SCORE TESTS FOR HOMOGENEITY OF VARIANCES IN LONGITUDINAL TIME SERIES DATA VIA WAVELETS

NALINI MAHALINGAM

Library and Archives Canada

Published Heritage Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-33433-1
Our file Notre référence
ISBN: 978-0-494-33433-1

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

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

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

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

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

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

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

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

Canada

Score Tests for Homogeneity of Variances in Longitudinal Time Series Data via Wavelets

by

©Nalini Mahalingam

A thesis submitted to the School of Graduate Studies in partial fulfillment of the requirement for the Degree of Master of Science in Statistics

Department of Mathematics and Statistics
Memorial University of Newfoundland

St. John's Newfoundland, Canada August 2007

Abstract

We discuss Neyman's partial score test for homogeneity of variances in nonparametric models. We considered two data structures. First, we consider longitudinal data where the observations from each subject are generated from a nonparametric model with heteroscedastic errors. In this context, we found that the discrete wavelet transform approach used by Cai, Hurvich and Tsai (1998) does not lead to a consistent estimate of the mean response function which in turn affects the score statistic. Second, we consider longitudinal data where the observed response from each subject is assumed to be a time series that is nonstationary in mean and variance. The trend component of each series is estimated by a wavelet version of weighted least squares and the residuals are used in estimating the local variances. These estimates are used in a simulation study of the score statistic we construct for testing homoscedasticity in the longitudinal set-up. In the simulation study we examine the size and power of the test.

Acknowledgements

I would like to express my most sincerely gratitude to my supervisor Dr. A. J. Oyet for his valuable guidance, constant encouragement, unbelievable patience, helpful suggestions and arrangement of adequate financial support during my program. Without his assistance, this dissertation could not have been completed.

I am grateful to the School of Graduate Studies and Department of Mathematics and Statistics for their financial support in the form of Graduate Fellowship and Teaching Assistantship. I would like to thank all staff of the Department of Mathematics and Statistics for providing me a friendly atmosphere and the necessary facilities to complete my program.

My sincere thanks go to my mother and brothers for their love, constant encouragement and moral support throughout my study.

I would also like to thank all my friends and well-wishers who encouraged and helped me in numerous ways to complete this study.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The problem of testing for heteroscedasticity in nonparametric regression models or in a time series with deterministic trend has been discussed and very well motivated by several authors. In general, observed data from a nonparametric regression model or time series data with deterministic trend can be represented in the form

$$y_i = f(x_i) + \varepsilon_i \qquad i = 1, ..., n \tag{1.1}$$

where x_i's are equally spaced points. In nonparametric regression, ε_i's are random noise and usually assumed to be normally distributed with mean 0 and constant variance σ^2. For time series data ε_i's are correlated and usually assumed to have the correlation structure of a seasonal autoregressive moving average process. Cai, Hurvich and Tsai (1998) and Kovac and Silverman (2000) argue that the errors may not have constant variance. In the nonparametric regression set-up Kovac and Silverman (2000) assumed that the variance of ε_i is σ_i^2. Cai, Hurvich and Tsai (1998) assumed that the ε_i are independent normal random variables with mean zero and variance $g_i\sigma^2$, where $g_i = g(\mathbf{z}_i, \boldsymbol{\delta})$ is a twice differentiable function of a $p \times 1$ vector of parameters $\boldsymbol{\delta}$ and $\mathbf{z}_i$ is a $p \times 1$ vector of covariates. For examining homogeneity, Cai, Hurvich and Tsai (1998) defined the null hypothesis as $H_0 : \boldsymbol{\delta} = \boldsymbol{\delta_0}$ with the requirement that $g(\mathbf{z}_i, \boldsymbol{\delta_0}) = 1$, $i = 1, 2, ..., n$. In their simulation study, they used $g(\mathbf{z}_i, \boldsymbol{\delta}) = \exp(\mathbf{z}_i\boldsymbol{\delta})$. It follows that $g(\mathbf{z}_i, \boldsymbol{\delta}) = 1$ when $\boldsymbol{\delta} = \boldsymbol{\delta_0} = \mathbf{0}$. Therefore, they considered the score test statistic for testing the null hypothesis $H_0 : \boldsymbol{\delta} = \boldsymbol{\delta_0} = \mathbf{0}$, where the function is estimated by discrete wavelet transformation. They compared the performance of the score statistic and studentised score statistic when the error distribution is normal and non-Gaussian. According to their simulation results, the score statistic has satisfactory power when the sample size is large, but the score statistic performed poorly in controlling the size of the test. Oyet and Sutradhar (2003) found that this result may be due to the fact that the formulation of the null hypothesis $H_0 : \boldsymbol{\delta} = \boldsymbol{\delta_0} = \mathbf{0}$ appears to have serious limitations. They stated that, if it is considered that the noise variances depend on the covariates, it is then reasonable to assume that for $\mathbf{z}_i = \mathbf{z}$, say, for all $i = 1, 2, \ldots, n$ one would expect that $\sigma_i^2 = g(\mathbf{z}_i, \boldsymbol{\delta})\sigma^2$ will be the same. Now, for $\mathbf{z}_i = \mathbf{z}$, it is still possible to have $g(\mathbf{z}_i, \boldsymbol{\delta_0} = \mathbf{0}) = g(\mathbf{z}, \boldsymbol{\delta_0} \neq \mathbf{0}) = 1$.

As opposed to the homogeneity regression model (1.1), Oyet and Sutradhar (2003) defined a heterogeneity regression model as

$$y_i = f(x_i) + \sigma_i\xi_i \qquad i = 1, ..., n \tag{1.2}$$

where ξ_i's are assumed to be normally distributed with mean 0 and variance 1 and $\mathrm{Var}(\varepsilon_i) = \mathrm{Var}(\sigma_i\xi_i) = \sigma_i^2$. They assumed that a group of observations have the same local variance. Suppose that there are q such local groups with variances $\sigma^2_{(1)}, \sigma^2_{(2)}, ..., \sigma^2_{(q)}$ and the jth group has n_j observations, so that $\sum_{j=1}^{q} n_j = n$. Then the null hypothesis for testing for homogeneity in the model (1.2) becomes

$$H_0 \ : \ \sigma^2_{(1)} = \sigma^2_{(2)} = \cdots = \sigma^2_{(q)}. \tag{1.3}$$

Oyet and Sutradhar (2003) then developed a partial score test for testing the null hypothesis H_0 in (1.3) based on Neyman's (1959) score test which requires only a consistent estimate for the nuisance parameter. Contrary to the claim of Cai, Hurvich and Tsai (1998), they found that under this new formulation of the null hypothesis, if the function $f(x)$ is estimated by a wavelet version of weighted least squares instead of discrete wavelet transformation, the statistic performed well in controlling the size of the test. Oyet and Sutradhar (2003) noted that in the context of time series models, the y_i could be treated as arising from a deterministic time series model with trend or seasonal effects represented by $f(x)$.

Earlier, Sutradhar (1996) had discussed the hypothesis (1.3) for q independent time series where each series follows a seasonal autoregressive moving average process. The author assumed that $\boldsymbol{Y}_i \sim N(0, \sigma_i^2 \Sigma)$ where $\boldsymbol{Y}_i = (y_{i1}, y_{i2}, ..., y_{ir})'$, $i = 1, 2, ..., q$ and Σ is a $r \times r$ scalar matrix. In his study, the author compared the performance of Neyman's (1959) score test and Bartlett's test when each of the q time series is assumed to come from an autoregressive process of order 1 (AR(1)) and a moving average process of order 1 (MA(1)).

In this thesis, we study the performance of Neyman's (1959) score statistic under three situations. First, we follow the set-up of Oyet and Sutradhar (2003) by considering observed data from a nonparametric regression model. We then follow Cai, Hurvich and Tsai (1998) to estimate the regression function by discrete wavelet transformation and use the wavelet coefficients at the finest scale to estimate the variance parameter in Chapter 2. We found that the test performed poorly in controlling the size when the function was estimated by discrete wavelet transformation. However, when the true value of the function was used in the test and the variance parameter estimated by discrete wavelet transformation, we found that the test performed well in controlling the size of the test. This result appear to suggest that when using data from a time series containing a deterministic trend component or data generated by a mean response function in a nonparametric regression model, Neyman's score statistic requires a consistent estimate of the nuisance parameter $\sigma^2_{(q)}$ and also of the mean response function $f(x)$ to perform well in both size and power. In Chapter 3, we extend the work of Oyet and Sutradhar (2003) to data arising from a nonparametric regression model with correlated errors. In the context of a time series, we consider k independent time series with length r where each series follows a seasonal autoregressive moving average process. We assume that a group of series follow the same seasonal auto regressive moving average process. Suppose that there are q such groups with different seasonal autoregressive moving average processes, $j = 1, 2, \ldots, q$, where each group j contains p subgroups. The model for the data can be represented as

$$\boldsymbol{Y}_{ij} \;=\; \boldsymbol{f} + \boldsymbol{\varepsilon}_{ij} \qquad i = 1, 2, ..., p \qquad j = 1, 2, ..., q, \tag{1.4}$$

where $\boldsymbol{Y}_{ij} = (y_{ij1}, y_{ij2}, \ldots, y_{ijr})'$, $\boldsymbol{f} = (f(x_1), f(x_2), \ldots, f(x_r))'$ and $\boldsymbol{\varepsilon}_{ij} = (\varepsilon_{ij1}, \varepsilon_{ij2}, \ldots, \varepsilon_{ijr})'$. Here $\boldsymbol{\varepsilon}_{ij}$ follows a seasonal autoregressive moving average process with correlation structure $\sigma^2_{(yj)}\mathbf{R}$. Here $\sigma^2_{(yj)}$ is the variance of ε_{ijh} for a given group j ($j = 1, 2, \ldots, q$) and all $i = 1, 2, \ldots, p$, $h = 1, 2, \ldots, r$, and $\mathbf{R}$ is a $r \times r$ correlation matrix of $\boldsymbol{\varepsilon}_{ij}$.

Neyman's (1959) score test is well known to be asymptotically unbiased in estimating a preassigned level of significance. Also, this test is asymptotically locally most powerful and in general asymptotically equivalent to the likelihood ratio and Wald's tests (Moran (1970)) and does not have any convergence problem for highly correlated data. We prefer Neyman's (1959) score test as this requires $\sqrt{n}$ consistent estimates of the nuisance parameters, which need not be maximum likelihood estimates.

1.1 Some Background on Wavelets

This section is devoted to a brief introduction to the definition and theory of wavelets that will be used in our study. Additional details can be found in Mallat (1989), Meyer (1992), Daubechies (1992), Erlebacher, Hussaini, Jameson (1996), Härdle, Kerkyacharian, Picard, Tsybakov (1998) and Vidakovic (1999).

To define wavelets, we consider two functions: $\phi(x)$, referred to as the scaling function, and $\psi(x)$, commonly called the primary wavelet. Wavelets arise naturally from the multiresolution analysis of the space of square integrable functions $L^2(\mathbb{R})$. From the multiresolution analysis, a dilation equation (two-scale equation or refinement equation) for the scaling function given by

$$\phi(x) = \sum_{k\in\mathbb{Z}} \sqrt{2}\, h_k\, \phi(2x-k), \tag{1.5}$$

is obtained, where $\phi(x)$ is normalized so that $\int_{-\infty}^{\infty} \phi(x)\, dx = 1$. Once the scaling function is found as a solution to (1.5), the primary wavelet $\psi(x)$ is then defined in terms of the scaling function as

$$\psi(x) = \sum_{k\in\mathbb{Z}} \sqrt{2}\, g_k\, \phi(2x-k), \tag{1.6}$$

where $\psi(x)$ satisfies $\int_{-\infty}^{\infty} \psi(x)\, dx = 0$. The coefficients $\{h_k, k \in \mathbb{Z}\}$ are called *filter coefficients* and $\{h_k, k \in \mathbb{Z}\}$ and $\{g_k, k \in \mathbb{Z}\}$ are related by

$$g_k = (-1)^k\, h_{1-k}.$$

By definition, a wavelet system is the collection of translated and dilated versions $\{\phi_{j,k}(x), \psi_{j,k}(x),\ j,k \in \mathbb{Z}\}$ of a scaling function $\phi(x)$ and the primary wavelet $\psi(x)$

with

$$\phi_{j,k}(x) = 2^{j/2}\,\phi(2^j x - k), \qquad j,k \in \mathbb{Z}, \tag{1.7}$$

and

$$\psi_{j,k}(x) = 2^{j/2}\,\psi(2^j x - k), \qquad j,k \in \mathbb{Z}, \tag{1.8}$$

where j is the dilation parameter and k is the translation parameter. The dilations and translations of the wavelet, $\{\psi_{j,k}(x),\ j,k \in \mathbb{Z}\}$, form an orthonormal basis of $L^2(\mathbb{R})$. Therefore, any $f(x) \in L^2(\mathbb{R})$ can be written as

$$f(x) = \sum_{j\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} d_{jk}\,\psi_{j,k}(x). \tag{1.9}$$

Two important properties of filter coefficients are normalization and orthogonality. The normalization condition

$$\sum_{k\in\mathbb{Z}} h_k = \sqrt{2},$$

ensures the existence of a unique solution to (1.5) and (1.6). The orthogonality condition, for any $l \in \mathbb{Z}$

$$\sum_{k\in\mathbb{Z}} h_k\,h_{k-2l} = \delta_l,$$

ensures the orthogonality of the translate of $\phi(x)$. Let the spaces spanned by $\phi_{j,k}(x)$

and $\psi_{j,k}(x)$ over the parameter k, with j fixed, be denoted by V_j and W_j respectively,

$$\begin{aligned} V_j &= span_{k\in\mathbb{Z}}\, \phi_{j,k}(x), \\ W_j &= span_{k\in\mathbb{Z}}\, \psi_{j,k}(x). \end{aligned}$$

It has been shown that (see Vidakovic (1999)) the spaces V_j and W_j are related by

$$\cdots\cdots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \cdots\cdots \tag{1.10}$$

The nested spaces V_j have the following properties.

(i) An intersection that is trivial. That is,

$$\bigcap_{j\in\mathbb{Z}} V_j = \{0\}. \tag{1.11}$$

(ii) A union that is dense in $L^2(\mathbb{R})$. That is,

$$\overline{\bigcup_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R}). \tag{1.12}$$

The spaces V_j and W_j are also related by

$$V_{j+1} = V_j \bigoplus W_j.$$

Here W_j is the orthogonal complement of V_j within the larger space V_{j+1}. That is, any function $f(x) \in V_{j+1}$ can be written as a linear combination or direct sums of functions in V_j and W_j. By iteration, it is easily verified that

$$V_{j+1} = V_0 \oplus \bigoplus_{i=0}^{j} W_i.$$

For any fixed j_0 the decomposition $L^2(\mathbb{R}) = V_{j_0} \oplus \bigoplus_{j=j_0}^{\infty} W_j$ corresponds to the representation

$$f(x) = \sum_{k\in\mathbb{Z}} c_{j_0k}\, \phi_{j_0,k}(x) + \sum_{j\geq j_0} \sum_{k\in\mathbb{Z}} d_{jk}\, \psi_{j,k}(x), \tag{1.13}$$

where due to orthonormality of the wavelets, the coefficients are given by

$$c_{j_0k} = \int_{-\infty}^{\infty} f(x)\, \phi_{j_0,k}(x)\, dx, \qquad d_{jk} = \int_{-\infty}^{\infty} f(x)\, \psi_{j,k}(x)\, dx.$$

The relation (1.9) is called homogeneous wavelet expansion and (1.13) is called inhomogeneous wavelet expansion.

The idea of multiresolution analysis was introduced by Mallat(1989) to obtain the scaling function $\phi(x)$ and the primary wavelet $\psi(x)$. This is one of the most important concepts in discrete wavelet theory. A multiresolution analysis of the space $L^2(\mathbb{R})$ consists of a sequence of nested, closed subspaces as in (1.10).

Definition 1.1.1. A multiresolution analysis of $L^2(\mathbb{R})$ consists of an increasing sequence of closed subspaces V_j, $j \in \mathbb{Z}$, of $L^2(\mathbb{R})$ such that

(a) $\bigcap V_j = \{0\}$,

(b) $\overline{\bigcup V_j} = L^2(\mathbb{R})$,

(c) there exists a scaling function $\phi(x) \in V_0$ such that $\{\phi(x-k),\ k \in \mathbb{Z}\}$ is an orthonormal basis of V_0,

(d) $f(2^j x) \in V_j \Rightarrow f(2^j x - k) \in V_j, \quad \forall k \in \mathbb{Z}$

(e) $f(x) \in V_j \Leftrightarrow f(2x) \in V_{j+1}, \quad \forall j \in \mathbb{Z}$.

The intuitive meaning of (e) is that in passing from V_j to V_{j+1}, the resolution of the approximation is doubled. Mallat (1989) has shown that given any multiresolution analysis, it is possible to derive any function $\psi(x)$ such that the family $\{\psi_{j,k}(x), k \in \mathbb{Z}\}$ is an orthonormal basis of the orthogonal complement W_j of V_j in V_{j+1}, so that

$\{\psi_{j,k}(x), j, k \in \mathbb{Z}\}$ is an orthonormal basis of $L^2(\mathbb{R})$.

Once $\{\psi_{j,k}(x), k \in \mathbb{Z}\}$ is a general basis for W_j, the relation (1.13) is called a multiresolution expansion of $f(x)$. The space W_j is called resolution level of multiresolution analysis. To turn (1.13) into a wavelet expansion one needs to justify the use of (1.8) in (1.13). We have only one resolution level in Fourier analysis and there are many resolution levels in multiresolution analysis.

1.1.1 Some Important Wavelet Bases

Several families of wavelets have been introduced in the wavelet literature by several authors. In this section we discuss only two important families of wavelets. These are the Haar and Daubechies' wavelets.

Haar Wavelet

The Haar wavelet is the simplest wavelet system. The disadvantage of the Haar wavelet is that it is not continuous and therefore not differentiable. The Haar scaling function is defined by

$$\phi(x) = \begin{cases} 1, & 0 \leq x < 1; \\ 0, & \text{othewise.} \end{cases}$$

and the primary wavelet can also be described by a step function:

$$\psi(x) = \begin{cases} 1, & 0 \leq x < 1/2; \\ -1, & 1/2 \leq x < 1; \\ 0, & \text{othewise.} \end{cases}$$

The dilation equation for the Haar scaling function is given by

$$\phi(x) = \phi(2x) + \phi(2x - 1).$$

From (1.5) we deduce that $h_0 = \frac{1}{\sqrt{2}}$, $h_1 = \frac{1}{\sqrt{2}}$ and $h_k = 0$, otherwise. It then follows that $g_0 = \frac{1}{\sqrt{2}}$, $g_1 = -\frac{1}{\sqrt{2}}$ and $g_k = 0$, otherwise. Therefore, from (1.6) the primary wavelet $\psi(x)$ is defined as

$$\psi(x) = \phi(2x) - \phi(2x - 1).$$

Daubechies' Wavelet

Daubechies' was the first to construct compactly supported orthogonal wavelets with a preassigned degree of smoothness. The set $\{\phi_{j,k}(x), j, k \in \mathbb{Z}\}$ of Daubechies' compactly supported dilated and translated versions of the scaling function is an orthonormal system, and the set $\{\psi_{j,k}(x), j, k \in \mathbb{Z}\}$ formed by Daubechies' compactly supported dilated and translated versions of the wavelet function is also an orthonormal basis in $L^2(\mathbb{R})$. The primary wavelet $\psi(x)$ has N vanishing moments which determines the accuracy of approximations based on the wavelet. That is,

$$\int x^n \psi(x)\, dx = 0, \quad n = 0,\, 1,\, ...,\, N - 1.$$

For all versions of the Daubechies' wavelet, the length of the filter coefficients is related to the number of vanishing moments. That is, if L is the length of the filter coefficient then L = $2N$ and $\phi(x)$ and $\psi(x)$ have compact support suppϕ = [0, $2N$ - 1], suppψ = [-N + 1, N].

Daubechies' wavelets are usually denoted by DAUBN, where N is the number of vanishing moments. We note that the DAUB1 wavelet coincides with the Haar wavelet. Figure 1.1 show the plots of DAUBN, (N = 2, 4, 8) scaling and wavelet functions and Table 1.1 gives the filter coefficients for DAUB2-DAUB10 wavelets.

Table 1.1: The $\boldsymbol{h}$ filters for Daubechies' wavelets for $N = 2, \ldots, 10$ vanishing moments

k	DAUB2	DAUB3	DAUB4
0	0.4829629131445342	0.3326705529500827	0.2303778133088966
1	0.8365163037378080	0.8068915093110930	0.7148465705529161
2	0.2241438680420134	0.4598775021184915	0.6308807679298592
3	-0.1294095225512604	-0.1350110200102548	-0.0279837694168604
4		-0.0854412738820267	-0.1870348117190935
5		0.0352262918857096	0.0308413818355607
6			0.0328830116668852
7			-0.0105974017850690

k	DAUB5	DAUB6	DAUB7
0	0.1601023979741926	0.1115407433501095	0.0778520540850092
1	0.6038292697971887	0.4946238903984531	0.3965393194819173
2	0.7243085284377723	0.7511339080210954	0.7291320908462351
3	0.1384281459013216	0.3152503517091976	0.4697822874051931
4	-0.2422948870663808	-0.2262646939654398	-0.1439060039285650
5	-0.0322448695846383	-0.1297668675672619	-0.2240361849938750
6	0.0775714938400454	0.0975016055873230	0.0713092192668303
7	-0.0062414902127983	0.0275228655303057	0.0806126091510831
8	-0.0125807519990819	-0.0315820393174860	-0.0380299369350144
9	0.0033357252854738	0.0005538422011615	-0.0165745416306669
10		0.0047772575109455	0.0125509985560998
11		-0.0010773010853085	0.0004295779729214
12			-0.0018016407040475
13			0.0003537137999745

k	DAUB8	DAUB9	DAUB10
0	0.0544158422431070	0.0380779473638881	0.0266700579005487
1	0.3128715909143165	0.2438346746126514	0.1881768000776480
2	0.6756307362973218	0.6048231236902548	0.5272011889316280
3	0.5853546836542239	0.6572880780514298	0.6884590394535462
4	-0.0158291052563724	0.1331973858249681	0.2811723436606982
5	-0.2840155429615815	-0.2932737832793372	-0.2498464243271048
6	0.0004724845739030	-0.0968407832230689	-0.1959462743773243
7	0.1287474266204823	0.1485407493381040	0.1273693403356940
8	-0.0173693010018109	0.0307256814793158	0.0930573646035142
9	-0.0440882539307979	-0.0676328290613591	-0.0713941471663802
10	0.0139810279173996	0.0002509471148278	-0.0294575368218849
11	0.0087460940474065	0.0223616621236844	0.0332126740593155
12	-0.0048703529934519	-0.0047232047577528	0.0036065535669515
13	-0.0003917403733769	-0.0042815036824646	-0.0107331754833277
14	0.0006754494064506	0.0018476468830567	0.0013953517470513
15	-0.0001174767841248	0.0002303857635232	0.0019924052951842
16		-0.0002519631889428	-0.0006858566949593
17		0.0000393473203163	-0.0001164668551292
18			0.0000935886703200
19			-0.0000132642028945

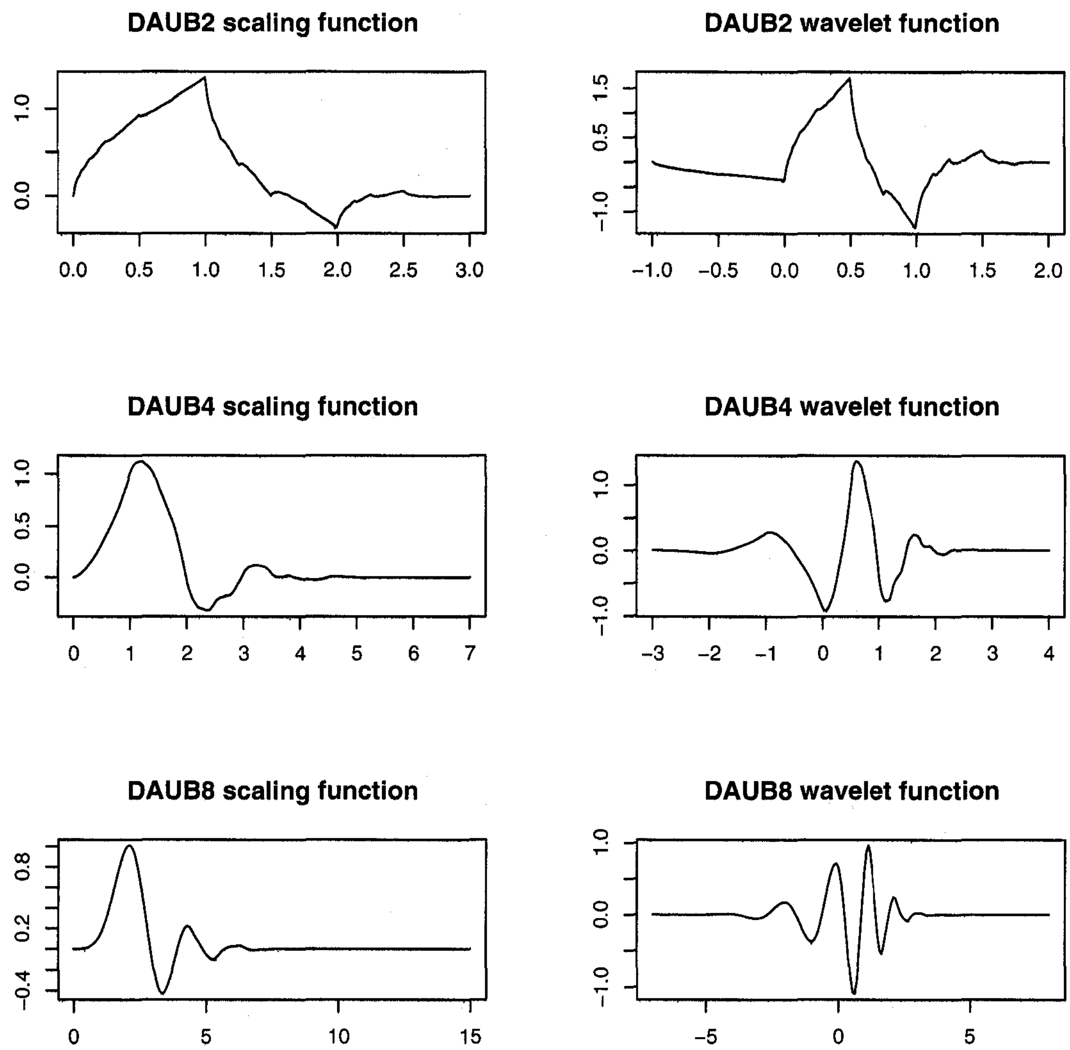

Figure 1.1: Graph of scaling and wavelets functions from Daubechies' family, N = 2, 4, and 8

1.1.2 Wavelet System Construction

The general framework for wavelet system construction is given as follows,

1. Pick a scaling function $\phi(x)$ such that $\{\phi_{0,k}(x), k \in \mathbb{Z}\}$ is an orthonormal basis of V_0, and the relations (1.10), (1.11) and (1.12) are satisfied. Then, $\phi(x)$ generates a multiresolution analysis of $L^2(\mathbb{R})$.

2. Find a primary wavelet $\psi(x) \in W_0$ such that $\{\psi_{0,k}(x), k \in \mathbb{Z}\}$ is an orthonormal basis of W_0. Consequently, $\{\psi_{j,k}(x), k \in \mathbb{Z}\}$ becomes an orthonormal basis of W_j.

3. Conclude that any $f(x) \in L^2(\mathbb{R})$ has the unique representation in terms of an L_2-convergent series:

$$f(x) \;=\; \sum_{k\in\mathbb{Z}} c_{0k}\,\phi_{0,k}(x) + \sum_{j=0}^{\infty}\sum_{k\in\mathbb{Z}} d_{jk}\,\psi_{j,k}(x). \tag{1.14}$$

   The expansion (1.14) starts with the reference space V_0. One can also choose V_{j_0}, for some $j_0 \in \mathbb{Z}$, in place of V_0. Then the inhomogeneous wavelet expansion is of the form (1.13).

Strang (1989) and Pinheiro and Vidakovic (1997) have outlined techniques for the constructions of the scaling function $\phi(x)$. Once $\phi(x)$ is known, we can compute the primary wavelet $\psi(x)$. **Construction 1 - Construction 4** are described below.

Construction 1. Here, we iterate $\phi_j(x) = \sum \sqrt{2}\, h_k \phi_{j-1}(2x - k)$ with the box function as $\phi_0(x)$, that is, $\phi_0(x) = I_{[0,1]}(x)$. When $h_0 = \sqrt{2}$ the boxes get taller and thinner, approximating the delta function. For $h_0 = h_1 = \frac{1}{\sqrt{2}}$ the box is invariant: $\phi_1 = \phi_0$. For $\frac{1}{2\sqrt{2}}, \frac{1}{\sqrt{2}}, \frac{1}{2\sqrt{2}}$ the hat function appears as $j \to \infty$, and $\frac{1}{8\sqrt{2}}, \frac{4}{8\sqrt{2}}, \frac{6}{8\sqrt{2}}, \frac{4}{8\sqrt{2}}, \frac{1}{8\sqrt{2}}$ yields the cubic B-spline. The DAUB2 wavelet has four filter coefficients $\frac{1}{4\sqrt{2}}(1+\sqrt{3})$, $\frac{1}{4\sqrt{2}}(3+\sqrt{3})$, $\frac{1}{4\sqrt{2}}(3-\sqrt{3})$, and $\frac{1}{4\sqrt{2}}(1-\sqrt{3})$. This scaling function $\phi(x)$ leads to orthogonal wavelets.

Construction 2. The second construction takes the Fourier transform of (1.5). To transform the equation, multiply $\phi(x)$ by $e^{-\imath\xi x}$ and integrate with respect to x to obtain

$$\begin{aligned} \int_{-\infty}^{\infty} \phi(x)\, e^{-\imath\xi x}\, dx &= \sum_{k\in\mathbb{Z}} h_k \int_{-\infty}^{\infty} \sqrt{2}\,\phi(2x-k)\, e^{-\imath\xi x}\, dx, \\ \hat{\phi}(\xi) &= \frac{1}{\sqrt{2}} \left(\sum_{k\in\mathbb{Z}} h_k\, e^{-\imath k\xi/2} \right) \int_{-\infty}^{\infty} \phi(y)\, e^{-\imath y\xi/2}\, dy, \\ &= P\left(\frac{\xi}{2}\right) \hat{\phi}\left(\frac{\xi}{2}\right). \end{aligned} \tag{1.15}$$

The symbol $P(\xi) = \frac{1}{\sqrt{2}} \sum_{k\in\mathbb{Z}} h_k\, e^{-\imath k\xi}$ is the crucial function in this theory. Note that $P(0) = 1$ is the normalization condition. Now iterate equation (1.15) at $\xi/2$ to obtain

$$\hat{\phi}(\xi) = P\left(\frac{\xi}{2}\right) \left[P\left(\frac{\xi}{4}\right) \hat{\phi}\left(\frac{\xi}{4}\right) \right].$$

After N iterations, this becomes

$$\hat{\phi}(\xi) = \left[\prod_{j=1}^{N} P\left(\frac{\xi}{2^j}\right) \right] \hat{\phi}\left(\frac{\xi}{2^N}\right).$$

As N $\to \infty$, $\xi/2^N$ is approaching zero, and $\hat{\phi}(0) = \int \phi(x)\, dx = 1$(see (1.15)). Then the iteration leads to the infinite product

$$\hat{\phi}(\xi) = \prod_{j=1}^{\infty} P\left(\frac{\xi}{2^j}\right). \tag{1.16}$$

For $h_0 = \sqrt{2}$ we find P $\equiv 1$ and $\hat{\phi} \equiv 1$, the transform of the delta function. For $h_0 = h_1 = \frac{1}{\sqrt{2}}$ the product of the P's is a geometric series:

$$P\left(\frac{\xi}{2}\right)P\left(\frac{\xi}{4}\right) = \frac{1}{4}(1+e^{-\imath\xi/2})(1+e^{-\imath\xi/4}) = \frac{1-e^{-\imath\xi}}{4(1-e^{-\imath\xi/4})}$$

and

$$\prod_{j=1}^{N} P\left(\frac{\xi}{2^j}\right) = \frac{1-e^{-\imath\xi}}{2^N(1-e^{-\imath\xi/2^N})}.$$

As N $\to \infty$, this approaches the infinite product $(1-e^{-\imath\xi})/(\imath\xi)$. This is $\int_0^1 e^{-\imath\xi x}dx$, the transform of the box function. The hat function comes from squaring $P(\xi)$ which by (1.16) also squares $\hat{\phi}(\xi)$. The cubic B-spline comes from squaring again.

Construction 3. This construction of $\phi(x)$ works directly with the dilation equation (1.5). Suppose ϕ is known at all integers $x = n$, the dilation equation (1.5) gives ϕ at the half-integers (just use the dilation equation (1.5) at $x = n/2$):

$$\phi\left(\frac{n}{2}\right) = \sum_{k\in\mathbb{Z}} \sqrt{2}\,h_k\,\phi(n-k). \tag{1.17}$$

Then the relation (1.17) gives ϕ at the quarter-integers:

$$\phi\left(\frac{n}{4}\right) = \sum_{k\in\mathbb{Z}} \sqrt{2}\,h_k\,\phi(\frac{n}{2}-k)$$

and ultimately at all dyadic points $x = n/2^j$. This is fast to program.

With the four Daubechies coefficients $(h_0, h_1, h_2, h_3) = (\frac{1}{4\sqrt{2}}(1+\sqrt{3}), \frac{1}{4\sqrt{2}}(3+\sqrt{3}), \frac{1}{4\sqrt{2}}(3-\sqrt{3}), \frac{1}{4\sqrt{2}}(1-\sqrt{3}))$, it is usual to set x = 1 and x = 2 in the dilation equation (1.5) and use the fact that $\phi = 0$ unless $0 < x < 3$:

$$\begin{aligned}\phi(1) &= \frac{1}{4}(3+\sqrt{3})\,\phi(1) + \frac{1}{4}(1+\sqrt{3})\,\phi(2),\\ \phi(2) &= \frac{1}{4}(1-\sqrt{3})\,\phi(1) + \frac{1}{4}(3-\sqrt{3})\,\phi(2).\end{aligned}$$

This is $\phi = L\phi$, with matrix entries $L_{\imath j} = h_{2\imath-j}$, which is an eigenvalue problem. The eigenvalues are 1 and $\frac{1}{2}$. The eigenvector for $\lambda = 1$ has components $\phi(1) = \frac{1}{2}(1+\sqrt{3})$, $\phi(2) = \frac{1}{2}(1-\sqrt{3})$. The other eigenvalue $\lambda = \frac{1}{2}$ means that the dilation equation can be differentiated: $\phi'(x) = \sum 2\sqrt{2}\, h_k\, \phi'(2x-k)$ leads similarly to $\phi'(1)$ and $\phi'(2)$. For the hat function, the dilation matrix again has $\lambda = 1,\ \frac{1}{2}$. For the cubic spline the eigenvalues are $1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}$.

Construction 4. In this construction, we describe an algorithm for fast numerical calculation of wavelet values at a given point, based on the Daubechies and Lagarias (1992) local pyramidal algorithm. The Daubechies and Lagarias algorithm enables us to evaluate ϕ and ψ at a point with preassigned precision. We will illustrate the algorithm with wavelets from the Daubechies family. However, the algorithm works for all finite impulse response quadrature mirror filters.

Let ϕ be the scaling function of DAUBN wavelet with support [0, $2N$ - 1]. Let $x \in (0,1)$ and denote the subset of the first n 0-1 digits in the dyadic expansion of x ($x = \sum_{j=1}^{\infty} d_j 2^{-j}$) by *dyad(x,n)*. That is, *dyad(x,n)* $= \{d_1, d_2, ..., d_n\}$.

Let $\boldsymbol{h} = (h_0, h_1, ..., h_{2N-1})$ be the wavelet filter coefficients. Define two ($2N$-1) $\times$ ($2N$-1) matrices as:

$$T_0 = (\sqrt{2}\cdot h_{2i-j-1})_{1\le i,j\le 2N-1} \quad \text{and} \quad T_1 = (\sqrt{2}\cdot h_{2i-j})_{1\le i,j\le 2N-1}. \tag{1.18}$$

Then, the local pyramid algorithm can be constructed based on Theorem 1.

Theorem 1. *(Daubechies and Lagarias, 1992)*

$$\lim_{n\to\infty} T_{d_1}\cdot T_{d_2}\cdots T_{d_n} = \begin{pmatrix} \phi(x) & \phi(x) & \cdots & \phi(x) \\ \phi(x+1) & \phi(x+1) & \cdots & \phi(x+1) \\ \vdots & \vdots & \vdots & \vdots \\ \phi(x+2N-2) & \phi(x+2N-2) & \cdots & \phi(x+2N-2) \end{pmatrix}.$$

The convergence of $\|T_{d_1} \cdot T_{d_2} \cdots T_{d_n} - T_{d_1} \cdot T_{d_2} \cdots T_{d_{n+m}}\|$ to zero, for fixed m, is exponential and constructive, That is, effective decreasing bounds on the error can be established. See Vidakovic (1999) and Pinheiro and Vidakovic (1997) for details.

Example 1. Consider the DAUB2 scaling function ($N = 2$). The corresponding filter is $\boldsymbol{h} = \left(\frac{1+\sqrt{3}}{4\sqrt{2}}, \frac{3+\sqrt{3}}{4\sqrt{2}}, \frac{3-\sqrt{3}}{4\sqrt{2}}, \frac{1-\sqrt{3}}{4\sqrt{2}}\right)$. According to (1.18) the matrices T_0 and T_1 are given as

$$T_0 = \begin{bmatrix} \frac{1+\sqrt{3}}{4} & 0 & 0 \\ \frac{3-\sqrt{3}}{4} & \frac{3+\sqrt{3}}{4} & \frac{1+\sqrt{3}}{4} \\ 0 & \frac{1-\sqrt{3}}{4} & \frac{3-\sqrt{3}}{4} \end{bmatrix} \quad and \quad T_1 = \begin{bmatrix} \frac{3+\sqrt{3}}{4} & \frac{1+\sqrt{3}}{4} & 0 \\ \frac{1-\sqrt{3}}{4} & \frac{3-\sqrt{3}}{4} & \frac{3+\sqrt{3}}{4} \\ 0 & 0 & \frac{1-\sqrt{3}}{4} \end{bmatrix}.$$

If we evaluate the scaling function at an arbitrary point, say $x = 0.45$, then the twenty "decimals" in the dyadic representation of 0.45 are $dyad(0.45{,}20) = \{0,1,1,1,0,0,1,1,0,0,1,1,0,0,1,1,0,0,1,1\}$. The advantage of this procedure is that in addition to the value at 0.45, we also get the values at 1.45 and 2.45. The values $\phi(0.45), \phi(1.45)$, and $\phi(2.45)$ may be approximated as averages of the first, second, and third row, respectively in the matrix

$$\prod_{i \in dyad(0.45,20)} T_i = \begin{bmatrix} 0.86480582 & 0.86480459 & 0.86480336 \\ 0.08641418 & 0.08641568 & 0.08641719 \\ 0.04878000 & 0.04877973 & 0.04877945 \end{bmatrix}.$$

The Daubechies and Lagarias algorithm gives only the values of the scaling function. To find the values of the primary wavelet, we either use (1.6) or apply Theorem 2.

Theorem 2. Let x be an arbitrary real number and let the DAUBN wavelet be given by its filter coefficients $\{h_0, h_1, h_2, ..., h_{2N-1}\}$. Define a vector $\mathbf{u}$ with $2N$ - 1 components as

$$\mathbf{u}(x) = \{(-1)^{1-[2x]}\, h_{i+1-[2x]}, i = 0, ..., 2N-2\}.$$

If for some i the index $i+1-[2x]$ is negative or larger than $2N$ - 1, then the corresponding component of $\mathbf{u}$ is equal to 0.
Let the vector $\mathbf{v}$ be

$$\mathbf{v}(x,n) = \frac{1}{2N-1}\mathbf{1}' \prod_{i \in dyad(\{2x\},n)} T_i,$$

where $\mathbf{1}' = (1, 1, ..., 1)$ is the row-vector of ones. Then,

$$\psi(x) = \lim_{n\to\infty} \mathbf{u}(x)'\, \mathbf{v}(x,n),$$

and the limit is constructive.

We used Construction 4 in this thesis to construct the Daubechies' wavelet system.

1.2 Some Wavelet Methods for Estimating Functions

Wavelet expansion and discrete wavelet transformation are two main wavelet methods for estimating the functions.

1.2.1 Wavelet Expansion

The function f may be expanded in a generalized Fourier series of the form

$$f(x) \;=\; \sum_{k\in\mathbb{Z}} c_{j_0k}\,\phi_{j_0,k}(x) + \sum_{j\geq j_0}\sum_{k\in\mathbb{Z}} d_{jk}\,\psi_{j,k}(x),$$

where the coefficients are given by

$$c_{j_0k} = \int_{-\infty}^{\infty} f(x)\,\phi_{j_0,k}(x)\,dx \qquad\qquad d_{jk} = \int_{-\infty}^{\infty} f(x)\,\psi_{j,k}(x)\,dx.$$

Oyet and Sutradhar (2003) have shown that the coefficients c_{j_0k} and d_{jk} can be estimated by the weighted least squares method and by a modified Gasser-Müller method. In this thesis, we are going to use the weighted least squares method to estimate the wavelet coefficients.

1.2.2 Discrete Wavelet Transformations

An estimator of f by discrete wavelet transformation can be obtained by performing the following three steps.

Step 1. Transform the observations by applying the discrete wavelet transformation.

Discrete wavelet transformations (DWT) are applied to discrete data sets and produce discrete outputs. Discrete wavelet transformations map data from the time domain of the original or input vector to the wavelet domain. The result is a vector of wavelet coefficients of the same size as the input data vector. These transformations are linear and can be defined by matrices of dimension $n \times n$ if they are applied to inputs of size n. When the matrix is orthogonal, the corresponding transformation is a rotation in $\mathbb{R}^n$ in which the data is a point in $\mathbb{R}^n$. The coordinates of the point in the rotated space comprise the discrete wavelet transformation of the original coordinates. Here we provide one example.

Example 2. Let the vector be (1, 2) and let M(1, 2) be the point in $\mathbb{R}^2$ with coordinates given by the data vector. The rotation of the coordinate axes by an angle of $\pi/4$ can be interpreted as a DWT in the Haar wavelet basis. The rotation matrix is

$$W = \begin{pmatrix} \cos\frac{\pi}{4} & \sin\frac{\pi}{4} \\ \cos\frac{\pi}{4} & -\sin\frac{\pi}{4} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}$$

and the discrete wavelet tansformation of $(1,2)'$ is $W \cdot (1,2)' = \left(\frac{3}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)'$.

The change of basis can be performed by matrix multiplication. Therefore, it is possible to define discrete wavelet transformation by matrices. That is, first we have to construct the orthogonal wavelet transformation matrix W. We have already seen a transformation matrix corresponding to Haar forward transformation in Example 2. The construction of W is given as follows:

Let the length of the input observations be 2^J, $\boldsymbol{h} = \{h_s, s \in \mathbb{Z}\}$ be the wavelet filter and N be an appropriate chosen constant. Denote by H_k a matrix of size $(2^{J-k} \times 2^{J-k+1}), k = 1, 2, ...$ with entries

$$h_s, \quad s = (N-1) + (j-1) - 2(i-1) \textit{ modulo } 2^{J-k+1},$$

at the position (i, j). The matrix G_k which is corresponding to the already defined H_k can be obtained by changing h_i by $(-1)^i h_{N+1-i}$. For filters from the Daubechies family, a standard choice for N is the number of vanishing moments.

The matrix $\begin{bmatrix} H_k \\ G_k \end{bmatrix}$ is a basis-changing matrix in 2^{J-k+1} dimensional space; consequently, it is unitary. Therefore,

$$I_{2^{J-k+1}} = [\; H_k' \quad G_k' \;] \cdot \begin{bmatrix} H_k \\ G_k \end{bmatrix} = H_k' \cdot H_k + G_k' \cdot G_k$$

and

$$I = \begin{bmatrix} H_k \\ G_k \end{bmatrix} \cdot [\; H_k' \quad G_k' \;] = \begin{bmatrix} H_k \cdot H_k' & H_k \cdot G_k' \\ G_k \cdot H_k' & G_k \cdot G_k' \end{bmatrix}.$$

This implies,

$$H_k \cdot H_k' = I, \;\; G_k \cdot G_k' = I, \;\; G_k \cdot H_k' = H_k \cdot G_k' = 0 \text{ and } H_k' \cdot H_k + G_k' \cdot G_k = I.$$

Now, for a sequence y the J-step wavelet transformation is $\boldsymbol{d} = W_J \cdot \boldsymbol{y}$, where

$$W_1 = \begin{bmatrix} H_1 \\ G_1 \end{bmatrix}, \quad W_2 = \begin{bmatrix} \begin{bmatrix} H_2 \\ G_2 \end{bmatrix} \cdot H_1 \\ G_1 \end{bmatrix},$$

$$W_3 = \begin{bmatrix} \begin{bmatrix} \begin{bmatrix} H_3 \\ G_3 \end{bmatrix} \cdot H_2 \\ G_2 \end{bmatrix} \cdot H_1 \\ G_1 \end{bmatrix}, \ldots$$

Example 3. Suppose that $\boldsymbol{y} = \{1,\ 0,\ \text{-}3,\ 2,\ 1,\ 0,\ 1,\ 2\}$ and the filter is $\boldsymbol{h} =$

$(h_0, h_1, h_2, h_3) = \left(\frac{1+\sqrt{3}}{4\sqrt{2}}, \frac{3+\sqrt{3}}{4\sqrt{2}}, \frac{3-\sqrt{3}}{4\sqrt{2}}, \frac{1-\sqrt{3}}{4\sqrt{2}}\right)$. Then, $J = 3$ and the matrices H_k and G_k are of dimension $2^{3-k} \times 2^{3-k+1}$.

$$H_1 = \begin{bmatrix} h_1 & h_2 & h_3 & 0 & 0 & 0 & 0 & h_0 \\ 0 & h_0 & h_1 & h_2 & h_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & h_0 & h_1 & h_2 & h_3 & 0 \\ h_3 & 0 & 0 & 0 & 0 & h_0 & h_1 & h_2 \end{bmatrix}$$

$$G_1 = \begin{bmatrix} -h_2 & h_1 & -h_0 & 0 & 0 & 0 & 0 & h_3 \\ 0 & h_3 & -h_2 & h_1 & -h_0 & 0 & 0 & 0 \\ 0 & 0 & 0 & h_3 & -h_2 & h_1 & -h_0 & 0 \\ -h_0 & 0 & 0 & 0 & 0 & h_3 & -h_2 & h_1 \end{bmatrix}.$$

Since,

$$\begin{aligned} H_1 \cdot \boldsymbol{y} &= \{2.19067, -2.19067, 1.67303, 1.15539\} \\ G_1 \cdot \boldsymbol{y} &= \{0.96593, 1.86250, -0.96593, 0.96593\}. \end{aligned}$$

The one-step DAUB2 discrete wavelet transformation of $\boldsymbol{y}$ is
$W_1 \cdot \boldsymbol{y} = \{2.19067, -2.19067, 1.67303, 1.15539 \,|\, 0.96593, 1.86250, -0.96593, 0.96593\}$,

$$H_2 = \begin{bmatrix} h_1 & h_2 & h_3 & h_0 \\ h_3 & h_0 & h_1 & h_2 \end{bmatrix}, \qquad G_2 = \begin{bmatrix} -h_2 & h_1 & -h_0 & h_3 \\ -h_0 & h_3 & -h_2 & h_1 \end{bmatrix}.$$

Since,

$$\begin{aligned} H_2 \cdot H_1 \cdot \boldsymbol{y} &= H_2 \cdot \{2.19067, -2.19067, 1.67303, 1.15539\} \\ &= \{1.68301, 0.31699\}, \\ G_2 \cdot H_1 \cdot \boldsymbol{y} &= G_2 \cdot \{2.19067, -2.19067, 1.67303, 1.15539\} \\ &= \{-3.28109, -0.18301\}, \end{aligned}$$

the two-step DAUB2 discrete wavelet transformation of $\boldsymbol{y}$ is
$W_1 \cdot \boldsymbol{y} = \{1.68301, 0.31699 \,|\, -3.28109, -0.18301 \,|\, 0.96593, 1.86250, -0.96593, 0.96593\}$.

In this example, due to the lengths of the filter and the data, we can perform the transformation for two steps only, W_1 and W_2.

Step 2. Threshold the wavelet coefficients.

Donoho and Johnstone (1994) and Donoho, Johnstone, Kerkyacharian and Picard (1995) proposed the thresholding technique in wavelet analysis. The idea behind thresholding is the removal of small wavelet coefficients considered to be noise. That is, set to 0 the coordinates of a vector $\boldsymbol{d}$ if they are smaller in absolute value than a fixed non-negative number-the threshold λ . There are two thresholding methods frequently used in wavelet theory which are referred to as hard and soft. The expressions for the hard- and soft- thresholding rules are

$$\delta^h(d, \lambda) = d\,\mathbf{1}(|d| > \lambda), \quad \lambda \geq 0,\, d \in \mathbb{R}$$

and

$$\delta^s(d, \lambda) = (d - sgn(d) \cdot \lambda)\,\mathbf{1}(|d| > \lambda), \quad \lambda \geq 0,\, d \in \mathbb{R}$$

respectively. The hard thresholding method keeps some coefficients fixed and sets others to 0 and the soft thresholding method either shrinks coefficients or sets them to 0.

There are several choices for the threshold λ. If the threshold is too small or too large, then the wavelet shrinkage estimator will tend to overfit or underfit the data. Donoho and Johnstone(1994) proposed the universal threshold $\lambda = \sigma\sqrt{2\log n}$. The universal threshold removes noise with high probability that no noise is present in the data after thresholding.

We observe that an estimate of the variance σ^2 is needed for computing the threshold λ. There are several choices for the estimator of σ. Almost all methods involve the wavelet coefficients at the finest scale. The finest scale is only used to estimate the variance of noise. The signal-to-noise ratio $(\mathrm{SNR})^2$ is usually small at high resolutions and, if the signal is not too irregular, the finest scale should contain mainly noise. Moreover, the finest scale contains 50% of all coefficients. Some estimators of σ are

$$s = \sqrt{\frac{1}{n/2-1}\sum_{i=1}^{n/2}\left[d_i^{(J-1)} - \bar{d}^{(J-1)}\right]^2}, \tag{1.19}$$

or a more robust MAD (median absolute deviation from the median) estimator

$$\begin{aligned}\hat{\sigma} &= \frac{MAD[\mathbf{d}^{(J-1)}]}{0.6745},\\ &= \frac{median[|\mathbf{d}^{(J-1)} - median(\mathbf{d}^{(J-1)})|]}{0.6745},\end{aligned} \tag{1.20}$$

where $\mathbf{d}^{(J-1)}$ is the vector of finest detail coefficients associated to the multiresolution subspace W_{J-1}.

Step 3. Invert the DWT to obtain an estimate of the function. Let

$$W = \begin{pmatrix} w_{11} & \cdots & w_{1n} \\ & \vdots & \\ w_{n1} & \cdots & w_{nn} \end{pmatrix}$$

be the discrete wavelet transformation matrix. In terms of W, the wavelet shrinkage estimator of $\boldsymbol{f}$ can be written as

$$\hat{\boldsymbol{f}} = W^{-1}(\hat{\sigma}\,\delta_\lambda(W\mathbf{y}/\hat{\sigma})).$$

Therefore, the component $\hat{f}_i$ can be written as $\sum_k w_{ki}\,(\hat{\sigma}\,\delta_\lambda(W\mathbf{y}/\hat{\sigma})_k)$, $i = 1, ..., n$.

Chapter 2

Partial Score Test for Homogeneity in the Model with Uncorrelated Errors

In this chapter, we assume that a group of observations from (1.2) have the same local variance. Suppose that there are q such local groups with variances $\sigma^2_{(1)}, \sigma^2_{(2)}, ..., \sigma^2_{(q)}$. Let the number of observations in each group be n_j (j = 1, 2, ..., q), so that $\sum_{j=1}^{q} n_j = n$, where n is the total number of observations. In terms of new groupings, x_i in (1.2) represent the hth (h = 1, 2, ..., n_j) time point of the jth (j = 1, 2, ..., q) group such that $x_{jh} \equiv x_{\sum_{u=1}^{j-1} n_u + h} \equiv x_i$. The model (1.2) can then be written as

$$y_{jh} \;=\; f(x_{jh}) + \sigma_{(j)}\xi_{jh} \quad h = 1, ..., n_j, \quad j = 1, ..., q, \qquad (2.1)$$

where ξ_{jh}'s are assumed to be normally distributed with mean 0 and variance 1. Therefore, the null hypothesis for testing the homogeneity in the model can be written as $H_0 : \sigma^2_{(1)} = \sigma^2_{(2)} = \cdots = \sigma^2_{(q)}$. By the reparametrization, $\gamma_j = \frac{\sigma^2_{(j)}}{\sigma^2_{(q)}}$, the null hypothesis reduces to

$$H_0 \ : \ \gamma_1 = \gamma_2 = \cdots = \gamma_{q-1} = 1. \tag{2.2}$$

In what follows, we develop a partial score test statistic for testing the null hypothesis in (2.2). We begin with a definition of the score function and score statistic.

Definition 2.0.1. ***(Rao's Score)*** Given a statistical model $\{f_X(x;\gamma) : \gamma \in \Gamma\}$ with likelihood function $L(\gamma;x)$, a score or score function is defined to be the partial derivative of the logarithm of the likelihood function with respect to the parameter γ. Then the score function U is given by

$$\begin{aligned} U &= \frac{\partial}{\partial\gamma} \log L(\gamma;x), \\ &= \frac{1}{L(\gamma;x)} \frac{\partial}{\partial\gamma} L(\gamma;x). \end{aligned}$$

See Hogg, McKean and Craig (2005) for details. The expected value of U, written $E(U|\gamma)$, is zero. To see this, rewrite the definition of expectation, using the fact that the probability mass function is just $L(\gamma;x)$, which is conventionally denoted by $f(x;\gamma)$ (in which the dependence on x is more explicit). The corresponding cumulative distribution function is denoted as $F(x;\gamma)$. With this change of notation and writing $f_\gamma(x;\gamma)$ for the partial derivative with respect to γ,

$$\begin{aligned} E(U|\gamma) &= \int_{[0,1]} \frac{f'_\gamma(x;\gamma)}{f(x;\gamma)} dF(x;\gamma) \\ &= \int_X \frac{f'_\gamma(x;\gamma)}{f(x;\gamma)} f(x;\gamma)dx \\ &= \int_X \frac{\partial}{\partial\gamma} f(x;\gamma)dx, \end{aligned}$$

where the integral runs over the whole of the probability space of X and a prime denotes partial differentiation with respect to γ. If certain differentiability conditions are met, the integral may be rewritten as

$$\frac{\partial}{\partial\gamma}\int_X f(x;\gamma)dx = \frac{\partial}{\partial\gamma}1 = 0.$$

Because the expectation of the score function is zero, the variance of the score function may be written as

$$\begin{aligned} Var(U|\gamma) &= E\left\{\left[\frac{\partial}{\partial\gamma}\log L(\gamma;x)\right]^2\Bigg|_\gamma\right\}, \\ &= -E\left\{\left[\frac{\partial^2}{\partial\gamma^2}\log L(\gamma;x)\right]\Bigg|_\gamma\right\}. \end{aligned}$$

This is called the Fisher information $I(\gamma)$. When there are p parameters, so that $\boldsymbol{\gamma}$ is a $p \times 1$ vector of parameters ($\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_p)'$) and $\mathbf{U}$ is a $p \times 1$ vector of score function, then the variance of the score function $\mathbf{U}$ is known as the Fisher information matrix $I(\boldsymbol{\gamma})$. The (i,j)th element of the Fisher information matrix can be written

$$\begin{aligned} (I(\boldsymbol{\gamma}))_{i,j} &= E\left[\frac{\partial}{\partial\gamma_i}\log L(\boldsymbol{\gamma};x)\frac{\partial}{\partial\gamma_j}\log L(\boldsymbol{\gamma};X)\right], \\ &= -E\left[\frac{\partial^2}{\partial\gamma_i\partial\gamma_j}\log L(\boldsymbol{\gamma};x)\right]. \end{aligned}$$

Definition 2.0.2. Let $\mathbf{U}(\boldsymbol{\gamma})$ be the vector of first partial derivatives of the log likelihood function with respect to the parameter vector $\boldsymbol{\gamma}$, and let $\mathbf{H}(\boldsymbol{\gamma})$ be the matrix of second partial derivatives of the log likelihood function with respect to $\boldsymbol{\gamma}$. Let $\mathbf{I}(\boldsymbol{\gamma})$ be the expected value of $-\mathbf{H}(\boldsymbol{\gamma})$. Consider a null hypothesis H_o. Let $\hat{\gamma}_o$ be the MLE of $\boldsymbol{\gamma}$ under H_o. The chi-square score statistic for testing H_o is defined by

$$\mathbf{U}'(\hat{\gamma}_o)\mathbf{I}^{-1}(\hat{\gamma}_o)\mathbf{U}(\hat{\gamma}_o)$$

and it has an asymptotic χ^2 distribution with r degrees of freedom under H_o, where r is the number of restrictions on γ by H_o. See Cox and Hinkley (1974) for details.

2.1 Partial Score Statistic

We begin the mathematical development of the score statistic with the likelihood function of the parameters $\sigma^2_{(j)}$ which may be written as

$$L = \prod_{j=1}^{q} \prod_{h=1}^{n_j} \frac{1}{\sqrt{2\pi\sigma^2_{(j)}}} \exp\left[-\frac{1}{2}\left(\frac{(y_{jh}-f(x_{jh}))}{\sigma_{(j)}}\right)^2\right].$$

Clearly, the log likelihood function can then be written as

$$\ell = -\frac{1}{2}\left[n \log(2\pi) + \sum_{j=1}^{q} n_j \log \sigma^2_{(j)} + \sum_{j=1}^{q}\sum_{h=1}^{n_j}\left(\frac{y_{jh}-f(x_{jh})}{\sigma_{(j)}}\right)^2\right].$$

Under the reparametrization (2.2), the log likelihood becomes a function of the parameters γ_j (j = 1, 2, ..., q-1) and $\sigma^2_{(q)}$ given by

$$\begin{aligned}
\ell &= -\frac{1}{2}\left[n \log(2\pi) + \sum_{j=1}^{q} n_j \log(\gamma_j\,\sigma^2_{(q)}) + \sum_{j=1}^{q}\sum_{h=1}^{n_j}\left(\frac{(y_{jh}-f(x_{jh}))^2}{\gamma_j\,\sigma^2_{(q)}}\right)\right], \\
&= -\frac{1}{2}\left[n \log(2\pi) + n \log \sigma^2_{(q)} + \sum_{j=1}^{q} n_j \log \gamma_j + (\sigma^2_{(q)})^{-1}\sum_{j=1}^{q}\sum_{h=1}^{n_j}\gamma_j^{-1}(y_{jh}-f(x_{jh}))^2\right],
\end{aligned}$$

where $\gamma_q = 1$ and y_{jh} and $f(x_{jh})$ are the observations and the value of the trend at time point x_{jh} respectively. Since $\sigma^2_{(q)}$ in γ_j is unknown in practice, it becomes a nuisance parameter and γ_j, j = 1, 2, ..., q-1, are the main parameters of interest for the test.

Define $\gamma = (\gamma_1, \gamma_2, ..., \gamma_{q-1})'$ and let $\hat{\sigma}^2_{(q)}$ be a consistent estimator of $\sigma^2_{(q)}$. Neyman's partial score test is based on the score function

$$U_j(\hat{\sigma}^2_{(q)}) \;=\; \eta_j(\hat{\sigma}^2_{(q)}) - \psi_{j1}\,\xi_1(\hat{\sigma}^2_{(q)}), \quad j = 1, ..., q-1 \tag{2.3}$$

where ψ_{j1} (j = 1, 2, ..., q-1) is the partial regression coefficient of η_j on ξ_1. In (2.3) η_j and ξ_1 are given by

$$\eta_j = \left.\frac{\partial \ell}{\partial \gamma_j}\right|_{\boldsymbol{\gamma}=\mathbf{1}_{q-1}} \qquad \text{and} \qquad \xi_1 = \left.\frac{\partial \ell}{\partial \sigma^2_{(q)}}\right|_{\boldsymbol{\gamma}=\mathbf{1}_{q-1}},$$

respectively. Now,

$$\frac{\partial \ell}{\partial \gamma_j} \;=\; \frac{1}{2}\left[\frac{\sum_{h=1}^{n_j}(y_{jh} - f(x_{jh}))^2}{\gamma_j^2\sigma^2_{(q)}} - \frac{n_j}{\gamma_j}\right].$$

So that

$$\eta_j \;=\; \frac{1}{2}\left[\frac{\sum_{h=1}^{n_j}(y_{jh} - f(x_{jh}))^2}{\sigma^2_{(q)}} - n_j\right]. \tag{2.4}$$

Similarly,

$$\frac{\partial \ell}{\partial \sigma^2_{(q)}} \;=\; \frac{1}{2}\left[\frac{\sum_{j=1}^{q}\sum_{h=1}^{n_j}\gamma_j^{-1}(y_{jh} - f(x_{jh}))^2}{(\sigma^2_{(q)})^2} - \frac{n}{\sigma^2_{(q)}}\right],$$

which leads to

$$\xi_1 = \frac{1}{2\sigma^2_{(q)}} \left[\frac{\sum_{j=1}^{q} \sum_{h=1}^{n_j} (y_{jh} - f(x_{jh}))^2}{\sigma^2_{(q)}} - n \right].$$

The score function in (2.3) may be used to construct the score vector

$$\mathbf{u} = [u_1(\sigma^2_{(q)}), u_2(\sigma^2_{(q)}), ..., u_{q-1}(\sigma^2_{(q)})]' = \boldsymbol{\eta} - \mathbf{d} g^{-1} \xi_1,$$

where $\boldsymbol{\eta} = (\eta_1, ..., \eta_{q-1})'$ is a (q - 1)-dimensional vector with its jth (j = 1, 2, ..., q-1) element given by (2.4). The vector $\mathbf{d} = (d_1, ..., d_{q-1})'$ is also a (q - 1)-dimensional vector with its jth (j = 1, 2, ..., q-1) element and the scalar quantity g defined as

$$d_j = -E\left[\frac{\partial^2 \ell}{\partial \gamma_j \, \partial \sigma^2_{(q)}}\right]_{\boldsymbol{\gamma}=\mathbf{1}_{q-1}} \qquad \text{and} \qquad g = -E\left[\frac{\partial^2 \ell}{\partial (\sigma^2_{(q)})^2}\right]_{\boldsymbol{\gamma}=\mathbf{1}_{q-1}},$$

respectively. By differentiating the log likelihood function, we find that

$$\frac{\partial^2 \ell}{\partial \gamma_j \, \partial \sigma^2_{(q)}} = -\frac{1}{2} \left[\frac{\sum_{h=1}^{n_j} (y_{jh} - f(x_{jh}))^2}{\gamma_j^2 \, (\sigma^2_{(q)})^2} \right].$$

It follows that,

$$E\left[\frac{\partial^2 \ell}{\partial \gamma_j \, \partial \sigma^2_{(q)}}\right] = -\frac{1}{2} \left[\frac{\sum_{h=1}^{n_j} E[(y_{jh} - f(x_{jh}))^2]}{\gamma_j^2 \, (\sigma^2_{(q)})^2} \right]$$

$$= -\frac{1}{2}\left[\frac{\sum_{h=1}^{n_j}\sigma^2_{(j)}}{\gamma_j^2\,(\sigma^2_{(q)})^2}\right]$$

$$= -\frac{1}{2}\left[\frac{n_j}{\gamma_j\,\sigma^2_{(q)}}\right].$$

Therefore, it can be verified that

$$d_j = \frac{n_j}{2\sigma^2_{(q)}}.$$

Similarly, we obtain

$$\frac{\partial^2 \ell}{\partial(\sigma^2_{(q)})^2} = -\frac{1}{2}\left[\frac{2\sum_{j=1}^{q}\sum_{h=1}^{n_j}\gamma_j^{-1}\,(y_{jh}-f(x_{jh}))^2}{(\sigma^2_{(q)})^3} - \frac{n}{(\sigma^2_{(q)})^2}\right].$$

So that,

$$E\left[\frac{\partial^2 \ell}{\partial(\sigma^2_{(q)})^2}\right] = -\frac{1}{2}\left[\frac{2\sum_{j=1}^{q}\sum_{h=1}^{n_j}\gamma_j^{-1}\,E[(y_{jh}-f(x_{jh}))^2]}{(\sigma^2_{(q)})^3} - \frac{n}{(\sigma^2_{(q)})^2}\right]$$

$$= -\frac{1}{2}\left[\frac{2\sum_{j=1}^{q}\sum_{h=1}^{n_j}\gamma_j^{-1}\,\sigma^2_{(j)}}{(\sigma^2_{(q)})^3} - \frac{n}{(\sigma^2_{(q)})^2}\right]$$

$$= -\frac{n}{2\sigma^4_{(q)}},$$

and

$$g = \frac{n}{2\sigma^4_{(q)}}. \tag{2.5}$$

Lemma 1. The variance-covariance matrix of $\mathbf{u}$ is given by $\mathbf{C}$ - $\mathbf{d}g^{-1}\mathbf{d}'$, where $\mathbf{C}_{q-1\times q-1}$ is a diagonal matrix with elements

$$c_{jk} = -E\left[\frac{\partial^2 \ell}{\partial\gamma_j\,\partial\gamma_k}\right]_{\boldsymbol{\gamma}=\mathbf{1}_{q-1}}, \qquad \text{for } j = k.$$

Proof. To compute the variance-covariance matrix of $\mathbf{u}$, we first show that $E(\mathbf{u}) = 0$. Now,

$$E(\mathbf{u}) = E(\boldsymbol{\eta}) - \mathbf{d}g^{-1}E(\xi_1).$$

It is clear that $E(\mathbf{u}) = 0$, since $E(\boldsymbol{\eta}) = 0$ and $E(\xi_1) = 0$. Therefore,

$$\begin{aligned} Cov(\mathbf{u}) &= E\left[(\boldsymbol{\eta} - \mathbf{d}g^{-1}\xi_1)(\boldsymbol{\eta} - \mathbf{d}g^{-1}\xi_1)'\right], \\ &= E\left[\boldsymbol{\eta}\boldsymbol{\eta}' - \boldsymbol{\eta}\mathbf{d}'g^{-1}\xi_1 - \mathbf{d}g^{-1}\xi_1\boldsymbol{\eta}' + \mathbf{d}g^{-2}\xi_1^2\mathbf{d}'\right], \\ &= Cov(\boldsymbol{\eta}) - E(\boldsymbol{\eta}\xi_1)\,g^{-1}\mathbf{d}' - \mathbf{d}g^{-1}\,E(\xi_1\boldsymbol{\eta}') + \mathbf{d}g^{-2}\mathbf{d}'\,Var(\xi_1), \end{aligned} \tag{2.6}$$

where from Definition 2.0.1 $Var(\xi_1) = -E\left[\frac{\partial^2\,\ell}{\partial(\sigma^2_{(q)})^2}\right]_{\boldsymbol{\gamma}=\mathbf{1}_{q-1}} = g$ is given in (2.5) and $Cov(\boldsymbol{\eta}) = \mathrm{C}$ is a matrix with its (j,k)th element defined by

$$c_{jk} = -E\left[\frac{\partial^2\ell}{\partial\gamma_j\,\partial\gamma_k}\right]_{\boldsymbol{\gamma}=\mathbf{1}_{q-1}}.$$

Now, it is easy to verify that $c_{jk} = 0$, $j \neq k$ and

$$\frac{\partial^2 \ell}{\partial \gamma_j^2} = -\frac{1}{2}\left[\frac{2\sum_{h=1}^{n_j}(y_{jh} - f(x_{jh}))^2}{\gamma_j^3\,\sigma_{(q)}^2} - \frac{n_j}{\gamma_j^2}\right].$$

Therefore,

$$\begin{aligned} E\left[\frac{\partial^2 \ell}{\partial \gamma_j^2}\right] &= -\frac{1}{2}\left[\frac{2\sum_{h=1}^{n_j}E[(y_{jh} - f(x_{jh}))^2]}{\gamma_j^3\,\sigma_{(q)}^2} - \frac{n_j}{\gamma_j^2}\right] \\ &= -\frac{1}{2}\left[\frac{2\sum_{h=1}^{n_j}\sigma_{(j)}^2}{\gamma_j^3\,\sigma_{(q)}^2} - \frac{n_j}{\gamma_j^2}\right] \\ &= -\frac{n_j}{2\gamma_j^2}, \end{aligned}$$

and

$$c_{jk} = \frac{n_j}{2}, \qquad \text{for } j = k.$$

Using the fact that $\xi_1 = \dfrac{\sum_{j=1}^{q}\eta_j}{\sigma_{(q)}^2}$, we obtain

$$\begin{aligned} E(\boldsymbol{\eta}\xi_1) &= \frac{1}{\sigma_{(q)}^2}\,E\left[\boldsymbol{\eta}\sum_{j=1}^{q}\eta_j\right], \\ &= \frac{1}{\sigma_{(q)}^2}\,\mathbf{c}, \qquad \textit{where} \quad \mathbf{c} = (c_{11}, ..., c_{q-1,q-1})' \\ &= \mathbf{d}. \end{aligned}$$

By applying these results to (2.6), we obtain

$$Cov(\mathbf{u}) = \mathbf{C} - \mathbf{d}g^{-1}\mathbf{d}'.$$

□

After replacing $\sigma^2_{(q)}$ in $\mathbf{u}$, $\mathbf{C}$, $\mathbf{d}$ and g by $\hat{\sigma}^2_{(q)}$, we obtain Neyman's partial score test statistic as

$$S(\hat{\sigma}^2_{(q)}) = \mathbf{u}'[\mathbf{C} - \mathbf{d}g^{-1}\mathbf{d}']^{-1}\mathbf{u}, \tag{2.7}$$

which is asymptotically distributed as χ^2 with q - 1 degrees of freedom.

2.2 Simulation Study

In this section, we conduct a numerical study to examine the size and power performances of the score test statistic $S(\hat{\sigma}^2_{(q)})$. In our simulation study, we used four mean response functions. These are

(i) A Constant function:
$f(x) = 3$

(ii) A Balanced block function consisting of 16 means:
$f(x) = \sum_{j=1}^{16} h_j \, I_{[2^{-p}(j-1), 2^{-p}j]}(x),$
where $p = 4, h_j = (-1.2, 0.5, 3, 2, 4, 1.5, -1.2, 0.5, 3, 2, 4, 1.5, 3, 2, 4, 1.5)$

(iii) The HeaviSine function:
$f(x) = 4 \sin 4\pi x - sgn(x - 0.3) - sgn(0.72 - x)$ and

(iv) The Doppler test function:
$f(x) = \sqrt{x(1-x)} \sin \frac{2.1\pi}{x + 0.05}$

Figure 2.1 shows the plots of these mean functions. Several authors, including Oyet and Sutradhar (2003), Cai, Hurvich and Tsai (1998) and Donoho and Johnson (1994) have used these functions in their study of the score statistic and in other examples.

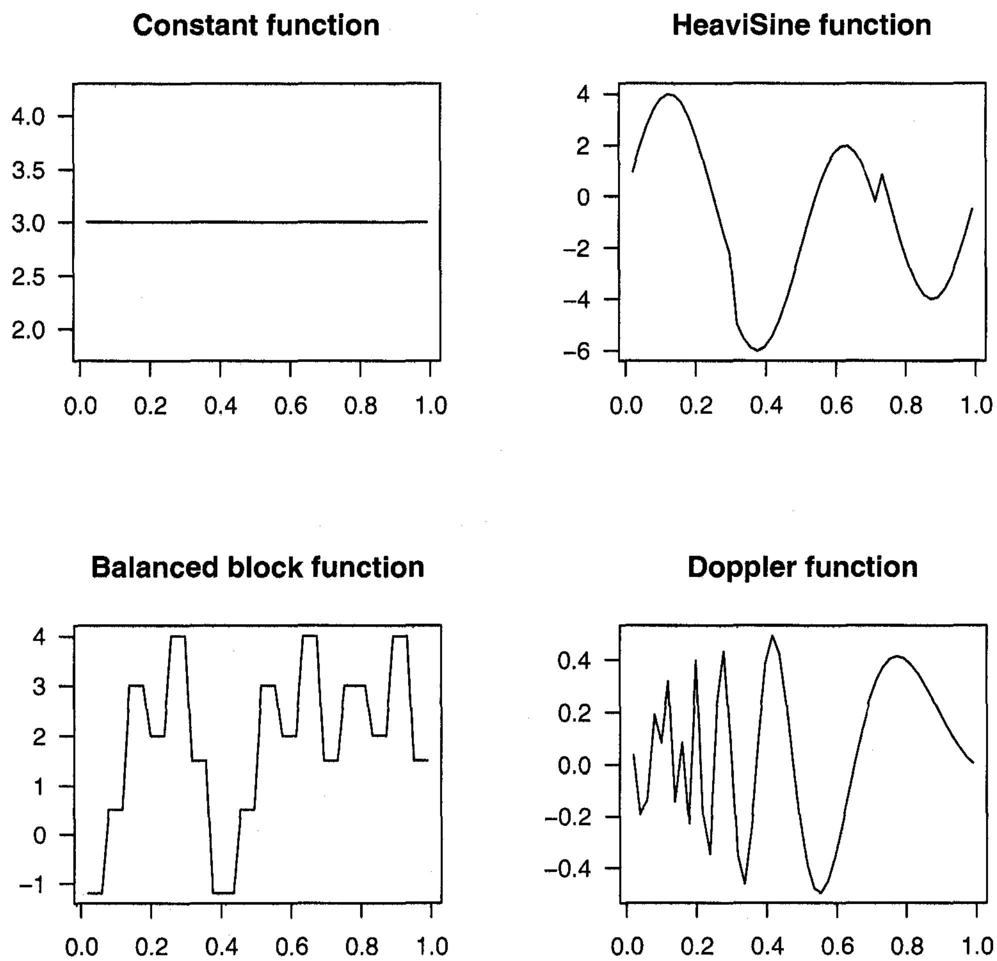

Figure 2.1: Plots of mean functions

It is clear that a simulation study requires the estimation of the mean response function and the nuisance parameter $\sigma^2_{(q)}$. We consider two estimators for the variance

$\sigma^2_{(q)}$ based on discrete wavelet transformation. The first estimator is the standard estimator given in (1.19) and the second estimator is the median absolute deviation (MAD) given in (1.20). The DWT approach described in Chapter 1 was also used in estimating the function. Here, we use the DAUB8 filter coefficient to estimate the variance and function.

2.2.1 Size of the Score Test

To compute the size of the score test, we consider $q = 4$ groups of observations. The distribution of the error terms $\varepsilon_{jh}, h = 1, \ldots, n_j$ and $j = 1, \ldots, 4$ in (2.1) is chosen to be the normal distribution. Therefore, we generate the observations, $y_{jh}, h = 1, \ldots, n_j$ and $j = 1, \ldots, 4$, in (2.1) from a normal distribution with $\sigma^2_{(1)} = \sigma^2_{(2)} = \sigma^2_{(3)} = \sigma^2_{(4)}$ $= 0.8$. Using these observations, we estimate $\sigma^2_{(q)}$ and compute the value of the score statistic $S(\hat{\sigma}^2_{(q)})$ for testing the null hypothesis $H_0 : \gamma_1 = \gamma_2 = \gamma_3 = 1$ under two conditions. These are (a) when the function is known; and (b) when the function is estimated by DWT. The null hypothesis was rejected if the value of score statistic in (2.7) exceeded the 95th percentile of a χ^2 distribution with q - 1 = 3 degrees of freedom. We repeat this process 1000 times and compute the proportion of rejections for the test for a nominal significance level of 5%. The results are shown in Table 2.1.

It is clear from Table 2.1 the score test statistic performed well in controlling the size of the test with the constant, HeaviSine and Doppler mean response functions when the mean functions are assumed to be known and the variance is estimated by the standard method. However, when the Balanced block function is used, more time points are needed for the test statistic to perform well in controlling the size of the test. If the variance is estimated by median absolute deviation method, the score test statistic performed poorly in controlling the size except for the Balanced block function.

Table 2.1: Estimates of size of $H_0 : \gamma_1 = \gamma_2 = \gamma_3 = 1$ at nominal 5% level, with $\sigma^2_{(4)}$ = 0.8

Mean function	n	Known function		Estimated function	
		S-estimate	MAD-estimate	S-estimate	MAD-estimate
Constant	128	0.055	0.107	0.033	0.054
	256	0.054	0.083	0.038	0.047
	512	0.049	0.068	0.035	0.045
	1024	0.055	0.068	0.039	0.045
Balanced block	128	0.009	0.042	0.118	0.125
	256	0.02	0.046	0.161	0.164
	512	0.032	0.047	0.165	0.169
	1024	0.04	0.054	0.159	0.154
HeaviSine	128	0.044	0.095	0.056	0.074
	256	0.044	0.086	0.111	0.139
	512	0.048	0.065	0.151	0.162
	1024	0.054	0.065	0.188	0.208
Doppler	128	0.052	0.096	0.03	0.056
	256	0.056	0.085	0.058	0.076
	512	0.048	0.068	0.048	0.062
	1024	0.055	0.068	0.098	0.104

In all cases, the score test statistic performed poorly in controlling the size when the mean functions are estimated by DWT and the variance is estimated by the standard method. The variance is estimated by median absolute deviation method, the score test statistic performed poorly in controlling the size except the constant function.

According to these results, we can conclude that the score test statistic performed well in controlling the size of the test when the function is known but performed poorly in controlling the size of the test when the function is unknown and estimated by DWT. It follows that the score test statistic performed poorly due to the estimation by DWT. Oyet and Sutradhar(2003) have shown that the test statistic performs well if the function is estimated by a wavelet version of weighted least squares method.

We infer from this that the poor performance of the score statistic in terms of size is as a result of the estimation by DWT. We will therefore use the wavelet version of weighted least squares method in Chapter 3.

2.2.2 Power of the Score Test

To compute the power, we consider $q = 4$ groups of observations with different group variances. We choose four different set of group variances. These are $\sigma^2_{(1)} = 3.2$, $\sigma^2_{(2)} = 2.4$, $\sigma^2_{(3)} = 1.6$ and $\sigma^2_{(4)} = 0.8$; $\sigma^2_{(1)} = 0.8$, $\sigma^2_{(2)} = 0.8$, $\sigma^2_{(3)} = 1.6$ and $\sigma^2_{(4)} = 0.8$; $\sigma^2_{(1)} = 0.8$, $\sigma^2_{(2)} = 0.8$, $\sigma^2_{(3)} = 3.2$ and $\sigma^2_{(4)} = 0.8$; $\sigma^2_{(1)} = 0.8$, $\sigma^2_{(2)} = 2.4$, $\sigma^2_{(3)} = 2.4$ and $\sigma^2_{(4)} = 0.8$. The distribution of the error terms $\varepsilon_{jh}, h = 1, \ldots, n_j$ and $j = 1, \ldots, 4$ in (2.1) is chosen to be normally distributed. We then generate the observations, $y_{jh}, h = 1, \ldots, n_j$ and $j = 1, \ldots, 4$, in (2.1) from a normal distribution with different set of group variances. Using these observations, we compute $\sigma^2_{(q)}$ under the null hypothesis and then compute the value of the score statistic $S(\hat{\sigma}^2_{(q)})$ for testing the null hypothesis $H_0 : \gamma_1 = \gamma_2 = \gamma_3 = 1$ against the specified alternatives $H_1 : \gamma_1 = 4$, $\gamma_2 = 3$, $\gamma_3 = 2$; $H_2 : \gamma_1 = 1$, $\gamma_2 = 1$, $\gamma_3 = 2$; $H_3 : \gamma_1 = 1$, $\gamma_2 = 1$, $\gamma_3 = 4$; H_{41} : $\gamma_1 = 1$, $\gamma_2 = 3$, $\gamma_3 = 3$ under two conditions. These are (a) when the function is known; and (b) when the function is estimated by DWT. The null hypothesis was rejected if the value of score statistic in (2.7) exceeded the 95th percentile of a χ^2 distribution with q - 1 = 3 degrees of freedom. We repeat this process 1000 times. Finally, we compute the proportion of rejections for the test for a nominal significance level of 5%. The results are shown in Table 2.2.

It is clear from Table 2.2 that the power of the score test statistic is higher under constant, HeaviSine and Doppler mean response functions than the Balanced block function for either of the variance estimators. Even for a small sample size of n = 128, the power exceeds 93% with the alternative hypotheses H_3 and H_4 except for the Balanced block function. However, the power is about 86% with alternative hypothesis H_1. We observe that for alternative hypothesis H_2, the score statistic required more time points for the power of the test to be high because the variation between the group variances is small. From these results, we can conclude that the performance of the power of the score test statistic is satisfactory for a small sample size when the variation between the group variances is large.

Table 2.2: Estimates of power of $H_0 : \gamma_1 = \gamma_2 = \gamma_3 = 1$ versus specified alternatives at nominal 5% level

Alternative set	Mean function	n	Known function	
			S-estimate	MAD-estimate
$H_1 : \gamma_1 = 4, \gamma_2 = 3,$	Constant	128	0.87	0.886
$\gamma_3 = 2$		256	1	0.999
	Balanced block	128	0.758	0.788
		256	0.998	0.997
	HeaviSine	128	0.86	0.875
		256	1	0.999
	Doppler	128	0.867	0.879
		256	1	0.999
$H_2 : \gamma_1 = 1, \gamma_2 = 1,$	Constant	128	0.549	0.582
$\gamma_3 = 2$		256	0.857	0.841
		512	0.994	0.993
		1024	1	1
	Balanced block	128	0.325	0.391
		256	0.762	0.775
		512	0.985	0.986
		1024	1	1
	HeaviSine	128	0.512	0.557
		256	0.845	0.845
		512	0.989	0.99
		1024	1	1
	Doppler	128	0.535	0.574
		256	0.854	0.842
		512	0.994	0.992
		1024	1	1

(Table 2.2 Contd....)

Alternative set	Mean function	n	Known function	
			S-estimate	MAD-estimate
$H_3 : \gamma_1 = 1, \gamma_2 = 1, \gamma_3 = 4$	Constant	128	0.986	0.985
		256	1	0.999
	Balanced block	128	0.965	0.959
		256	1	1
	HeaviSine	128	0.985	0.989
		256	1	0.999
	Doppler	128	0.986	0.988
		256	1	1
$H_4 : \gamma_1 = 1, \gamma_2 = 3, \gamma_3 = 3$	Constant	128	0.937	0.939
		256	1	0.999
	Balanced block	128	0.836	0.875
		256	1	0.998
	HeaviSine	128	0.944	0.932
		256	1	1
	Doppler	128	0.932	0.929
		256	1	1

It is clear from Table 2.3 the score test statistic has higher power with constant, HeaviSine and Doppler mean response functions than the Balanced block function for either of the variance estimators. The alternative hypotheses H_3 and H_4 produced the power of the test in the range of 0.777 - 0.87 except for the Balanced block function when the sample size is n = 128. However, the power of the test is about 0.7 under the alternative hypothesis H_1. For alternative hypothesis H_2 with n = 128 the range of the power is 0.303 - 0.383 which confirm that the score statistic needs more time points to achieve a high power because the variation between the group variances is small. From these results, we can conclude that the score test statistic achieves good power with small sample size when the variation between the group variances is large.

Table 2.3: Estimates of power of $H_0 : \gamma_1 = \gamma_2 = \gamma_3 = 1$ versus specified alternatives at nominal 5% level

Alternative set	Mean function	n	Estimated function	
			S-estimate	MAD-estimate
$H_1 : \gamma_1 = 4, \gamma_2 = 3,$ $\gamma_3 = 2$	Constant	128	0.69	0.732
		256	0.993	0.995
		512	1	1
	Balanced block	128	0.573	0.598
		256	0.955	0.951
		512	1	1
	HeaviSine	128	0.7	0.725
		256	0.994	0.993
		512	1	1
	Doppler	128	0.722	0.739
		256	0.995	0.994
		512	1	1
$H_2 : \gamma_1 = 1, \gamma_2 = 1,$ $\gamma_3 = 2$	Constant	128	0.329	0.376
		256	0.731	0.71
		512	0.977	0.971
		1024	1	1
	Balanced block	128	0.38	0.368
		256	0.394	0.44
		512	0.855	0.861
		1024	0.999	0.996
	HeaviSine	128	0.383	0.382
		256	0.761	0.765
		512	0.98	0.982
		1024	1	1
	Doppler	128	0.303	0.324
		256	0.691	0.687
		512	0.955	0.954
		1024	1	1

(Table 2.3 Contd....)

Alternative set	Mean function	n	Estimated function	
			S-estimate	MAD-estimate
$H_3 : \gamma_1 = 1, \gamma_2 = 1,$	Constant	128	0.842	0.804
$\gamma_3 = 4$		256	0.999	0.999
	Balanced block	128	0.743	0.706
		256	0.947	0.931
		512	1	1
	HeaviSine	128	0.87	0.812
		256	0.995	0.994
		512	1	1
	Doppler	128	0.824	0.77
		256	0.99	0.99
$H_4 : \gamma_1 = 1, \gamma_2 = 3,$	Constant	128	0.795	0.807
$\gamma_3 = 3$		256	0.998	0.998
	Balanced block	128	0.633	0.644
		256	0.966	0.972
		512	1	1
	HeaviSine	128	0.831	0.821
		256	0.999	0.999
	Doppler	128	0.777	0.777
		256	0.999	0.999

The power performance when the function is known is high compared to when the function is estimated by DWT. However, the power performance of the test is quite satisfactory when the function is estimated by DWT. But this statistic may be too conservative in controlling the size.

Chapter 3

Partial Score Test for Homogeneity in the Model with Correlated Errors

In this chapter, we discuss the problem of testing for homogeneity in k independent time series each of length r with the ith ($i = 1, 2, \ldots, k$) series represented by a seasonal autoregressive moving average (SARMA) process. Let y_{ih} represent the hth observation in the ith time series. Then, in operator notation, we can write

$$\Phi_U(B^s)\,\phi_u(B)\,y_{ih} \;=\; \theta_v(B)\,\Theta_V(B^s)\,a_{ih}$$

where $\phi_u(B)$ is the autoregressive polynomial of order u in nonnegative power of B, $\theta_v(B)$ is the moving average of polynomial of order v in nonnegative power of B, $\Phi_U(B^s)$ is the autoregressive polynomial of order U in nonnegative power of B^s and $\Theta_V(B^s)$ is the moving average of polynomial of order V in nonnegative power of B^s. That is,

$$
\begin{aligned}
\phi_u(B) &= 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_u B^u \\
\theta_v(B) &= 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_v B^v \\
\Phi_U(B^s) &= 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_U B^{Us} \\
\Theta_V(B^s) &= 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \cdots - \Theta_V B^{Vs}
\end{aligned}
$$

The notation B is the backshift operator such that $B^j y_{ih} = y_{ih-j}$, s is the seasonal period and a_{ih}'s are uncorrelated random variables and assumed to be normally distributed with mean 0 and variance σ_i^2. Throughout this chapter, we will assume that a group of series have the same seasonal auto regressive moving average process. Similar to the grouping in Chapter 2, suppose that there are q such groups with different seasonal auto regressive moving average processes, j = 1, 2, ..., q, where each group j contains p time series. Let $\boldsymbol{Y}_{ij} = (y_{ij1}, y_{ij2}, \ldots, y_{ijr})'$ be the $r \times 1$ vector representing the observations of the ijth (i = 1, 2, ..., p and j = 1, 2, ..., q) time series and $\boldsymbol{f} = (f(x_1), f(x_2), \ldots, f(x_r))'$ be the $r \times 1$ vector of trend values. Then it follows that the time series can be modelled as in (1.4). We assume that $\boldsymbol{\varepsilon}_{ij} \sim N_r(\boldsymbol{0}, \Sigma_j)$ for a given group j and all $i = 1, 2, \ldots, p$, where Σ_j is a $r \times r$ scalar matrix whose elements are functions of ϕ, θ, Φ and Θ. We observe that Σ_j can be written as

$$\Sigma_j = \sigma^2_{(yj)} \boldsymbol{R}_j$$

where $\boldsymbol{R}_j$ is the correlation matrix for a given group j and all $i = 1, 2, \ldots, p$. For simplicity, we have assumed that $\boldsymbol{R}_j = \boldsymbol{R}$ for all $i = 1, 2, \ldots, p$ and $j = 1, 2, \ldots, q$. Then, the data can be modelled as

$$y_{ijh} = f(x_h) + \varepsilon_{ijh}, \quad i = 1, 2, ..., p, \quad j = 1, 2, ..., q, \quad h = 1, 2, ..., r, \tag{3.1}$$

where ε_{ijh} follows the SARMA process.

$$\Phi_U(B^s)\,\phi_u(B)\,\varepsilon_{ijh} \;=\; \theta_v(B)\,\Theta_V(B^s)\,a_{ijh}$$

where a_{ijh} is a sequence of normally distributed white noise processes with mean 0 and variance $\sigma^2_{(j)}$ for a given group j and all i = 1, 2, ..., p and h = 1, 2,..., r. Under the model (3.1), $\boldsymbol{R}$ is a function of ϕ, θ, Φ and Θ. We note that $\sigma^2_{(yj)}$ can be written as $b^*(\phi, \theta, \Phi, \Theta)\,\sigma^2_{(j)}$. That is, $\sigma^2_{(yj)} = b^*(\phi, \theta, \Phi, \Theta)\,\sigma^2_{(j)}$. It follows that $\boldsymbol{\varepsilon}_{ij} \sim N_r(\boldsymbol{0}, b^*\,\sigma^2_{(j)}\boldsymbol{R})$. The null hypothesis for testing homogeneity can then be expressed as in (1.3). By the reparametrization, $\gamma_j = \frac{\sigma^2_{(j)}}{\sigma^2_{(q)}}$, the null hypothesis reduces to (2.2). In what follows, we develop a partial score test statistic for testing the null hypothesis in (2.2), under the model (3.1).

3.1 Partial Score Test for Homogeneity under SARMA Model

When the observations in a time series are highly correlated, the likelihood ratio and Wald's test are known to have convergence problems (see Sutradhar and Bartlett (1993)) in testing the hypothesis (2.2) due to the large values of the correlation parameters (ϕ, θ, Φ or Θ). However, as opposed to the likelihood ratio and Wald's tests, Neyman's (1959) score test requires only $\sqrt{n}$ consistent estimates for $\phi, \theta, \Phi, \Theta$ and $\sigma^2_{(q)}$ under the null hypothesis. We also saw in Chapter 2 that for the score test to perform well in controlling the size of the test a consistent estimator of the trend function is needed.

Based on (3.1), the likelihood function of the parameters $\phi, \theta, \Phi, \Theta$ and $\sigma^2_{(j)}$ may be written as

$$L = \prod_{i=1}^{p} \prod_{j=1}^{q} (2\pi\, b^* \sigma^2_{(j)})^{-\frac{t}{2}}\, |\boldsymbol{R}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2\, b^* \sigma^2_{(j)}} (\boldsymbol{Y}_{ij} - \boldsymbol{f})' \boldsymbol{R}^{-1} (\boldsymbol{Y}_{ij} - \boldsymbol{f}) \right].$$

The log likelihood function becomes

$$\begin{aligned} \ell &= -\frac{1}{2} \left[n \log(2\pi) + n \log(b^*) + pr \sum_{j=1}^{q} \log(\sigma^2_{(j)}) + pq \log|\boldsymbol{R}| + \right. \\ &\quad \left. \sum_{i=1}^{p} \sum_{j=1}^{q} \frac{(\boldsymbol{Y}_{ij} - \boldsymbol{f})' \boldsymbol{R}^{-1} (\boldsymbol{Y}_{ij} - \boldsymbol{f})}{b^* \sigma^2_{(j)}} \right], \end{aligned}$$

where $n = pqr$.

Under the reparametrization (2.2), the log likelihood of the parameters γ_j ($j = 1$, 2, ..., q-1), $\phi, \theta, \Phi, \Theta$ and $\sigma^2_{(q)}$ is written as

$$\begin{aligned} \ell &= -\frac{1}{2} \left[n \log(2\pi) + n \log(b^*) + pr \sum_{j=1}^{q} \log(\gamma_j\, \sigma^2_{(q)}) + pq \log|\boldsymbol{R}| + \right. \\ &\quad \left. trace \left(\sum_{i=1}^{p} \sum_{j=1}^{q} \frac{(\boldsymbol{Y}_{ij} - \boldsymbol{f})' \boldsymbol{R}^{-1} (\boldsymbol{Y}_{ij} - \boldsymbol{f})}{b^* \gamma_j\, \sigma^2_{(q)}} \right) \right], \\ &= -\frac{1}{2} \left[n \log(2\pi) + n \log(b^*) + pr \sum_{j=1}^{q} \log(\gamma_j) + n \log(\sigma^2_{(q)}) + pq \log|\boldsymbol{R}| + \right. \\ &\quad \left. trace \left(\boldsymbol{R}^{-1} \sum_{i=1}^{p} \sum_{j=1}^{q} \frac{(\boldsymbol{Y}_{ij} - \boldsymbol{f})(\boldsymbol{Y}_{ij} - \boldsymbol{f})'}{b^* \gamma_j\, \sigma^2_{(q)}} \right) \right], \\ &= -\frac{1}{2} \left[n \log(2\pi) + n \log(b^*) + pr \sum_{j=1}^{q} \log(\gamma_j) + n \log(\sigma^2_{(q)}) + \right. \\ &\quad \left. pq \log|\boldsymbol{R}| + \frac{1}{b^* \sigma^2_{(q)}} trace\, (\boldsymbol{R}^{-1} \mathrm{W}) \right], \end{aligned}$$

where $\gamma_q = 1$ and $\mathbf{W} = \sum_{i=1}^{p}\sum_{j=1}^{q} \frac{(\boldsymbol{Y}_{ij} - \boldsymbol{f})(\boldsymbol{Y}_{ij} - \boldsymbol{f})'}{\gamma_j}$ is a symmetric matrix of order $r \times r$.

We define $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, ..., \gamma_{q-1})'$ and

$$\begin{aligned}
\boldsymbol{\beta} &= (\beta_1, \ldots, \beta_u, \beta_{u+1}, \ldots, \beta_{u+v}, \beta_{u+v+1}, \ldots, \beta_{u+v+U}, \\
&\quad \beta_{u+v+U+1}, \ldots, \beta_{u+v+U+V}, \beta_{u+v+U+V+1})', \\
&\equiv (\phi_1, \ldots, \phi_u, \theta_1, \ldots, \theta_v, \Phi_1, \ldots, \Phi_U, \Theta_1, \ldots, \Theta_V, \sigma^2_{(q)})'.
\end{aligned}$$

Let $\hat{\boldsymbol{\beta}}$ be some consistent estimator of $\boldsymbol{\beta}$. Neyman's partial score test is then based on

$$U_j(\hat{\boldsymbol{\beta}}) = \eta_j(\hat{\boldsymbol{\beta}}) - \sum_{k=1}^{b} \psi_{jk}\, \xi_k(\hat{\boldsymbol{\beta}}), \quad j = 1, ..., q-1, \tag{3.2}$$

where ψ_{jk} (j = 1, 2, ..., q-1, k = 1, 2, ..., b; b = u+v+U+V+1) is the partial regression coefficient of η_j on ξ_k. We recall that η_j and ξ_k are defined as

$$\eta_j = \left.\frac{\partial \ell}{\partial \gamma_j}\right|_{\boldsymbol{\gamma}=\mathbf{1}_{q-1}} \tag{3.3}$$

and

$$\xi_k = \left.\frac{\partial \ell}{\partial \beta_k}\right|_{\boldsymbol{\gamma}=\mathbf{1}_{q-1}} \tag{3.4}$$

respectively. Next, we use (3.2) to construct the score vector

$$\mathbf{u} = [u_1(\boldsymbol{\beta}), u_2(\boldsymbol{\beta}), ..., u_{q-1}(\boldsymbol{\beta})]' = \boldsymbol{\eta} - \mathbf{D}\mathbf{G}^{-1}\boldsymbol{\xi}, \tag{3.5}$$

where $\boldsymbol{\eta} = (\eta_1, ..., \eta_{q-1})'$ is a (q - 1)-dimensional vector with its jth (j = 1, 2, ..., q-1) element given by (3.3), $\boldsymbol{\xi} = (\xi_1, ..., \xi_b)'$ is also a b-dimensional vector with its kth (k = 1, 2, ..., b) element given by (3.4) and the *(j,k)*th element of $\mathbf{D}_{q-1\times b}$ and $\mathbf{G}_{b\times b}$ defined as

$$d_{jk} = -E\left[\frac{\partial^2 \ell}{\partial\gamma_j\,\partial\beta_k}\right]_{\boldsymbol{\gamma}=\mathbf{1}_{q-1}} \tag{3.6}$$

and

$$g_{jk} = -E\left[\frac{\partial^2 \ell}{\partial\beta_j\,\partial\beta_k}\right]_{\boldsymbol{\gamma}=\mathbf{1}_{q-1}} \tag{3.7}$$

respectively. It can be shown that the variance-covariance matrix of $\mathbf{u}$ is given by $\mathbf{C}$ - $\mathbf{D}\mathbf{G}^{-1}\mathbf{D}'$, where the elements of $\mathbf{C}_{q-1\times q-1}$ are given by

$$c_{jk} = -E\left[\frac{\partial^2 \ell}{\partial\gamma_j\,\partial\gamma_k}\right]_{\boldsymbol{\gamma}=\mathbf{1}_{q-1}}.$$

After replacing $\boldsymbol{\beta}$ in $\mathbf{u}$, $\mathbf{C}$, $\mathbf{D}$ and $\mathbf{G}$ by $\hat{\boldsymbol{\beta}}$, we obtain Neyman's partial score test statistic as

$$S(\hat{\boldsymbol{\beta}}) = \mathbf{u}'[\mathbf{C} - \mathbf{D}\mathbf{G}^{-1}\mathbf{D}']^{-1}\mathbf{u}, \tag{3.8}$$

which is asymptotically distributed as χ^2 with q - 1 degrees of freedom.

3.2 Partial Score Test for Homogeneity under AR(1) Model

In this Section, we consider a special case of the SARMA model in which $\phi_2 = \phi_3 = \cdots = \phi_u = 0, \theta_i = 0$ for all $i = 1, 2, \ldots, v$, $\Phi_j = 0$ for all $j = 1, 2, \ldots, U$ and $\Theta_k = 0$

for all $k = 1, 2, \ldots, V$. In this case the SARMA model reduces to the autoregressive model of order 1, AR(1). Here, we consider q groups of heterogeneous time series where each group contains p time series and each series follows the AR(1) process. That is, the model for the term ε_{ijh} in (3.1) reduces to $\varepsilon_{ijh} = \phi\,\varepsilon_{ijh-1} + a_{ijh}$ where -1 $< \phi <$ 1 is the parameter of the process and a_{ijh}'s are independent and identically normally distributed random variables with mean zero and variance $\sigma^2_{(j)}$ for a given group j (j = 1, 2, ..., q) and all i = 1, 2, ..., p, and h = 1, 2, ..., r. Since the AR(1) process is stationary, it is straightforward to see that $E(\varepsilon_{ijh}) = 0$ and

$$\begin{aligned}
Var(\varepsilon_{ijh}) &= E(\varepsilon^2_{ijh}),\\
&= E[(\phi\,\varepsilon_{ijh-1} + a_{ijh})^2],\\
&= E(\phi^2\,\varepsilon^2_{ijh-1} + 2\,\phi\,\varepsilon_{ijh-1}\,a_{ijh} + a^2_{ijh}),\\
&= \phi^2 Var(\varepsilon_{ijh}) + 0 + \sigma^2_{(j)},\\
&= \frac{\sigma^2_{(j)}}{1-\phi^2}.
\end{aligned}$$

with auto-covariance function given by

$$Cov(\varepsilon_{ijh}, \varepsilon_{ijh+k}) = E(\varepsilon_{ijh}\,\varepsilon_{ijh+k}).$$

For k = 1,

$$\begin{aligned}
Cov(\varepsilon_{ijh}, \varepsilon_{ijh+1}) &= E(\varepsilon_{ijh}\,\varepsilon_{ijh+1}),\\
&= E[\varepsilon_{ijh}(\phi\,\varepsilon_{ijh} + a_{ijh+1})],\\
&= \phi\,Var(\varepsilon_{ijh}),\\
&= \frac{\phi\,\sigma^2_{(j)}}{1-\phi^2}.
\end{aligned}$$

For k = 2,

$$\begin{aligned} Cov(\varepsilon_{ijh}, \varepsilon_{ijh+2}) &= E(\varepsilon_{ijh}\, \varepsilon_{ijh+2}), \\ &= E[\varepsilon_{ijh}(\phi\, \varepsilon_{ijh+1} + a_{ijh+2})], \\ &= \phi\, Cov(\varepsilon_{ijh}, \varepsilon_{ijh+1}), \\ &= \frac{\phi^2\, \sigma^2_{(j)}}{1-\phi^2}. \end{aligned}$$

In general, for the AR(1) process ε_{ijh}, we find that the autocovariance function can be written as

$$Cov(\varepsilon_{ijh}, \varepsilon_{ijh'}) = \frac{\phi^{|h-h'|}\, \sigma^2_{(j)}}{1-\phi^2}.$$

It follows that $\boldsymbol{\varepsilon}_{ij} \sim N_r\left(\mathbf{0}, \dfrac{\sigma^2_{(j)}}{1-\phi^2}\boldsymbol{R}(\phi)\right)$, where $\boldsymbol{R}(\phi) = \phi^{|h-h'|}$ is the $r \times r$ correlation matrix.

We noted earlier that for the SARMA model $\sigma^2_{(yj)} = b^*(\phi, \theta, \Phi, \Theta)\, \sigma^2_{(j)}$. For the AR(1) model, $b^*(\phi, \theta, \Phi, \Theta) = \dfrac{1}{1-\phi^2}$. Therefore, replacing $b^*(\phi, \theta, \Phi, \Theta)$ by $\dfrac{1}{1-\phi^2}$ in the loglikelihood function in Section 3.1 we obtain

$$\begin{aligned} \ell &= -\frac{1}{2}\Bigg[n \log(2\pi) - n \log(1-\phi^2) + pr\sum_{j=1}^{q} \log(\gamma_j) + n \log(\sigma^2_{(q)}) + pq \log|\boldsymbol{R}(\phi)| + \\ &\quad \left(\frac{1-\phi^2}{\sigma^2_{(q)}}\right) trace\,(\boldsymbol{R}^{-1}(\phi)\mathbf{W})\Bigg], \end{aligned}$$

where $\gamma_q = 1$ and $\mathbf{W} = \sum_{i=1}^{p}\sum_{j=1}^{q} \dfrac{(\boldsymbol{Y}_{ij} - \boldsymbol{f})(\boldsymbol{Y}_{ij} - \boldsymbol{f})'}{\gamma_j}$ is a symmetric matrix of order $r \times r$. Let $\mathbf{V}_{ij} = (\boldsymbol{Y}_{ij} - \boldsymbol{f})(\boldsymbol{Y}_{ij} - \boldsymbol{f})'$ and $\mathbf{V}_j = \sum_{i=1}^{p} \mathbf{V}_{ij}$. Define

$$\begin{aligned}\Omega(\phi) &= \frac{\partial\, \boldsymbol{R}(\phi)}{\partial\, \phi}, \\ &= |h-h'|\, \phi^{|h-h'|-1},\end{aligned}$$

and the vectors $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, ..., \gamma_{q-1})'$ and $\boldsymbol{\beta} = (\beta_1, \beta_2)' = (\sigma^2_{(q)}, \phi)'$. Let $\hat{\boldsymbol{\beta}}$ be the consistent estimator of $\boldsymbol{\beta}$. In the special case of an AR(1) process we find that

$$\begin{aligned}\frac{\partial \ell}{\partial \gamma_j} &= -\frac{1}{2}\left[\frac{rp}{\gamma_j} + \left(\frac{1-\phi^2}{\sigma^2_{(q)}}\right)\frac{\partial}{\partial \gamma_j}\, trace\,(\boldsymbol{R}^{-1}(\phi)\mathbf{W})\right] \\ &= -\frac{1}{2}\left[\frac{rp}{\gamma_j} - \left(\frac{1-\phi^2}{\gamma_j^2\,\sigma^2_{(q)}}\right) trace\,(\boldsymbol{R}^{-1}(\phi)\mathbf{V}_j)\right],\end{aligned}$$

$$\frac{\partial \ell}{\partial \sigma^2_{(q)}} = -\frac{1}{2}\left[\frac{n}{\sigma^2_{(q)}} - \left(\frac{1-\phi^2}{\sigma^4_{(q)}}\right) trace\,(\boldsymbol{R}^{-1}(\phi)\mathbf{W})\right],$$

and

$$\begin{aligned}\frac{\partial \ell}{\partial \phi} &= -\frac{n\phi}{(1-\phi^2)} - \frac{pq}{2}\, trace\,(\boldsymbol{R}^{-1}(\phi)\Omega(\phi)) + \\ &\quad \frac{1}{2\sigma^2_{(q)}}[(1-\phi^2)\, trace\,(\boldsymbol{R}^{-1}(\phi)\Omega(\phi)\boldsymbol{R}^{-1}(\phi)\mathbf{W}) + 2\phi\, trace\,(\boldsymbol{R}^{-1}(\phi)\mathbf{W})].\end{aligned}$$

It then follows from (3.3) and (3.4) that

$$\eta_j = \frac{1}{2\sigma^2_{(q)}}[(1-\phi^2)\, trace\,(\boldsymbol{R}^{-1}(\phi)\mathbf{V}_j) - rp\sigma^2_{(q)}], \tag{3.9}$$

$$\xi_1 = \frac{1}{2\sigma^4_{(q)}}[(1-\phi^2)\, trace\,(\boldsymbol{R}^{-1}(\phi)\mathbf{V}) - n\sigma^2_{(q)}], \tag{3.10}$$

and

$$\xi_2 = -\frac{n\phi}{(1-\phi^2)} - \frac{pq}{2} trace\,(\boldsymbol{R}^{-1}(\phi)\Omega(\phi)) + \frac{1}{2\sigma^2_{(q)}}[2\phi\, trace\,(\boldsymbol{R}^{-1}(\phi)\mathbf{V}) + (1-\phi^2)\, trace\,(\boldsymbol{R}^{-1}(\phi)\Omega(\phi)\boldsymbol{R}^{-1}(\phi)\mathbf{V})], \tag{3.11}$$

where $\mathbf{V} = \sum_{j=1}^{q} \mathbf{V}_j$. These expressions are then used to construct the vectors $\boldsymbol{\eta} = (\eta_1, ..., \eta_{q-1})'$ and $\boldsymbol{\xi} = (\xi_1, \xi_2)'$ in the score vector

$$\mathbf{u} = [u_1(\boldsymbol{\beta}), u_2(\boldsymbol{\beta}), ..., u_{q-1}(\boldsymbol{\beta})]' = \boldsymbol{\eta} - \mathbf{D}\mathbf{G}^{-1}\boldsymbol{\xi}.$$

To compute the evaluation of $\mathbf{u}$, we obtain the elements of the matrices $\mathbf{D}$ and $\mathbf{G}$ using (3.6) and (3.7). Now it is quite straightforward to show that

$$\frac{\partial^2 \ell}{\partial\gamma_j\,\partial\sigma^2_{(q)}} = -\left(\frac{(1-\phi^2)}{2\,\gamma_j^2\,\sigma^4_{(q)}}\right) trace\,(\boldsymbol{R}^{-1}(\phi)\mathbf{V}_j).$$

By noting that

$$\begin{aligned} E(\mathbf{V}_j) &= E\left(\sum_{i=1}^{p} \mathbf{V}_{ij}\right), \\ &= \sum_{i=1}^{p} E[(\boldsymbol{Y}_{ij} - \boldsymbol{f})(\boldsymbol{Y}_{ij} - \boldsymbol{f})'], \\ &= \sum_{i=1}^{p} Var(\boldsymbol{\varepsilon}_{ij}), \\ &= \frac{p\,\sigma^2_{(j)}\boldsymbol{R}(\phi)}{1-\phi^2}, \end{aligned}$$

we obtain

$$E\left[\frac{\partial^2 \ell}{\partial \gamma_j \, \partial \sigma^2_{(q)}}\right] = -\left(\frac{p\,\sigma^2_{(j)}}{2\,\gamma_j^2\,\sigma^4_{(q)}}\right) trace\,(I_r),$$
$$= -\frac{pr}{2\,\gamma_j\,\sigma^2_{(q)}}.$$

Therefore,

$$\begin{aligned} d_{j1} &= -E\left[\frac{\partial^2 \ell}{\partial \gamma_j \, \partial \sigma^2_{(q)}}\right]_{\boldsymbol{\gamma}=\mathbf{1}_{q-1}} \\ &= \frac{pr}{2\,\sigma^2_{(q)}}, \qquad j = 1, 2, \ldots, q-1. \end{aligned} \tag{3.12}$$

Similarly, it can be shown that

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \gamma_j \, \partial \phi} &= \left(\frac{1-\phi^2}{2\,\sigma^2_{(q)}}\right) \frac{\partial}{\partial \gamma_j} trace\,(\boldsymbol{R}^{-1}(\phi)\Omega(\phi)\boldsymbol{R}^{-1}(\phi)\mathbf{W}) + \\ &\quad \left(\frac{\phi}{\sigma^2_{(q)}}\right) \frac{\partial}{\partial \gamma_j} trace\,(\boldsymbol{R}^{-1}(\phi)\mathbf{W}), \\ &= -\left(\frac{1-\phi^2}{2\,\sigma^2_{(q)}\gamma_j^2}\right) trace\,(\boldsymbol{R}^{-1}(\phi)\Omega(\phi)\boldsymbol{R}^{-1}(\phi)\mathbf{V}_j) - \\ &\quad \left(\frac{\phi}{\sigma^2_{(q)}\gamma_j^2}\right) trace\,(\boldsymbol{R}^{-1}(\phi)\mathbf{V}_j). \end{aligned}$$

Thus,

$$\begin{aligned} E\left[\frac{\partial^2 \ell}{\partial \gamma_j \, \partial \phi}\right] &= -\left(\frac{1-\phi^2}{2\,\sigma^2_{(q)}\gamma_j^2}\right) trace\,[\boldsymbol{R}^{-1}(\phi)\Omega(\phi)\boldsymbol{R}^{-1}(\phi)E(\mathbf{V}_j)] - \\ &\quad \left(\frac{\phi}{\sigma^2_{(q)}\gamma_j^2}\right) trace\,[\boldsymbol{R}^{-1}(\phi)E(\mathbf{V}_j)], \end{aligned}$$

$$
\begin{aligned}
&= -\left(\frac{p\,\sigma^2_{(j)}}{2\,\sigma^2_{(q)}\gamma_j^2}\right) trace\,[\boldsymbol{R}^{-1}(\phi)\Omega(\phi)] - \left(\frac{\phi\, p\,\sigma^2_{(j)}}{(1-\phi^2)\sigma^2_{(q)}\gamma_j^2}\right) trace(I_r),\\
&= -\left(\frac{p}{2\,\gamma_j}\right) trace\,[\boldsymbol{R}^{-1}(\phi)\Omega(\phi)] - \frac{pr\,\phi}{(1-\phi^2)\,\gamma_j}.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
d_{j2} &= -E\left[\frac{\partial^2 \ell}{\partial\gamma_j\,\partial\phi}\right]_{\boldsymbol{\gamma}=\mathbf{1}_{q-1}}\\
&= \left(\frac{p}{2}\right) trace\,[\boldsymbol{R}^{-1}(\phi)\Omega(\phi)] + \frac{pr\,\phi}{(1-\phi^2)}, \qquad j = 1,2,\ldots,q-1. \quad (3.13)
\end{aligned}
$$

From (3.7) it is clear that

$$
\begin{aligned}
g_{11} &= -E\left[\frac{\partial^2 \ell}{\partial(\sigma^2_{(q)})^2}\right]_{\boldsymbol{\gamma}=\mathbf{1}_{q-1}},\\
g_{12} &= -E\left[\frac{\partial^2 \ell}{\partial\phi\,\partial\sigma^2_{(q)}}\right]_{\boldsymbol{\gamma}=\mathbf{1}_{q-1}}
\end{aligned}
$$

and

$$
g_{22} = -E\left[\frac{\partial^2 \ell}{\partial\phi^2}\right]_{\boldsymbol{\gamma}=\mathbf{1}_{q-1}}.
$$

Using the loglikelihood function it is easily verified that

$$
\frac{\partial^2 \ell}{\partial(\sigma^2_{(q)})^2} = \frac{n}{2\,\sigma^4_{(q)}} - \left(\frac{1-\phi^2}{\sigma^6_{(q)}}\right) trace\,[\boldsymbol{R}^{-1}(\phi)\mathbf{W}],
$$

$$
\begin{aligned}
\frac{\partial^2 \ell}{\partial\phi\,\partial\sigma^2_{(q)}} &= -\frac{1}{2\sigma^4_{(q)}}\left[(1-\phi^2)\,trace\,(\boldsymbol{R}^{-1}(\phi)\Omega(\phi)\boldsymbol{R}^{-1}(\phi)\mathbf{W})+\right.\\
&\quad \left. 2\phi\, trace\,(\boldsymbol{R}^{-1}(\phi)\mathbf{W})\right]
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial^2 \ell}{\partial \phi^2} &= -\frac{n(1+\phi^2)}{(1-\phi^2)^2} + \left(\frac{pq}{2}\right) trace\left[\boldsymbol{R}^{-1}(\phi)\Omega(\phi)\boldsymbol{R}^{-1}(\phi)\Omega(\phi)\right] - \\
&\quad \left(\frac{pq}{2}\right) trace\left[\boldsymbol{R}^{-1}(\phi)\frac{\partial\Omega(\phi)}{\partial\phi}\right] + \\
&\quad \frac{1}{2\sigma^2_{(q)}}\{(1-\phi^2)[-trace\,(\boldsymbol{R}^{-1}(\phi)\Omega(\phi)\boldsymbol{R}^{-1}(\phi)\Omega(\phi)\boldsymbol{R}^{-1}(\phi)\mathbf{W}) + \\
&\quad trace\left(\boldsymbol{R}^{-1}(\phi)\frac{\partial\Omega(\phi)}{\partial\phi}\boldsymbol{R}^{-1}(\phi)\mathbf{W}\right) - \\
&\quad trace\,(\boldsymbol{R}^{-1}(\phi)\Omega(\phi)\boldsymbol{R}^{-1}(\phi)\Omega(\phi)\boldsymbol{R}^{-1}(\phi)\mathbf{W})] + 2\,trace\,(\boldsymbol{R}^{-1}(\phi)\mathbf{W}) - \\
&\quad 2\phi\,trace\,(\boldsymbol{R}^{-1}(\phi)\Omega(\phi)\boldsymbol{R}^{-1}(\phi)\mathbf{W}) - 2\phi\,trace\,(\boldsymbol{R}^{-1}(\phi)\Omega(\phi)\boldsymbol{R}^{-1}(\phi)\mathbf{W})\}.
\end{aligned}
$$

Now,

$$
\begin{aligned}
E[\mathbf{W}] &= \sum_{j=1}^{q}\frac{1}{\gamma_j}E(\mathbf{V}_j), \\
&= \sum_{j=1}^{q}\frac{p\,\sigma^2_{(j)}\boldsymbol{R}(\phi)}{(1-\phi^2)\gamma_j}, \\
&= \frac{pq\,\sigma^2_{(q)}\boldsymbol{R}(\phi)}{(1-\phi^2)}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
g_{11} &= -E\left[\frac{\partial^2\ell}{\partial(\sigma^2_{(q)})^2}\right]_{\boldsymbol{\gamma}=\mathbf{1}_{q-1}} \\
&= \left[-\frac{n}{2\,\sigma^4_{(q)}} + \left(\frac{pq}{\sigma^4_{(q)}}\right) trace\,(I_r)\right]_{\boldsymbol{\gamma}=\mathbf{1}_{q-1}} \\
&= \frac{n}{2\,\sigma^4_{(q)}}.
\end{aligned}
$$

Similarly,

$$\begin{aligned}
g_{12} &= -E\left[\frac{\partial^2 \ell}{\partial\phi\,\partial\sigma^2_{(q)}}\right]_{\boldsymbol{\gamma}=\mathbf{1}_{q-1}} \\
&= 4\left[\frac{1}{8\sigma^4_{(q)}}\{(1-\phi^2)\,trace\,(\boldsymbol{R}^{-1}(\phi)\Omega(\phi)\boldsymbol{R}^{-1}(\phi)E[\mathbf{W}]) + \right. \\
&\quad \left. \frac{1}{2}\phi\,trace\,(\boldsymbol{R}^{-1}(\phi)E[\mathbf{W}])\}\right]_{\boldsymbol{\gamma}=\mathbf{1}_{q-1}} \\
&= \left[\frac{1}{2\sigma^4_{(q)}}\{pq\,\sigma^2_{(q)}\,trace\,(\boldsymbol{R}^{-1}(\phi)\Omega(\phi)) + \frac{2n\phi\,\sigma^2_{(q)}}{1-\phi^2}\}\right]_{\boldsymbol{\gamma}=\mathbf{1}_{q-1}} \\
&= \frac{1}{2\sigma^2_{(q)}}\left[pq\,trace\,(\boldsymbol{R}^{-1}(\phi)\Omega(\phi)) + \frac{2n\phi}{1-\phi^2}\right]. \qquad (3.14)
\end{aligned}$$

$$\begin{aligned}
g_{22} &= -E\left[\frac{\partial^2 \ell}{\partial\phi^2}\right]_{\boldsymbol{\gamma}=\mathbf{1}_{q-1}} \\
&= \left[\frac{n(1+\phi^2)}{(1-\phi^2)^2} - \frac{n}{1-\phi^2} - \left(\frac{pq}{2}\right)\,trace\,[\boldsymbol{R}^{-1}(\phi)\Omega(\phi)\boldsymbol{R}^{-1}(\phi)\Omega(\phi)] + \right. \\
&\quad \left(\frac{pq}{2}\right)\,trace\,\left[\boldsymbol{R}^{-1}(\phi)\frac{\partial\Omega(\phi)}{\partial\phi}\right] + \left(\frac{pq}{2}\right)\,trace\,[\boldsymbol{R}^{-1}(\phi)\Omega(\phi)\boldsymbol{R}^{-1}(\phi)\Omega(\phi)] - \\
&\quad \left(\frac{pq}{2}\right)\,trace\,\left[\boldsymbol{R}^{-1}(\phi)\frac{\partial\Omega(\phi)}{\partial\phi}\right] + \left(\frac{pq}{2}\right)\,trace\,[\boldsymbol{R}^{-1}(\phi)\Omega(\phi)\boldsymbol{R}^{-1}(\phi)\Omega(\phi)] + \\
&\quad \left. \left(\frac{pq\,\phi}{1-\phi^2}\right)\,trace\,[\boldsymbol{R}^{-1}(\phi)\Omega(\phi)] + \left(\frac{pq\,\phi}{1-\phi^2}\right)\,trace\,[\boldsymbol{R}^{-1}(\phi)\Omega(\phi)]\right]_{\boldsymbol{\gamma}=\mathbf{1}_{q-1}} \\
&= \left[\frac{n(1+\phi^2)}{(1-\phi^2)^2} - \frac{n}{1-\phi^2} + \left(\frac{pq}{2}\right)\,trace\,[\boldsymbol{R}^{-1}(\phi)\Omega(\phi)\boldsymbol{R}^{-1}(\phi)\Omega(\phi)] + \right. \\
&\quad \left. \left(\frac{2pq\,\phi}{1-\phi^2}\right)\,trace\,[\boldsymbol{R}^{-1}(\phi)\Omega(\phi)]\right]_{\boldsymbol{\gamma}=\mathbf{1}_{q-1}} \\
&= \frac{2n\phi^2}{(1-\phi^2)^2} + \left(\frac{pq}{2}\right)\,trace\,[\boldsymbol{R}^{-1}(\phi)\Omega(\phi)\boldsymbol{R}^{-1}(\phi)\Omega(\phi)] + \\
&\quad \left(\frac{2pq\,\phi}{1-\phi^2}\right)\,trace\,[\boldsymbol{R}^{-1}(\phi)\Omega(\phi)]. \qquad (3.15)
\end{aligned}$$

Lemma 2. The variance-covariance matrix of $\mathbf{u}$ is given by $\mathbf{C}$ - $\mathbf{D}\mathbf{G}^{-1}\mathbf{D}'$, where $\mathbf{C}_{q-1\times q-1}$ is a diagonal matrix with elements

$$c_{jk} = -E\left[\frac{\partial^2 \ell}{\partial\gamma_j\,\partial\gamma_k}\right]_{\boldsymbol{\gamma}=\mathbf{1}_{q-1}}, \qquad \textit{for } j = k.$$

Proof. From Lemma 1, it is clear that $E(\mathbf{u}) = 0$. From (2.6) we write

$$\begin{aligned} Cov(\mathbf{u}) &= Cov(\boldsymbol{\eta}) - Cov(\boldsymbol{\eta},\boldsymbol{\xi})\,\mathbf{G}^{-1}\mathbf{D}' - \mathbf{D}\mathbf{G}^{-1}\,Cov(\boldsymbol{\xi},\boldsymbol{\eta}) + \\ &\quad \mathbf{D}\mathbf{G}^{-1}\,[Cov(\boldsymbol{\xi})]\mathbf{G}^{-1}\mathbf{D}'. \end{aligned} \tag{3.16}$$

By definition of a score function, $Cov(\boldsymbol{\eta}) = \mathbf{C}$ is a diagonal matrix with its (j,k)th element defined as

$$c_{jk} = -E\left[\frac{\partial^2 \ell}{\partial\gamma_j\,\partial\gamma_k}\right]_{\boldsymbol{\gamma}=\mathbf{1}_{q-1}}.$$

Now, it is easy to verify that $c_{jk} = 0$, $j \neq k$ and

$$\begin{aligned} \frac{\partial^2 \ell}{\partial\gamma_j^2} &= \frac{pr}{2\gamma_j^2} - \left(\frac{1-\phi^2}{\sigma_{(q)}^2\gamma_j^3}\right) trace\,(\boldsymbol{R}^{-1}(\phi)\mathbf{V}_j), \\ E\left[\frac{\partial^2 \ell}{\partial\gamma_j^2}\right] &= \frac{pr}{2\gamma_j^2} - \left(\frac{1-\phi^2}{\sigma_{(q)}^2\gamma_j^3}\right) trace\,[\boldsymbol{R}^{-1}(\phi)E(\mathbf{V}_j)], \\ &= \frac{pr}{2\gamma_j^2} - \left(\frac{p\sigma_{(j)}^2}{\sigma_{(q)}^2\gamma_j^3}\right) trace\,(I_r), \\ &= -\frac{pr}{2\gamma_j^2}. \end{aligned}$$

It follows that

$$c_{jk} = \frac{pr}{2}, \qquad \textit{for } j = k. \tag{3.17}$$

Now $Cov(\boldsymbol{\xi})$ and $Cov(\boldsymbol{\eta}, \boldsymbol{\xi})$ are, by definition of a score function, given by

$$
\begin{aligned}
Cov(\boldsymbol{\xi}) &= \begin{pmatrix} -E\left[\dfrac{\partial^2 \ell}{\partial(\sigma^2_{(q)})^2}\right]_{\boldsymbol{\gamma}=\mathbf{1}_{q-1}} & -E\left[\dfrac{\partial^2 \ell}{\partial\phi\,\partial\sigma^2_{(q)}}\right]_{\boldsymbol{\gamma}=\mathbf{1}_{q-1}} \\ -E\left[\dfrac{\partial^2 \ell}{\partial\phi\,\partial\sigma^2_{(q)}}\right]_{\boldsymbol{\gamma}=\mathbf{1}_{q-1}} & -E\left[\dfrac{\partial^2 \ell}{\partial\phi^2}\right]_{\boldsymbol{\gamma}=\mathbf{1}_{q-1}} \end{pmatrix}, \\
&= \mathbf{G},
\end{aligned}
$$

and

$$
\begin{aligned}
Cov(\boldsymbol{\eta}, \boldsymbol{\xi}) &= \begin{pmatrix} -E\left[\dfrac{\partial^2 \ell}{\partial\gamma_1\,\partial\sigma^2_{(q)}}\right]_{\boldsymbol{\gamma}=\mathbf{1}_{q-1}} & -E\left[\dfrac{\partial^2 \ell}{\partial\gamma_1\,\partial\phi}\right]_{\boldsymbol{\gamma}=\mathbf{1}_{q-1}} \\ -E\left[\dfrac{\partial^2 \ell}{\partial\gamma_2\,\partial\sigma^2_{(q)}}\right]_{\boldsymbol{\gamma}=\mathbf{1}_{q-1}} & -E\left[\dfrac{\partial^2 \ell}{\partial\gamma_2\,\partial\phi}\right]_{\boldsymbol{\gamma}=\mathbf{1}_{q-1}} \\ \vdots & \vdots \\ -E\left[\dfrac{\partial^2 \ell}{\partial\gamma_{q-1}\,\partial\sigma^2_{(q)}}\right]_{\boldsymbol{\gamma}=\mathbf{1}_{q-1}} & -E\left[\dfrac{\partial^2 \ell}{\partial\gamma_{q-1}\,\partial\phi}\right]_{\boldsymbol{\gamma}=\mathbf{1}_{q-1}} \end{pmatrix} \\
&= \mathbf{D},
\end{aligned}
$$

respectively. By applying these results to (3.16), we obtain

$$
Cov(\mathbf{u}) = \mathbf{C} - \mathbf{D}\mathbf{G}^{-1}\mathbf{D}'.
$$

$\square$

After replacing $\boldsymbol{\beta}$ in $\mathbf{u}$, $\mathbf{C}$, $\mathbf{D}$ and $\mathbf{G}$ by $\hat{\boldsymbol{\beta}}$, we obtain Neyman's partial score test statistic as in (3.8), which is asymptotically distributed as χ^2 with q - 1 degrees of freedom.

3.2.1 Estimation of Function

Previously, we used the DWT approach to estimate the trend function which led to the poor performance of the size of the score statistic. In this section, we use a wavelet version of weighted least squares to estimate the trend $f(x)$. First, we express $f(x)$ in terms of its finite order m wavelet expansion, (see Section 1.2.1) for all $x \in [0,1]$,

$$\begin{aligned} f(x) &= c\,\phi(x) + \sum_{j=0}^{m}\sum_{k=0}^{2^j-1} d_{jk}\,\psi_{j,k}(x) + \sum_{j=m+1}^{\infty}\sum_{k=0}^{2^j-1} d_{jk}\,\psi_{j,k}(x), \\ &= c\,\phi(x) + \sum_{j=0}^{m}\sum_{k=0}^{2^j-1} d_{jk}\,\psi_{j,k}(x) + g(x), \end{aligned}$$

where g(x) = $\sum_{j=m+1}^{\infty}\sum_{k=0}^{2^j-1} d_{jk}\,\psi_{j,k}(x)$ is the remainder term in the wavelet expansion.

Define $\boldsymbol{\theta} = \{c, d_{00}, d_{10}, d_{11}, \ldots, d_{m2^m-1}\}'$ and $\mathbf{q}_m(x) = \{\phi(x), \psi_{0,0}(x), \psi_{1,0}(x), \ldots, \psi_{m,2^m-1}(x)\}'$. Then, we can rewrite the model (3.1) as

$$y_{ijh} = \mathbf{q}_m'(x_h)\boldsymbol{\theta} + g(x_h) + \varepsilon_{ijh}, \tag{3.18}$$

where $\boldsymbol{\theta} = \int_0^1 f(x)\mathbf{q}_m(x)dx$. We note that in actual computations only the first term in (3.18) can be estimated.

The vector of filter coefficients $\boldsymbol{\theta}_{ij}$ is then estimated by weighted least squares method as

$$\hat{\boldsymbol{\theta}}_{ij} = (\mathbf{X}'\mathbf{A}\mathbf{X})^{-1}\mathbf{X}'\mathbf{A}\,\boldsymbol{Y}_{ij},$$

where $\mathbf{X}$ is the model matrix with rows $\mathbf{q}_m'(x_h)$, $\mathbf{A}$ is the $r \times r$ diagonal matrix with diagonal elements $a(x_h)$, the weighted least square weights. For the construction of the weighted least squares weights, we follow Oyet and Wiens (2000). Oyet and Wiens

(2000) derived the minimum variance unbiased weights given by

$$a(x_h) = \frac{\int_0^1 ||\mathbf{q}_m(x)||dx}{||\mathbf{q}_m(x_h)||},$$

where $||\mathbf{q}(\mathbf{x})||$ denotes the Euclidean norm. It follows that

$$\hat{\boldsymbol{\theta}} = \frac{\sum_{i=1}^{p}\sum_{j=1}^{q}\hat{\boldsymbol{\theta}}_{ij}}{pq}.$$

Therefore the trend function is estimated by

$$\hat{f}(x) = \mathbf{q}'_m(x)\,\hat{\boldsymbol{\theta}}$$

In our simulation studies, we have used the DAUB8 filter coefficients to construct the function.

3.2.2 Estimation of Nuisance Parameters

To compute Neyman's partial score test statistic $S(\hat{\boldsymbol{\beta}})$, we need to obtain a consistent estimates of $\sigma^2_{(q)}$ and ϕ. We use the residuals and the weights from the weighted least squares method to compute the estimates of $\sigma^2_{(q)}$ and ϕ as

$$\hat{\sigma}^2_{(q)} = \frac{\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{h=1}^{r} a(x_h)\,(y_{ijh} - \hat{f}(x_h))^2}{pq\sum_{h=1}^{r} a(x_h)}, \tag{3.19}$$

and

$$\hat{\phi} = \frac{\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{h=1}^{r-1}(y_{ijh} - \hat{f}(x_h))(y_{ijh+1} - \hat{f}(x_{h+1}))/pq(r-1)}{\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{h=1}^{r}(y_{ijh} - \hat{f}(x_h))^2/pqr}, \tag{3.20}$$

respectively.

3.3 Size and Power Performance of the Score Test: Simulation Study

In this section, we performed a simulation study to examine the size and power performance of Neyman's partial score test statistic $S(\hat{\boldsymbol{\beta}})$. Here we used the same mean response functions used in Chapter 2 to examine the size and power performance of Neyman's partial score test statistic $S(\hat{\boldsymbol{\beta}})$. For each choice of mean response function, we considered time series of length $r = 8$ and $r = 16$. Due to the nature of weighted least squares, when $r = 8$, we set $m = 2$ and when $r = 16$, we obtained the best result with $m = 3$. We compute the proportion of rejections for the test, based on the number of simulations, for a nominal significance level of 5%. For each simulation, the null hypothesis was rejected if the test statistic exceeded the 95th percentile of a χ^2 distribution.

3.3.1 Size of the Score Test

We consider $q = 4$ groups of time series to compute the values of score test statistic $S(\hat{\boldsymbol{\beta}})$ in 5000 simulations, when $\sigma^2_{(q)}$ is estimated by (3.19) and ϕ is estimated by (3.20). The data used in the computations were generated as described earlier in

Section 2.2.1 with the addition of a correlated error term. We compute the size for testing the null hypothesis $H_0 : \gamma_1 = \gamma_2 = \gamma_3 = 1$ at nominal level 5% under the χ^2 distribution with q - 1 = 3 degrees of freedom, when the mean response function is estimated by wavelet version of weighted least squares method. The results are shown in Table 3.1.

Table 3.1: Estimates of size of $H_0 : \gamma_1 = \gamma_2 = \gamma_3 = 1$ at nominal 5% level, with $\sigma^2_{(4)}$ = 0.8, ϕ = 0.9

Mean function	k	r = 8	k	r = 16
Constant	16	0.318	16	0.3352
	64	0.0834	64	0.071
	128	0.0628	112	0.0598
	144	0.0534	120	0.0532
Balanced block	16	0.318	16	0.3352
	64	0.0834	64	0.071
	128	0.0628	80	0.0572
	144	0.0534	120	0.0532
HeaviSine	16	0.308	16	0.3352
	64	0.0834	80	0.0572
	128	0.0628	96	0.0612
	144	0.0534	120	0.0532
Doppler	16	0.318	16	0.3352
	64	0.0834	96	0.0612
	128	0.0628	112	0.0598
	144	0.0534	120	0.0532

It is clear from Table 3.1, for all mean response functions that a large sample size is needed for the score test statistic to control the size of the test due to the correlation between the observations. According to this result, we can conclude that when the sample size is sufficiently large, the score statistic performs well in controlling the size of the test when the function is estimated by wavelet version of weighted least squares method.

3.3.2 Power of the Score Test

Again, the data is generated as described earlier with the addition of correlation error terms. We compute the power for testing the null hypothesis $H_0 : \gamma_1 = \gamma_2 = \gamma_3 = 1$ against the same alternatives hypothesis which was used in Chapter 2 at nominal level 5% under the χ^2 distribution with q - 1 = 3 degrees of freedom in 5000 simulations, when the mean response function is estimated by wavelet version of weighted least squares method. The results are shown in Table 3.2.

It is clear from Table 3.2 that for all mean response functions, the score test statistic has satisfactory power. However, the score statistic needs more time points to achieve the high power under the alternative hypothesis H_2 because the variation between the group variances is small. The alternative hypotheses H_1, H_3 and H_4 produced a power of the test which was above 96% when the number of time series k = 16 with length r = 16. But the power exceeds 87% with the alternative hypotheses H_3 and H_4 when the number of time series k = 16 with length r = 8. However the power is about 80% with alternative hypothesis H_1. From these results, we can conclude that the score test statistic achieve high power when the variation between the group variances is large.

Table 3.2: Estimates of power of $H_0 : \gamma_1 = \gamma_2 = \gamma_3 = 1$ versus specified alternatives at nominal 5% level

Alternative set	Mean function	k	r = 8	k	r = 16
$H_1 : \gamma_1 = 4, \gamma_2 = 3,$	Constant	16	0.8094	16	0.9648
$\gamma_3 = 2$		32	0.9922	24	0.9948
		40	0.999	32	0.9988
	Balanced block	16	0.8094	16	0.9648
		32	0.9922	24	0.9948
		40	0.999	32	0.9988
	HeaviSine	16	0.8206	16	0.9648
		32	0.9922	24	0.9948
		40	0.9978	32	0.9988
	Doppler	16	0.8094	16	0.9648
		32	0.9922	24	0.9948
		40	0.9978	32	0.9988
$H_2 : \gamma_1 = 1, \gamma_2 = 1,$	Constant	16	0.587	16	0.8
$\gamma_3 = 2$		64	0.9894	32	0.9806
		80	0.9972	40	0.9942
		88	0.9992	48	0.9994
	Balanced block	16	0.5874	16	0.8
		64	0.9854	32	0.9806
		80	0.9968	40	0.9942
		88	0.999	48	0.9994
	HeaviSine	16	0.5874	16	0.8
		64	0.9854	32	0.9806
		80	0.9972	40	0.9942
		88	0.9992	48	0.9994
	Doppler	16	0.5874	16	0.8
		64	0.9854	32	0.9806
		80	0.9968	40	0.9942
		88	0.9992	48	0.9994

(Table 3.2 Contd....)

Alternative set	Mean function	k	r = 8	k	r = 16
$H_3 : \gamma_1 = 1, \gamma_2 = 1,$ $\gamma_3 = 4$	Constant	16	0.9222	16	0.9784
		40	0.998	24	0.994
		48	0.9992	32	0.999
	Balanced block	16	0.9222	16	0.9784
		40	0.998	24	0.994
		48	0.9992	32	0.999
	HeaviSine	16	0.9222	16	0.9784
		40	0.998	24	0.994
		48	0.9992	32	0.999
	Doppler	16	0.9222	16	0.9784
		40	0.998	24	0.994
		48	0.9992	32	0.999
$H_4 : \gamma_1 = 1, \gamma_2 = 3,$ $\gamma_3 = 3$	Constant	16	0.8724	16	0.9802
		48	0.9994	24	0.9956
		56	1	32	0.9992
	Balanced block	16	0.8724	16	0.9802
		48	0.9994	24	0.9956
		56	0.9998	32	0.9992
	HeaviSine	16	0.8724	16	0.9802
		48	0.9994	24	0.995
		56	1	32	0.9992
	Doppler	16	0.8724	16	0.9802
		48	0.9994	24	0.9956
		56	1	32	0.9992

Chapter 4

Concluding Remarks

In this thesis, we have constructed partial score test statistics for testing homogeneity of variances when the data is from a nonparametric regression model with uncorrelated and correlated errors.

In Chapter 2, we have described the construction of Neyman's partial score test for testing the homogeneity of variances in nonparametric model with uncorrelated errors. Neyman's partial score test statistic only requires the consistent estimate of nuisance parameters. We follow Cai, Hurvich and Tsai (1998) to estimate the mean response function by DWT and use the wavelet coefficients at the finest scale to estimate the variance parameter. In the simulation study, we found that the score test statistic performed well in controlling the size of the test when the function is known but performed poorly in controlling the size of the test when the function is unknown and estimated by DWT. This finding appeared to suggest that the score test statistic performed poorly due to the estimation by DWT. That is, the discrete wavelet transform approach may not lead to a consistent estimate of the mean response function which will in turn affect the performance of the score statistic. Cai, Hurvich and Tsai (1998) in their study found that the score test performed poorly in controlling the size of the test under the assumption of normality even when the sample size is large. The results in this thesis show that the poor performance of the score statistic in their study can be attributed to two factors. These are

(i) the formulation of the hypothesis in their paper which leads to identifiability problems.

(ii) the fact that they used the DWT approach in estimating the response function.

We note that if the hypothesis is properly formulated and a wavelet version of the weighted least squares approach is used in estimating the response, the score statistic performs well in controlling the size of the test. See Oyet and Sutradhar (2003) for details. We also found that if the sample size is small, the power of the score test is high when the variation between the group variances is large.

In Chapter 3, we have also described the construction of Neyman's partial score test for testing the homogeneity of variances when the data arises from a nonparametric model with correlated errors. We followed Oyet and Sutradhar (2003) to estimate the mean response function by a wavelet version of weighted least squares. Here, we use the residuals from the weighted least squares estimation to estimate the correlation and variance parameters. In the simulation study, we found that the score statistic performed well in controlling the size of the test. We also found that the power performance of the score statistic is best when the variation between the group variances is large.

Bibliography

[1] Cai, Z., Hurvich, C.M. and Tsai, C., (1998), Score tests for heteroscedasticity in wavelet regression, Biometrika, 85, 229-234.

[2] Cox, D. R. and Hinkley, D. V., (1974), Theoretical Statistics, Chapman and Hall Ltd, London.

[3] Daubechies, I., (1992), Ten Lectures on Wavelets, Number 61 in CBMS-NSF Series in Applied mathematics, Society for industrial and Applied Mathematics, Philadelphia.

[4] Daubechies, I. and Lagarias, J.C., (1992), Two-scale difference equations II. Local regularity, infinite products of matrices and fractals, SIAM Journal on Mathematical Analysis, 23, No.4, 1031-1079.

[5] Donoho, D. L., and Johnstone, I. M., (1994), Ideal spatial adaptation by wavelet shrinkage, Biometrika, 81(3), 425-455.

[6] Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D., (1995), Wavelet shrinkage: Asymptopia?(with discussion), Journal of Royal statistical society, Series B, 57(2), 301-369.

[7] Erlebacher, G., Hussaini, M. Y. and Jameson, L. M., (1996), Wavelets, Theory and Applications, Oxford University Press, Oxford, New York.

[8] Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A., (1998), Wavelets, Approximation, and Statistical Applications, Lecture Notes in Statistics 129, Springer-Verlag, New York.

[9] Hogg, R. V., McKean, J. W. and Craig, A. T., (2005), Introduction to Mathematical Statistics, Upper Saddle River, New Jersey.

[10] Kovac, A. and Silverman, B.W., (2000), Extending the scope of wavelet regression methods by coefficient dependent thresholding, J. Amer. Statis. Assoc. 95, 172-182.

[11] Mallat, S. G., (1989), Multiresolution approximations and wavelet orthonormal bases of $L^2(\mathbb{R})$, Transactions of the American Mathematical Society, 315, 69-87.

[12] Meyer, Y., (1992), Wavelets and Operators, Cambridge University Press, Cambridge.

[13] Moran, P. A. P., (1970), On asymptotically optimal tests of composite hypothesis, Biometrika, 57, 47-55.

[14] Neyman, J., (1959), Optimal Asymptotic Tests of Composite Hypothesis, Probability and Statistics, The Harald Cramer Volume, Ulf Grenander, Wiley, New York, pp. 213-234.

[15] Oyet, A.J. and Sutradhar, B., (2003), Testing variances in wavelet regression models, Statistics and Probability Letters, 61, 97-109.

[16] Oyet, A.J., and Wiens, D.P., (2000), Robust designs for wavelet approximations of regression models, J. Nonparametric Statist., 12: 837-859.

[17] Piheiro, A. and Vidakovic, B., (1997), Estimating the square root of a density via compactly supported wavelets, Computational Statistics and Data Analysis, 25, no. 4, 399-415.

[18] Strang, G., (1989), Wavelets and dilation equations: A brief introduction, SIAM Review, 31, no. 4, 614-627.

[19] Sutradhar, B.C., (1996), Score test versus Bartlett type modified test for testing homogeneity of variances with autocorrelated errors, Sankhya B, 58, 10-27.

[20] Sutradhar, B. C. and Bartlett, M. S. (1993), A small and large sample comparison of Wald's, likelihood ratio and Rao's test for testing linear regressin with autocorrelated errors, Sankhya B, 55, 186-196.

[21] Vidakovic, B., (1999), Statistical Modeling by Wavelets, Wiley, New York.