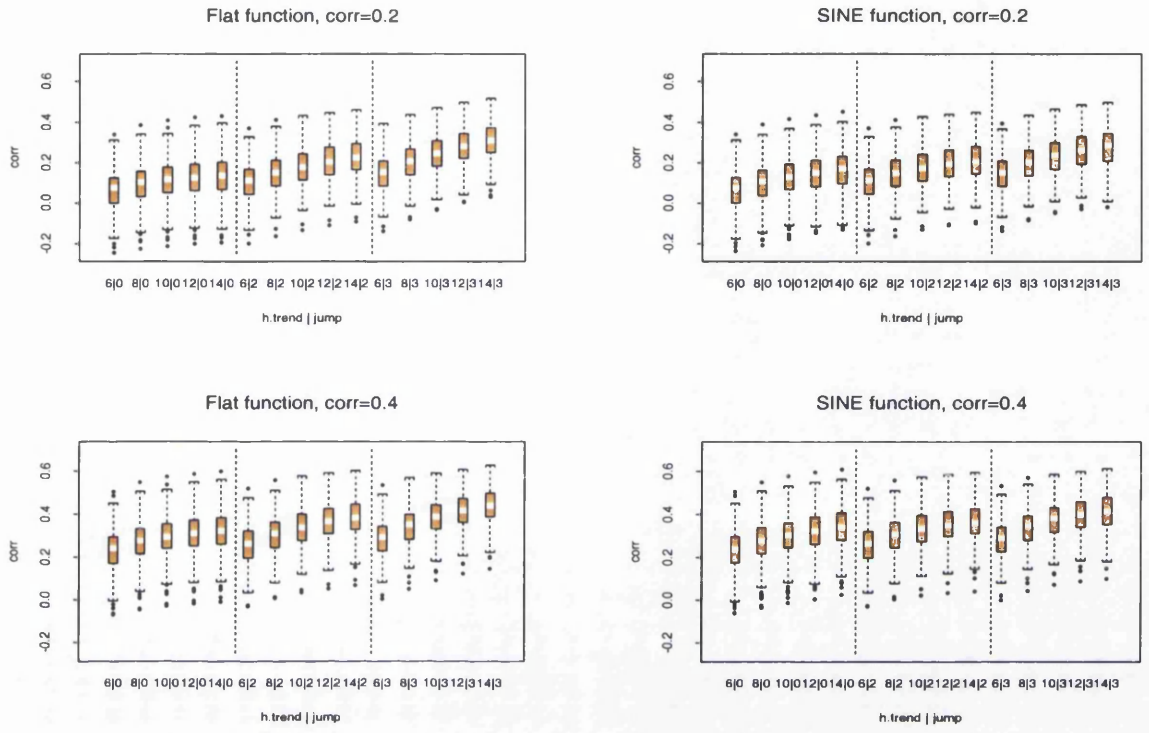

Figure 3.18. Boxplots of estimated correlations of 0.2 and 0.4 using the residual approach for both flat and sine underlying trends. The constants, *h.trend* and *jump* in the symbol "*h.trend*|*jump*" represent $h.trend = h.trend \times 10^{-2}$ with a jump size of *jump*.

and b increase. The medians of the estimated correlations using different *h.trend* and b are generally slightly lower than the true correlations under the null hypothesis when $jump = 0$. The test seems to be able to cope quite well with a slight underestimation of the correlation such as with $h.trend = 0.1, 0.12$ and $b = n/4$, but the size does go outside limits if it is greatly underestimated.

Keeping *h.trend* and b fixed, it can also be observed that the estimated correlation increases as the jump sizes increases. The estimated correlations for the sine function are only marginally higher than those in the flat trend, for the residual approach in comparison to the moving window approach. This implies that the residual approach copes better when a trend is present. To examine this closer, Table 3.3 presents the medians of the estimated

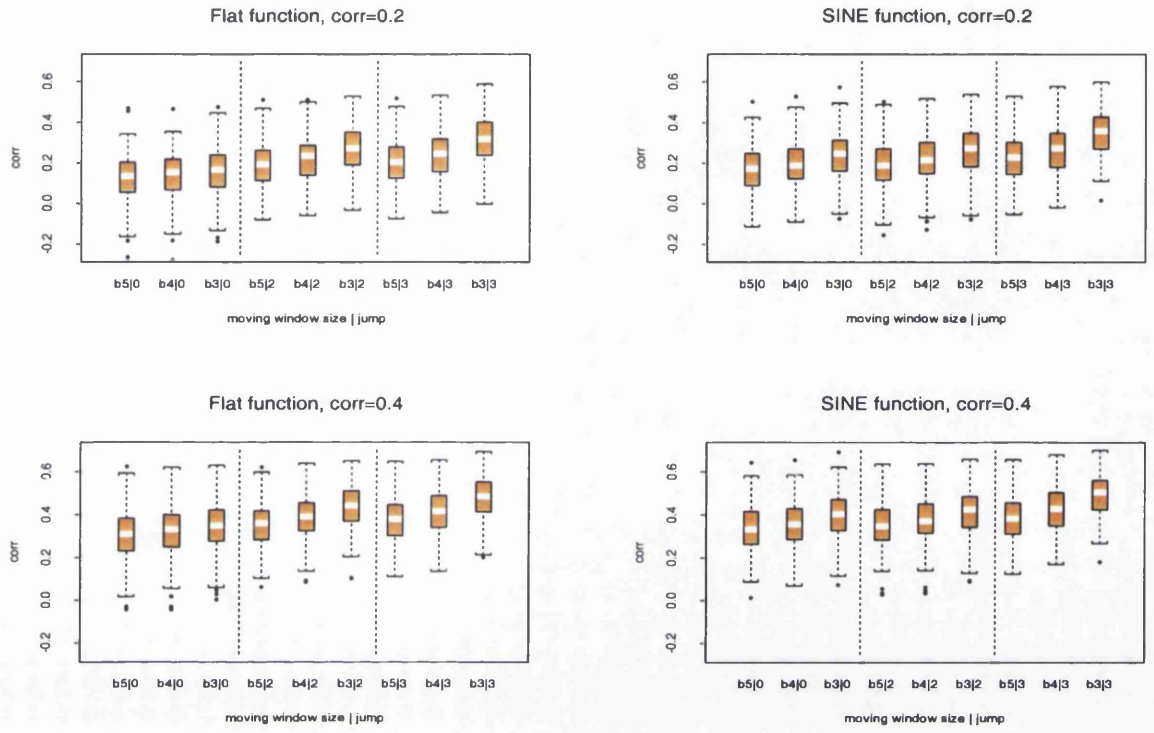

Figure 3.19. Boxplots of estimated correlations of 0.2 and 0.4 using the moving window approach for both flat and sine underlying trends. The constants, b and j in the symbol "$b|j$" represent size of moving window, b with $jump = j$.

correlation 0.2 using the two approaches with their suggested parameters, *h.trend* and b. We can observe that the residual approach can handle trend quite well, since the difference between the estimated correlations under the flat and sine trends is negligible. The moving window, on the other hand, shows a slight difference between that of the flat and sine trend.

However, in the presence of a jump, the residual approach performs worse than the moving window approach. Under both trends, its median correlation rises by a higher margin as the jump size increases compared to the moving window approach. The correlations estimated by the residual approach using $h.trend = 0.1$ are further from the true value of 0.2 and hence are much worse than those estimated by the moving window approach, as it is not an optimum choice. Nonetheless, this table illustrates that the residual approach operates

Trend	Median of estimated correlation 0.2					
	Residual Approach ($h.trend = 0.1$)			Moving Window Approach ($b = n/5$)		
	$jump = 0$	$jump = 2$	$jump = 3$	$jump = 0$	$jump = 2$	$jump = 3$
Flat	0.114	0.177	0.244	0.131	0.191	0.201
Sine	0.127	0.172	0.235	0.168	0.185	0.224

Table 3.3. Median (of 200 simulated samples) of estimated correlation 0.2, using the *residual approach* with $h.trend = 0.1$ and the *moving window approach* with $b = n/5$, for a flat and sine trend.

slightly better in the presence of trend compared to the moving window approach. However, it is less robust to jumps than the moving window approach.

In summary, both the residual approach and the moving window approach, proposed to incorporate the estimation of correlation in the discontinuity test, are sensitive to the parameters of $h.trend$ and b used respectively. The former is shown to perform slightly better in the presence of trend, while the latter copes better in the presence of jumps. Another advantage of the moving window approach over the residual approach, is that it seems more stable across the different $h.test$ values, as the power curves tend to be approximately parallel to the true power curve over different $h.test$ values used. Hence, it is the preferred choice.

3.6 Comparison with the isotonic regression test

After looking at the various properties of the nonparametric regression test in the previous sections, in this section we will examine how it performs compared to an another change-point detection test using isotonic regression (Wu et al., 2001), that has also been devised for detecting change-point with correlated data. This test was briefly mentioned in Chapter 2. It will be discussed in greater detail in Chapter 5, where it will also be used to apply to three real-life data-sets.

A simulation study is carried out to compare the performances of the two tests. Using the data settings as described in Table 3.4, we will present three different sets of simulation. The first, in Figure 3.20, presents a simulation study with flat and sine trends assuming that the parameters required for both tests are known. The second, in Figure 3.21, does not

Data conditions	
Design Space	Equally-Spaced
No. of simulations	200
Trend Function	flat, sine, $-4x^2$, $-2x^2$ trends
Error	$\varepsilon \sim N(0, \sigma^2\Sigma)$ where $\varepsilon_i = \alpha\varepsilon_{i-1} + Z_i$ with $Var(Z_i) = \sigma_z^2 = 1$
Correlation	$\alpha = corr = \{0, 0.4\}$
Test Approaches	
Smoothing Parameter	$h.test = \{0.08, 0.12, 0.16, 0.2\}$ (T1)
Estimator of Error Variance	Rice (T1) and Lag window estimators (T2)
Estimator of Correlation	moving window approach (T1)

Table 3.4. Simulation settings for comparison of two discontinuity tests. T1 and T2 refer to the proposed nonparametric regression test and Wu's test respectively.

make the assumption that the correlation and error variance are known. These unknown parameters, correlation (for T1) and error variance (for T2), are estimated according to the procedures that the tests' creators proposed. The last set of simulations, displayed in Figure 3.22, shows how the two tests perform under two underlying trends, which are not monotonically increasing though the jump is positive. True parameters are assumed here. Correlations of 0 and 0.4 are used for all the above, and are displayed at the top and bottom panels of the figures respectively.

The graphs are plotted over a range of magnitudes of jump to see how well the two tests perform. As our test is very much dependent on the smoothing parameter used for the discontinuity test, *h.test*, a few power curves obtained using different *h.test* are also plotted.

From the top left plot in Figure 3.20, we can observe that the power of the test for $jump = 0$ (i.e. size), under a flat trend, is approximately 0.05 for the two tests considered. The power of the isotonic test for different jumps is higher than our test.

In the presence of a sine trend, Wu's test for power with $jump = 0$ drops to zero, but using moderate smoothing parameters of 0.08 and 0.12, the size of our test remains at approximately 0.05. Wu's test only starts to pick up as the magnitude of jump increases

and performs better than our test when the jump is 1.5 or greater. We can observe a similar pattern for $corr = 0.4$ shown at the bottom panel of the figure for both trends. The decrease in power for our test as correlation increases from 0 to 0.4, is greater than that of Wu's test, which remains reasonably high.

In the previous study, we have plugged-in the true variance for the isotonic regression test. Next, let us investigate how it performs when this parameter has to be estimated. This study is similar to the previous ones, except that all the variances and correlations are estimated. We used the moving window approach with $b = n/4$ to estimate the correlation.

From Figure 3.21, we can observe that the test results using the isotonic regression approach, has size that goes off limits (greater than 0.15) for both $corr = 0$ and 0.4. This is due to the great variability of their estimated variance. In fact, it is underestimated most of the time (its median is lower than the true value.). Our test, however, performs very well for all *h.test* used in the flat trend and for moderate *h.test* values in the sine trend situation, for both correlations using $b = n/4$.

Next, it would be interesting to examine how the isotonic test performs if the underlying trend is of a decreasing function. The results using trend functions $-4x^2$ and $-2x^2$ for both correlations 0 and 0.4 are presented in Figure 3.22. Here, we will plug-in the true variance for Wu's test and the true correlation for our test (similar to the first study).

The isotonic test performs badly for a trend function of $-4x^2$. The power remains at 0 over various jumps, and only starts to increase when $jump > 2.5$. However our test performs very well and has generally higher power than Wu's test over all *h.test* and jumps. With a trend function of $-2x^2$, the power for the isotonic test is lower than the nonparametric regression test for $jump < 2$, but is better at higher jumps. A similar pattern can be seen for $corr = 0.4$.

3.6.1 Comments

In summary, if the variance is correctly estimated and the underlying function is flat, the isotonic regression test is generally better in detecting a change-point than the nonparametric regression test. However, in the presence of a sine trend, the isotonic test performs poorly

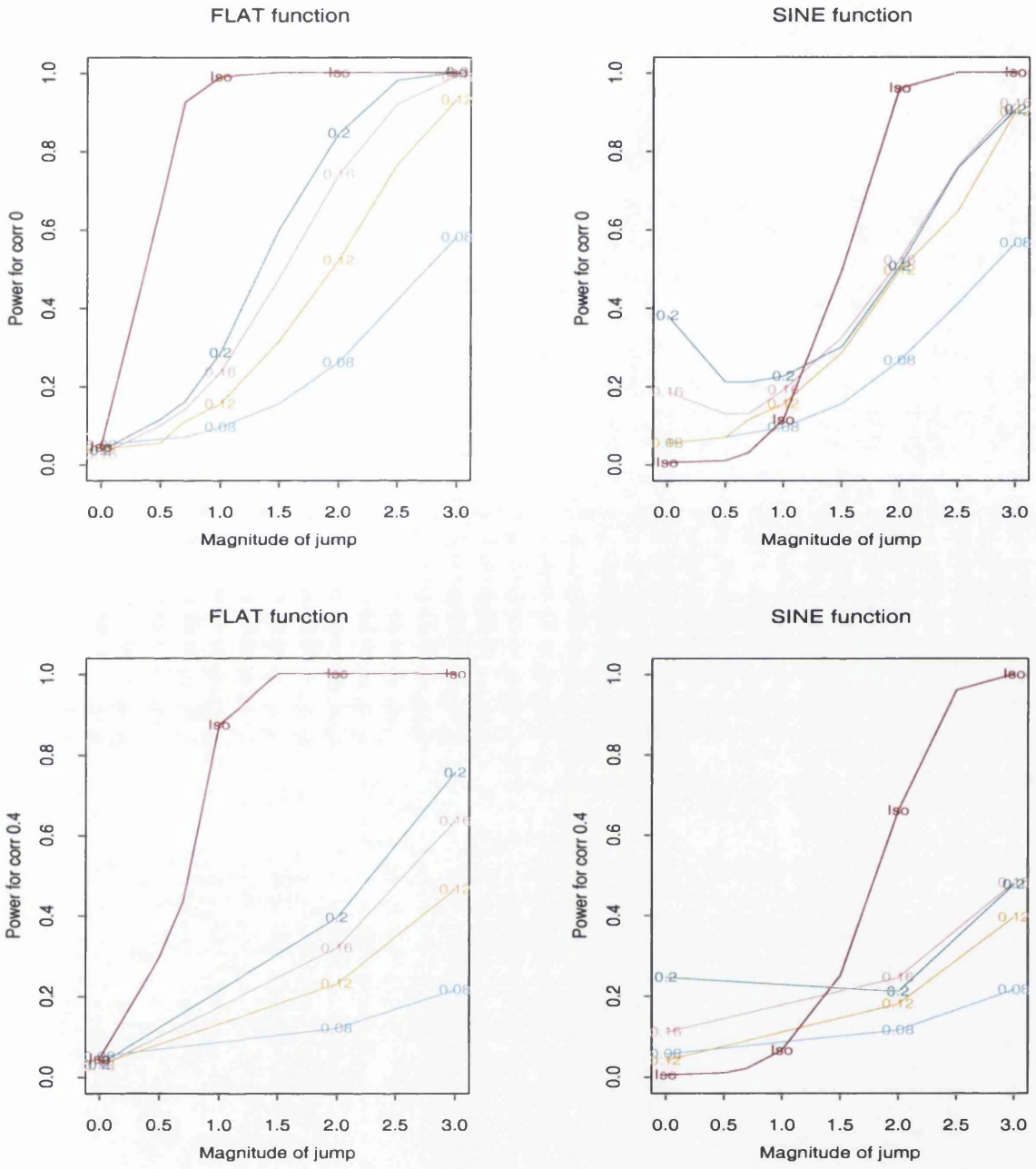

Figure 3.20. Power curves for the isotonic regression test (red, "Iso") and nonparametric regression test with flat and sine trends (with different *h.test* values indicated on the curves). True correlations of 0 (first row) and 0.4 (2nd row) are used.

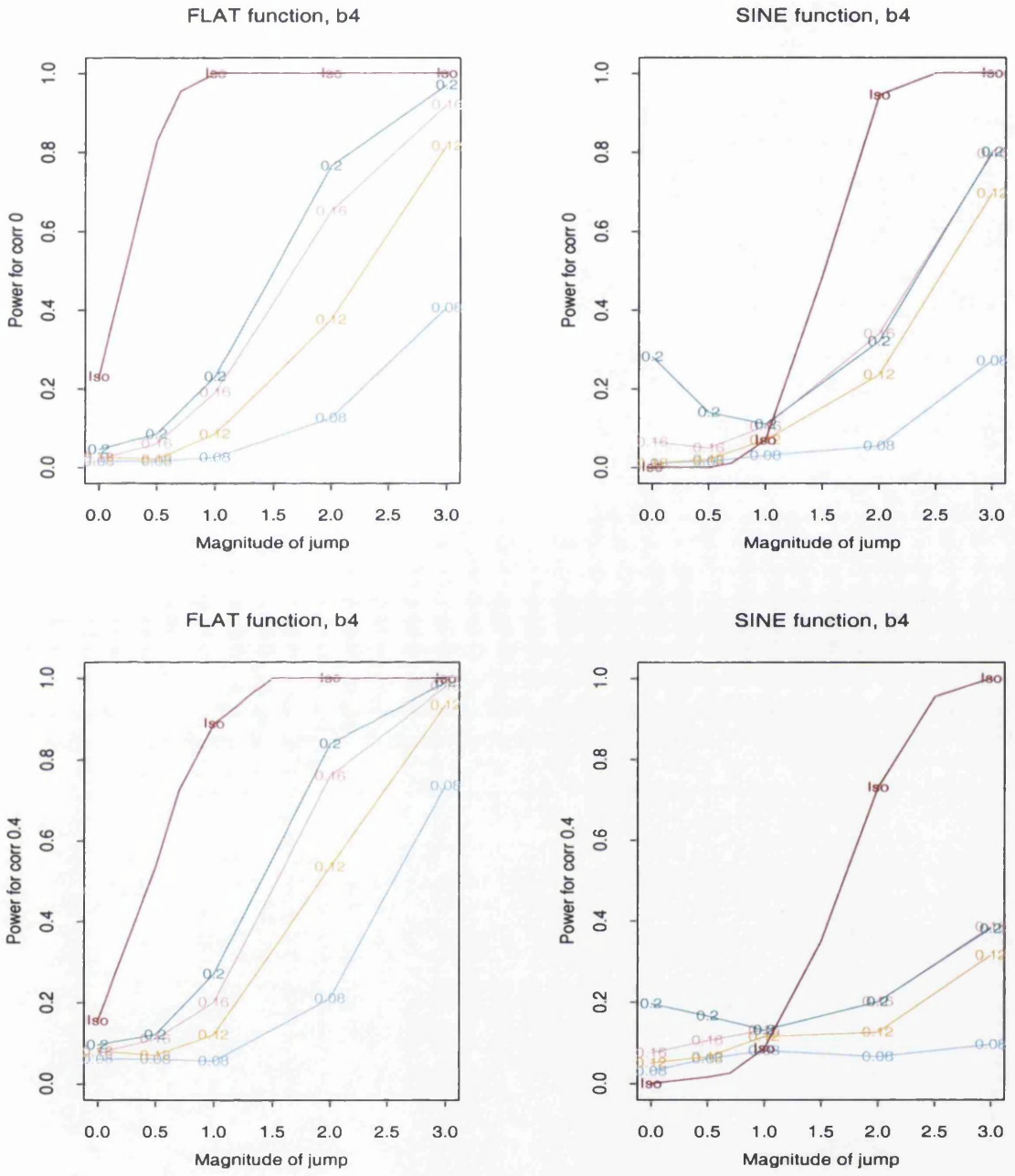

Figure 3.21. Power curves for the isotonic regression test (red, "Iso") and nonparametric regression test with flat and sine trends (with different *h.test* values indicated on the curves). Correlations and error variance are unknown. Correlation of 0 (first row) and 0.4 (2nd row) are used.

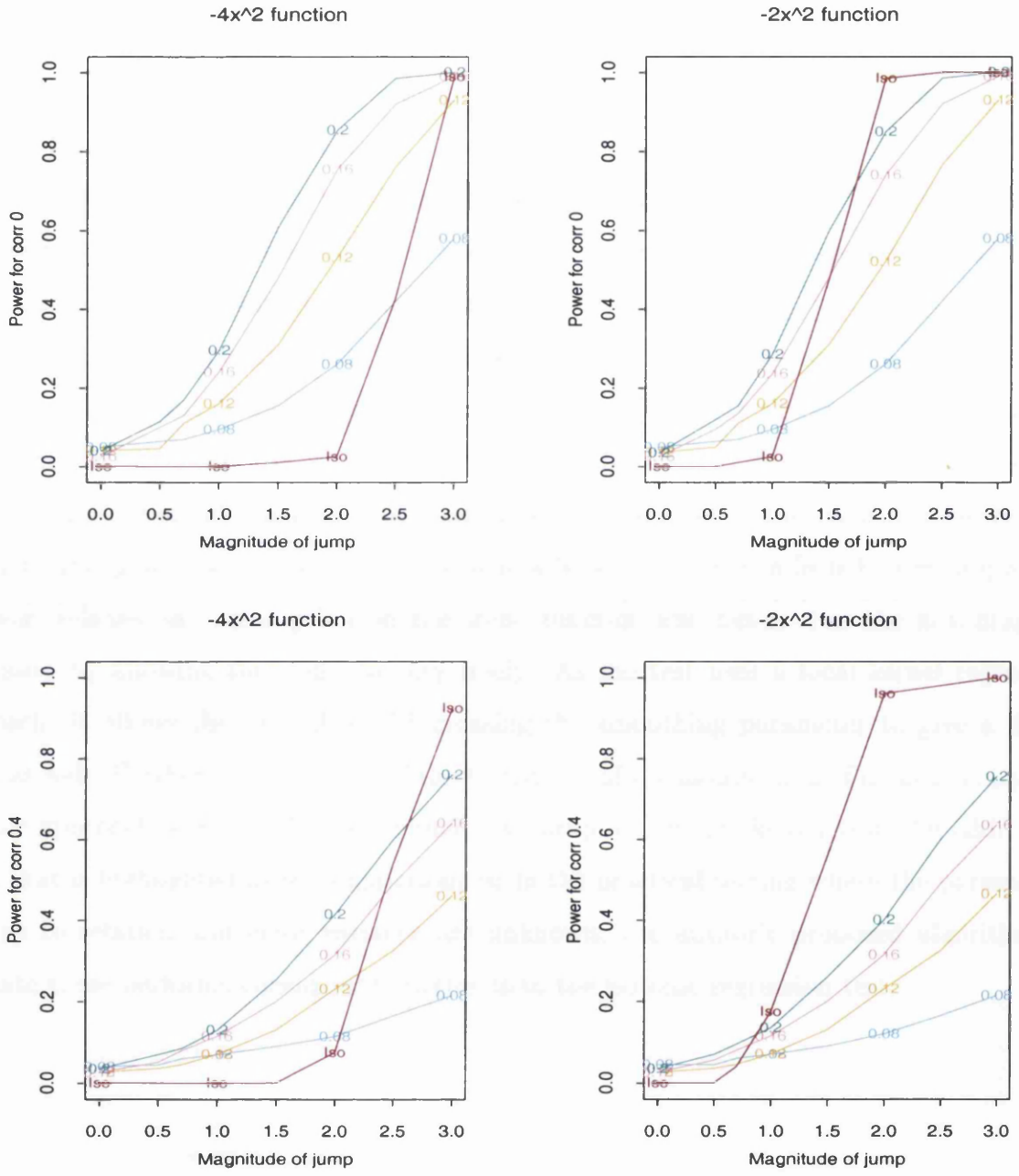

Figure 3.22. Power curves for the isotonic regression test (red, "Iso") and nonparametric regression test with quadratic trends, $-4x^2$ and $-2x^2$ (with different *h.test* values indicated on the curves). True correlations of 0 (first row) and 0.4 (2nd row) are used.

for low jumps, since one required assumption of the test is that the underlying trend is monotonically increasing. This assumption inevitably breaks down in the case of a sine trend. Hence, to be able to detect any discontinuities, it requires any jump present to be more than 1. If the trend is decreasing, such as $-2x^2$ or $-4x^2$, an even larger jump is required for the test to be significant.

Nevertheless, Wu's test is more capable of detecting change-points when the data are positively correlated. The decrease in power as the data become more correlated is reasonably small, compared to a substantial decrease for our test.

If it is valid to assume that the underlying trend and jump if present, are monotonically increasing, Wu's test is more powerful than our test. This is because it takes advantage of the correct assumption of a monotonically increasing trend. However, in many real life applications, the trend function is seldom known and if it is of the form of a decreasing function, the power of the isotonic regression will be very much affected. Our approach, however, relaxes any assumption on the trend function and hence, has the advantage of generality by allowing the trend to vary freely. As the test uses a local kernel regression approach, it allows the flexibility of increasing the smoothing parameter to give a linear trend as well. Furthermore, because of the restriction of the assumption of monotonicity, the isotonic approach is incapable of detecting any abrupt drops, unlike our test. An additional point that is highlighted in the simulations is: in the practical setting where the parameters such as correlation and error variance are unknown, the author's proposed algorithm to estimate these performs considerably better than the isotonic regression test.

Chapter 4

Testing for Discontinuities in One-Dimensional Correlated Data: Unequally Spaced Setting

4.1 Introduction

In the previous chapter, we dealt with data that are equally spaced in the one-dimensional setting. However, in many real-life settings, observations might not be collected over regular intervals. This chapter discusses how our methodology can be extended to detect discontinuities in an unequally spaced setting for one-dimensional data. This is then followed by an extensive simulation study to examine the properties of our test under the influence of various factors. Structurally, this chapter is similar to Chapter 3.

As the observations are irregularly spaced over time or distance, a traditional tool known as the variogram will be used to describe the underlying dependence or correlation structure. Thus, before we proceed further, let us first consider some elementary concepts and definitions in spatial analysis, including the variogram.

4.1.1 Estimation of the Variogram

Let D be a subset of $\Re^d$ (d-dimensional Euclidean space). Suppose $\{Z(s), s \in D\}$ is a collection of random variables indexed by s, whose mean $\mu(s)$ is a constant and hence, the differences $Z(s+h) - Z(s)$ have zero mean and the variance of the differences depends only on h, for any $h \in \Re^d$,

$$\begin{aligned} E(Z(s+h) - Z(s)) &= 0 \\ Var(Z(s+h) - Z(s)) &= 2\gamma((s+h) - s) \\ &= 2\gamma(h) \end{aligned}$$

where $\gamma(h)$ is the semivariogram. A process which satisfies this property is considered to be intrinsically stationary. The function $\gamma(h)$ is known as the semivariogram and $2\gamma(h)$ as the variogram which is a function only of the increment h (Cressie, 1991).

Since $E(Z(s+h) - Z(s)) = 0$,

$$\begin{aligned} Var(Z(s+h) - Z(s)) &= E[Z(s+h) - Z(s)]^2 - [E(Z(s+h) - Z(s))]^2 \\ 2\gamma(h) &= E[Z(s+h) - Z(s)]^2 \end{aligned}$$

Furthermore, if $2\gamma(h) = 2\gamma(|h|)$, i.e. a function only of the length h, then the process is also known as *isotropic*.

Another important property that the variogram must possess is called *conditional negative definiteness* (Cressie, 1991).

$$\sum_{i=1}^{m}\sum_{j=1}^{m} a_i a_j 2\gamma(h) \leq 0$$

where the a_i are real numbers, $\{a_i : i = 1, \ldots, m\}$ such that $\sum_{i=1}^{m} a_i = 0$

A natural estimator of the variogram is to take the average squared differences between points separated by a distance h. This method of moments estimator is often known as the classical variogram estimator (Matheron, 1963).

$$2\hat{\gamma}(h) = \frac{1}{|N(h)|}\sum_{N(h)}(Z(s_i) - Z(s_j))^2 \tag{4.1}$$

where $N(h) = \{(s_i, s_j) : |s_i - s_j| = h,\ i, j = 1, \ldots, n\}$ are the pairs of observed data points separated by a distance of h, $|N(h)|$ is the number of pairs in $N(h)$ and $Z(s_i)$, $Z(s_j)$ are the data values at spatial locations (or times) s_i and s_j respectively. This classical variogram estimator in Equation (4.1) is unbiased for $2\gamma(.)$ when $Z(.)$ is *intrinsically stationary* .

However, one concern about using this method of estimation is that it is not robust against outliers or contamination of the data (Cressie, 1991). We note that if the process $Z(t)$ is assumed to be Gaussian, then $(Z(t_i) - Z(t_j))^2$ will be distributed as $2\gamma(h)\chi^2(1)$, which is highly skewed. Hence, it may be useful to transform it to obtain a more Gaussian-like distribution.

Cressie and Hawkins (1980) observed that if $Y \sim \chi^2$, then taking $Y^{\frac{1}{4}}$ will give a nearly symmetric distribution. This implies that taking sample averages of $|Z(s_i) - Z(s_j)|^{\frac{1}{2}}$ is much more robust against outliers than $|Z(s_i) - Z(s_j)|^2$. Their proposed variogram estimator, which is approximately unbiased, is given as

$$2\overline{\gamma}(h) = \frac{1}{0.457 + 0.494/|N(h)|}\left\{\frac{1}{|N(h)|}\sum_{(s_i,s_j)\in N(h)}|Z(s_i) - Z(s_j)|^{\frac{1}{2}}\right\}^4 \tag{4.2}$$

where the constant $1/(0.457 + 0.494/|N(h)|)$ is a bias correction factor.

Another delightful property of using Equation (4.2), instead of (4.1), is that it has higher efficiency in estimating the variogram. This is because, although the summands in both estimators are not independent, those in Equation (4.2) are less dependent. In other words, if the normal random variables, X_1, X_2 have zero means and unit variances, and have a correlation coefficient $\rho = Corr(X_1, X_2)$, then $Corr({X_1}^2, {X_2}^2) = \rho^2$, but $Corr(|X_1|^{1/2}, |X_2|^{1/2}) < \rho^2$. (Cressie, 1991)

They also proposed another estimator based on the median of the fourth root of the

squared differences.

$$2\tilde{\gamma}(h) = \left[med\left\{|Z(s_i) - Z(s_j)|^{\frac{1}{2}} : (s_i, s_j) \in N(h)\right\}\right]^4 / B(h) \tag{4.3}$$

where the bias correction factor $B(h)$ is 0.4571 asymptotically (Cressie and Hawkins, 1980).

Other possible robust estimators of the variogram are the scale and quantile estimators and the slope estimator. The scale and quantile estimators are shown to be equivalent to Equation (4.3) asymptotically, while properties such as the bias and variance of the slope estimator are not clear (Cressie, 1991). Through simulations, it was demonstrated that the robust estimator, $2\overline{\gamma}(h)$ is more efficient than $2\tilde{\gamma}(h)$ in Equation (4.3) (Taylor, 1987, Cressie and Hawkins, 1980).

For the rest of the thesis, the robust variogram estimator is chosen because of its slightly more superior properties, especially in the presence of outliers, which might occur in the presence of change-points. However, accurate estimation of our correlation structure is not of utmost concern to us here as it is just a means to an inferential end. The main objective of a variogram analysis is to construct a variogram that gives a reasonable estimate of the underlying covariance structure or correlation. Note that when the variance is constant, the relationship between the variogram and the covariance or correlation function is

$$\begin{aligned} \text{Cov}(Z(s), Z(s+h)) &= \sigma^2 - \gamma(h), \\ \text{Corr}(Z(s), Z(s+h)) &= \frac{\sigma^2 - \gamma(h)}{\sigma^2} \end{aligned}$$

There are other factors which have to be considered in obtaining the construction of the empirical variogram. These are:

- lag or bin width of the variogram, e.g. $lag = 0.01$ means pairs which are of a distance 0.01 +/-0.005 apart are grouped in one bin;
- lag tolerance, which is the distance tolerance and in the above example, the value is 0.005. It is often taken as $lag/2$, allowing for unequally spaced data. In the above example where $lag = 0.01$, the lag tolerance is then 0.005;

- number of lags to be computed, *nlag*;
- maximum distance, *maxdist*, where the empirical variogram is computed, normally taken as half the maximum distance of the data as that is the distance of reliability suggested by Journel and Huijbregts (1978). However, if we have highly correlated data, this maximum distance of evaluation might have to be increased, in order to obtain a better estimate of the correlation structure.
- minimum number of pairs allowed is suggested to be $minpairs = 30$ (Journel and Huijbregts, 1978).

4.1.2 Modelling the Empirical Variogram

Figure 4.1 illustrates an idealised empirical variogram. The semivariogram is plotted against the distance between two spatial locations. It increases smoothly as a function of the distance. The basic structural parameters of the variogram are:

- apparent range: representing the distance at which the data are no longer correlated
- nugget variance: representing measurement error or micro-scale variation
- sill: representing the variance of the random field, i.e. the limit of $\gamma(h)$ at $h = range$ or as $h \to \infty$

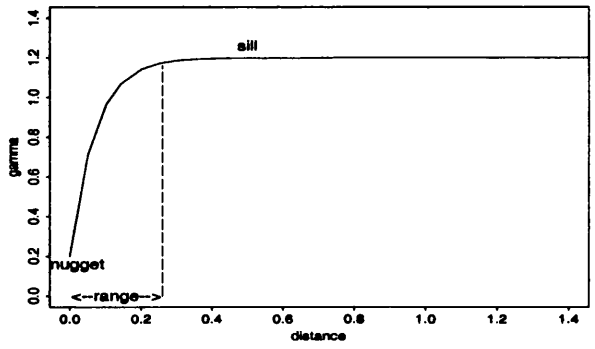

Figure 4.1. An example of a semivariogram showing the sill, range and the nugget variance.

To ensure that the necessary properties of the variogram are satisfied, the empirical variogram can be modelled using theoretical variogram functions. There are various isotropic variogram models proposed by Journel and Huijbregts (1978, pg.161-195) that have the necessary properties. Some of these are the spherical, linear, exponential and gaussian models. In our simulation study, we have considered the exponential model, whose range is approximately one-third the *apparent* range, where the apparent range is the distance at which the data are no longer correlated. The exponential variogram model is expressed as

$$\gamma(h) = c_0 + c_1[1 - \exp(-h/a)] \tag{4.4}$$

where the nugget variance and the sill are denoted as c_0 and c_1 respectively, and the range parameter is a, but the effect of correlation disappears at distance $3a$. In the case where the nugget, c_0 is taken as 0, as in the setting here, the total variance is simply the sill, c_1, i.e. $c_1 = \sigma^2$. One of the main advantages of using the exponential model is that it is a relatively simple positive correlation function which decreases with increasing distance, and should be sufficient to capture most of the dependence present in nearby observations. (For the spherical and gaussian models, the range is the apparent range.)

To fit this theoretical variogram model, various procedures such as *maximum likelihood* (ML), *restricted maximum likelihood* (REML), *ordinary least squares*, *weighted least squares* (WLE) or *generalized least squares* (GLE) can be used. The topic of modelling the variogram is very interesting in its own right and various authors have compared the performances of these procedures (see Cressie, 1991, pg. 99-100 for a set of references). Since this concerns the estimation of the correlation of the data, which is a factor that affects our discontinuity test, it is indeed an important issue. However it is not the principal focus of the work. We have decided to adopt the weighted least squares approach, because on the overall performance scale it fares reasonably well in terms of bias, mean-squared error and coverage probability of the classical 95% prediction interval (Zimmerman and Zimmerman, 1991). Furthermore, it not only does well for the Gaussian model, but its performance is comparable for non-Gaussian models too (Cressie, 1991).

We made use of the various functions provided from the Splus add-on module, Spatial Stats, to construct the empirical variogram. The estimation of the parameters, such as range, sill and nugget, are undoubtedly influenced by the various factors such as *lag*, *nlag* and *maxdist* that we have discussed earlier in Section 4.1.1. In our simulation study, we have set the suggested minimum number of pairs, maximum distance, and lag tolerance as stated above, and *nlag* is set to the sample size, n. Lags or bin widths of 0.01 and 0.02 are considered in the estimation of the variogram.

4.2 Methodology

4.2.1 Statistical Model

Just as in Chapter 3, the basic nonparametric model is considered:

$$y_i = g(x_i) + \varepsilon_i \tag{4.5}$$

where g is the nonparametric regression function, allowing the mean to be smoothly varying, and ε_i are correlated and normally distributed errors $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n) \sim N_n(0, \sigma^2\Sigma)$ for $i = 1, \ldots, n$ and Σ is the correlation matrix. Here we take our covariates, x_i to be unequally spaced fixed design points over 0 to 1. The assumptions about the true underlying regression function are similar to those for the equally spaced design in Section 3.2.1.

4.2.2 The test statistic

Our methodology for the unequally spaced setting is again based on the differences between the left and right smooths as discussed in Section 3.2.2. The global test will first be introduced in Section 4.2.2, followed by the local test in Section 4.4.2.

4.2.2.1 Global test

Here, there is no prior knowledge of the possible change-point locations. Our interest lies in testing whether there are any discontinuities in the smooth regression function, g. The

hypotheses are as listed below.

H_0: g is continuous over the interval (0,1).

H_1: g is discontinuous at least one point in the interval (0,1).

Following the same notation as that for the equally spaced setting, in computational terms, the vector of the differences of the left and right smooths, r can be written as DY, where D is an $n \times n$ matrix, and has its ith row $(W_{L_i} - W_{R_i})$.

Two main forms of test statistic can be considered here. The first form is the same as the test statistic that have been used in the equally spaced setting, $F_3(h)$ in Equation (3.9) from Section 3.2.2, where the error variance is estimated by expressing it in quadratic form, $\frac{Y^T BY}{tr(B)}$. It shall be denoted as the original test statistic, F_O. It is expressed as

$$\begin{aligned} F_O(h) &= \frac{Y^T D^T \Lambda' DY}{\hat{\sigma}^2} \\ &= \frac{Y^T D^T \Lambda' DY}{\frac{Y^T BY}{tr(B)}} \end{aligned} \tag{4.6}$$

where Λ' is the inverse of the diagonal matrix whose diagonal entries are those of $D\Sigma D^T$; Σ is obtained via the variogram approach and $\hat{\sigma}^2$ is expressed in quadratic form and substituted as $\frac{Y^T BY}{tr(B)}$.

As discussed previously in Chapter 3, although $\hat{\sigma}^2$ does not give a good estimation of σ^2 when the data are correlated, this does not invalidate the test. By incorporating the estimation of σ^2 via quadratic forms into the test statistic, a ratio of quadratic forms is created which makes the test statistic independent of the true value of σ^2. This test statistic also computes the variation of the data by taking into account the amount of data lying on the left and right smooths of each evaluation point, by introducing Λ' in its form. Hence it is slightly superior compared to $F_1(h)$ (in Equation 3.6) which is of a similar form but does not include the weighted variances, Λ'. As in the equally spaced setting, we shall focus our attention on the weighted version as expressed in Equation (4.6) because of its more favourable properties.

In the previous section, we introduced the variogram as a tool to model the dependence structure of unequally spaced data. In fact, the modelling of the underlying correlation

structure is one of the main differences between the equally and unequally spaced designs in our test. The former is assumed to follow an AR(1) correlation structure, while the latter is modelled via a theoretical exponential variogram function, discussed earlier in Section 4.1.2.

Since, in fitting the variogram, we not only obtain the range of the data, which describes the dependence structure, but also acquire an estimate of the error variance through the sill of the variogram, we could consider an alternative test approach. This is to regard the variance as a "plug-in" estimator of σ^2 obtained from the sill of the variogram, which we then treat as the true value. Hence, the random nature of the denominator is not incorporated into the distributional calculations with the test statistic. This new form of test statistic is denoted as F_N. It can be expressed explicitly as,

$$F_N(h) = \frac{Y^T D^T \Lambda' D Y}{\hat{\sigma}^2} \tag{4.7}$$

where Λ' is the inverse of the diagonal matrix whose diagonal entries are those of $D\Sigma D^T$; $\hat{\sigma}^2$ and Σ are plug-in estimates, estimated from the sill and range of the variogram respectively.

4.2.2.2 Local test

This is a simplified version of the global test, where there is prior information about the possible location of a change-point, τ. Here, our focus zooms in onto that particular point of interest, τ, to investigate if the change is a significant abrupt shift, or if it can be attributed to random variation.

The hypotheses can be stated as:

H_0: g is continuous at τ over the interval $(0, 1)$.

H_1: g has a jump discontinuity at τ in the interval $(0, 1)$.

The first form of the test statistic of the global test, Equation (4.6), can be simplified to

$$F_{O.local}(h) = \frac{Y^T d^T d Y}{\frac{Y^T B Y}{tr(B)} v_\tau} \tag{4.8}$$

where d, w_L and w_R are vectors of length n, $d = (w_L - w_R)$ is evaluated at τ, and $v_\tau = d\Sigma d^T$.

The new test statistic which was introduced in Equation (4.7) for the global test can also be simplified to give

$$F_{N.local}(h) = \frac{Y^T d^T dY}{\hat{\sigma}^2 v_\tau} \tag{4.9}$$

where σ^2 is estimated from the sill of the variogram.

4.2.3 Reference distribution of test statistic

To assess the significance of the observed differences, we use the observed value of $F(h)$, which is of the form $F(h) = \frac{Y^T AY}{Y^T BY}$. This arises when σ^2 can be expressed in quadratic form, as in the equally spaced setting. H_0 is rejected if the observed value of the test statistic, $F(h)$ is bigger than some critical value, c_α when H_0 is true, i.e.

$$P(F(h) > c_\alpha) = \alpha$$

Then the p-value of the test for the first form of test statistic, F_O, is expressed as in Equation (3.12) for the equally spaced setting.

However for the new test statistic, F_N, there will be some slight changes in the evaluation of the p-value. Let k be the observed value of the test statistic, then the p-value of the test is

$$\begin{aligned} p &= P(F(h) > k | H_0 \text{ is true}) \\ &= P\left(Y^T \frac{D^T \Lambda' D}{\hat{\sigma}^2} Y > k\right) \qquad (4.10) \\ &= P\left(Y^T D^T \Lambda' DY > \hat{\sigma}^2 k\right) \qquad (4.11) \\ &= P(Y^T QY > \hat{\sigma}^2 k) \end{aligned}$$

where Q is a symmetric matrix, $Q = D^T \Lambda' D$, and Λ' is the inverse of the diagonal matrix whose diagonal entries are those of $D\Sigma D^T$.

If the bias term in $Y^T QY$ is approximately zero, then $F(h)$ has the distribution of a

quadratic form where the random variables are normal with means approximately zero, i.e.

$$Y^TQY \approx \varepsilon^TQ\varepsilon \qquad \text{where } \varepsilon \sim N(0, \sigma^2\Sigma),$$

where Σ is the correlation matrix of Y.

Once again, we make use of the general results about the distribution of quadratic forms (Johnson and Kotz, 1972), to compute the probability p. This is done by matching the moments of the final quadratic expression, Y^TQY, to a χ^2 distribution of the form $a\chi^2_b + c$. The cumulants of $\varepsilon^TQ\varepsilon$ are obtained as functions of $tr(Q\hat{\sigma}^2\Sigma)^j$. This is given in the expression below.

$$\kappa_j = 2^{j-1}(j-1)!tr(Q\hat{\sigma}^2\Sigma)^j, \tag{4.12}$$

The p-value for F_N is estimated as $1-q$, where q is the probability lying below the point $(-c+\hat{\sigma}^2k)/a$ in a χ^2 distribution with b degrees of freedom. This is different from the procedure for the p-value for F_Q, where q is simply the probability lying below the point $-c/a$ in a χ^2 distribution with b degrees of freedom.

Alternatively, from Equation (4.10), we can also express Q as $\frac{D^T\Lambda' D}{\hat{\sigma}^2}$. The cumulants of $\varepsilon^TQ\varepsilon$ are still obtained as functions of $tr(Q\hat{\sigma}^2\Sigma)^j$, using the covariance matrix, as above. The p-value for F_N can be evaluated as $1-q$, where q is the probability lying below the point $(-c+k)/a$ in a χ^2 distribution with b degrees of freedom. This gives the same p-value as above.

4.2.4 Choice of variance estimator

As σ^2 is unknown, as is almost always the case in practice, an estimator of σ^2 has to be incorporated into the test procedure. Two possible estimators, Rice and Gasser, that we have seen earlier in the previous chapter can be considered here for F_Q.

As the data are irregularly spaced, the variogram has been proposed to model the dependence structure of the data. This not only gives an estimate of the strength of dependence of the data from the estimated range, it also provides an estimate of the error variance from the

estimated sill. Hence a plug-in estimate of the sill of the variogram can also be considered as the variance estimator. This is used in the new test statistic, F_N.

In the situation where the range of dependence is known, the new test statistic then requires only the estimation of the sill. However, before computing the empirical variogram model of the data, it is critical that the mean of the data is constant, in order for the stationary assumption of the variogram to be valid. Hence it is important that we first remove any trend or jump that might be present in the data. To do this, the data are first detrended by fitting a nonparametric smooth curve using a smoothing parameter, *h.trend*. The residuals thus obtained are then used for the variogram estimation via Equation (4.2). An exponential theoretical variogram model is subsequently fitted to the empirical variogram by weighted least squares to obtain the sill. Finally, the discontinuity test is carried out using a smoothing parameter, *h.test*.

However, if both the range and sill are unknown, the estimation of both of these quantities is required for F_N. On the other hand, F_O only requires a plug-in estimate of the range. The estimation of the correlation will be discussed in greater detail in the following section.

4.2.5 Adjusting the test for correlation

Here, we face a similar challenging task as that in the equally spaced setting in estimating the correlation of the data. As the data are dependent, we need to estimate the correlation structure before carrying out the discontinuity test. Hence, this requires us to carry out a two stage procedure, whereby the aim of the first stage is to estimate the correlation structure, Σ of the data. The second stage involves carrying out the discontinuity test, taking Σ into account.

In this section, various ways for incorporating the estimation of the correlation through the fitting of an empirical variogram model in the first stage of the discontinuity test will be discussed.

The Residual-Variogram procedure
1st stage:

- Remove the trend in the data by fitting a smooth curve using a particular smoothing parameter, *h.trend.*
- Using the residuals, we obtain an estimate of the correlation matrix, Σ, by weighted nonlinear least squares fitting of the variogram to get the range and sill.

2nd stage:

- Insert the estimate of Σ into the discontinuity test using either F_O or F_N as the test statistic, with bandwidth, *h.test*, and obtain the p-value of the test.

Figure 4.2. The Residual-Variogram procedure

4.2.5.1 The Residual-Variogram approach

One particular proposed technique here is the *residual-variogram approach*, which is very similar to the residual approach that we have discussed in the equally spaced setting in Section 3.2.7. The only difference is that the dependence structure is estimated using the variogram, and hence is computationally more intensive than computing the sample correlation coefficient of an AR(1) model.

After detrending the data with a smoothing parameter *h.trend*, the residuals are then used to obtain the empirical variogram via Equation (4.2). Subsequently, an exponential variogram model is then fitted to model the underlying correlation structure of the data. This procedure is referred to as the *residual-variogram approach* and is summarised in Figure 4.2.

The possible difficulty we might encounter using this first method to estimate correlation is that the resulting residuals are highly sensitive to the smoothing parameter, *h.trend*, used to remove the trend. Furthermore, if a considerable jump is present in the data, using a smooth curve might not succeed in removing it. This might result in the correlation being overestimated, hence a decrease in power. This is similar to the difficulties we have discussed in the equally spaced setting in Section 3.2.7, where it is suggested that a slightly lower *h.trend* should be used in the first stage, compared to the second stage.

4.2.5.2 The Double Discontinuity Test approach (DDT)

The second suggested technique evaluates the significance of the hypothesis that there is discontinuity by carrying out the discontinuity test twice. Hence we will term it the *double discontinuity test*. It tries to remove any possible jump that might be present by adding an extra step in this first stage of correlation estimation. This is done by first carrying out the discontinuity test, treating the data as independent. As we have seen in the earlier chapter, treating positively correlated data as independent (as in the case here), will increase the probability of detecting a jump, even if it is absent.

If the test is significant, the regression function is then adjusted for the jump identified at the position that has the highest standardised jump. The same smoothing parameter is used for both the first stage discontinuity testing where the data are treated as independent, and for the trend removal of the adjusted data, since the selection of these is dependent on how smooth we believe the underlying trend is, under the null hypothesis of no jumps. We will denote this by h_1. The smoothing parameter for the final discontinuity test is denoted by *h.test*. The algorithm is summarised in Figure 4.3.

4.2.5.3 Remarks

The moving variogram approach, an analogy of the moving window approach discussed in Section 3.2.7, could also be considered here. The idea is similar, whereby an empirical variogram is fitted to each "moving" set of b observations, assumed to have constant trend and without discontinuity. An alternative but similar approach is to use a segmented variogram approach. This involves dividing the data into several segments and obtaining an estimated Σ for each of these segments, which are assumed to be stationary. For both techniques, the final estimated Σ is obtained by taking the median of all the estimated ranges (and sill, if using F_N) in each of the separate segments.

However, the main drawbacks are that they are not only computationally intensive, but a relatively large data set will be required to obtain a good estimate of Σ. Hence, we will proceed to the estimation of the dependence structure in an unequally spaced setting via the residual-variogram approach and the double discontinuity test in our simulation study. Both

The Double Discontinuity Test procedure, DDT
1st stage:

- Step 1:
  - Treating the data as independent, the discontinuity test is carried out with a smoothing parameter, h_1. If $p < 0.05$, then adjust the data at the change-point location (with the highest standardised jump). This step removes any jump that might be present.
- Step 2:
  - Remove the trend in the data by fitting a smooth curve using the same smoothing parameter as before, h_1
  - Using the residuals, we obtain an estimate of the correlation matrix, Σ, by weighted nonlinear least squares fitting of the variogram to get the range and sill. (This step is similar to the residual-variogram approach)

2nd stage:

- Insert the estimate of Σ into the discontinuity test using either F_O or F_N as the test statistic, with bandwidth, *h.test*, and obtain the p-value of the test.

Figure 4.3. The Double-Discontinuity Test procedure, DDT

the original test statistic, F_O or the new test statistic F_N can be used. The latter makes use of both the range and sill estimates from the variogram fitting.

4.3 Simulation Study

In this section, we will report on a simulation study to investigate how the discontinuity test performs when it is applied to a set of unequally spaced correlated data. The global test is used for the rest of this chapter, except for Section 4.4.2 where we investigate the size and power using the local test.

We consider a model

$$y = g(x) + \varepsilon$$

where $g(x)$ is the regression function of x or trend of the data, x is irregularly spaced between

0 and 1 and ε is the correlated noise.

To obtain our simulated data, we first simulate an irregularly spaced set of 100 points between 0 and 1, denoted as x, from the uniform distribution. Figure 4.4 shows the simulated locations x. We then proceed to simulate a series of correlated data, y, which has an exponential covariance structure over the set of points, x. The point 'm' denotes the median of the data, which is taken as the change-point location for the power simulations.

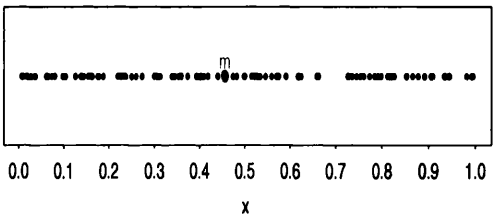

Figure 4.4. The figure shows the locations of the unequally spaced observations, x. The letter 'm' denotes the median of the data, located at 0.456.

As in the simulations done in the equally spaced setting, the performance of the test under a variety of factors, namely the effects of correlation size, variance estimator and its performance in a local test will be investigated. However, the study will be of a smaller scale as factors such as effects of trend function, treating correlated data as independent, and sample size will have similar results to those reported in the equally spaced setting. Similar to Chapter 3, the simulations are repeated at several levels of the smoothing parameter and plots are made of the empirical size and the power against the smoothing parameter values.

Table 4.1 shows the data conditions of the simulation study. Both forms of test statistic, F_O and F_N, will be used. We will first investigate how the various factors affect the performance of our test, by assuming that the correlation is known. In the later section, we will then investigate the performance of the suggested algorithms to adjust the test for correlation, by incorporating the correlation estimation in the procedure. To estimate the range and/or sill, the robust estimator of the variogram introduced in Section 4.1.2 is produced using the factors as suggested at the end of the section. Lags or bin widths of 0.01 and 0.02,

which bin observations into bins that are 0.01 and 0.02 distance apart respectively, are used as they do affect the estimation of the variogram to a certain extent.

Data conditions	
Design Space	**Unequally-spaced**, generated from Un(0,1)
No. of simulations	**200**
Trend Function	**flat** and **sine**
Sample size	**n = 100**
Error	$\varepsilon \sim N(0, \sigma^2\Sigma)$
Error Variance	**1**
Covariance Structure	**exponential**
Correlation	Range, $r = \{\mathbf{0.005}, \mathbf{0.01}, 0.02, 0.04, 0.08, 0.12, 0.16\}$
Size of Jump	$jump = 2$, **3** at median of data, m
Test Approaches	
Smoothing Parameter	$h.test = \{\mathbf{0.08}, \mathbf{0.12}, \mathbf{0.16}, \mathbf{0.2}\}$
Estimator of Error Variance	**Rice** (F_O), Gasser (F_O), **sill** (F_N)
Estimator of Correlation	weighted least squares fitting of variogram using maxdist=0.5, nlag=n and lag=0.01,0.02.
Algorithm to adjust for correlation	- none (for Section 4.4); - Residual-Variogram approach using $h.trend = \{0.06, 0.08, 0.1, 0.12\}$ - Double Discontinuity Test (DDT) using h_1

Table 4.1. Simulation settings for one-dimensional unequally spaced data. The parameters in bold are the ones used by default in the simulations, unless specified otherwise.

To provide a better understanding of the degree of dependence of the different ranges of the exponential variogram as compared to the correlation coefficient of an equally spaced AR(1) model, let us first consider the theoretical connections between these models.

For an exponential variogram model as expressed in Equation (4.4),

$$\begin{aligned} \text{Cov}(Z(s), Z(s+h)) &= \sigma^2 - \gamma(h) \\ &= \sigma^2[1 - (1 - \exp(-h/a))]. \\ \text{Therefore } \text{Corr}(Z(s), Z(s+h)) &= \rho(h) = \exp(-h/a). \end{aligned}$$

If one samples a continuous-time process with autocorrelation function $\rho(h) = \exp(-h/a)$

at equal intervals of t time units, the resulting discrete-time process is an AR(1) model with parameter $\alpha = \exp(-t/a)$.

To provide an illustration of this result, Table 4.2 displays the approximate equivalent correlation of an AR(1) model for the different ranges when the data are equally spaced with $n = 50$ and 100. These results are obtained by simulating 200 sets of 50 and 100 data-points with different ranges over an equally spaced interval, and then computing the median of the 200 corresponding correlation coeffients of an AR(1) model.

For $n = 100$, the data are sampled at 0.01 time units. Hence $\alpha_{100} = \exp(-0.01/a)$. The time interval for $n = 50$ is 0.02. Hence $\alpha_{50} = \exp(-0.02/a) = \alpha_{100}{}^2$ for the same range a. Thus, keeping the range a fixed, the correlation coefficient of an AR(1) model for $n = 50$ is the square of the correlation coefficient for $n = 100$. For instance, for range of 0.02, the estimated correlation coefficient for $n = 100$ is 0.59 and the correlation coefficient for $n = 50$ is 0.34, which is the approximate square of 0.59 as observed from Table 4.2.

Conversely, to obtain the same correlation coefficient for $n = 50$ and $n = 100$, i.e. $\alpha_{50} = \alpha_{100}$, the range for the former has to be twice that of the latter. This can be observed from Table 4.2, where for $r = 0.01$, the correlation for $n = 100$ is approximately the same as for $r = 0.02$ for $n = 50$.

Corresponding AR(1) correlation coefficient, *corr* for range, r		
Range, r	*corr* for $n = 100$	*corr* for $n = 50$
0.01	0.36	0.11
0.015	0.50	0.23
0.02	0.59	0.34
0.03	0.69	0.47
0.04	0.76	0.57
0.08	0.86	0.72
0.12	0.89	0.79
0.16	0.91	0.82

Table 4.2. Corresponding estimated correlation coefficients of an AR(1) model of different ranges of an exponential covariance model for $n = 50$ and 100.

4.4 Unequally Spaced Setting: Correlation known

This section investigates the influence of various factors on the discontinuity test, assuming that the correlation of the data, corresponding to the range of dependence, r, is given in the unequally spaced setting. The effects of different ranges, r on the performance of the test, using both forms of test statistic, will be investigated in Section 4.4.1. This section also investigates the performance of the discontinuity test using different variance estimators, namely Rice or Gasser for F_O, where it is incorporated into the test statistic via quadratic forms; and the plug-in sill estimate of the variogram fitting for F_N.

Lastly we have also displayed results using a local test, which we would expect to perform better than the global test as seen earlier in the equally spaced setting.

4.4.1 Effects of correlation and variance estimators

In this section, we investigate the effects of different degrees of dependence, namely the range of dependence, denoted by r, and the variance estimators, on the power and size of the discontinuity test.

For illustrative purposes, let us first look at Figure 4.5 which consists of sets of data with different ranges r, comprising of $r = 0.005$ to $r = 0.16$ simulated over locations x (as shown in Figure 4.4), with error variance of 1. These data are simulated from the same seed, which explains the apparent decreasing trend in all the plots. As the data are unequally spaced, and are quite sparse around 0.7, some of the data appear to have a discontinuity even though it is absent. This is very evident especially for data with stronger dependence structures, such as $r \geq 0.01$.

Figure 4.6 displays the size and power of the test using F_O and F_N, for both the flat and sine functions over different smoothing parameters, with range of dependence, r of 0.005, 0.01, 0.08 and 0.16. The first two values of r represent low and moderate dependence structures, while the last two correspond to highly dependent data. The results obtained using Rice and Gasser in F_O, and sill from variogram estimation in F_N are represented by red, blue and green lines respectively. To obtain the sill for the latter, the procedure as mentioned earlier in Section 4.2.4 is used. Lags or bin widths of 0.01 and 0.02 are used and

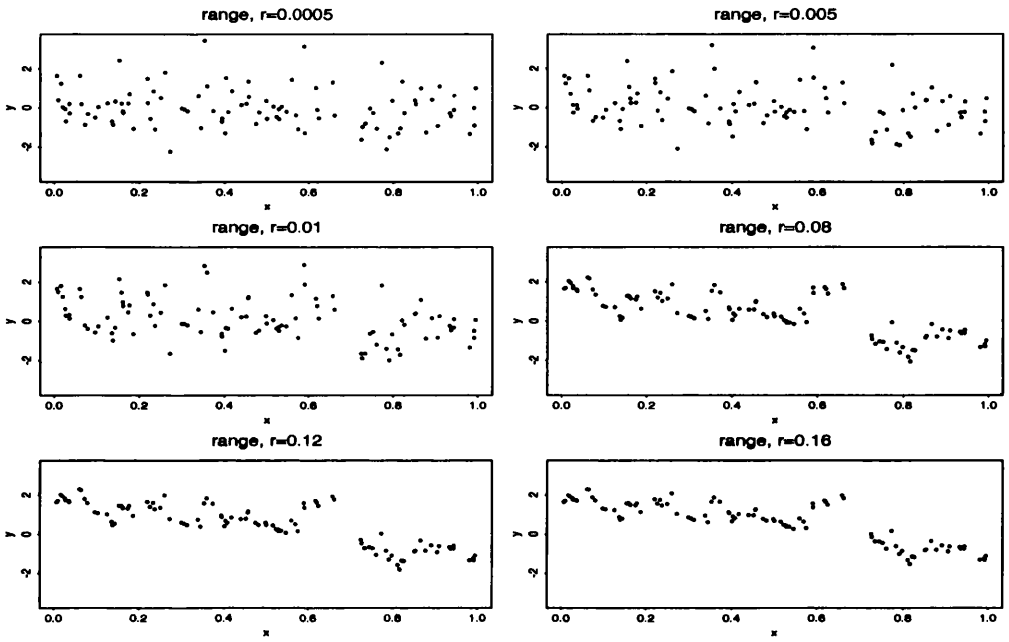

Figure 4.5. An example of sets of simulated 100 data-points with different ranges over fixed unequally spaced locations x (under the null hypothesis of no discontinuity). A flat trend function is used, with error variance of 1.

represented by solid and dotted green lines respectively.

In the situation where the trend is flat, the size is approximately 0.05 over all *h.test* and r for both variance estimators of Rice and Gasser using F_O. On the other hand, using F_N, the size of the test is within limits for $r = 0.005$ and 0.01, but is greatly inflated for high ranges such as 0.08 and 0.16 at low *h.test* but conservative at higher *h.test*.

The corresponding power using Rice is higher than Gasser over all r and *h.test*. They both have substantially better power than F_N for $r = 0.005$ and 0.01. For large r, the power using F_O is low. On the other hand, power for F_N cannot be interpreted as the size is not well controlled. The power using lag 0.02 is generally slightly higher than that using lag 0.01.

The bottom panel of Figure 4.6 displays the size and power for the sine trend. A similar pattern as described above can be found for the size using both F_O and F_N. The only difference is that the size is only within limits for low to moderate *h.test* for F_O. The corresponding power for the sine trend is lower than that for the flat function. The Rice

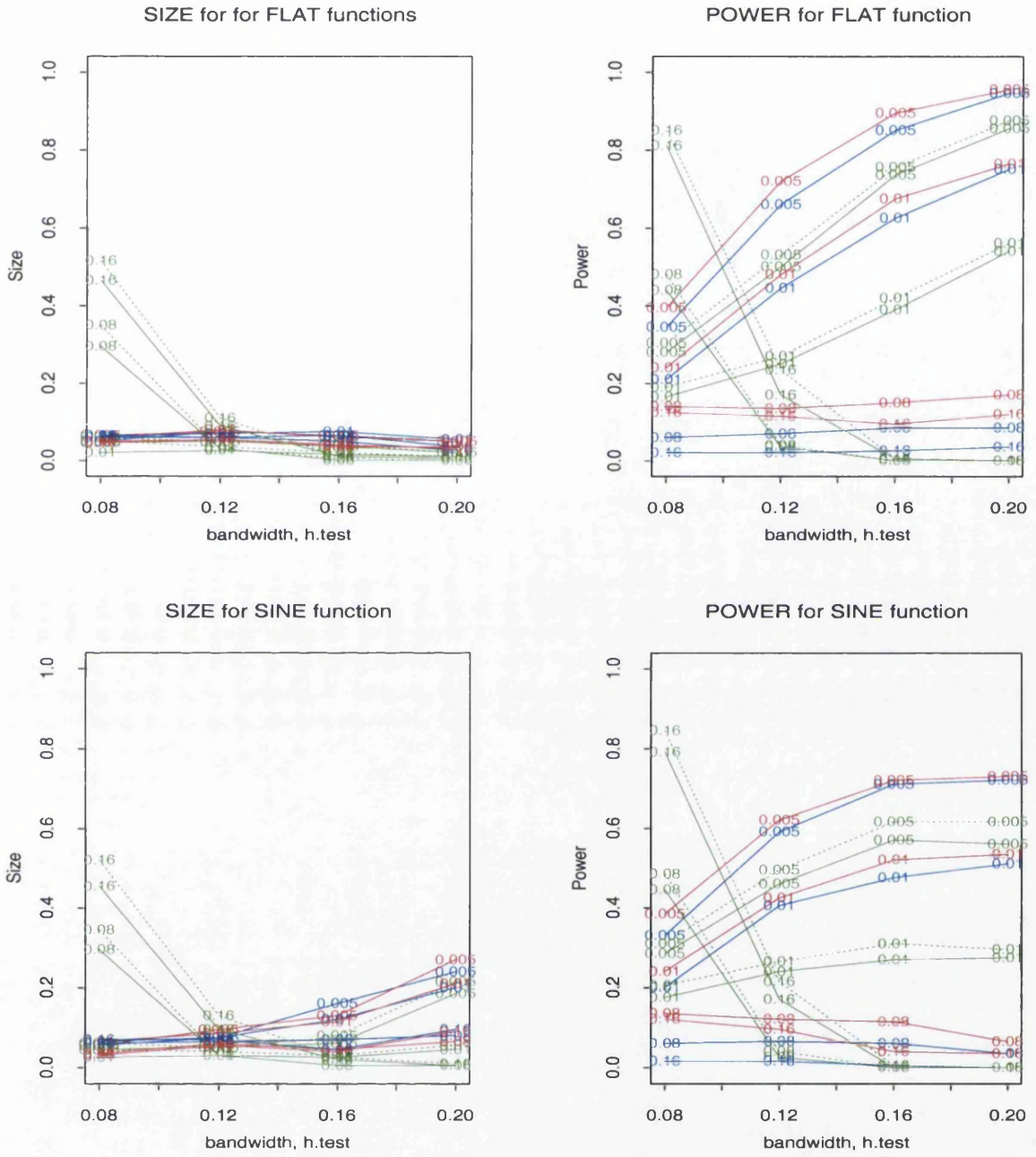

Figure 4.6. Effects of Variance Estimators. Size and Power using Rice (red), Gasser (blue) variance estimators for F_O and estimated sill (green) for F_N, assuming correlation is known. Both flat and sine trends are used with $jump = 3$ for power calculations. The numbers on the curve indicate the range, r. Solid: lag 0.01; dotted: lag 0.02.

variance estimator still performs better than the Gasser one. Both have very much more superior power than the new test statistic with the sill.

The poor performance of the new test statistic can be due to the fact that the sill is not well estimated. Next, we shall look at the graphical summaries of the estimated sill under H0 and H1 for both trend functions, and for $r = 0.005$ and $r = 0.08$ in Figures 4.7 and 4.8 respectively, to investigate the possible causes.

From Figure 4.7, it can be observed that the sill is quite well estimated for $r = 0.005$ under H0 with both the flat and sine trends, using different *h.trend*. However, it is greatly inflated under H1 for both trends, with a worse performance in the flat function, thus resulting in low power. On the other hand, using a small smoothing parameter such as $h.trend = 0.08$ for $r = 0.08$ causes the sill to be very much underestimated, because this traces the data too closely. This causes the size to be greatly inflated as seen in Figure 4.6. The inflation of the sill estimate under H1 for $r = 0.08$ increases as *h.trend* increases. This results in the test

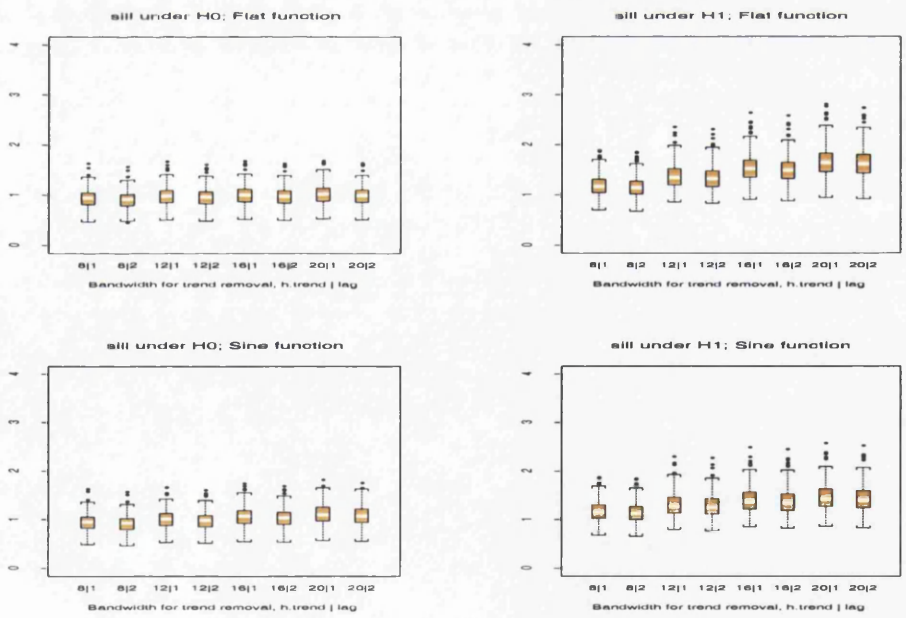

Figure 4.7. Boxplots of estimated sill from variogram used in F_N, for models under H0 and H1. Both flat and sine trends and range, $r = 0.005$ (assumed known) are used.

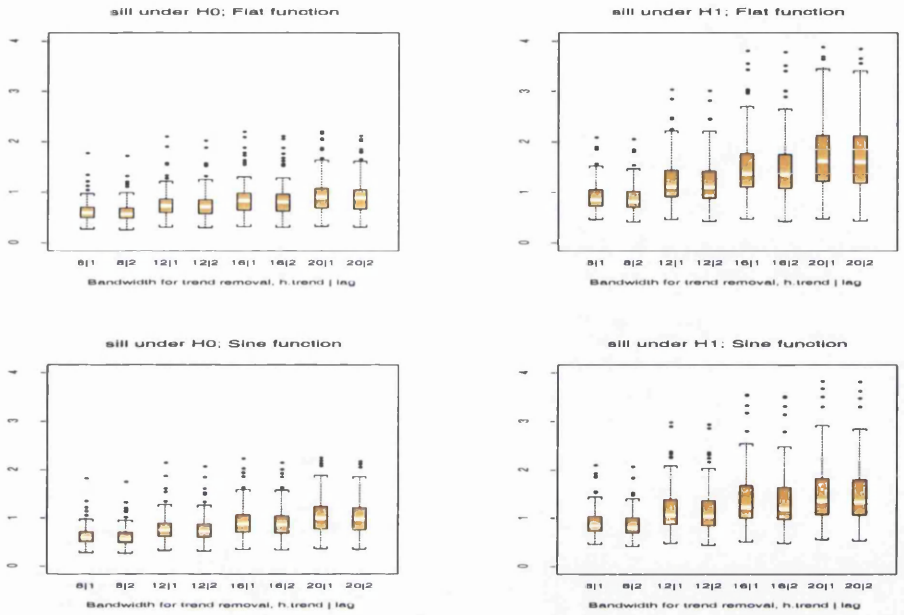

Figure 4.8. Boxplots of estimated sill from variogram used in F_N, for models under H0 and H1. Both flat and sine trends and range, $r = 0.08$ (assumed known) are used.

being very conservative, having power close to 0 at high $h.test = h.trend$. These features explain why the new test statistic has not performed well as the sill estimate is very sensitive to the $h.trend$ value used. Hence its performance is poorer than F_O which does not require the plug-in estimate.

Another general notable feature in all the boxplots is that the estimated sill is slightly lower using lag 0.02, as there is more averaging involved in binning observations that are 0.02 apart, compared to 0.01. Hence the size and power for lag 0.02 are generally higher than that of 0.01 for F_N.

4.4.2 Local test

In this section, we will investigate the test performance for a local test using F_O. Here, the only evaluation point is at the median of the data, which is at location m, as shown in Figure 4.4.

From Figure 4.9, we can observe that the size for the local test shows similar features to that for the global test for both trend functons. The power, however, for the local test for both test statistics is very much higher than that for the global test (compare with Figure 4.6). For instance, power for a very high range such as 0.16 with a flat trend is approximately 0.08 in the global test, but has increased to 0.85 for F_O (at $h.test = 12$). This indicates that it is greatly beneficial to the detection of a change-point if additional information as to where the possible change-point might occur is given, especially for data with a very high range.

From Figure 4.9, a very interesting feature that is not apparent in the global test for F_O is observed. At higher *h.test*, the power decreases as the degree of dependence, r increases as seen earlier in the global test. However, at lower *h.test*, power for large r, such as 0.08, 0.12 and 0.16, is higher than at 0.02 and 0.04. In fact, for these three high ranges, power appears to decrease as *h.test* increases. This feature can also be observed in the sine trend.

4.4.3 Remarks

In the situation where the correlation is assumed to be given, the simulations have shown the following features in an unequally spaced design.

1. The newly introduced test statistic, F_N, is found to perform differently to the original test statistic, F_O. The difference is effectively due to the different estimators of error variance used and how they have been incorporated into the test. The original test statistic incorporates the variance estimation via quadratic forms in its test statistic, and hence does not depend on its true value, while the F_N uses a plug-in estimate obtained from the sill of the variogram estimation. Both the Rice and Gasser estimators for F_O perform better than the plug-in sill for F_N over all r. The estimation of the sill is investigated and is shown to be inflated for high range and sensitive to the value of *h.trend* used to remove the trend. It also does not cope well in the presence of jumps. This explains why the size is not well controlled and the power is lower.

   However, the estimation of the sill might be linked with the range and in this case, we have assumed the latter to be known. Hence, it would be worthwhile to revisit this

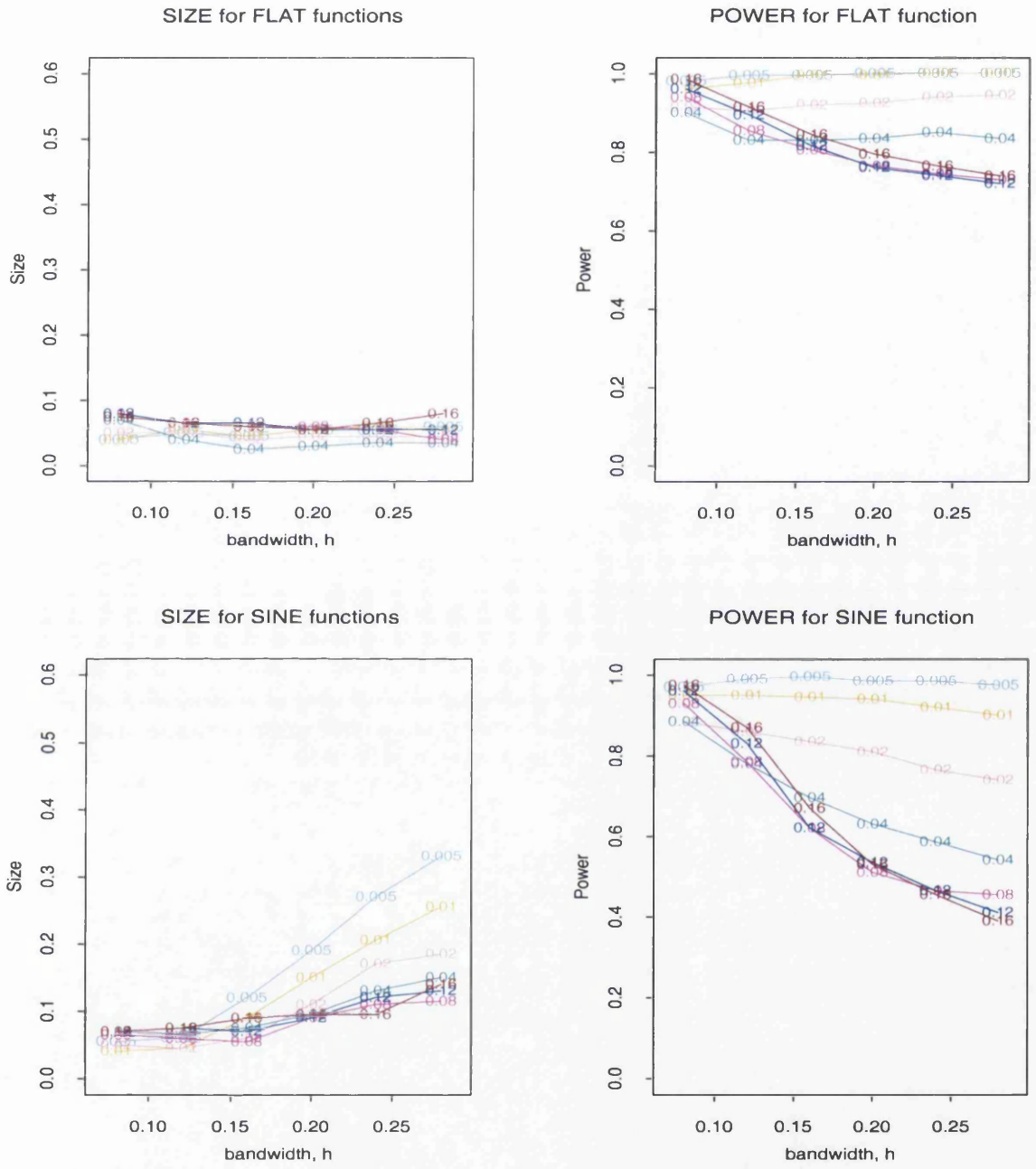

Figure 4.9. Effects of correlation and smoothing parameter for local test using F_O (with Rice). Both flat and sine trends are used with correlation assumed known. The values on the curves denote the range, r.

new test statistic again in the later section where both of these unknown parameters are estimated.

2. The Gasser variance estimator performs slightly poorer than the Rice one for both the flat and sine trends.

3. For both test statistics, there is a general pattern that as the degree of dependence r increases, power decreases. This is because a large range corresponds to a greater dependence structure in the data, and hence will result in lower power since it is equivalent to having a smaller number of independent observations. Another explanation of this phenomenon is that the correlation in the data results in a less precise estimate of the underlying regression function, as shown in Equation (3.31), resulting in lower power. The power for high dependence of $r \geq 0.08$ is very low at < 0.2 for both trends using F_O.

4. Similarly as in the equally spaced setting, the size is within limits over all smoothing parameters for a flat trend. It is however only within limits for low to moderate smoothing parameters for a sine trend. The power is also higher for the flat trend.

5. The power of the local test with F_O is investigated and shown to be substantially higher than that of the global test, particularly for the higher range of dependence.

4.5 Unequally-spaced setting: correlation unknown

In this section, the correlation is *not* assumed to be known and has to be estimated. The correlation/covariance structure of the data will be modelled via the exponential variogram model, using the two proposed techniques, namely the residual-variogram approach and the double discontinuity test, which were discussed in Section 4.2.5, to investigate how they perform, using both F_O and F_N. In the situation where it is impossible to fit the exponential variogram via nonlinear weighted least squares fitting, which often occurs when the range of dependence is low, the range is then set to zero. Only the sill is estimated.

The results obtained using the residual-variogram approach are presented in Section 4.5.1, followed by the double discontinuity test in Section 4.5.2.

4.5.1 Residual-variogram approach

Figure 4.10 presents the results for data with a range of 0.005 using F_O. The values on the figure indicate the values of *h.trend* that are used to remove the trend. The coloured solid lines represent the size and power produced using lag 0.01, and the dotted lines represent those using lag 0.02 for the variogram estimation. Graphical summaries of the estimated range and sill under the null and alternative hypothesis (with $jump = 3$) using different *h.trend* for both trends, are also provided in Figures 4.18 and 4.19.

Let us first focus on the size of the test under both trend functions displayed on the left panel of Figure 4.10. The size is within limits over all *h.trend* evaluated, though it increases as *h.trend* decreases for both trends. The size is highest for $h.trend = 0.06$, but is still within sensible limits.

From the right panel, it can be observed that as *h.test* increases, power increases as expected for both trends. Other notable features are that as *h.trend* decreases, power increases; and power for lag 0.01 is generally greater than that for 0.02. Using a value of $h.trend = 0.06$ give results that are quite close to the size and power curves with the true Σ (represented by a black solid line) for both trends, while using a higher value of *h.trend* will result in a decrease in power.

Next, we will investigate how the new test statistic, F_N perform. Figure 4.11 presents the results with the same 200 simulated sets of 100 data-points. Here, both the estimates of the sill and range of the exponential variogram model are required.

The power using F_N is higher than that of F_O for each *h.trend*, but it should be noted that the size is also higher. For the flat trend with F_N, using a value of $h.trend = 0.06$ gives power that is close to the true power (i.e. with true Σ), but the size is quite high at approximately 0.1 over the range of *h.test* considered. Hence a higher *h.trend* value should be recommended in this case. The 'best' power, with controlled size is then approximately the same as that using F_O. It can be noted too that the power using the true covariance in

F_N is higher than that using only true correlation in F_O.

Subsequently, the performance of the two test statistics, F_O and F_N, is also investigated for a higher range of $r = 0.01$. Figures 4.12 and 4.13 present the size and power curves using F_O and F_N respectively. From Figure 4.12, we can observe that the most suitable *h.trend* value is 0.08 for the flat function, and 0.1 for the sine function. The best power for the flat and sine functions, with appropriate size, is about 0.6 (with $h.trend = 0.08$ and $h.test = 0.16$) and 0.24 (with $h.trend = 0.1$ and $h.test = 0.14$) respectively. However, the resulting power using these estimators is lower than that when the true correlation is known, especially in the presence of the sine trend. We can note that the values of recommended *h.trend* for $r = 0.01$ are higher than that for a lower range, $r = 0.005$.

In the case where F_N is used for the discontinuity test, the power is generally higher than for F_O. However, the size is higher too, as is seen in the earlier case for $r = 0.005$.

From our simulations, we can see that the residual-variogram approach is very sensitive to the values of the smoothing parameter, *h.trend*, which is used to remove the trend in the first stage of the discontinuity test, because this determines the estimated correlation/covariance. The size and power increase as the value of *h.trend* decreases. This is because a lower value of *h.trend* will give a smaller estimate of the correlation, and hence increase the power to detect abrupt changes. However, if too low a value of *h.trend* is used, the correlation might be underestimated, resulting in an inflation of the size of test. Conversely, using a large value of *h.trend* might cause an overestimation of the correlation, resulting in the test being more conservative. On the whole, the proposal to use a slightly lower *h.trend* than *h.test* for the first stage, does perform reasonably well for both trend functions and test statistics.

The algorithm performs reasonably well in the situation for low range, $r = 0.005$ when suitable *h.trend* values are used, but fares poorly for a higher range, $r = 0.01$. In the latter case, the highest power, with controlled size, is very much lower than that for the power when the true correlation is known. This indicates that it is not able to capture the underlying correlation of the data well.

If the dependence structure in the data is strong, a higher value of *h.trend* is required to give an appropriate size. However, some form of compromise would be inevitable here, because using too high a *h.trend* value will result in the correlation being overestimated,

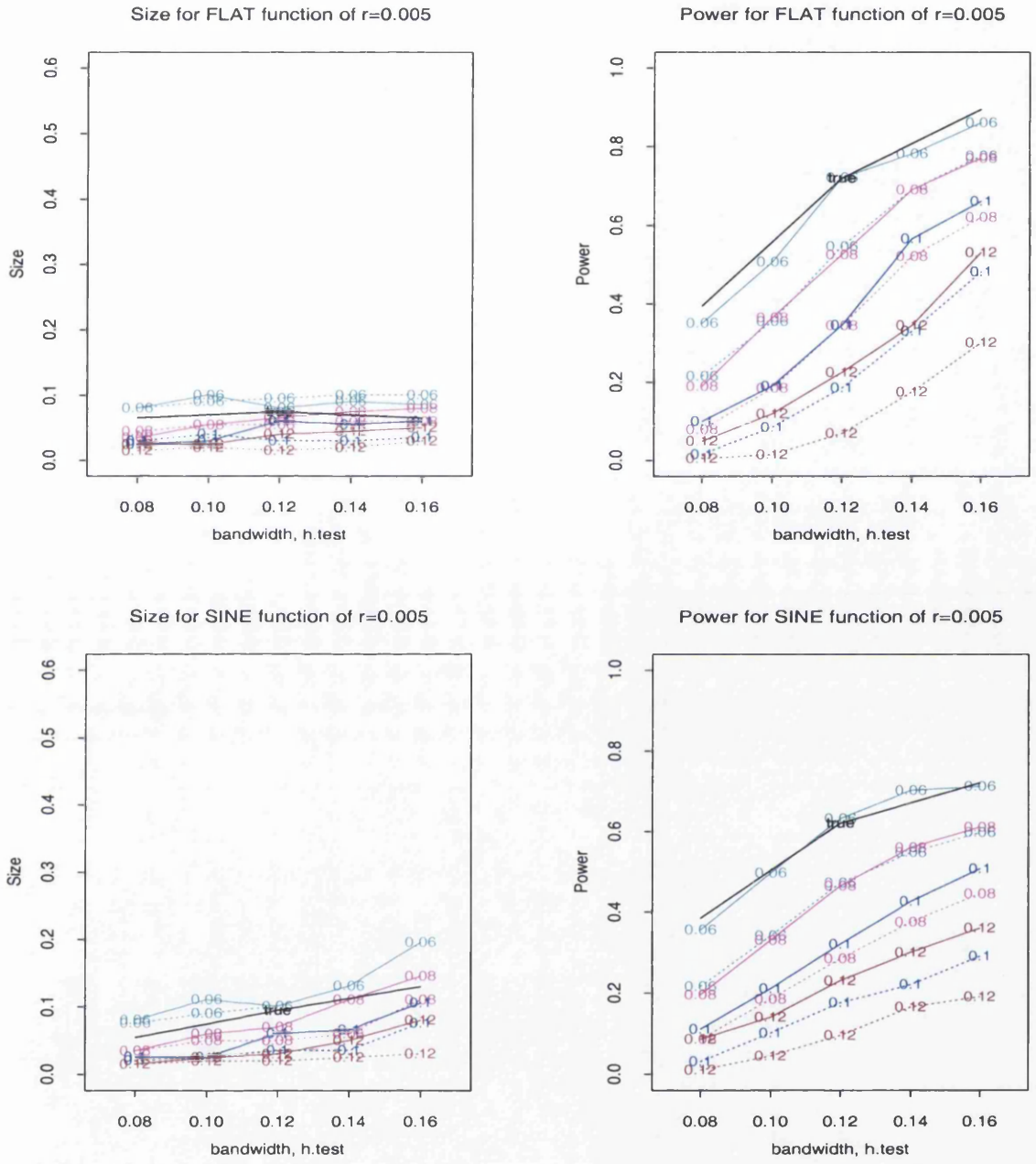

Figure 4.10. Size and Power of the residual-variogram approach with estimated Σ using F_O (with Rice), for $r = 0.005$ and $jump = 3$. solid: lag 0.01, dotted: lag 0.02. The values on the curves denote the values of *h.trend*. The size and power using true Σ are represented by black, solid lines.

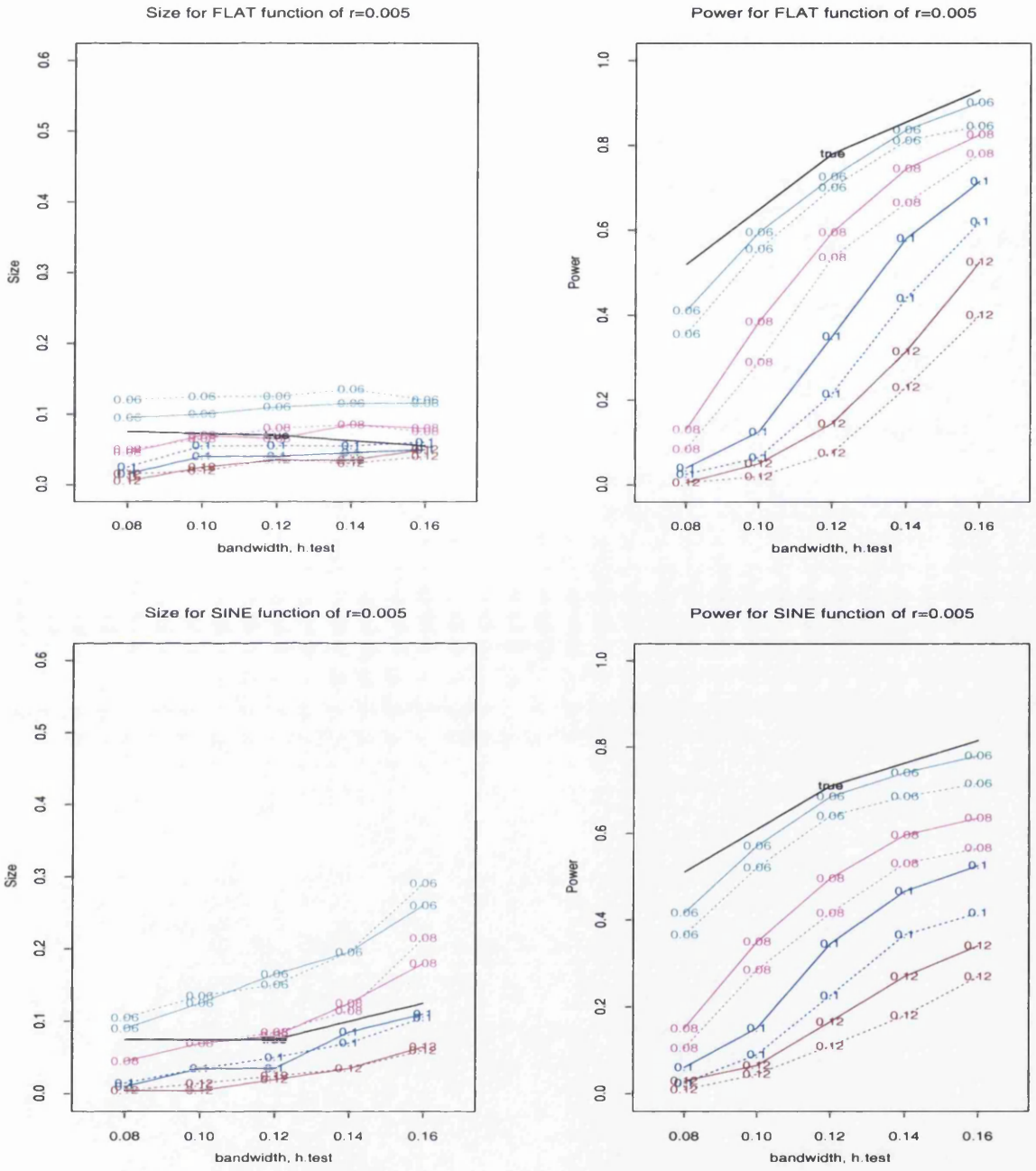

Figure 4.11. Size and Power of the residual-variogram approach with estimated Σ and sill using F_N, for $r = 0.005$ and $jump = 3$. solid: lag 0.01, dotted: lag 0.02. The values on the curves denote the values of *h.trend*. The size and power using true $\sigma^2\Sigma$ are represented by black, solid lines.

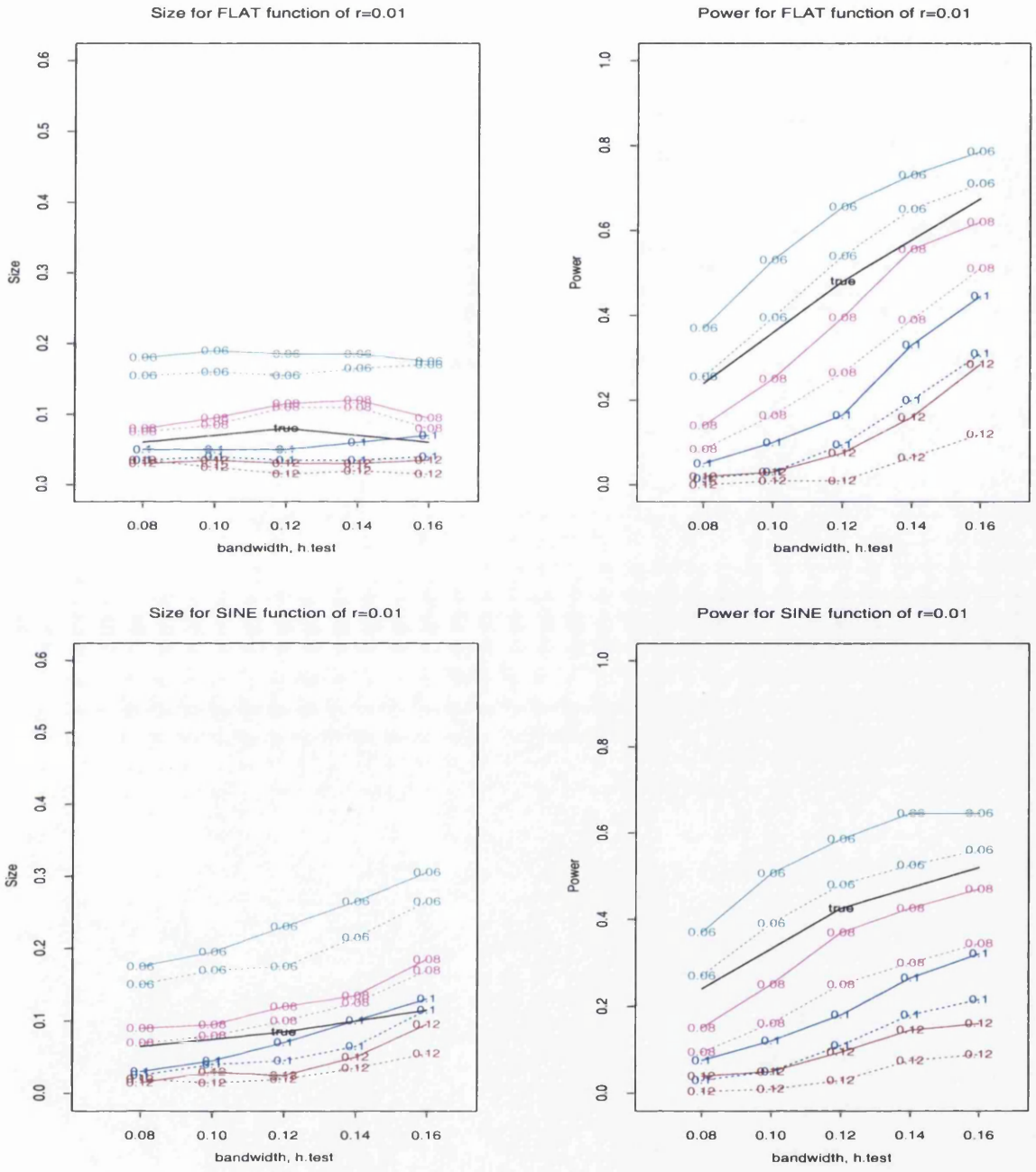

Figure 4.12. Size and Power of the residual-variogram approach with estimated Σ using F_O (with Rice), for $r = 0.01$ and $jump = 3$. solid: lag 0.01, dotted: lag 0.02. The values on the curves denote the values of *h.trend*. The size and power using true Σ are represented by black, solid lines.

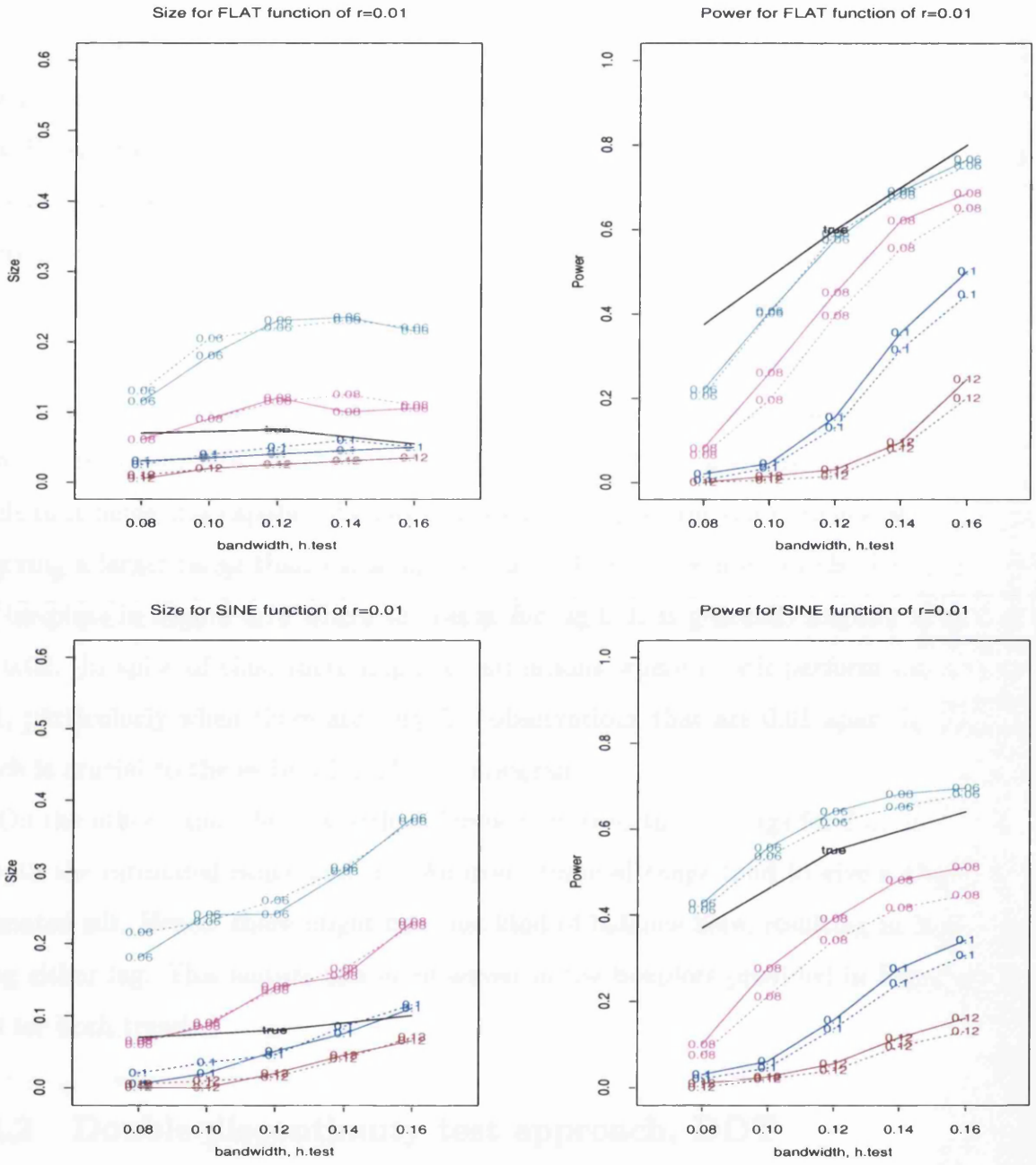

Figure 4.13. Size and Power of the residual-variogram approach with estimated Σ and sill using F_N, for $r = 0.01$ and $jump = 3$. solid: lag 0.01, dotted: lag 0.02. The values on the curves denote the values of *h.trend*. The size and power using true $\sigma^2\Sigma$ are represented by black, solid lines.

partly due to jumps that have not been completely removed, as well as any trend that might still be present.

Overall, F_O seems to be more stable and perform slightly better than F_N, when the correlation and the error variance have to be estimated. This could be due to the fact that F_N is dependent on both the range and sill estimates, which are in turn dependent on *h.trend*. If either or both the range and sill are underestimated, the size might be inflated, whereas in F_O, inflation of the size will only occur if the range is underestimated.

Another notable difference between F_O and F_N for both trends, is that the difference in size and power using lags 0.01 and 0.02 in the variogram estimation, differs more in magnitude for the former than the latter. For F_O, the size and power for lag 0.01 are usually higher than for lag 0.02. Using a larger lag of 0.02 involves more averaging and this might result in it being less capable of capturing a good range of the data. Hence, it is more prone to giving a larger range than when lag 0.01 is used. This feature can also be observed from the boxplots in Figure 4.18 where the range for lag 0.01 is generally slightly lower than for lag 0.02. In spite of this, there might be situations where it will perform better than lag 0.01, particularly when there are very few observations that are 0.01 apart in the first bin which is crucial to the estimation of the variogram.

On the other hand, there is little difference between the two lags for F_N, as it makes use of both the estimated range and sill. An overestimated range tend to give a slightly underestimated sill. Hence, there might be some kind of balance here, resulting in less difference using either lag. This feature can be observed in the boxplots provided in Figures 4.18 and 4.19 for both trends.

4.5.2 Double discontinuity test approach, DDT

This section will consider the double discontinuity test, DDT, as proposed in Section 4.2.5, to incorporate the estimation of the correlation in the discontinuity test. Both F_O and F_N will again be used with two similar low and moderate ranges as before, $r = 0.005, 0.01$. It is of interest to investigate how this algorithm performs and how it compares with the residual-variogram approach.

Figures 4.14 and 4.15 display the simulation results for $r = 0.005$ using F_O and F_N respectively for both the flat and sine trends. A wide range of smoothing parameters, $h_1 = 0.1$ to 0.18 is used for the first stage, with different *h.test* for the second stage. The first stage comprises of carrying out the discontinuity test with h_1, treating the data as independent to find the possible change-point location. It then removes the trend using the same smoothing parameter after adjusting the data by the jump size.

From the left panel of Figure 4.14, we can observe that the size increases only marginally with decreasing h_1 for both trends, but is well-calibrated regardless of what smoothing parameter is used in the first stage. It is also consistent over all *h.test*, for both lags 0.01 and 0.02 for the flat trend, and consistent over moderate *h.test* for the sine trend.

The right panel shows that the power increases marginally as h_1 decreases for both trends. The power for lag 0.01 is slightly higher than that for lag 0.02. Interestingly, the power curves using lag 0.01 for most h_1 follow the true power curve very closely and some are even slightly higher. The size and power results for both trends illustrate that the approach is only slightly sensitive to the h_1 value used for the first stage (in estimating the correlation).

Next, we will look at the performance using F_N. From the left panel of Figure 4.15, we can see that the size for F_N is generally higher than that for F_O. Larger h_1 values of 0.14 and above are required to give good size. The power differs marginally for the different h_1 used, and is somewhat higher than the power curve for the true plugged-in estimate of $\sigma^2\Sigma$, except for a large value of 0.18, where there is a considerable decrease (particularly for the flat trend). The feature of negligible difference in using lag 0.01 or 0.02 for all the h_1 values, except 0.18, for F_N can be observed again.

The simulation results for $r = 0.01$ are presented graphically in Figures 4.16 and 4.17 using F_O and F_N respectively for both the flat and sine trends.

The size for $r = 0.01$ using F_O is generally higher than that for $r = 0.005$. A larger h_1 value, such as 0.16 or 0.18, is required to keep the size within limits for the first stage, under both the flat and sine trend functions. One interesting feature of the power curves that is evident here again, is that the power for using $h_1 = 0.16$ or 0.18 is considerably higher than that for the true power curve. This feature is observed for the sine trend too.

As for F_N in Figure 4.17, the inflation of size is greater. It appears that a large *h.trend* =

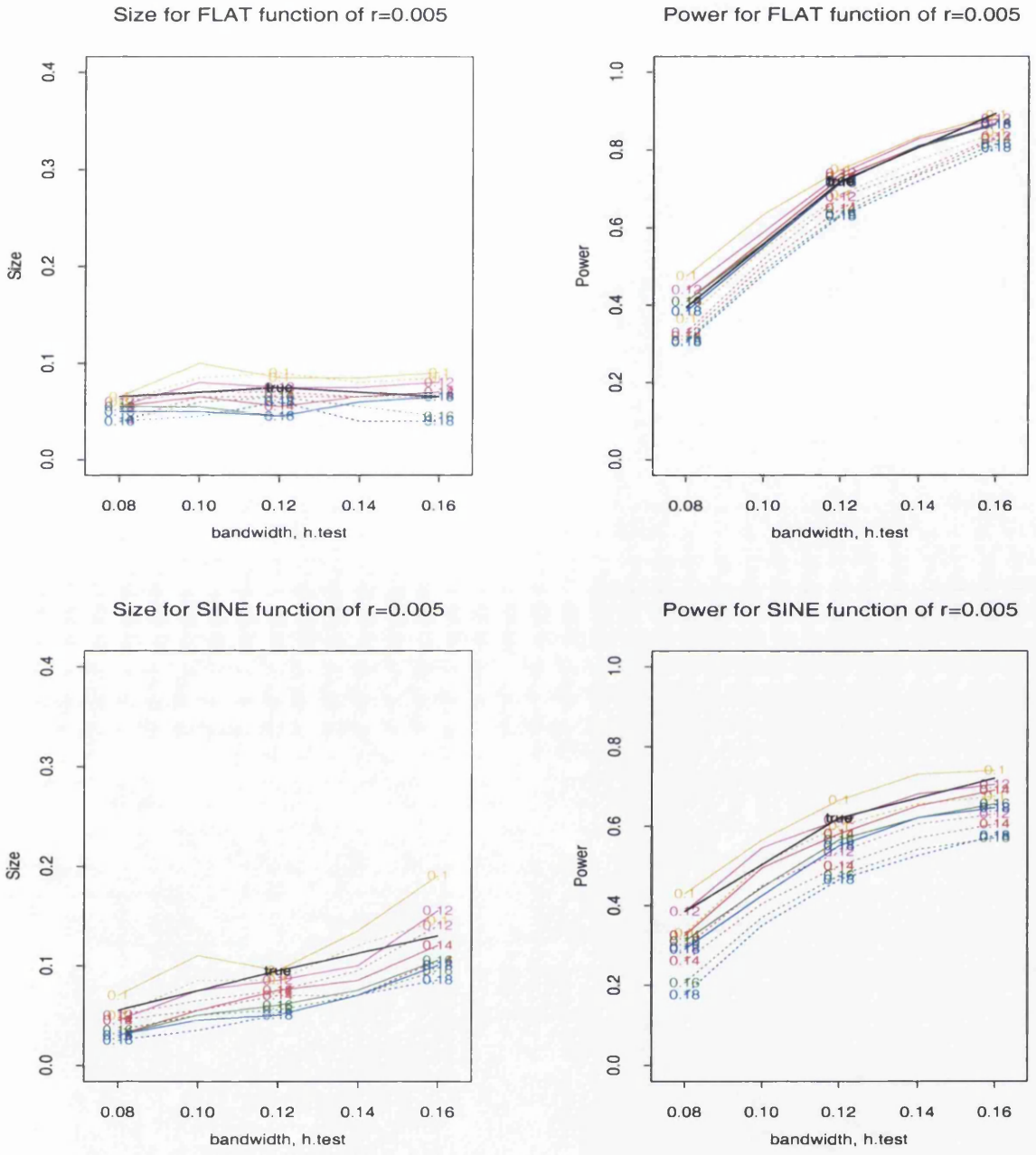

Figure 4.14. Size and Power of the double discontinuity test with estimated Σ using F_O (with Rice), with $r = 0.005$ and $jump = 3$. solid: lag 0.01, dotted: lag 0.02. The values on the curves denote the values of h_1 used for the 1st stage. The size and power using true Σ are represented by a black, solid lines.

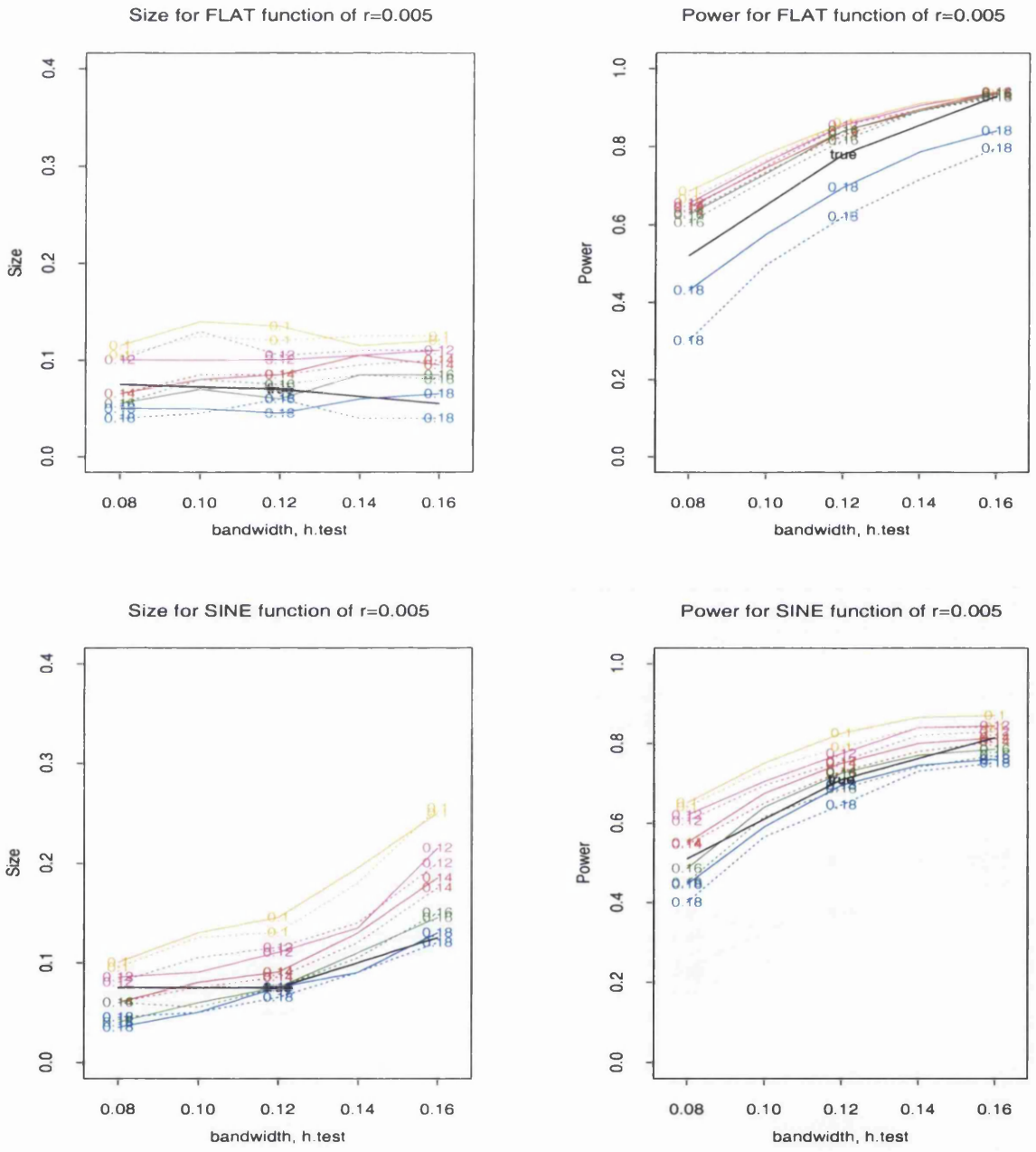

Figure 4.15. Size and Power of the double discontinuity test with estimated Σ and sill using F_N, with $r = 0.005$ and $jump = 3$. solid: lag 0.01, dotted: lag 0.02. The values on the curves denote the values of h_1 used for the 1st stage. The size and power using true $\sigma^2\Sigma$ are represented by black, solid lines.

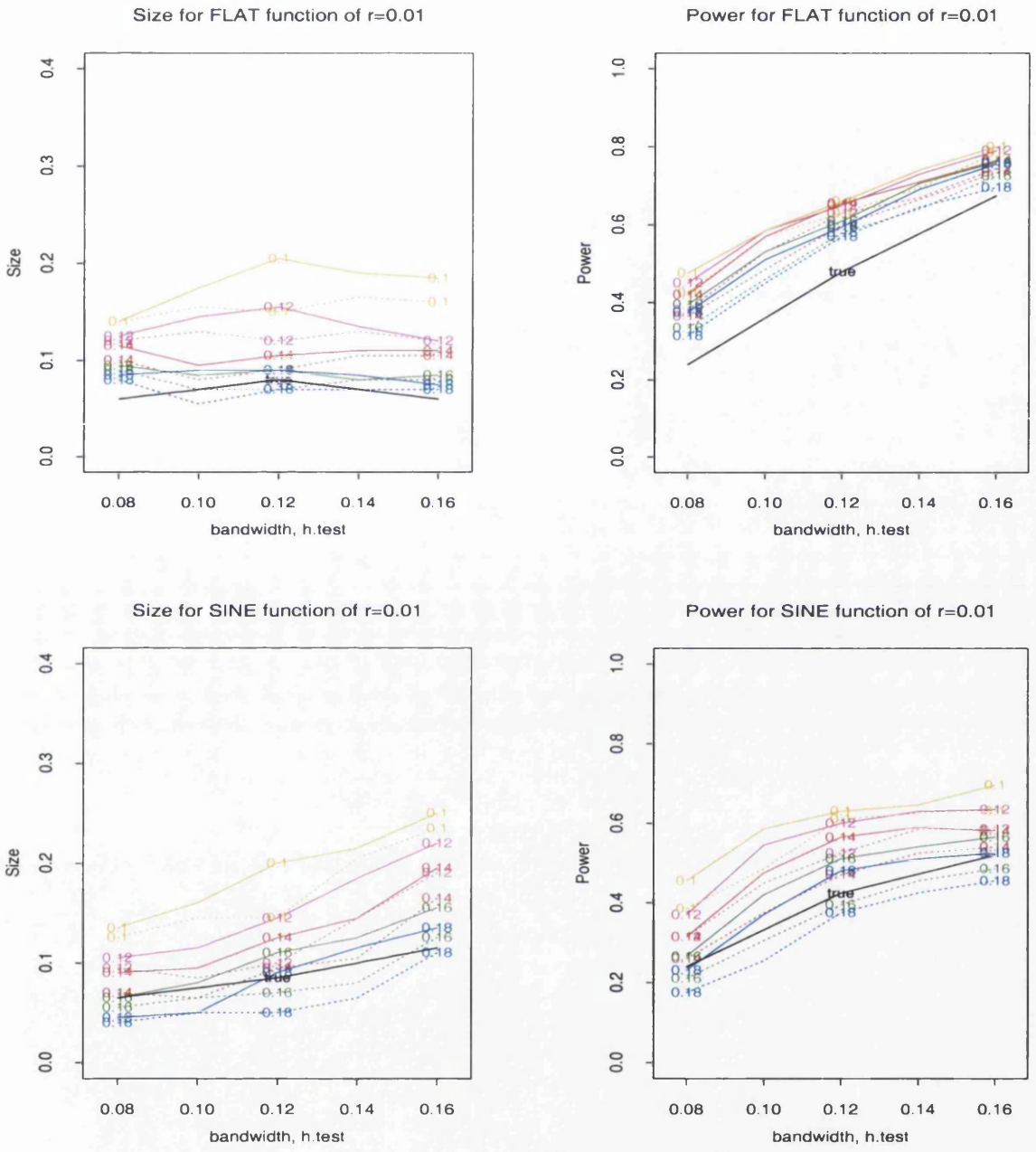

Figure 4.16. Size and Power of the double discontinuity test with estimated Σ using F_O (with Rice), with $r = 0.01$ and $jump = 3$. solid: lag 0.01, dotted: lag 0.02. The values on the curves denote the values of h_1 used for the 1st stage. The size and power using true Σ are represented by black, solid lines.

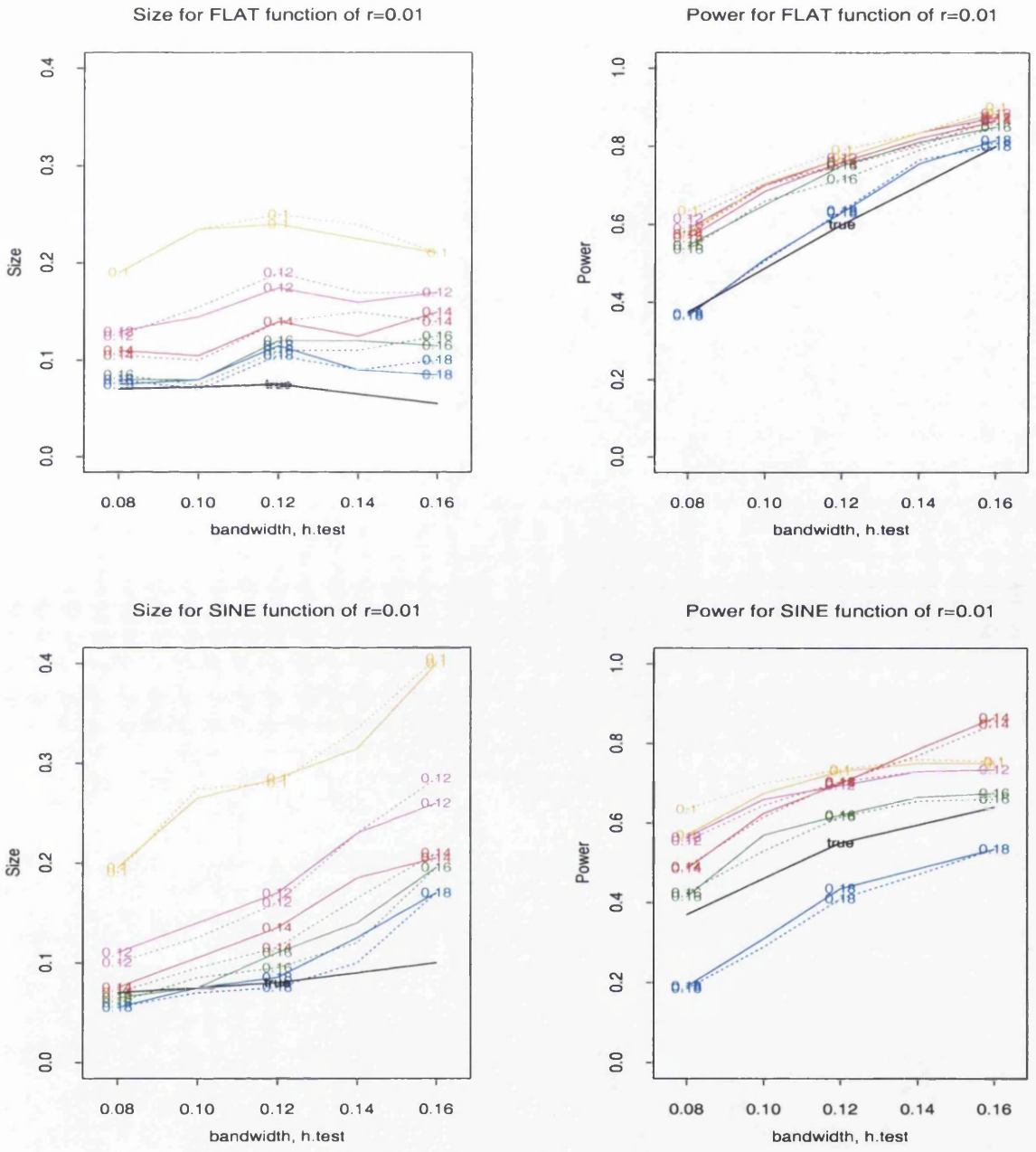

Figure 4.17. Size and Power of the double discontinuity test with estimated Σ and sill using F_N, with $r = 0.01$ and $jump = 3$. solid: lag 0.01, dotted: lag 0.02. The values on the curves denote the values of h_1 used for the 1st stage. The size and power using true $\sigma^2\Sigma$ are represented by a black, solid lines.

0.18 is needed to ensure a good size for both the flat and sine trends. This value gives a power that is close to the true power curve.

In summary, the results have shown that the double discontinuity test approach returns reasonably well-calibrated sizes for a range of smoothing parameters. The clear advantage over the residual-variogram approach is that though a slightly larger h_1 value is required for both techniques when the dependence of the data is strong, it is generally less sensitive to the choice of smoothing parameter used in the 1st stage to estimate the correlation. Therefore, it reduces the need of choosing an appropriate but unknown parameter from the procedure. Using a larger h_1 for the first stage only decreases the power of the test slightly, but might improve the size of the test considerably. Hence it is recommended that a slightly larger smoothing parameter, h_1, is used for the first stage, especially if one suspects that the data may be highly correlated.

Another obvious advantage is that the power of the double-discontinuity test for both low and moderate ranges that are considered, not only outperforms the residual-variogram approach but is comparable and sometimes higher than the power curves which have the true plugged-in correlation/covariance matrix.

Though F_N performs better than F_O using certain h_1 values, it is less stable than F_O. This can be due to the fact that it involves the plug-in estimate of both the range and sill of the variogram, while F_O incorporates the variation of the error variance using Rice estimator in its distributional calculations. This feature is also seen in the residual variogram approach.

4.5.3 Discussion

The key factor that causes the difference in the performance of the two techniques, the residual-variogram and the double discontinuity test, lies in their estimation of the range and sill of the variogram under the null and alternative hypothesis. Next, we shall investigate this difference by looking at the estimated parameters using these approaches. To allow for comparison, all the graphs for the same range are drawn on the same scale. As the features are quite similar for both ranges, we shall just focus our discussion on $r = 0.01$.

Figures 4.20 and 4.21 show the boxplots of the range and sill of the 200 simulated data

sets under H_0 and H_1 for $r = 0.01$, using the residual-variogram approach for the flat and sine trends respectively, with different *h.trend* and lags 0.01 and 0.02. Similarly, Figures 4.24 and 4.25 display the boxplots using the double discontinuity test. The first number on the symbol "$a|b$" represents a smoothing parameter of $a \times 10^{-2}$ for the first stage, and the second number represents the lag or bin width, $b \times 10^{-2}$, where a and b are constants.

There are not many differences in the features observed from the estimated range and sill of the variogram under the two different trends, flat and sine, hence we shall summarise them together. Table 4.3 presents the summary results for the residual-variogram approach, while Table 4.4 presents that for the double discontinuity test.

From the summary tables, one clear reason why the residual-variogram test performs worse than the double discontinuity test, is because at well-calibrated sizes, where the range is well-estimated under H_0, it is often overestimated under H_1. Hence the power is low, and in fact, it is often lower than the power with the true plugged-in range. On the other hand, the range for the double discontinuity test is fairly similar for the model under H_0 and H_1, where its median is slightly lower than the true range of 0.01 over all h_1 considered. Hence, this helps the double discontinuity test to have comparable powers as that of the true power.

	Model under H_0	**Model under H_1**
Range	mostly underestimated at low *h.trend*, but increases marginally as *h.trend* increases	mostly underestimated at how *h.trend*, but as *h.trend* increases, range is over-estimated. Genearally higher than H_0.
	lag 0.02 is slightly higher than lag 0.01	lag 0.02 is slightly higher than lag 0.01
	variability of estimate stays fairly constant.	variability increases as *h.trend* increases.
Sill	slightly underestimated but increases , marginally as *h.trend* increases	well-estimated at low *h.trend*, but increases and is overestimated at high *h.trend* values. Generally higher than H_0.
	lag 0.02 is slightly lower than lag 0.01	lag 0.02 is slightly lower than lag 0.01
	variability of estimate stays fairly constant.	variability of estimate increases as *h.trend* increases

Table 4.3. Summary of estimated range and sill using the residual variogram approach for both the flat and sine trends.

	Model under H_0	Model under H_1
Range	mostly underestimated at low h_1, but increases marginally as h_1 increases	mostly underestimated at low h_1, but improves marginally as h_1 increases. The estimate is approximately the same as under H_0
	lag 0.02 is slightly higher than lag 0.01	lag 0.02 is slightly higher than lag 0.01
	variability of estimate stays fairly constant.	variability is similar to H_0 and stays fairly constant
Sill	slightly underestimated thoroughout range of h_1	slightly underestimated throughtout range of h_1. The estimate is approximately the same as those under H_0.
	lag 0.02 is slightly lower than lag 0.01	lag 0.02 is slightly lower than lag 0.01
	variability of estimate stays fairly constant.	variability is similar to H_0 and stays fairly constant

Table 4.4. Summary of estimated range and sill using the double discontinuity test approach for both the flat and sine trends.

Moreover, using the residual variogram approach, there is a substantial increase in the estimate of the range with increasing *h.trend*, while there is only marginal increase for the double discontinuity approach with h_1. The variability of the estimate also increases with increasing *h.trend* for the former, but is fairly similar for the latter approach with increasing h_1. This explains why the double discontinuity test is less sensitive to the smoothing parameter used in the first stage.

Another feature to highlight is that lag 0.02 tends to give a larger estimate of range, but a lower estimate of sill, than lag 0.01. Consequently, this results in a slightly lower power for lag 0.02 than lag 0.01 for both algorithms when F_O is used. On the other hand, the difference in using these two lags is negligible for F_N.

We noted in our previous discussion that F_N somehow tends to give a higher size compared to F_O, using the same smoothing parameter. From the boxplots, it can be observed that when too low a value of *h.trend* is used, both the range and sill are greatly underestimated. This will affect F_N more as it considers both plug-in estimates of the range and sill.

As the F_O performs better overall, we have also investigated how the two algorithms perform with a smaller jump of 2 using this test statistic. The size remains the same as discussed earlier. Figures 4.26 and 4.27 display the size and power for the residual-variogram and DDT algorithms with a jump of 2. The power with well controlled size for the residual-variogram is much lower than the true power, but for the DDT, the power is comparable to the true power.

Though the double discontinuity test is computationally more intensive than the residual variogram approach, it has certainly proven itself to be a much superior technique to detect discontinuity in correlated data. It is not only robust against jumps and trends, but also capable of producing powers that are comparable or higher than the power obtained using the true range. Furthermore, it has the favourable property of not being too dependent on the smoothing parameter used in the first stage.

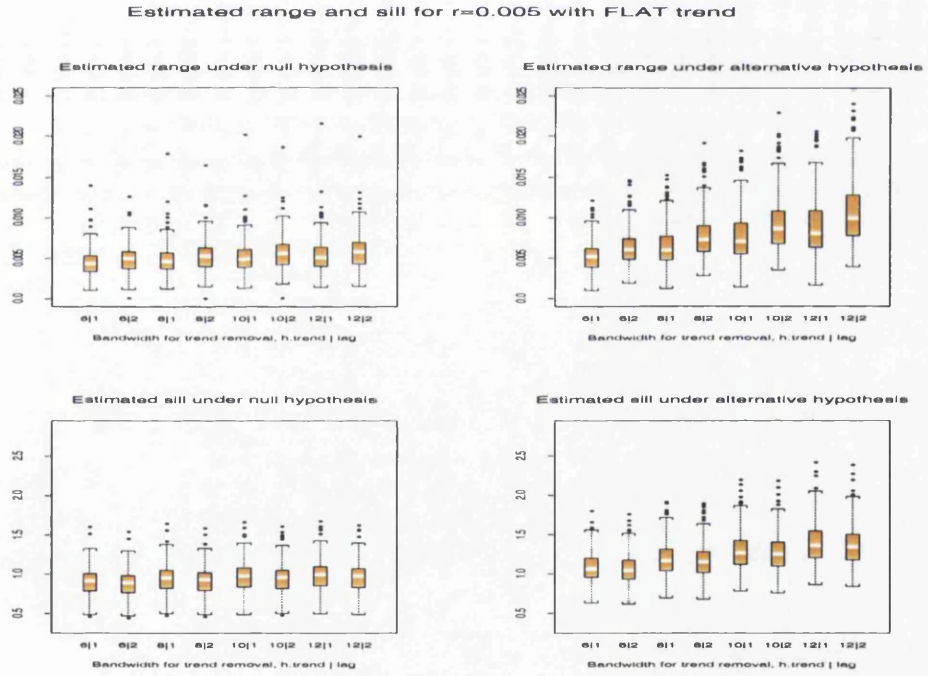

Figure 4.18. Boxplots of estimated range and sill using residual-variogram approach with a flat trend for $r = 0.005$.

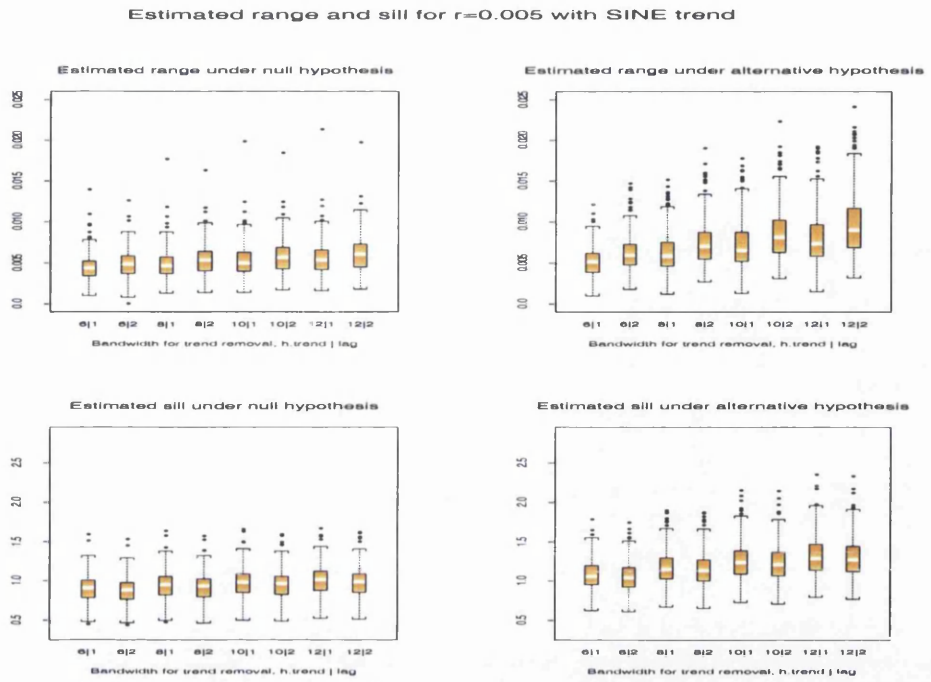

Figure 4.19. Boxplots of estimated range and sill using residual-variogram approach with a sine trend for $r = 0.005$.

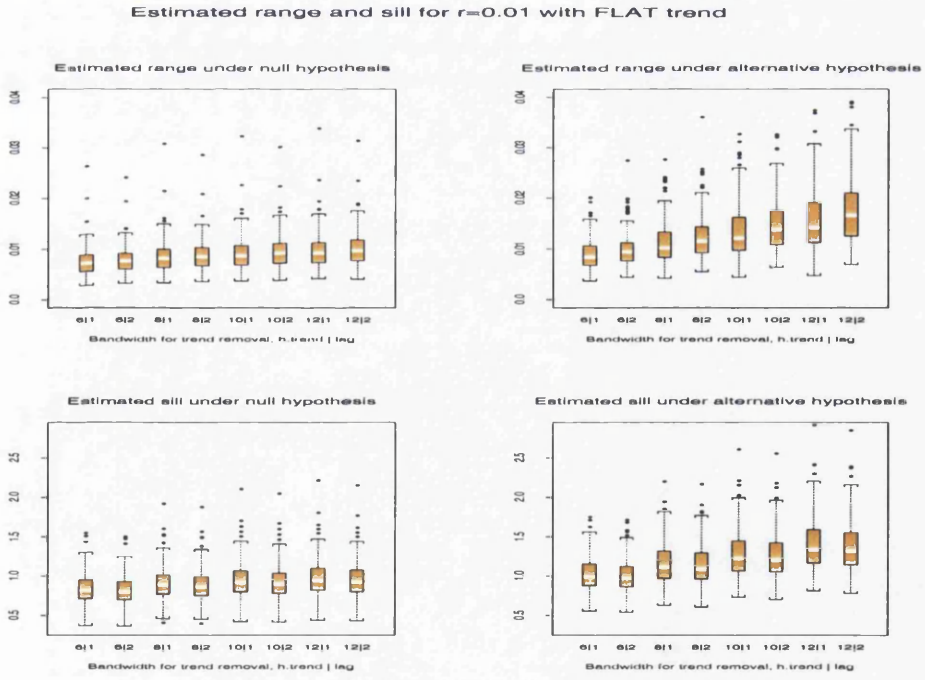

Figure 4.20. Boxplots of estimated range and sill using residual-variogram approach with a flat trend for $r = 0.01$.

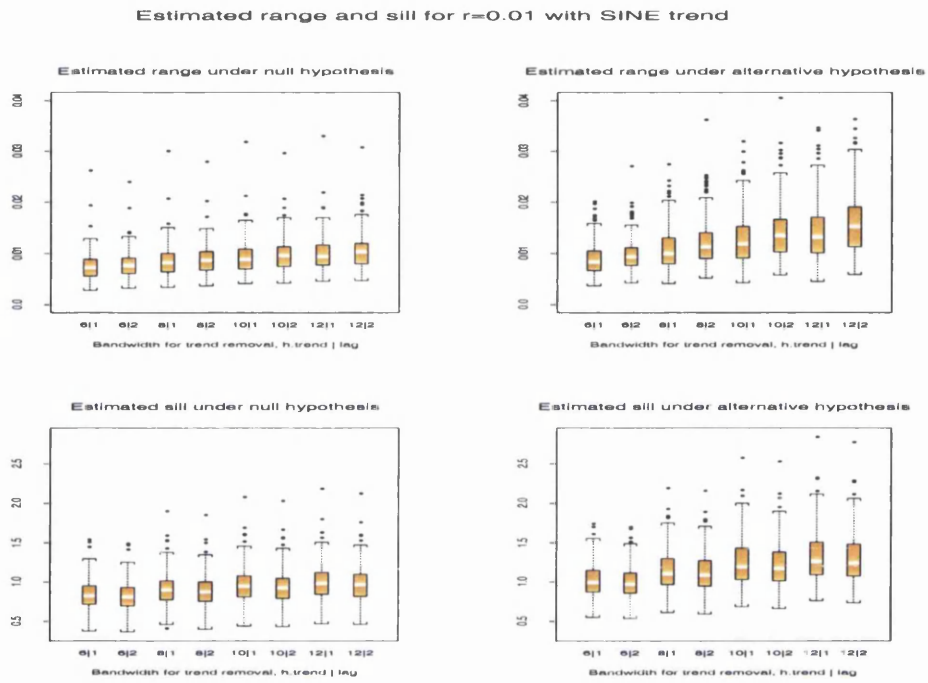

Figure 4.21. Boxplots of estimated range and sill using residual-variogram approach with a sine trend for $r = 0.01$.

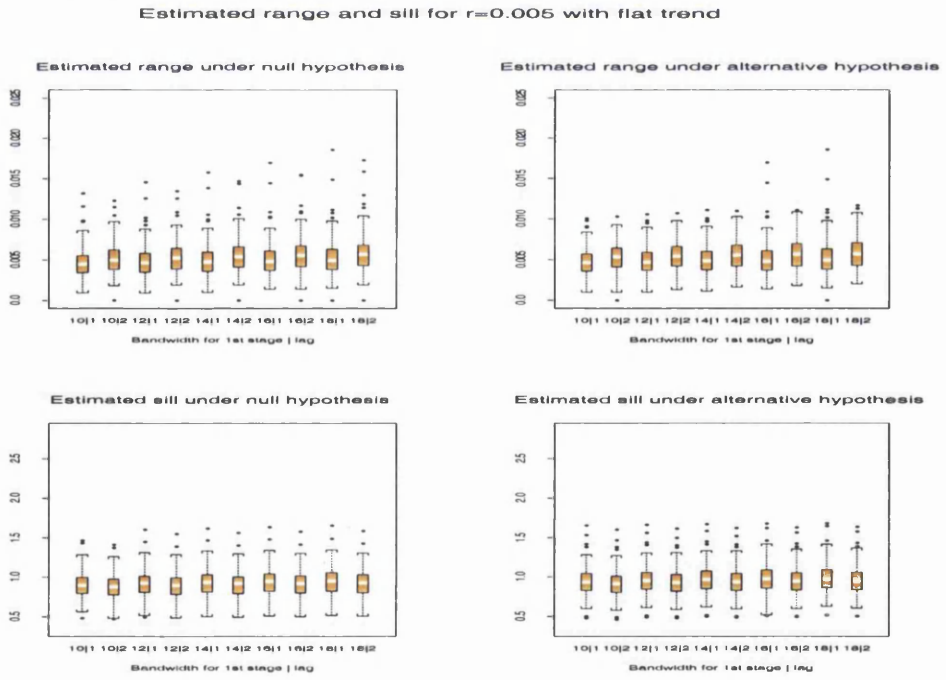

Figure 4.22. Boxplots of estimated range and sill using double discontinuity test with a flat trend for $r = 0.005$.

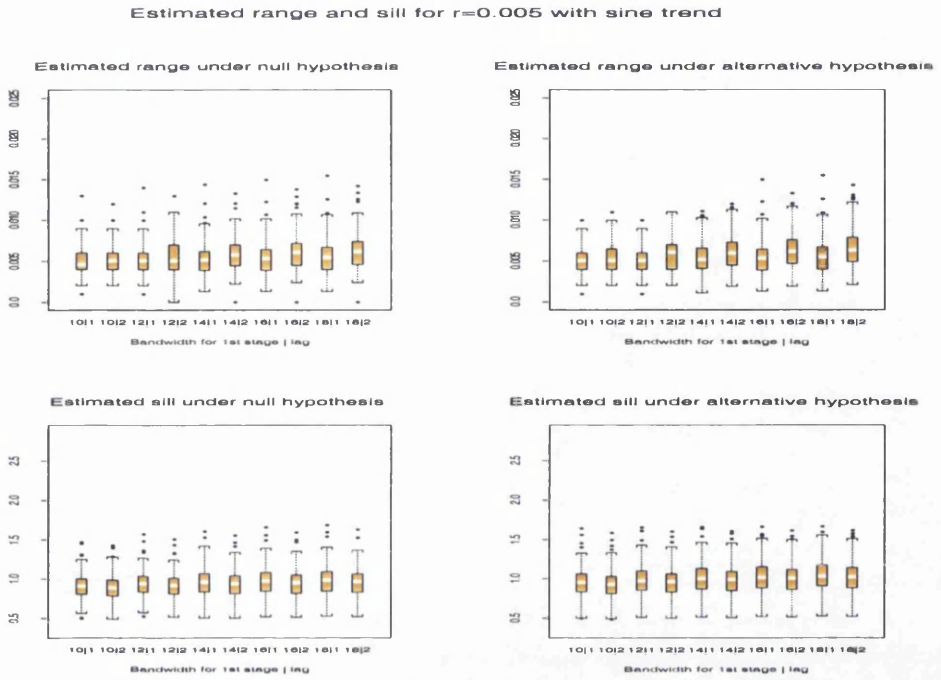

Figure 4.23. Boxplots of estimated range and sill using double discontinuity test with a sine trend for $r = 0.005$.

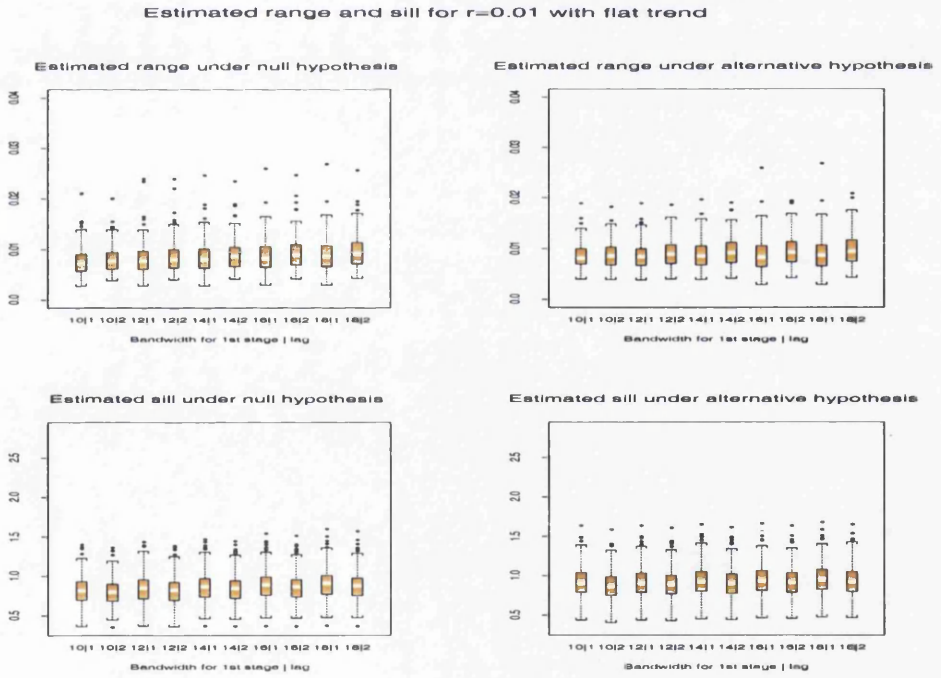

Figure 4.24. Boxplots of estimated range and sill using double discontinuity test with a flat trend for $r = 0.01$.

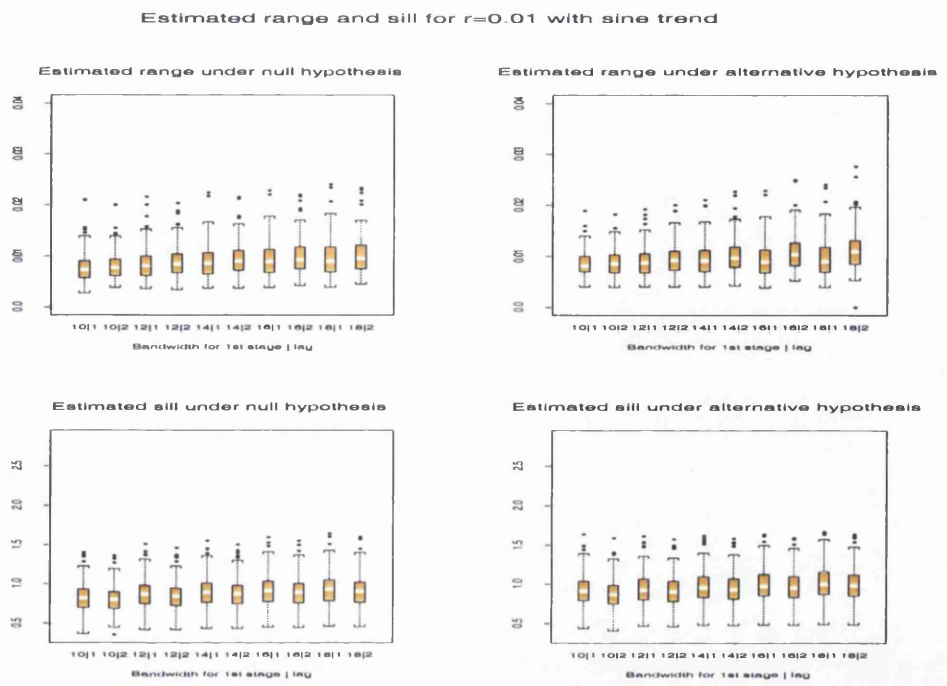

Figure 4.25. Boxplots of estimated range and sill using double discontinuity test with a sine trend for $r = 0.01$.

4.6 Different sets of unequally spaced locations

In all the previous sections, we have only looked at data that are simulated over a single set of unequally spaced locations, denoted here as set 1. In this section, we will investigate whether there is any difference in the size and power if different sets of unequally spaced locations are used.

Four other sets of locations are simulated and denoted as set 2 to set 5. Figure 4.28 displays all five different sets, with "m" indicating the median of the data.

A simulation study is carried out using $r = \{0.005, 0.01, 0.02\}$ for each of these sets over a range of *h.test*, assuming that r is known. The results using F_O are displayed in Figure 4.29. The size curves are similar for the different sets. However, power curves (with $jump = 3$) appear to have two "clusters"', with set 5 being the highest, followed by set 1 and then set 4. The other cluster consists of set 2 and 3, where power of set 3 is slightly lower.

The most probable explanation for the differing powers in these 5 sets lies in the vicinity of the change-point. From Figure 4.28, it can be observed that the neighbourhood where the

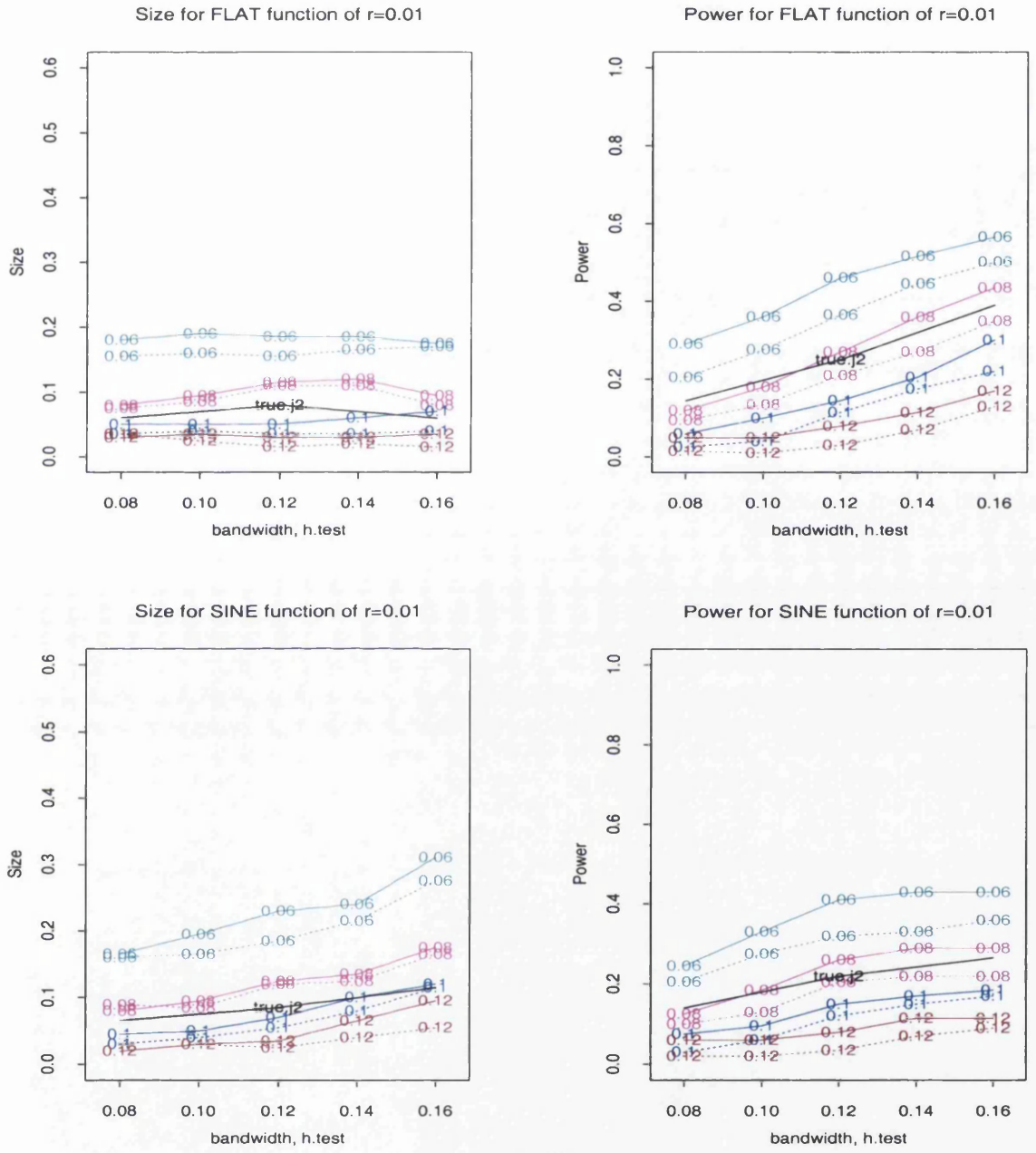

Figure 4.26. Size and Power of the residual-variogram approach for estimated Σ using F_O (with Rice) for $r = 0.01$ and $jump = 2$. solid: lag 0.01, dotted: lag 0.02. The values on the curves denote the values of *h.trend*. The size and power using true Σ are represented by black, solid lines.

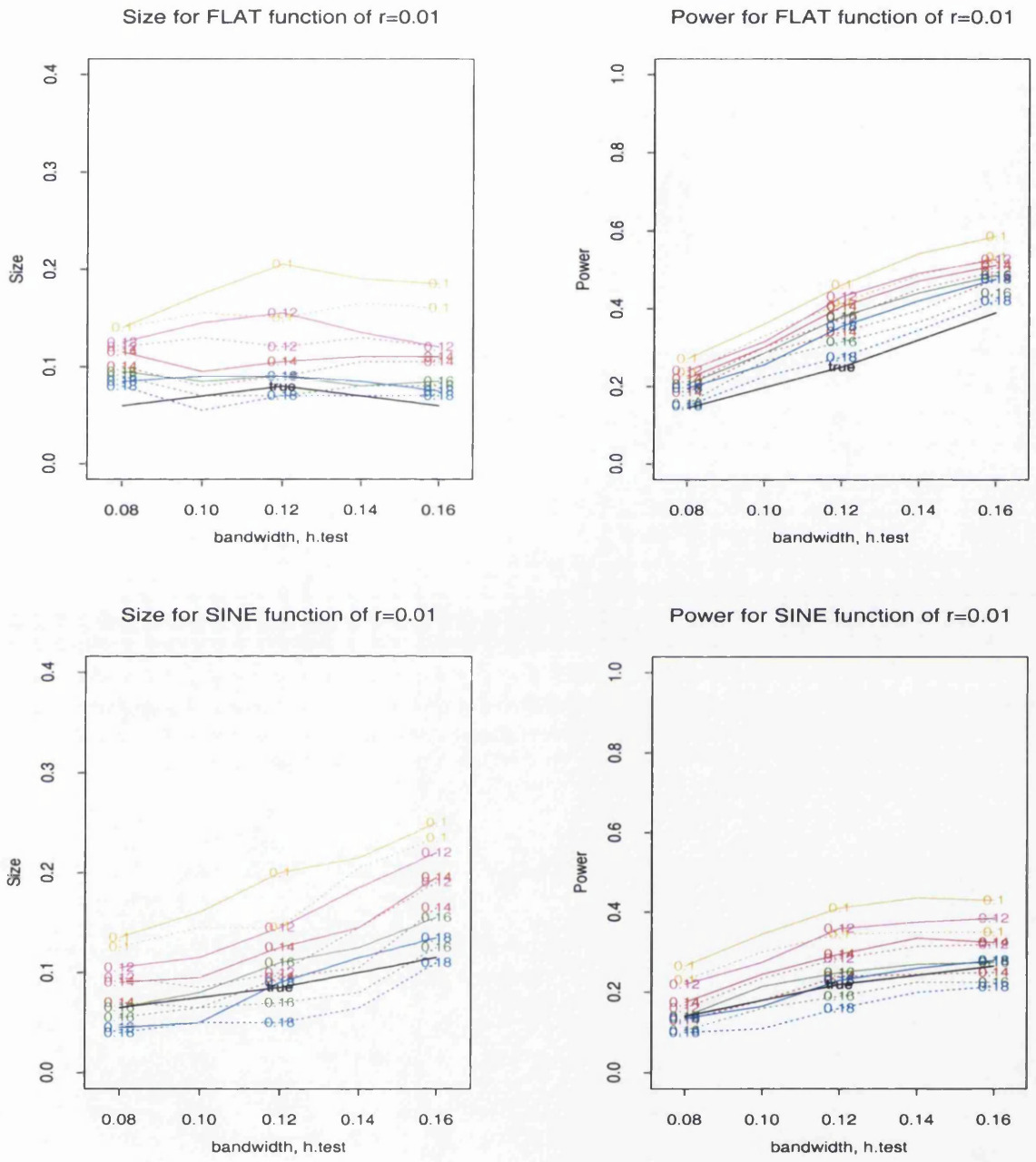

Figure 4.27. Size and Power of the double discontinuity test with estimated Σ and sill using F_O, with $r = 0.01$ and $jump = 2$. solid: lag 0.01, dotted: lag 0.02. The values on the curves denote the values of h_1 used for the 1st stage. The size and power using true Σ are represented by black, solid lines.

location of the change-point is located, is sparse for both sets 2 and 3. The neighbourhood for the rest, however, are less sparse, hence the power of detecting a discontinuity is higher. This is because our test statistic consists of the sum of the standardised differences between the left and right smooths. If there are more points located near a change-point, then the increase in the observed test statistic will be greater, as the standardised differences between the left and right smooths will be higher. Hence this improves the performance of the test.

4.7 Concluding Remarks

This chapter brings to an end the development and the study of the discontinuity test in the one-dimensional setting. We started off by developing the test statistic that is suited for an unequally spaced design, and introduced the variogram as a tool to describe the dependence structure of the data in such a setting.

The two test statistics introduced here are F_O and F_N. The former is the same as that used in the equally spaced design, where the error variance is unknown and is incorporated into the test statistic, by placing it in quadratic form. Since fitting a theoretical (exponential)

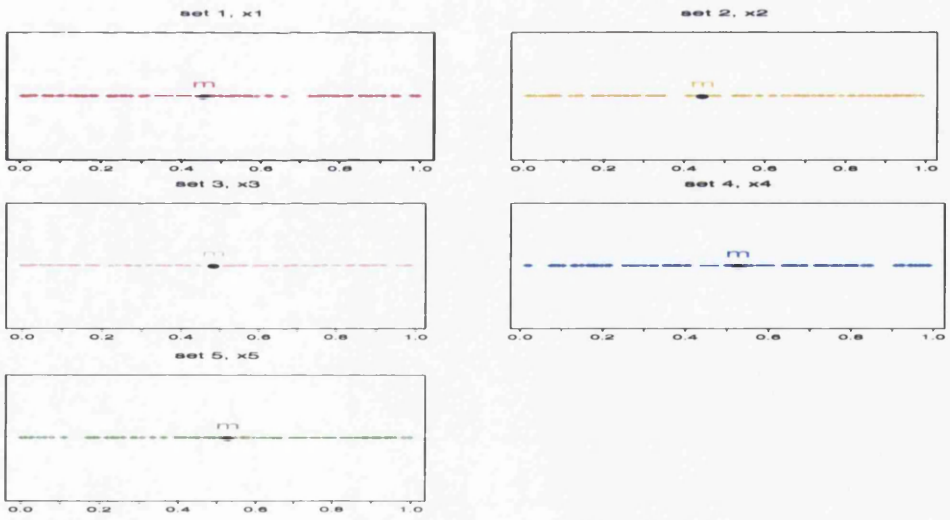

Figure 4.28. Plots of 5 different sets of unequally spaced observations, $x1$, $x2$, $x3$, $x4$ and $x5$. The letter 'm' denotes the position of the median of the data.

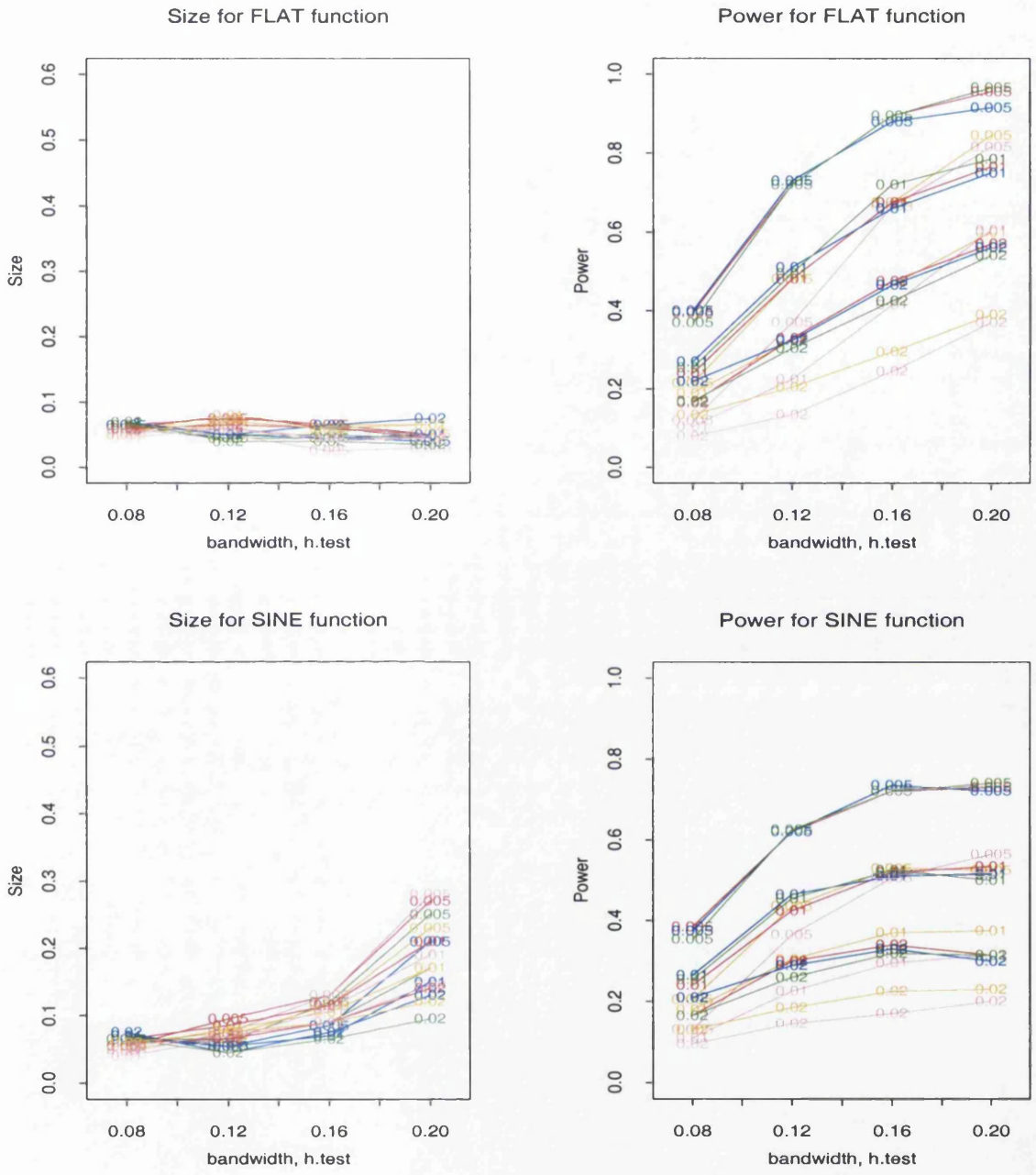

Figure 4.29. Effects of different sets of unequally spaced locations x. Size and Power for four different sets of locations with $r = \{0.005, 0.01, 0.02\}$. Set 1 (original): red, Set 2: orange; set 3: pink; set 4: blue; set 5: green.

variogram model gives an estimate of both the range (strength of dependence) as well as the sill (error variance) of the data, F_N considers a plug-in estimate of error variance from the sill of the variogram, which it then treats as the true value. Both the global and local tests were considered.

An extensive simulation was then carried out to study the effects of the different factors. It is also of interest to compare and contrast the performance of the two test statistics introduced here. The difference is mainly attributed to the different variance estimators and how they are incorporated into the test statistic.

It is not surprising that the effects of most of the factors, such as shape of trend function, smoothing parameter and strength of dependence, are similar to the equally spaced design. Rice generally performs better than Gasser over different ranges and smoothing parameters. and the difference between these two estimators increases as the range increases.

When the range is assumed known, but the error variance is unknown, the new test statistic performs worse than the original one. This is because it is very sensitive to the value of *h.trend* used to remove the trend first before obtaining the estimate of the sill from the variogram. Using too low a value will cause inflation of the test, while using a high value will cause a substantial decrease in power.

Two different techniques, namely the residual-variogram approach and the double discontinuity test, are proposed to estimate the dependence structure of the data. The latter adds an extra step to remove any potential jump present first before removing the trend. This is done by carrying out the discontinuity test in the first stage, treating the data as independent. This extra effort appears to pay off as the double discontinuity test performs better than the residual variogram approach. The results demonstrate that the former is less sensitive to the smoothing parameter used in the 1st stage. It gives better and, approximately, similar estimates of range and sill under H_0 and H_1, hence keeping the size well-calibrated and, at the same time, produces good power at appropriate smoothing parameters. Nonetheless, a larger smoothing parameter is required for the first stage to capture the dependence structure for the two approaches if the data are highly correlated.

It is not surprising that when the error variance is known, F_N performs better that F_O. However, when both the error variance and the correlation are unknown, F_O is recommended

as it is more stable. This is because F_N is more sensitive to the smoothing parameter used in the first stage as the latter not only influences the estimate of the strength of dependence structure (the range) but also the error variance (the sill). Hence if an unsuitable smoothing parameter is used, F_N will be more affected than F_O, which is only influenced by the estimation of the range.

In the last section, the size and power properties for other sets of unequally spaced observations are investigated. The simulation results using true ranges, revealed that while the sizes are similar, the power properties are dependent on the vicinity where the change-point is located. If the data are quite sparse near the change-point, the power might be lower. It is expected that the effects of other factors discussed earlier would also apply for different sets of unequally spaced observations, though the magnitude of the power might differ.

In the following chapter, the discontinuity test is applied to various real-life data sets that are collected over time, to examine if there are any discontinuities present.

Chapter 5

Applications

In this chapter, we will be looking at three sets of data: rainfall data in Argentina in Section 5.2, data on global warming in Section 5.3 and water quality data in the River Clyde in Section 5.4. The first two data sets have been analysed by Wu et al. (2001), while the third data has not yet been published. We will investigate if there are any abrupt changes in these three sets of data using the isotonic regression approach (Wu et al., 2001) and a Bayesian approach (Thomas, 2001), and compare them with the nonparametric regression approach that we have proposed in Chapters 3 and 4. The former two tests are selected as they are representatives of some of the change-point detection approaches that have been developed to cope with correlated data.

The Argentina and Global Warming case studies are both *equally-spaced*, whilst the River Clyde case study is *unequally spaced*. The isotonic regression and Bayesian approaches are developed to cope with equally-spaced data only. The author's proposed nonparametric regression approach, however, is developed to cope with both equally and unequally spaced observations.

Both the isotonic regression and nonparametric regression approaches allow for a nonparametric trend in the data, but in the former, only a nondecreasing step function is allowed for the mean trend while the latter allows for a smoothly varying mean. The Bayesian approach is more restricted as it assumes a constant mean. However, while the variance and autocorrelation structure of the data are assumed to remain constant throughout the study

for our nonparametric test and that by isotonic regression, the Bayesian approach allows for a change in all three: variance, autocorrelation or mean.

All three approaches assume normality of errors which might not be completely satisfied in practice. The assumptions will be checked to see if they are valid. If necessary, transformation of the data can be considered for the normality assumption to be valid.

5.1 Different approaches to change-point detection

In this section, we discuss the different features of the three change-point algorithms, namely isotonic regression (Wu et al., 2001), a Bayesian approach (Thomas, 2001) and nonparametric regression.

5.1.1 Features of the Isotonic Regression Approach

Let us first consider Wu et al's (2001) isotonic regression test, which is a test capable of detecting the existence of monotonic trends and/or abrupt changes in the mean level, in short range dependent sequences.

Isotonic regression is a nonparametric method appropriately used when a dependent response variable is *monotonically* related to an independent explanatory variable. The fitted model consists of level sets of increasing value, for which the estimated response is constant. The model used is similar to the nonparametric regression model, except that the means μ_t are modelled using an isotonic regression function.

The model is assumed to be

$$y_t = \mu_t + \varepsilon_t, \qquad t = 1, \ldots, n, \tag{5.1}$$

where ε_t is normally distributed with mean 0 and finite covariance $\gamma(t) = cov(\varepsilon_i, \varepsilon_{i+t}) = E(\varepsilon_0 \varepsilon_t)$. Suppose that the means are ordered $\mu_1 \leq \mu_2 \leq \ldots \leq \mu_n$ and consider testing the hypothesis $H_0 : \mu_1 = \mu_2 = \ldots = \mu_n$. This is the null hypothesis encountered in many change point problems, but the alternative is different. The latter can be due to the presence of a monotonic trend, or abrupt monotonic changes.

Wu et al. (2001) considered the ***unpenalised*** likelihood ratio test of H_0, and also developed the ***penalised*** likelihood ratio test of H_0 and its asymptotic distribution. The asymptotic distribution is derived under the more general assumption that the errors are part of a zero mean, square integrable, stationary ergodic process that exhibits suitable short range dependence.

The motivation to develop the penalised likelihood ratio test was to cope with the *"spiking"* problem in large samples, which might have a great influence on the test statistic, when the first isotonic mean $\hat{\mu}_1$ is too small and the last isotonic mean, $\hat{\mu}_n$ is too large. Wu et al. (2001) introduced a penalised term to rectify this problem. This is done by increasing the value of the first observation and decreasing the value of the last observation. This gives rise to the modified test statistic known as the **penalised log likelihood ratio test statistic** (see Wu et al., 2001, Equation (8)).

To estimate the error variance, the trend is first removed by taking the difference between the raw data and the isotonic unpenalised/penalised means, $\hat{\mu}_t$ to give the residuals, $\hat{\varepsilon}_t = y_t - \hat{\mu}_t$. Then using lag window estimators (Brockwell and Davis, 1991), σ^2 can be estimated as

$$\hat{\sigma}^2 = \hat{\gamma}(0) + 2\sum_{k=1}^{m_n} w\left(\frac{k}{m_n}\right)\hat{\gamma}(k); \tag{5.2}$$

(see Wu et al., 2001, Equation (19)), where the lag window $w(x)$ is an even, piecewise continuous function of x satisfying the conditions $w(0) = 1$, $w(x) = 0$ for $|x| \geq 1$ and $|w(x)| \leq 1$ for all x; and $\hat{\gamma}(k)$ is the sample autocovariance function, obtained by $\hat{\gamma}(k) = \frac{1}{n}\sum_{i=1}^{n-k}\hat{\varepsilon}_i\hat{\varepsilon}_{i+k}$. . The size of the truncated window is controlled by m_n, which is recommended to be $m_n = n^{1/3}$. We will use this same value for all the sets of data that we will look at in the rest of this chapter. (We have seen earlier in the brief simulation study in Section 3.6, that the isotonic regression test is sensitive to the estimation of this error variance).

As the asymptotic null distribution of the penalised ratio test statistic is complicated, the percentiles of the asymptotic null distribution of the test statistic is computed by simulating 30,000 Brownian motions instead. This is provided in Table 1 of Wu et al. (2001), which we will refer to in our analysis in later sections.

Taking order restrictions into account makes the inference carried out by using isotonic regression favourable, as it is a one-sided test which is certainly more powerful than a two-sided test. However, great caution has to be taken as the hypothesized order restriction may not necessarily hold.

5.1.2 Features of the Bayesian approach

The next method that we shall look at is Thomas' Bayesian approach. One of the main differences between the frequentist and the Bayesian approaches is that the Bayesians view unknown quantities as random parameters and assign prior distributions to them, while the frequentists view the parameters as fixed (though unknown) and not random. For instance, in the Bayesian approach to change-point analysis, the change-point locations are random and probability distributions are assigned to them.

Thomas' (2001) approach is developed to cope not only with abrupt changes in the mean level, but also possible combinations of abrupt changes in the variance and autocorrelation structure of the data. His model is assumed to be piecewise stationary, and the presence of a change-point indicates a switch to a different process, which we can refer to as an abrupt change in the mean level, variance level and/or autocorrelation structure. The motivating application of his work is in the field of road maintenance; to partition roads into homogeneous parts.

The model used by Thomas is

$$y_t = \alpha + \beta y_{t-1} + \varepsilon_t, \qquad t = 1, \ldots, n, \tag{5.3}$$

where α and β are unknown parameters; with α representing a constant term and β denoting the correlation coefficient of an AR(1) model with $|\beta| < 1$. The ε_t are independent and normally distributed errors, with mean 0 and variance σ^2, i.e. $\varepsilon_t \sim N(0, \sigma^2)$.

Thomas' method addresses the question of whether an abrupt change is present somewhere in the data by evaluating the Bayes factor using improper priors. The indeterminacy of the Bayes factor due to comparisions of models with different dimensions (the dimension of a model without a change-point is lower than that which has a change-point) is resolved

using the minimal imaginary training sample approach, introduced by Smith and Spiegelhalter (1981), Speigelhalter and Smith (1982). Exact analytical solutions can be obtained, conditional on the starting values of the process, if the autocorrelation is known. In more practical situations where the autocorrelation structure and the starting values are unknown, exact inference is still possible using Markov Chain Monte Carlo techniques (see Thomas, 2001, Paper III). An alternative approach that he proposed is to use approximate inference which has the favourable property of being computationally faster and more efficient.

By making use of the fact that AR(1) processes are reversible, the approximate model with no change-point is given as

$$\begin{aligned} & y_t = \alpha + \beta y_{t-1} + \varepsilon_t, && t = 2, \ldots, n-1, \\ \text{or} \quad & y_t = \alpha + \beta y_{t+1} + \varepsilon_t, && t = 2, \ldots, n-1 \end{aligned} \tag{5.4}$$

and the approximate model with change-point r, is given as

$$\begin{aligned} y_t &= \alpha_1 + \beta_1 y_{t-1} + \varepsilon_t, && t = 2, \ldots, r, \\ y_t &= \alpha_2 + \beta_2 y_{t+1} + \varepsilon_t, && t = r+1, \ldots, n-1 \end{aligned} \tag{5.5}$$

Noninformative priors are used for the unknown parameters. The overall evidence of a change somewhere in the series can then be computed by comparing the evidence pooled from all the models placing one change-point in the model with no change-point present. This is denoted as $p(change)$, the posterior probability of a change. No rules have been set as to when $p(change)$ is high enough to conclude that a change-point exists. However a posterior probability of greater than 0.7 might be sufficient since it implies that there is stronger evidence of a single change in the series, compared to a probability of 0.3 of no change.

If there is sufficient evidence to indicate a change has occurred, the next step will be to locate the change-point. The posterior distribution of the possible locations of a change-point conditional on the assumption that it exists, allows inference about the probable locations of the change-point. The posterior mode can be taken as a natural estimate of the possible

change-point location.

Thomas's model assumes that the trend is a constant. That results in one of the disadvantages of this approach since if the data do possess a trend or are not reasonably smooth with well-defined abrupt changes in level, this method will not perform well. This is because it assumes that the segments are homogeneous and hence exchangeable, but this is not the case above. The method gives very different results too when different segments of the whole data are subjected to analysis, as each observation contains information about the existence and location of a change-point. Thomas commented too that the approach might not be numerically stable in situations where the datasets are extremely large, with very large or small values.

The assumption that the trend is constant might not be feasible for all three of our datasets. Nevertheless, we will carry out his test on the raw dataset (where no trend is removed), as well as on the residuals obtained after a nonparametric smooth curve is use to remove the trend. It is not surprising that the latter might affect the value of $p(change)$. In situations where $p(change)$ is lowered, this could be due to the trend removal, as performing the latter might conceal the very location and impact the size of the jump, hence affecting the discontinuity test.

Another possible setback to Thomas' approach is that it assumes that there is at most one change-point in the model and hence might not be able to cope with data that have several change-points. In situations where numerous change-points exist, the overall evidence of a change, $p(change)$ might be quite unpredictable, where it either increases or decreases in value. For instance, there can be cases whereby the value of $p(change)$ is low when several change-points are actually present in the data. (see Thomas, 2001, Paper IV)

Two different iterative *ad hoc* procedures for carrying out the algorithm can be considered. The first procedure is carried out by analysing the full dataset and, if it is justifiable to insert a change-point, the data are partitioned into two separate parts and reanalysed to investigate if there are any further change-points. In cases where there might be multiple change-points in a considerably large dataset, Thomas suggested Procedure 2, which uses overlapping segments. If $p(change)$ is high in a segment, it is then partitioned and checked for further change-points. The reason for using the overlapping segments is to avoid finding

no change-point when there are in fact several present. We will carry out both procedures in our analysis.

5.1.3 Features of the Nonparametric Regression Approach

This methodology has been introduced and described in great detail in Chapters 3 and 4 and simulation studies have been carried out extensively on both the equally and unequally spaced settings. In this section, we will describe in further detail the different features and steps of our algorithm, informed by the results of the simulation study.

The model for both the former two methods under the null hypothesis is that the mean is constant. Wu's test has the alternative that there is a monotonically increasing trend and/or an abrupt change occurring in the data. Thomas' test has as alternative that there is an abrupt change present. Our algorithm tests the alternative that there is at least one abrupt change, in the presence of a smoothly varying trend and correlated data.

Firstly, before carrying out the discontinuity test, we should check if the data can be assumed to be independent. This will make the algorithm much simpler, as we will not need to estimate the underlying correlation structure of the data. For the equally spaced setting, this can be done simply by fitting a smooth trend to obtain the residuals. From the ACF plot of the residuals, we will be able to deduce if the independence assumption is reasonable. In the case of unequally spaced data, the test of constant variogram devised by Diblasi and Bowman (2001) can be used to examine the independence assumption. This is in fact applicable for both equally and unequally spaced data. As the constant variogram test assumes a linear trend, we first have to remove any underlying smooth trend present in the data. This can be done by fitting a nonparametric smooth curve to the data. There are four parameters in this algorithm, namely the bandwidth to remove the trend in the data, *h.trend*; the distances at which the empirical variogram is computed, *maxdist*; the number of bins to group the pairwise differences, *nbins*, and the bandwidth to use to test for a constant variogram, *h.test*. The sensitivity of the test to the different parameters is considered. A test with p-value greater than 0.05 will indicate that there is insufficient evidence to reject the null hypothesis that there is no spatial dependence in the data. Reference bands can

be used as a graphical tool to give an indication of the nature of any spatial dependence or correlation. It is expected that under the null hypothesis, at every evaluated distance of the variogram, the smoothed variogram should be contained within the reference bands, with a probability of 95% at the 0.05 level.

If the independence assumption is feasible, the discontinuity test is then carried out, treating the data as independent. However, if it is not reasonable to do so, the correlation structure has to be estimated, as neglecting this aspect will increase the size of the test, resulting in the detection of false change-points.

The main difference in the algorithm for the equally and unequally spaced observations is in the way the correlation is estimated. For the equally spaced setting, to incorporate an estimated correlation, the moving window approach is used as it has been shown to outperform the residual approach in the simulation study in Chapter 3. The correlation structure is modelled using an AR(1) model. For the unequally spaced setting, both the residual variogram and the Double Discontinuity Test (DDT) algorithms are used. The correlation of the residuals is estimated via the variogram by fitting an exponential covariance structure. Since the correlation is a nuisance parameter, it is not crucial that we model it exactly. It seems reasonable to adopt a simple model of an AR(1) correlation structure or an exponential model as it is convenient and should capture most of the main structure.

If the test is significant, this indicates that there is at least one abrupt change in the presence of a smoothly varying trend. The sensitivity of the test to the different parameters involved (namely the size of moving window, b, the trend to remove the data, *h.trend*, and the bandwidth for the discontinuity test, *h.test*) is then investigated.

5.2 Application 1: The Argentina data

The volume of yearly rainfall in Argentina, collected from 1884 to 1996, was provided by Eng Cesar Lamelas, a meteorologist from the Agricultural Experimental Station Obispo Colombres, Tucumán. During 1952 to 1962, a dam was constructed, which is believed to have affected the volume of rainfall. It is thus of interest to investigate if there has been a change in the mean rainfall volume over this period. This analysis has been done previously

by Wu et al. (2001). The dataset is made up of 113 equally spaced observations.

5.2.1 Using the Isotonic Regression Approach

Figure 5.1 shows the volume of yearly Argentina rainfall, collected from 1884 to 1996, with both the unpenalised and penalised isotonic regression functions, using a penalty factor of 0.15 for the latter. The top left panel gives the raw data, which seems to show a decrease from 1880 to about 1950s and an increase after that. However, Wu's approach assumes that the mean trend is monotonically increasing (i.e. it ignores the possibility of any decrease).

The top and bottom right panels display the same plot but with fitted unpenalised and penalised functions respectively. The difference between the fitted unpenalised and penalised isotonic estimators is that the first and last values are slightly suppressed for the penalised estimators. It is noticeable that in 1884, the first observation is low, and this is suppressed in the penalised isotonic regression. Using the unpenalised and penalised likelihood ratio tests, the test statistic values are 21.27 and 18.0 respectively, which are both significant at the 0.05 level, indicating the presence of a change-point. From both plots, we can observe that there is an abrupt jump sometime between 1955 and 1956 in the annual rainfall volume.

The bottom left panel displays the ACF plot of the residuals after removal of trend by isotonic regression. Most of the ACF for the different lags are low and insignificant, hence indicating a very weak short range dependence. These residuals are used to estimate the error variance, taking into account the short range dependence in the data. This gives $\hat{\sigma}^2 = 253.57$.

The assumption of normality is checked in Figure 5.2. The raw data appears positively skewed. There is a slight indication that the detrended data is also slightly positively skewed, but an assumption of normality is reasonable.

Wu et al. (2001) claimed that the conclusion of the presence of an abrupt change at 1955 is in line with what the meteorologist expected, due to the construction of the dam.

5.2.2 Using the Bayesian Approach

We will first attempt to run the Bayesian change-point algorithm taking the trend as constant and modelling the correlation using an AR(1) model. The top panel of Figure 5.3 gives the volume of rainfall for each year over the whole period. The dotted line indicates where the most probable location of the change-point. This corresponds to the posterior mode (peak) of the posterior probabilities conditional on the presence of a change-point.

The posterior probability of an abrupt change occurring somewhere in the series, $p(change)$,

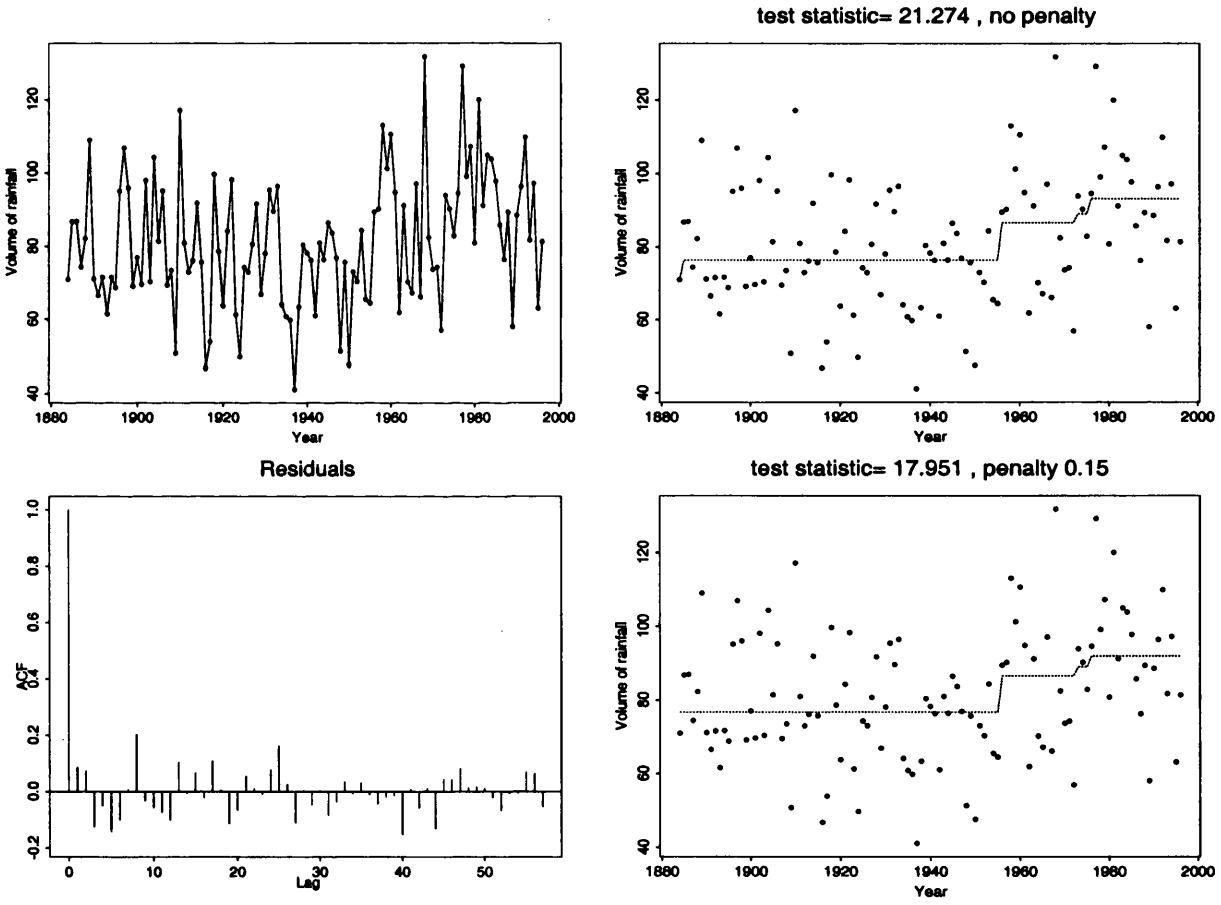

Figure 5.1. Argentina data: Detecting a change-point using isotonic regression with unpenalised and penalised functions with penalty factor 0.15. The top left panel displays the raw data; the top and bottom right panels display the raw data with the fitted unpenalised and penalised isotonic regression functions respectively. The bottom left panel presents the ACF plot of the residuals after removal of trend by isotonic regression.

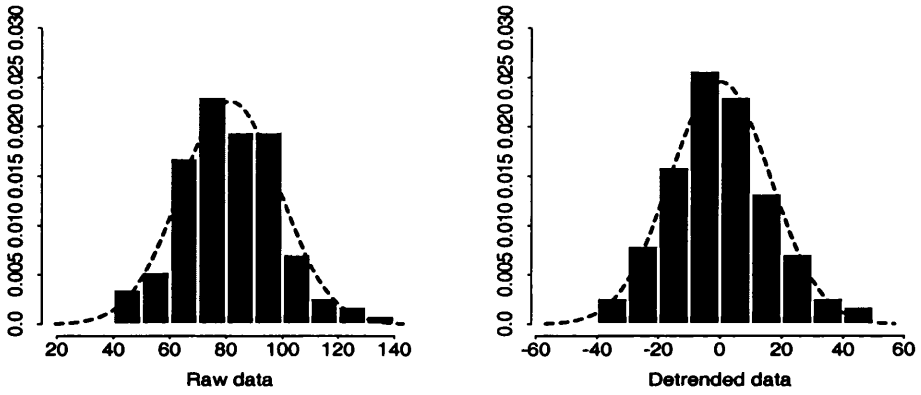

Figure 5.2. Argentina data: Checking normality of data. The left and right panels show histograms of the raw and detrended data using isotonic regression respectively, superimposed with their corresponding expected normal density curves.

is very low, at only 0.039. Hence, there is very little evidence to suggest any abrupt change in the data.

Next we will look at the plot of the posterior probabilities of the possible locations of a change-point conditional on its existence, shown at the bottom of the figure. This plot is however not helpful here as the value of $p(change)$ is too low to suggest any abrupt change occuring in the data. Nevertheless, for illustrative purposes, from the plot, the highest peak (posterior mode) indicates that the most probable change-point occurs at 1955. This is the same as that suggested by Wu's approach in section 5.2.1. The second highest is 1967.

We will consider Procedure 2 which uses overlapping segments of the data, with years 1887-1953 and 1933-1996. The value of $p(change)$ is low at 0.0276 for the first segment, indicating no change-point. However, for the second segment, $p(change)$ is moderate at 0.48. Though not convincing, there is now a much stronger indication of the presence of a change. Figure 5.4 shows the results using the second segment and from the bottom panel, the most probable change-point location indicated by the posterior mode is 1955, which is similar to that found by Wu.

In Table 5.1, we present the results using different combinations of overlapping segments. It is found that the result is very sensitive to the segments considered. Besides the two

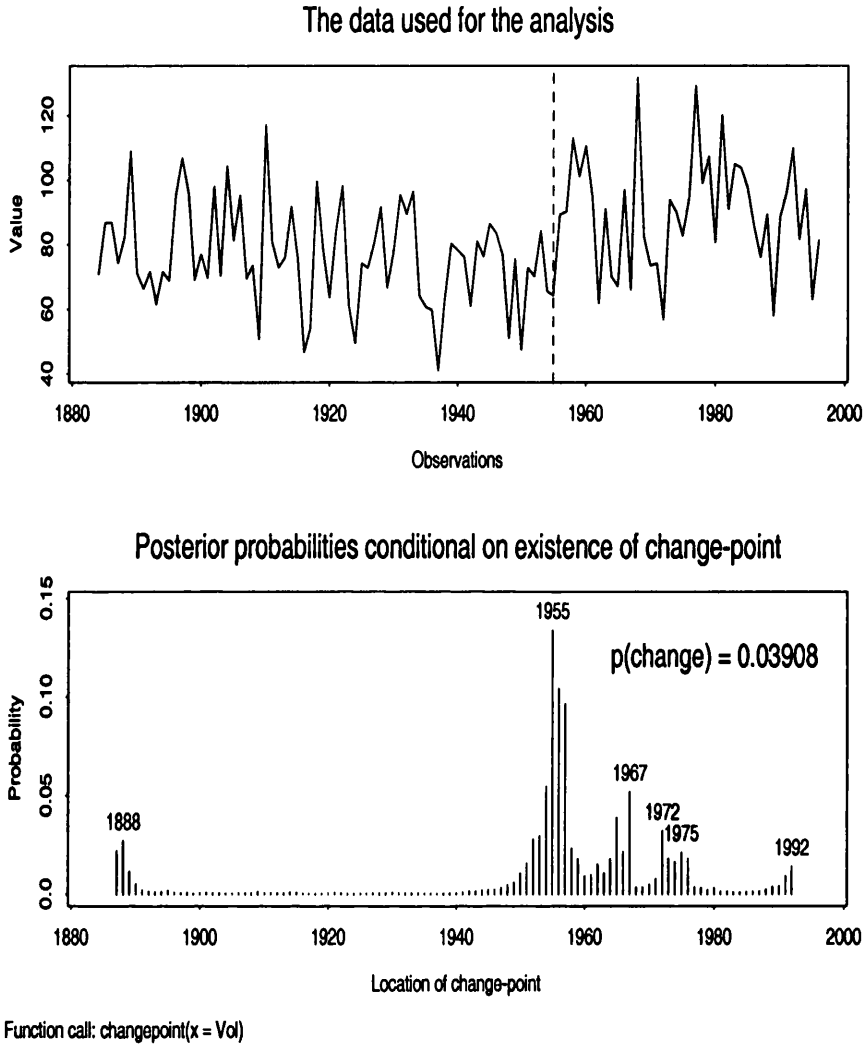

Figure 5.3. Argentina data: A Bayesian approach to detect a change-point, assuming a constant trend. The top panel shows the full data under study, with a dotted line indicating where the most probable change-point might occur if one assumes its existence. The bottom panel shows the resulting probability distribution over the location of a change-point conditional on its existence.

overlapping segments that are considered (highlighted in bold), the rest do not have sufficient evidence to indicate any abrupt change.

Next, let us now assume that there is an underlying trend in the data, and attempt to remove it by first fitting a nonparametric smooth curve to obtain the residuals. We then carry out Thomas's test on the residuals to see if there is a change-point.

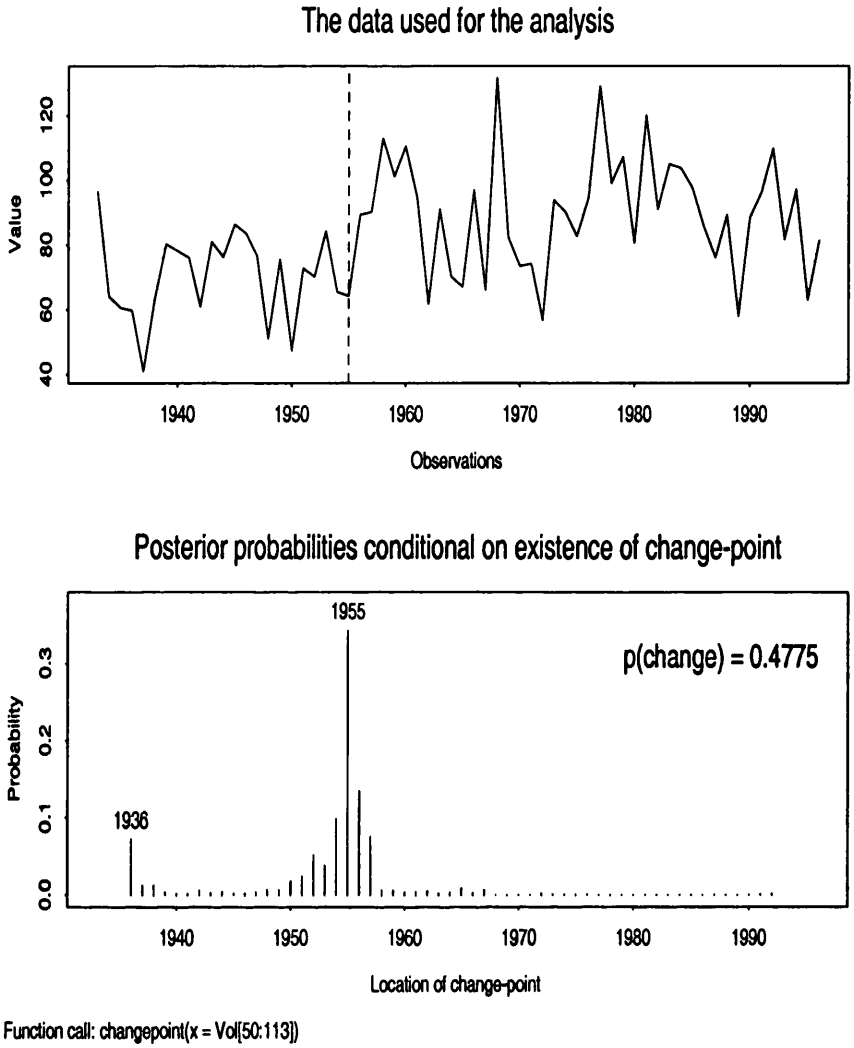

Figure 5.4. Argentina data: The top panel shows the second segment of the data under study, from year 1933-1996. The bottom panel shows the resulting probability distribution over the location of a change-point conditional on its existence.

Using a bandwidth of 13 to remove the trend gives $p(change)$ of 0.0053. As $p(change)$ is so low, it again suggests that the presence of a change-point in the series is unlikely. However, this result might be dependent on the bandwidth used to remove the trend to obtain the residuals. The use of a variety of values of *h.trend* is thus investigated. As presented in Table 5.2, *h.trend* does not affect the values of $p(change)$, which remains less than 0.01 for all values of *h.trend* considered. Using Procedure 2 with different overlapping segments of the detrended data does not alter the value of $p(change)$ much.

Segment	$p(change)$	posterior mode
1884-1933	0.0318	1929
1884-1943	0.0301	1933
1884-1953	**0.0276**	**1949**
1884-1963	0.0173	1957
1923-1996	0.2209	1955
1933-1996	**0.4775**	**1955**
1943-1996	0.0852	1955
1953-1996	0.0233	1956

Table 5.1. This table presents the results of $p(change)$ and possible change-point location indicated by the posterior mode, using different segments of the Argentina data.

h.trend	$p(change)$	posterior mode
1	0.005695	1888
10	0.005114	1888
15	0.005448	1992
20	0.005889	1888
40	0.007013	1888

Table 5.2. Argentina data: This table gives the $p(change)$ and possible change-point location indicated by the posterior mode, using a variety of *h.trend* values to remove the trend before carrying out the discontinuity test.

The assumption of normality of the detrended data (as shown in Figure 5.5) seems reasonable. Overall, the Bayesian approach does not provide sufficient evidence to support the presence of any abrupt change.

5.2.3 Using Nonparametric Regression

In this section, we will apply our nonparametric regression approach to detect discontinuities. We will first test for independence in the data via the ACF plot. If the independence assumption is valid, the discontinuity test is then carried out, without any incorporation of correlation. On the other hand, if the data show any possibility of correlation, we will then carry out the discontinuity test, accounting for correlation. As the Argentina data are equally spaced, the moving window approach will be used to model the correlation using an

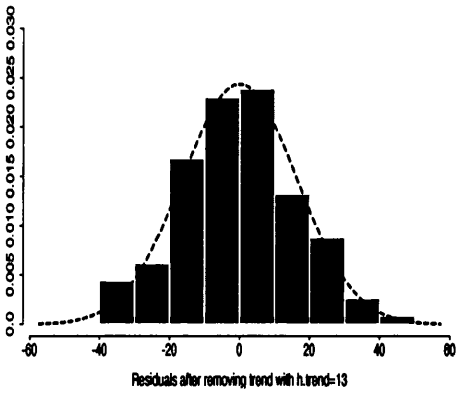

Figure 5.5. Argentina data: Checking normality of residuals after removal of trend using smoothing parameter of 13. The figure show a histogram of the residuals, superimposed with its corresponding expected normal density curves.

AR(1) model.

5.2.3.1 Test for independence

In this section, we are interested to test if there is any dependence structure in the data. As the data are equally spaced, we can simply look at the ACF plot to deduce if the independent data assumption is reasonable.

To evaluate the correlation of the data, we first remove the trend present in the data by fitting a smooth curve of a particular bandwidth, *h.trend*. Fig.5.6 shows the ACF plots of the residuals obtained using $h.trend = 7$ and $h.trend = 14$. Most of the lags are insignificant for both, implying that the data are almost independent. (This agrees with the ACF plot produced earlier for the isotonic regression approach, displayed in Figure 5.1.) Hence it seems reasonable to treat our data as independent in Section 5.2.3.2.

5.2.3.2 Treating the data as independent

In this section, we will consider carrying out the discontinuity test based on independence. Figure 5.7 shows the left (solid) and right (dotted) smooths of the data using smoothing parameters $h.test = 7$ and $h.test = 14$ for the discontinuity testing, and their corresponding

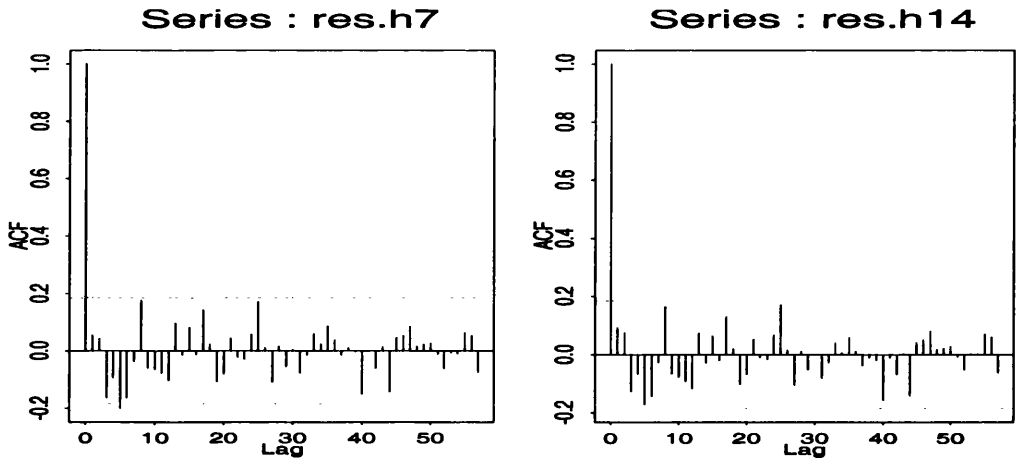

Figure 5.6. Argentina data: ACF plots of residuals obtained using *h.trend* = 7 (left diagram) and *h.trend* = 14 (right diagram).

reference bands. The latter are obtained by taking the average of the left and right smooths at each evaluation points, +/− the estimated standard error, *e.se.*, of the differences between the two smooths. This provides a useful guide as to whether change-points are present. We would expect that if there are change-points, the left and right smooths will not both be contained within the reference bands. (This implies that the two smooths are more than 2 standard errors apart). Hence, a good estimate of the locations of change-point are positions that have very large standardised difference $st.diff$. The latter is the difference between the two smooths divided by its corresponding estimated standard error of the differences between the two smooths.

The discontinuity test for *h.test* = 7 gives a p-value of 0.03. This indicates that the test is significant and hence there is sufficient evidence to support the claim that at least one abrupt change is present in the data. The years 1933, 1961 and 1955 (in decreasing order of $st.diff$) are not within the reference bands, suggesting possible change-point locations. The bottom panel displays the test with *h.test* = 14. Although the years 1955, 1975 and 1933 are not contained within the reference band, the p-value is 0.17, which is not significant. Therefore there is not sufficient evidence to support the claim of any abrupt change here. Notice that the left and right smooths here are smoother compared to those in the top plot, as a larger bandwidth is used.

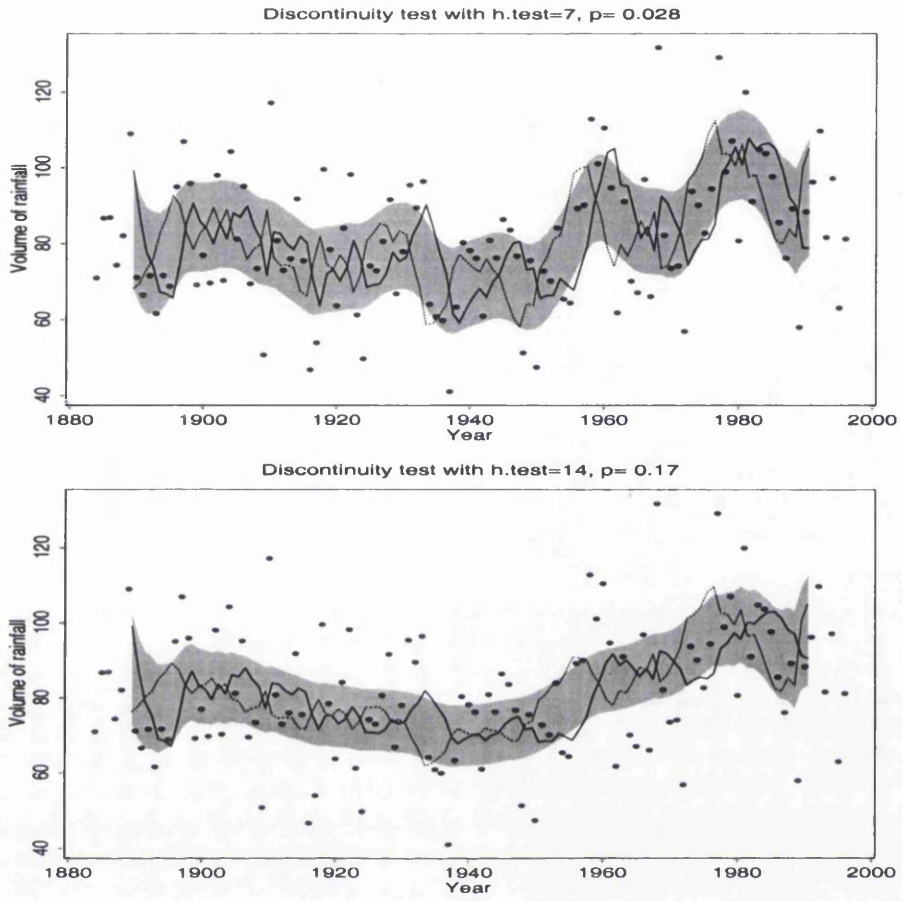

Figure 5.7. Argentina data: Treating data as independent with different bandwidths. The top panel shows the graph with its left (solid) and right (dotted) smooths and its corresponding reference bands, with bandwidth, $h.test = 7$. The bottom panel was obtained using $h.test = 14$.

The difference in p-values for $h.test = 7$ and $h.test = 14$ suggests that the test is very sensitive to the value of $h.test$ used. The change-point significance trace (CPST) that is proposed in Section 3.2.9 is displayed in Figure 5.8. From this plot, we can visualise the performance of the test in terms of how the significance of the test, the possible change-point locations, and the size of the abrupt change at the most probable location vary across different values of the smoothing parameters, $h.test$.

From the significance trace in the left panel, we can see that the test is significant only for $h.test = 4$ to 8. It is interesting to note from the change-point locator plot that the

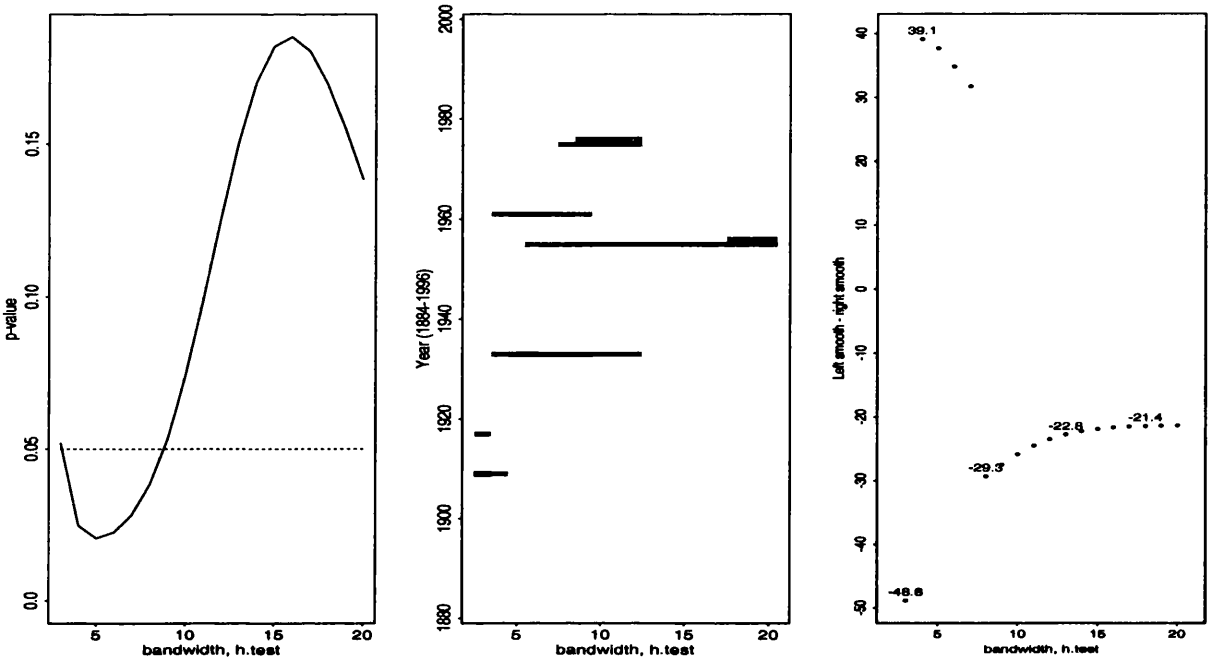

Figure 5.8. Argentina data: The left panel is a significance trace for a range of *h.test* from 3 to 20. The corresponding possible change-point locations are provided in the change-point location plot. The darker the tone of the line, the larger is the standardised difference between the left and right smooths, $st.diff$. Only values of $|st.diff| > 2.5$ are plotted. The plot at the far right shows the corresponding differences between the left and right smooths with the highest $st.diff$, at each bandwidth, *h.test*

change-point locations vary as *h.test* increases. For instance, at $h.test = 7$, the standardised differences of abrupt changes in descending order occur at 1933 (↓), 1961 (↓) and 1955 (↑), where the first two are abrupt drops, denoted by (↓) and only year 1955 shows an abrupt increase, denoted by (↑). However as *h.test* increases above 12, the only year that has high standardised differences is 1955 (↑), but the test is not significant. The change-point locations of the test when the p-value is significant are summarised in Table 5.3. Lastly, from the right panel, for $p < 0.05$, the abrupt changes as *h.test* increases from 4 to 7 are all positive, indicating abrupt drops in volume with decreasing magnitudes. The only exception is with $h.test = 8$ which shows an abrupt increase.

Figure 5.9 displays the fitted left and right smooths at the most probable change-point locations for various bandwidths, $h.test = 4$ to 14. It is quite clear that when a small

h.test	p value	possible change-point locations arranged in decreasing order
4	0.025	1933 (↓), 1961 (↓), 1909 (↑)
5	0.021	1933 (↓), 1961 (↓)
6	0.023	1933 (↓), 1961 (↓), 1955 (↑)
7	0.028	1933 (↓), 1961 (↓), 1955 (↑)
8	0.038	1955 (↑), 1933 (↓), 1961 (↓)

Table 5.3. Argentina data: Treating the data as independent using our nonparametric regression approach. This table gives the possible change-point locations for different h.test, where the discontinuity test is significant ($p < 0.05$). (↓): abrupt drop, (↑): abrupt increase

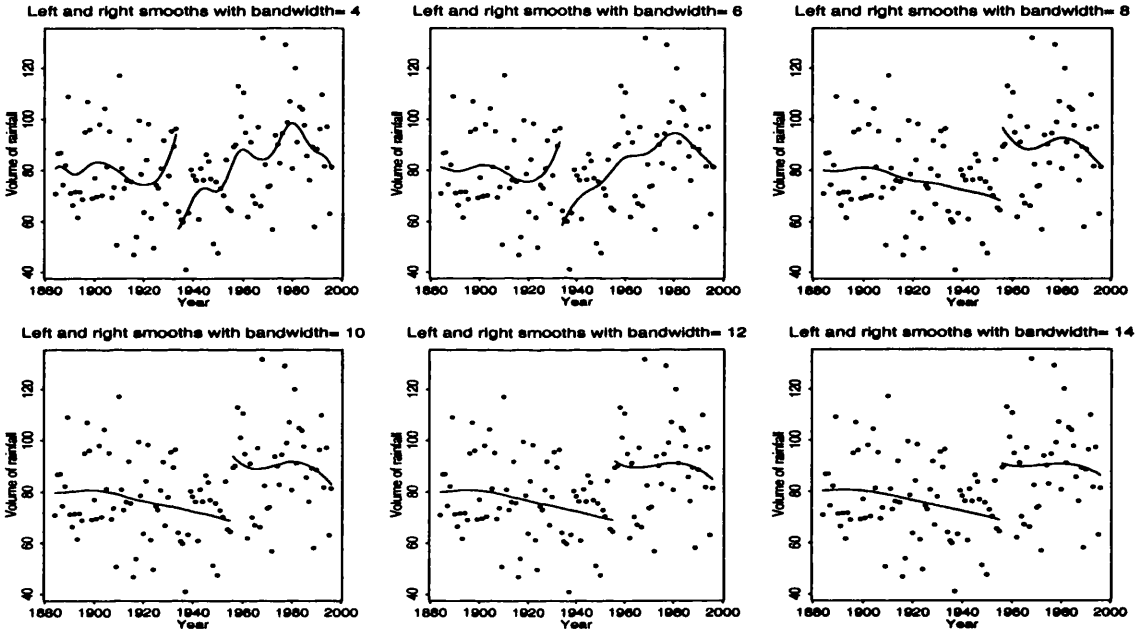

Figure 5.9. Argentina data: Fitting left and right smooths at the estimated change-point locations. The jump sizes are given in the right panel of Figure 5.8

bandwidth, *h.test* is used, the nonparametric curve traces the data closely. The left and right smooths become much smoother when higher values of *h.test* are used. The choice of *h.test* thus rests upon the degree of smoothness we would like to attribute to the underlying trend.

Since the test is only significant at $4 \leq h.test \leq 8$, our results are similar to those of

Wu's only when $h.test = 8$ is used where it gives an abrupt increase at 1955 with a jump size of 30. The other significant results suggest the most probable abrupt change is a drop at 1933 instead. Using larger bandwidths from 9 to 20, the largest abrupt standardised change is also registered at 1955 with a jump size of about 21.5, but this is non-significant. The error variance, $\hat{\sigma}^2$ is estimated by using the Rice's variance estimator. Treating the data as independent, it is obtained as 244.21, which is lower than that estimated by Wu.

It is justifiable to treat the data as independent as the correlation in the data is very low. However, to allow for comparison to Wu's test which accounted for correlation, we shall proceed to carry out our discontinuity test using the moving window approach, to account for correlation in the next section.

5.2.3.3 Modelling Correlation - Equally Spaced Setting

In this section, the moving window approach will be used to incorporate the estimation of correlation in the discontinuity test. This can also serve as a check on the sensitivity to the assumptions about the correlation structure.

Figure 5.10 shows the Argentina data with the left and right smooths with its corresponding reference bands, that is obtained with a smoothing parameter $h.test = 7$ and using moving window of size $b = 7$, which gives an estimated correlation of 0.027. (Notice that the reference band is slightly wider than that in the independent case using the same smoothing parameter in Figure 5.7.) The p-value of the test is just significant at 0.048, indicating the presence of an abrupt change. The largest abrupt change is a decrease at 1933 ($\downarrow$), followed by 1961 ($\downarrow$) and 1955 ($\uparrow$). The estimated error variance is 257.9 which is quite close to that estimated by Wu.

The change-point significance trace is displayed in Figure 5.11. The left panel shows the significance trace using $b = 7$, with its corresponding plots of change-point locations and size of abrupt change. The test is significant only for $h.test = 4$ to 7, which is a subset of that for the independent case ($h.test = 4$ to 8). Similarly, the change-point locator plot produced here is simply a subset of that for the independent case. The most abrupt change picked out by the test for $h.test = 4$ to 7 is a drop in 1933 ($\downarrow$). This is different from that obtained by

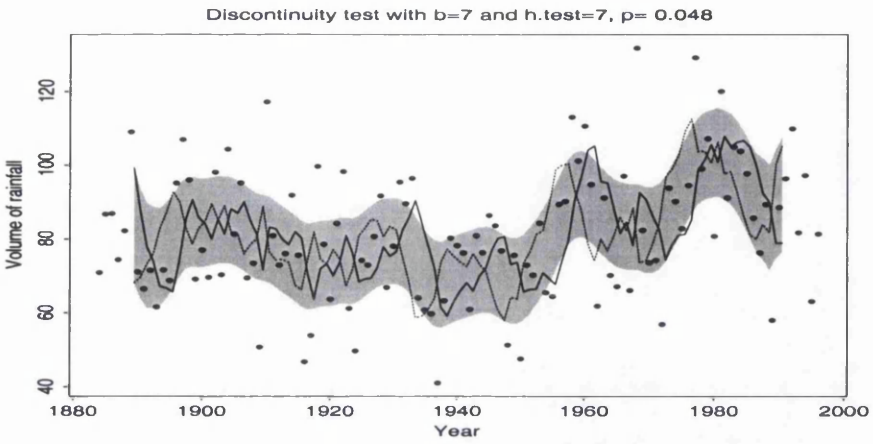

Figure 5.10. Argentina data: This figure shows the raw data with the left and right smooths with its reference bands, taking $h.test = 7$ and $b = 7$. This gives a significant result for the discontinuity test with $p = 0.048$.

Wu. However, the size of the most probable abrupt change over different *h.test*, is similar in both the correlated and independent cases.

As the test is sensitive to both the smoothing parameter, *h.test*, as well as the estimated correlation b, we will proceed next to do a sensitivity analysis, varying the size of b over different *h.test*. Figure 5.12 shows several significance traces to indicate how the p-values change if different sizes of moving windows b are used to estimate the correlation, over different *h.test*. The test is significant only when small values of b are used with small bandwidths, *h.test*.

The choice of b does affect the result of the discontinuity test. The test is significant only when a small moving window of less than 10 observations is used, i.e. $b < 10$ with $h.test \leq 8$. This corresponds to lower estimated correlation as shown in Table 5.4. As b increases, the estimated correlation increases. This is as expected because as the size of the moving window increases, it will include observations where there might be abrupt jumps or it might be influenced by the presence of trend, resulting in an overestimation of correlation. This in turn will decrease the probability of detecting a jump. In light of the fact that the data are almost independent, and allowing for moderate smoothing, the use of $b = 7$ and $h.test = 7$ in our earlier analysis seems appropriate.

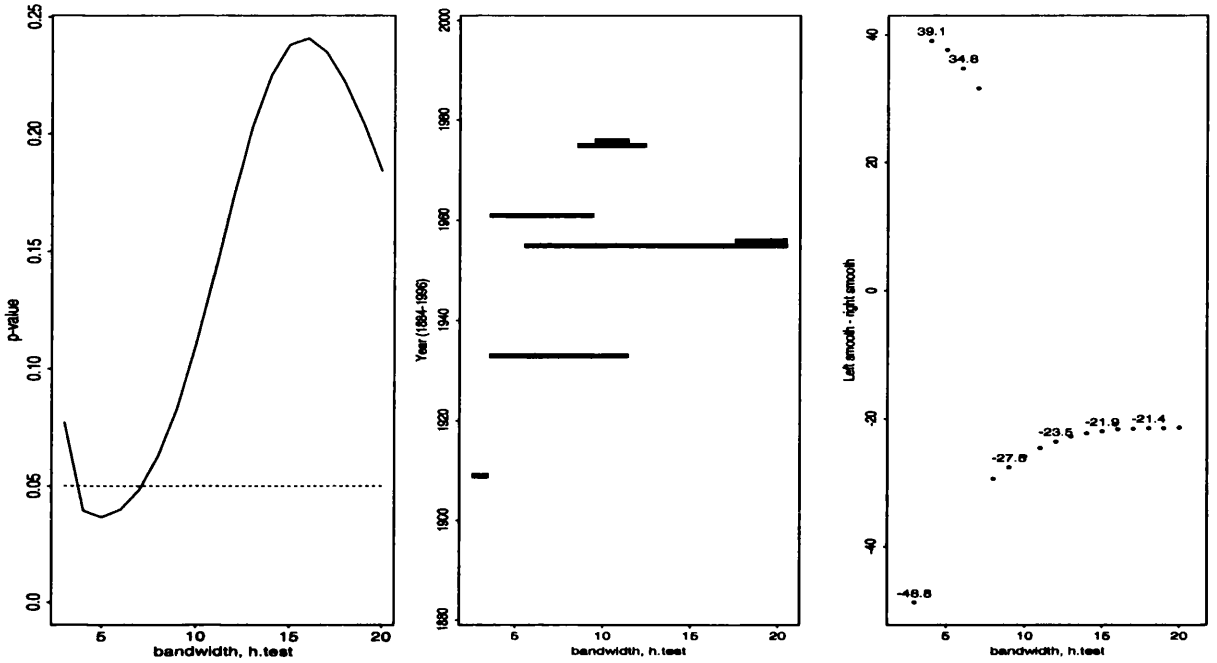

Figure 5.11. Argentina data: A moving window of size $b = 7$ is used to estimate the correlation of the data. The left plot is a significance trace over a range of *h.test* from 3 to 20. The middle plot is the change-point locator plot. The plot at the far right shows the corresponding differences between the left and right smooths which have the highest $st.diff$, at each bandwidth, *h.test*.

window size, b	estimated correlation
5	-0.03
6	0
7	0.03
8	0.03
10	0.04
15	0.07
20	0.08
25	0.11
30	0.12

Table 5.4. Argentina data: Estimated correlations using different moving window sizes, b

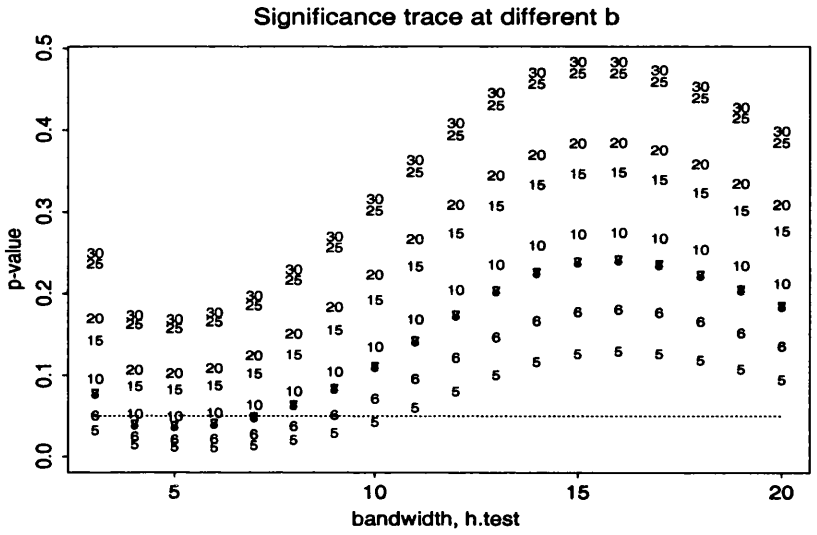

Figure 5.12. Argentina data: This figure shows the significance traces computed with different values of b indicated on the plot, over a range of *h.test*.

5.2.4 Summary of Argentina data analysis

In the analysis of the Argentina data, we have applied three different approaches to detect discontinuities.

Firstly, Wu's isotonic regression test is significant, which indicates that the mean is not constant. The rejection of his null hypothesis does not necessarily imply the presence of an abrupt change, as it could also be due to an increasing trend. In this set of data however, it does seem tenable to suggest that the test for both the penalised and unpenalised functions are significant due to the presence of an abrupt change in 1955 (refer to Figure 5.1).

Using the Bayesian approach, there seems to be insufficient evidence to support the presence of any change-point. We first carry out the analysis assuming that the underlying trend is piecewise constant. There is no evidence of any change when the full dataset is used since the overall evidence of a change in the series, $p(change)$, is very low. Using the segment 1933-1996, $p(change)$ is at the borderline case of 0.48, suggesting a possible change-point at 1955, which is the same as that obtained by Wu. The test is however very sensitive to the length of the segment chosen. The values of $p(change)$ are much lower for the other range of

segments considered. We also consider removing a smooth trend in the data before carrying out the test. This results in even weaker evidence of any abrupt change. Hence there is no convincing evidence to support any abrupt change using this approach.

Next, for the nonparametric regression approach, we first look at the ACF plots of the detrended data using different *h.trend* to examine if there is any evidence of correlation present in the data. There is insufficient evidence to support any dependence in the data, hence it is justifiable to carry out the test treating the data as independent. This gives a significant result for low to moderate smoothing parameters, $4 \leq h.test \leq 8$. The most probable change-points is at year 1933 ($\downarrow$) for $4 \leq h.test \leq 7$, which is different from Wu's. Our test gives the same significant result as Wu's with a jump at 1955 ($\uparrow$) only when $h.test = 8$ is used.

There is not much difference in the conclusion when the data are analysed with the moving window approach using a suitable moving window size b, which gives an estimated correlation of less than 0.03. The test is however significant at a smaller range of *h.test*, i.e. $4 \leq h.test \leq 7$, which picks out the most abrupt change as a drop in 1933. As there appears to be a trend in the data, using a moderate smoothing parameter in this case, seems appropriate.

In summary, both the isotonic regression and our nonparametric regression approaches have sufficient evidence to suggest that the data are discontinuous, but not the Bayesian approach. However, the former two approaches differ in their proposed change-point locations. Wu's test picks up 1955 ($\uparrow$) while our test suggests either 1933 ($\downarrow$) or 1955 ($\uparrow$) as the most probable change-points allowing for a moderately smooth trend. The year 1933 might not have been picked up by Wu as it is a monotonic test. A possible reason why the Bayesian approach has failed to detect any abrupt changes could be because it assumes a constant trend, which might not be the case in this data.

5.3 Application 2: The Global Warming Data

For at least the past decade, the warming of our climate has been a topic of great concern for many scientists, policy-makers, and citizens. Some scientists perceive that the changes in

global-mean temperatures are due to anthropogenic forcing influences (greenhouse warming), which are produced as a result of human activities and technology, and natural variability (Wigley et al., 1997, Jones et al., 2001). Others have remarked that it could be due to several other forces. Factors such as increase in solar luminosity, decrease in volcanic stratospheric aerosols, warming of carbon dioxide and increase in anthropogenic carbon aerosols (soot emissions) can result in an increase in the global temperature, while factors such as increased anthropogenic sulphur aerosols (Kaufman and Fraser, 1997) and stratospheric cooling due to decrease in ozone concentrations in the stratosphere can result in a decrease in global temperature.

Several IPCC Climate Models attribute the warming of 1940 to anthropogenic carbon dioxide, and cooling from 1940 to 1970 to anthropogenic sulphate aerosols, and resumed warming from 1970 to the present to the anthropogenic carbon dioxide (IPCC, 1995, 2001).

Most people think of global changes in temperature as varying smoothly, especially when we look at the gradual increase in temperatures over the past 30 years. However, history has revealed the harsh reality that there have been many instances where there were sudden shifts in temperatures, and these can possibly happen again, in the imminent future. The sudden impact that these could have on us and our environment might be costly, if we are caught unprepared (Abrupt Climate Change - Inevitable Surprises, 2002).

Hence, one of the first steps, which is of great importance, is to analyse our past data to examine if abrupt changes have indeed occurred, and to investigate what the possible causes could be. That will certainly advance our understanding of the climate, and help us to be more prepared to face the unavoidable surprises.

The global warming data considered here, which are a combination of both land and marine data, are provided by Jones et al. (2001). They consist of 145 equally spaced annual temperature anomalies from 1856-2000, which are expressed in degrees Celsius and are relative to the 1961-1990 mean.

We shall apply the three different discontinuity tests to the data. Though it is apparent that there is a general increase in the temperature, there are also periods of cooling (sometimes, abrupt) and thus, using isotonic regression might not be the most appropriate method. In the work presented by Wu et al. (2001), the natural variability is attributed

to the short range dependent background noise; while the mean trend is attributed to the different external forces that are causing changes in the temperature. He tested nonparametrically for an increase. In the author's approach, natural variability is also attributed to short range dependent background noise, but the trend is allowed to be smoothly varying, with no restriction that it has to be an increase, and interest lies in whether there is an abrupt shift in the smooth trend. On the other hand, the Bayesian approach assumes the trend to be piecewise constant and tests for an abrupt change in either the mean, correlation or variance simultaneously.

5.3.1 Using the Isotonic Regression Approach

Figure 5.13 shows the yearly global warming temperature anomalies, collected from 1856 to 2000, with both the unpenalised and penalised isotonic regression functions by Wu et al. (2001), using a penalty of 0.15 for the latter.

Looking at the top left panel which shows the raw data against year, the overall impression is that there is an increase in the annual year anomalies from 1856 to 2000. However, there also seems to be some abrupt drops in temperature around 1910 and 1950. The top right panel gives the same plot with the addition of a fitted unpenalised function. The test statistic value is 383.52 which is highly significant, indicating that the mean is not constant. Looking at the unpenalised isotonic regression function, there seems to be intervals where the temperatures are fairly constant and intervals where there are gradual increases over a period of time. For instance, there is an increasing trend from 1856 to the mid 1860s. The temperatures then remain fairly constant, before another increasing trend, occurring from 1920 to 1940, and another one from 1980 to 2000. (This is very different from the unpenalised or penalised function in the Argentina data, which displays an abrupt jump in 1955.) The autocorrelation plot in the bottom left panel indicates the presence of short range dependence. Only the first lag is significant. The bottom right panel shows the plot of the raw data with the penalised isotonic regressions using a penalty factor of 0.15. Wu et al. (2001) estimated $\hat{\sigma}^2 = 0.01558$ and showed that their test statistic is highly significant at both the 0.05 and 0.01 level.

From Figure 5.14, the raw data appears positively skewed, which could be due to the presence of both trend and possibly abrupt changes. However, the normality assumption for the detrended data using isotonic regression seems reasonable.

The significant result of Wu's test is only able to tell us that there is sufficient evidence to reject the null hypothesis and conclude that there is a change in the mean level of the global temperatures. However, it is not able to tell us if that is due to the presence of a monotonic trend or an abrupt change.

Looking at Figure 5.13, it is not obvious that the increase in temperatures occurs abruptly.

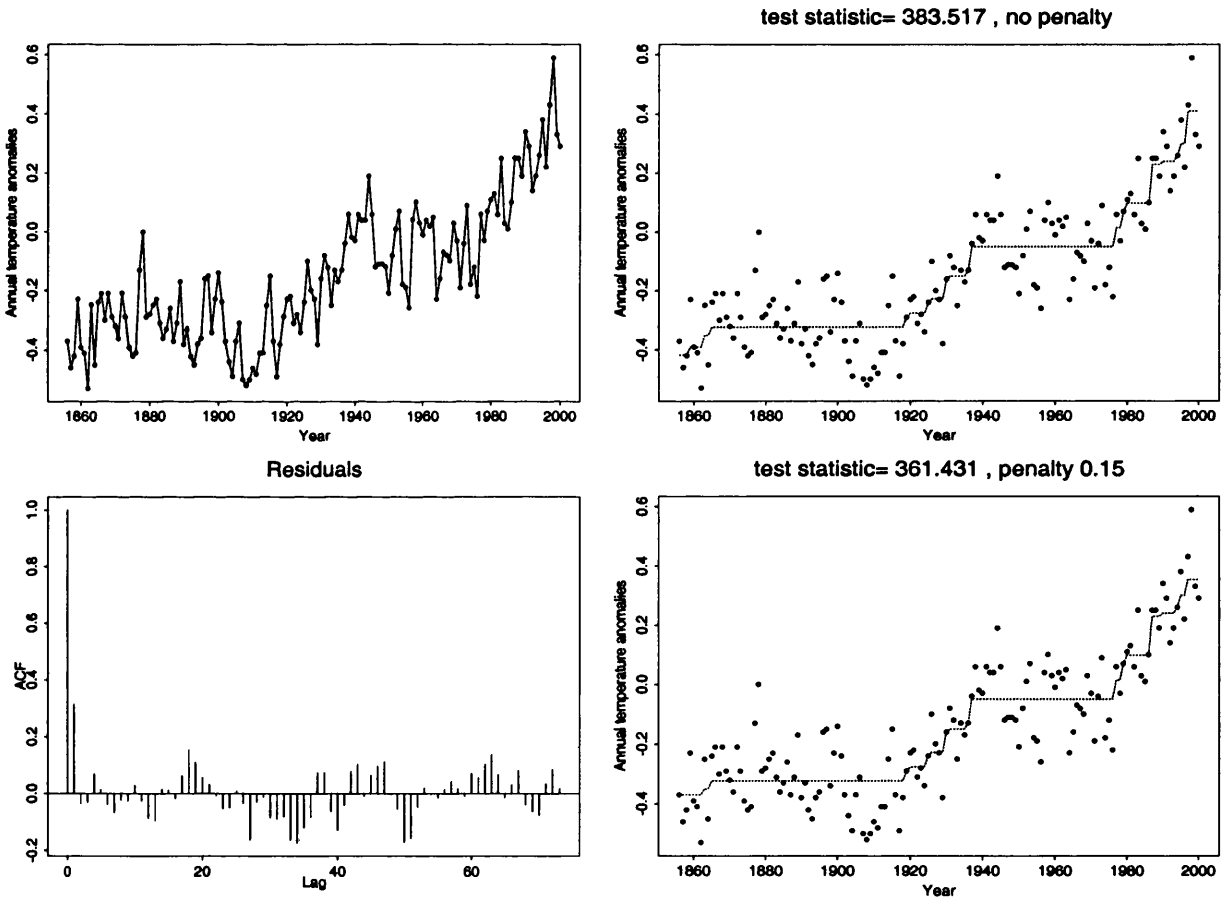

Figure 5.13. Global Warming data: Detecting a change-point using isotonic regression with unpenalised and penalised functions of penalty factor 0.15. The top left panel displays the raw data; the top and bottom right panels display the raw data with the fitted unpenalised and penalised isotonic regression functions. The bottom left panel presents the ACF plot of the residuals after removal of the trend by isotonic regression.

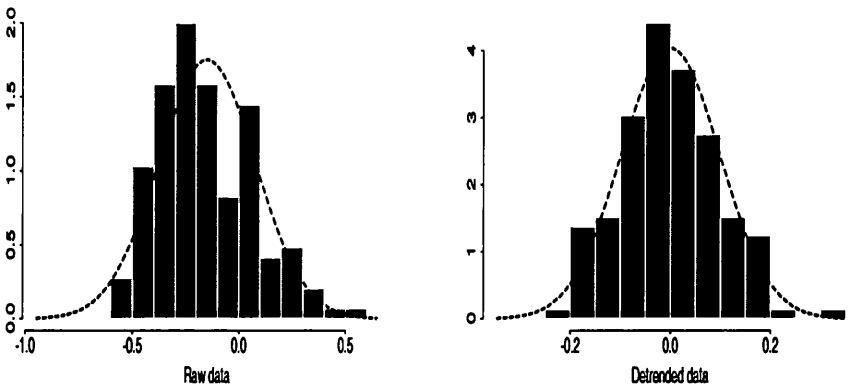

Figure 5.14. Global Warming data: Checking normality of data. The left and right panels show histograms of the raw and detrended data using isotonic regression respectively, superimposed with their corresponding expected normal density curves.

Instead, it appears to have increased over a certain period of time. Hence it is plausible that the significance of the isotonic test is due to the presence of monotonically increasing trends.

5.3.2 Using the Bayesian Approach

In this section, the Bayesian change-point algorithm is first carried out on the raw data, as shown in Figure 5.15. The overall posterior probability of an abrupt change occurring somewhere in the series is 0.9428, which is very high. Hence it is highly likely that there is an abrupt change in the data.

The bottom plot of Figure 5.15 shows the posterior probabilities of the possible locations of a change-point conditional on its existence. The peak or posterior mode occurs at 1929, which gives the possible change-point location. Examining the data before and after 1929, the abrupt change can be attributed to a shift in the mean, rather than the correlation or variance.

As Thomas' approach only assumes that the model has at most one change-point, the presence of several change-points might affect the analysis in an unpredictable way, as mentioned in Section 5.1. We will apply both Procedures 1 and 2 mentioned in Section 5.1 to investigate if there are any further change-points.

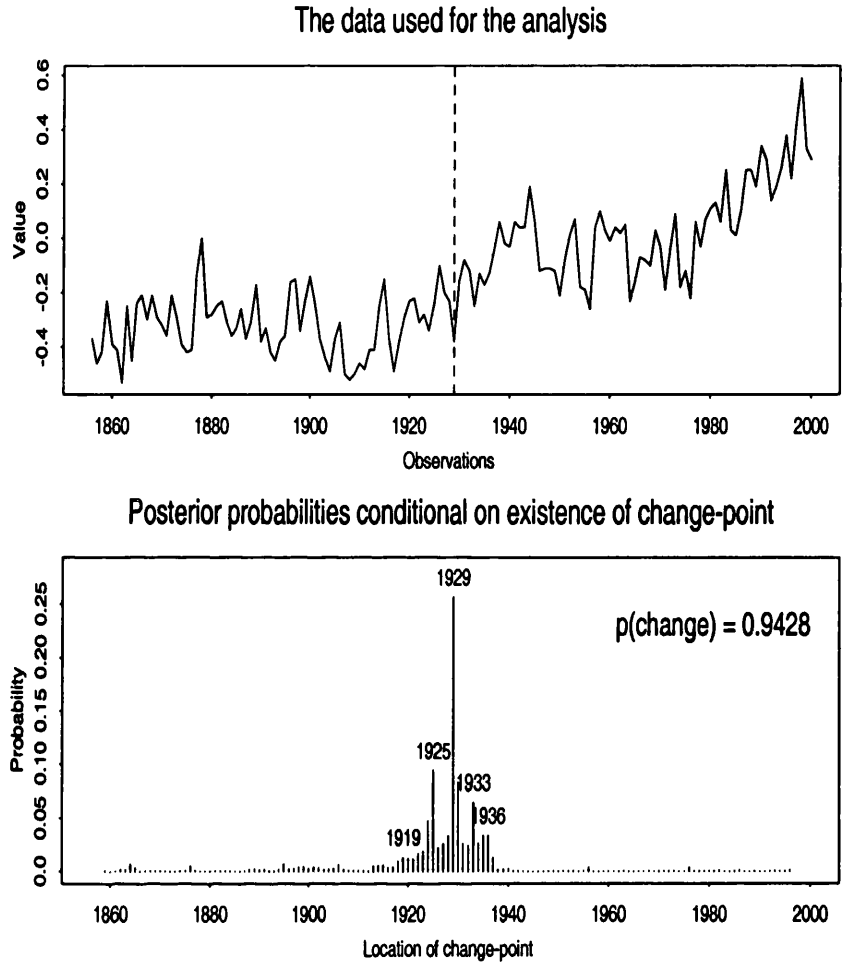

Figure 5.15. Global Warming data: A Bayesian approach to detect a change-point, assuming a constant trend. The top panel shows the full data under study, with a dotted line indicating where the most probable change-point might occur if one assumes its existence. From the plot of posterior probabilities conditional on a change-point at the bottom panel, the posterior mode gives a possible change-point location at year 1929.

For Procedure 1, the analysis is carried out using the first 74 observations. The value of $p(change)$ is low at 0.02969, and hence there is unlikely to be any change-point in this segment (from 1856-1929). However, if one insists on inserting a change-point, the most probable location is 1864. The $p(change)$ for the second segment, from 1930-2000, is moderate at 0.6093, and the most probable change-point location is 1986. This is displayed in Figure 5.16.

Further segmentation of the data gives very low $p(change)$ values, hence we can conclude that there could possibly be just two abrupt changes, occurring at 1929 and 1986.

Using Procedure 2, with overlapping segments 1856-1945 and 1935-2000 gives $p(change)$ of 0.7706 and 0.5856 respectively, and similar change-point locations as Procedure 1. The data are then partitioned at their change-point location and subjected to the change-point algorithm again. Subsequent analysis give very low $p(change)$, implying no other change point present.

The test is however sensitive to the segment that is being used for the change-point analysis. Table 5.5 shows how the values of $p(change)$ and the change-point location vary as different segments of the data are used, but most segments include a possible change-point at 1929. The other possible change-points are 1986 and 1948, though arguments against those points are that the overall evidence is not very strong for the former, while the latter is not very consistent when different ranges of segments are used.

segment	$p(change)$	posterior mode
1856-1940	0.343	1864
1856-1945	0.771	1929
1856-1955	0.572	1929
1856-1965	0.694	1929
1856-1975	0.913	1929
1856-1985	0.969	1929
1856-1995	0.976	1929
1856-2000	0.943	1929
1925-2000	0.500	1986
1930-2000	0.609	1986
1935-2000	0.586	1986
1945-2000	0.828	1948
1955-2000	0.384	1986
1965-2000	0.482	1978

Table 5.5. Analysis of change-point using different segments of the Global Warming data, via the Bayesian approach.

Next, we attempt to remove the trend in the data, using $h.trend = 13$, and carry out the change-point algorithm again. The value of $p(change)$ is very low at 0.004815, indicating no support to reject the hypothesis that the mean is constant. Table 5.6 presents the results

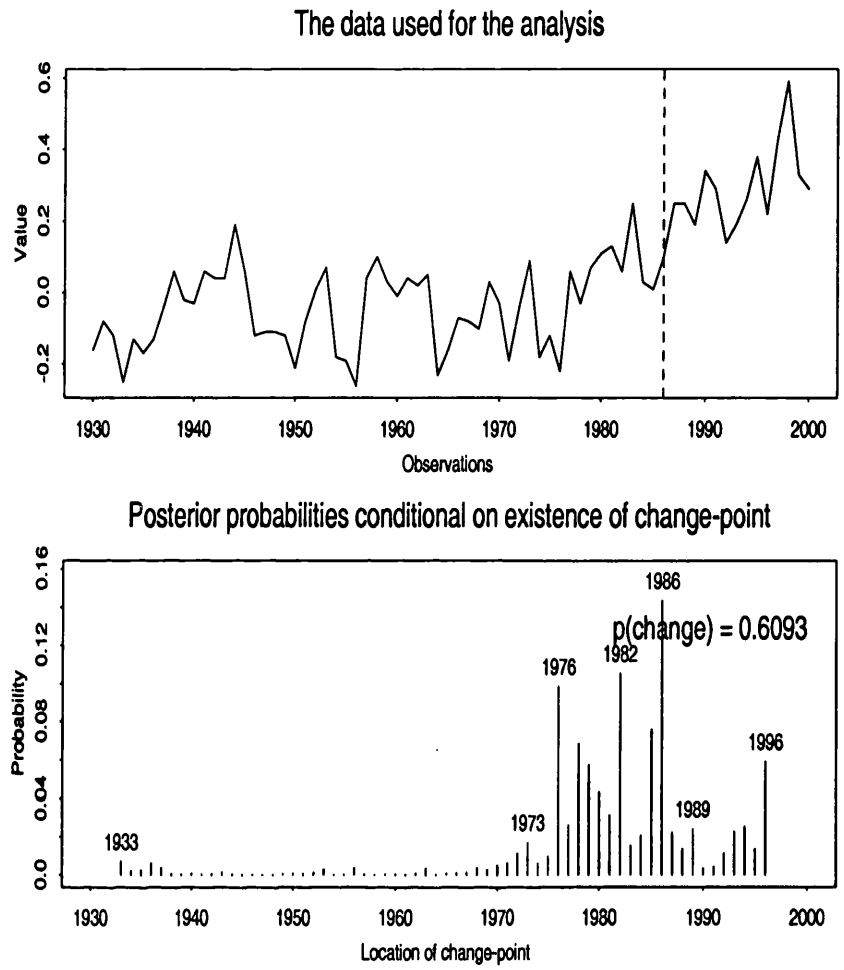

Figure 5.16. Global Warming data: Bayesian approach to detect change-point using part of the data from 1930 to 2000. The overall posterior probability of a change is 0.6093, which is relatively high. The bottom plot displays the posterior probabilities conditional on a change-point. The posterior mode gives a possible change-point location at the year 1986.

when other values of *h.trend* are used. The overall evidence of any change remains very low.

If the underlying trend in the data is assumed to be constant, then the change-point algorithm detects presence of an abrupt change at 1929. Another change could be present, but the evidence for that is not that strong. However, if the trend of the data is removed and the test carried out on the residuals, there is no evidence that there is any possible change-point in the series since $p(change)$ is low.

h.trend	*p(change)*	posterior mode
10	0.005114	1888
15	0.005448	1992
20	0.005889	1888
40	0.007013	1888

Table 5.6. Global Warming data: This table gives the *p(change)* and possible change-point location indicated by the posterior mode, using a variety of *h.trend* values to remove the trend before carrying out the discontinuity test.

5.3.3 Using a Nonparametric Regression Approach

We will now consider the approach that we have developed, a nonparametric regression approach to change-point detection. We will first consider the ACF plot of the residuals to see if it is justifiable to assume that the data are independent.

5.3.3.1 Test for independence

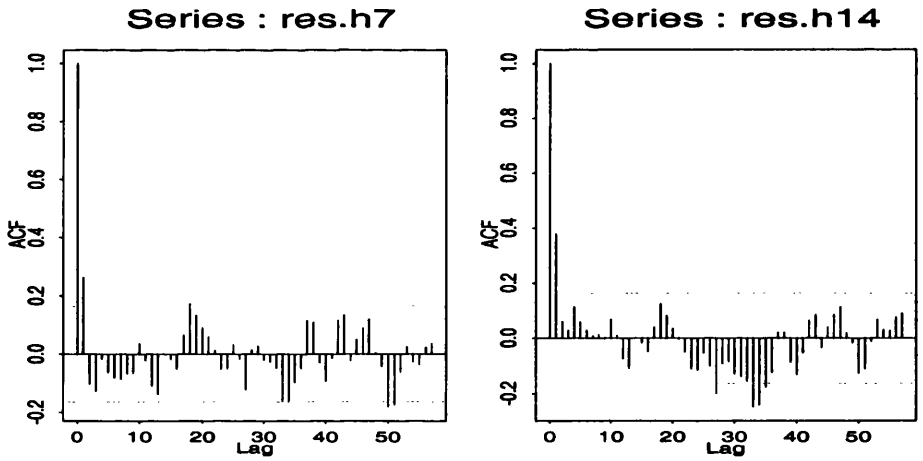

Figure 5.17. Global Warming data: ACF plots of residuals obtained using *h.trend* = 7 and *h.trend* = 14

Figure 5.17 displays the ACF plots of the residuals obtained using *h.trend* = 7 and *h.trend* = 14. Only the first lag is significant for both. (This is quite similar to the ACF plot obtained using isotonic regression in Figure 5.13.) This can imply either that there

is correlation present in the data (if we have successfully removed the trend), or that the correlation can be due to the trend which has not been completely removed, and hence it 'flows' into the residuals. Abrupt changes in the data might also contribute.

It seems reasonable to assume the presence of short range dependence structure, and proceed to carry out the discontinuity test, using the moving window approach to account for correlation.

5.3.3.3 Modelling Correlation - Equally Spaced Setting

In this section, the moving window approach is used to model the underlying correlation in the discontinuity test.

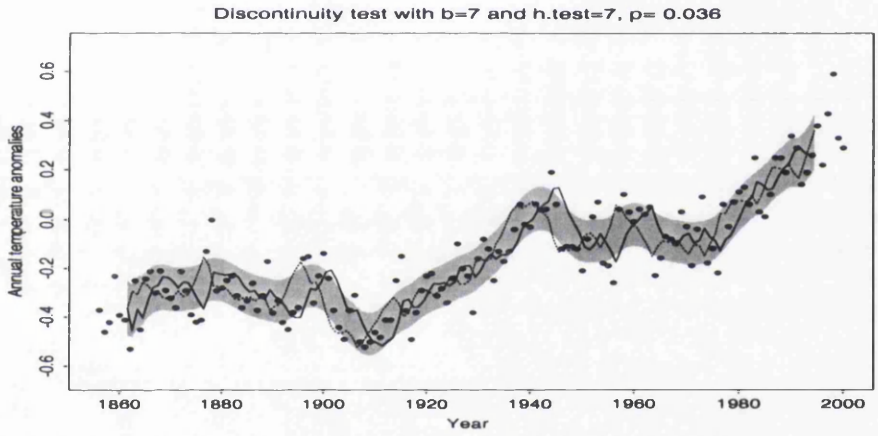

Figure 5.18. Global Warming data: This figure shows the raw data with the left and right smooths with its reference bands, taking $h.test = 7$ and $b = 7$. This gives a significant result for the discontinuity test with $p = 0.036$.

Figure 5.18 shows the global warming data with the left and right smooths with their corresponding reference band, which is obtained using moving window of size, $b = 7$ and a smoothing parameter, $h.test = 7$. The estimated correlation coefficient is 0.2. The p-value of the test is significant at 0.036, indicating the presence of an abrupt change. The largest abrupt change is a decrease at 1945 ($\downarrow$), followed by 1956 ($\uparrow$), 1901 ($\downarrow$) and 1895 ($\uparrow$).

The change-point significance trace is displayed in Figure 5.19. The left panel shows the significance trace using $b = 7$ and its corresponding plots of change-point locations and

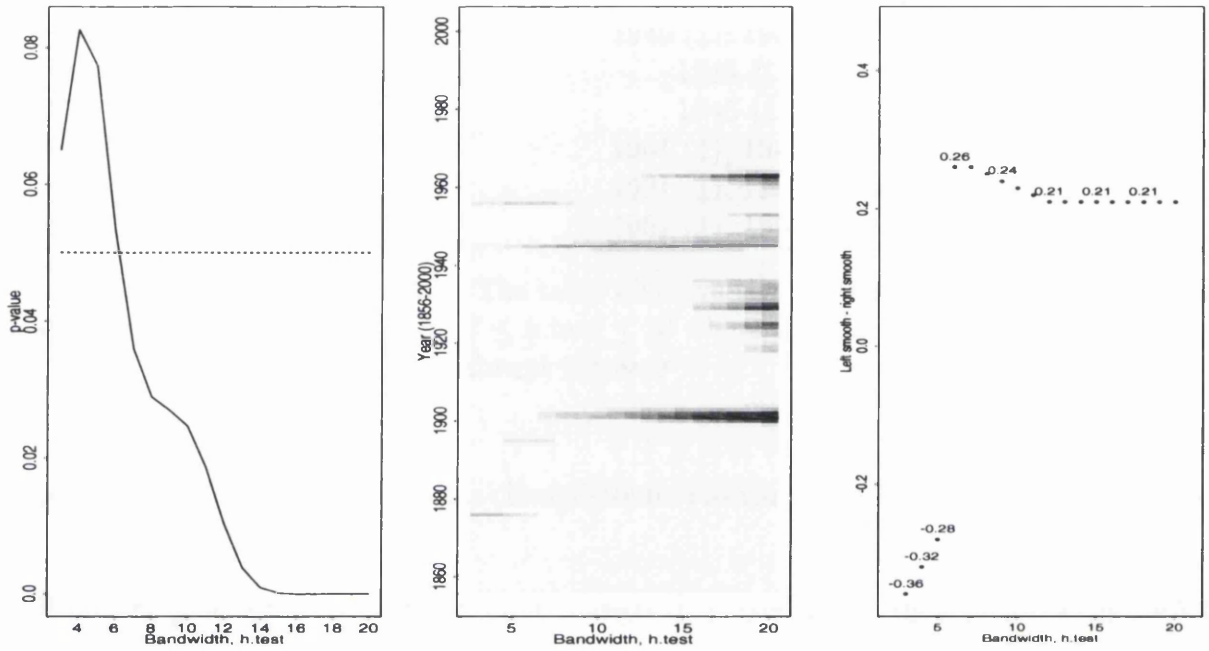

Figure 5.19. Global Warming data: A moving window of size $b = 7$ is used to estimate the correlation of the data. The left plot is a significance trace over a range of *h.test* from 3 to 20. The middle plot is the change-point locator plot, showing the possible change-point locations. The plot at the far right shows the corresponding differences between the left and right smooths which have the highest $st.diff$, at each bandwidth, *h.test*

size of abrupt change. The test is significant for moderate to large smoothing parameters, $7 \leq h.test \leq 20$.

The change-point locator plot in the middle panel displays the possible positions of the change-points, and shows that the change-point locations vary as *h.test* varies. For $h.test = 7$, the largest drop occurs at 1945 (↓), followed by abrupt changes in 1956 (↑) and 1901 (↓). For $h.test = 8$ and 9, the largest drop occurs at 1945 (↓), followed by another abrupt cooling between 1901- 1902 (↓), and an abrupt jump at 1956 (↑). For $h.test = 10-12$, the three plausible change-points are all abrupt coolings at around 1944-1947 (↓), 1901-1903 (↓), and 1963 (↓), in descending order. As *h.test* increases further, years 1901-1903 (↓) take over as the most probable change-point of abrupt cooling, followed by 1945-1947 (↓) and 1963 (↓). It is interesting to note that for $h.test > 9$, the three most probable change-points with highest standardised differences greater than 2.5, are all points of abrupt cooling. Table 5.7

h.test	$p-value$	possible change-point locations arranged in decreasing order
7	0.036	1945 (↓), 1956 (↑), 1901 (↓)
8,9	0.029,0.027	1945 (↓), 1901 (↓)
10-12	0.025,0.019, 0.01	1945 (↓), 1901 (↓)
13-16	0.004, 0.001, 0,0	1901 (↓), 1945 (↓), 1963 (↓)
17	0	1901 (↓), 1963 (↓), 1945 (↓)
18-20	0	1901 (↓), 1963 (↓), 1929 (↑)

Table 5.7. Global warming data: The table displays the three most possible change-point locations where $|st.diff| > 2.5$ for $7 \leq h.test \leq 20$ where the discontinuity test is significant ($p < 0.05$). (↓): abrupt drop, (↑): abrupt increase

summarises the three most plausible change-points (which are at least 5 years apart), whose $|st.diff| > 2.5$ and $p < 0.05$.

From the plot of the size of the abrupt changes displayed at the right panel of Figure 5.19, we can see that most of the abrupt changes over the smoothing parameters are positive, which indicate rapid drops in temperature anomalies of about 0.21-0.26 degree Celsius.

Next, we do a sensitivity analysis by varying the size of the moving window b over different *h.test*. The estimated correlations using different b are provided in Table 5.8. As the size of b increases, the estimated correlation increases. Using a small moving window is almost equivalent to assuming that the data are independent. From Figure 5.20, we can observe that the test is significant for almost all *h.test* for low b (less than 8). For moderate b, the test is significant for moderate *h.test* and above.

We have selected the use of a small moving window of size $b = 7$ in our analysis. It seems justifiable to use this value since by looking at the data, there appears to be the presence of both trend and abrupt changes. In such a situation, a small moving window would be recommended as it is more robust to the presence of trends and jumps. However, by doing this, we are also assuming that the correlation is not very large. This seems a reasonable assumption, since by looking at the ACF plots in Figure 5.17, the correlation of the first lag is only about 0.3, after quite a large smoothing parameter is used to remove the trend. Considering the fact that this value might be overestimated due to the presence of abrupt changes that might not be fully removed, the underlying correlation may be less than that.

window size, b	estimated correlation
5	0.07
6	0.13
7	0.20
8	0.23
9	0.30
10	0.31
15	0.36
20	0.46
25	0.50
30	0.54

Table 5.8. Global Warming data: Estimated correlations using different sizes of moving windows, b.

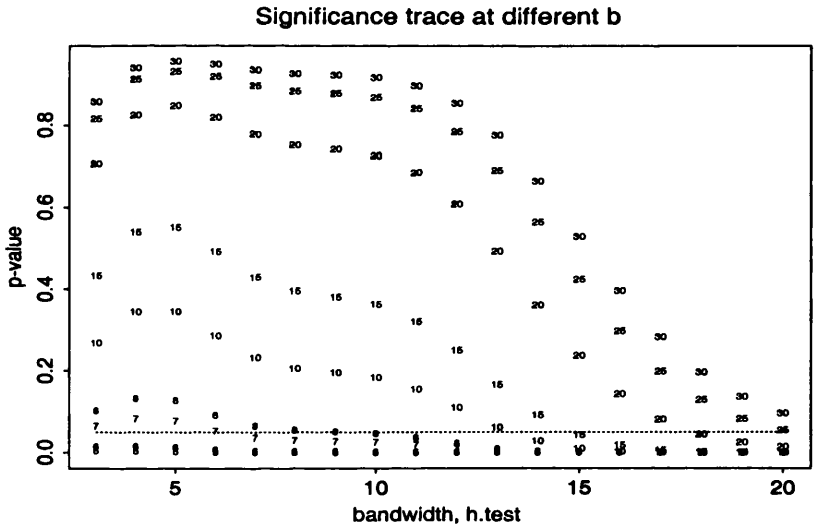

Figure 5.20. Global warming data: This figure shows the significance traces computed with different values of b indicated on the plot, over a range of *h.test*.

Hence from Table 5.8, the recommended values of b will be less than 10.

The normality assumption of the detrended data using $h.trend = 7$ and 14 is checked and it appears reasonable (refer to Figure 5.21). Using our discontinuity test, accounting for correlation, there is convincing evidence to suggest that the data are discontinuous, as

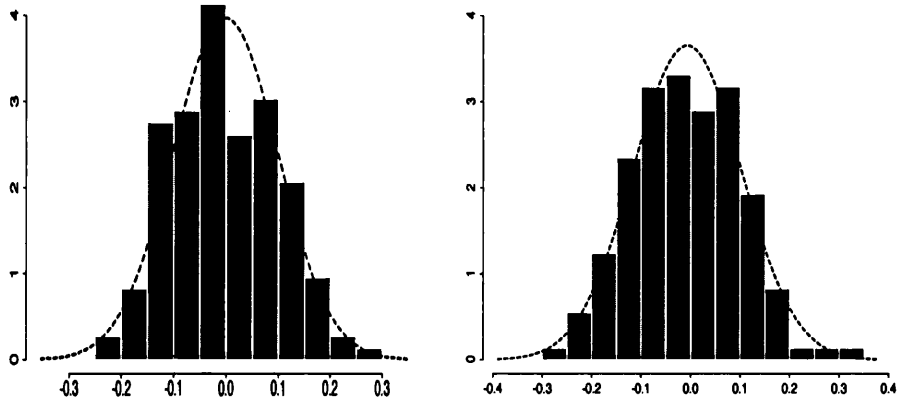

Figure 5.21. Global Warming data: Checking normality of data. The left and right panels show histograms of detrended data using smoothing parameter $h.trend = 7$ and 14 respectively, superimposed with their corresponding expected normal density curves.

the test is significant over a wide range of smoothing parameters, $7 \leq h.test \leq 20$. Though the locations of the change-points are sensitive to the value of $h.test$ used, which is the degree of smoothness attributed to the underlying trend, they all indicate abrupt coolings. This suggests that though global temperatures are increasing over the years, there have been periods of abrupt coolings.

5.3.4 Summary of Global Warming data analysis

The results of all three tests are significant. For the isotonic regression approach, the test using both penalised and unpenalised loglikelihood ratio test statistics, gives significant results, and so we reject the null hypothesis that the mean level is constant. The fitted isotonic regression function seems to suggest that this might be due more to monotonically increasing trends over certain periods of time, rather than abrupt changes.

For the Bayesian approach, the test gives significant results, implying that the underlying mean structure is not constant. The algorithm is first applied to the raw dataset with no trend removed. Procedures 1 and 2 give indications that there are two abrupt changes present in the series, namely an abrupt increase in 1929, followed by possibly another one in 1986, although the evidence for the latter is not strong. However, when overlapping segments are

used in Procedure 2, the test is sensitive to how the segments are partitioned. A smooth curve is also fitted to remove the trend in the data, and the test is conducted again on the residuals. This gives a non-significant result, possibly because the smooth curve might have removed any possible abrupt changes present.

For the nonparametric regression approach, we first investigate if the data can be assumed to be independent. From the ACF plot of the residuals, obtained after fitting a smooth curve of a particular smoothing parameter to remove the trend, only the first lag is shown to be significant.

Accounting for correlation in the data, the discontinuity test is carried out using the moving window approach with an appropriate window of size $b = 7$. The test is significant over a wide range of smoothing parameters. Hence there is convincing evidence to suggest that there are abrupt changes in the data, in the presence of a smoothly varying trend. The change-point locations are however sensitive to the smoothing parameter used, though they do reach a common agreement that the most probable change-point at each smoothing parameter is an abrupt drop in temperature. Moreover, it is possible that there might be more than one change-point, as indicated from the change-point locator plot, which shows mainly abrupt drops in temperature, except for a few exceptions. Hence, allowing for a moderate to very smooth underlying trend, there is convincing evidence that there have been periods of abrupt coolings in global temperatures. Nonetheless, the test is also sensitive to the moving window size that is used to estimate the correlation in the data. A small moving window is recommended in this case as it is more robust to the presence of both trends and jumps.

In conclusion, though the three discontinuity tests are all significant, they differ in the interpretation of the results. Wu's test suggests that the mean is not constant, possibly due to the monotonic trend of increasing global temperatures. As for the Bayesian approach, the overall evidence of an abrupt change present is very high and it gives the most probable change-point at 1929. It is not surprising that this change-point is also picked up by our test when we use a large smoothing parameter of 18 and above to give a very smooth (close to linear) trend. However, allowing for a moderate to a very smooth underlying trend, our discontinuity test has managed to pick up the presence of other abrupt changes in the data.

In fact, the more prominent ones are actually abrupt coolings in 1901, 1945 and 1963. This is in line with what several scientists have been concerned about regarding sudden shifts in temperatures (Abrupt Climate Change - Inevitable Surprises, 2002). The isotonic regression test is obviously unable to pick this up since it is monotonic. The Bayesian approach, on the other hand, might have been affected by the presence of positive trend in the data, which in turn hindered the detection of the abrupt drops in temperature.

5.4 Application 3: The River Clyde Data

Water quality in the River Clyde in Scotland is routinely monitored by the Scottish Environment Protection Agency (SEPA), which is the public body in charge of environmental protection in Scotland. In 1985, a sewage treatment plant was upgraded. It is of interest to investigate if this new treatment has improved the quality of the water, by analysing the level of the dissolved oxygen in the river, which is a variable that is closely linked to the quality of the water. Measurements of the surface level dissolved oxygen at one station on the river have been collected from 1982 to 1988. There are 70 data-points which are unequally spaced over time.

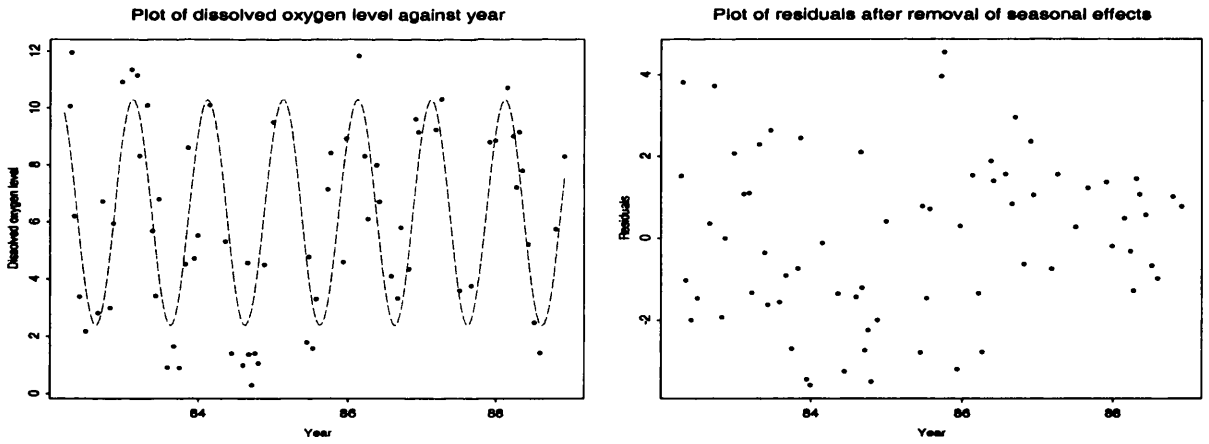

Figure 5.22. River Clyde data: The left panel shows the original dissolved oxygen with a fitted cosine curve to remove the seasonal effects present, while the right panel shows the residuals obtained after removal of seasonal effects.

The left panel in Figure 5.22 shows the original dissolved oxygen level against year (1982 to 1988), with a cosine curve fitted by nonlinear least squares to remove the seasonal effects present in the data. The right panel shows the residuals after seasonal effects have been removed. These will be the observations used for analysis in the following sections, to investigate if there is any abrupt change in the mean level of dissolved oxygen.

Subjectively, there appears to be a gradual decrease in dissolved oxygen level from 1982 to 1985 and possibly an abrupt increase mid 1985, and a continual decrease from 1985 to 1988. There seems to be a considerable amount of variation in the data, which might suggest that the seemingly abrupt increase we observe in 1985 could be due to random variation, instead of an abrupt change. We would expect, however, that it will be difficult to detect a jump, in the presence of such high variation.

Though the River Clyde data are unequally spaced, they are quite regularly spaced in terms of seasons and thus it can be said to be predominantly a regularly-spaced dataset. Hence we will first look at the data, treating them as equally spaced and compare the analysis, using the three approaches in Sections 5.4.1, 5.4.2 and 5.4.3, as we have done in the earlier two applications. This also allows a comparison of the three methods. Lastly, we will also carry out our nonparametric regression approach that is adapted for unequally spaced observations in discontinuity testing, to see if there is any difference in the results.

5.4.1 Using the Isotonic Regression Approach

In this section, we will apply the isotonic regression approach to the data where the seasonal effects have been removed.

The top left panel in Figure 5.23 shows the plot of the dissolved oxygen level after the removal of the seasonal effects, measured from 1982 to 1988. The top and bottom right panels display the unpenalised and penalised isotonic regression functions respectively. The latter uses a penalty factor of 0.15. The error variance is estimated as 3.50. The unpenalised test statistic takes the value 7.075, which is significant at the 0.05 level, hence allowing us to reject the hypothesis that the underlying mean trend is constant. However, using a penalty factor of 0.15 for the penalised isotonic regression gives a test statistic value of 5.483, which

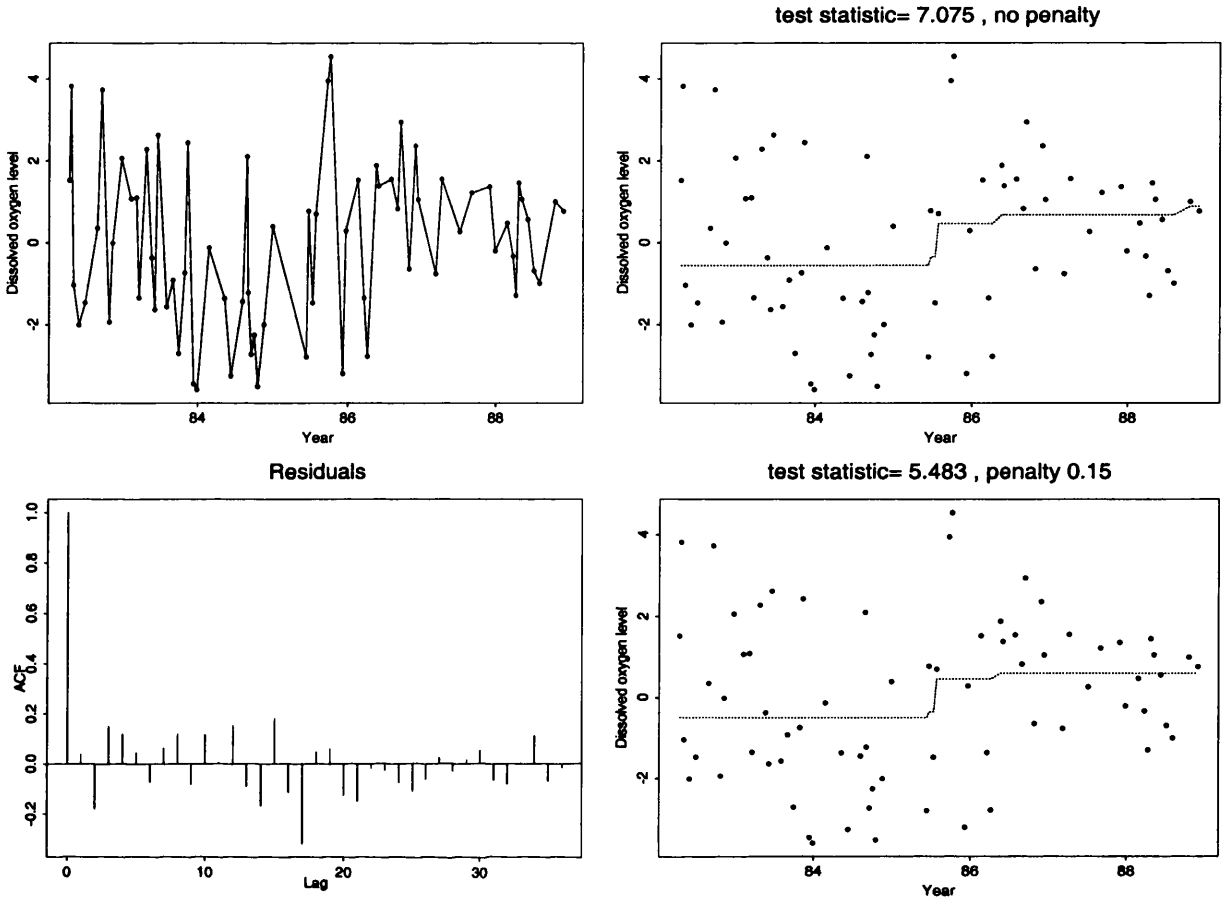

Figure 5.23. River Clyde data: Detecting a change-point using isotonic regression with unpenalised and penalised functions of penalty factor 0.15. The top left panel displays the raw data; the top and bottom right panels display the raw data with the fitted unpenalised and penalised isotonic regression functions. The bottom left panel presents the ACF plot of the residuals after removal of the trend by isotonic regression.

is not significant at the 0.05 level, with reference to the critical values generated by Wu et al. (2001). Subsequent values of the test statistic using smaller penalty factors of 0.05 and 0.1, are 6.506 and 5.982 respectively, which are both not significant at the 0.05 level. Hence, using penalised isotonic regression suggests no evidence to reject a constant mean.

From the ACF plot at the bottom left panel of Figure 5.23, almost all the lags are insignificant, indicating that the data can actually be assumed to be independent. From Figure 5.24, we can observe that the raw data on the left panel appears to be positively skewed. The assumption of normality is more reasonable for the detrended data using

isotonic regression from the adjacent plot.

Using the isotonic regression approach, the test is only just significant if we use the unpenalised log likelihood ratio test statistics, which suggests an abrupt change at mid 1985. On the other hand, using a penalty factor of 0.05 and above gives insignificant results, implying that there is insufficient evidence to reject the hypothesis that the underlying mean structure is constant.

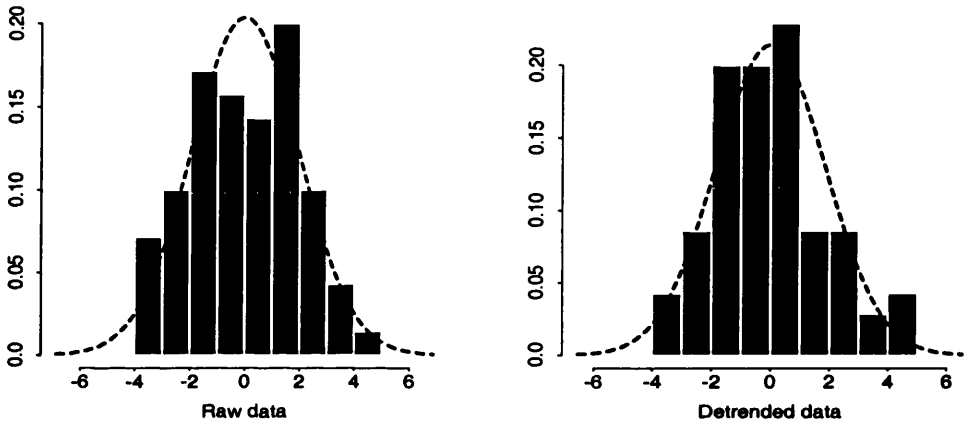

Figure 5.24. River Clyde data: Checking normality of data. The left and right panels show histograms of the raw and detrended data using isotonic regression respectively, superimposed with their corresponding expected normal density curves.

5.4.2 Using the Bayesian Approach

In this section, we apply Thomas' Bayesian change-point algorithm to the data. As the sample size is quite small, only Procedure 1 which uses the whole dataset is considered for the analysis. His model assumes that the underlying trend is constant which might not be applicable in this data.

We will first attempt to run his change-point algorithm, taking the trend as constant. The top plot of Figure 5.25 gives the volume of dissolved oxygen level against year. The dotted line indicates where the most probable change-point occurs if one assumes its existence, which is about March 1986. It appears that there might be some differences in the variances

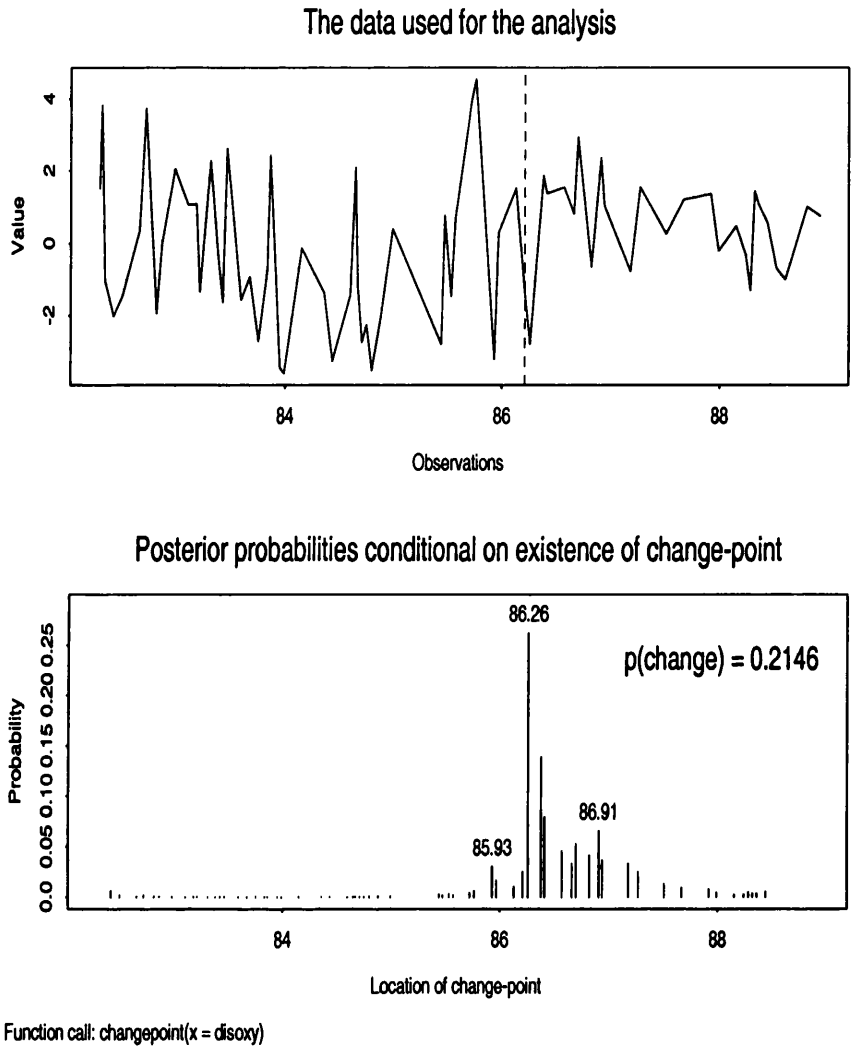

Figure 5.25. River Clyde data: A Bayesian approach to detect a change-point. The top panel shows the dissolved oxygen level with a dotted line indicating where the most probable change-point might occur if one assumes its existence. From the plot of posterior probabilities conditional on a change-point at the bottom panel, the posterior mode gives a possible change-point location at year 1986.26

between the two segments before and after the change-point. However, the evidence of the presence of any change is too weak to reach any conclusion.

Next, let us attempt to remove any trend in the data before carrying out Thomas' algorithm. Using $h.trend = 1$, the value of $p(change)$ obtained is very low at 0.091, indicating no support of a change in the series. The analysis is repeated with a range of *h.trend* and

h.trend	*p(change)*	posterior mode
0.4	0.092	47
0.7	0.095	46
1.0	0.091	46
1.3	0.081	46
1.6	0.070	53
1.9	0.064	53
2.2	0.052	53
3.0	0.065	53

Table 5.9. River Clyde data: This table gives the *p(change)* and possible change-point location indicated by the posterior mode, using a variety of *h.trend* values to remove the trend before carrying out the discontinuity test.

the results are displayed in Table 5.9. The value of *p(change)* remains very low.

Applying the Bayesian approach to both the raw and detrended data, there appears to be no evidence to support the presence of a change-point in the series.

5.4.3 Using the Nonparametric Regression Approach

We will now consider the approach that we have developed, using nonparametric regression approach for change-point detection. The structure of this section follows that of Section 5.2.3, with the addition of the modelling of the correlation structure in an unequally spaced setting.

5.4.3.1 Test for independence

As the River Clyde data are unequally spaced, we will just focus on the test of constant variogram devised by Diblasi and Bowman (2001), to test if there is a dependence structure in the data, instead of using the ACF plot. This test involves the choice of a few parameters, namely, *h.trend*, *maxdist*, *nbins* and *h.test* as mentioned earlier in Section 5.1.3. We first look at how the test performs as we vary the distance at which the empirical variogram is evaluated with $maxdist = 6, 5, 4, 3$, fixing the others at $nbins = 50$ and $h.trend = h.test = 1.4$.

The top left panel in Figure 5.26 displays the square root of the pairwise differences

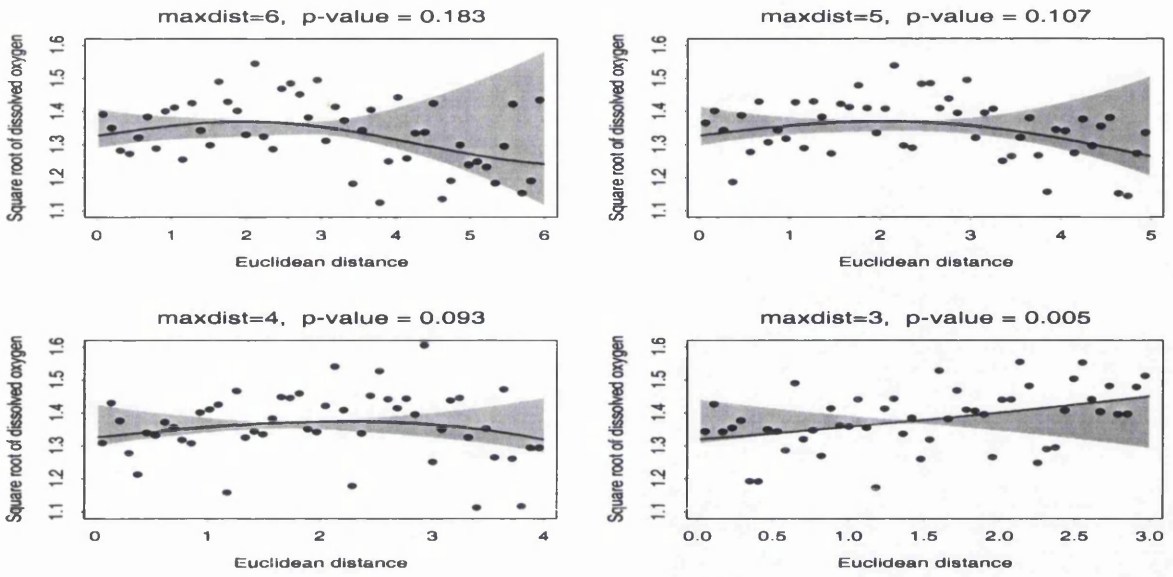

Figure 5.26. River Clyde data: A test for constant variogram using $maxdist = 3, 4, 5, 6$, coupled with $nbins = 50$, $h.test = h.trend = 1.4$.

of the absolute values of the residuals used to obtain the empirical variogram after the removal of trend with $h.trend = 1.4$ at $maxdist = 6$. The p-value of the test is 0.18, which is not significant, and hence there is insufficient evidence to reject the hypothesis that the underlying data are independent. As in the conclusions drawn from the p-value, the smoothed variogram is indeed contained within the 95% reference bands. However, as the value of *maxdist* decreases, the p-value also decreases and is significant at 0.005 for $maxdist = 3$. As expected, the smoothed variogram is not contained within the reference bands as can be observed from the bottom right panel. Hence the distance at which the empirical variogram is evaluated does affect the test to a certain extent.

Besides varying the value of *maxdist*, the sensitivity of the test to the other parameters, *h.trend*, *nbins*, over different *h.test* is also investigated. The significance traces using different *h.trend* and *nbins* are plotted in the left and right panels of Figure 5.27 respectively. We have fixed $nbins = 50$ for the former, and $h.trend = 1.4$ for the latter.

From the left panel, we can observe that as *h.trend* increases, the p-value decreases from above to below the 0.05 level. This implies that if a large smoothing parameter is used to

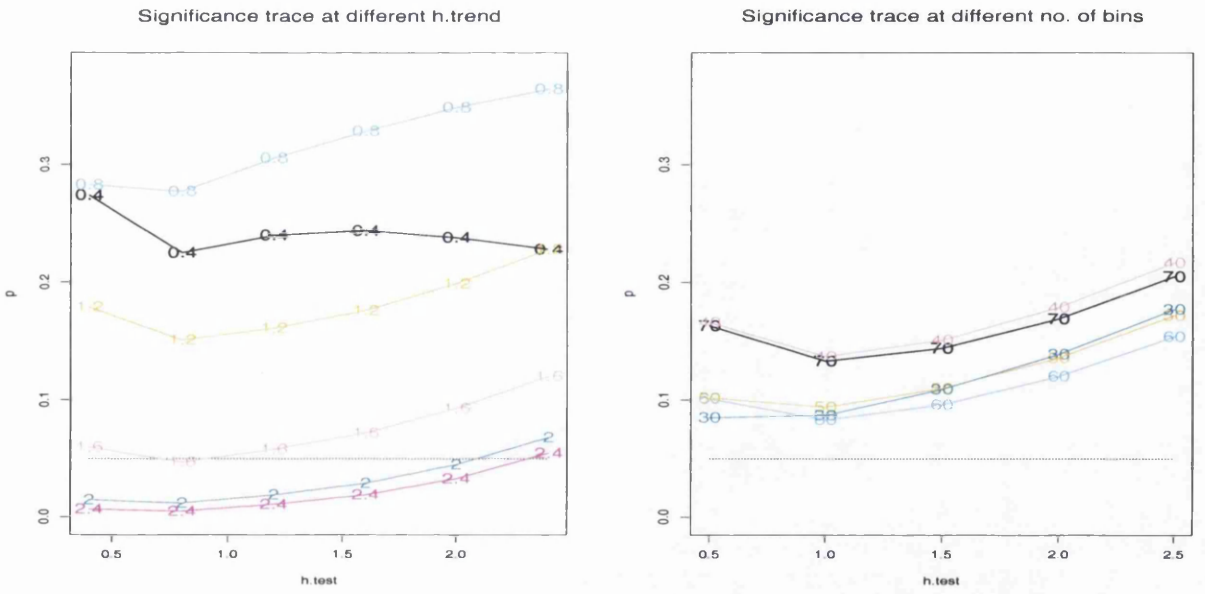

Figure 5.27. River Clyde data: The left panel shows the significance trace using different *h.trend* as denoted on the plot, with *nbins* = 50 for a test of constant variogram. The right panel displays a significance trace using different *nbins*, with *h.trend* = 1.4. Both have *maxdist* fixed at 5.

remove the trend in the data, the test will be significant, over almost all *h.test* < 2. This is not too surprising as using a small smoothing parameter to remove the trend, might also remove any correlation that is present in the data. However, the issue that correlation might be falsely introduced if the trend or jump is not properly removed, remains. The test is however not sensitive to the choice of *nbins*.

Most of the p-values are non-significant for a wide range of parameters that we have investigated, indicating no dependence structure is present. Nevertheless, the result is not very conclusive, as it seems to be dependent on some of the parameters used in the algorithm, namely *h.trend* and to a lesser degree *maxdist* and *h.test*. Hence, we will proceed to analyse the data, treating it as independent as well as modelling the correlation.

5.4.3.2 Treating the data as independent

In this section, the data are assumed to be independent. Figure 5.28 shows the discontinuity test being carried out for $h.test = 0.7$ and 1.7. Their left and right smooths are superimposed onto the plot with the corresponding reference band. The p-values are 0.68 and 0.045 respectively, whereby only the latter is significant, indicating the presence of at least one discontinuity. The presence of a discontinuity is more obvious from the bottom panel where the widths of the reference bands are narrower and the left and right smooths are not contained within the 95% reference band around 1985-1986. The error variance is estimated as 3.12, slightly lower than that estimated by Wu et al. (2001).

The great difference in p-values for $h.test = 0.7$ and $h.test = 1.7$ suggests that the test is very sensitive to the value of $h.test$ used. The change-point significance trace, CPST, shown in Figure 5.29 illustrates how the test performs when a variety of $h.test$ values are used, together with the possible change-point locations and the sizes of the most abrupt changes.

We can observe from the left panel that the test is significant for values of $h.test$ greater than 1.6, and is insignificant otherwise. The corresponding change-point locations for the former are around mid 1985 with a jump of about 3. In fact, there is only a slight variation in the change-point locations, all being around mid to late 1985, over the range of $h.test$ considered. The abrupt changes are mainly abrupt increases, except at lower bandwidths, $h.test < 1$, which are abrupt drops.

Figure 5.30 illustrates the possible change-point locations with the fitted left and right smoothers, using $h.test = 0.4$ to 2.2. When bandwidths $h.test$ of 1.6 and 2.2 are used, the underlying trend of the data is close to linear. As expected, using a smaller bandwidth gives a more irregular trend. However, since the test is significant only for $h.test > 1.6$, there is convincing evidence that there is an abrupt increase in mid 1985 only if we are willing to assume that the underlying trend is very smooth.

5.4.3.3 Modelling Correlation - Equally Spaced Setting

In this section, we will incorporate the estimation of the correlation in the data using the moving window approach in our discontinuity test. Though the data are irregularly spaced,

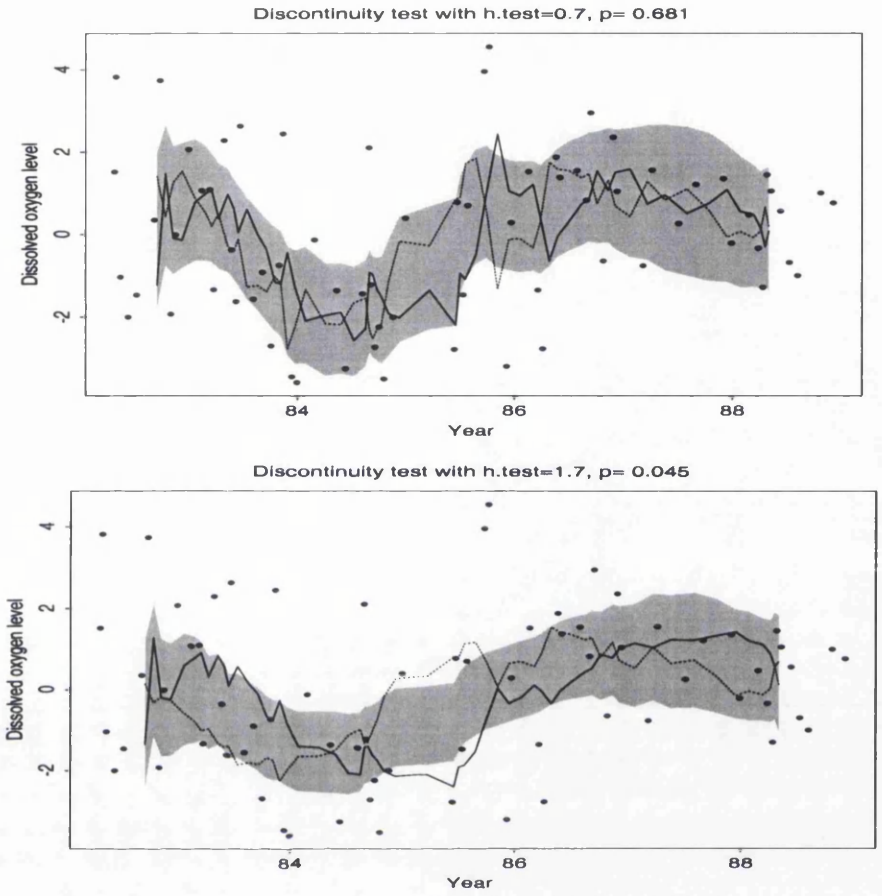

Figure 5.28. River Clyde data: Treating data as independent with different bandwidths. The top panel shows the graph with its left (solid) and right (dotted) smooths and its corresponding reference bands, with bandwidth, $h.test = 0.7$. The bottom panel was obtained using $h.test = 1.7$.

the correlation is still modelled using an AR(1) model to allow for comparisons with the other two tests.

Figure 5.31 presents the data with the left and right smooths and their reference band, using $b = 15$ and the same smoothing parameter as used earlier, $h.test = 1.7$. The estimated correlation coefficient is negative at -0.081. This gives a significant result with a lower p-value of 0.011 than in the independent case. The largest abrupt change is still registered at mid 1985. The error variance is estimated at 2.88. This results in the reference bands being narrower than those in the independent case with the same smoothing parameter.

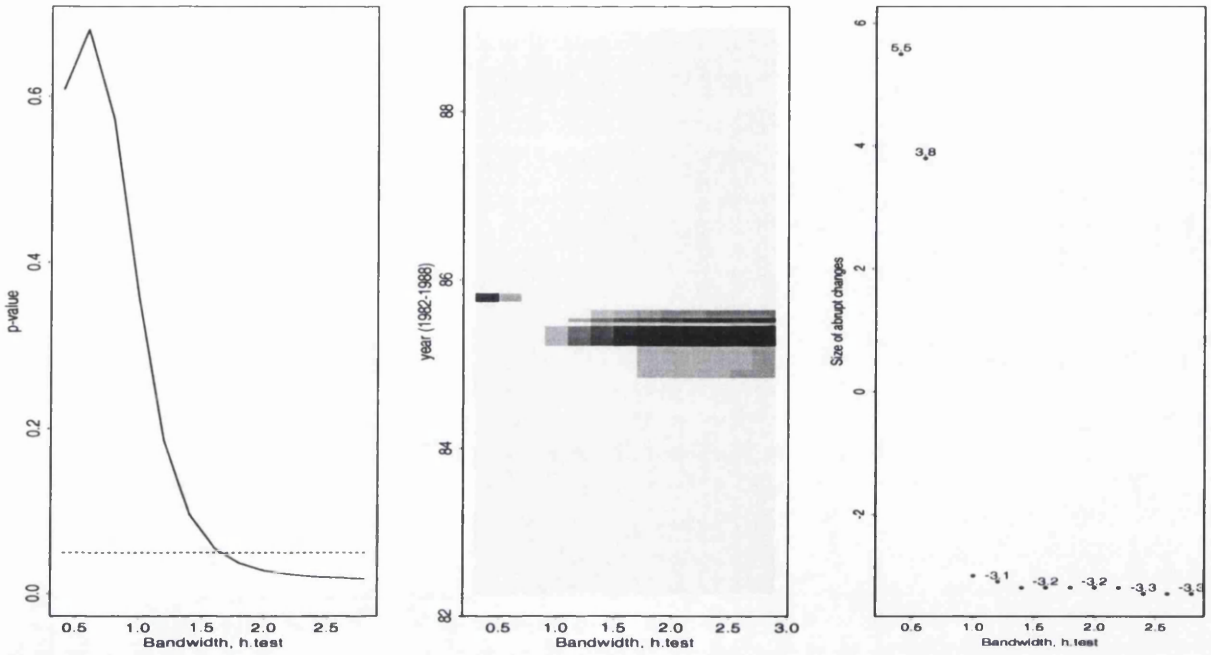

Figure 5.29. River Clyde data: The left plot is a significance trace which is a plot of the p-values over a range of *h.test* from 0.4 to 2.8. The corresponding possible change-point locations are provided in the adjacent plot. The plot at the far right shows the size of the most abrupt change.

The change point significance trace (CPST) is displayed in Figure 5.32, which is quite similar to Figure 5.29 for the independent case. The differing features are that the test is significant over a wider range of smoothing parameters, $h.test > 1.2$ and with the addition of a few other possible change-point locations.

Next, we perform a sensitivity analysis by varying the size of the moving window b over different values of *h.test*. From Figure 5.33, we can observe that the test is significant over a wider range of *h.test* for lower b. However, the test is not significant over all *h.test* only when a large b of size 28 is used, which corresponds to an estimated correlation of 0.069. The estimated correlations using different b are provided in Table 5.10. As the size of b increases, the estimated correlation increases. It is interesting to note that using a moving window size of less than 25 gives a negative correlation, while a size of 25 and above gives positive correlations.

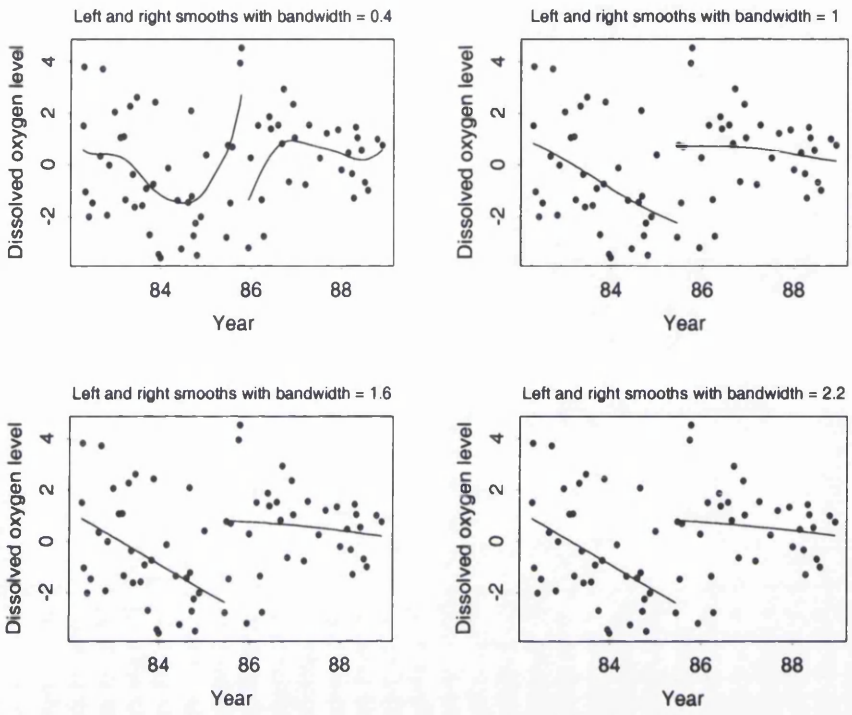

Figure 5.30. River Clyde data: Fitting left and right smooths at the estimated change-point locations. The jump sizes are given in right plot in Figure 5.29.

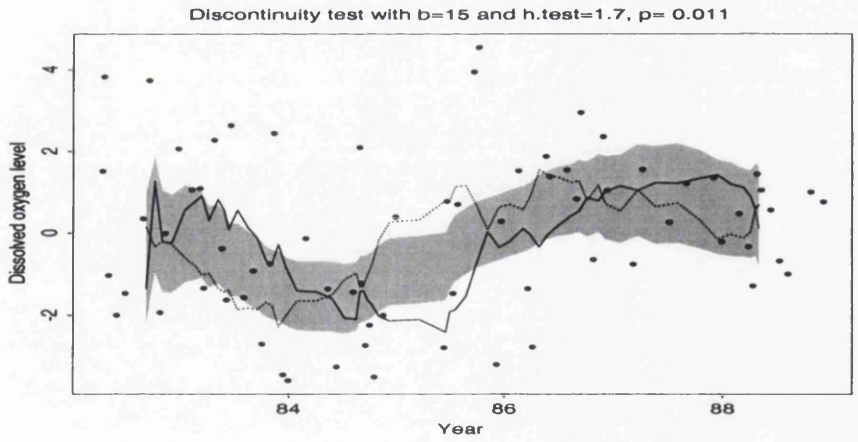

Figure 5.31. River Clyde data: This figure shows the raw data with the left and right smooths and its reference bands, taking $h.test = 1.7$ and $b = 15$. This gives a significant result for the discontinuity test with $p = 0.011$.

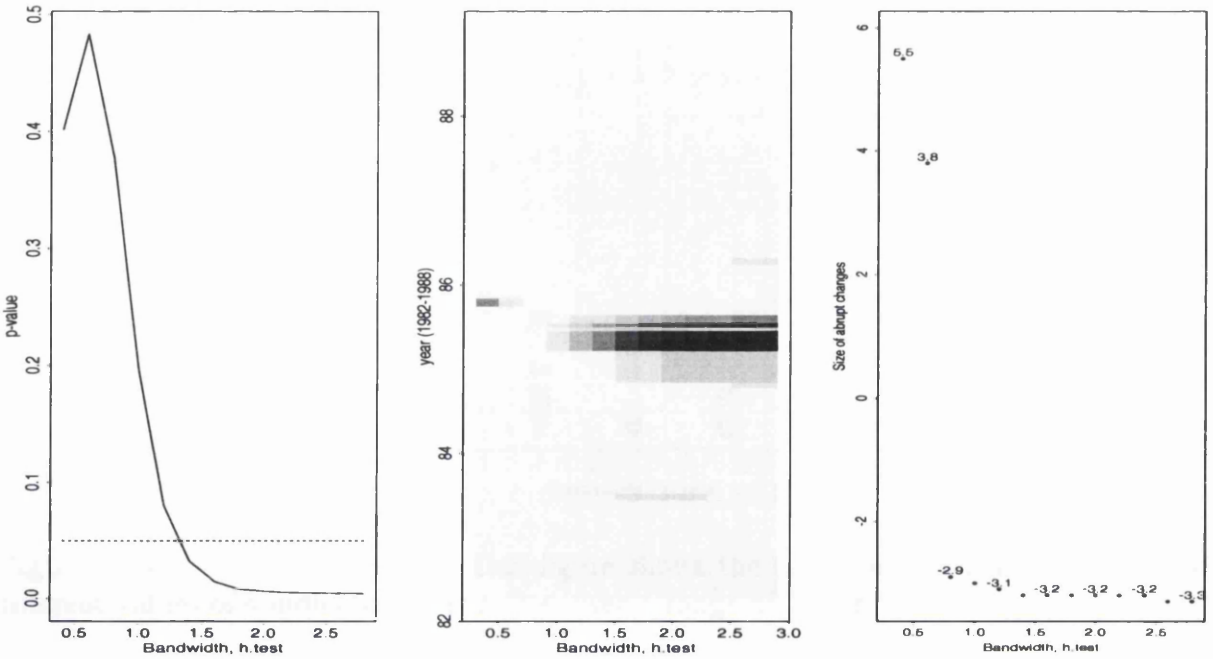

Figure 5.32. River Clyde data: A moving window of size $b = 15$ is used to estimate the correlation of the data. The left plot is a significance trace over a range of *h.test* from 0.4 to 2.8. The middle plot is the change-point locator plot. The plot at the far right shows the size of the most probable abrupt change.

size of moving window	estimated correlation
4	-0.293
7	-0.134
10	-0.088
13	-0.089
15	-0.081
16	-0.069
19	-0.067
22	-0.038
25	0.001
28	0.069

Table 5.10. River Clyde data: This table shows the estimated correlation coefficient of an AR(1) model, using the moving window approach. The estimated correlations are all relatively low, and are negative for $b < 25$ and positive for $b \geq 25$.

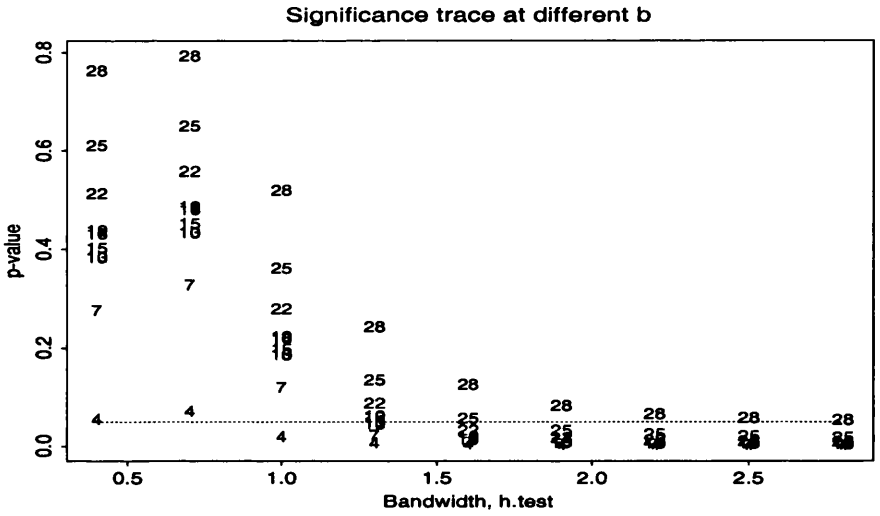

Figure 5.33. River Clyde data: This figure shows the significance traces computed with different values of b indicated on the plot, over a range of *h.test*.

At first sight, the significant results for $b < 25$ for a wider range of *h.test* might seem surprising in comparison to the independent case. However, this is because using $b < 25$ gives negative correlations. The power of the test to detect discontinuities is better if the data are negatively correlated. This gives more convincing evidence of the presence of an abrupt change in mid 1985, as the test is significant over a wider range of smoothing parameters.

5.4.3.4 Modelling Correlation - Unequally Spaced Setting

As the data are in fact unequally spaced, it would be more appropriate to model the underlying correlation using the variogram approach, as we will attempt to do in this section. First, we will use the residual variogram approach introduced in Section 3.2.7.

As displayed at the top left panel of Figure 5.34, a small bandwidth, $h.trend = 0.4$, is used to remove the underlying trend of the data. This gives the residuals as seen in the adjacent plot. An empirical variogram is then produced where a theoretical exponential covariance structure is fitted by non-linear least squares to obtain an estimate of the underlying correlation structure. The range and sill are estimated as 0.016 and 3.23 respectively, using $lag = 0.2$, $maxdist = 5$ and $minpairs = 30$. For a range of 0.016 of an exponential

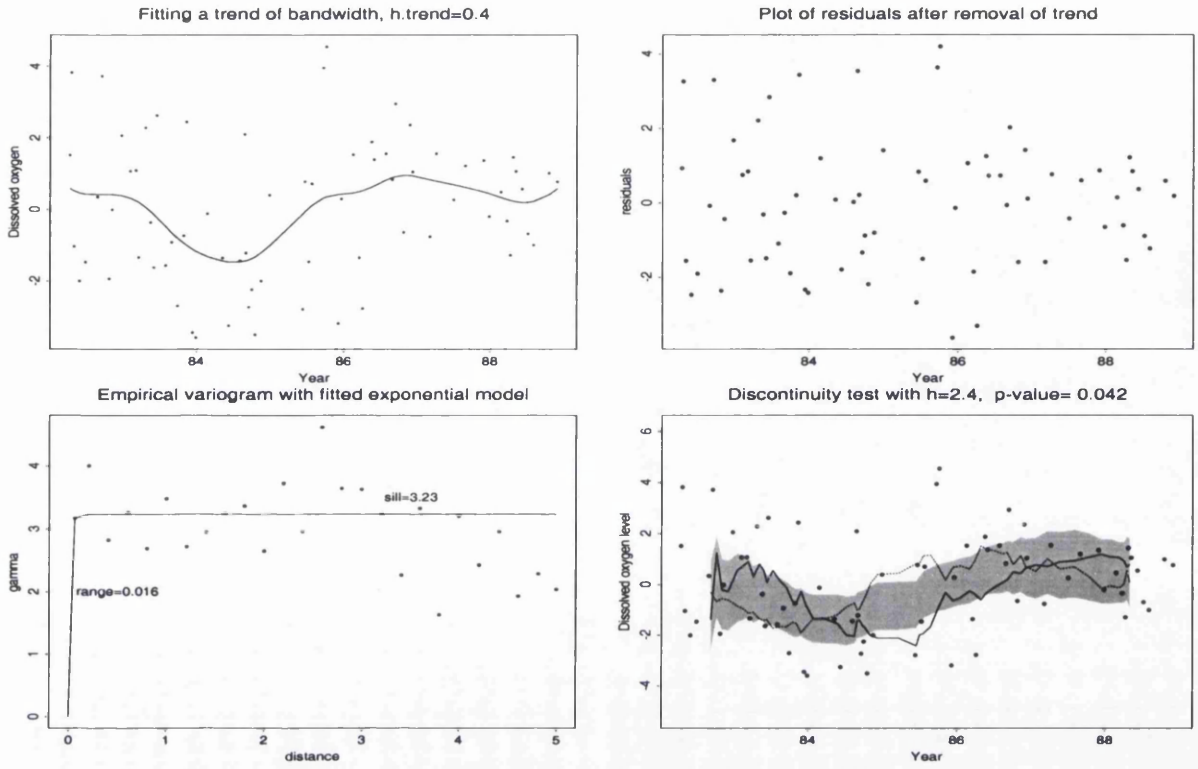

Figure 5.34. River Clyde data: An example to illustrate the estimation of the dependence structure via the variogram approach. We use *h.trend* = 0.4 and *h.test* = 2.5 for the discontinuity test. This gives a p-value of 0.042, which is significant indicating the presence of an abrupt change.

model, the apparent range is 0.048 indicating a low correlation where only data within half a month are dependent.

The discontinuity test is then carried out, accounting for correlation, to see if there is any presence of discontinuity. A p-value of 0.042 is obtained, thus providing evidence of a discontinuity which appears to occur in mid 1985, as can be seen in the bottom right panel of Figure 5.34.

As shown earlier in the simulation study, the bandwidth to remove trend, *h.trend*, has a great influence on the power to detect a change-point. Since our discontinuity test is only dependent on the range and not the sill of the variogram, we will just concentrate on the range that is estimated when different *h.trend* are used to remove the trend. This is illustrated in Figure 5.35 with the corresponding estimated ranges for each *h.trend* given

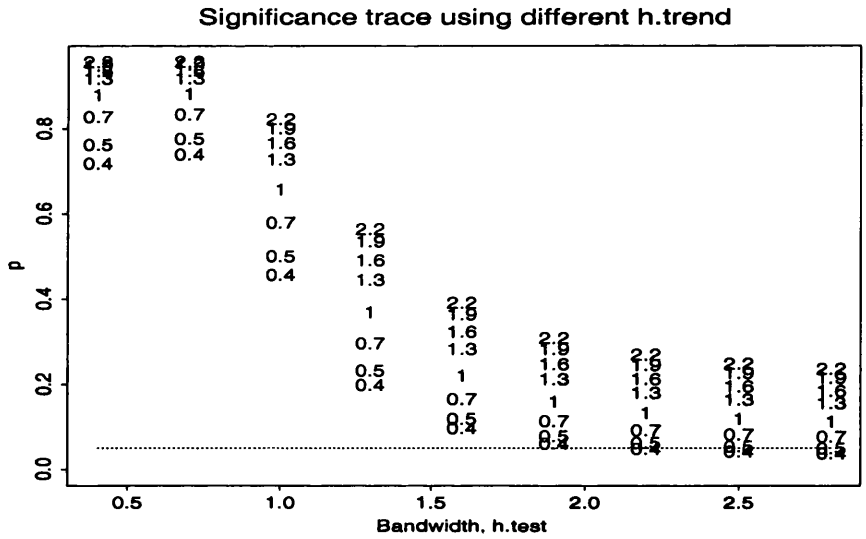

Figure 5.35. River Clyde data: The figure displays a significance trace using different *h.trend*, as denoted on the plot, to obtain residuals which are then used to estimate an exponential covariance structure.

bandwidth, h.trend	estimated range
0.4	0.016
0.5	0.019
0.7	0.024
1	0.029
1.3	0.034
1.6	0.037
1.9	0.040
2.2	0.042

Table 5.11. River Clyde data: This table shows the estimated ranges of an exponential ocvariance structure using different h.trend

in Table 5.11. The test is only significant for *h.test* > 2 when a small *h.trend* of 0.4 is used. Using the other values of *h.trend*, the test is not significant over the range of *h.test* considered.

An alternative approach to incorporating the estimation of correlation in an unequally spaced setting is to use the DDT algorithm. The estimated correlation using different smoothing parameters for the first stage, $h1$ = *h.trend*, is provided in Table 5.12. We

bandwidth, *h.trend*	**estimated range**	
	$lag = 0.16$	$lag = 0.2$
0.4	0.017	0.016
0.7	0.021	0.024
1	0.025	0.029
1.3	0.028	0.034
1.6	0.030	0.037
1.9	0.025	0.028
2.2	0.026	0.029

Table 5.12. River Clyde data: This table shows the estimated ranges of an exponential ocvariance structure using different $h1 = h.trend$ for the first stage of the DDT algorithm.

have also investigated the use of different lags to examine how much this factor influences the estimation. The rest of the parameters, *maxdist* and *minpairs*, are similar to those used earlier.

As we can observe from Table 5.12, the estimated ranges with $lag = 0.2$ are similar to Table 5.11 (obtained earlier using the residual variogram approach) until $h.trend = 1.6$, and are much lower for subsequent *h.trend.* This is not surprising since the adjustment of the data occurs only when the test is significant for $h1 > 1.6$ for the first stage of the DDT algorithm. This demonstrates the usefulness of the DDT algorithm and how it avoids overestimation of the correlation in the presence of abrupt changes.

Nevertheless, using the DDT algorithm in this data set, does not improve the result of the discontinuity test already obtained using the residual variogram approach, as the test is only significant if the estimated range is less than 0.02 over large *h.test.*

From Table 5.12, the estimation of the correlation is shown to be very sensitive to the parameter *lag* that is used. It is not possible to estimate the range at all using lags of 0.1 to 0.15, but it is possible with values of 0.16 and 0.2, which involve more averaging and may not give a good estimation of the actual range. There are some differences in the ranges estimated by the latter too.

The difficulty in obtaining a good estimate of the range may be due to the fact that the data are almost independent; or possibly negatively rather than positively correlated; or simply that the dataset is too small, resulting in too few observations nearer the origin, to

aid the estimation of the range.

An example to illustrate the use of the DDT algorithm is featured in Figure 5.36 with $h1 = h.trend = h2 = 1.7$, using lag of 0.2. The left panel shows the first step of the algorithm, carried out assuming that the data is independent with $h.test = 1.7$, which gives a significant p-value of 0.028. The data are then adjusted at the most probable change-point located at mid 1985. The top right panel shows a smooth curve fitted using the same bandwidth used to remove the trend in the data. The residuals thus obtained are used to compute the empirical variogram as shown at the bottom left panel. An exponential variogram model is fitted by nonlinear least squares to obtain the range and sill as 0.028 and 3.31 respectively. The discontinuity test is carried out again, accounting for correlation. This gives a non-significant p-value of 0.18. From the significance trace, the test is not significant over a wide range of $h.test$ values considered. However, when $lag = 0.12$ is used instead, the data are almost independent. (It is not possible to fit the exponential variogram model as the variogram estimate for the first bin is higher than that of its three nearest neighbours). This is displayed in Figure 5.37. The data are then assumed to be independent. The corresponding significance trace is similar to that in the independent case seen earlier.

The assumption of normality of the detrended data using $h.trend = 0.7$ and 1.7 seems reasonable, as displayed in Figure 5.38.

5.4.4 Summary of River Clyde data analysis

In the analysis of the River Clyde data, which is irregularly spaced over time, the seasonal patterns are first removed before the residuals examined for any change-point in the data.

The evidence from the isotonic regression is not very strong. Results from the tests using the unpenalised and penalised isotonic regression functions do not agree. Only the former is just significant, suggesting a change-point in mid 1985. The test might have been affected by the decreasing trend before 1985. There is also insufficient evidence to support any presence of a discontinuity using the Bayesian approach.

Lastly, the nonparametric regression approach was applied to the data. From the test

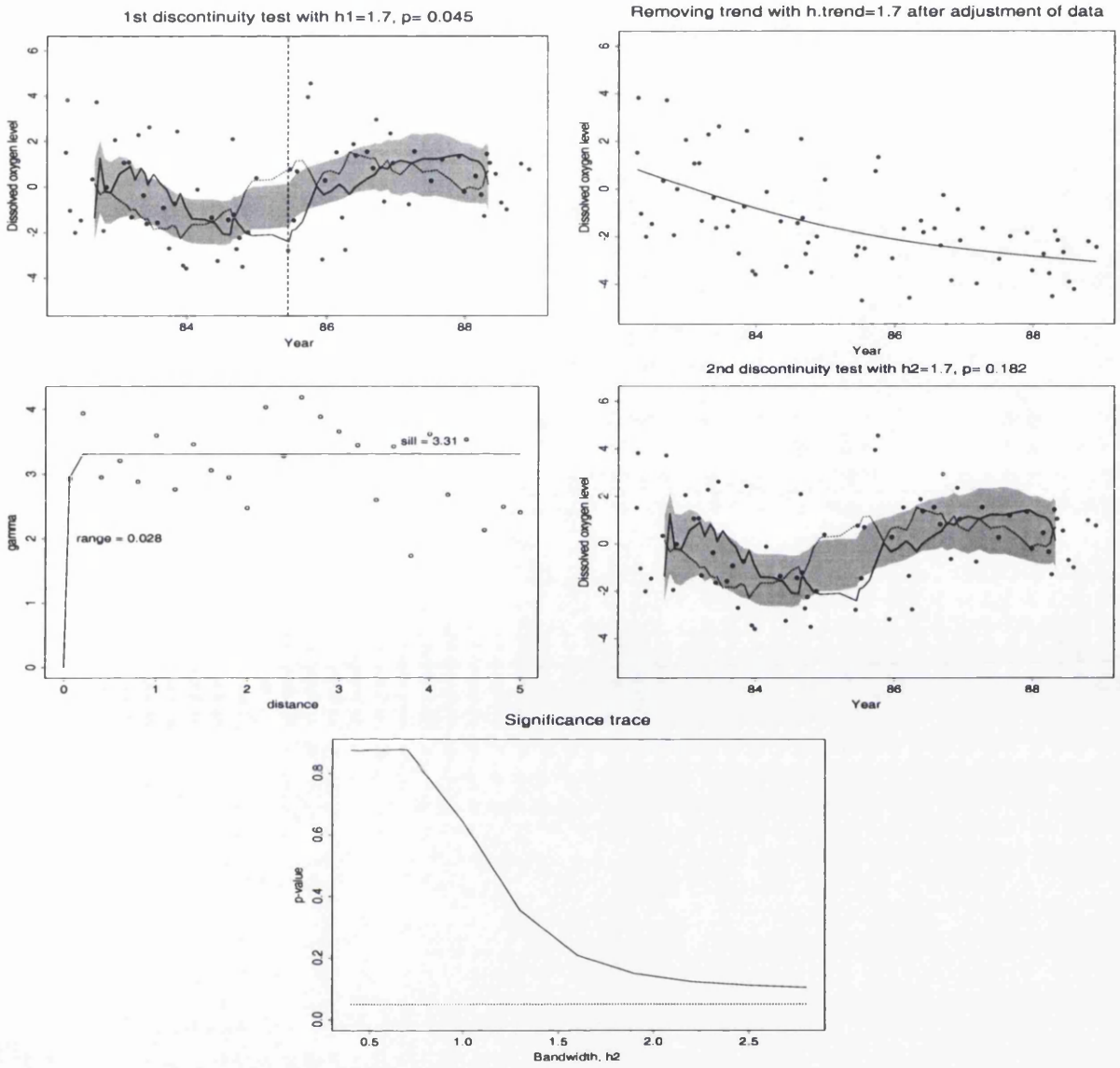

Figure 5.36. River Clyde data: Using the DDT approach with $h1 = h.trend = h2 = 1.7$. The top left panel shows the 1st stage of DDT being carried out, which gives a significant p-value. The vertical line indicates the most probable change-point located, i.e. mid 1985. The top right panel shows the adjusted data with a fitted smooth trend. The middle left panel shows the empirical variogram with a fitted exponential model. The middle right panel displays the 2nd stage of the DDT test. The bottom panel shows the significance trace for different values of $h2$

of constant variogram, most of the p-values using different values of the parameters are not significant, hence suggesting no presence of dependence structure. This in fact agrees with the ACF plot of the detrended data obtained via the isotonic regression approach. However,

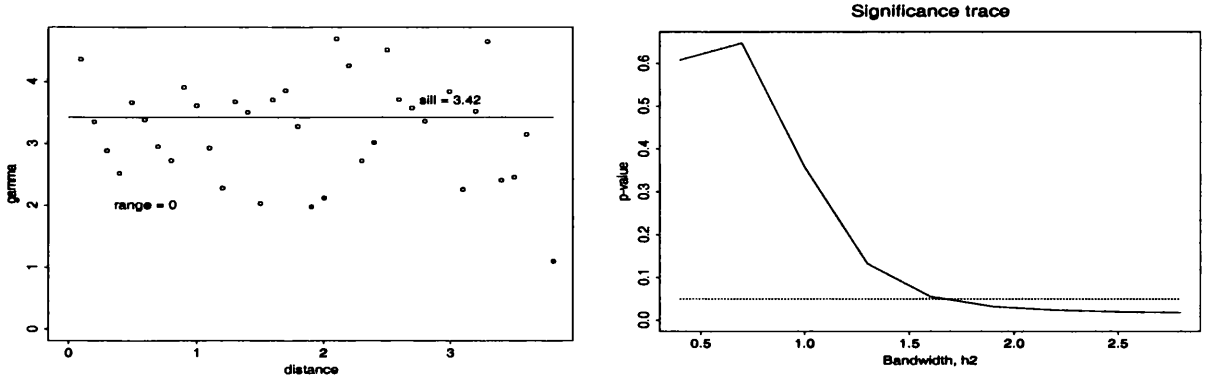

Figure 5.37. River Clyde data: The left panel shows the empirical variogram with a fitted exponential model using $lag = 0.12$. The right panel displays the significance trace over different $h2$.

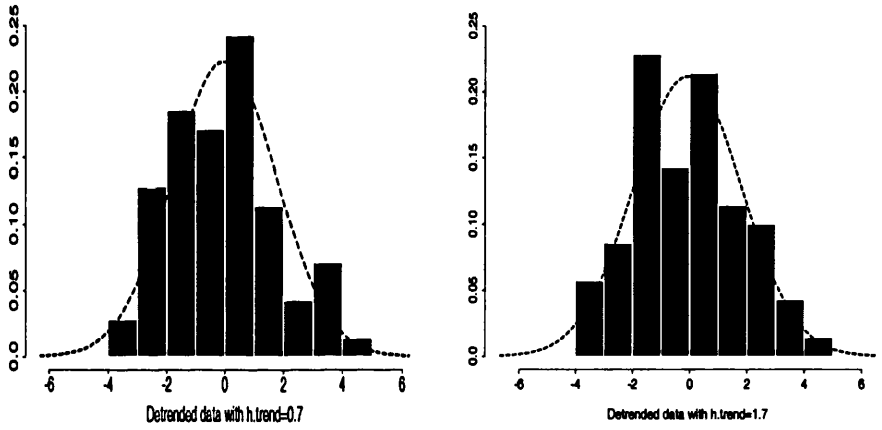

Figure 5.38. River Clyde data: Checking normality of data. The left and right panels show histograms of detrended data using smoothing parameter $h.trend = 0.7$ and 1.7 respectively, superimposed with their corresponding expected normal density curves.

the sensitivity of the test to the parameters used in the algorithm makes it not entirely conclusive. Hence, the test is then carried out both assuming that the data are independent, as well as incorporating the modelling of the correlation.

There is not much difference in the analysis assuming that the data are independent and modelling the correlation using the moving window approach, though the latter shows slightly more convincing significant results if the estimated correlations are negative. There

is some difficulty in modelling the correlation using an exponential covariance structure in this setting, due to its high sensitivity to the lag parameter that is used. Taking into account the irregularity of the observations across time with the very small number of observations may complicate the estimation procedure using the variogram. This is particulary the case if the dependence is weak, as is possible in this situation.

Overall, the difficulty in detecting any abrupt changes in this analysis could be because the dataset is relatively small with only about 70 observations. There also appears to be quite a lot of variation. The fact that it is irregularly spaced over time also complicates the issue further. Nonetheless, overall, using the nonparametric regression approach, there is convincing evidence to suggest that there is a discontinuity in mid 1985 if the trend can be assumed to be relatively smooth and in the presence of low positive or negative correlation. Hence, we can conclude that the improvement in the sewage treatment plant has significantly improved the quality of the water in the River Clyde. A similar conclusion is established by the isotonic regression approach using the unpenalised isotonic function. This improvement is however, not detected by the Bayesian approach, which is possibly affected by the decreasing monotonic trend present before and after the change-point.

5.5 Discussion

In this chapter, we have looked at three different approaches that have been adapted to analyse correlated data in three data-sets.

As there are differing features in each approach, the conclusion drawn from each test will undoubtedly be different. For instance, the mean is expected to be constant under the null hypothesis for both the models used by Wu et al. (2001) and Thomas (2001), while the underlying model used by the nonparametric regression approach is smoothly varying.

Concurring with Wu's analysis, the isotonic regression test, using both the unpenalised and penalised test statistics, is significant for both the Argentina data and global warming data, where the former is due to an abrupt change, whilst the latter is more likely due to an increasing trend in temperature over the years. The test is however just significant for the River Clyde data using the unpenalised likelihood ratio test, but not with the penalised one.

We note that the applications of the isotonic regression test as implemented by Wu et al. (2001) on both the Argentina rainfall data and the global warming data are unlikely to be appropriate as the assumption of monotonicity might be in doubt. In particular, it is obvious that though it is apparent that there is a general increase in global temperatures over 145 years, there are also periods of cooling. It is important that the restrictive assumption of monotonicity should be carefully checked to determine the suitability of this technique before carrying out the analysis.

On the other hand, using the Bayesian approach, a change-point is interpreted as an abrupt change in one or more combinations of the mean, variance, or correlation. It is shown to be very sensitive to the segments considered. There is only sufficient evidence to suggest that an abrupt change has occurred in the global warming data, but not in the other two data-sets. This could possibly be due to the fact that the Bayesian test is affected by the trend present in the data.

The nonparametric regression approach, however, interprets a change-point as an abrupt change in the mean structure that is given the flexibility to be smoothly varying. The degree of smoothness is controlled by the smoothing parameter. The flexibility given to the underlying trend increases its applicability and seems more appropriate in the environmental applications seen in this chapter, rather than the assumption of monotonicity or constant trend. As expected, the test is sensitive to the smoothness that one is willing to attribute to the trend function. In the case of the Argentina data, the test is only significant if the trend is moderate, whereas for the River Clyde data, only a high smoothing parameter give a significant result. The global warming data however, is less sensitive to this parameter, as it is significant over a wider range of smoothing parameter.

The test will also undoubtedly be sensitive to the values of the parameters used in the estimation of the correlation. Without detailed knowledge of the context of the problem, it is very difficult to decide on the level of correlation in the data. An overestimated correlation will reduce information in the data, and hence increases the difficulty in detecting the discontinuities in the trend. On the other hand, an underestimated correlation will result in false alarms. In the analysis of the Argentina and global warming data, the choice of the moving window size is recommended to be small, about 7, so that it will be robust to the

trend and abrupt changes present. The choice of b is not so crucial in the River Clyde data, as most values give negative correlation.

Nonetheless, the main aim of the proposed test is not to give a definitive answer as to the choice of the smoothing parameter, or the parameters used in the estimation of the correlation. It is hard to determine what is the best value of smoothing parameter to use. Methods for bandwidth selections such as cross validation, break down even in the simple case of independent data when abrupt changes are present. This is because it will select a smoothing parameter that is smaller than it really should be. Without detailed knowledge of the context of the problem, such as the meteorological systems that might be involved, it is very difficult to attribute a level of correlation to the data, as well as the degree of smoothness of the trend. Instead, we have introduced the significance traces to allow the user to visualise and interpret the sensitivity of the test to the different parameters that are used for the estimation of correlation. The change-point significance trace is also proposed as a graphical tool to illustrate how the significance of the test, the locations of change-points and the size of the most probable change might vary, over different smoothing parameters, at a certain level of estimated correlation.

Chapter 6

Testing for Discontinuities in Two-dimensional Correlated Data

6.1 Introduction

In the previous chapters, we have so far been confined to looking at discontinuities in the univariate setting, where the response y depends on only one covariate, x. In many practical applications, the behaviour of y might be dependent on more than one covariate. There has been increasing interest in detecting and estimating discontinuities in regression with two covariates, due to its wide variety of applications, for instance in image processing, spatial analysis, oceanography, meteorology, and other areas. As it is often difficult to obtain a suitable parametric form to model the functional relationship between the responses and the two covariates in surface regression analysis, which may be very complicated, smoothing by nonparametric regression is a very useful tool to adopt here.

Discontinuities in the one-dimensional setting are very different from the two-dimensional case. The search in the univariate setting is for just one or more discontinuities in x, but in the bivariate setting the discontinuities are considered to take the form of an abrupt change along an entire curve of unknown location and shape. Identification of the discontinuity thus involves estimation of both its position as well as its direction. The locus of the discontinuity is frequently referred to as the *jump location curve.*

In image analysis, there has been a lot of research on edge detection and edge preserving image reconstruction as they are very important issues (Bracewell, 1995). To obtain a good overview of computer edge detection, one could refer to Peli and Malah (1982), Torre and Poggio (1986) and Gonzalez and Woods (1992). In fact, edge detection is closely linked to jump detection, and edge preserving image reconstruction is similar to jump preserving surface recovery in nonparametric statistics. At the boundary of the object, the image intensity function can be regarded as a regression surface having step discontinuities between the object and its background. This is commonly referred to as step edges. There are many other forms of edges discussed in this field, but we will focus on the step edge which is similar to the jump location curve of the regression surface.

Much of the earlier work in the statistical literature has focused on the estimation of the jump location curve. Korostelev and Tsybakov (1993) and Tsybakov (1994) assumed the presence of a jump location curve and investigated the estimation of different boundaries of the object in an image using piecewise polynomials, where the polynomial coefficients are estimated by maximum likelihood. Their estimators are shown to reach the minimax optimal rate of convergence. Rudemo and Stryhn (1994) studied two types of two region image models with univariate boundary representation and suggested a "nonparametric histogram-like contour estimator". O'Sullivan and Qian (1994) detect object boundaries by defining a contrast statistic. They assumed that there is only one jump location curve which is a "smooth, closed" curve. Wang (1998) proposed a method based on wavelet transformation to estimate jump and sharp cusp curves of a function on a plane. Hall and Rau (2000) suggested a different approach, using a sequential, "tracking" algorithm to estimate a smooth fault line.

Other work includes Müller and Song (1994) which assumed that the number of jump location curves is known and that they are closed curved sets. Qiu (1997a) proposed the rotational difference kernel estimator when there is only one discontinuity curve. Both of these approaches use two one-sided kernel smoothers constructed from data on either side of a line of a particular direction at a point of interest. As this procedure involves a search over all directions for a possible jump direction at each design point, it is computationally very intensive. To present a technique that is less demanding computationally, Qiu and Yandell (1997) and Qiu (1998) suggested using coefficients obtained by fitting a local least squares

plane at the point of interest, giving a possible angle of orientation of the discontinuity.

Jose and Ismail (2001) assumed that the jump curve is at least first order differentiable and that it does not need to be of any explicit form. They proposed a simpler procedure, using information from opposites sides of the four quadrants of a neighbourhood of a point of interest, to fit kernel weighted least squares regression estimators. In other words, two regression smoothers are fitted at directions perpendicular to each other.

In a recent paper, Qiu (2002) also avoided the maximization approach by considering differences in kernel estimators at a number of specified directions at a particular point of interest. They performed a simulation study to investigate how the estimator of the jump location curve performs as the number of directions is increased. The number of jump location curves can be unknown.

Qiu and Bhandarkar (1996) proposed an edge detection algorithm based on the combination of local smoothing and hypothesis testing. At each pixel location, a 9x9 mask is centered, and a search for a grey-level discontinuity at four different directions is conducted. If it exceeds the evaluated threshold value, the point is then flagged as a step edge pixel. In this setting, the jump location curve is regarded as a *point set* in the design space and the aim is to find these step edge pixels. This strategy to detect jumps by regarding the discontinuity curve to be a set of points, gives more flexibility in the shape of the jump location curves and it often does not require the number of curves to be assumed. Other papers such as Qiu and Yandell (1997), Qiu and Yandell (1998), Qiu (2002) and Jose and Ismail (2001) have also adopted this approach. The other main strategy to detect discontinuities is to search for a particular form of jump location curve from a population of all possible candidates (Müller and Song, 1994, Qiu, 1997b, O'Sullivan and Qian, 1994, Korostelev and Tsybakov, 1993, Rudemo and Stryhn, 1994). This, however, requires stricter assumptions on the jump location curve, such as the number of curves is assumed known.

Other popular edge detection techniques include the Markov random field approach (see Geman and Geman, 1984, Besag, 1984, Besag et al., 1995). Qiu (2002) provides a good list of references to other techniques.

The papers above are concerned with the estimation of the jump location curve, (commonly referred to as jump or edge detection) of a discontinuous regression surface, assuming

that the data are independent. Very little work has been devoted to inference in the global context; to test if there is a discontinuity. Moreover, in the spatial setting, there is likely to be spatial dependence. Earlier work in Chapter 3 on discontinuity testing in autocorrelated time series data has shown that not accounting for the correlation present in the data or underestimating the underlying correlation will result in a higher rate of false alarms in hypothesis testing. It might also result in an inaccurate estimation of the position of discontinuity.

In an unpublished paper, Bowman et al. (2003) explored a global test to detect discontinuities in the two-dimensional setting, assuming that the data are independent. In view of the wide applications, it makes it very profitable to propose a test procedure which is able to account for dependence in the data. We will build on their work to detect discontinuities, taking spatial dependence into account.

6.1.1 Concept of spatial dependence

In the two-dimensional setting, we are dealing with observations of spatial data, which may be dependent as with time series data. In the latter case, later observations may depend on the earlier ones; this dependence is commonly referred to as autocorrelation. The ordering of the data is clear here. On the other hand, spatial data may be viewed as observations taken at points on a surface, since this occurs in two-dimensions. One can expect that data that are located closer together in space tend to exhibit greater similarity than those separated by larger distances. We can thus define spatial dependence as the tendency of a variable to exhibit similar values as a function of distance between the spatial locations at which they are obtained.

There is a body of tools specifically designed to explore spatial information, collectively known as geostatistics (Issaks and Srivastava, 1990). Geostatistics has a very wide domain. We will focus on modelling spatial dependence. In Chapter 4, we used the variogram to model the correlation/covariance structure of the one-dimensional unequally spaced data. The variogram is in fact more commonly used to model spatial dependence in the two-dimensional setting. By fitting a theoretical variogram model to the empirical variogram,

we can obtain an estimate of the range and sill which give the degree of spatial dependence present in the data. The higher the range and sill, the stronger is the correlation and the variability of the data respectively. In this work, the term "spatial dependence" refers to the covariance structure, comprising both the range and sill, while the term "spatial correlation" simply refers to the correlation structure, comprising only of the range. We will adopt the same technique as discussed in Section 4.1.2 to model the spatial dependence structure, using an exponential model.

Similar to the one-dimensional case, a positive correlation function like the isotropic exponential model should be widely applicable since it models a smooth decrease in spatial correlation as the distance between the observations increases. Moreover, as this is just for an inference end, a relatively simple model that is sufficient to capture most of the spatial correlation present is appealing.

The format of this chapter is, firstly, to give an introduction to the methodology, along with the test statistic and reference distribution. This is then followed by an extensive simulation study in the equally and unequally spaced setting. Lastly, the methodology is illustrated by analysis of a set of spatial radioactivity data undertaken to map cesium fallout in a chosen area in Finland after the nuclear reactor accident in Chernobyl (RESUME-95, 1997).

6.2 Methodology - Nonparametric Regression Approach

In this section, we will extend our discontinuity testing to a nonparametric regression model with two covariates. The bivariate case has numerous features which makes it very profitable to explore. From a methodological point of view, theoretical properties such as asymptotic bias are known. On a practical point, the scope of this class of models is very wide, comprising different forms of nonparametric regression function, ranging from a complete linear plane to a smoothly varying surface, with different degrees of smoothness. This pleasant property increases the flexibility and applicability of the model and makes it very attractive in real life applications where the underlying trend structure is rarely known over the spatial locations involved.

6.2.1 Statistical Model

Consider samples of n triplets of variables (x_{1i}, x_{2i}, y_i), where observations $\{y_i\}$ are related by

$$y_i = m(x_{1i}, x_{2i}) + \varepsilon_i \tag{6.1}$$

where (x_{1i}, x_{2i}), $i = 1, 2, \ldots, n$ are regularly spaced design points in $[0,1] \times [0,1]$, and ε_i are correlated and normally distributed errors $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n) \sim N_n(0, \sigma^2\Sigma)$, where σ^2 is the unknown (constant) error variance, and Σ is an $n \times n$ correlation matrix. The smooth bivariate regression function, $m(x_1, x_2)$ allows the surface to be smoothly varying, and is continuous over $[0,1]^2$ except at the jump location curves.

6.2.2 The test statistic

The main idea of the methodology proposed for the two covariate setting is similar to the univariate setting, which involves the differences between two smooth estimates. However, there is an added element of direction here. At a particular point, the difference between the left and right smooths changes as the direction changes. Even if the point lies on the jump location curve, the differences between the two smooths might not be large if it is not taken along the right direction (which is perpendicular to the angle of orientation of the discontinuity curve). Any test statistic proposed in this spatial setting will have to take this new parameter into account. As a result of this, there is an added complication to the whole test procedure. Any approach to estimating these directions using data-based methods might create problems in the distributional calculations.

Bowman et al. (2003) proposed a test statistic introduced in the next Section 6.2.2.1, which involves only a particular number of directions, θ, since the support of a jump is unlikely to change very much with θ. They suggested it will be sufficient to just consider four different directions, namely $\theta_k = \{k\pi/4, k = 1, \ldots 4\}$. The test developed here will build on their work, which is constructed to cope only with independent errors, and extend it to incorporate a dependence structure.

Letting the evaluation points lie on a regular grid, contained within the data points, expressed as $\{(z_{1r}, z_{2s}) : r, s = 1, \ldots g\}$, then using a linear smoother, the fitted values can be written simply as $\hat{m}(z_1, z_2) = Sy$ where S is a $g^2 \times n$ smoothing matrix of known constants. The computation of S will be discussed in greater detail in Section 6.2.4.

At each evaluation point (z_{1r}, z_{2s}), the differences between the two bivariate smooths is obtained at either side of the dividing line with direction θ_k, $0 \leq \theta_k \leq \pi$, where $k = 1, \ldots, p$. If there is a discontinuity at that particular point of evaluation and direction, one would expect the difference between the left and right smooths, $\hat{m}_{L\theta_k}(z_{1r}, z_{2s}) - \hat{m}_{R\theta_k}(z_{1r}, z_{2s})$, to be large. The values of the differences at the evaluation points $\{(z_{1r}, z_{2s}) : r, s = 1, \ldots g\}$ with direction, θ_k can then be expressed in matrix notation as $(S_{L\theta_k} - S_{R\theta_k})y = D_{\theta_k} y$.

6.2.2.1 Global test

The global test in the two-dimensional setting is similar to the one-dimensional case. The main difference is that the search for any sign of discontinuity in the form of a jump location curve in the spatial context, involves both the position as well as the direction.

The hypotheses can be listed as below.

H_0: m is continuous over $(0, 1) \times (0, 1)$.

H_1: m is discontinuous on at least one particular jump location curve in $(0, 1) \times (0, 1)$.

The proposed test statistic consists of the sum of the squared standardised differences between the left and right smooths at each evaluation point, over four different angles of orientation, θ_k, where $\{\theta_k = k\pi/4, k = 1, \ldots, 4\}$. The idea is that if there is no discontinuity present, then all the differences will be small. However, if there is a discontinuity, some of the differences will be large. The test statistic is given by

$$TS_1(h) = \sum_{k=1}^{4} \sum_{r,s=1}^{g} \left(\frac{\hat{m}_{L\theta_k}(z_{1r}, z_{2s}) - \hat{m}_{R\theta_k}(z_{1r}, z_{2s})}{\hat{\sigma}} \right)^2 \tag{6.2}$$

where $\hat{\sigma}^2$ is the estimator of the noise variance, chosen such that it can be expressed in quadratic form $y^T B y$, where B is an $n \times n$ matrix as before. This will be discussed in greater details in Section 6.2.5. In matrix notation, the test statistic in Equation (6.2) can

be expressed as

$$\begin{aligned} TS_1(h) &= \frac{\sum_{k=1}^{4} y^T {D_{\theta_k}}^T D_{\theta_k} y}{\hat{\sigma}^2} \\ &= \frac{y^T \left[\sum_{k=1}^{4} {D_{\theta_k}}^T D_{\theta_k}\right] y}{\hat{\sigma}^2} \end{aligned} \tag{6.3}$$

The final expression $y^T \left[\sum_{k=1}^{4} {D_{\theta_k}}^T D_{\theta_k}\right] y$ is a quadratic form, and this allows the usual quadratic form calculations that we have seen in the earlier chapters to be applied.

As proposed earlier in Section 3.2.2, we can consider another test statistic whose variances at each evaluation point vary, depending on the amount of information that is available to produce the differences between the two smooths. These variances are given by the diagonal elements of $\hat{\sigma}^2(D_{\theta_k} {D_{\theta_k}}^T)$.

The test statistic is then formulated as

$$TS_2(h) = \frac{y^T \left[\sum_{k=1}^{4} {D_{\theta_k}}^T \Lambda' D_{\theta_k}\right] y}{\hat{\sigma}^2} \tag{6.4}$$

Λ' is the inverse of the diagonal matrix whose diagonal entries are those of $D_{\theta_k} {D_{\theta_k}}^T$.

Since the estimator of σ^2 can also be expressed in quadratic form, we will then have a test statistic which is a ratio of two quadratic forms

$$\frac{y^T A y}{y^T B y},$$

where the unknown variance of the data is scaled out.

6.2.2.2 Local test

In the situation where there is prior knowledge of a possible jump location curve, we can perform the discontinuity test, taking this knowledge into account. This simplifies the test, as the jump location curve with the appropriate angles of orientation in the comparisons are given. We then just obtain the differences between the two smoothers at the points and directions on the curve where the discontinuity exist.

The hypotheses can be listed as below.

H_0: m is continuous at the jump location curve over an interval $(0,1) \times (0,1)$.

H_1: m is discontinuous at the jump location curve over an interval $(0,1) \times (0,1)$

The test statistic is similar to the global one in Equation (6.3), but the matrix D will be different here, as the computation of differences between the two smooths is performed only at or near the proposed jump location curve.

6.2.3 Reference distribution of test statistic

As the test statistic can be expressed in the form $\frac{y^T A y}{y^T B y}$, we can make use of the quadratic form calculations to compute the p-value of the test, taking the correlation matrix Σ into account, for dependent data in the reference distribution calculations. This is previously mentioned in Section 3.2.3.

6.2.4 Choice of smoothing techniques

The same principle of local linear smoothing as described in the one-dimensional case in Section 3.2.4, can be extended to the two covariate case without much difficulty (see Fan and Gijbels, 1996). This presents an advantage of borrowing information from neighbouring points to obtain a smooth fit. However, with the increase in dimensions, we are faced with the widely known issue of the curse of dimensionality (Bellman, 1961). This causes the efficiency of the nonparametric surface estimator to decrease rapidly as the dimensions of the design space increases. In fact, as the number of dimensions increases, the sample size has to increase exponentially to maintain a good fit.

At a particular evaluation point, say (z_{1r}, z_{2s}), a local *plane* is fitted and the fitted value of $\hat{\beta}_0$ from the least square problem in Equation (6.5) gives the bivariate smooth estimate of the regression function at that point (z_{1r}, z_{2s}).

$$\min_{\beta_0,\beta_1,\beta_2} \sum_{i=1}^{n} \{y_i - \beta_0 - \beta_1(x_{1i} - z_{1r}) - \beta_2(x_{2i} - z_{2s})\}^2 w(x_{1i} - z_{1r}; h_1) w(x_{2i} - z_{2s}; h_2) \quad (6.5)$$

The kernel function w denotes a probability function with standard deviation h. A normal

density function is used here. In matrix notation, we denote by X_z an $n \times 3$ matrix whose ith row is $(1, x_{1i} - z_{1r}, x_{2i} - z_{2s})$, and W_z contains the normal product kernels. It is a diagonal matrix with its (i, i) element as

$$w_i = w(x_{1i} - z_{1r}; h_1)w(x_{2i} - z_{2s}; h_2);$$

then the solution of the least square problem to obtain $\hat{\beta}_0$ is expressed as

$$[1, 0, 0](X_z^T W_z X_z)^{-1} X_z^T W_z y$$

In the case where equal smoothing applied along the two coordinate axes of the covariates, then $h_1 = h_2$.

Ruppert and Wand (1994) derived the asymptotic conditional bias of $\hat{m}(x; H)$ at the boundary +in Theorem 2.2 of their paper, given by Equation (6.6).

$$E(\hat{m}(x; H) - m(x)|X_1, \ldots, X_n) \doteq \frac{e_1^T N_x^{-1}}{2} \int_{D_x, H} \begin{bmatrix} 1 \\ u \end{bmatrix} K(u) u^T H^{1/2} \mathcal{H}_m(x) H^{1/2} u du + o_p tr(H) \tag{6.6}$$

where $x = (x_1, x_2)$ in our notation; $H^{1/2}$ is the bandwidth matrix which denotes the amount of smoothing; $\mathcal{H}_m$ is the Hessian matrix which is an indication of the curvature of m at x in different directions. Provided that all the second order derivatives of m are continuous, then the leading term of the above equation should cancel when the expected value of the differences between the left and right smoothers under the null hypothesis is taken, since $m_L(x) = m_R(x)$. This gives

$$E(\hat{m}_L(x; H) - \hat{m}_R(x; H)) = o(tr(H)) = o(h_1^2 + h_2^2) \tag{6.7}$$

It follows that, since the sequence of bandwidth $H^{1/2}$ taken is such that $n^{-1}|H|$ and each entry of H tends to 0 as $n \to \infty$, with H being symmetric and positive definite, then $E(\hat{m}_L(x; H) - \hat{m}_R(x; H))$ is approximately 0. It follows then that the test statistic in Equation (6.4) is a

ratio of quadratic forms in Normal random variables with means approximately zero and with the same variance.

6.2.5 Choice of variance estimator

Similar to what was discussed in the univariate setting in Section 3.2.6, we hope to consider a variance estimator that can be written in quadratic form Y^TBY, so that the test will be independent of the true value of σ^2. Extending the approaches used in the univariate to the bivariate case is not straightforward, as we do not have a simple ordering of the response y_i in the latter setting.

In an unpublished paper, Bock et al. (2001) suggested a simple variance estimator making use of the residual sum of squares approach. Recall that in the univariate setting, from Equation (3.20),

$$RSS = \sum_{i=1}^{n}\{y_i - \hat{m}(x_i)\}^2$$

This can be easily generalised to a bivariate setting as

$$RSS = \sum_{i=1}^{n}\{y_i - \hat{m}(x_{1i}, x_{2i})\}^2$$

In matrix notation, this can be expressed as $y^T(I-S)^T(I-S)y$ where S is the smoothing matrix and $\hat{m} = Sy$ is the vector of bivariate smooth estimates. From the expression of the expected values of the RSS estimator given in Equation (3.22), the first term is the sum of the squared biases of the design spaces. These can also be written as $\sum_{i=1}^{n}\{m_i - E(\hat{m}_i)\}^2$ where m_i and $\hat{m}_i$ denote the regression surface and its estimate at the *ith* location.

In order to reduce the bias in the univariate setting, it was suggested that a small smoothing parameter should be used. The same idea follows in the two-dimensional case to reduce the bias. Instead of fitting a regression surface using a global bandwidth, Bock et al (2001) suggested assigning only substantial weights to the three nearest observations at a point

(x_{1i}, x_{2i}). These three nearest neighbours can be defined by evaluating the neighbour distances in simple Euclidean form and picking up those whose distances from the point of interest are within the third shortest, $d3$. In the situation where the two dimensions have different units, the distance between points should be standardised by the the standard deviations, s_1 and s_2, of the data in each dimension. This is expressed as

$$\sqrt{\left(\frac{x_{1i} - x_{1j}}{s_1}\right)^2 + \left(\frac{x_{2i} - x_{2j}}{s_2}\right)^2}$$

By setting the smoothing parameter of a normal kernel function as $(s_1 d_3/2, s_2 d_3/2)$, only the three nearest neighbours will be given substantial weights in computing an estimate $\hat{m}(x_{1i}, x_{2i})$ at the ith point of interest. Note that in this approach, different smoothing parameters have been used at each point, as this is determined by the distance of its third nearest neighbour, $d3$.

This RSS based estimator which only makes use of three nearest neighbours can then be standardised either via an external standardisation procedure or by standardising each individual contribution (Bock et al., 2001). For the latter case, this is expressed as

$$\hat{\sigma}_3^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{m}(x_i)}{v(x_i)}\right)^2 \tag{6.8}$$

where $v(x_i)$ denotes the weights attributed to the variance of $y_i - \hat{m}(x_i)$. In this way, a point which has neighbours closer to it, will contribute more to the overall estimated variance since it will contain more information. The weights $v(x_i)$ can be easily computed as the diagonal elements of $(I - S)(I - S)^T$.

Equation (6.8) can be written in matrix notation as

$$\hat{\sigma}_3^2 = \frac{y^T(I - S)^T \Upsilon'(I - S)y}{n} \tag{6.9}$$

where Υ' is a matrix filled with zero but with diagonal elements that is obtained by the inverse of a matrix with diagonal elements of $(I - S)(I - S)^T$.

Bock et al. (2001) performed a simulation study to compare this RSS based estimator

with other bivariate variance estimators, such as the difference based estimator proposed by Herrmann et al. (1995), which was an extension of Gasser's approach (Gasser et al., 1986) given in Equation (3.27). They showed that the RSS based estimator in Equation (6.8) performs better than the difference based estimator. This is because the latter is based on local interpolation using the Dirichlet tesselation which might be subject to greater variability if it draws in neighbouring observations that might in fact be quite far from the point of interest. Because of the favourable property of the RSS based estimator using three nearest neighbours, we have chosen to incorporate it in our test statistic.

6.2.6 Estimation of jump location curve

As mentioned earlier in Section 6.1, many authors have focused attention on obtaining a good estimate of the jump location curve. Though the main aim of this chapter is inference, it would be useful to produce an estimate of the jump location curve if the test is significant, to indicate where the possible discontinuities might be. This section is set aside for this purpose, not with the aim of obtaining an accurate estimation, but to provide a useful and user-friendly graphical visualisation of the location of the jump curves.

There are some similarities and differences between the techniques used to carry out the test procedure and to obtain the jump location curve. The former involves finding the differences between the left and right smooths on either side of the dividing line at a set number of directions θ where $\theta = \{0, \pi/4, \pi/2, 3\pi/4\}$. However, the latter involves firstly, finding the angle of the maximum slope at each evaluation point, (z_{1r}, z_{2s}), where $r, s = \{1, \ldots, g\}$ are located on a regular grid. The differences of the smoothers are then taken at the direction that is perpendicular to that obtained angle.

To obtain the angle of the maximum slope, we fit a local least square plane at each evaluation point (z_{1r}, z_{2s}).

$$\hat{y}_{rs} = \hat{\beta}_0 + \hat{\beta}_1(x_1 - z_{1r}) + \hat{\beta}_2(x_2 - z_{2s}) \tag{6.10}$$

By considering the polar coordinates where $x_1 = a\cos\theta_{rs}$ and $x_2 = a\sin\theta_{rs}$, with a as a

constant, we have

$$\hat{y}_{rs} = \hat{\beta}_0 + \hat{\beta}_1(a\cos\theta_{rs} - z_{1r}) + \hat{\beta}_2(a\sin\theta_{rs} - z_{2s}) \tag{6.11}$$

To attain the direction of the maximum slope, $\hat{\theta}_{rs}$ where $0 \leq \hat{\theta}_{rs} \leq \pi$, we find the turning points of $\hat{y}_{rs}$ in Equation (6.11) by differentiating it with respect to θ_{rs}, and setting the derivative equal to zero to obtain

$$\begin{aligned} -\hat{\beta}_1 a\sin\theta_{rs} + \hat{\beta}_2 a\cos\theta_{rs} &= 0 \\ \tan\theta_{rs} &= \frac{\hat{\beta}_2}{\hat{\beta}_1} \\ \hat{\theta}_{rs} &= \tan^{-1}\frac{\hat{\beta}_2}{\hat{\beta}_1} \end{aligned}$$

The direction of the "dividing line" at each evaluation point is then at $\hat{\theta}_{rs} + \pi/2$. The corresponding difference between the left and right smooths on either side of this dividing line is then given as $\hat{m}_{L\hat{\theta}_{rs}} - \hat{m}_{R\hat{\theta}_{rs}}$ In matrix notation, the estimates of the differences over all the evaluation points can be expressed as $(S_{L\hat{\theta}} - S_{R\hat{\theta}})y = D_{\hat{\theta}}y$. To obtain a smoother jump location curve, we then smooth these differences again, giving $SD_{\hat{\theta}}y$. We refer to this approach of smoothing the data twice as *double smoothing* (dsT). Its main function is to obtain a better estimate of the jump location curve, as it smooths out the differences between the left and right smooths with the view that the data are more sparse in the two-dimensional than in the one-dimensional case. However, there is a danger of oversmoothing if the underlying trend or jump location curve is quite irregular.

The standardised differences for this doublesmoothed data are then given by

$$\frac{SD_{\hat{\theta}}y}{\sqrt{Var(SD_{\hat{\theta}}y)}} = \frac{SD_{\hat{\theta}}\xi' y}{\hat{\sigma}} \tag{6.12}$$

where ξ' is the inverse of the matrix that is filled with zero but with diagonal elements $SD_{\hat{\theta}}\Sigma D_{\hat{\theta}}^T S^T$ and Σ is the correlation matrix. This computation takes into account the dependence structure of the data. If spatial dependence exists, but the data are treated as

independent, these standardised differences would be overestimated.

At locations where discontinuities exist, we would expect the differences between the two smooths to be high. To obtain a graphical illustration of where the discontinuities might lie, a contour plot of different levels of standardised differences with magnitude greater than 2 can be produced.

Finally, to illustrate this technique, a data set with range of 0.03 and variance of 1 is simulated with a jump of 2 for data within a circular jump location curve. Figure 6.1 displays the data of size 17×17. At each evaluation point, the direction of maximum slope, $\hat{\theta}_{rs}$ is evaluated, and the perpendicular to it gives the "dividing line", indicated by the small blue lines. The standardised differences of the left and right smooths are then taken on either side of the dividing line. The test is carried out assuming knowledge of the range of 0.03. It is significant at 0.006. Any standardised differences greater than or equal to 2.5 are then indicated by the contours (solid black lines). We can observe that the contour of level 2.5 is very close to the actual jump location curve.

The computation of these standardised differences as shown in Equation (6.12), involves the covariance structure, i.e. both the correlation and the variance of the data, which are often unknown and have to be estimated. The next Section 6.2.7 describes an estimation of this quantity which can simply be plugged into Equation (6.12) to obtain the standardised differences. Alternatively, we could use the estimated correlation but estimate the variance by means of the RSS based estimator instead of using the sill from the variogram fitting. The variance estimators mentioned in Section 6.2.5 are, however, only applicable for independent data. A slight adjustment for the estimator in Equation (6.9) in the correlated setting is easily computed as

$$\hat{\sigma}_4^2 = \frac{y^T(I-S)^T\Upsilon'_\Sigma(I-S)y}{n} \tag{6.13}$$

where Υ'_Σ is a matrix filled with zero but with diagonal elements that is obtained by the inverse of a matrix with diagonal elements of $(I-S)\Sigma(I-S)^T$.

This method of estimating the jump location curve in interval form is coordinate free

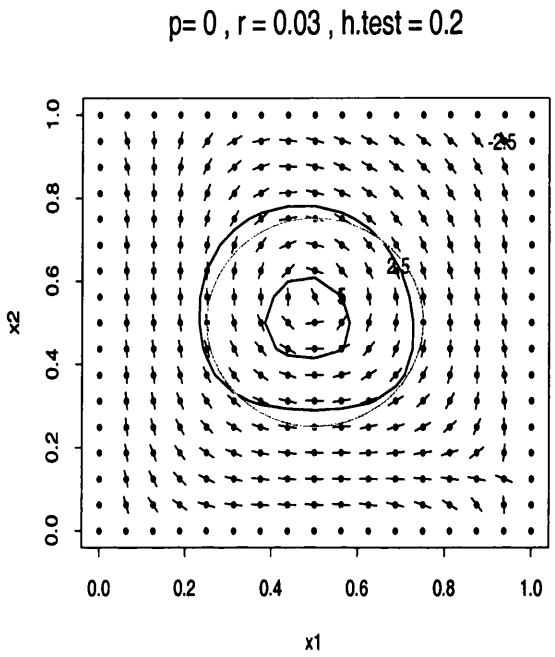

Figure 6.1. The figure displays a set of simulated data with range, $r = 0.03$. The blue lines indicate the dividing line at each evaluation point. Perpendicular to the line is the direction of maximum slope, $\hat{\theta}_{rs}$. The red curve indicates the actual jump location curve. Superimposed on them are the contours of standardised differences of 2.5 and 5, indicated by solid black lines.

and does not require that the functional form of the discontinuity curve is of any parametric structure or of closed form. Moreover this approach makes use of fitting least square planes (similar to Qiu and Yandell, 1997), which avoids the computationally very intensive maximization of the differences between the two one-sided kernels with respect to direction at each design point (Muller and Song, 1994; Qiu, 1997a). However, as mentioned earlier in Section 6.2.2, in the computation of the test statistic, it would not be possible to estimate suitable directions using data-based methods, as this might create problems in the distributional calculations.

6.2.7 Adjusting the test for spatial dependence

This section is devoted to introducing an algorithm to adjust the test, taking into account the spatial dependence that might be present in the data, by incorporating the estimation of this unknown quantity obtained via variogram fitting, into the test.

In the one-dimensional unequally spaced setting in Chapter 4, we proposed the two-stage double discontinuity test (DDT) algorithm to carry out the discontinuity test twice. However, estimation of the dependence structure in a two-dimensional spatial setting, where there might be both trend and discontinuities, is a highly complicated task. There is a need to create a new algorithm. This is because, unlike the one-dimensional case with DDT, it is impossible to adjust for the magnitude of the jump in the first stage in this more complex setting, if we do not have any knowledge of the jump location curve. Even if we do assume knowledge of the latter, we would also have to assume a constant jump enclosed within the jump location curve if we are to be able to adjust the size of the jump before estimating the correlation via variogram fitting. However, this would introduce more restrictions to the model, which might not be as applicable in real-life settings, compared to allowing the jump to vary smoothly. The algorithm that is introduced here therefore does not assume that the model has a constant jump. It is intended to cope with both the presence of trend and discontinuities which might affect the estimation of the dependence structure.

Similar to the DDT algorithm, the proposed method is a two-stage process, where the discontinuity test is carried out twice. The first stage involves carrying out the discontinuity test, treating the data as independent. A smooth surface is then fitted to remove the trend in the data, giving what is termed the "detrended data". If the p-value of the discontinuity test is less than 0.05, a partitioning process is performed, whereby detrended data that have standardised differences greater than a particular set value, say 2.5, are removed. This step is intended to remove any points that might be located near the jump location curve, which would affect the estimation of the correlation. The remaining residuals are then used to generate the empirical variogram which is modeled using nonlinear weighted least squares fitting to obtain the range and sill. The discontinuity test is then carried out again, taking the correlation into account, to obtain the p-value of the adjusted test. This algorithm is

The Double Discontinuity Test with Partition (DDTP)
1st stage:
Aim: Estimation of correlation

- Step 1:
  - Treating the data as independent, the discontinuity test is carried out with double smoothing, using a smoothing parameter of $h1$. The p-value and the standardised differences, $st.diff1$ obtained are carried over to the next step.
- Step 2:
  - The trend in the data is removed by fitting a smooth two-dimensional plane using the same smoothing parameter as before, $h.trend = h1$, to obtain the residuals.
  - If $p < 0.05$, we will partition the data into two different segments, according to whether the standardised differences, $st.diff1$ are smaller or larger than a particular set limit, *std*. Only the detrended data with $st.diff1$ less than *std*, will be used. This step removes any data points which might be located near the jump curve.
  - Using the detrended and partitioned data, we obtain an estimate of the correlation matrix, Σ, by weighted nonlinear least squares fitting of the variogram to get the range and sill.

2nd stage:

- Accounting for the estimated Σ, the discontinuity test is carried out with single smoothing using bandwidth, $h2$, to obtain the p-value of the test.
- If a graphical plot is required, the resulting standardised differences, $st.diff2$ which incorporates the spatial covariance structure (obtained via the variogram modelling), are used to produce a contour plot of possible locations of the jump location curve (see Section 6.2.6).

Figure 6.2. The Double-Discontinuity Test with Partition (DDTP)

referred to as the Double Discontinuity Test with Partition and the details are provided in Figure 6.2.

The algorithm recommends double smoothing for the first stage of the algorithm to give a better smooth. However, if the trend in the data is quite rough, double smoothing

might cause false claims of possible discontinuities at certain locations due to oversmoothing. Nonetheless, the algorithm uses this for the first stage, as by performing the first discontinuity test with double smoothing, a better estimate of the jump location curve is obtained. This is useful as it results in a more accurate removal of observations (with standardised differences greater than a specified value) near the jump location curve. These observations, if not properly removed, might result in an overestimation of the correlation, resulting in low power. Even if there might be some incorrectly removed observations, this would not prove to be much of a problem as long as there are sufficient remaining to be used for variogram fitting.

Doublesmoothing has also been proposed for the trend removal, as we have assumed the trend to be relatively smooth, in the presence of correlated data. The presence of correlation in the data creates a "false trend" in the data. As we want to avoid tracking the data too closely, it would be better to oversmooth. There is a valid concern that, if discontinuities exist, oversmoothing might result in an overestimated range, amidst other concerns. However, it is hoped, with the combination of the partitioning process, that observations near the jump location curve are removed.

In modelling the empirical variogram, we have used nonlinear weighted least squares to fit an exponential variogram, with $lag = 0.05$, lag tolerance $= lag/2$, $nlag = n$, distance used, $maxdist$=half of maximum pairwise distance or maximum distance, minimum number of pairs, $minpairs = 30$.

The algorithm has chosen to remove the trend using all the data before partitioning the data. The partitioning process could also be considered first, followed by trend removal using only the chosen segment of the data. However the former approach has been performed as it is of interest to look at sample sizes which are quite moderate in size. If a large amount of data is removed during the partitioning process, there may be problem in obtaining a good estimate of the trend present. In the case of a larger sample size, partitioning the data before removing the trend can be considered. The second stage of the test then involves carrying out the discontinuity test, adjusting for the estimated correlation, without double smoothing. The latter will be referred to as single smoothing (dsF).

Note that the standardised differences here, $st.diff1$ do not have any statistical meaning

if the data are dependent, as the covariance of the data is not taken into account, when computing this value. It is simply part of the algorithm to remove any unwanted observations. However, $st.diff2$ takes into consideration the spatial dependence and hence gives a better estimation of the possible locations of discontinuity.

The effectiveness of this algorithm depends not only on the sample size, the jump location curves, trend function, but also the size of the jump and its error variance. Its performance might differ even though the data might have the same signal to noise ratio. For instance, it might be easier to cope with data with a jump of 1 and noise of 0.5, compared to a jump of 2 and noise of 1. In other words, if the magnitude of jump is smaller, it might produce a more accurate estimate of the range.

6.2.8 Remarks

The methodology can be easily extended to irregularly spaced observations, or a regularly spaced grid with missing observations. The regular grid of evaluation points (z_{1r}, z_{2s}) can be used for the discontinuity test as well as to locate the jump location curve. Nonetheless, the standardised differences can also be computed on the actual data-points, rather than the evaluation points.

However, in order to carry out the partitioning process of the DDTP algorithm, it is essential that standardised differences, $st.diff1$ are computed at the actual data locations for both settings.

6.3 Simulation Study

In this section, we will investigate the influence of various factors on the discontinuity test. It is subdivided into two main sections: an equally spaced setting (Section 6.4) and an unequally spaced setting (Section 6.5), where both the scenarios of correlation known and unknown are considered.

The size of each test is assessed by generating sets of data where no discontinuity is present, while the power of each test is obtained from data with discontinuities at the jump

location curve. A 99% confidence interval based on a binomial distribution with $nsim$ =no of simulations and $p = 0.05$ can be used to assess the empirical size of the test for substantial departures from the specified $\alpha = 0.05$ level. In the setting where $nsim = 200$ and $p = 0.05$, the 99% confidence interval for the empirical size is $(0.015, 0.095)$.

We have kept certain parameters constant throughout the whole simulation study. These include using an isotropic exponential variogram model with error variance of 1, and a jump location curve, given by $(x1 - 0.5)^2 + (x2 - 0.5)^2 = 0.25^2$. The latter is a circle centered at $(0.5, 0.5)$ with radius of 0.25. The data conditions and the parameters used in the test approaches are summarised in Table 6.1. Those highlighted in bold are used by default in the simulations, unless specified otherwise. Note that single smoothing is used in the discontinuity test and double smoothing is only considered during the 1st stage of the DDTP algorithm.

6.4 Equally Spaced Setting

This section will investigate the performance of the test, when the data are equally spaced on a square grid. In Section 6.4.1, we will first assume knowledge of the correlation, and vary the other factors involved in the discontinuity test, while in Section 6.4.2, we will embark on the challenging task of estimating the unknown spatial correlation in different settings, using the DDTP algorithm introduced in Section 6.2.7. The data conditions and parameters used in the test are stated in Table 6.1.

6.4.1 Equally Spaced Setting: Correlation known

This section will assume knowledge of the correlation in an equally spaced grid. In each of the following sections, only the parameters of interest are varied, with the rest of the factors remaining unchanged. Section 6.4.1.1 examines the effects of trend while Sections 6.4.1.2 and 6.4.1.3 investigate the effects of correlation and sample size respectively.

Data conditions	
Design Space	- Equally-Spaced (for Section 6.4) - Unequally-Spaced (for Section 6.5) simulated from Uniform distribution, different for both axis
No. of simulations	**200**
Error	$\epsilon \sim N(\mathbf{0}, \sigma^2\Sigma)$
Error Variance	**1**
Trend Function	**Flat, Qiu**, Ismail
Correlation	exponential model, Σ, with range, $r = \{\mathbf{0.03}, 0.05\}$
Sample size	$n = 11 \times 11, \mathbf{17 \times 17}, 21 \times 21, 24 \times 24$
Size of Jump	*jump*, $jp = \mathbf{2}$, 3
Jump curve	$(x1 - 0.5)^2 + (x2 - 0.5)^2 = 0.25^2$, **red dotted line** as shown in Figure 6.1
Test Approaches	
No. of evaluation points	**ngrid = 15 × 15**
Smoothing Parameter	$h.test = \{\mathbf{0.1}, \mathbf{0.15}, \mathbf{0.2}, \mathbf{0.25}, \mathbf{0.3}\}$
Estimator of Error Variance	RSS based estimator, $\hat{\sigma}_3$
Estimator of Correlation	- None (for Section 6.4.1 and 6.5.1); - Double Discontinuity with Partition (DDTP) using $h1 = h.trend = \{0.2, 0.3, 0.4, 0.5, 0.6\}$, $h2 \in \{h1, h1 - 0.1, h1 - 0.2\}$, $std = 2.5, 3.5$, and *maxdist*= 0.7,1.4.

Table 6.1. Simulation setting for equally and unequally paced two-dimensional data. The parameters in bold are the ones used by default in the simulations, unless specified otherwise.

6.4.1.1 Effects of trend function and smoothing parameter

It is expected that the underlying trend in the data, and the degree of smoothness would affect the result of the discontinuity test. We have also seen this feature in earlier simulation studies in the one-dimensional case. Here, we will look at data with range 0.03, and with three different trend functions: Flat, Qiu and Ismail. These are given by $trend = 0, -0.5 - x2 + 3(x1 - 0.5)^2$ and $1 + 2sin(3x1 + 2x2)$ respectively. (The last two trend functions are used by Qiu and Yandell, 1997 and Jose and Ismail, 2001 respectively). These three functions are

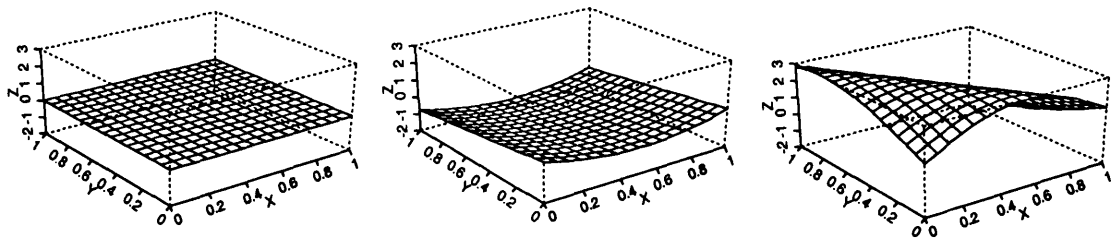

Figure 6.3. The figure displays the different trend functions: Flat, Qiu and Ismail, drawn over the same scale on a 17×17 evaluation points.

h.test	Size			Power		
	Flat	Qiu	Ismail	Flat	Qiu	Ismail
0.10	0.065	0.065	0.070	0.420	0.420	0.440
0.15	0.060	0.060	0.060	0.795	0.795	0.865
0.20	0.050	0.050	0.180	0.965	0.960	0.995
0.25	0.055	0.065	0.485	0.995	0.995	1.000
0.30	0.055	0.120	0.785	0.995	0.995	1.000

Table 6.2. Simulation results for $n = 17 \times 17$, $ngrid = 15 \times 15$, $range = 0.03$ and $jump = 2$ with three different trend functions. True correlation is assumed. Single smoothing is used.

displayed in Figure 6.3, drawn on the same scale. It can be observed that the Ismail trend is the roughest among the three, followed by the Qiu trend.

The simulation results are presented in Table 6.2. The size for the flat function is approximately 0.05 over all smoothing parameters, *h.test*. It remains within limits for the Qiu trend, except at $h.test = 0.3$. However, for a rougher function such as the Ismail trend, the size is only within limits for low smoothing parameter values of 0.10 and 0.15.

Powers for the Flat and Qiu trends are similar. They increase as the smoothing parameter increases. At sizes which are within limits, the powers for the Ismail trend are higher compared to the former two.

6.4.1.2 Effects of correlation

Here, we are interested to investigate how different degrees of spatial correlation would affect the performance of the test. This section also serves to give a useful reference to the later Section 6.4.2, where no knowledge of correlation is assumed, and has to be estimated. In situations where the correlation is estimated fairly well, the size and power should be similar to that if the true correlation is assumed known.

The simulation results with $r = 0.03$ and 0.05 under both trend functions are displayed in Table 6.3. Under the flat trend, the size for both ranges is within limits, and power increases as smoothing parameter increases. Power for $r = 0.03$ is however higher than for $r = 0.05$. Under the Qiu trend, size remains within limits for both ranges for all *h.test*, except at *h.test* $= 0.3$. Similar features for power as in the flat function can be observed, where power is higher for lower range. In fact, power for both ranges is similar under the Flat and Qiu trends, highlighting that the Qiu trend is fairly smooth.

h.test	**Flat Function**				**Qiu Function**			
	Size		**Power**		**Size**		**Power**	
	$r = 0.03$	$r = 0.05$	$r = 0.03$	$r = 0.05$	$r = 0.03$	$r = 0.05$	$r = 0.03$	$r = 0.05$
0.10	0.065	0.055	0.420	0.330	0.065	0.055	0.420	0.330
0.15	0.060	0.060	0.795	0.580	0.060	0.060	0.795	0.580
0.20	0.050	0.045	0.965	0.800	0.050	0.045	0.960	0.800
0.25	0.055	0.050	0.995	0.885	0.065	0.065	0.995	0.865
0.30	0.055	0.065	0.995	0.890	0.120	0.120	0.995	0.885

Table 6.3. Simulation results for $n = 17 \times 17$, range, $r = 0.03, 0.05$, $jump = 2$, $\sigma^2 = 1$ for both flat and Qiu's trend functions, with single smoothing.

6.4.1.3 Effects of sample size

In this section, the focus is on the effects of sample size on the test. Four sample sizes are considered here, namely $n = 11 \times 11, 17 \times 17, 21 \times 21$ and 24×24. Figure 6.4 shows the data points for these various sample sizes and their evaluation points (indicated by green dots). The number of evaluation points is constant at 15×15, which are regularly spaced out from

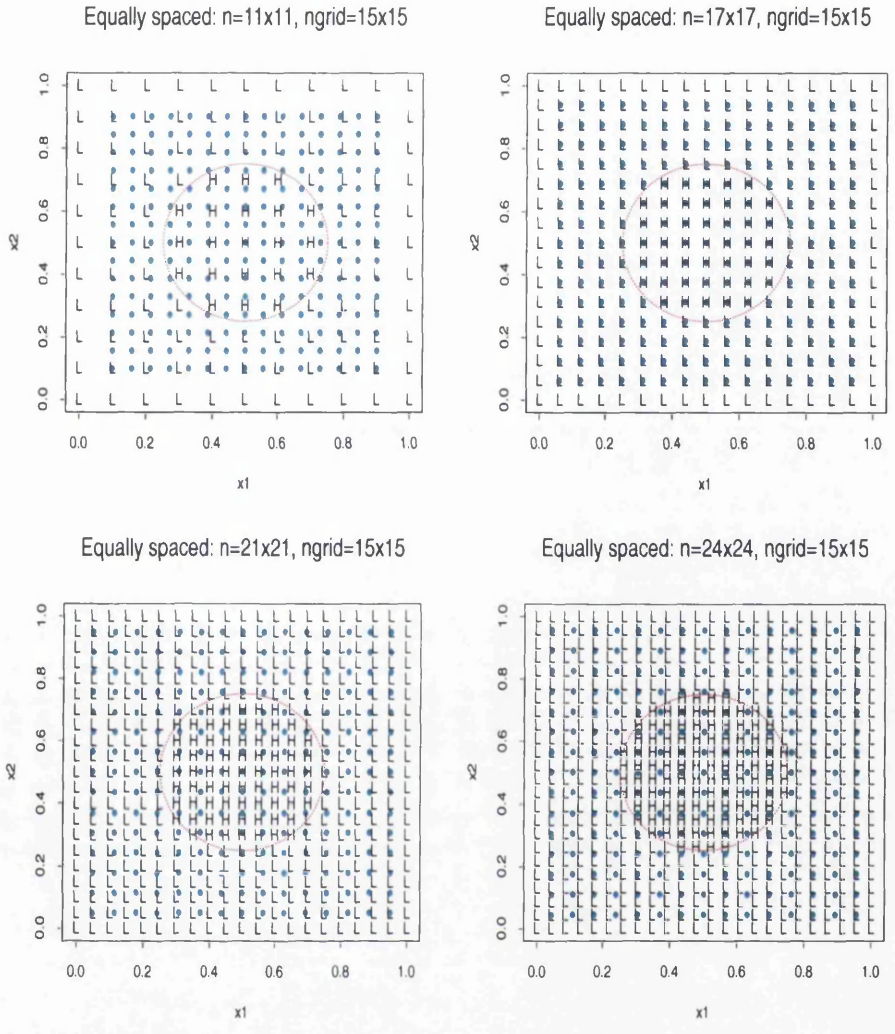

Figure 6.4. The figure displays three different sample sizes, $n = 11 \times 11, 17 \times 17$, 21×21 and 24×24. The evaluation points ($ngrid = 15 \times 15$) are represented by green dots. The red dotted line denotes the jump location curve, and the points with jump added are denoted as "H", while the rest are denoted as "L".

the second row and column to the second last row and column. We shall consider only the flat trend here, with two different correlations, over different smoothing parameters.

From Table 6.4, we can observe that for all sample sizes, the size for $r = 0.03$ remains within limits over all smoothing parameters. An interesting feature in the results for power is that the power for $n = 17 \times 17$ is higher compared to $n = 21 \times 21$. This effect is more prominent at lower *h.test*. However, both are considerably higher than at 11×11. Similar

h.test	n=11x11		n=17x17		n=21x21		n=24x24	
	Size	Power	Size	Power	Size	Power	Size	Power
0.10	0.045	0.110	0.065	0.420	0.025	0.260	0.050	0.590
0.15	0.055	0.335	0.060	0.795	0.040	0.770	0.035	0.885
0.20	0.055	0.755	0.050	0.965	0.055	0.970	0.050	1.000
0.25	0.050	0.800	0.055	0.995	0.055	0.990	0.050	1.000
0.30	0.070	0.960	0.055	0.995	0.050	0.995	0.050	1.000

Table 6.4. Simulation results for three different sample sizes, $n = 11 \times 11, 17 \times 17, 21 \times 21$, 24×24, range, $r = 0.03$ and $jump = 2$, with a flat trend function, with single smoothing.

h.test	n=11x11		n=17x17		n=21x21		n=24x24	
	Size	Power	Size	Power	Size	Power	Size	Power
0.10	0.040	0.095	0.055	0.330	0.025	0.160	0.050	0.370
0.15	0.045	0.285	0.060	0.580	0.040	0.445	0.040	0.585
0.20	0.050	0.655	0.045	0.800	0.055	0.715	0.045	0.820
0.25	0.075	0.800	0.050	0.885	0.065	0.850	0.060	0.945
0.30	0.075	0.855	0.065	0.890	0.065	0.875	0.080	0.960

Table 6.5. Simulation results for three different sample sizes, $n = 11 \times 11, 17 \times 17, 21 \times 21$, 24×24, range, $r = 0.05$ and $jump = 2$, with a flat trend function, with single smoothing.

features can also be observed for $r = 0.05$ as displayed in Table 6.5. This feature is indeed surprising as we would intuitively expect power to increase with an increase in sample size.

The same range has been used for all three sample sizes. This has different implications to the degree of correlation in each case. A larger sample size corresponds to a finer resolution, where the spatial autocorrelation between neighbouring observations is stronger than one with a smaller sample size. This might affect the power of the test. However a more obvious cause which might have a more prominent effect, could be the positions of the evaluation points for the four samples. More evaluation points for $n = 17 \times 17$ sit exactly at the data points which have been raised to a particular height, compared to $n = 11 \times 11$ and $n = 21 \times 21$. Hence this might result in the power being higher for the former especially at lower *h.test*.

To investigate if it is due to the location or number of the evaluation points, the experiment is repeated with $ngrid = (n-2) \times (n-2)$. In this setting, the evaluation points sit

h.test	**n=11x11**		**n=17x17**		**n=21x21**		**n=24x24**	
	Size	Power	Size	Power	Size	Power	Size	Power
0.10	0.020	0.090	0.065	0.420	0.055	0.395	0.065	0.565
0.15	0.060	0.275	0.060	0.795	0.060	0.750	0.035	0.855
0.20	0.065	0.630	0.050	0.965	0.055	0.980	0.045	1.000
0.25	0.045	0.850	0.055	0.995	0.055	0.995	0.045	1.000
0.30	0.045	0.890	0.055	0.995	0.055	0.995	0.060	1.000

Table 6.6. Simulation results for $n = 11 \times 11$, 17×17, 21×21 and $n = 24 \times 24$ with range, $r = 0.03$ and $jump = 2$ with a flat trend function, with single smoothing. The no. of evaluation points, $ngrid = (n-2) \times (n-2)$ is used for to compare to the results obtained in Table 6.4.

h.test	**n=11x11**		**n=17x17**		**n=21x21**		**n=24x24**	
	Size	Power	Size	Power	Size	Power	Size	Power
0.10	0.020	0.090	0.055	0.330	0.045	0.320	0.045	0.380
0.15	0.070	0.250	0.060	0.580	0.035	0.575	0.040	0.575
0.20	0.045	0.495	0.045	0.800	0.070	0.805	0.035	0.810
0.25	0.050	0.685	0.050	0.885	0.070	0.880	0.060	0.945
0.30	0.045	0.750	0.065	0.890	0.065	0.925	0.060	0.960

Table 6.7. Simulation results for $n = 11 \times 11$, 17×17, 21×21 and $n = 24 \times 24$ with range, $r = 0.05$ and $jump = 2$ with a flat trend function, without double smoothing. The no. of evaluation points $ngrid = (n-2) \times (n-2)$ is used for to compare to the results obtained in Table 6.5. .

exactly at the data points, excluding those at the first and last row and column. The results are shown in Tables 6.6 and 6.7 for $r = 0.03$ and $r = 0.05$ respectively. For $r = 0.03$, the powers for $n = 21 \times 21$ are much higher especially at lower bandwidths than those for the same sample size shown in Table 6.4, which uses $ngrid = 15 \times 15$. In fact, they are comparable to those produced for $n = 17 \times 17$. The power for $n = 24 \times 24$ is approximately the same, but for $n = 11 \times 11$, the power is slightly lower when less evaluation points, $ngrid = 9 \times 9$ are used instead. The features are similar for $r = 0.05$. This indicates that the test may be sensitive to the position and number of evaluation points especially at lower bandwidths.

6.4.2 Equally Spaced Setting: Correlation unknown

In a real-life setting, the spatial correlation in the data is rarely known. A fairly good estimate of the unknown correlation is quite crucial to the performance of the test. This section investigates how the DDTP algorithm that was proposed in Section 6.2.7 performs in various data conditions of different degrees of spatial correlation ($r = 0.03, 0.05$), under the null and alternative hypothesis. The same data that were simulated in the Section 6.4.1 were used here. Only the Flat and Qiu trends are considered. To estimate correlation in the presence of a rougher function like the Ismail trend for such a small sample size would be difficult.

The first stage of the DDTP uses the same smoothing parameter, i.e. $h1 = h.trend$ and is carried out with double smoothing (dsT), while the second stage explores different smoothing parameters, $h2 \in \{h1, h1 - 0.1, h1 - 0.2\}$ with just single smoothing (dsF), as applied in the previous sections. A wide range of smoothing parameters has been used, from 0.1 to 0.6 and a distance of 0.7 for variogram modelling. The value of *std* used for the partitioning process is set first at 2.5. If the results for the size are unsatisfactory, higher values of $std = 3.5$ are then considered. Only those simulation results that give reasonable size and power, denoted here as Size1 and Power1, are displayed in Tables 6.8 - 6.11. Simulations which give sizes that are within limits are highlighted in bold.

Tables 6.8 and 6.9 display results for $r = 0.03$, with Flat and Qiu trends respectively. A large smoothing parameter $h1 = h.trend \geq 0.4$, is required to give a good size. Size is inflated for lower $h1 = h.trend$. The power of the test for those with appropriate sizes are comparable to the true power. Using $std = 2.5$ gives reasonable results.

As with the data of $r = 0.03$ with a Qiu trend, a slightly smaller smoothing parameter (eg. $h1 = 0.3$) than that for the flat function can be used to give a good size. However, using a smoothing parameter of 0.4 gives very low power. The best power of 0.85 is obtained using a smoothing parameter of 0.3 for both stages. This is slightly lower but still comparable to the true power. Using $std = 2.5$ gives reasonable results. For both trend functions, at the same value of $h1$, power stays the same or increases as $h2$ increases, while size remains approximately the same. Hence a recommended value of $h2$ is $h2 \in \{h1, h1 - 0.1\}$. Size and

Flat function for $r = 0.03$						
std	$h1 = h.trend$ (dsT)	$h2$ (dsF)	Size1	Power1	Size2	Power2
2.5	0.30	0.10	0.170	0.505	-	-
2.5	0.30	0.20	0.150	0.945	-	-
2.5	0.30	0.30	0.145	0.985	-	-
2.5	0.40	0.20	0.105	0.875	-	-
2.5	**0.40**	**0.30**	**0.085**	**0.930**	0.120	0.980
2.5	**0.40**	**0.40**	**0.085**	**0.960**	0.105	0.995
2.5	**0.50**	**0.30**	**0.070**	**0.840**	-	-
2.5	**0.50**	**0.40**	**0.085**	**0.870**	**0.085**	**0.920**
2.5	**0.50**	**0.50**	**0.085**	**0.860**	**0.085**	**0.920**

Table 6.8. Table shows results for $n = 17 \times 17$, range, $r = 0.03$ and $jump = 2$ with a flat trend, when correlation is estimated using the DDTP approach. Size1, Power1: size and power with $maxdist = 0.7$; Size2, Power2: size and power with $maxdist = 1.4$

Qiu function for $r = 0.03$						
std	$h1 = h.trend$ (dsT)	$h2$ (dsF)	Size1	Power1	Size2	Power2
2.5	0.20	0.10	0.205	0.450	-	-
2.5	0.20	0.20	0.235	0.940	-	-
2.5	**0.30**	**0.10**	**0.080**	**0.260**	-	-
2.5	**0.30**	**0.20**	**0.060**	**0.750**	**0.060**	**0.750**
2.5	**0.30**	**0.30**	**0.090**	**0.850**	**0.090**	**0.935**
2.5	**0.40**	**0.20**	**0.035**	**0.160**	-	-
2.5	**0.40**	**0.30**	**0.035**	**0.215**	**0.040**	**0.375**
2.5	**0.40**	**0.40**	**0.055**	**0.215**	**0.055**	**0.400**

Table 6.9. Table shows results for $n = 17 \times 17$, range, $r = 0.03$ and $jump = 2$ with a Qiu trend, when correlation is estimated using the DDTP approach. Size1, Power1: size and power with $maxdist = 0.7$; Size2, Power2: size and power with $maxdist = 1.4$

power increase as $h1 = h.trend$ decreases.

Next, we shall look at the ability of the DDTP algorithm to capture a higher range of 0.05 in the presence of both trend functions. Tables 6.10 - 6.11 display the results for $r = 0.05$ with Flat and Qiu trends respectively. Using $std = 2.5$ does not give a good size for a wide range of smoothing parameters. However, using $std = 3.5$ gives appropriate sizes at high smoothing parameters. The power with appropriate sizes at $h1 = h.trend = 0.6, h2 = 0.4$ is

Flat function for $r = 0.05$						
std	$h1 = h.trend$ (dsT)	$h2$ (dsF)	Size1	Power1	Size2	Power2
2.5	0.40	0.20	0.285	0.775	-	-
2.5	0.40	0.30	0.260	0.830	-	-
2.5	0.40	0.40	0.225	0.830	-	-
2.5	0.50	0.30	0.165	0.685	-	-
2.5	0.50	0.40	0.165	0.685	-	-
2.5	0.50	0.50	0.160	0.685	-	-
2.5	0.60	0.40	0.125	0.545	-	-
2.5	0.60	0.50	0.125	0.545	-	-
2.5	0.60	0.60	0.120	0.530	-	-
3.5	0.40	0.20	0.210	0.790	-	-
3.5	0.40	0.30	0.160	0.840	-	-
3.5	0.40	0.40	0.160	0.835	-	-
3.5	0.50	0.30	0.160	0.675	-	-
3.5	0.50	0.40	0.150	0.675	-	-
3.5	0.50	0.50	0.150	0.675	-	-
3.5	**0.60**	**0.40**	**0.095**	**0.490**	-	-
3.5	**0.60**	**0.50**	**0.090**	**0.490**	**0.095**	**0.615**
3.5	**0.60**	**0.60**	**0.080**	**0.490**	**0.095**	**0.605**

Table 6.10. Table shows results for $n = 17 \times 17$, range, $r = 0.05$ and $jump = 2$ with a Flat trend, where correlation is estimated using the DDTP approach. Size1, Power1: size and power with $maxdist = 0.7$; Size2, Power2: size and power with $maxdist = 1.4$

constant at 0.49, regardless of the value of $h2$ used.

As with the data of $r = 0.05$ with a Qiu trend, using $std = 2.5$ also does not give a good size for a wide range of smoothing parameters. The only one that has size within the 99% CI is with $h1 = h.trend = 0.4$, $h2 = 0.3$, which gives a power of 0.300. This is very much lower compared to the true power of 0.885. The value of $std = 3.5$ is then used, and this gives more controlled sizes at the same smoothing parameters. The best power of 0.48 can be obtained using $h1 = h.trend = 0.35$, $h2 = 0.25$. This is still much lower than the true power of 0.87. We see the same feature as observed earlier for $r = 0.03$, whereby power increases as $h2$ increases for the same $h1$, while size remains approximately the same. Hence using $h2 = h1$ or $h1 - 0.1$ is recommended for good power. Size and power decrease with increasing $h1 = h.trend$. However, a larger $h1 = h.trend$ is at times necessary to get a good

Qiu function for $r = 0.05$						
std	$h1 = h.trend$ (dsT)	$h2$ (dsF)	Size1	Power1	Size2	Power2
2.5	0.30	0.10	0.220	0.410	-	-
2.5	0.30	0.20	0.230	0.730	-	-
2.5	0.30	0.30	0.235	0.795	-	-
2.5	0.35	0.15	0.190	0.360	-	-
2.5	0.35	0.25	0.160	0.540	-	-
2.5	0.35	0.35	0.190	0.570	-	-
2.5	0.40	0.20	0.100	0.250	-	-
2.5	**0.40**	**0.30**	**0.095**	**0.300**	0.115	0.575
2.5	0.40	0.40	0.100	0.300	-	-
3.5	0.30	0.10	0.175	0.335	-	-
3.5	0.30	0.20	0.155	0.670	-	-
3.5	0.30	0.30	0.130	0.770	-	-
3.5	0.35	0.15	0.115	0.250	-	-
3.5	**0.35**	**0.25**	**0.090**	**0.475**	0.110	0.665
3.5	0.35	0.35	0.120	0.480	-	-
3.5	**0.40**	**0.20**	**0.060**	**0.195**	-	-
3.5	**0.40**	**0.30**	**0.055**	**0.250**	**0.095**	**0.52**
3.5	**0.40**	**0.40**	**0.070**	**0.235**	0.110	0.500

Table 6.11. Table shows results for $n = 17 \times 17$, range, $r = 0.05$ and $jump = 2$ with a Qiu trend, where correlation is estimated using the DDTP approach.

size. Using a larger *std* gives a better size, but this results in a decrease in power.

In ths first round of simulations, the variogram distance to model the correlation was set at half the maximum pairwise distance, which is $maxdist = 0.7$. As *maxdist* might be a factor that influences the estimation of the range and sill, the simulations have been repeated for those which give appropriate size (< 0.1) and better powers (using $h2 = h1, h1 - 0.1$) with $maxdist = 1.4$, which is the maximum pairwise distance. The other conditions stay the same. The corresponding size and power are denoted as "Size2" and "Power2" in Tables 6.8 - 6.11.

Using $maxdist = 1.4$ gives sizes that are slightly inflated at certain smoothing parameters for $r = 0.03$ with a flat trend, but are within limits for the Qiu trend. For those with appropriate sizes, the powers are higher than when $maxdist = 0.7$ is used. Sizes for data with range 0.05 are within limits for the Flat function, but most are inflated for the Qiu function. A similar feature where power is higher for those with appropriate size can be

observed. Using a larger distance of $maxdist = 1.4$ gives a better power when the size is appropriate. However, it appears to be less stable, giving inflated sizes. Nonetheless, it copes better with a lower range.

As the algorithm does not perform very well for a jump of 2 for a high range of 0.05, the jump size is increased to 3 to investigate if this will improve its performance. Only those simulations with conditions that give appropriate sizes are carried out with $maxdist = 0.7$. Tables 6.12 and 6.13 display the power with jump 3, along with the results obtained earlier with jump 2.

From the simulation results, power for the Flat function is increased from 0.49 to 0.795, while power is increased by approximately 40% for the Qiu function at various combinations of smoothing parameters. The best is with $h1 = h.trend = 0.35, h2 = 0.25$, giving a high power of 0.9.

Flat function for $r = 0.05$					
std	$h1 = h.trend$ (dsT)	$h2$ (dsF)	Size1	Power1 (jp=2)	Power1 (jp=3)
3.5	**0.60**	**0.40**	**0.095**	**0.490**	**0.795**
3.5	**0.60**	**0.50**	**0.090**	**0.490**	**0.795**
3.5	**0.60**	**0.60**	**0.080**	**0.490**	**0.795**

Table 6.12. Table shows results for $n = 17 \times 17$ and range, $r = 0.05$, with a FLAT trend, where correlation is estimated using the DDTP approach for Power with jumps of 2 and 3 with controlled sizes.

Qiu function for $r = 0.05$					
std	$h1 = h.trend$ (dsT)	$h2$ (dsF)	Size1	Power1 (jp=2)	Power1 (jp=3)
2.5	**0.40**	**0.30**	**0.095**	**0.300**	**0.715**
3.5	**0.35**	**0.25**	**0.090**	**0.475**	**0.905**
3.5	**0.40**	**0.20**	**0.060**	**0.195**	**0.665**
3.5	**0.40**	**0.30**	**0.055**	**0.250**	**0.695**
3.5	**0.40**	**0.40**	**0.070**	**0.235**	**0.67**

Table 6.13. Table shows results for $n = 17 \times 17$ and range, $r = 0.05$, with a Qiu trend, where correlation is estimated using the DDTP approach, for Power with jump of 2 and 3 with well controlled sizes.

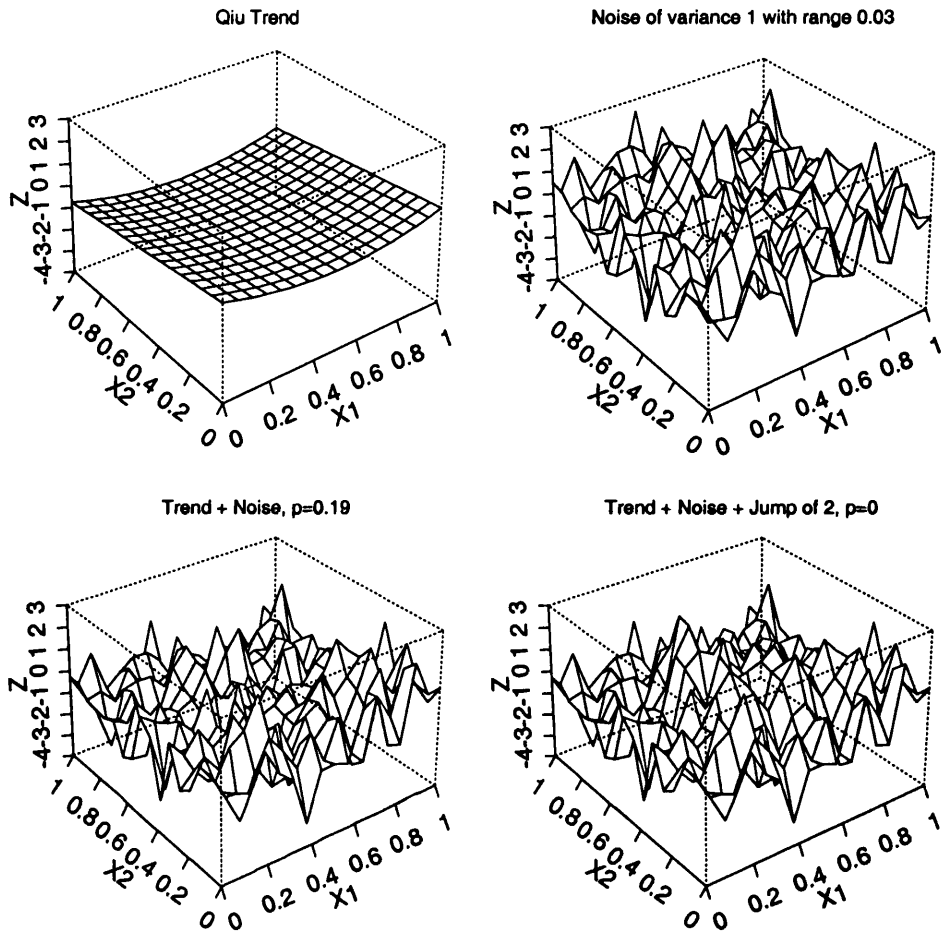

Figure 6.5. The figure displays the Qiu trend, the correlated noise of range 0.03 and variance of 1, data with trend and noise, and data with trend and noise and jump of size 2 within the jump location curve. $h1 = h.trend = h2 = 0.3$, $std = 2.5$.

Next, as an illustration, Figures 6.5 and 6.6 display examples of a simulated set with the Qiu trend function and a normal error of variance 1 with dependence structure of range, $r = 0.03$ and 0.05 respectively, drawn over the same scale to allow for ease of comparison. The top left and right panels of Figure 6.5, displays the Qiu trend and a simulated set of noise of variance 1 with range 0.03 respectively. The bottom left and right panel display the data with trend and noise (i.e. model under the null hypothesis), and the model with trend, noise and jump of 2 added within the jump location curve (i.e. model under the alternative hypothesis) respectively.

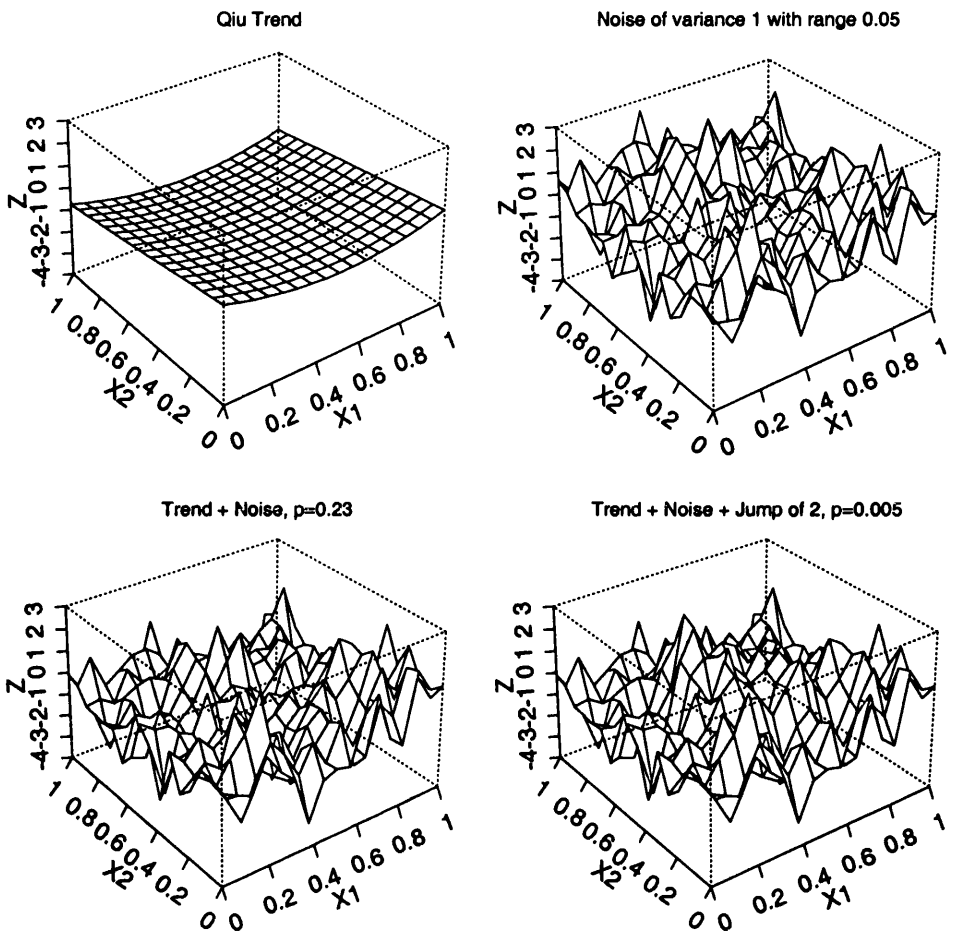

Figure 6.6. The figure displays the Qiu trend, the correlated noise of range 0.05 and variance of 1, data with trend and noise, and data with trend and noise and jump of size 2 within the jump location curve. $h1 = h.trend = 0.35$, $h2 = 0.25$, $std = 3.5$.

Using $h1 = h.trend = h2 = 0.3$, $std = 2.5$, the DDTP algorithm is applied, which gives an estimated range, sill and p-value of 0.0279, 0.985 and 0.19 respectively, for the model under the null hypothesis. It gives an estimated range, sill and p-value of 0.0276, 0.991 and 0.005 respectively, for the model under the alternative hypothesis. Hence the test has correctly given the right conclusions of whether any discontinuities existed.

Figure 6.6 shows a similar simulated set, but with range of 0.05. The DDTP is applied with $h1 = h.trend = 0.35, h2 = 0.25$ and $std = 3.5$, which gives estimated range and sill of 0.045 and 0.967 respectively, and a p-value which is not significant at 0.23, under the

model with no discontinuity. As for the model with discontinuity, it correctly rejects the null hypothesis with a p-value of 0.005. Its estimated range and sill are 0.045, and 1.02.

From Figures 6.5 and 6.6, as the sample size is small there appears to be quite a lot of variation with noise variance of 1 for both dependence structures, as shown in the top right panel. From the bottom right panel of both figures, it is not obvious to the eye whether the irregularity could be due to the presence of discontinuity in the data, or due to trend and/or random variation. However, the DDTP has correctly detected discontinuities in both cases when they are present. It gives quite a good estimate of the range and sill, under both the null and alternative hypothesis, highlighting the effectiveness of introducing the partitioning process in the algorithm.

6.5 Unequally Spaced Setting

Often, data might not be collected at regular intervals. In this section, it is of interest to see how the discontinuity test performs and whether its performance would be affected if the data are unequally spaced. This could apply to cases where the intervals between data-points are not equal, or it could refer to locations that form a regular grid, but there might be missing data at certain locations.

In the following simulations, we obtained two sets of unequally spaced locations in both directions by simulating from a $Un(0, 1)$ distribution. The irregular spatial locations are kept constant throughout the study. Figure 6.7 shows the position of the unequally spaced data, the evaluation points and the jump location curve. The data conditions and parameters used in the test approaches are similar to those used in the equally spaced case. These are displayed earlier in Table 6.1.

Assuming knowledge of the true correlation, we will explore how the test works in this design setting in Section 6.5.1, and then apply the DDTP algorithm in Section 6.5.2 to estimate the unknown spatial correlation. A smaller scale of simulation study is done here compared to the equally spaced setting, as we would expect the features of the influence of various factors on the discontinuity test to be quite similar. (We have seen this in the one-dimensional case in Chapters 3 and 4.)

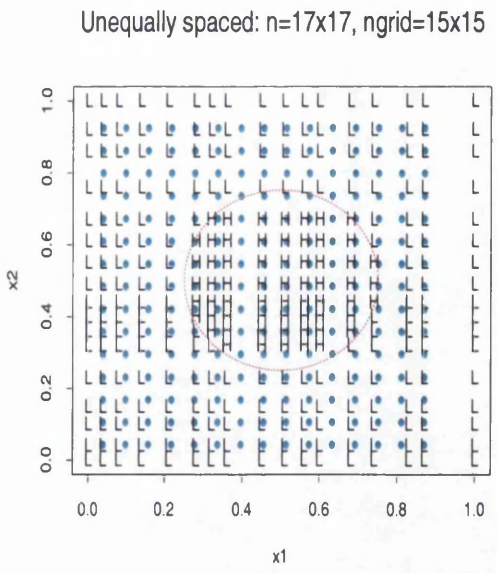

Figure 6.7. The figure displays the unequally spaced data points ($n = 17 \times 17$). The evaluation points ($ngrid = 15 \times 15$) are represented by green "*". The red dotted line represents the jump curve, and the points with jump added are denoted as "H", while the rest are denoted as "L".

6.5.1 Unequally Spaced Setting: Correlation known

Here, knowledge of both ranges, 0.03 and 0.05 are assumed. Table 6.14 presents the simulation results under both trend functions. This table can be used as a reference to the expected size and power of the test if the range and sill are accurately estimated.

h.test	**Flat Function**				**Qiu Function**			
	Size		**Power**		**Size**		**Power**	
	$r = 0.03$	$r = 0.05$	$r = 0.03$	$r = 0.05$	$r = 0.03$	$r = 0.05$	$r = 0.03$	$r = 0.05$
0.10	0.055	0.080	0.470	0.405	0.060	0.080	0.470	0.410
0.15	0.070	0.085	0.755	0.520	0.070	0.080	0.755	0.515
0.20	0.040	0.055	0.970	0.730	0.050	0.060	0.965	0.710
0.25	0.035	0.050	1.000	0.855	0.050	0.060	1.000	0.850
0.30	0.045	0.050	1.000	0.930	0.080	0.075	1.000	0.910

Table 6.14. Simulation results for $n = 17 \times 17$, range, $r = 0.03, 0.05$ and $jump = 2$ for both flat and Qiu's trend functions, with single smoothing, for unequally spaced data.

In accordance to Table 6.14, there is a similar pattern in the results as in the equally spaced setting, displayed in Table 6.3, where the power is higher for data with lower range, and increases with increasing smoothing parameter, under both trend functions. The size however remains within limit over all *h.test* for both trends, unlike that in the equally spaced setting, where it goes off limit at higher *h.test* for the Qiu trend.

6.5.2 Unequally Spaced Setting: Correlation unknown

This section examines how the DDTP algorithm copes with unequally spaced data. Only those simulations with various smoothing parameters used in the DDTP algorithm, that give well-controlled size in the equally spaced setting (Section 6.4.2) are used. Tables 6.15- 6.18 display the results for both $r = 0.03$ and 0.05, under Flat and Qiu trends. Comparisons will be made to the results obtained here with those in Section 6.4.2 obtained in the equally spaced setting. As mentioned earlier, Size1, Power1 denote the size and power obtained using $maxdist = 0.7$, while Size2, Power2 denote the size and power obtained with $maxdist = 1.4$.

From Tables 6.15 and 6.16, we can observe that some of the sizes are inflated for $r = 0.03$ under both Flat and Qiu trends respectively. Power is comparable to that of the equally spaced setting for the Flat trend. It appears to perform better for the Qiu trend at higher smoothing parameters. In accordance to Table 6.17, size for $r = 0.05$ is within limits for the Flat trend, and power is slightly higher than that for the equally spaced setting. The sizes are however more inflated for the Qiu trend as shown in Table 6.18.

Using the maximum distance of 1.4 gives a better power and the size generally remains well-controlled under different smoothing parameters.

6.6 Discussion of Simulation Results

The simulation study is divided into two main sections: the equally spaced setting in Section 6.4 and the unequally spaced setting in Section 6.5. In the simulation study for the equally spaced setting, knowledge of correlation is first assumed to investigate the effects

Flat function for $r = 0.03$						
std	$h1 = h.trend$ (dsT)	$h2$ (dsF)	Size1	Power1	Size2	Power2
2.5	**0.40**	**0.30**	**0.075**	**0.905**	**0.095**	**0.905**
2.5	**0.40**	**0.40**	**0.090**	**0.935**	**0.095**	**0.925**
2.5	**0.50**	**0.40**	**0.080**	**0.885**	0.100	0.900
2.5	**0.50**	**0.50**	**0.090**	**0.890**	0.110	0.920

Table 6.15. Table shows results for $n = 17 \times 17$, range, $r = 0.03$ and $jump = 2$, with a flat trend, when correlation is estimated using the DDTP approach, for unequally spaced data.

Qiu function for $r = 0.03$						
std	$h1 = h.trend$ (dsT)	$h2$ (dsF)	Size1	Power1	Size2	Power2
2.5	**0.30**	**0.20**	**0.065**	**0.725**	0.105	0.755
2.5	**0.30**	**0.30**	0.100	0.915	0.100	0.935
2.5	**0.40**	**0.30**	**0.065**	**0.595**	**0.065**	**0.735**
2.5	**0.40**	**0.40**	**0.070**	**0.655**	**0.075**	**0.795**

Table 6.16. Table shows results for $n = 17 \times 17$, range, $r = 0.03$, and $jump = 2$ with a Qiu trend, when correlation is estimated using the DDTP approach, for unequally spaced data.

of other factors, such as trend function and smoothing parameter (Section 6.4.1.1), correlation (Section 6.4.1.2) and sample size (Section 6.4.1.3). In the later section, correlation is unknown, and is estimated using the DDTP algorithm for different ranges and trends. A similar structure of simulation study for the unequally spaced setting is also used.

As we have seen in previous simulation studies, the smoothness of the underlying function is determined by the smoothing parameter. Both of these factors play an individual as well as combined role in the performance of the test. If the trend is very smooth, as in the case with the Flat and Qiu trend, power increases with increasing value of the smoothing parameter and the size stays within limits. However, if the trend is irregular, as in the Ismail trend, oversmoothing will result in an inflated size.

As contemplated, if the range is increased, power is decreased. Power, however, is high for both ranges. At $h.test = 0.2$, power for $r = 0.03$ and 0.05 with an underlying Flat or Qiu trend, are about 0.96 and 0.8 respectively.

Flat function for $r = 0.05$						
std	$h1 = h.trend$ (dsT)	$h2$ (dsF)	Size1	Power1	Size2	Power2
3.5	**0.60**	**0.50**	**0.090**	**0.560**	**0.075**	**0.690**
3.5	**0.60**	**0.60**	**0.090**	**0.565**	**0.075**	**0.690**

Table 6.17. Table shows results for $n = 17 \times 17$, range, $r = 0.05$ and $jump = 2$ with a Flat trend, where correlation is estimated using the DDTP approach, for unequally spaced data

Qiu function for $r = 0.05$						
std	$h1 = h.trend$ (dsT)	$h2$ (dsF)	Size1	Power1	Size2	Power2
2.5	0.40	0.30	0.110	0.605	0.125	0.735
2.5	0.40	0.40	0.120	0.625	0.140	0.770
3.5	0.35	0.25	0.120	0.670	0.130	0.780
3.5	0.35	0.35	0.125	0.755	0.130	0.850
3.5	**0.40**	**0.30**	**0.085**	**0.540**	**0.090**	**0.695**
3.5	**0.40**	**0.40**	**0.095**	**0.585**	0.115	0.760

Table 6.18. Table shows results for $n = 17 \times 17$, range, $r = 0.05$ and $jump = 2$, with a Qiu trend, where correlation is estimated using the DDTP approach, for unequally spaced data

Lastly, the effects of sample size are investigated. Generally, there is an increase in power as sample size increases. However the powers for $n = 17 \times 17, 21 \times 21$ and 24×24 differ only slightly for low smoothing parameter but is approximately the same for moderate to large smoothing parameter. From the simulation results, it is also shown that the test may be sensitive to the position and number of evaluation points, *ngrid* used. For instance, the use of $ngrid = 19 \times 19$ instead of 15×15 makes a considerable difference in power for a sample of size $n = 21 \times 21$.

It is interesting to note that the power for a moderate dependence structure of $r = 0.05$ with a jump of 2 is much higher in this two-dimensional setting compared to the one-dimensional setting that we have seen earlier in Chapters 3 and 4. Though the curse of dimensionality exists, where the data are considered more sparse in a higher dimension, the power of the discontinuity test is not affected. This is because the discontinuity in the former case occurs over a curve, hence there are more evaluation points which have high

standardised differences compared to the latter which is just at a point.

The above has so far assumed that the correlation is known, but this is rarely the case in practice. Moreover, in the spatial context, there is often trend present which complicates the issues. To be able to obtain a precise estimate of the spatial dependence of the data in the presence of both an unknown trend and possible discontinuities is a challenging task. The DDTP algorithm is proposed in Section 6.2.7 to tackle this issue. It comprises estimating and removing the trend in the data using local linear smoothing; and secondly removing observations near a possible jump location curve. Both of these, if unaccounted for, will cause an increase in the estimated range and hence result in low probability of any success to detect discontinuities if present. The capability of the algorithm to be able to handle different data settings is investigated in Section 6.4.2.

Overall, DDTP seems to work markedly better for a lower range, eg. $r = 0.03$ than 0.05. Using appropriate smoothing parameters for both stages gives power that is quite close to the true power for $r = 0.03$.

For a flat function, a fairly large smoothing parameter has to be used in the first stage in order to get a good estimate of the correlation, giving good size for both ranges. Using a low smoothing parameter will result in tracking the data too closely, hence underestimating the range in the data. This causes the size to go off limits.

Increasing the smoothing parameter for the first stage will result in a decrease in power, while an increasing smoothing parameter in the second stage will result in an increase in power. This is because the first stage involves trend removal. Oversmoothing might result in overestimation of the range, hence giving lower power. On the other hand, increasing the smoothing parameter for the second stage will cause an increase in power as both trend functions, Flat and Qiu are relatively smooth.

The DDTP procedure does not work well for a higher range like $r = 0.05$ when $std = 2.5$ is used (as in the earlier simulations for $r = 0.03$). However, if *std* is increased to 3.5, the size is more well-controlled. This is because data of such a high dependence, when treated as independent, will give very high standardised differences. Hence a lot of observations are inaccurately removed during the partitioning process in the first stage. To avoid this, *std* has to be increased. However there has to be a balance, as having too large a *std* might

result in not having an efficient removal of the observations near the jump location curve. This would in turn cause an overestimation of the range of the model under the alternative hypothesis, hence decreasing the power.

Nonetheless, powers for both the Flat and Qiu trends for $r = 0.05$ of a jump size of 2 are only about half of their true power. By increasing the jump of size 2 to 3 for a higher range of 0.05, the power for the DDTP increases considerably. Hence for the DDTP algorithm to be capable of a higher success rate in detecting discontinuities, accounting for the presence of both trend and moderately high correlation, a higher jump to noise ratio is helpful.

Another feature that we can observe from the simulation study is that a larger smoothing parameter is required to estimate a higher range. The reasoning is similar to what we have seen earlier in our previous simulation study in the one-dimensional case. Using a low *h.trend* will result in tracking the data too closely, giving a lower estimate of correlation. However, while the flat function works better when a large smoothing parameter is used, in the presence of a less smooth function such as Qiu, a smaller smoothing parameter is preferred.

Another factor that was investigated was the effects of the distance used for variogram fitting on the estimated correlation, which would affect the discontinuity test to a certain extent. The simulation study was repeated using a larger distance of 1.4, and the results show that it does affect the size and power of the test. A larger distance tends to result in inflation of the sizes, giving non-conservative tests. However, at well-controlled sizes, it gives better power if suitable conditions in the test approach are used. This could be partly due to the location of the discontinuities. In our simulations, this is located in the middle. If the maximum pairwise distance is considered, it allows observations that are far apart and not located near the discontinuity to be included in the variogram fitting, giving a more accurate range estimate in the model under the alternative hypothesis. Hence though it might not be a very crucial factor, it might be useful to vary the distance of the variogram fitting when analysing data for discontinuities.

In the unequally spaced setting, the performance of the DDTP algorithm is comparable to that in the equally spaced setting for low range of 0.03 and for a flat function with a high range of 0.05. However, in the presence of a Qiu trend, the sizes are mostly inflated.

Increasing the distance for variogram fitting is also investigated. Results indicate a better power for all, but sizes are slightly inflated for some.

In conclusion, the DDTP works well in the presence of both the Flat and Qiu trends if the range is lower at 0.03. The latter gives powers that are comparable to the true power under suitable conditions. The best power is about 0.9. However, for a higher range of 0.05, the power is somewhat lower than that of the true power. A larger jump is required to increase the success rate of detecting a discontinuity when present.

The DDTP algorithm might not work very well for a high range, as the sample size is very small, and in the presence of both trend and correlation, this problem becomes even more complicated. With a larger set of data, it will certainly work better, as we would have more data during the partitioning process in the first stage, which would provide a better fitting of the trend, before the estimation of correlation.

The sensitivity of the algorithm to the smoothing parameter used is inevitable. To allow for flexibility and wider applications, the DDTP algorithm has not made any assumptions about the underlying trend functions, allowing it to be smoothly varying. The shape of the jump location curve is also not assumed. Hence it faces a complicated and challenging task to estimate correlation under these conditions. One of the trend functions in the simulation study is the Flat function, which basically means that no trend is present. By trying to remove this nonexistent trend, with a small smoothing parameter that is not sufficiently large, the data would be tracked too closely, causing an underestimation of the range. This will prove to be a bigger problem if the range of dependence is high, as we have seen in our simulation study.

6.7 An Application to Mapping Radioactivity

6.7.1 Introduction & Background

This section describes the analysis of a data set collected during a field test RESUME-95 (Rapid Environmental Surveying Using Mobile Equipment) exercise, organised by Nordic Nuclear Safety Research (NKS). Two of the main objectives of the RESUME-95 exercise

were to test the capability of several airborne, car-borne and in-situ instruments to map radioactive contamination, in this case from Chernobyl fallout; and to compare the results obtained using different monitoring tools, under similar controlled conditions.

The study location, denoted as Area II in the exercise RESUME-95 (1997), was rectangular in shape and 6×3 km^2 area. It is made up of typical Finnish countryside, with a mixture of forest lakes, marsh ground, rocky terrain, and some open agricultural fields. The general topography was quite flat, with most higher ground less than 50m in height. However, some of the lake sides were quite steep. The area for this exercise was selected as it was contaminated with man-made radionuclides in the days following the Chernobyl nuclear reactor accident. It was found to have high Cs-137 deposition based on earlier airborne survey results collected by the Geological Survey of Finland (GSF).

Due to the diverse mixture of terrains present in Area II, there might be considerable variation in the deposition of Cesium-137 over the whole area. The aim of this analysis is to investigate if there are any significant sharply varying depositions in such a terrain.

6.7.2 Description of Data

In order to analyse in detail Cs-137 activities, 21 parallel survey lines of 6 kilometres were flown by a helicopter carrying the detector in August 1995; line spacing was 150 metres. Each observation consists of the average helicopter position and the counts per second from a spectrum representing an energy scale, (X1,X2,Y). A coloured map of the original data which had undergone some smoothing can be found in Plate 7 of RESUME-95 (1997).

To compensate for differences in the sampling time and variations in the flight tracks of different monitoring tools, the data set has been re-gridded. The original flight lines had a bearing of 22/202 degrees and were 150 meters apart. To obtain a new right-angular grid, the observations were shifted 22 degrees anti-clockwise and arranged in cells of size 150m and 300m perpendicular and along the lines respectively. The final rectangular grid is regularly spaced with 22 rows $\times$ 21 columns. The mean value of all the observations within each cell is taken as the value for that grid cell. The unit is given as Cps, which denotes counts per second (Hovgaard and Scott, 1997).

There are 411 data points with 51 missing values. The maximum was 173 Cps, minimum was 22 with standard deviation of 25.5. The summary statistics of these data are provided in Table 6.19.

Summary Statistics of airborne data							
N	Number missing	Min	1st Quartile	Median	Mean	3rd Quartile	Max
411	51	22.00	108.75	124.50	121.65	139.50	173.25

Table 6.19. Summary statistics for cesium data

We first do a bivariate interpolation over the x1 and x2 coordinates to obtain a fitted value for the missing z to be used for a contour plot. Another visualization technique that would be useful to see the spatial distribution is an image plot, where the darker the tone in a particular region, the higher the concentration of the airborne data. The contour and image plots are displayed in Figure 6.8, which highlight the fact that there is considerable variation in the whole area. Another feature that can be observed is that the highest levels of Cs-137 occur on the central and northern parts of the area, which were largely forested areas. The lower levels are mainly associated with cultivated areas or villages and lakes. The Cs-137 activity is lower in villages where it has been washed away from paved surfaces, and in cultivated areas where the soil has been turned over. The radioactivity map does produce results that indicate interesting correlations with land use.

Perspective plots of the smoothed airborne data using different degrees of smoothing are displayed in Figure 6.9. It can be observed from the diagram that using a small smoothing parameter tracks the data closely and is more capable of picking up fine features, while using a large smoothing parameter results in a much smoother trend. There are considerable differences in the features that are being picked out using $h = 0.1$ and 0.2, but the differences are minimal as we increase from $h = 0.3$ to 0.4.

Note that in this section, the imputation of missing observations using bivariate interpolation, is for illustrative purposes. The subsequent sections make use of the original data, without any imputed values.

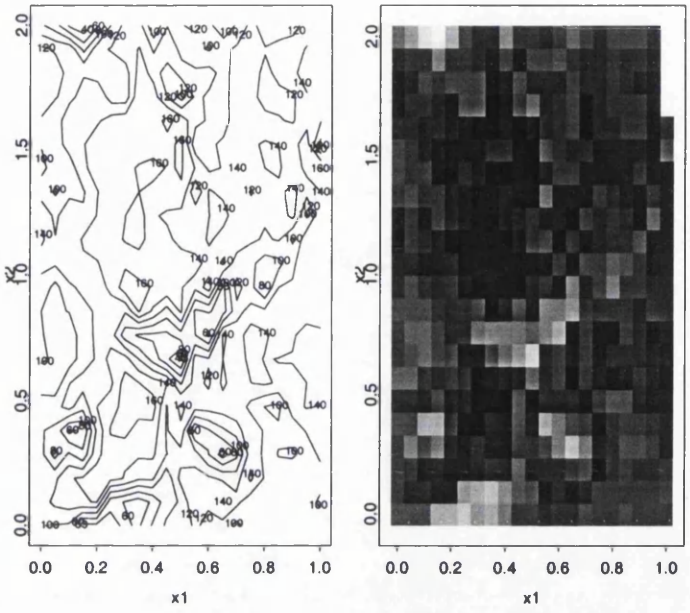

Figure 6.8. Cesium data: A contour and image plot of the airborne data (HPGe) measured in Cps (Counts per second) after interpolation over exact x1 and x2 locations

6.7.3 Detecting discontinuities: Analyses & Results

In this section, we will investigate if there are any discontinuities by carrying out the proposed discontinuity test, taking into account the spatial correlation that might be present in the data. There is a motivation to consider the structure of dependence of deposition of Cs-137 over their spatial locations as that is supported not only by the geographical characterisation of the area, but also because the area has been re-gridded. The value of the regridded cell is obtained by taking the mean of all the observations that fall in that cell. The raw data has thus been somewhat smoothed.

As it is not obvious how a parametric regression model can be used to model the data, nonparametric smoothing using local linear regression is particularly appropriate here. The smoothed data are obtained by taking a weighted average of their nearby spatial neighbours.

The Double Discontinuity Test with Partition (DDTP) algorithm introduced in Section 6.2.7, is applied. This involves the incorporation of the estimation of the correlation in

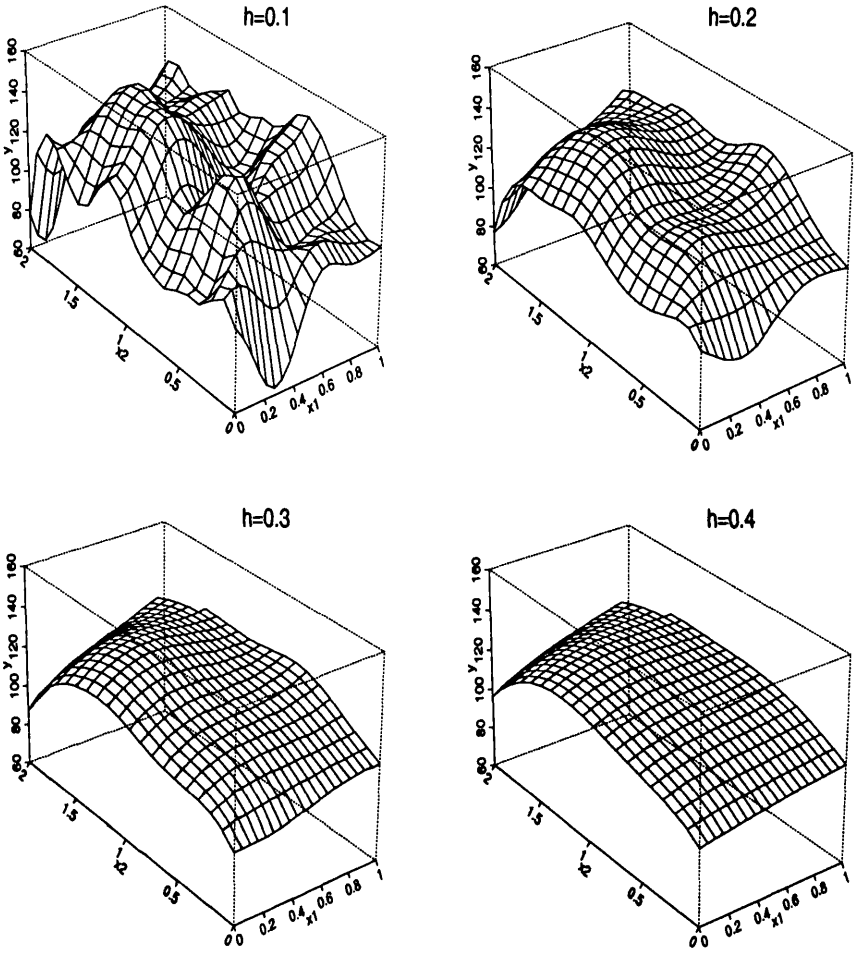

Figure 6.9. Cesium data: Perspective plots of smoothing original data using different smoothing parameters.

the discontinuity test. In the test approach, 17×17 equally spaced evaluation points (as shown in Figure 6.10), have been selected as this gives 289 evaluation points, which is of a reasonably high resolution. The parameters used for a robust variogram fitting by nonlinear weighted least squares fitting, are the maximum pairwise distance, lag of 0.05 and minimum pairs of 30 in each lag. The value for the partitioning process, *std*, was set at 2.5.

The first stage of the algorithm involves carrying out the discontinuity test with double smoothing, treating the data as independent. A smoothing parameter of 0.1 was selected as the area is made up of many different terrains, which one would suspect to result in quite an irregular trend. Hence, a small smoothing parameter might be more appropriate

in order to prevent oversmoothing the fine features. Using this smoothing parameter gives a p-value of 0, indicating that there is strong evidence of discontinuities when the data are treated as independent. This is then followed by using a similar smoothing parameter, $h.trend = 0.1$, with double smoothing , to detrend the data. As the p-value is less than 0.05, only observations with standardised differences, $st.diff1$, less than $std = 2.5$ are used to fit the variogram and obtain the range and sill. The 2nd stage of DDTP is then carried out using $h2 = 0.1$, taking into account the estimated correlation in the data. The test is significant at 0.003.

Figure 6.11 illustrates the different stages of the test. From left, the first plot shows the $(x1, x2)$ locations of the airborne data, with several missing values. Contours of various levels illustrate the locations where the values of the standardised differences, $st.diff1$ are greater than or equal to 2.5. It is important to stress here that the values of $st.diff1$ are obtained assuming the data are independent. The next plot shows step 2 of the 1st stage process, i.e. the empirical variogram of the detrended and partitioned data used to obtain the range and sill, which are 0.057 and 326 respectively. The discontinuity test is then carried out again,

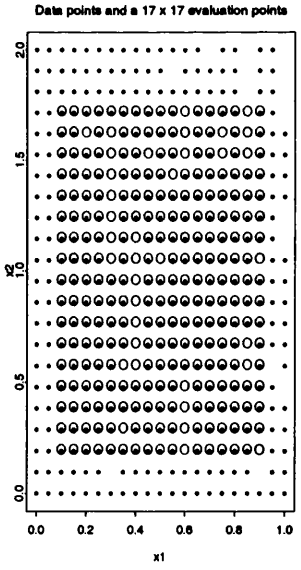

Figure 6.10. Cesium data: Location of data points with a regular grid of 17×17 evaluation points indicated by "o".

taking into account the spatial correlation of the data. Contours of $st.diff2$ at levels 2.5 and 5 are displayed in the next diagram.

We can observe that the contours are different for the first and third plots, as the contours in the latter plot takes into consideration the spatial covariance structure obtained from the estimated range and sill of the variogram. The former, however, assumes independent data, and hence the standardised differences are overestimated. In fact, the particular region within the contour of level 5 is where the cesium concentration is low, with the surrounding area being substantially higher (refer to Figure 6.8). By looking at the colour maps in Plates 7 and 19 of RESUME95, we can see that this area actually corresponds to a lake surrounded by forest.

Finally, the last plot is the significance trace, showing how the significance of the final discontinuity test depends on the smoothing parameter, $h2$. In our simulations, we have suggested using either the same smoothing parameter as the first stage, i.e. $h1 = h.trend = h2$, or a slightly smaller one. However, for completeness, we have shown how the p-values vary with a range of smoothing parameters. From the plot, we have convincing evidence to conclude that discontinuities do exist in the data, regardless of the amount of smoothing used in this latter stage.

6.7.3.1 Checking the normality Assumption

Our test assumes that the errors are normally distributed. This section is devoted to investigating if this assumption is valid. Figure 6.12 displays three different plots of the data at different stages of the test. The left plot shows the data that have undergone both the detrending and partitioning process in stage 1 with double smoothing ; while the next plot displays the data that have undergone the same process as the former one but with single smoothing. The right plot shows the original data, that have only undergone the partitioning process. A smoothing parameter of 0.1 has been used throughout. An expected normal curve is superimposed on all the plots

Though there is a slight indication that it might be negatively skewed, from the plots it seems reasonable to assume normality.

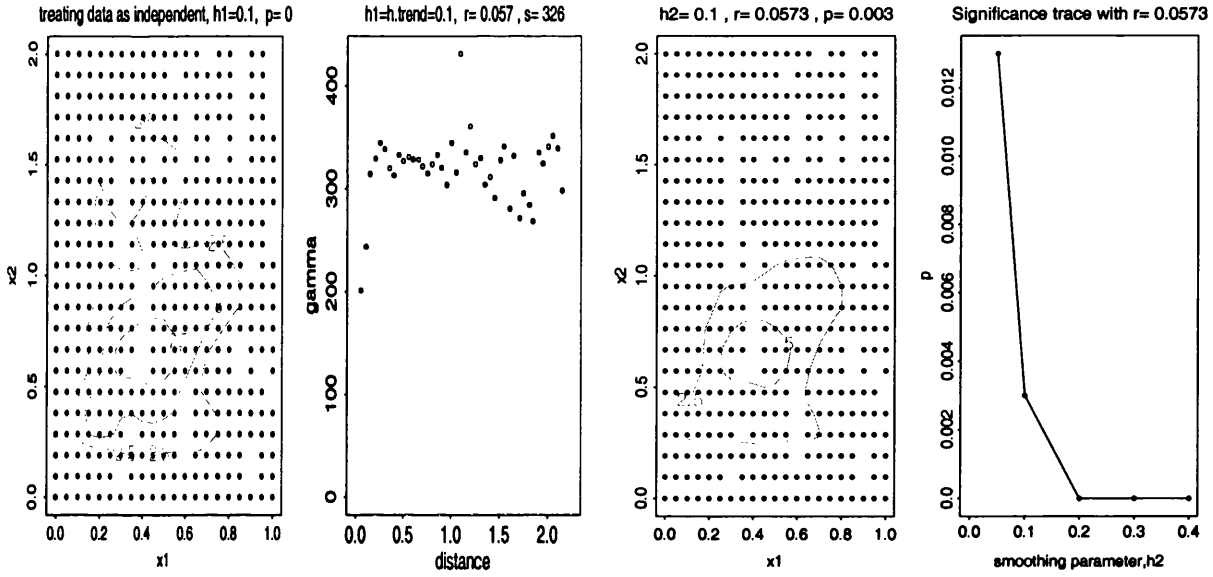

Figure 6.11. Cesium data: Different stages of the DDTP algorithm with $h1 = h2 = 0.1$ and significance trace for different $h2$. From left: contour plot of levels with $st.diff1 \geq 2.5$; empirical variogram of detrended and partitioned data; contour plot with $st.diff2 \geq 2.5$; significance trace obtained using different $h2$.

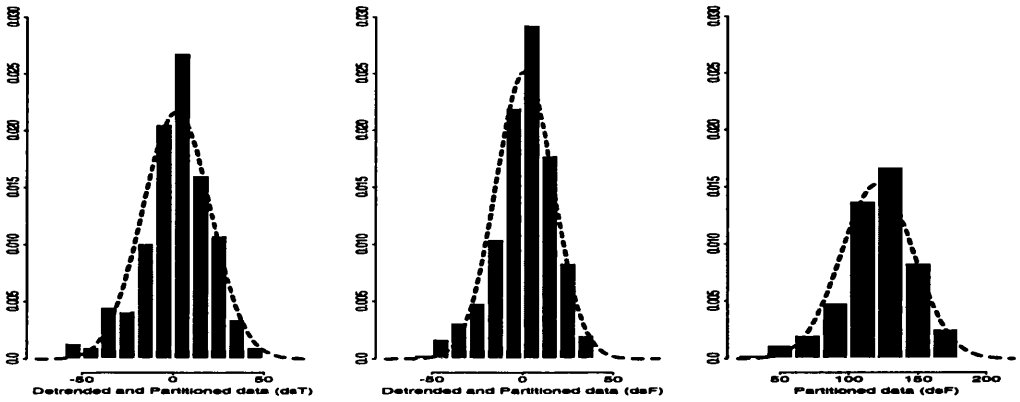

Figure 6.12. Cesium data: Checking normality of the data. From left: data that have undergone detrending (with double smoothing , dsT) and partitioning; data that have been detrended (with single smoothing, dsF), and data that have only been through partitioning. The smoothing parameter used is $h1 = h.trend = 0.1$, with $std = 2.5$ for partitioning.

6.7.4 Sensitivity Analysis

In this section, we investigate how the different parameters used by our test approach will affect the performance of our test. It is subdivided into four different parts. The first part

involves plugging in a variety of dependence structures and performing the test using various smoothing parameters. It does not involve any estimation of the correlation structure. The subsequent parts, however, involve the different parameters used in the DDTP algorithm: each investigating the aspect of varying a particular factor in the algorithm. Sections 6.7.4.2 - 6.7.4.4 respectively consider: varying the degree of smoothing in the DDTP algorithm; the value of *std* during the partitioning process and the distance used for variogram fitting. In each of these subsections, only the parameter of interest is varied, the rest stay the same as specified earlier.

6.7.4.1 Effects of dependence structure and smoothing parameter

In this section, different degrees of correlation determined by the range, r, with several smoothing parameters, $h2$ are introduced to examine how the discontinuity test performs. Table 6.20 shows the p-value of the test under different conditions. Most of the p-values are significant for $r \leq 0.08$ at all smoothing parameters. The test is only insignificant when $r = 0.10$.

In fact, $r = 0.10$ of an exponential variogram model gives a very high dependence structure, as the data are correlated up to a distance of 0.3 apart. From this table, we have convincing evidence that under low to fairly high dependence structure of $r = 0.08$, the test is significant regardless of the smoothing parameter used. However, for a high range of $r = 0.09$ and 0.10, the test is only significant at certain $h2$.

	p-value							
$h2$	**r=0.02**	**r=0.04**	**r=0.05**	**r=0.06**	**r=0.07**	**r=0.08**	**r=0.09**	**r=0.10**
0.05	0.039	0.02	0.015	0.012	0.010	0.008	0.007	0.007
0.10	0	0	0.001	0.005	0.018	0.043	0.075	0.112
0.20	0	0	0	0	0	0.000	0.003	0.014
0.30	0	0	0	0	0	0.003	0.018	0.060
0.40	0	0	0	0	0.001	0.012	0.054	0.139

Table 6.20. Cesium data: Effects of different correlation structure and smoothing parameter

This section highlights the feature that if our estimated range is less than or equal

to 0.08, then the test will be significant regardless of the amount of smoothing used in the discontinuity test. This will be a useful source of reference for the following sections, where we investigate introducing different parameters in the DDTP algorithm to estimate correlation and account for it in our discontinuity testing.

6.7.4.2 Effects of varying the smoothing parameter in the DDTP algorithm

As the DDTP algorithm involves the use of local linear regression to remove the trend which requires the specification of a smoothing parameter to govern the amount of smoothing involved, this in turn affects the estimation of our dependence structure in the first stage. Overestimation of this dependence structure might result in the test not being significant. In fact, the second stage also requires a smoothing parameter, h2, which would normally be taken as the same as h1, or slightly smaller, but to give a more complete picture, we have chosen to use the significance trace.

In the earlier analysis, we have used a smoothing parameter of $h1 = h.trend = 0.1$. On one hand, one could argue that by looking at the contour and image plot in Fig 6.8, it appears that there is quite a lot of variation in the data, which could be due to the different terrains. Hence it might be better to use a small smoothing parameter such as 0.1, to pick out any fine features. On the other hand, it is also reasonable to assume a very smooth trend of Cs-137 over the region as the topology is relatively flat. As discussed before in previous chapters, the trend and the dependence structure might combine to give the overall "movement" of the data, hence it might be safer to apply a larger smoothing parameter, so that the rest of the variation in the data can be attributed to the correlation present.

In this section, keeping the rest of the parameters as used earlier, we use different smoothing parameters in the DDTP algorithm, and investigate if the performance of our test is sensitive to this parameter. In other words, the question to be answered here is whether using different $h1 = h.trend$ will change the conclusions of the presence of abrupt changes in the Cesium-137 activity in the region in our earlier analysis in Section 6.7.3.

Table 6.21 provides a summary of the estimated range and sill of the variogram when different smoothing parameters, $h1 = h.trend$, are used in the first stage. The corresponding

$h1 = h.trend$	r	s	p-value
0.10	0.0573	331	0.003
0.20	0.0711	457	0
0.30	0.0790	478	0.002
0.40	0.078	516	0.008

Table 6.21. Cesium data: Estimated range and sill for different $h1 = h.trend$ in the DDTP algorithm

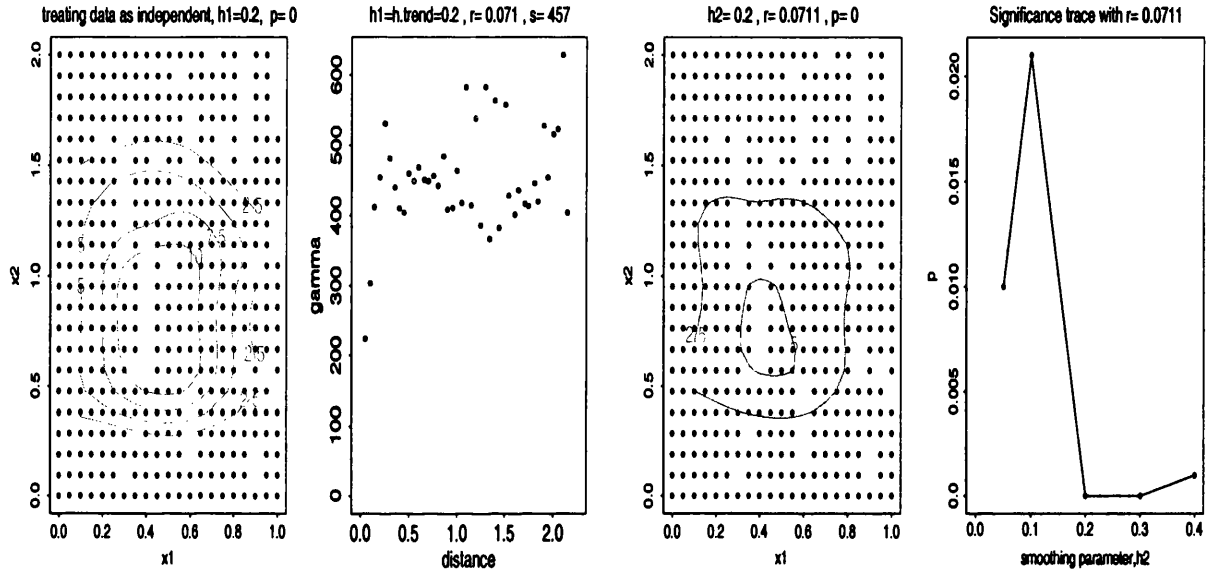

Figure 6.13. Cesium data: Different stages of the DDTP algorithm with $h1 = h2 = 0.2$ and significance trace for different $h2$. From left: contour plot of levels with $st.diff1 \geq 2.5$; empirical variogram of detrended and partitioned data; contour plot with $st.diff2 \geq 2.5$; significance trace obtained using different $h2$.

p-values using h2=h1 are also given. We can observe that the estimates of the range and sill generally increase as $h1 = h.trend$ increases. There is comparably a higher jump in the range as $h.trend$ increases from 0.1 to 0.2. However the test is still significant, giving a very low p-value. (Though not shown on this table, different $h2$ have also been investigated for the second stage, and the result stays significant.) Hence regardless of the amount of smoothing used for the first or second stage in the DDTP algorithm, the result is significant, giving convincing evidence of the presence of discontinuities.

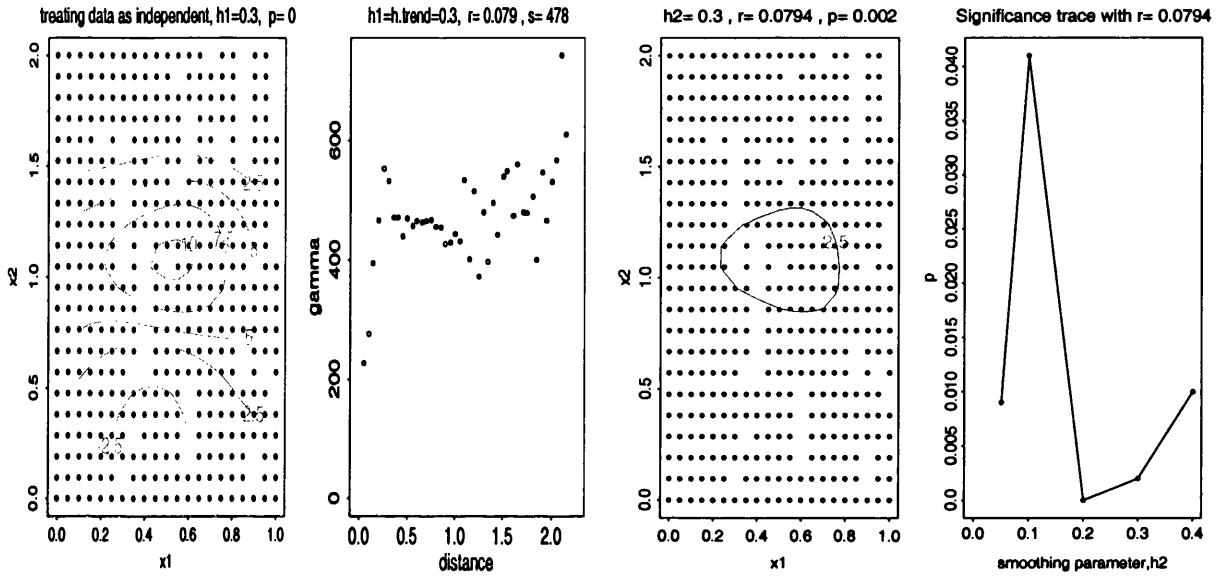

Figure 6.14. Cesium data: Different stages of the DDTP algorithm with $h1 = h2 = 0.3$ and significance trace for different $h2$. From left: contour plot of levels with $st.diff1 \geq 2.5$; empirical variogram of detrended and partitioned data; contour plot with $st.diff2 \geq 2.5$; significance trace obtained using different $h2$.

Figures 6.13 and 6.14 display the results of the test when a smoothing parameter of 0.2 and 0.3 are used in the DDTP algorithm respectively. The final contour plot which takes into account the dependence structure in Figure 6.11 is similar to the one in Figure 6.13, but is different when a higher smoothing parameter of 0.3 is used as displayed in Figure 6.14.

Hence, though using different smoothing parameters does not affect the significance of the test, it does give very different pictures of possible jump location curves. Using a higher smoothing parameter will give observations that have lower final standardised differences. This could be attributed to the fact that the range and sill might be overestimated in the latter.

6.7.4.3 Effects of varying *std*

In this section, the value of *std* that is used for the partitioning process, is varied. Values of $std = 2.5 - 8$ are used and the estimated range and sill are provided in Table 6.22. As the value of *std* increases, more data are included in the variogram fitting.

std	r	s	p-value
2.5	0.0573	331	0.003
3	0.0599	457	0
4	0.0695	478	0.002
5	0.0674	516	0.008
6	0.0634	516	0.008
8	0.0576	516	0.008

Table 6.22. Cesium data: Estimated range and sill for $h1 = h.trend = 0.1$ in the DDTP algorithm using different values of *std*

We can observe that as *std* increases from 2.5 to 4, the estimates of the range increases, but as it increases from 5 to 8, the range decreases. The sill, on the other hand, increases from 331 to 516 with increasing *std*.

As the value of *std* increases, more data are included in the variogram fitting. These might include observations that are quite near to the jump location curve, causing an overestimation of the dependence structure.

However, in this example, the p-value stays significant over a wide range of *std*. Hence the test is not sensitive to this value. A similar study was done using $h1 = h.trend = 0.2$, and the result remains significant regardless of the value of *std* used.

6.7.4.4 Effects of varying the distance of the variogram fitting

The next parameter that might cause a difference in the estimation of the dependence structure, is the distance to model the empirical variogram. We have used four different distances, $maxdist = \{1.12, 1.49, 1.68, 2.4\}$ which are $1/2, 2/3, 3/4, 1$ of the maximum pairwise distance of the spatial locations.

Table 6.23 displays the estimation of the range and sill for $h1 = h.trend = 0.1$ in the DDTP algorithm using different distances, and their corresponding p-value with $h2 = 0.1$. There are only slight differences between the estimation of the range and sill. The p-value remains significant regardless of the distance used for the variogram fitting. Hence, the result is not sensitive to this parameter of modelling the dependence structure.

A similar study was done with $h1 = h.trend = h2 = 0.2$ and the result stays significant

maxdist	r	s	p-value
1.12	0.0601	326	0.005
1.49	0.0589	337	0.004
1.68	0.0586	340	0.004
2.24	0.0573	351	0.003

Table 6.23. Cesium data: Estimated range and sill for $h1 = h.trend = h2 = 0.1$ in the DDTP algorithm using different values of distances for variogram fitting

with differing distances used.

6.7.5 Conclusions

We first used a smoothing parameter of 0.1 in our DDTP algorithm to investigate the presence of discontinuities. The result was significant with p-value less than 0.003. We then did a sensitivity analysis by varying the different parameters used in the algorithm which might affect the test. These include the smoothing parameter; the value to partition the data, *std* and the distance for variogram modeling. As all the various combinations of parameters give rise to estimated ranges of less than 0.09, the test is significant with very low p-values for most cases.

There is strong evidence to suggest that there are discontinuities in this radioactivity map of cesium. This is hardly surprising as the studied area involves different terrains, which 10 years after the accident will reflect different environmental behaviours of cesium fallout. The particular region with high standardised differences that is being picked out in our analysis as the possible jump location curve corresponds to a lake which has very low cesium deposition, compared to its surrounding forestry areas.

Chapter 7

Conclusions

7.1 Introduction

There is no doubt that nonparametric smoothing offers a flexible graphical tool for exploring relationships in data. It is also well-known that ignoring discontinuities can lead to false conclusions in statistical analysis. In the area of change-point analysis, there has been quite a lot of focus in recent years on the estimation of change-points or the discontinuous but otherwise smooth trend. However, what is lacking is an inferential tool to address a primary and vital question of whether the model is continuous or discontinuous. Moreover, in many real-life applications, data that are collected over time or space are often correlated. Therefore, for the purpose of correct inference, incorporating a satisfactory estimation of the correlation into the discontinuity test is of vital importance because it has a considerable influence on the performance of the test. The challenge of this thesis was to develop a discontinuity test in dependent data for one and two-dimensional settings, allowing the trend to be smoothly varying.

Even though the previous chapters contained their respective discussions and conclusions of particular issues, it is helpful to summarise and further discuss those key findings together with suggestions for areas of future research. This concluding chapter is organised as follows: the findings of the one-dimensional case and its proposed future work, followed by the two-dimensional case discussed in a similar manner.

7.2 Discontinuity testing in one-dimensional setting

7.2.1 Summary and Discussion

The discontinuity test proposed by Bowman et al. (2003) exhibits good promise for detecting discontinuities and for inference. By expressing the test statistic as a ratio of quadratic forms, it not only does not need an accurate estimate of error variance, it also does not rely on asymptotic results. Nevertheless, it does require the normality assumption. Furthermore, this test assumes independent errors, which might not be applicable in many real-life applications. As demonstrated in the simulation study in Section 3.4.2, assuming independence of the data when there exists some degree of correlation will inflate the size of the test, thus causing greater false alarms.

Hence, one of the key objectives of this dissertation is to extend the global discontinuity test for correlated data of Bowman et al. (2003) to that of a correlated setting. This was developed and studied in Chapter 3 for the equally spaced setting. An extensive simulation study was conducted in settings of small and moderate sample sizes, to investigate how the various factors influence the test. The underlying correlation was modelled using an AR(1) model. Both the cases of known and unknown correlation were considered. The main findings of the simulation study for the global test were that the power increases in cases where there is an increase in bandwidth; a smoother trend function (flat compared to sine trends); an increase in ratio of jump to standard deviation; a decrease in correlation for positively correlated data, but an increase in correlation for negatively correlated data; and where Rice's variance estimator is used rather than Gasser's particularly for small sample size. Unsurprisingly, the power for the local test is higher than that for the global test. On the other hand, the size of the test is well-behaved for the flat trend and is not sensitive to the choice of bandwidth. However, if the trend is less smooth like the sine trend, oversmoothing will cause a higher rate of false alarm.

In the case where the correlation is unknown, the consequences of inaccurate estimation of the correlation in the possible presence of trend and discontinuities were noted for finite samples. Two algorithms, namely the moving window approach and the residual approach,

were devised to tackle this issue by incorporating a robust estimation of correlation in the discontinuity test. Both approaches performed reasonably well when suitable values of the size of moving window b, and the bandwidth for trend removal, *h.trend* were used respectively, though it was also noted that they were quite sensitive to the choice of these values. The residual approach copes better in the presence of trend as it attempts to first remove the trend before estimating the underlying correlation. However, when a considerable jump is present in the data, using a smooth curve might not be successful in removing it. Its correlation might thus be overestimated, resulting in a decrease in power. Overall, the moving window approach demonstrates a slightly superior performance as it is more stable and performs better than the residual approach when the jump size is large.

After examining the properties of the proposed discontinuity test, its performance was also compared with the isotonic regression test (Wu et al., 2001) via a small simulation study. Different trend functions were studied. The main findings are: if the variance is known and the underlying function and abrupt change is monotonic, then the isotonic regression test is more powerful than the author's discontinuity test, as it takes advantage of the correct but more restrictive assumption of monotonicity. However, its performance is affected if the trend or jump is not monotonic. Furthermore, it is noteworthy that in a practical setting where the variance is unknown, the isotonic regression test with their recommended way of estimating variance, does not work well. The size of their test is inflated under the flat trend. On the other hand, the size of the author's proposed technique using a moderate size of moving window is within limit.

Chapter 4 extended the discontinuity test further to the situation where the data are unequally spaced. The main difference of the test here compared to the equally spaced case is the modelling of the correlation using a variogram instead of an AR(1) model. In addition to the original test statistic, a new test statistic which considers a plug-in error variance estimated via the sill of the variogram, was also developed here. Two new algorithms, the residual-variogram and the double discontinuity test (DDT), which incorporate the modelling of the correlation structure via variogram fitting, were proposed.

From the simulation studies, the performance of the original test was generally superior to the new test. This is most likely due to the fact that the latter requires estimation of both

the unknown correlation and error variance, while the former only requires a satisfactory estimate of the correlation. The key conclusion is that in practice, it is highly recommended that the original test is used unless both the unknown parameters of variance and correlation are known in advance.

The DDT algorithm has shown itself to be an improvement over the ordinary residual-variogram approach. Despite the fact that there is a slight increase in computational load for the former, this is worthwhile as it is not only robust against jumps and trends, but is also capable of producing powers that are comparable or higher than the power obtained using the true correlation. In addition, in the various settings that were considered, it was shown to be relatively insensitive to the smoothing parameter that was used in the initial stage. Though the DDT algorithm has not been used in the simulations for Chapter 3, it can certainly be considered in the equally spaced setting as well. The mechanism of the DDT algorithm is essentially the same, except for the correlation model, which can still be an AR(1).

In fact, the novel idea behind the DDT algorithm was inspired by the findings that the test tends to give a false alarm when the data are treated as independent when they are in fact positively correlated. Hence by including an initial stage of adjusting the data at the most probable change-point if the test is significant under the independence assumption, before removing the trend, this will give an improved estimate of the underlying correlation structure.

There are several difficulties encountered when attempting to model the empirical variogram using nonlinear weighted least squares, to fit an exponential variogram model. In the case where the correlation is low, the algorithm has difficulty in obtaining an estimate of the range, as it is too close to zero. Moreover, the estimation of the correlation appears to be quite sensitive to the different lags used to bin the observations to produce the empirical variogram. Hence, though using the variogram increases the diversity of correlation structures that can be considered, it does bring along with it an additional level of complexity and uncertainty, and is computationally more intensive.

In modelling the underlying correlation, a relatively simple AR(1) and exponential model have been selected for the equally and unequally spaced settings respectively. For the equally

spaced case, higher orders of an autoregressive model can also be considered. However this will involve estimating more parameters which makes the problem more complicated. Since the estimation of the correlation, though crucial, is only a means to an inferential end, a simpler correlation model such as an AR(1) should be adequate. Similar arguments can be given for the unequally spaced case.

The primary goals in Chapter 5 are to apply the developed nonparametric smoothing technique to analyse three real-life data sets, and to compare its results with those of the other two approaches that have also been developed for correlated data, namely the isotonic regression test (Wu et al., 2001), and the Bayesian test (Thomas, 2001). The conclusions drawn from each test will undoubtedly be different, as there are differing features in each approach.

There is however common agreement among all three tests that there is convincing evidence to reject the null hypothesis and conclude that the mean is not constant for the global warming data. The main difference to highlight is that Wu's test suggests that the mean is not constant, possibly due to the monotonic trend of increasing global temperatures. The Bayesian approach on the other hand suggests a plausible abrupt increase in 1929. However, using the nonparametric smoothing approach, the more prominent abrupt changes over a wide range of smoothing parameter were actually abrupt coolings.

In fact, there was no convincing evidence to indicate any presence of abrupt change using the Bayesian test for the other two data-sets. This could possibly be due to the fact that it assumes a constant mean which might not have been applicable in these data-sets. On the other hand, for the Argentina rainfall data, there was convincing evidence of an abrupt change using Wu's test, and the author's proposed test with a small moving window, if a moderate trend is undertaken. As for the River Clyde data, the isotonic regression test was only just significant when the unpenalised likelihood ratio test was used, but was nonsignificant with the penalised one. Its results could possibly be affected by the decreasing trend before 1985. However, using the nonparametric smoothing approach (which is not restricted by the assumption of monotonicity), there was convincing evidence to conclude that there was an abrupt increase in the level of dissolved oxygen level in mid 1985, if the estimated correlation is low and the underlying trend is relatively smooth. This suggests

that the sewage treatment plant has significantly improved the quality of the water after its implementation in 1985.

The significance of the nonparametric regression test will unavoidably be affected by the values of the parameters used in the estimation of the correlation, as well as by the smoothing parameter which attributes the degree of smoothness to the underlying trend. Nonetheless, the main aim of the test is not to give a definitive answer to the choice of these parameters. In many real-life settings, where we are confronted by limited knowledge of the context of the problem, it is difficult to attribute a level of correlation to the data, as well as the smoothness of the trend, in the complex situations of possible discontinuities. Moreover, using a fully data-driven technique might increase the computational load to a substantial magnitude. (This will be raised again in Section 7.2.2.)

Approaching the problem from a different perspective, the preference is to utilise the idea of significance trace (Bowman and Azzalini, 1999) and introduce a graphical plot of significance traces to allow the user to assess and interpret the sensitivity of the test to the different parameters that are used for the estimation of the correlation. In addition, the user-friendly three-in-one graphic referred to as the change-point significance trace (CPST) was devised. This multipurpose tool enables the user to visualise and assess how the significance of the test, the possible change-point locations and the size of the most probable abrupt change, vary with the choice of the smoothing parameter at an attributed level of correlation.

7.2.2 Future Work

In this section, we will discuss some of the possible developments of the proposed discontinuity test in the one-dimensional setting.

First of all, making use of yet another attractive feature of local polynomial regression, a natural extension of the discontinuity test can be to detect a jump in the vth derivative of the regression function. In this case, a degree of at least v of the local polynomials has to be fitted. The correlation can still be estimated using the moving window and residual approaches. However, for the DDT algorithm, we would have to modify the test slightly to consider only the segments that are partitioned by the change-points separately.

As one would expect, though the simulation studies performed in this work have been considerably extensive, one could further explore the influence of various factors on the proposed algorithms. Factors such as greater variety of smooth trend functions, different error variance, multiple change-points and even the effects of misspecification of correlation structure could be considered.

The DDT algorithm indeed shows a lot of promise. Various extensions and improvements can be made to widen its practicality and utility. The following outlines some of the possible future developments.

Despite the drawbacks of the cross-validation method being computationally very intensive and often suffering from high variability (see Hall and Johnstone, 1992), it can still be very appealing to acquire an automatic choice of bandwidth. Moreover, with the improvement in computer technology, the issue with computational load might no longer be such a restrictive factor.

The following outlines how a cross validation approach to select a suitable bandwidth for correlated data can be incorporated into the DDT algorithm.

- First stage
  - Carry out the discontinuity test, assuming that the data are independent. If $p < 0.05$, the most probable change-point location is located and the data are adjusted accordingly.
  - After adjustment, the trend can be presumed to be continuous and smoothly varying. A suitable cross-validation criteria for correlated data (see e.g. Altman (1990) and Yao and Tong (1998)) can be employed to obtain a suitable smoothing parameter, h_{cv}, which is then used to fit a smooth curve to remove the trend.
  - Subsequently, the residuals are used to estimate the underlying correlation via variogram fitting.
- Second stage
  - The discontinuity test is carried out again, accounting for correlation, using the same smoothing parameter h_{cv} as obtained earlier.

This proposal seems to have quite a lot of potential and may emerge to be very useful in practice if proven to work well. Hence it would be worthwhile to explore further simulations to determine its performance in various settings.

Currently, the DDT algorithm only allows for one adjustment. Nevertheless, this can easily be extended to cope better in the presence of multiple change-points, by allowing for several adjustments at locations which have high standardized differences, e.g. above 2.5, though this will only perform well if the data are large enough. This will increase the capability of the DDT algorithm to obtain a better estimate of correlation in the presence of multiple jumps, and hence improve the power of the test.

The algorithm is also specifically devised for positively correlated data. Though, in practice positively correlated data are more common, as one would normally expect correlation to decrease with increasing distance, it may be advantageous to extend this algorithm further to cope with negatively correlated data as well. Instead of carrying out the discontinuity test in the first stage, assuming that the data are independent, one possible strategy could be to first assume that at least one discontinuity exists and estimate the possible locations by means of the standardized differences. The locations which have high standardized differences are removed. Smooth trends are then fitted to each segment sub-divided by the change-points and the residuals thus obtained are used for the estimation of the underlying correlation. Further simulations will be required to ascertain the performance of this proposal.

In addition to the above, a minor modification will be needed in the modelling of the correlation structure. An alternative variogram function which possesses the necessary properties can be used to model the negatively correlated data, instead of the relatively simple exponential variogram function that has been used to model positively correlated data.

Following the rejection of the null hypothesis that there is at least one discontinuity present, it might be of interest to find out the number of change-points present. The reference band and the change-point significance trace have been employed as helpful graphical tools to observe the possible locations of the abrupt changes, and give a subjective idea of the number of possible change-points (i.e. those observations whose standardised differences are greater than 2 or 2.5). Subsequently, smooth trends can be fitted to the segments separately. However, an alternative approach using cross-validation, suggested by Müller and Stadmüller

(1999), can be utilised here too. The concept behind their proposal is simply to choose the number of change-points, d, which would minimise the cross-validation sum of squares

$$CVSS(d) = \sum_{i=1}^{n} (y_i - \hat{g}_d^{(-i)}(x_i))^2 \tag{7.1}$$

at a pre-defined smoothing parameter (which could be selected via a cross validation criteria). The smooth fit $\hat{g}_d^{(-i)}(x_i)$ is estimated at x_i, assuming that there are d change-points, with the exclusion of the data point (x_i, y_i).

7.3 Discontinuity testing in two-dimensional setting

7.3.1 Summary and Discussion

In Chapter 6, the discontinuity test was developed and extended further to a two covariates dependent setting. The aim is to detect discontinuities in regression surfaces using local linear smoothing techniques. In this two-dimensional setting, the discontinuity is considered to take the form of an abrupt change in the mean level over a smooth curve, referred to as the "jump location curve". Identification of the discontinuity thus involves estimation of both its position as well as direction.

The Double Discontinuity Test with Partition (DDTP) algorithm is devised to incorporate the estimation of correlation into the test. The idea is similar to that of the DDT algorithm, where the discontinuity test is carried out in the first stage, assuming that the data are independent. However, instead of then adjusting the data at the detected change-point location, which will be difficult in a spatial setting, the DDTP algorithm opts to remove the data that are near the jump location curve, which presumably would have high standardised differences. Only the remaining data are used for the estimation of the correlation. In finite samples, one possible drawback might be that very few observations are left to estimate the dependence structure. However, one of the appealing properties of the algorithm is its capability to cope with multiple jump locations, which could be of various unknown shapes.

Simulations were conducted both in the equally and unequally spaced setting using finite

sample sizes. Overall, the DDTP works reasonably well in tackling a highly challenging task of detecting discontinuities in a small sample size of 17×17 data-points. It performs better for a lower but moderate range of 0.03. Using appropriate smoothing parameters for both stages gives good size and power that are quite close to the true power for both trend functions, particularly for moderate range. As expected, the algorithm is dependent on a good estimate of the correlation and the smoothing parameter used to obtain the correlation, as well as that attributed to the trend in the final discontinuity test. The test will certainly be more effective in detecting discontinuities if the sample size increases. However, this will also increase its computational workload.

A motivating application for our approach is the analysis of cesium (Cs-137) fallout data in a particular area in Finland. Applying the DDTP algorithm using the same smoothing parameter of 0.1 for both stages, the test was highly significant. A sensitivity analysis was also carried out, by varying the various parameters used in the algorithm, including the smoothing parameter, and the test remained significant. Hence there is convincing evidence to suggest that there are sharply varying areas of deposition. The particular region with high standardized differences that was picked out by the contour plot corresponds to a lake which has very low cesium activity.

7.3.2 Future Work

The above-mentioned suggestions for future work in the one-dimensional case can also be considered here for future research. For instance, extension to discontinuities in the derivatives and more simulations in various settings to evaluate the performance of the DDTP algorithm can be considered. In the case of larger samples sizes, modifications to the DDTP algorithm can be adopted, by revising the order of the trend removal and the partitioning process in the first stage. A suitable cross-validation criteria (e.g. see Francisco-Fernández and Opsomer (2003)) to choose an appropriate smoothing parameter can also be utilised, by only taking into consideration the fitting of smooth trends to the remaining data after the partitioning process in the DDTP algorithm. The simulations have used the exponential variogram model, but similarly, a negative correlation structure can be employed instead.

Whilst this is not a main focus of this work, the estimation of the jump location curve could be further explored. It might also be useful to identify the number of jump location curves. However, extending the cross validation criteria by Müller and Stadmüller (1999) for selecting the number of change-points in the one-dimensional case to the number of jump location curves in the two-dimensional case may not be straightforward. The major hindrance may well be that accurate estimates of the jump location curves are required, and that might only be feasible with a larger sample size.

However, inevitably, the computational load for the two-dimensional setting is substantially higher than the one-dimensional setting. Its future developments might be dependent on improvement in computational power and speed.

7.4 Final Comments

The above ideas for future research are just a small portion of the questions that can be tackled using the proposed discontinuity test, adjusting for correlation using various devised algorithms. On the whole, the research objectives (as stated at the end of Chapter 2) of (1)-(3): developing a discontinuity test that incoporates a robust estimation of correlation in one and two dimensional settings; and (4): providing useful graphical tools to infer their likely locations, have been achieved. Nevertheless, this test still contains immense potential for further development and application.

The importance of inference in detecting discontinuities using nonparametric regression is progressively becoming more recognised today. This thesis has in turn endeavoured to add to the currently emerging interest, by extending it further to cope with correlated data in small to medium sized samples.

Bibliography

Abrupt Climate Change - Inevitable Surprises (2002). National Research Council, National Academy Press. Washington DC.

Altman, N. (1990). Kernel smoothing of data with correlated errors. *Journal of American Statistical Association*, 85:749–759.

Antoch, J., Hušková, M., and Prášková, Z. (1997). Effect of dependence on statistics for determination of change. *Journal of Statistical Planning and Inference*, 60:291–310.

Antoniadis, A. and Gijbels, I. (2002). Detecting abrupt changes by wavelet methods. *Journal of Nonparametric Statistics*, 14:7–29.

Azzalini, A. and Bowman, A. (1993). On the use of nonparametric regression for checking linear relationships. *Journal of Royal Statistical Society Series B*, 55:549–557.

Bagshaw, M. and Johnson, R. (1975). The effect of serial correlation on the performance of cusum tests. *Technometrics*, 17:73–80.

Bai, J. (1994). Least squares estimation of a shift in linear processes. *Journal of Time Series Analysis*, 15:453–472.

Bai, J. (1997). Estimation of a change point in multiple regression models. *The Review of Economics and Statistics*, 79:551–563.

Bai, J. (1999). Likelihood ratio tests for multiple structural changes. *Journal of Econometrics*, 91:299–323.

Bai, J. and Perron, P. (1996). Estimating and testing linear models with multiple structural changes. *Econometrica*, 66:47–78.

Barnard, G. (1959). Control charts and stochastic processes. *Journal of the Royal Statistical Society B*, 21:239–271.

Basseville, M. and Nikiforov, I. (1993). *Detection of abrupt changes: theory and application.* Prentice Hall, Englewood Cliffs, New Jersey.

Bellman, R. (1961). *Adaptive Control Processes.* Princeton University Press, Princeton, NJ.

Besag, J. (1984). On the statistical analysis of dirty pictures (with discussion). *Journal of the Royal Statistical Society Series B*, 48:259–302.

Besag, J., Green, P., Higdon, D., and Mengersen, K. (1995). Bayesian computation and stochastic systems (with discussion). *Statistical Science*, 10:3–66.

Bhattacharya, G. K. and Johnson, R. A. (1968). Nonparametric tests for shift at an unknown time point. *The Annals of Mathematical Statistics*, 39:1731–1743.

Bock, M. (1999). *Methods of inference for nonparametric curves and surfaces.* PhD thesis, University of Glasgow.

Bock, M., Beary, I., and Bowman, A. (2001). Estimation and inference for error variance in univariate and bivariate nonparametric regression. Unpublished manuscript. University of Glasgow.

Bowman, A., Pope, A., and Ismail, B. (2003). Diagnosing discontinuities in nonparametric regression. Unpublished manuscript. University of Glasgow.

Bowman, A. W. and Azzalini, A. (1999). *Applied Smoothing Techniques for Data Analysis.* Oxford Statistical Science Series, Oxford.

Box, G., Jenkins, G., and Reinsel, G. (1995). *Time Series Analysis: Forecasting and Control.* Prentice-Hall, Englewood Cliffs, N.J., 3rd edition.

Box, G. and Tiao, G. (1965). A change in level of a nonstationary time series. *Biometrika*, 52:181–192.

Bracewell, R. (1995). *Two-Dimensional Imaging.* Prentice-Hall, Inc: New Jersey.

Brockwell, P. and Davis, R. (1991). *Time Series: Theory and Methods.* Springer-Verlag, New York, 2nd edition.

Brodsky, B. and Darkhovsky, B. (1993). *Nonparametric Methods in Change-Point Problem.* Kluwer Academic Publishers, The Netherlands.

Brodsky, B. and Darkhovsky, B. (2000). *Non-Parametric Statistical Diagnosis, Problems and Methods.* Kluwer Academic Publishers, The Netherlands.

Burt, D. (2000). *Bandwidth selection concerns for jump discontinuity preservation in the Regression setting using M-smoother and the extension to hypothesis testing.* PhD thesis, State University.

Carlstein, E. (1988). Nonparametric change-point estimation. *The Annals of Statistics*, 16:188–197.

Chen, J. and Gupta, A. (2000). *Parametric statistical change point analysis.* Birkhauser, Boston.

Chernoff, H. and Zacks, S. (1964). Estimating the current mean of a normal distribution which is subject to changes in time. *Annals of Mathematical Statistics*, 35:99–1018.

Chu, C. (1994). Estimation of change points in a nonparametric regression function through kernel density estimation. *Communications Statist - Theory Meth.*, 23:3037–3062.

Chu, C., Glad, I., Godtliebsen, F., and Marron, J. (1998). Edge-preserving smoothers for image processing. *Journal of American Statistical Association*, 93:526–541.

Cline, D., Eubank, R., and Speckman, P. (1995). Nonparametric estimation of regression curves with discontinuous derivatives. *Journal of Statistical Research*, 29:17–30.

Cobb, G. (1978). The problem of the nile: Conditional solution to a changepoint problem. *Biometrika*, 65:243–251.

Cressie, N. (1991). *Statistics for Spatial Data*, volume 2. Wiley, New York.

Cressie, N. and Hawkins, D. (1980). Robust estimation of the variogram. *Mathematical Geology*, 12:115–125.

Csörgő, M. and Horváth, L. (1988). Nonparametric methods for changepoint problems. In Krishnaiah, P. and Rao, C., editors, *Handbook of Statistics*, pages 403–425, Amsterdam. Elsevier.

Darkhovshky, B. (1976). A non-parametric method for the a posteriori detection of the "disorder" time of a sequence of independent random variables. *Annals of Mathematical Statistics*, 39:1731–1741.

Darkhovsky, B. (1994). Nonparametric methods in change-point problems: A general approach and some concrete algorithms. *Change-point problems, IMS Lecture Notes- Monograph series*, 23:99–107.

Davis, W. (1979). Robust methods for detection of shifts of the innovation variance of a time series. *Technometrics*, 21:313–320.

Denison, D., Mallick, B., and Smith, A. (1998). Automatic bayesian curve fitting. *Journal of Royal Statistical Society Series B*, 60:333–350.

Dette, H., Munk, A., and Wagner, T. (1998). Estimating the variance in nonparametric regression - what is a reasonable choice? *Journal of Royal Statistical Society Series B*, 60:751–764.

Diblasi, A. and Bowman, A. (2001). On the use of the variogram in checking for independence in spatial data. *Biometrics*, 1:211–218.

Dümbgen, L. (1991). The asymptotic behaviour of some nonparametric change-point estimators. *The Annals of Statistics*, 19:1471–1495.

Epps, T. (1988). Testing that a gaussian process is stationary. *The Annals of Statistics*, 16:1667–1683.

Eubank, R. and Speckman, P. (1994). Nonparametric estimation of functions with jump discontinuities. In E. Carlstein, H.-G. M. and D.Siegmund, editors, *Change-Point Problems*, volume 23 of *Institute of Mathematical Statistics Lecture Note Series*, pages 130–144.

Fan, J. (1993). Local linear regression smoothers and their minimax efficiencies. *Annals of Statistics*, 21:196–216.

Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications.* Chapman and Hall, London.

Francisco-Fernández, M. and Opsomer, J. (2003). Smoothing parameter selection methods for nonparametric regression with spatially correlated errors. Universidad de A Coruña and Iowa State University.

Francisco-Fernandez, M. and Villar-Fernandez (2001). Local polynomial regression estimation with correlated errors. *Communcations in Statistics - Theory and Methods*, 30:1271–1294.

Gardner, J. L. (1969). On detecting changes in the mean of normal variables. *Annals of Mathematical Statistics*, 40:116–126.

Gasser, T. and Müller, H. (1979). Kernel estimation of regression functions on estimating regression. In Gasser, T. and Rosenblatt, editors, *Smoothing Techniques for Curve Estimation*, pages 23–68, Berlin. Springer-Verlag.

Gasser, T., Sroka, L., and Jennen-Steinmetz, C. (1986). Residual variance and residual pattern in nonlinear estimation. *Biometrika*, 73:625–633.

Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distributions and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741.

Gijbels, I. and Goderniaux, A.-C. (2002a). Bandwidth selection for change point estimation in nonparametric regression. University of Louvain.

Gijbels, I. and Goderniaux, A.-C. (2002b). Bootstrap test for detecting change points in nonparametric regression. University of Louvain.

Gijbels, I., Hall, P., and Kneip, A. (1999). On the estimation of jump points in smooth curves. *Annals of the Institute of Statistical Mathematics*, 51:231–251.

Giraitis, L., Leipus, R., and Surgailis, D. (1996). The change-point problem for dependent observations. *Journal of Statistical Planning and Inference*, 53:297–310.

Gonzalez, R. and Woods, R. (1992). *Digital Image Processing.* Addison-Wesley, Reading, MA.

Green, P. (1995). Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, 82:711–732.

Grégoire, G. and Hamrouni, Z. (2002a). Change-point estimation by local linear smoothing. *Journal of Multivariate Analysis*, 83:56–83.

Grégoire, G. and Hamrouni, Z. (2002b). Two nonparameteric tests for change-point problem. *Journal of Nonparametric Statistics*, 14:87–112.

Hall, P. and Johnstone, I. (1992). Empirical functionals and efficient smoothing parameter selection. *Journal of the Royal Statistical Society B*, 54:475–530.

Hall, P. and Keilegom, I. (2002). Using difference-based methods for inference in nonparameteric regression with time series errors. University of Louvain.

Hall, P. and Marron, J. (1990). On variance estimation in nonparametric regression. *Biometrika*, 77:415–419.

Hall, P. and Rau, C. (2000). Tracking a smooth fault line in a response surface. *Annals of Statistics*, 28:713–733.

Hall, P. and Titterington, P. (1992). Edge-preserving and peak-preserving smoothing. *Technometrics*, 34:429–440.

Hart, J. (1991). Kernel regression estimation with time series errors. *Journal of Royal Statistical Society Series B*, 53:173–187.

Hart, J. (1994). Automated kernel smoothing of dependent data by using time series cross-validation. *Journal of Royal Statistical Society Series B*, 56:529–542.

Hastie, T. and Loader, C. (1993). Discussion. *Statistical Science*, 8:120–143.

Hastie, T. and Tibshirani, R. (1990). *Generalised Additive Models.* Chapman and Hall, London.

Hawkins, D. (2001). Fitting multiple change-point models to data. *Computational Statistics & Data Analysis*, 37:323–341.

Heinrichs, C., Munson, P., Counts, D., Cutler, G., and Baron, J. (1995). Patterns of human growth. *Science*, 268:442–445.

Henderson, R. (1986). Change-point problems with correlated observations, with an application in material accountancy. *Technometrics*, 28:381–389.

Herrmann, E., Gasser, T., and Kneip, A. (1992). Choice of bandwidth for kernel regression when residuals are corelated. *Biometrika*, 79:783–795.

Herrmann, E., Wand, M. P., Engel, J., and Gasser, T. (1995). A bandwidth selector for bivariate kernel regression. *Journal of the Royal Statistical Society Series B*, 57:171–180.

Hinkley, D. (1970). Inference about the change-point in a sequence of random variables. *Biometrika*, 57:1–17.

Horváth, L. and Kokoszka, P. (2002). Change-point detection with non-parameteric regression. *Statistics*, 36:9–31.

Hovgaard, J. and Scott, E. (1997). Comparison of the results of the resume-95 exercise. In *RESUME95: Rapid Environmetal Surveying Using Mobile Equipment*, Copenhagen. NKS, NKS-secretariat.

Inclan, C. and Tiao, G. (1994). Use of cumulative sums of squares for retrospective detection of changes of variances. *Journal of American Statistical Association*, 89:913–923.

IPCC (1995). *Climate Change 1994: Radiative Forcing of Climate Change, and An Evaluation of the IPCC IS92 Emission Scenarios.* Cambridge University Press. Houghton, J.T., Meira Filho, L.G., Bruce, J., Lee, H., Callander, B.A., Haites, E., Harris, N. and Maskell, K. eds.

IPCC (2001). *Climate Change 2001: Working Group I: The Scientific Basis.* Intergovernmental Panel on Climate Change (IPCC).

Issaks, E. and Srivastava, R. (1990). *An Introduction to Applied Geostatistics.* Oxford University Press.

James, B., James, K., and Siegmund, D. (1987). Tests for a change-point. *Biometrika*, 74:71–83.

Jandhyala, V. and MacNeill, I. (1989). Residual partial sum limit processes for regression models with applications to detecting parameter changes at unknown times. *Stochastic Process*, 33:309–323.

Jandhyala, V. and MacNeill, I. (1991). Test for parameter changes at unknown times in linear regression models. *Journal of Statistical Planning and Inference*, 27:291–316.

Jandhyala, V. and MacNeill, I. (1999). Iterated partial sum sequences of regression residuals and tests for changepoints with continuity constraints. *Journal of the Royal Statistical Society Series B*, 59:147–157.

Jarrett, R. (1979). Time intervals between coal mining disasters. *Biometrika*, 66:191–193.

Johnson, N. and Kotz, S. (1972). *Distributions in Statistics: Continuous Univariate Distributions*, volume 2. Wiley, New York.

Jones, P., Parker, D., Osborn, T., and 2001, K. B. (2001). Global and hemispheric temperature anomalies - land and marine instrumental records, In Trends: A Compendium of Data on Global Change. Oak Ridge National Laboratory, U.S. Department of Energy.

Jose, C. and Ismail, B. (1997). Estimation of jump points in nonparametric regression through residual analysis. *Communications in Statistics - Theory and Methods*, 26:2583–2607.

Jose, C. and Ismail, B. (1999). Change points in nonparametric regression functions. *Communications in Statistics - Theory and Methods*, 28:1883–1902.

Jose, C. and Ismail, B. (2001). Nonparametric inference on jump regression surface. *Nonparametric Statistics*, 13:791–813.

Journel, A. and Huijbregts, C. (1978). *Mining Geostatistics.* Academic Press, London.

Kander, Z. and Zacks, S. (1966). Test procedures for possible changes in parameters of statistical distributions ocurring at unknown time points. *The Annals of Mathematical Statistics*, 37:1196–1210.

Kang, K., Koo, J., and Park, C. (2000). Kernel estimation of discontinuous regression function. *Statistics & Probability Letters*, 47:277–285.

Kao, C. and Ross, S. (1995). A cusum test in the linear regression model with serially correlated disturbances. *Economic Reviews*, 14:331–346.

Kauermann, G. (2000). Edge preserving smoothing by local nonparametric maximum likelihood estimation. University of Glasgow.

Kaufman, Y. and Fraser, R. (1997). The effect of smoke particles on clouds and climate forcing. *Science*, 94:1636–1639.

Kim, H. and Siegmund, D. (1989). The likelihood ratio test for a change-point in simple linear regression. *Biometrika*, 79:409–423.

Kim, H.-J. (1996). Change-point detection for correlated observations. *Statistica Sinica*, 6:275–287.

Kim, J. and Hart, J. (1998). Tests for change in a mean function when the data are dependent. *Journal of Time Series Analysis*, 19 No.4:399–424.

Koo, J.-Y. (1997). Spline estimation of discontinuous function. *Journal of Computational and Graphical Statistics*, 6:266–284.

Korostelev, A. and Tsybakov, A. (1993). Minimax theory of image reconstruction. *Lecture Notes in Statistics*, 82.

Krishnaiah, P. and Miao, B. (1988). Review about estimation of change points. In Krishnaiah, P. and Rao, C., editors, *Handbook of Statistics*, pages 375–402, Amsterdam. Elsevier.

Kulperger, R. (1985). On the residuals of autoregressive processes and polynomial regression. *Stochastic Processes and their Applications*, 21:107–118.

Kumar, K. and Wu, B. (2001). Detection of change points in time series analysis with fuzzy statistics. *International Journal of Systems and Science*, 32:1185–1192.

Lai, T. (2001). Sequential analysis: Some classical problems and new challenges. *Statistica Sinica*, 11:303–408.

Lampl, M., Veldhuis, J., and Johnson, M. (1992). Saltation and stasis: A model of human growth. *Science*, 258:801–803.

Lavielle, M. (1999). Detection of multiple changes in a sequence of dependent variables. *Stochastic Processes and their Applications*, 83:79–102.

Lavielle, M. and Moulines, E. (2000). Least-squares estimation of an unknown number of shifts in a time series. *Journal of Time Series Analysis*, 21:35–39.

Lee, S. and Park, S. (2001). The cusum of squares test for scale changes in infinite order moving average processes. *Scandinavian Journal of Statistics*, 28:625–644.

Lee, T. (2002). Automatic smoothing for discontinuous regression function. *Statistica Sinica*, 12:823–842.

Loader, C. (1996). Change-points estimation using nonparametric regression. *Annals of Statistics*, 24:1667–1678.

Lomard, F. (1987). Rank tests for changepoint problems. *Biometrika*, 74:615–624.

MacNeill, I. (1978). Properties of sequences of partial sums of polynomial regression residuals with applications to tests for change in regression at unknown times. *The Annals of Statistics*, 6:422–433.

Matheron, G. (1963). Principles of geostatistics. *Economic Geology*, 58:1246–1266.

McDonald, J. and Owen, A. (1986). Smoothing with split linear fits. *Technometrics*, 28:195–208.

Moskvina, V. (2001). *Application of the singular-spectrum analysis for change-point detection in time series.* PhD thesis, Cardiff University.

Müller, H.-G. (1992). Change-points in nonparametric regression analysis. *Annals of Statistics*, 20:737–761.

Müller, H.-G. (1993). Discussion. *Statistical Science*, 8:134–139.

Müller, H.-G. and Song, K.-S. (1994). Maximin estimation of multidimensional boundaries. *Journal of Multivariate Analysis*, 50:265–281.

Müller, H.-G. and Song, K.-S. (1997). Two-stage change-point estimators in smooth regression models. *Statistics & Probabilty Letters*, 34:323–355.

Müller, H.-G. and Stadmüller, U. (1999). Discontinuous versus smooth regression. *Annals of Statistics*, 27:299–337.

Nadaraya, E. (1964). On estimating regression. *Theory of probability and its applications*, 10:186–190.

Nagaraj, N. (1990). Two-sided tests for change in level for correlated data. *Statistical Papers*, 31:181–194.

Opsomer, J., Wang, Y., and Yang, Y. (2001). Nonparametric regression with correlated errors. *Statistical Science*, 16:134–153.

O'Sullivan and Qian, M. (1994). A regularized contrast statistic for object boundary estimation implementation and statistical evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16:561–570.

Oudshoorn, C. (1998). Asymptotically minimax estimation of a function with jumps. *Bernoulli*, 4:15–33.

Page, E. (1954). Continuous inspection schemes. *Biometrika*, 41:100–115.

Page, E. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika*, 42:523–527.

Peli, T. and Malah, D. (1982). A study of edge detection algorithms. *Computer Graphics and Image Processing*, 20:1–21.

Pettitt, A. (1980). A simple cumulative sum type statistic for the change-point problem. *Biometrika*, 67:79–84.

Picard, D. (1985). Testing and estimating change-points in time series. *Adv. Applied Prob.*, 17:841–867.

Qiu, P. (1994). Estimation of the number of jumps of the jump regression functions. *Communcations in Statistics - Theory and Methods*, 23:2141–2155.

Qiu, P. (1997a). Nonparametric estimation of jump surface. *Sankhya (Series A)*, 59:268–294.

Qiu, P. (1997b). Nonparametric estimation of jump surface. *Change-Point Problems (eds. E. Carlstein, H.-G. Müller and D.Siegmund) Institute of Mathematical Statistics Lecture Note Series*, 23:317–329.

Qiu, P. (1998). Discontinuous regression surface fitting. *Annals of Statistics*, 26:2218–2245.

Qiu, P. (2002). A nonparametric procedure to detect jumps in regression surfaces. *Journal of Computational and Graphical Statistics*, 11:799–822.

Qiu, P. (2003). A jump preserving curve fitting procedure based on local piecewise-linear kernel estimation. *Journal of Nonparametric Statistics*, accepted.

Qiu, P. and Bhandarkar, S. (1996). An edge detection technique using local smoothing and statistical hypothesis testing. *Pattern Recognition Letters*, 17:849–872.

Qiu, P. and Yandell, B. (1997). Jump detection in regression surfaces. *J. Comput. Graph. Statist.*, 6:332–354.

Qiu, P. and Yandell, B. (1998). A local polynomial jump-detection algorithm in nonparametric regression. *Technometrics*, 40:141–152.

Raimondo, M. (1998). Minimax estimation of sharp change points. *Annals of Statistics*, 26:1379–1397.

RESUME-95 (1997). *RESUME-95:Rapid Environmental Surveying Using Mobile Equipment.* Nordic Nuclear Safety Research, NKS-secretariat, Copenhagen.

Rice, J. (1984). Bandwidth choice for nonparametric kernel regression. *Annals of Statistics*, 12:1215–1230.

Rotondi, R. (2002). On the influence of the proposal distributions on a reversible jump mcmc algorithm applied to the detection of multiple change-points. *Computational Statistics & Data Analysis*, 40:633–653.

Rudemo, M. and Stryhn, H. (1994). Boundary estimation for star-shaped objects. *Change-Point Problems, Institute of Mathematical Statistics Lecture Note Series (eds. E. Carlstein, H.-G. Müller and D.Siegmund)*, 23:276–283.

Ruppert, D. and Wand, M. (1994). Multivariate locally weighted least squares regression. *Annals of Statistics*, 22:1346–1370.

Schechtman, E. (1982). A nonparametric test for detecting changes in location. *Communications Statist. - Theory Meth.*, 11,13:1475–1482.

Schechtman, E. and Wolfe, D. (1985). Multiple change points problem - nonparametric procedures for estimation of the points of change. *Communications in Statistics - Simulation and Computation*, 14:615–631.

Shaban, S. (1980). Change point problem and two-phase regression: an annotated bibliography. *International Statistical Review*, 48:83–93.

Shewhart, W. (1931). *Economic Control of Quality of Manufactured Product.* American Society for Quality Control, Milwaukee, Wisconsin.

Siegmund, D. (1986). Boundary crossing probabilities and statistical applications. *The Annals of Statistics*, 14:361–404.

Simonoff, J. (1996). *Smoothing Methods in Statistics.* Springer-Verlag, New York.

Smith, A. (1975). A bayesian approach to inference about a change-point in a sequence of random variables. *Biometrika*, 62:407–416.

Smith, A. and Spiegelhalter, D. (1981). Bayesian approaches to multivariate structure. In V.Barnett, editor, *Interpreting Multivariate Data*, pages 335–348. Wiley, Chicester.

Speckman, P. (1994). Detection of change-points in nonparametric regression. University of Missouri-Columbia.

Speckman, P. (1995). Fitting curves with features: semiparametric change-point methods. *Computing Science and Statistics*, 26:257–264.

Speigelhalter, D. and Smith, A. (1982). Bayes factor for linear and log-linear models with vague prior information. *Journal of the Royal Statistical Society Series B*, 44:377–387.

Spokoiny, V. (1998). Estimation of a function with discontinuities via local polynomial fit with an adaptive window choice. *Annals of Statistics*, 26:1356–1378.

Tang, S. and MacNeill, I. (1993). The effect of serial correlation on tests for parameter change at unknown time. *The Annals of Statistics*, 21:552–575.

Taylor, J. (1987). Using a generalised mean as a measure of location. *Biometrical Journal*, 29:731–738.

Thomas, F. (2001). *A bayesian approach to retrospective detection of change-points in road surface measurements.* PhD thesis, Stockholm University.

Torre, V. and Poggio, T. (1986). On edge detection. *IEEE Trans, Pattern Anal. Mach. Intell.*, 8:147–163.

Trevor, J. H. and Loader, C. (1993). Local regression: automatic kernel carpentry (with discussion). *Statistical Science*, 8:120–143.

Tsay, R. (1986). Nonlinearity tests for time series. *Biometrics*, 73:461–466.

Tsay, R. (1990). Testing and modelling threshold autoregressive processes. *Journal of the American Statistical Association*, 84:231–240.

Tsybakov, A. B. (1994). Multidimensional change-point problems and boundary estimation. *Change-Point Problems (eds. E. Carlstein, H.-G. Müller and D.Siegmund)*, 23:317–329.

Wand, M. and Jones, M. (1995). *Kernel Smoothing.* Chapman and Hall.

Wang, Y. (1995). Jump and sharp cusp detection by wavelets. *Biometrika*, 82:385–397.

Wang, Y. (1998). Change curve estimation via wavelets. *Journal of the American Statistical Association*, 93:163–172.

Wiche, D., Miller, R., and Hsu, D. (1976). Changes of variance in first-order autoregressive time series models - with an application. *Appl. Statist.*, 25:248–256.

Wigley, T., Jones, P., and Raper, S. (1997). The observed global warming record: What does it tell us? *Proc. Natl. Acad. Sci.*, 94:8314–8320.

Wolfe, D. and Schechtman, E. (1984). Nonparametric statistical procedures for the change-point problem. *Journal of Statistical Planning and Inferences*, 9:389–396.

Worsley, K. (1983). The power of likelihood ratio and cumulative sum tests for a change in a bimonial probability. *Biometrika*, 70:455–464.

Worsley, K. (1986). Confidence regions and tests for a change-point in a sequence of exponential family random variables. *Biometrika*, 73:91–104.

Wu, B. and Chen, M.-H. (1999). Use of fuzzy statistical technique in change periods detection of nonlinear time series. *Applied Mathematics and Computation*, 99:241–254.

Wu, J. and Chu, C. (1993). Kernel-type estimators of jump points and values of a regression function. *Annals of Statistics*, 21:1545–1566.

Wu, W., Woodroofe, M., and Mentz, G. (2001). Isotonic regression: another look at the change point problem. *Biometrika*, 88(3):793–804.

Yao, Q. and Tong, H. (1998). Cross-validatory bandwidth selections for regression estimation based on dependent data. *Journal of Statistical Planning and Inferences*, 68:387–415.

Yao, Y. (1988). Estimating the number of change-point by schwarz's criterion. *Statistics & Probability Letters*, 6:181–187.

Yashchin, E. (1993). Performance of cusum control schemes for serially correlated observations. *Technometrics*, 35:37–52.

Zacks, S. (1983). Survey of classical and bayesian approaches to the change-point problem: Fixed sample and sequential procedures of testing and estimation. In Rizvi, M. H., Rustagi, J. S., and Siegmund, D., editors, *Recent Advances in Statistics Papers in Honor of Herman Chernoff on His Sixtieth Birthday*, pages 245–269.

Zimmerman, D. and Zimmerman, M. (1991). A comparison of spatial semivariogram estimators and corresponding ordinary kriging predictors. *Technometrics*, 33:77–92.