New Methods for Spectral Estimation with Confidence Intervals for Locally Stationary Wavelet Processes

Kara Nicola Stevens

School of Mathematics

June 2013

A Dissertation Submitted to the University of Bristol in accordance with the requirements of the degree Doctor of Philosophy in the Faculty of Science

Abstract

In this thesis we develop new methods of estimating the second order structure of a locally stationary process with associated confidence intervals (CI) to quantify the estimator. In particular we focus on smoothing the estimate of the evolutionary wavelet spectrum (EWS) of a locally stationary wavelet process. Similarly to the stationary spectrum, the raw estimate of the EWS is unbiased but not consistent. Therefore, de-noising methods are required to improve our estimate. We have developed three such methods and used simulated data to test the methods.

We begin by investigating the statistical properties of the raw estimate, which we use to establish theoretical models. Our first method uses a simple running mean estimator. This method produces reasonable results, but due to its strong dependence the translation invariant de-noising estimator, there are limitations in the type of spectra's which can be successfully recovered.

Our next two methods uses the benefits of a posterior distribution from Bayesian analysis to produce estimators with associated CI. We use the log and Haar-Fisz transformations to stabilise the variance of our raw estimates. The model for the log transformed data is more complex model than the Haar-Fisz, and required more numerical approximations during the estimation process. The Haar-Fisz transformation pulls the distribution of the data closer to Gaussianity, which greatly simplified the modelling assumptions. The Haar-Fisz estimator and CI were computationally more efficient and produced superior results compared to the previous methods. It appeared to be fairly robust to wavelet selection, but again, was heavily dependent on the hyperparameters.

Using different methods to determine the hyperparameters greatly influenced the results. With further research we hope to improve the efficiency of determining the hyperparameters.

Dedication

I would like to dedicate this thesis to my Mum, Alison Stevens, gone but not forgotten.

Acknowledgements

Firstly I would like to thank my supervisor, Professor Guy Nason, for all the inspiration, guidance and encouragement given to me throughout my degree. Without his support, this PhD would not have been possible.

I would like to extend my gratitude to the EPSRC and SuSTaIn initiative for funding me throughout my time at the University of Bristol.

I would to thank all my friends and family for their continuous support, advice, and encouragement. In particular Linda Lietchmann for taking the time to proof read my thesis; Lesley Barwell, Becky Hunt and Lisa Giles for always believing in me and listening to my complaints. I would especially like to thank Daniel Barwell, for his enduring patient and love through some very turbulent times.

Finally, I would like to thank my Mum, for all her sacrifices in life so I could even have the chance of attending University in the first place. I am what I am today because of her.

Without all these people, I would not have been able to accomplish what I have.

Author's Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: .....

Kara Nicola Stevens

DATE: .....

Contents

List of Figures

Chapter 1

Introduction

A time series is a collection of observations made sequentially through time. Time series analysis is used to understand the underlying features of the data and to make forecasts (predictions). Data of this form occur in many different fields, such as signal analysis, medicine and economics. There is no single methodology to analyse time series data that is suitable across all fields. Usually, the field from which the data were obtained and aim of the analysis will dictate the appropriate methodology.

If the series satisfies certain properties, such as observed at regular intervals and stationarity (where the mean and variance are constant, and the covariance only depends on time difference), there are an abundance of analysis methods and supporting probability theory. For example, see methods described in Priestley (1994) and Chatfield (2004). For real-life series it is often difficult to determine whether the process is really stationary with only observations from a section of the process. Often the behaviour of a real-life process can change from one stationary regime to another and this may only become apparent with continued observation. If the assumption of stationarity is disregarded, there is a plethora of different models which can be considered. One class of non-stationary models are the locally stationary processes with slowly evolving second-order structures, and, in particular, studied here, the locally stationary wavelet (LSW) process of Nason et al. (2000).

An important feature of locally stationary time series is their second-order structure. The second-order structure can be assessed via the covariance or spectrum of the time series, and accurate estimation of these quantities is particularly important. If we assume the series is generated from an LSW process, to obtain an unbiased and consistent estimate of the local spectrum, we use smoothing techniques upon the raw wavelet periodogram.

There are many methods of estimating the local spectrum of an LSW process via smoothing (Nason et al., 2000; Fryzlewicz and Nason, 2006; Van Bellegem and von Sachs, 2008), however few of these methods possess a way of quantifying the result.

In this thesis we propose three new methods of estimating the time varying spectra of an LSW process, and develop associated confidence intervals. Chapter 2 reviews some basic wavelet theory and time series analysis. In particular, we review the application of wavelets in locally stationary time series analysis and nonparametric regression.

In chapter 3 we investigate some of the statistical properties of the wavelet coefficients and raw wavelet periodogram, which we use to estimate the time varying spectra from data. This provides us with the assumptions we need to develop our smoothing techniques later. We also numerically verify our findings using two different defined local spectra, for which we simulate a variety of LSW processes in the statistical package `R` (`R` Development Core Team, 2012), with innovations generated from three types of distributions: Gaussian, chi-square and Student's t distribution.

From the information in chapter 3, we develop our first method in chapter 4, which uses the simple smoothing technique of the central moving average. This chapter makes particular use of the linear form of the estimator and covariance structure in the data to produce confidence intervals. The estimate obtained from using the translation-invariant de-noising technique of (Nason et al., 2000) is used in the selection of bin widths. Using the simulated LSW processes from chapter 3, we test our method and compare our findings with the true local spectra.

Chapters 5 and 6 are all associated with using the nonparametric regression technique of wavelet shrinkage, to smooth the raw wavelet periodogram. We begin in chapter 5 by defining an additive model by using the log transformation. Using a Bayesian shrinkage rule, we determine the posterior distribution, mean and variance which were all used to create the estimate and corresponding confidence intervals. We conclude chapter 5 by determining the asymptotic convergence of our method under assumptions of independence and iid distributions. In chapter 6 we describe the computational steps needed to perform Bayesian wavelet shrinkage of the log transformed periodogram and numerically test the method by applying it to the two simulated LSW processes from chapter 3 in `R`.

Chapter 7 begins by describing the Haar-Fisz transformation of the raw wavelet periodogram and empirically investigating the effects on the covariance and distribution. Then, with our knowledge from chapter 5, we develop a Bayesian wavelet shrinkage method

using the Haar-Fisz transformed raw wavelet periodogram. We develop the method for use in R and test the technique on the simulated LSW processes.

Finally in chapter 8, we summarise our findings and conclusions from each chapter. We also suggest some possible solutions to problems we have encountered and future work on this area.

Chapter 2

Literature Review

The literature review will aim to introduce the area of research which has been the focus of this thesis. The review consists of six sections: Besov spaces, Fourier Analysis, Wavelet Theory, Nonparametric Regression, Time Series Data and Locally Stationary Time Series Analysis.

As wavelets are functions which are commonly used to represent functions, we will begin by introducing a space of functions known as Besov spaces. Although not all Besov spaces are function spaces. In section 2.2 we describe Fourier analysis which provides the natural building blocks of wavelets. The flexibility of wavelets has led to applications in signal processing, image and texture analysis, geophysics and time series analysis to name but a few. A good mathematical description of wavelet analysis can be obtained through multi-resolution analysis (MRA). Within section 2.3, we describe two key forms of wavelet transforms for this thesis, the Discrete Wavelet Transform (DWT) and the Non-decimated Wavelet Transform (NDWT). We shall conclude this section with a description of three types of wavelets: The Haar, Shannon and Daubechies wavelets.

Section 2.4 discusses nonparametric regression. Nonparametric regression is a form of regression analysis in which the form of the model is not predefined, but determined from information from the data. Hence, larger sample sizes than regression based on parametric models are required because the data must supply the model structure as well as the model estimates. Most importantly we shall describe a particular method of non-parametric regression known as wavelet shrinkage and variations of this method.

Section 2.5 describes the type of data we are applying our methodology too, time series data. We shall begin with stationary time series and associated models for this data. Second-order stationary models assume that the mean and variance are constant,

and the covariance structure only depends on the time difference. However, because real life scenarios rarely conform to the stationary ideal we shall relax our assumptions by looking at non-stationary time series. In particular, locally stationary processes with slowly evolving second order structure, i.e. a time series with a time dependent variance structure.

The main focus of this thesis is to develop techniques which can be used to analyse the second order structure of locally stationary wavelet processes. We describe in detail the form of this model and important components associated with the model, such as the evolutionary wavelet spectrum which is used to convey information on the second order structure.

2.1 Besov Spaces

We shall first introduce some notation and briefly review some important mathematical concepts required for the definition and derivation of the methods used in this thesis.

Let $\mathbb{Z}$, $\mathbb{N}$, $\mathbb{R}$ and $\mathbb{C}$ denote the sets of integer, natural, real and complex numbers. For a complex number $z \in \mathbb{C}$ let $|z|$ denote its modulus and $\overline{z}$ its complex conjugate. Also let $\delta_{a,b}$ be the Kronecker delta function, not to be confused with the Dirac delta function denoted by $\delta(x)$.

Let $\mathbf{x} = \{x_i\}$ and $\mathbf{y} = \{y_i\}$ be sequences in the l^p space for $1 \leq p < \infty$ if

1. $\sum_{i\in\mathbb{Z}} |x_i|^p < \infty$
2. $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i\in\mathbb{Z}} x_i \overline{y_i}$
3. $\|x\|_p = \left(\sum_{i\in\mathbb{Z}} |x_i|^p\right)^{1/p}$

A function f belongs to a *Lebesgue space*, denoted by $L^p(\mathbb{A})$ if

1. $\|f\|_p = \left(\int_{\mathbb{A}} |f(x)|^p \mathrm{d}x\right)^{1/p} < \infty$, for $1 \leq p < \infty$,
2. $\|f\|_\infty = \sup_{x\in\mathbb{A}} |f(x)| < \infty$, for $p = \infty$.

We shall briefly introduce some relevant aspects of the Besov space as in Abramovich et al. (1998). This will be useful for the discussion of wavelet shrinkage in section 2.4.2. For more detail we refer the reader to DeVore and Popov (1988); DeVore et al. (1992); Meyer (1992); Triebel (2000, 2006) or the seminal paper Besov (1959).

Let the s^{th} difference of a function f be

$$\Lambda_t^{(s)} f(x) = \sum_{i=0}^{s} \binom{s}{i} (-1)^i f(x + it),$$

and let the s^{th} modulus of smoothness of f in $L^p[0,1]$ be

$$V_{s,p}(f;x) = \sup_{t \le x} \left(\|\Lambda_t^s f\|_{L^p[0,1]} \right).$$

Then the Besov semi-norm, with index (p,q,r), is defined for $s > r$, where $1 \le p,\, q \le \infty$, by

$$|f|_{B^r_{p,q}} = \left[\int_0^1 \left\{ \frac{V_{r,p}(f;x)}{x^r} \right\}^q \frac{\mathrm{d}x}{x} \right]^{1/q}, \qquad \text{if } 1 \le q < \infty,$$

and

$$|f|_{B^r_{p,q}} = \sup_{0<x<1} \left(x^{-r}\, V_{r,p}(f;x) \right), \qquad \text{if } q = \infty,$$

where sup is the supremum. We define the Besov norm as

$$\|f\|_{B^r_{p,q}} = \|f\|_{L^p[0,1]} + |f|_{B^r_{p,q}}.$$

The Besov space, denoted by $B^r_{p,q}$, is a class of functions $f : [0,1] \to \mathbb{R}$, which satisfies $f \in L^p[0,1]$ and $|f|_{B^r_{p,q}} < \infty$. The parameter r measures the number of derivatives and the existence of the derivatives is required in the L^p-sense. The parameter q provides a further finer graduation. A Besov ball is a class of functions $B^r_{p,q}(A)$, which are bounded from above by A.

The Besov spaces are very general cases, and comprises of many other well known spaces as special cases. For example:

- the Sobolev space W_2^r is $B^r_{2,2}$, but $W_p^r \neq B^r_{p,p}$;
- the Holder space C^r is the Besov space $B^r_{\infty,\infty}$.

2.2 Fourier Analysis

This section is a brief overview of Fourier analysis and includes some basic results. In many aspects, wavelet and Fourier analysis are very similar. Therefore, it is useful to introduce some of the relevant concepts from Fourier theory to motivate wavelet analysis, which provides a useful benchmark comparison.

Fourier methods are a powerful tool for signal analysis. By transforming a signal from the time or spatial domain into the spectral or frequency domain, many characteristics of the signal are revealed. The Fourier transform uses sine and cosine waves to form basis functions. A good introduction to general Fourier analysis can be found in Stein and Shakarchi (2003) or Folland (2003).

Next, let us recall the definition of the Fourier series. Fourier series are very important for developing wavelet theory. For example, calculating filters and transfer functions are tasks which require Fourier series.

Definition 2.2.1. Let $f : [0, 2\pi) \to \mathbb{R}$ be a periodic function, with a period 2π, so that $f(x) = f(x - 2\pi)$. Then the Fourier representation of f is

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \{a_n \cos(nx) + b_n \sin(nx)\},$$

where the Fourier coefficients are calculated from

$$a_n = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos(nx)\, \mathrm{d}x \qquad \text{and} \qquad b_n = \frac{1}{\pi} \int_0^{2\pi} f(x) \sin(nx)\, \mathrm{d}x.$$

This representation is possible as $\{1, \cos(nx), \sin(nx)\}$ constitutes an orthonormal basis of $L^2[0, 2\pi)$. The magnitude of the Fourier coefficients conveys the frequency content of the function f. If a function is periodic, with finite discontinuities and extremes, and is absolutely integrable over the interval, then the Fourier series will converge point-wise to f, except at the discontinuities.

Definition 2.2.2. Let the inner product, $\langle \cdot, \cdot \rangle$, of two functions f and g be denoted by

$$\langle f(x), g(x) \rangle = \int f(x) \overline{g(x)}\, \mathrm{d}x.$$

Then the *Fourier transform* of the function $f \in L^1(\mathbb{R})$ is defined as

$$\hat{f}(\omega) = \mathcal{F}[f(x)] = \left\langle f(x), e^{i\omega x} \right\rangle = \int_{\mathbb{R}} f(x) \overline{e^{\mathrm{i}\omega x}}\, \mathrm{d}x = \int_{\mathbb{R}} f(x) e^{-\mathrm{i}\omega x}\, \mathrm{d}x. \tag{2.1}$$

If the function $\hat{f} \in L^1(\mathbb{R})$, then the *inverse Fourier transform* is defined as

$$f(x) = \mathcal{F}^{-1}[f(x)] = \int_{\mathbb{R}} \hat{f}(\omega) e^{\mathrm{i}\omega x}\, \mathrm{d}\omega. \tag{2.2}$$

The function $\hat{f}(\omega)$ is generally complex valued, and can be expressed as

$$\hat{f}(\omega) = |\hat{f}(\omega)|\, e^{\mathrm{i}\phi(\omega)},$$

where $|\hat{f}(\omega)|$ is known as the *magnitude* and $\phi(\omega)$ is called the *phase*. If $f(x)$ is real then

- $\hat{f}(-\omega) = \overline{\hat{f}(\omega)}$
- $|\hat{f}(\omega)|$ is an even function and $\phi(\omega)$ is an odd function of ω.

2.2.1 Properties

Vidakovic (1999, section 2.3, pages 30-31), provides a list of the important properties of the Fourier transform. Some of the properties which are directly relevant to wavelet analysis (Goswami and Chan, 2011) are briefly outlined next.

Linearity

If the function $f(x) = \alpha f_1(x) + \beta f_2(x)$, for some constants $\alpha, \beta \in \mathbb{C}$, then the Fourier transform of f is

$$\hat{f}(\omega) = \int_{\mathbb{R}} f(x)e^{-\mathrm{i}\omega x}\,\mathrm{d}x \;=\; \alpha \int_{\mathbb{R}} f_1(x)e^{-\mathrm{i}\omega x}\,\mathrm{d}x + \beta \int_{\mathbb{R}} f_2(x)e^{-\mathrm{i}\omega x}\,\mathrm{d}x \;=\; \alpha\hat{f}_1(x) + \beta\hat{f}_2(x). \tag{2.3}$$

This can be easily shown for the Fourier series.

Time Shifting and Scaling

If the function $f(x)$ is shifted by x_0, then there is a phase shift in the Fourier transform of the function. Let $f_0(x) = f(x - x_0)$, and $\hat{f}_0(\omega) \;=\; \int_{\mathbb{R}} f(x - x_0)e^{-\mathrm{i}\omega x}\,\mathrm{d}x$. Then substitute $u = x - x_0$ and

$$\hat{f}_0(\omega) \;=\; \int_{\mathbb{R}} f(u)e^{-\mathrm{i}\omega(u+x_0)}\,\mathrm{d}u = e^{-\mathrm{i}\omega x_0}\hat{f}(\omega) = \left|\hat{f}(\omega)\right| e^{\mathrm{i}\phi(\omega)-\mathrm{i}\omega x_0}.$$

The shift does not affect the magnitude spectrum as it is incorporated into the phase term. If $a \neq 0$ is a constant, then the Fourier transform of $f_a(x) = f(ax)$ is $\hat{f}_a(\omega) = \int_{\mathbb{R}} f(ax)e^{-\mathrm{i}\omega x}\,\mathrm{d}x$. Let $u = at$, then

$$\hat{f}_a(\omega) = \int_{\mathbb{R}} f(u)e^{-\mathrm{i}\omega u/a}\,\mathrm{d}\left(\frac{u}{a}\right) = \frac{1}{|a|}\hat{f}\left(\frac{\omega}{a}\right).$$

Frequency Shifting and Scaling

If you shift the Fourier transform frequency the result has a similar effect as a time shift. Let $\hat{f}_0(\omega) = \hat{f}(\omega - \omega_0)$, then

$$f_0(x) = f(x)\, e^{\mathrm{i}\omega_0 x}.$$

Similarly, a scale change where $\hat{f}_a(\omega) = \hat{f}(a\omega)$ implies

$$f_a(x) = \frac{1}{|a|} f\Big(\frac{x}{a}\Big).$$

Parseval's Theorem

Parseval's Theorem states that

$$\int_{\mathbb{R}} |f(x)|^2 \,\mathrm{d}x = \frac{1}{2\pi} \int_{\mathbb{R}} |\hat{f}(\omega)|^2 \,\mathrm{d}\omega. \tag{2.4}$$

Suppose we have two functions, $f, g \in L^2[0, 2\pi)$, and their Fourier transforms $\hat{f}$ and $\hat{g}$, then via Parseval's identity we have

$$\langle f(x), g(x) \rangle = \frac{1}{2\pi} \Big\langle \hat{f}(\omega), \hat{g}(\omega) \Big\rangle.$$

A proof of this can be found on Goswami and Chan (2011, page 40). If $f(x) = g(x)$, then we have Parsavel's theorem. Equation (2.4) states that the total energy computed in the time domain is equal to that in the frequency domain. So Parseval's theorem allows the energy of the signal to be computed in either domain, depending on the ease of computation.

2.2.2 Limitations

Sinusoids are only localised in frequency and not time. Therefore cancellations are required to suitably represent discontinuities in time. At the points of discontinuity the Fourier series will converge to the left or right of this point. The discontinuities will lead to *Gibbs phenomena*, where the n^{th} Fourier series partial sum has large oscillations near the jump, and may increase the maximum of this sum above the function f. As the frequency increases this effect does not diminish and can lead to what is known as the *ringing effect*. This requires a large number of basis function to produce an adequate representation.

Another problem with this methodology is the span of the bases. As the bases span

the whole real axis, there is no direct method to extract localised information. If there are very few non-zero Fourier coefficients the Fourier series yields great insight into the behaviour of the function f.

2.3 Wavelet Theory

A good description of wavelets can be found in Burrus et al. (1998). Usually, a *wave* is thought of as an oscillating function of time or space. A *wavelet* is a small wave that still has the same oscillating properties of a wave but its energy is concentrated in time. This allows for simultaneous time and frequency analysis. Wavelet analysis is a desirable method because

- *Sparsity* - wavelets can produce coefficients where many are zero and only a few are non-zeros.
- *Fast and memory efficient* - for N observations a Fast Fourier Transform computes at $\mathcal{O}(N \log(N))$ whereas wavelets achieve $\mathcal{O}(N)$.
- *Localisation* - wavelets are able to examine functions locally.

Unlike Fourier series, locality can be achieved in the frequency and time domains simultaneously. This is especially useful in the representation of *non-stationary* functions.

We shall begin our introduction of wavelets with multi-resolution analysis. This provides the mathematical foundation of the construction of wavelets and wavelet bases, and naturally leads to the description of the scaling and wavelet function.

2.3.1 Multi-resolution Analysis

Multi-resolution analysis (MRA) is attributed to Mallat (1989) and Meyer (1992). The idea of MRA is similar to sub-band decomposition and coding, which divides a signal into a set of frequency bands in a particular way for efficiency. MRA provides the means of looking at the fine detail of a signal, or one can obtain an overall sense of the behaviour. From MRA one can then develop the filters associated with wavelets, that leads naturally to the *discrete wavelet transform*, so we may construct an orthonormal basis for $L^2(\mathbb{R})$.

Definition 2.3.1. A multi-resolution analysis (Daubechies, 1992) on $L^2(\mathbb{R})$ is a sequence of closed subspaces $\{V_j\}_{j\in\mathbb{Z}}$ of $L^2(\mathbb{R})$, such that

(a) $V_j \subset V_{j+1},\ \forall\ j \in \mathbb{Z}$, i.e. they lie in a containment hierarchy.

(b) $\bigcap_{j\in\mathbb{Z}} V_j = \{0\}$, i.e. the intersection is trivial.

(c) $\overline{\bigcup_{j\in\mathbb{Z}}} = L^2(\mathbb{R})$, i.e. the union is dense in $L^2(\mathbb{R})$.

(d) $f(x) \in V_0 \iff f(2^j x) \in V_j,\ \forall\ j \in \mathbb{Z}$.

(e) $f \in V_0 \iff f(x-k) \in V_0,\ \forall\ k \in \mathbb{Z}$.

(f) There exists a scaling function $\phi \in V_0$ with $\int_{-\infty}^{\infty} \phi(x)\mathrm{d}x = 1$, such that $\{\phi_{0,k} := \phi(t-x)\}_{k\in\mathbb{Z}}$ is an orthonormal basis in V_0.

We will refer to the index j as the scale.

Properties (a) and (b) imply that any function $f \in L^2(\mathbb{R})$ can be gradually approximated by its projections on $\{V_j\}_{j\in\mathbb{Z}}$ spaces. The property in (d) means that any V_j is a scaled version of V_0. Together, properties (d) and (e), ensure that

$$\phi_{j,k}(x) := 2^{j/2}\phi(2^j x - k) \qquad \forall\ k \in \mathbb{Z} \tag{2.5}$$

is an orthonormal basis of V_j, $\forall\, j \in \mathbb{Z}$. The function $\phi_{j,k}$ are known as *translations* and *dilations* of the function ϕ.

The last property means any function $f \in V_0$ can be written as a linear combination,

$$f = \sum_k c_k\, \phi_{0,k},$$

where c_k is the inner product:

$$c_k = \langle f,\, \phi_{0,k}\rangle_{L^2(\mathbb{R})} = \int_{-\infty}^{\infty} f(x)\, \phi(x-k)\mathrm{d}x.$$

Let P_{V_j} denote the projection onto V_j, Then

$$(P_j f)(x) = \sum_k c_{j,k}\, \phi_{j,k}(x),$$

where $c_{j,k}$ are the *scaling coefficients* calculated from

$$\langle f,\, \phi_{j,k}\rangle_{L^2(\mathbb{R})} = \int_{-\infty}^{\infty} f(x)\, \phi_{j,k}(x)\mathrm{d}x,$$

and every function in $L^2(\mathbb{R})$ can be approximated by the elements of the subspaces V_j. From properties (b) and (c), the precision in this approximation increases as $j \to \infty$.

Now consider the scaling function ϕ. As $\phi \in V_0 \subset V_1$ and since $\{\phi_{1,k}\}_{k\in\mathbb{Z}}$ is an orthonormal basis for V_1, we can express the scaling function as

$$\phi(x) = \sum_{k\in\mathbb{Z}} h_k \, \phi_{1,k}(x),$$

where $h_k = \langle \phi, \phi_{1,k} \rangle$ for $k \in \mathbb{Z}$. Therefore, we find the *scaling equation*, which is defined as

$$\phi(x) = \sqrt{2} \sum_{k\in\mathbb{Z}} h_k \, \phi(2x - k), \tag{2.6}$$

and obtain

$$\begin{aligned}
\langle \phi_{j-1,k}, \phi_{j,n} \rangle &= \int \phi_{j-1,k}(x) \; \phi_{j,n}(x) \; \mathrm{d}x \\
&= \int 2^{(j-1)/2} \phi(2^{j-1}x - k) \; 2^{j/2} \phi(2^j x - n) \; \mathrm{d}x \\
&= \sqrt{2} \int \phi(t) \; \phi(2t + 2k - n) \; \mathrm{d}t \qquad\qquad \text{as } t = 2^{j-1}x - k \text{ and } \mathrm{d}t = 2^{j-1} \, \mathrm{d}x \\
&= h_{n-2k}.
\end{aligned}$$

The unique coefficients $\{h_k\}_{k\in\mathbb{Z}} \in l^2(\mathbb{Z})$ form a vector known as the *low pass filter* associated with ϕ. Using $\{\phi_{j,n}\}_{n\in\mathbb{Z}}$ as an orthonormal basis for V_j, and let $\phi_{j-1,k}$ be an element of V_j, we have

$$\begin{aligned}
\phi_{j-1,k}(x) &= \sum_{n\in\mathbb{Z}} \langle \phi_{j-1,k}, \phi_{j,n} \rangle \phi_{j,n}(x) \\
&= \sum_{n\in\mathbb{Z}} h_{n-2k} \, \phi_{j,n}(x).
\end{aligned} \tag{2.7}$$

Equation (2.7) is known as the *scaling function refinement relation.*

Construction of the Wavelet Function

Constructing the wavelet function is centred on the idea of representing the information lost when moving from a finer to a coarser space (i.e. V_j to V_{j-1}). The *detail subspace*, denoted by W_j, is the orthogonal component of V_{j-1} in V_j, where the direct sum of V_j and W_j equals V_{j-1} because V_j are closed subspaces. Any function in V_{j-1} can be uniquely represented as a sum of a function in V_j and a function in W_j.

The construction of W_j and property (a) in definition 2.3.1 of the MRA implies

$W_j \perp W_{j'}, \forall\ j \neq j'$. This is true for each j, therefore

$$V_J = V_{J_0} \oplus \left(\bigoplus_{j=J_0}^{J-1} W_j \right), \qquad \forall\ J > J_0\ \in \mathbb{Z}.$$

In other words, any function in V_J can be 'recovered' by the sum of its approximation at a lower scale J_0 and the functions of the detail lost between the scales J_0 and J. Hence, with the properties (b) and (c) of definition 2.3.1,

$$L^2(\mathbb{R}) = \bigoplus_{j \in \mathbb{Z}} W_j. \tag{2.8}$$

This implies that an L^2-function can be split into mutually exclusive parts, and each part is in one of the detail subspaces W_j. We achieve this by the Q_j orthogonal projection onto the subspace, W_j. The spaces W_j, possess the scaling property (d), of definition 2.3.1,

$$f(x) \in W_j \Longleftrightarrow f(2^j x) \in W_j\ \forall\ j.$$

Let $\psi \in W_0$, and as $W_0 \subset V_1$, a corresponding expression to the scaling equation is

$$\psi(x) = \sqrt{2} \sum_{k \in \mathbb{Z}} g_k\ \phi(2x - k), \tag{2.9}$$

where the unique set of coefficients $\{g_k\}_{k \in \mathbb{Z}} \in l^2(\mathbb{Z})$ are known as the *high pass filter* associated with ψ (the wavelet function) (see Daubechies (1992) for more details). The filters h_k and g_k can be linked through the relationship

$$g_k = (-1)^k\, h_{1-k},\ k\ \in\ \mathbb{Z}, \tag{2.10}$$

(Daubechies, 1992, page 326). The filters $\{h_k\}$ and $\{g_k\}$ are known as the *quadrature mirror filters.*

We define a family of functions

$$\psi_{j,k}(x) := 2^{j/2} \psi(2^j x - k), \qquad \text{for } j,\ k\ \in\ \mathbb{Z}, \tag{2.11}$$

similarly to $\phi_{j,k}(x)$, in equation (2.5). The set $\{\psi_{j,k}\}_{k \in \mathbb{Z}}$ forms a basis of $L^2(\mathbb{R})$, which implies that $\{\psi_{j,k}\}_{k \in \mathbb{Z}}$ forms an orthonormal basis for W_j for a fixed j. Hence, $\{\psi_{0,k}\} \in W_0$ is an orthonormal basis for W_0.

By following a similar argument to determine the scaling function refinement relation in equation (2.7), we can determine a similar relationship for the wavelet function,

$$\psi_{j,k}(x) = \sum_{n\in\mathbb{Z}} g_{n-2k}\ \phi_{j+1,n}(x). \tag{2.12}$$

From the representations of the scaling and wavelet coefficient in equations (2.7) and (2.12) and using the relationship between the mirror filters in (2.10), we can express the filters as

$$h_{n-2k} = \langle \phi_{j,k}, \phi_{j+1,n} \rangle, \qquad \text{and} \qquad g_{n-2k} = \langle \psi_{j,k}, \phi_{j+1,n} \rangle. \tag{2.13}$$

If the scaling function can be expressed in an analytical form, which is not always the case, the low and high pass filters in equation (2.13) can be obtained by setting $j = k = 0$.

From equation (2.13), we can see that if the scaling function has a compact support, then the low pass filter and consequently the high pass filter posses a finite number of non-zero coefficients. Therefore, the wavelet function can be expressed as a finite linear combination of compactly supported functions and consequently ψ is also compactly supported.

Vidakovic (1999, page 54) proves the filters satisfy the orthogonality relations,

$$\sum_{n\in\mathbb{Z}} h_n\, h_{n-2k} = \delta_{0,k} \qquad \text{and} \qquad \sum_{n\in\mathbb{Z}} g_n\, g_{n-2k} = \delta_{0,k},$$

and

$$\sum_{n\in\mathbb{Z}} h_n\, g_{n-2k} = 0,$$

where $\delta_{a,b}$ is the Kronecker delta,

$$\delta_{a,b} = \begin{cases} 1 & \text{if } a = b, \\ 0 & \text{otherwise.} \end{cases}.$$

An alternate proof can also be found in Daubechies (1992).

2.3.2 Function Representation

We know that as $\{\phi_{j,k}(x)\}_k$ and $\{\psi_{j,k}(x)\}_k$ are orthonormal bases for V_j and W_j, respectively, and using the multi-resolution representation of $L^2(\mathbb{R})$ in (2.8), we can deduce

that

$$\{\phi_{j_0,k}(x)|k \in \mathbb{Z}\} \bigcup \{\psi_{j,k}(x)|j \geq j_0,\, k \in \mathbb{Z}\}$$

is an orthonormal basis for $L^2(\mathbb{R})$. Therefore, from the relation in equation (2.8) and property 2.3.1(b), by projection onto the resolution subspace V_{j_0}, we can represent functions $f \in L^2(\mathbb{R})$ by

$$f(x) = \sum_{k\in\mathbb{Z}} c_{j_0,k}\, \phi_{j_0,k}(x) \; + \; \sum_{j\geq j_0} \sum_{k} d_{j,k}\, \psi_{j,k}(x), \qquad \forall\, f \in L^2(\mathbb{R}), \tag{2.14}$$

where $c_{j_0,k} = \langle f, \phi_{j_0,k}\rangle$ and $d_{j,k} = \langle f, \psi_{j,k}\rangle$. The first summation represents an overall behaviour at coarse scale, also known as the primary resolution of the function f. The second summation is the detail of f from the finest to the coarsest scale.

Any $f \in L^2(\mathbb{R})$ can be written, in the limit, as a linear combination of wavelet functions at different scales and locations, with the representation

$$f(x) = \sum_{j\in\mathbb{Z}} d_{j,k}\, \psi_{j,k}(x). \tag{2.15}$$

The components $\{d_{j,k}\}_{j,k\in\mathbb{Z}}$, are known as the *wavelet coefficients* and contain the information of the function f at scale j, near $2^{-j}k$. The wavelet coefficients and the decomposition wavelet completely characterise f, and accurate computation of these components are of vital importance.

2.3.3 The Discrete Wavelet Transform

Mallat (1989) proposed a fast cascade algorithm for computing the *discrete wavelet transform* (DWT). It is based upon the nested structure of MRA, whereby scaling and wavelet coefficients are computed from the scaling coefficients at the previously finer scale. Therefore, the calculation of the inner products $\langle f, \phi_{j,k}\rangle$ and $\langle f, \psi_{j,k}\rangle\ \forall\, j,\, k$ are not required.

Suppose $f \in L^2(\mathbb{R})$, then for a fixed j, take its approximation on the space V_j to be

$$(P_j f)(x) = \sum_{k\in\mathbb{Z}} c_{j,k}\phi_{j,k}(x),$$

where $\{c_{j,k} = \langle f, \phi_{j,k}\rangle\}_k$ are the scaling coefficients of scale j. The detail produced by

moving between two consecutive approximation spaces V_j and V_{j+1} is given by

$$(Q_j f)(x) = \sum_{k \in \mathbb{Z}} d_{j,k} \psi_{j,k}(x),$$

where $\{d_{j,k} = \langle f, \psi_{j,k} \rangle\}_k$ are the wavelet coefficients at scale j.

As $(P_{j+1} f)(x) = (P_j f)(x) + (Q_j f)(x)$, then we obtain

$$\sum_k c_{j+1,k}\, \phi_{j+1,k} \;=\; \sum_l c_{j,l}\, \phi_{j,l} \;+\; \sum_k d_{j,l}\, \psi_{j,l}.$$

From the relation in equation (2.13) and the orthogonality of the spaces V_j and W_j, implies that

$$c_{j,k} = \sum_n h_{n-2k}\, c_{j+1,n}, \tag{2.16}$$

$$d_{j,k} = \sum_n g_{n-2k}\, c_{j+1,n}. \tag{2.17}$$

Let $\mathbf{c}_j = \{c_{j,k}\}_{k \in \mathbb{Z}}$ and $\mathbf{d}_j = \{d_{j,k}\}_{k \in \mathbb{Z}}$. Then, if we have the coefficients for a particular scale ($\mathbf{c}_j$ and $\mathbf{d}_j$), we can obtain the scaling and wavelet coefficients at the next coarser scale ($\mathbf{c}_{j-1}$ and $\mathbf{d}_{j-1}$) by the application of equations (2.16) and (2.17).

In practice, usually the continuous function $f(x)$ is unknown, but instead we know a collection of discrete observations of the function values $f(x_i)$ for $i = 0, \ldots, T-1$. Suppose these observations have been made at equally spaced locations and the number of observations made are dyadic, (i.e. the sample size is $T = 2^J$ for a fixed $J \in \mathbb{Z}$). Since only the observations are known for $f(x_i)$ for $i = 0, \ldots, 2^J - 1$, it is not possible to calculate the scaling coefficients at the finest scale exactly; therefore, we need to approximate them.

It is assumed the observed data values can be used to approximate the finest scaling coefficients, giving

$$c_{J,i} = f(x_i).$$

Then, by applying the low and high pass filters, g and h, the coarser scale wavelet and scaling coefficients can be recursively calculated using (2.16) and (2.17). Applying the filters to each scale will produce a new sequence of coefficients which are half the length of the previous finer scale. For example, starting with the data $\mathbf{c}_J$ one application of the filters will produce the new sequences $\mathbf{c}_{J-1}$ and $\mathbf{d}_{J-1}$ of length $2^{J-1} = T/2$. This is repeated until the desired level of decomposition has been met, which gives the transformed data

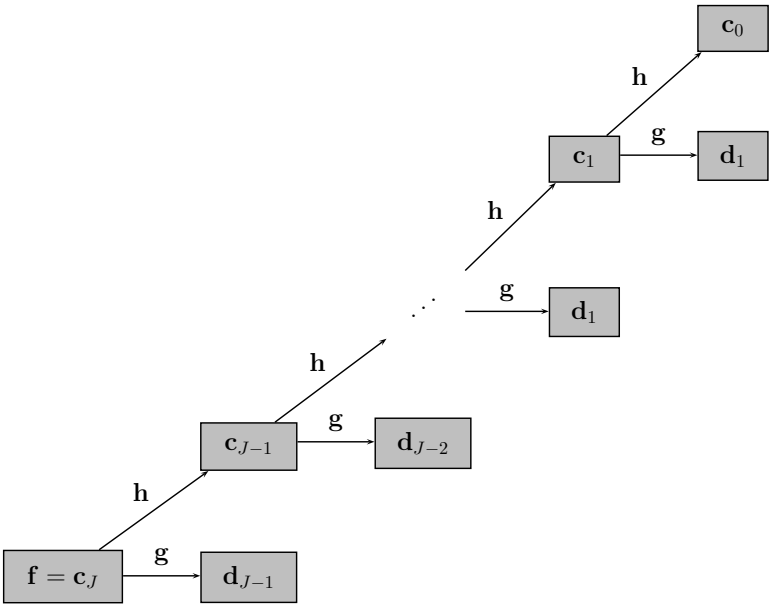


Figure 2.1: Flow diagram of the discrete wavelet transform of an observed data set using successive applications of the low and high pass filters g and h.

vector:

$$\mathrm{DWT}(\mathbf{f}) = (\mathbf{c}_{j_0}, \mathbf{d}_{j_0}, \mathbf{d}_{j_0+1}, \ldots, \mathbf{d}_{J-1}),$$

where $\mathbf{c}_{j_0}$ is a vector of the smooth coefficients at the *primary resolution level*. When the primary resolution $j_0 = 0$, a full decomposition is conducted upon the data and the vector $\mathbf{c}_0$ is length 1, see figure 2.1.

The DWT can also be recursively reversed. The coefficients for a particular scale can be obtained from the previously coarser scale using the relationship

$$c_{j,k} = \sum_{l=0}^{2^{j-1}} c_{j-1,l}\, h_{k-2l} + \sum_{l=0}^{2^{j-1}} d_{j-1,l}\, g_{k-2l}. \tag{2.18}$$

The inversion of the DWT results in a doubling in the number of coefficients at each scale.

Boundary Conditions

Problems occur when the support of the wavelet used in the decomposition extends beyond the length of the data. Some solutions were suggested by Nason and Silverman (1994):

- **Symmetry**: assume symmetry at the end points and reflect the observations at the beginning and the end of the observed data vector to extend the original length of

sampled function vector.

- **Periodic**: assume that the function is periodic over the range of the data, and loop back to the beginning of the observed data, i.e. $f_{k+T} = f_{k-T} = f_k$ for $k = 0, \ldots, T-1$.
- **Zero padding**: assume that the function values outside of the vector range are zero.

Other solutions include the *lifting scheme* (Sweldens, 1996) or to use specifically designed wavelets, such as those described in Cohen et al. (1993).

Operator Notation and Decimation

The low and high pass filters are sometime written in an operator form on the sequence space L^2 as follows. We define the convolution functions $\mathcal{G}, \mathcal{H} : L^2 \to L^2$, such that $\forall\, s \in L^2$

$$(\mathcal{G}s)_k = \sum_l g_{l-k}\, s_l, \tag{2.19}$$

and

$$(\mathcal{H}s)_k = \sum_l h_{l-k}\, s_l. \tag{2.20}$$

Let $\mathcal{D}_0$ be an operator that takes every even element of a sequence s, for example

$$(\mathcal{D}_0 s)_k = s_{2k}, \tag{2.21}$$

and let $\mathcal{D}_1$ be the operator which selects every odd element of a sequence. These operations are known as *dyadic decimation*.

If the operator $\mathcal{D}_0$ is combined with $\mathcal{G}$ and $\mathcal{H}$ on sequence c_j we obtain the decompositions in equations (2.16) and (2.17).

$$(\mathcal{D}_0(\mathcal{H}\mathbf{c}_j))_k \;=\; \left(\mathcal{D}_0\left(\sum_n h_{n-k}\, c_{j,n}\right)\right)_k \;=\; \sum_n h_{n-2k}\, c_{j,n} \;=\; c_{j-1,k},$$

and

$$(\mathcal{D}_0(\mathcal{G}\mathbf{c}_j))_k \;=\; \left(\mathcal{D}_0\left(\sum_n g_{n-k}\, c_{j,n}\right)\right)_k \;=\; \sum_n g_{n-2k}\, c_{j,n} \;=\; d_{j-1,k}.$$

Therefore, the DWT filtering steps can also be implemented using the filter convolutions in (2.19) and (2.20) and then applying the decimation at each step of the transform.

2.3.4 Non-decimated Wavelet Transform

One problem with the DWT is that it is not translation invariant, i.e. starting the transformation from an alternative location will result in a different decomposition of the data. An alternative would be to use the *translation invariant wavelet transform* (Coifman and Donoho, 1995) also known as the *non-decimated wavelet transform* (Nason and Silverman, 1995, NDWT). This uses the convolution operators (2.19) and (2.20), but does not decimate at any step of the DWT.

Therefore, the wavelet coefficients at all possible scales and locations are computed. As a complete representation of the function values can be obtained from the DWT, the NDWT is an example of an *over complete* system.

The non-decimated discrete family of wavelets are calculated by

$$\tilde{\psi}_{j,k}(x) = 2^{j/2}\psi(x-k), \qquad \text{for } j,\ k\ \in\ \mathbb{Z}. \tag{2.22}$$

2.3.5 Examples of Wavelets

We shall now give three examples of wavelet bases: the Haar, Shannon and Daubechies Extremal Phase Wavelets.

The Haar Wavelet Basis

One of the simplest wavelets one can use is the Haar wavelet (Haar, 1911). The Haar wavelet has little practical value, however it is often used to demonstrate theoretical findings because its simplicity makes it easily calculable.

Definition 2.3.2 (Haar Scaling and Wavelet Function)**.** The Haar scaling function is defined as

$$\phi^H(x) = \begin{cases} 1 & \text{if } 0 \le x < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{2.23}$$

The Haar wavelet function is defined as

$$\psi^H(x) = \begin{cases} 1 & \text{if } 0 \le x < 1/2, \\ -1 & \text{if } 1/2 \le x < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{2.24}$$

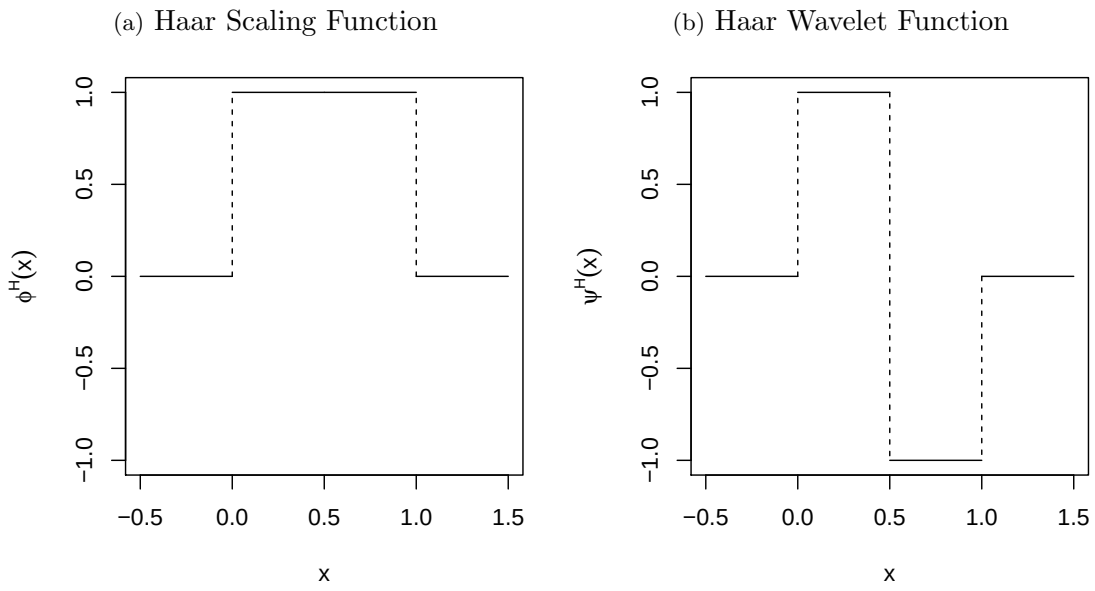


Figure 2.2: Plots of Haar scaling (a) and wavelet (b) function.

See figures 2.2(a) and 2.2(b) for plots of the scaling and wavelet function, respectively. Both are clearly compactly supported.

Applying equation (2.11) to the Haar wavelet function results in the continuous Haar wavelet of the form

$$\psi^H_{j,k}(x) = \begin{cases} 2^{-j/2} & \text{if } 2^j k \leq x < 2^j(k+1/2), \\ -2^{-j/2} & \text{if } 2^j(k+1/2) \leq x < 2^j(k+1), \\ 0 & \text{otherwise,} \end{cases}$$

for $j \in \mathbb{N}$ and $x, k \in \mathbb{R}$.

The Haar wavelet is a sampled version of the continuous-time wavelet, but this is not the case for all wavelets. If the wavelets are compactly supported, then this will be reflected in the discrete wavelets. For example the discrete Haar wavelet is

$$\psi^H_{j,k} = \begin{cases} 2^{-j/2} & \text{for } k = 0, \ldots, 2^{j-1} - 1, \\ -2^{-j/2} & \text{for } k = 2^{j-1}, \ldots, 2^j - 1, \\ 0 & \text{otherwise,} \end{cases} \tag{2.25}$$

for $k \in \mathbb{Z}$ and $j \in \mathbb{N}$.

The Haar wavelet is not an appropriate basis for many applications for several reasons.

- The discontinuities in the Haar's decomposition are not effective for approximating

smooth functions.

- Haar wavelets are well localised in the time domain but decay slowly at a rate of $\mathcal{O}(T^{-1})$ in the frequency domain, therefore more scales of detail are required.

The Shannon Wavelet Basis

Whereas the Haar wavelet and scaling functions are compactly supported in the time domain, the Shannon wavelet and scaling functions are compactly supported in the frequency domain. The origins of the Shannon wavelet stems from the *sinc* function in Claude Shannon's pioneering paper (Shannon, 1949).

Definition 2.3.3 (Shannon Scaling Function)**.** In the time domain the scaling function is defined as

$$\phi^S(x) = \frac{\sin(\pi x)}{\pi x}, \qquad x \in \mathbb{R}. \tag{2.26}$$

In the frequency domain the Shannon wavelet function is defined as

$$\hat{\phi}^S(\omega) = \mathbb{I}_{[-\pi,\pi]}(\omega), \qquad \omega \in \mathbb{R}, \tag{2.27}$$

as shown by Chui (1997, page 46).

Definition 2.3.4 (Shannon Wavelet function)**.** In the time domain the scaling function is defined as

$$\psi^S(x) = \frac{\sin(2\pi x) - \cos(\pi x)}{\pi(x - 1/2)}, \qquad x \in \mathbb{R}. \tag{2.28}$$

In the frequency domain the Shannon wavelet function is defined as

$$\tilde{\psi}^S(\omega) = \exp(-\mathrm{i}2^{-1}\omega)\mathbb{I}_{[-2\pi,-\pi)\cup(\pi,2\pi]}(\omega), \qquad \omega \in \mathbb{R}, \tag{2.29}$$

also shown by Chui (1997, page 64).

In the time domain the scaling function is infinitely differentiable, with an infinite support. The resulting wavelet function in the time domain is also smooth. Figures 2.3(a) and 2.3(b) show the Shannon scaling and wavelet function, respectively, demonstrate the smoothness of these functions in the time domain. However, the Shannon basis possess an infinite length filter, which induce poor localisation properties. This undesirable property means that in practice this wavelet is not often used. However, it can be very useful for

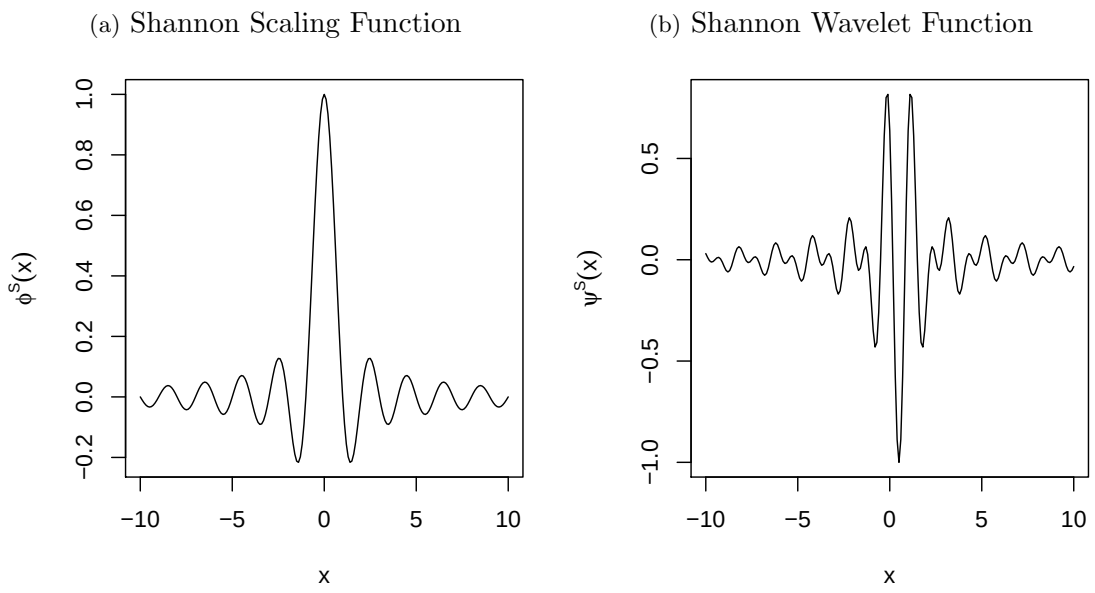


Figure 2.3: Plots of Shannon scaling (a) and wavelet (b) function in the time domain.

theoretical investigations, because it can demonstrate the versatility and robustness of wavelet based methods.

Daubechies Extremal Phase Wavelets

To describe Daubechies Extremal Phase wavelets we first define the vanishing moments of a wavelet.

Definition 2.3.5 (Vanishing Moments)**.** A wavelet has n vanishing moments if

$$\langle x^k, \psi(x) \rangle \;=\; \int x^k \, \psi(x) \, \mathrm{d}x = 0, \qquad \text{for } k \;\in\; \{0, \ldots, n-1\}.$$

The vanishing moment property of a wavelet will affect the smoothness of a wavelet, such that as the number of vanishing moments increases so does the smoothness of the wavelet.

Vanishing moments imply that the wavelet expansion coefficients will be small or zero on smooth parts of a signal and will be large at points of discontinuity. This leads to sparse function representations. A wavelet with n vanishing moments is sometimes referred to as a wavelet of order n.

Daubechies Extremal Phase family of wavelets (denoted by D_n) are compactly supported over the minimum support for n vanishing moments. The scaling functions possess a support of $[0, 2n-1]$ and the wavelet functions have a support over $[1-n, n]$. This family also includes the Haar wavelet, when $n = 1$. For vanishing moments $n > 1$ there

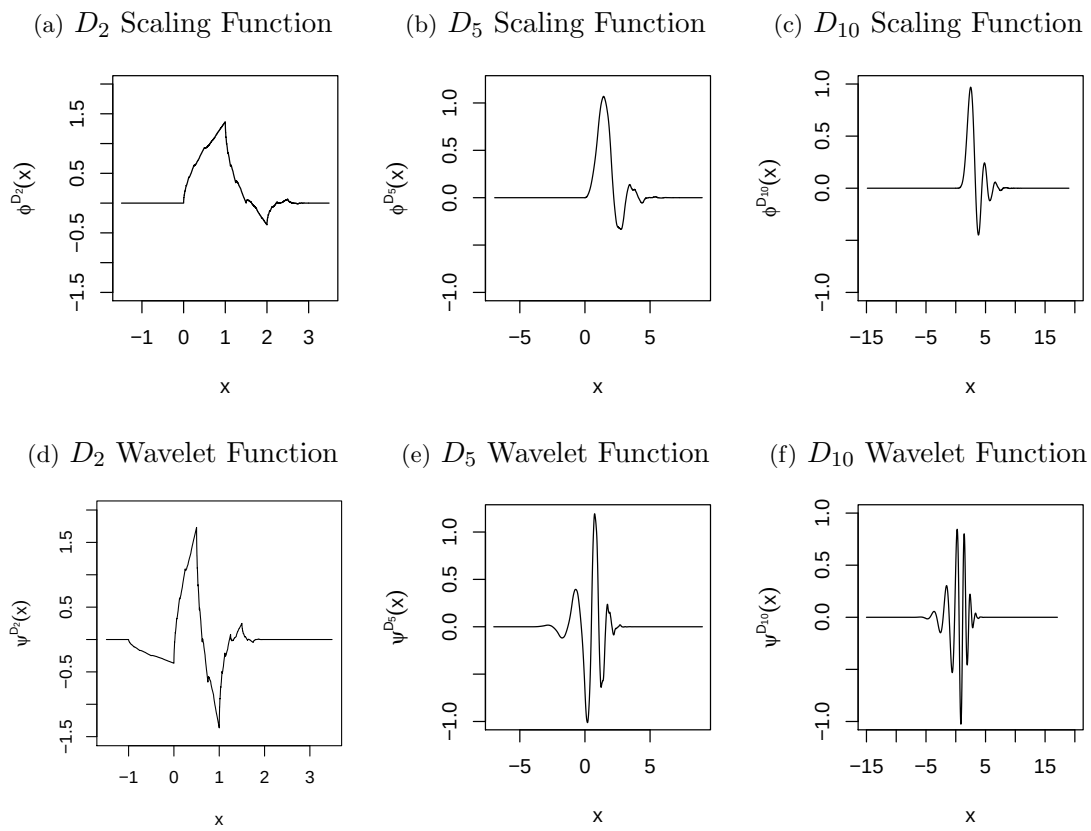


Figure 2.4: Plots of the scaling function and wavelet function of Daubechies Extremal Phase wavelet with 2, 5 and 10 vanishing moments.

is no closed form analytical representation in the time domain. More information on the construction and properties of these wavelets can be found in Daubechies (1992, Chapters 6 and 7), Vidakovic (1999, Chapter 3) and Percival and Walden (2000, Chapter 4).

See figures 2.4(a) to 2.4(f) for plots of the scaling and wavelet function of Daubechies Extremal Phase wavelets with $n = 2, 5$ and 10 vanishing moments. These plots demonstrate the compact support of the wavelets, and demonstrate how as the number of vanishing moments increase so does the smoothness of the functions.

2.4 Nonparametric Regression

Suppose we have a set of noisy observations, $y_1, \ldots, y_n$ of an unknown function f, taken at possibly irregularly spaced locations denoted by $x_1, \ldots, x_n$. This can be modelled as

$$y_i = f(x_i) + e_i, \qquad \text{for } i = 1, \ldots, n, \tag{2.30}$$

where $\mathbf{e} = \{e_1, \ldots, e_n\}$ are random variables which denote the noise. These are usually assumed to be independent and identically distributed (iid) with zero mean and variance

σ^2. The aim is to recover the signal f from the noisy observations y_i using the estimator $\hat{f}(x)$.

The quality of the estimator $\hat{f}(x)$ can be assessed through its mean integrated square error,

$$\text{MISE}(\hat{f}, f) \;=\; \mathbb{E}\left[\int_0^1 \{\hat{f}(x) \,-\, f(x)\}^2 \mathrm{d}x\right].$$

However, in practice $\hat{f}$ is estimated at the points $x_1, \ldots, x_n$, so we use the average mean square error to assess the performance

$$\text{AMSE}(\hat{f}, f) \;=\; \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\{\hat{f}(x) \,-\, f(x)\}^2\right].$$

We shall now discuss methods of estimating f when the form of f is not of a simple parametric form. These methods are known as nonparametric regression.

2.4.1 Linear Smoothing Methods

We shall briefly describe a few linear smoothing techniques. A comprehensive overview of nonparametric smoothing methods can be found in Simonoff (1996), Green and Silverman (1993) and Wahba (1990).

A popular technique to estimate f is to use a weighted average of the noisy data in a window around x,

$$\hat{f}(x) \;=\; \sum_{i=1}^{n} y_i \, w_i(x),$$

where $w_i(x)$ are the weight functions, which are only non zero for values of i, such that x_i are close to x. The window or bin width defines what we quantify as 'close' to x, and it also tunes the smoothness of the resulting function. The choice of this bin width is very important, because if it is too narrow then estimate will possess high variability, whereas if the bin width is too large then the data will be over-smoothed and the estimate will be biased. The weights can be constructed by using scaled and translated transformations of a kernel function. A kernel function will either have

- a concentration of mass at 0 and compact support, or
- rapid decay outside the interval $[-1, 1]$.

Some classic examples of kernel estimators are Watson (1964) and Nadaraya (1964).

An alternative method is to use basis expansions (see Ramsay and Silverman (1997)

for a review). The estimate is obtained from

$$\hat{f}(x) \;=\; \sum_{k=1}^{K} c_k\, \varphi_k(x),$$

which is tuned by the number K of basis functions $\{\varphi_k(\cdot)\}_k$ in the expansion and the chosen basis. Selecting the best basis can be based upon prior knowledge of the properties of function f. The coefficients in this approach can be estimated using a least squares approach, such as

$$\min_{c} \left\{ \sum_{i=1}^{n} \left(y_i \;-\; \sum_{k=1}^{K} c_k\, \varphi_k(x_i) \right)^2 \right\}.$$

Another technique is to use smoothing splines (Silverman, 1985), which balances goodness of fit to the noisy data but also a certain degree of smoothness. A common solution is to choose $\hat{f}$ such that

$$\min_{\hat{f} \in C^2} \left\{ \sum_{i=1}^{n} (y_i - \hat{f}(x_i))^2 + \lambda \|\hat{f}''\|_{L^2}^2 \right\},$$

where λ controls the balance between the smoothness of $\hat{f}$ and its similarity to the data. Smaller values for λ imply less penalty is paid for the roughness, and the curve becomes more wiggly and we have something closer to linear regression.

These methods work well when the function f is smooth. However, they perform poorly when there is evidence of discontinuities within f. Therefore, other non-linear methods have been developed. We shall review one technique relevant to this thesis, here.

2.4.2 Wavelet Shrinkage

In this section we shall discuss function estimation using wavelet shrinkage of a discretely sampled function. Excellent reviews of these methods can be found in Abramovich et al. (2000) and Antoniadis (2007). Unlike the methods described in section 2.4.1, there are fewer limitations to the form of the function f for which can be estimated via wavelet shrinkage. Further information on this topic is discussed later in this section *Wavelet Shrinkage and Besov Spaces*.

In statistics, wavelet shrinkage was first introduced by Donoho and Johnstone in the early 90's with a series of reports such as Donoho and Johnstone (1994*a*,*b*, 1995) and Donoho et al. (1996), to name but a few. A good overview of wavelet shrinkage can be found in Nason (1995) and Vidakovic (1999, Chapter 6) and methodology to perform

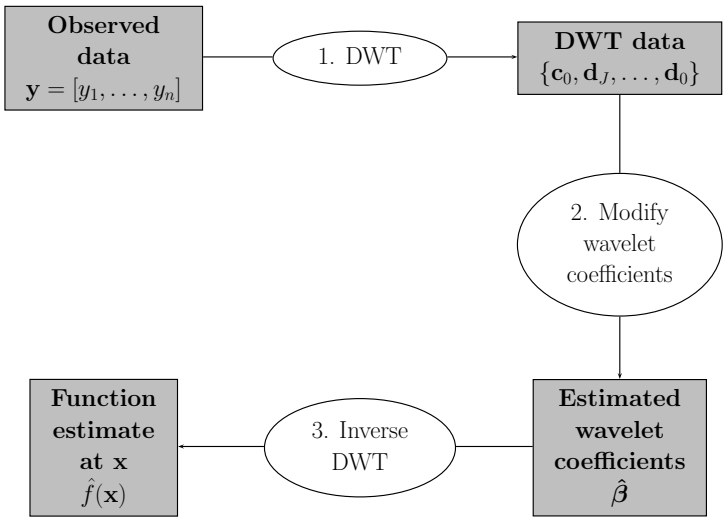


Figure 2.5: A flow diagram of wavelet shrinkage of (2.30).

wavelet shrinkage in `R` is described in Nason (2008, Chapter 3).

Assume we wish to find f from the noisy data $\mathbf{y} = \{y_i\}$ in equation (2.30). Wavelet shrinkage is very simple and can be performed in three steps, as shown in figure 2.5, with details next.

In step 1 we apply the DWT to noisy data $\mathbf{y}$. This results in the DWT of equation (2.30), which can be written as

$$\mathbf{d} = \boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2.31}$$

where $\mathbf{d} = W\mathbf{y}$, $\boldsymbol{\beta} = Wf(\mathbf{x})$, $\boldsymbol{\varepsilon} = W\mathbf{e}$ and W is the orthogonal discrete wavelet transform matrix. The vector of values denoted by β are considered to be the 'true' wavelet coefficients and $\mathbf{d}$ are the noisy empirical wavelet coefficients.

Step 2 is based on two consequences of the DWT

- If $\mathbf{e}$ is iid with mean zero and variance σ^2, then the DWT of $\mathbf{e}$, denoted by $\boldsymbol{\varepsilon}$, also has mean zero and variance σ^2. Furthermore, if the distribution of the noise is assumed Gaussian, then subsequently ε is also Gaussian.

- Many noiseless functions (f) possess a sparse wavelet representation. This implies that the true wavelet coefficients, $\boldsymbol{\beta}$, are mainly zero, with a few large coefficients that represent the signal f. We can conclude that small values of $\mathbf{d}$ represent noise in the data. Therefore, by using a *shrinkage technique*, which sets values below a certain *threshold* value to zero, and either keeping or shrinking the remaining values, we can obtain an proxy for the true coefficients.

Choosing the correct shrinkage technique and threshold value is of great importance, because if the cut off is too high then we can effectively over smooth the data, whereas if it is too low the recovered signal is too noisy.

Once we have obtained the new estimates of the wavelet coefficients, we apply the inverse wavelet transform. The result is the smoothed estimate of the true signal f at the locations x_i. As the wavelet matrix, W, is orthogonal, this implies that Parseval's theorem (as described in section 2.2.1) holds and the estimates in the different domains have the same risk, i.e.

$$\|\hat{f} - f\|_{L^2}^2 = \|W^T(\hat{\beta} - \beta)\|_{L^2}^2 = \|\hat{\beta} - \beta\|_{L^2}^2.$$

Variations in step 2 will yield different estimates of the function f, which we will describe next.

Shrinkage Techniques

There are different methods for shrinking the wavelet coefficients (Donoho and Johnstone, 1994*b*), such as hard shrinkage, which is defined as

$$\Delta^h(d_{j,k}, \delta) = d_{j,k}\, \mathbb{I}(|d_{j,k}| > \delta) \tag{2.32}$$

and soft shrinkage

$$\Delta^s(d_{j,k}, \delta) = \text{sign}(d_{j,k})(|d_{j,k}| - \delta)_+ \tag{2.33}$$

for the set threshold value $\delta \in [0, \infty)$, where

$$\text{sign}(x) = \begin{cases} -1 & \text{if } x < 0, \\ 1 & \text{if } x > 0, \\ 0 & \text{if } x = 0. \end{cases}$$

With hard shrinkage any coefficients with an absolute value less then the threshold is set to zero, and any coefficient value greater is kept. Soft shrinkage shrinks the wavelet coefficients by δ if it is greater than δ, and sets to zero the coefficients which are less than δ in absolute value. Hard shrinkage tends to have a bigger variance due to the shrinkage function's discontinuity, but this means it is better at detecting discontinuities in the signal. Soft shrinkage tends to have bigger bias because all of the large coefficients are

reduced in size (Bruce and Gao, 1996), but the results are visually smoother with smaller variance.

It was suggested by Gao and Bruce (1997) that a more appropriate method would be Firm shrinkage (also known as semi-soft shrinkage). In this case the shrinkage function is

$$\Delta^f(d_{j,k},\delta_1,\delta_2)=\begin{cases}0 & \text{if } |d_{j,k}| \leq \delta_1,\\ \text{sign}(d_{j,k})\frac{\delta_2(|d_{j,k}|-\delta_1)}{\delta_2-\delta_1} & \text{if } \delta_1 < |d_{j,k}| \leq \delta_2,\\ d_{j,k} & \text{if } |d_{j,k}| > \delta_2,\end{cases}$$

which is essentially, a combination of the hard and soft threshold.

Thresholds

There are many different methods to select δ, such as the *universal threshold* from Donoho and Johnstone (1994*b*)

$$\delta=\hat{\sigma}\sqrt{2\log(n)}, \tag{2.34}$$

which is only related to the data through the variance estimate $\hat{\sigma}^2$. This was motivated by

$$\lim_{T\to\infty}\mathbb{P}(\{\max_i|e_i|>\sigma\sqrt{2\log(T)}\})=0,$$

which assumes that the distribution of e_i is iid $\mathcal{N}(0,\sigma^2)$ for all i. Values of $d_{j,k}$ greater than $\sigma\sqrt{2\log(n)}$ are most likely to correspond to coefficients which contain signal, i.e. $\beta_{j,k}\neq 0$. Using the universal threshold in either the hard or soft shrinkage technique will remove coefficients due to noise with high probability. However, often the estimate is over-smoothed.

Another popular threshold is Stein's Unbiased Risk Estimator, also known as the *SURE threshold*, introduced by Donoho and Johnstone (1995), which is based on an unbiased estimate of the average mean-squared error of a multivariate Gaussian mean in Stein (1981).

Suppose we have a n dimensional vector $\mathbf{d}\sim\mathcal{N}_n(\mu,\Sigma)$, for which we wish to estimate the mean μ. An estimate $\hat{\mu}=\mathbf{d}+g(\mathbf{d})$ is constructed, where $g:\mathbb{R}^n\to\mathbb{R}^n$ is a weakly differentiable function, then an unbiased estimate of the average mean-squared error can be produced from

$$S(g,\mathbf{d})=Tr(\Sigma)+\|g(\mathbf{d})\|_{L^2}^2+2Tr(\Sigma Dg(\mathbf{d})),$$

where $Dg(\mathbf{d})$ is a $n \times n$ matrix with elements $\partial g_{j,k}(\mathbf{d})/\partial d_{j,k}$, $Tr(X)$ is the trace of matrix X. This result implies

$$\mathbb{E}_{\mu}[\|\hat{\mu} - \mu\|_{L^2}] = \mathbb{E}[S(g, \mathbf{d})]\,.$$

Soft shrinkage for a particular δ is essentially the same as the previously described method in Stein (1981), for $g(d_{j,k}) = \text{sign}(d_{j,k})(|d_{j,k}| - \delta)_+$. Therefore,

$$S(\delta, \mathbf{d}) = n\sigma^2 + \sum_{j,k} \min(d_{j,k}^2,\, \delta^2) - 2\sigma^2 \sum_{j,k} \mathbb{I}_{[-\delta,\delta]}(d_{j,k}),$$

which suggests the appropriate threshold choice of

$$\delta = \text{argmin}_{0 \le \delta \ge \sigma\sqrt{2\log(n)}} S(\delta, \mathbf{d}). \tag{2.35}$$

Equation (2.35) is the sure threshold. Although it usually exhibits smaller values than the universal threshold, there are problems when the wavelet coefficients are sparse.

However, a hybrid scheme, known as *SureShrink*, was suggested by Donoho and Johnstone (1995). The universal threshold is used when the coefficients are sparse for a particular scale, and a scale dependent SURE threshold otherwise. Donoho and Johnstone (1995) demonstrated that SureShrink is asymptotically minimax over a range of Besov spaces, such that an estimate is minimax over a particular function space $\mathcal{F}$, if $\sup_{g\in\mathcal{F}}\text{AMSE}(\hat{f}, f)$ coincides with the minimax risk, $\inf_{\hat{f}} \sup_{g\in\mathcal{F}}\text{AMSE}(\hat{f}, f)$.

Bayesian Wavelet Shrinkage

Bayesian statistical methods start with existing *prior* knowledge of model parameters (β), which are updated using the data ($\mathbf{x}$) to give *posterior* knowledge. The resulting posterior knowledge can be used to produce interpretations about these parameters. The model commonly used for Bayesian inference is

$$p(\beta|\mathbf{x}) = \frac{p(\mathbf{x}|\beta)p(\beta)}{\int_X p(\mathbf{x}|\beta)p(\beta)\mathrm{d}x},$$

where $p(\mathbf{x}|\beta)$ is known as the *likelihood*, $p(\beta)$ is known as the *prior density* function and $p(\beta|\mathbf{x})$ is the *posterior density* function. Bayesian inference provides a natural interpretation of parameter intervals which are known as credible intervals.

In general, Bayes rules are 'shrinkers', which has led to the Bayesian paradigm becoming increasingly popular in wavelet analysis. The Bayes rule can be constructed to slightly

shrink large coefficients and heavily shrink small coefficients. Most practicable Bayes rule can be easily computed by simulation or expressed in closed form. Good reviews on some Bayesian approaches in the wavelet domain can be found in (Ruggeri and Vidakovic, 2005; Abramovich et al., 2000; Vidakovic, 1998*b*).

We place a prior distribution upon each wavelet coefficient, therefore treating the true wavelet coefficients as 'model parameters' and the data are the noisy wavelet coefficients in the above scenario. Often, it is assumed that an accurate estimate of the error is available, and the distribution of error is known, with pdf $\zeta(\cdot)$. This allows for a closed-form expression of the posterior means and variance.

Due to the sparsity property of the wavelet transform, we know that there will be few 'true' wavelet coefficients which are non-zero. Therefore, noisy coefficients which are close to zero, are probably due to noise in the data. The likelihood is the distribution of the error from equation (2.30), in many cases, these are assumed iid Gaussian. The *Berger-Müller* prior is a popular choice (Abramovich et al., 1998; Clyde et al., 1998; Vidakovic, 1998*a*; Johnstone and Silverman, 2005), as it represents the sparsity present in the wavelet coefficients. The Berger-Müller prior is defined as

$$p(\beta_{j,k}) = \alpha_j \, \delta(0) \; + \; (1-\alpha_j)\tau_j^{-1}\xi\Big(\tau_j^{-1}\beta_{j,k}\Big)\,, \tag{2.36}$$

where $1-\alpha_j$ is the probability that $\beta_{j,k}$ is non zero, τ_j^2 is the prior variance and ξ is a symmetric distribution which will represent the distribution of the non-zero coefficients such as the Gaussian or Laplace distribution.

Often, the posterior distribution is difficult to calculate, so the posterior mean or median is used instead. Barber et al. (2001) suggested using cumulants of the posterior distribution and Johnson transformations (Johnson, 1949) to obtain credible intervals for wavelet regression estimates.

Hyperparameter Determination

Determining the hyperparameters is one of the most important and difficult aspects of any Bayesian analysis. In the case of Bayesian wavelet shrinkage, the correct hyperparameters will affect the capability of the method to recover the true signal and can either be determined in a full or empirical Bayesian approach. A fully Bayesian approach uses prior knowledge to determine all of the hyperparameters in the model, such as the methods described in Clyde et al. (1998) and Vidakovic (1998*a*). An empirical approach uses

the data to determine either some or all of these hyperparameters (Chipman et al., 1997; Abramovich et al., 1998; Clyde and George, 1998; Johnstone and Silverman, 1998).

The method of Abramovich et al. (1998) recommended that the precision and weights of the Berger-Müller prior should be ascertained using the marginal maximum likelihood estimates (MMLE). The hyperparameters, α_j and τ_j from equation (2.36) are defined as

$$\tau_j^2 = 2^{-aj}\mathcal{C}_1 \qquad \text{and} \qquad \alpha_j = 1 - \min(1, 2^{-bj}\mathcal{C}_2) \qquad \text{for } j = 0, \ldots, J-1, \tag{2.37}$$

where $J = \log_2(T)$ of sample size T, $\mathcal{C}_1$ and $\mathcal{C}_1$ are constants to be estimated using MMLE, and $a > 0$ and $0 < b \leq 1$ are hyper-hyperparameters to be chosen by the user.

Then, for a particular level j, suppose there are n_j wavelet coefficients which are greater than the threshold value δ. Let these wavelet coefficients be denoted by $u_{j,i}$, for $i = 1, \ldots n_j$. By conditioning on n_j, the $u_{j,i}$'s are independent realisations from the tail of $\xi(\cdot)$ beyond $\pm\delta$ with a variance of $\sigma^2 + \tau_j^2$, where σ^2 is the variance of the error distribution. By substituting in $\tau_j^2 = 2^{-aj}\mathcal{C}_1$ the log likelihood for $\xi(\cdot)$ can be calculated using $u_{j,i}$. Numerical maximisation can then be used to produce an estimate for $\mathcal{C}_1$, which is used to produce the estimate for τ_j^2.

The value of $\mathcal{C}_2$ can be determined using information obtained from estimation of the precision. Let the probability that a wavelet coefficient (β) is greater than the threshold (δ), conditional on the $\beta_{j,k} \neq 0$ be defined as

$$q_j = 2\,F\left(-\frac{\delta}{\sqrt{\sigma^2 + \tau_j^2}}\right),$$

where $F(\cdot)$ is the cumulative distribution function of the prior $\xi(\cdot)$. The estimated number of non zero wavelet coefficients at a particular scale j is n_j/q_j and the expected value of n_j/q_j is $2^{(1-b)j}\mathcal{C}_2$. Given b, $\mathcal{C}_2$ can be estimated by method of moments, therefore

$$\mathcal{C}_2 = \begin{cases} \frac{2^{1-b}-1}{2^{(1-b)J}-1} \sum_{j=0}^{J-1} \frac{n_j}{q_j} & \text{for } 0 \leq b < 1, \\ \frac{1}{J} \sum_{j=0}^{J-1} \frac{n_j}{q_j} & \text{for } b = 1. \end{cases}$$

Using this method it is important to select the correct values of the hyper-hyperparameters a and b, and threshold δ. Abramovich et al. (1998) suggested using the universal threshold value for δ, and the default values $a = 0.5$ and $b = 1$.

Rather than imposing any structure on the hyperparameters one could use empirical

Bayes and allow the form of the data to determine these values. One option would be to use *direct maximum likelihood estimation* of the log likelihood ($\mathcal{L}$) based upon the marginal distribution of the data. If we use the Berger-Müller prior, the marginal log likelihood is defined as

$$\mathcal{L}(\alpha_j, \tau_j) = \sum_k \log\Big\{\alpha_j \, \tau_j^{-1} \, \zeta(\tau_j^{-1} \, d_{j,k}) + (1 - \alpha_j)\gamma(d_{j,k}|\tau_j)\Big\}, \tag{2.38}$$

where $\zeta(\cdot)$ is the error pdf, and $\gamma(d_{j,k}|\tau_j)$ is the convolution of $\zeta(\cdot)$ and $\xi(\cdot)$ from (2.36), defined as

$$\gamma(d_{j,k}|\tau_j) = \int_{-\infty}^{\infty} \tau_j^{-1} \xi(\tau_j^{-1} x) \, \zeta(d_{j,k} - x) \, \mathrm{d}x.$$

Usually, numerical methods are required to produce the estimates of $\hat{\alpha}_j$ and $\hat{\tau}_j$.

Alternatively, George and Foster (1997) suggested a *conditional maximum likelihood* approach. Instead of maximising over both components, take the larger of the two components, i.e.

$$\mathcal{L}(\alpha_j, \tau_j) = \sum_k \log\Big\{\max\Big(\alpha_j \, \tau_j^{-1} \, \zeta(\tau_j^{-1} \, d_{j,k}), (1 - \alpha_j)\gamma(d_{j,k}|\tau_j)\Big)\Big\}. \tag{2.39}$$

This was further developed by Clyde and George (1998) in a wavelet context, who also suggesting using the expectation-maximisation (EM) algorithm (Dempster et al., 1977) as another possible method to determine the hyperparameters.

Wavelet selection

We shall define any wavelet used to perform wavelet shrinkage as the *smoothing* wavelet. Selecting an appropriate smoothing wavelet relies heavily on the characteristics of the function being estimated. For example, if a 'true' function is piecewise constant the same properties of the Haar wavelet will improve the estimation. However, the form of the function to be estimated is not usually known, so often prior knowledge from a field expert will be helpful, or as a default one of the smoother wavelets.

Primary Resolution

The primary resolution is the coarsest scale which we apply shrinkage to the wavelet coefficients. It is similar to the bin width parameter in linear smoothing methods and was first investigated by Hall and Turlach (1997). A variety of methods to select the primary resolution have been investigated, which rely on either the sample size, generalised cross validation or Bayesian inference (Nason, 2001; Lu et al., 2003; Park et al., 2008). Some

of these methods depend heavily upon the regression model, and most assume that the distribution of the errors is Gaussian. However, Bayesian wavelet shrinkage is considered an alternative method to selecting a primary resolution as the Bayes rule will automatically address this problem(Chipman et al., 1997).

Cycle Spinning

Using wavelet shrinkage to smooth data can sometimes lead to visual artefacts which, are caused by features either in the signal or basis or both. The resulting smoothed data can exhibit Gibbs phenomena, and/or alternating under and over estimation of a specific target. The size of the effect is directly related to the location, for example due to the nature of the Haar wavelet, an effect in or very close to a dyadic location will not cause any problems. However, at other locations this can cause quite a significant influence on the result.

One method to mitigate these issues is to use cycle spinning (Coifman and Donoho, 1995). It works by:

- Shifting the data a random number of positions.
- Curve estimation on the shifted data.
- Un-shifting the estimate to their original positions.

This process is repeated a certain amount of times, with different random shifts, and the average is taken as the final estimate. If the data were cycle spun for all possible positions, the result would yield the NDWT described in 2.3.4.

Wavelet Shrinkage and Besov Spaces

To enable non-trivial theory to be produced on the convergence rates of an estimator, we shall restrict f, such that it belongs to a ball, $B^r_{p,q}(A)$, in the Besov space, $B^r_{p,q}$ for $p \geq 1,\ q \leq \infty$ and non integer r.

The Besov norm of the function f is then related to a sequence space norm on the wavelet coefficients $(d_{j,k})$ of the function. The sequence space norm is given by

$$\|d\|_{B^r_{p,q}} = \begin{cases} |c_{0,0}| + \left(\sum_{j=0}^{\infty} 2^{jr'q} \left[\sum_{k=0}^{2^j-1} |d_{j,k}|^p \right]^{p/q} \right)^{1/q} & \text{if } 1 \leq q \leq \infty, \\ |c_{0,0}| + \sup_{j \geq 0} \left(2^{jr'q} \left[\sum_{k=0}^{2^j-1} |d_{j,k}|^p \right]^{1/p} \right) & \text{otherwise,} \end{cases} \tag{2.40}$$

where $r' = r + 1/2 - 1/p$ for $j \geq 0$ and $k = 0, \ldots 2^j - 1$ (Donoho et al., 1995).

If the wavelet function ψ is of regularity s, where $\max\left(0, 1/p - 1/2\right) < s < r$, then we have

$$C_1 \|f\|_{B^r_{p,q}} \leq \|d\|_{B^r_{p,q}} \leq C_2 \|f\|_{B^r_{p,q}},$$

where C_1 and C_2 are constants independent of f (Meyer, 1992; Donoho and Johnstone, 1995). Therefore, the Besov norm is equivalent to the corresponding sequence of norm in (2.40) depending on q. If $p = q = 1$, the sequence space norm in the first case in (2.40) becomes a weighted sum of the absolute wavelet coefficients $d_{j,k}$, and the Besov space norm is L^1-norm on the derivatives of f up to order r.

2.5 Time Series Data

Time series are a general class of processes consisting of a collection of observations made sequentially through time. Time series data occur in many different fields such as signal analysis, medical analysis, geology and economics to name a few. There are two types of time series, *discrete* and *continuous*. A discrete time series, represented by $\{X_t\}_{t=1}^T$ is a series which has been measured at discrete, successive time points and are often taken at uniformly spaced intervals. An example of a discrete time series might be a daily record of temperatures. A continuous time series is represented by $X(t)$, where observations are made continuously through time, i.e. for any $t \in \mathbb{R}$. An electroencephalograph (EEG) record measuring the electrical activity between two points in an individual's brain is an example of a continuous time series. Our research mainly addresses discrete time series. We shall assume, from this point onward that any trend and seasonality have been removed from the data.

Often an observation at one time point will influence subsequent observations, which makes the order of the data important. Therefore, the data is non-exchangeable and standard iid statistical analysis cannot always be utilised. However, simple summary statistics can be estimated from a time series such as the mean, $\mathbb{E}[X_t]$, and variance, $\mathrm{Var}[X_t]$ but these may change with time. One method of examining the dependence structure between observations is to calculate the covariance, known as the autocovariance function (acv.f). The acv.f is defined as

$$\gamma_t(\tau) = \mathrm{Cov}[X_t, X_{t+\tau}], \tag{2.41}$$

where τ is the time difference between observations and is known as lag. The size of the autocovariance coefficients will depend on the unit which X_t has been measured in. Sometimes it can be helpful to standardise the acv.f to produce the autocorrelation function (acf) denoted by $\rho_t(\tau)$. Assuming $\text{Var}[X_t] < \infty$, the acf is derived from the acv.f as

$$\rho_t(\tau) = \frac{\text{Cov}[X_t, X_{t+\tau}]}{\text{Var}[X_t]}. \tag{2.42}$$

Another important method of measuring the dependence structure between observations is through the partial autocorrelation function (p.acf). Given a time series X_t of lag τ is the autocorrelation between X_t and $X_{t+\tau}$ when the linear dependence of X_{t+1} through to $X_{t+\tau-1}$ has been removed. From equation (2.42) the partial autocorrelation function is defined as

$$\mathcal{P}_t(\tau, \tau) = \frac{\rho_t(\tau) - \sum_{i=1}^{\tau-1} \mathcal{P}_t(\tau-1, i)\ \rho_t(\tau-i)}{1 - \sum_{i=1}^{\tau-1} \mathcal{P}_t(\tau-1, i)\ \rho_t(i)}, \tag{2.43}$$

where

$$\mathcal{P}_t(\tau, i) = \mathcal{P}_t(\tau-1, i) - \mathcal{P}_t(\tau, \tau)\, \mathcal{P}_t(\tau-1, \tau-i). \tag{2.44}$$

For example, for $\tau = 2$ equation (2.43) can be written as

$$\mathcal{P}_t(2, 2) = \frac{\rho_t(2) - \rho_t(1)^2}{1 - \rho_t(1)^2}.$$

The two main goals of time series analysis are to:

- determine the nature of the phenomenon represented by the sequence of observations, and possibly provide a model for those observations,
- forecast (predict) future possible values.

Methods of time series analysis can be performed in either the frequency or time domain and there is no single methodology to analyse a time series data that is suitable across all fields. Usually the field from which the data was obtained and aim of the analysis will dictate the appropriate methodology.

2.5.1 Stationary Time Series

A useful class of stochastic processes are the *stationary* processes, where the statistical properties of the underlying random process do not change over time. Processes with this

property sometimes arise from a physical system which has achieved a steady state. Stationarity is a strong assumption postulated by classical time series analysis, it guarantees that increasing the sample size will lead to more information of the same nature. This is crucial for consistent asymptotic theory. The concepts of stationarity are well described in Priestley (1994, pages 105 – 106) and Chatfield (2004, pages 34 – 36).

The most widely used type of stationarity assumption is *second-order stationarity.* A process, X_t is second order stationary if it satisfies all of the following:

- $\mathbb{E}[X_t] = \mu < \infty$ for all t.
- $\gamma(\tau)$ does not depend on t, for all t.
- $\text{Var}[X_t] = \gamma(0) < \infty$ for all t.

Stationary Processes in the Time Domain

One of the most efficient techniques for fitting a model to a stationary time series is the Box-Jenkins methodology. For more detail on the Box-Jenkins method see Kirchgaässner et al. (2013); Chatfield (2004); Box et al. (2008), the latter is for the more advanced reader. Essentially the Box-Jenkins approach involves three steps, identification, estimation and diagnostic checking. This procedure uses the data's past behaviour to select the most appropriated forecasting model. If we have a discrete stationary time series, the Box-Jenkins methodology assumes that the time series data can be best represented by one of the following models:

- Autoregressive (AR) models: forecasts of a variable based on linear function of its past values,
- Moving Average (MA) models: forecasts based on linear combination of past errors,
- Autoregressive-Moving Average (ARMA) models: combination of the previous two categories.

ARMA(p, q) models consists of p AR components, and/or q MA components, defined as

$$X_t = \underbrace{\sum_{i=1}^{p} \alpha_i \, X_{t-i}}_{AR} + \underbrace{\sum_{j=0}^{q} \beta_j \, Z_{t-j}}_{MA}, \tag{2.45}$$

where Z_t is an iid, random zero mean process, and α_i and β_j are the model parameters to be estimated.

Given a stationary time series, the sample p.acf and acf can be used to estimate values for p and q. The p.acf of an AR(p) process becomes zero at lag greater than p, and the acf if an MA(q) process is zero for lag $q+1$ and greater. The 95% confidence interval of the sample p.acf and acf can be used to determine which values are non-significant departures from zero.

The AR parameters α_i can be estimated using least squares. However estimating the MA parameters β_i is more problematic as the errors are non-linear functions of the parameters, therefore numerical methods are required to minimise the residual sums of squares. Once an estimated ARMA(p, q) model has been found, the residuals of the model can be obtained by subtracting the fitted estimates from the observed values. There are a variety of residual analysis methods discussed in Box et al. (2008, chapter 8). In forecasting, if short term correlation is present in the time series, using the Box-Jenkins approach can produce some of the most accurate short term forecasts (Chatfield, 2004). Although it is advisable to accompany these forecasts with a confidence interval, these can often be too narrow.

Stationary Processes in the Frequency Domain

If a time series is stationary, the mirror of the acv.f, within the frequency domain, is the *power spectral density function* or *spectrum*. The spectrum describes how the variance or energy of a time series is distributed according to frequency. If the spectrum exists, it can be derived from the acv.f by

$$f(\omega) = \frac{1}{\pi} \sum_{\tau=-\infty}^{\infty} \gamma(\tau)\, e^{-i\omega\tau}, \qquad -\pi \leq \omega \leq \pi, \tag{2.46}$$

where ω is the frequency. Formula (2.46) is the discrete Fourier transform of the acv.f. The spectrum can then be used to find the acv.f by using the inverse Fourier transform,

$$\gamma(\tau) = \frac{1}{2} \int_{-\pi}^{\pi} f(\omega)\, e^{i\omega\tau}\, \mathrm{d}\omega, \qquad \tau \in \mathbb{Z}. \tag{2.47}$$

Every stationary stochastic process $\{X_t\}$ can be represented by the *Cramér representation*, which has the form

$$X_t = \int_{-\pi}^{\pi} A(\omega)\, e^{it\omega}\, dz(\omega), \qquad \text{for } t \in \mathbb{Z} \text{ and } \omega \in [-\pi, \pi], \tag{2.48}$$

where $A(\omega)$ is the amplitude of the process X_t, $e^{it\omega}$ is the oscillation and $dz(\omega)$ is an orthonormal increments process. This means that $dz(\omega) = z(\omega) - z(\omega-)$ and $z(\omega-) \in \mathbb{C}$ is a stochastic process with

$$\mathbb{E}[z(\omega)^2] = F(\omega), \tag{2.49}$$

where $f(\omega)$ is the spectral distribution function, where $F(\omega) = \int f(x)\mathrm{d}x$. The spectral distribution function is a positive, real function of the frequency variable, ω, associated with the stationary stochastic process. Also

$$\mathbb{E}[\{z(\omega_1) - z(\omega_2)\}\{\overline{z(\omega_3) - z(\omega_4)}\}] = 0,$$

for $\omega_1 \geq \omega_2 > \omega_3 \geq \omega_4$, is a property of the orthogonal increments process.

2.5.2 Non-Stationary Time Series

Many models from classical time series analysis rely heavily on the assumption of stationarity. For real-life time series it is often difficult to determine whether the process is really stationary using the available observations. Often the statistical behaviour of a real-life process can shift from one stationary regime to another or evolve slowly over time, which may only become apparent with continued observation. Time series of this form are often referred to as piecewise or locally stationary, respectively. Seismic data is a good example, the occurrence of an earthquake can cause a sudden change in the statistical properties of the process for a short period of time.

As the key assumption of stationarity is violated, none of the previous models described can be applied to data of this form. Analysing the data for small time intervals or other transformations might be utilised so that the assumption of stationarity still holds. However, analysing the data on smaller time intervals compromises accuracy and raises questions, such as

- How many time intervals are required?
- Where should the time intervals be placed?

Another possibility is to difference the data. This will reduce the sample size, and sometimes it is difficult to determine the number of times the data should be differenced to reach a stationary regime.

If it is inappropriate to assume stationarity, there is a large class of models available.

For example, one could model the ARMA process from equation (2.45) with time varying coefficients. These time dependent coefficients can be estimated using different techniques, such as those described by Grenier (1983) and Subba Rao (1970).

We shall describe non-stationary processes in the context of the *evolutionary spectra* as in Priestley (1965). Consider the stochastic process $\{X_t\}$ for $-\infty < t < \infty$ and assume $\mathbb{E}[X_t] = \mu(t),\ \forall\, t$. We restrict our attention to the class of processes for which there exists a family $\mathcal{F}$ of functions, $\vartheta(\omega)$, defined on the real line and indexed by t, and a measure $\lambda(\omega)$ also on the real line, such that for each t, τ the covariance function from (2.41), $\gamma_t(\tau) = \mathrm{Cov}[X_t, X_{t+\tau}]$, can be represented in the form

$$\gamma_t(\tau) = \int_{-\infty}^{\infty} \vartheta_t(\omega)\, \overline{\vartheta_{t-\tau}(\omega)}\ \mathrm{d}\lambda(\omega) \qquad -\infty < t, \tau < \infty. \tag{2.50}$$

For $\mathrm{Var}[X_t]$ to be finite for each t, $\vartheta_t(\omega)$ must be quadratically integrable with respect to the measure λ, for each t. If the autocovariance function can be represented as equation (2.50), then the process $\{X_t\}$ can be written as

$$X_t = \mu(t) + \int_{-\infty}^{\infty} \vartheta_t(\omega)\ \mathrm{d}\xi(\omega),$$

where $\mu(t)$ is the trend and $\xi(\omega)$ is the orthogonal process, such that $\mathbb{E}\big[|\mathrm{d}\xi(\omega)|^2\big] = \mathrm{d}\lambda(\omega)$, see Bartlett (1955, page 143) or Grenander and Rosenblatt (1957, page 27). The measure $\lambda(\omega)$ is similar to the stationary integrated spectrum $F(\omega)$ in (2.49). The analogy to the absolutely continuous spectrum is obtained if we assume $\lambda(\omega)$ *is absolutely continuous with respect to Lebesgue measure.*

Suppose, for each ω, $\vartheta_t(\omega)$ (as a function of t), possesses a generalised Fourier transform whose modulus has an absolute maximum at frequency $\theta(\omega)$. Then we can consider $\vartheta(\omega)$ as an amplitude modulated sine wave with frequency $\theta(\omega)$ and

$$\vartheta(\omega) = A(t, \omega) e^{\mathrm{i}\theta(\omega)t},$$

where the t dependent amplitude function, $A(t, \omega)$, is such that its Fourier transform has an absolute maximum at the origin (zero frequency). Furthermore if $\{\vartheta(\omega)\}$ is such that $\theta(\omega)$ is a single-valued function of ω (i.e. only one family member with a Fourier transform

with a maxima at the same point), then $\theta(\omega) = \omega$ with

$$\gamma_t(\tau) = \int_{-\infty}^{\infty} A(t,\omega)\,\overline{A(t-\tau,\omega)}\,e^{\mathrm{i}\omega\tau}\,\mathrm{d}\lambda(\omega) \qquad -\infty < t,\, \tau < \infty$$

and

$$X_t = \mu(t) + \int_{-\infty}^{\infty} A(t,\omega)\,e^{\mathrm{i}\omega t}\,\mathrm{d}\xi(\omega). \tag{2.51}$$

Further details on non-stationary processes with time dependent spectral representation can be found in Priestley (1994, chapter 11).

Asymptotic Theory

One of the main problems with non-stationary time series analysis is how to tackle the problem of asymptotic theory. For stationary processes, as the sample size increases more information of the same nature is obtained. However, increasing the number of future observations of a non-stationary process will not necessarily increase the amount of statistical information on the process. Therefore, an alternative approach needs to be considered for asymptotic theory.

Instead, fix the time interval and let the number of times we sample the process within that interval tend to infinity. This leads to rescaled time, first introduced by Dahlhaus (1997), which provides a framework for asymptotic theory in non-stationary time series analysis.

Remark 2.5.1 (Rescaled Time)**.** Given a time series of length T the rescaled time at location k is defined as

$$z = \frac{k}{T} \quad \in \ (0,1)\,.$$

For example

$$z_0 = \frac{0}{T}, z_1 = \frac{1}{T}, z_2 = \frac{2}{T}, \ldots, z_{T-1} = \frac{T-1}{T}.$$

2.5.3 Locally Stationary Fourier Processes

The locally stationary Fourier process is a natural extension of the time varying spectral representation given in equation (2.51). We shall apply the concept of rescaled time and consider for $\omega \in [-\pi, \pi]$.

Definition 2.5.1 (Locally Stationary Fourier Process (Dahlhaus, 1997))**.** The stochastic process $\{X_t\}_{t=0}^{T-1}$ with transfer function $A_t^*(\omega)$ and trend μ is locally stationary if there exists a representation

$$X_{t,T} = \mu\left(\frac{t}{T}\right) + \int_{-\pi}^{\pi} A_{t,T}^*(\omega)\ e^{\mathrm{i}\omega t}\ d\xi(\omega). \tag{2.52}$$

where

(i) $\xi(\omega)$ is a stochastic process on $[-\pi, \pi]$ with $\overline{\xi(\omega)} = \xi(-\omega)$ and the cumulants of the k^{th} order

$$\operatorname{cum}\{\mathrm{d}\xi(\omega_1), \ldots, \mathrm{d}\xi(\omega_k)\} = \eta\left(\sum_{j=1}^{k}\omega_j\right)\ g_k(\omega_1, \ldots, \omega_{k-1})\ \mathrm{d}\omega_1, \ldots, \mathrm{d}\omega_k,$$

where

- $g_1 = 0$,
- $g_1(\omega) = 1$,
- $|g_k(\omega_1, \ldots, \omega_{k-1})| \leq \mathcal{C}_k$, where $\mathcal{C}_k$ is a constant $\ \forall\ k$,
- $\eta(\omega) = \sum_{j=1}^{k}\delta(\omega + 2\pi j)$.

$g_1 = 0$, $g_1(\omega) = 1$ and $|g_k(\omega_1, \ldots, \omega_{k-1})| \leq \mathcal{C}_k\ \forall\ k$,

(ii) There exists a constant $\mathcal{K}$ and 2π-periodic function $A : [0,1]\times\mathbb{R} \to \mathbb{C}$ with $A(z, -\omega) = \overline{A(z,\omega)}$ and

$$\sup_{t,\omega}\left|A_{t,T}^*(\omega) - A\left({}^{t}\!/_{T}, \omega\right)\right| \ \leq\ \mathcal{K}T^{-1} \qquad\qquad \forall\ T$$

where we assume $A(z,\omega)$ and $\mu(z)$ are continuous in z.

The locally stationary behaviour of the process is ensured by the smoothness of A in z, in other words the amplitude function should not be too irregular.

2.6 Locally Stationary Wavelet Processes

Locally stationary wavelet (LSW) processes were introduced by Nason et al. (2000), and extended to encompass a larger range of processes in Van Bellegem and von Sachs (2008). We use the extension here. Assume that the wavelets used in this thesis are compactly supported, and that the length of the support for any wavelet $\psi_{j,0}$ is equal to $\mathcal{L}_j :=$

$|\text{supp}\psi_{j,0}|$. Therefore, if we have J scales, where $J-1$ is the finest scale, then

$$|\text{supp}\psi_{j,k}| = \mathcal{L}_j = (2^{J-j}-1)(\mathcal{L}_{J-1}-1)+1 \;\forall\, j \geq 0,$$

where $\mathcal{L}_{J-1}$ is the support at the finest scale.

Definition 2.6.1 (The Locally Stationary Wavelet Process)**.** A LSW process is a sequence of doubly indexed stochastic processes, $\{X_{t,T}\}_{t=0,\ldots,T-1}$, where $T = 2^J$ for some $J \in \mathbb{N}$. This process has the representation

$$X_{t,T} = \sum_{j=0}^{\infty} \sum_{k=-\infty}^{\infty} w_{j,k:T}\; \tilde{\psi}_{j,k-t}\; \xi_{j,k}, \tag{2.53}$$

where $\tilde{\psi}_{j,k-T}$ is a discrete non-decimated family of wavelets as in (2.22), for scales $j = 0,\ldots,\infty$, $k = -\infty,\ldots,\infty$ based on a mother wavelet, $\psi(t)$, of compact support and $\xi_{j,k}$ is a random zero mean orthonormal increments sequence. The component $w_{j,k:T}\,\xi_{j,k}$ can be thought of as a random amplitude of the oscillation $\tilde{\psi}_{j,k-t}$.

The quantities in equation (2.53) possesses the properties:

(a) $\mathbb{E}[\xi_{j,k}] = 0 \quad \forall\;\; j,\; k \;\Rightarrow\; \mathbb{E}[X_t] = 0.$

(b) $\mathbb{E}[\xi_{j,k}, \xi_{j',k'}] = \delta_{j,j'}\,\delta_{k,k'}$.

(c) There exists for each $j \geq 1$ a function $W_j(z)$ for $z = {}^{k}/_{T} \in (0,1)$, that possesses the following properties

i.
$$\sum_{j=1}^{\infty} |W_j(z)|^2 < \bar{C} \quad \text{uniformly in } z \in (0,1)\,.$$

ii. There exists a sequence of constants C_j such that for each T

$$\sup_k |w_{j,k;T} - W_j(z)| \leq \frac{C_j}{T}.$$

iii. The total variation (TV) of $W_j^2(z)$ is bounded by L_j, that is

$$\begin{aligned} TV(W_j^2) &:= \sup\left\{ \sum_{i=1}^{I} |W_j^2(a_i) - W_j^2(a_{i-1})| : 0 < a_0 < \ldots < a_I < 1, I \in \mathbb{N} \right\} \\ &\leq\; L_j. \end{aligned}$$

iv. The constants C_j and L_j are such that

$$\sum_{j=1}^{\infty} \mathcal{L}_j(\mathcal{L}_j L_j + C_j) \leq \rho < \infty.$$

2.6.1 The Evolutionary Wavelet Spectrum

The time evolution of properties of X_t depend on how fast $w_{j,k;T}$ changes as a function of k. Since $\tilde{\psi}_{j,k-t}$ is only 'significantly' non-zero near $k - t = 0$, the behaviour of X_t is directly related to the behaviour of $w_{j,k;T}$. Property (c) ii. from definition 2.6.1, implies that the evolution of $w_{j,k;T}$ is tied to the evolution of $W_j(z)$, where $z = k/T$. In other words, as $T \to \infty$, $w_{j,k;T} \approx W_j(z)$. Hence, the evolution of $w_{j,k;T}$ as a function of k depends on the smoothness of $W_j(z)$, which is controlled by the Lipschitz continuity of $W_j(z)$ as a function of z. The smoother $W_j(z)$ is, the slower the evolution of X_t.

The Evolutionary Wavelet Spectrum (EWS) measures the 'contribution to the variance' of the LSW process $X_{t,T}$ at scale level j and location z. We shall define the EWS next.

Definition 2.6.2 (Evolutionary Wavelet Spectrum)**.** The EWS is a function of the location z and scale $j \in \mathbb{Z}^+$ and is defined by

$$S_j(z) = |W_j(z)|^2, \tag{2.54}$$

for all $j \in \mathbb{N}$ and $z \in (0, 1)$.

Localised Autocovariance Function

The localised autocovariance (LACV) function for LSW process will provide information about the covariance around a particular location $z = k/T \in (0, 1)$. In stationary time series analysis the spectrum is the Fourier transform of the acv.f, and in LSW analysis we observe a similar relationship between the EWS and LACV. To describe this relationship, we require the autocorrelation wavelet (ACW).

The ACW is the autocorrelation function of a non-decimated wavelet. We shall first define the continuous ACW function which is calculated from the wavelet function, and secondly the discrete ACW obtained through the discrete wavelets.

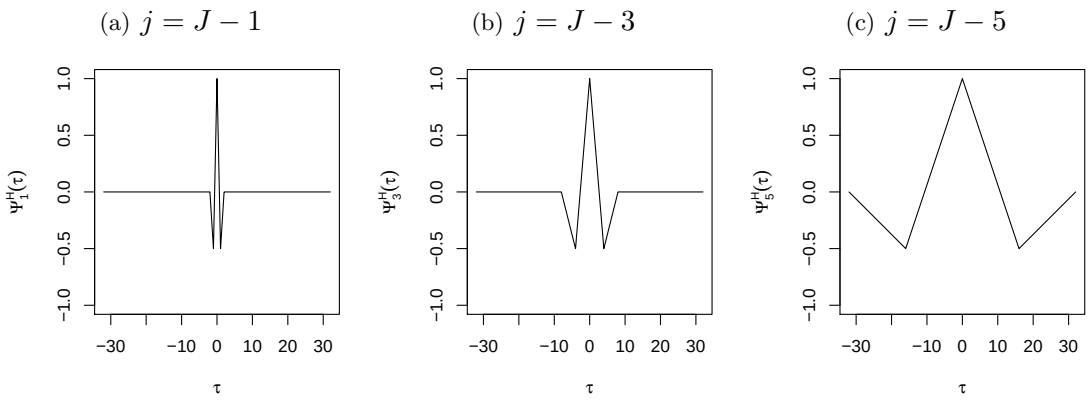


Figure 2.6: Plots of discrete Haar autocorrelation wavelet for at the finest scale $(J-1)$ in (a), third finest scale $(J-3)$ in (b) and fifth finest scale $(J-5)$ in (c), over lag $\tau = -32, \ldots, 32$.

Definition 2.6.3 (Continuous Autocorrelation Wavelet Function)**.** The Continuous ACW is defined by

$$\Psi(u) = \int_{-\infty}^{\infty} \psi(x)\ \psi(x-u)\mathrm{d}x,$$

where $u \in \mathbb{R}$ and support of $\Psi(u)$ is $\mathbb{R}$.

Definition 2.6.4 (Discrete Autocorrelation Wavelet)**.** The ACW at scale $j \in \mathbb{Z}^+$ at lag $\tau \in \mathbb{Z}$ is defined by

$$\Psi_j(\tau) = \sum_{k=-\infty}^{\infty} \tilde{\psi}_{j,k}\ \tilde{\psi}_{j,k-\tau}.$$

The discrete ACW determines the autocorrelation of a wavelet at a particular scale and different locations. We can also obtain the discrete ACW, $\Psi_j(\tau)$, using the continuous ACW function, $\Psi(u)$, for $j \geq 0$ by

$$\Psi_j(\tau) = \Psi\left(2^{-j}|\tau|\right), \tag{2.55}$$

for all Daubechies compactly supported wavelets. The discrete ACW provides a family of symmetric, compactly supported, positive semi-definite functions on $\tau \in \mathbb{Z}$.

Examples of the discrete Haar autocorrelation wavelet for scales $j = 1, 3, 5$ can be seen in figures 2.6. These represent the first $(J-1)$, third $(J-3)$ and fifth $(J-5)$ finest scales. As support of the wavelet increases, the overlap between wavelets at different locations increases. Therefore there is a reduction in the rate of decay in the correlation as the lag $|\tau|$ increases.

A component which is closely related to the ACW is the Cross-Autocorrelation Wavelet

(C-ACW), also known as the cross-scale autocorrelation wavelet (Fryzlewicz and Nason, 2006). It is used to examine the correlation structure of the wavelets between two scales as the difference (lag) between locations varies

Definition 2.6.5 (Cross-Autocorrelation Wavelet)**.** The C-ACW wavelet is defined by

$$\Psi_{l,m}(\tau) = \sum_{k=-\infty}^{\infty} \tilde{\psi}_{l,k}\ \tilde{\psi}_{m,k-\tau},$$

between scales l, $m \in \mathbb{Z}^+$ for a location difference of $\tau \in \mathbb{Z}$.

Note. If $l = m$ in definition 2.6.5 for the C-ACW, we have the standard discrete autocorrelation wavelet.

As we have now defined the required components, we shall now describe how the LACV is obtained via the EWS when combined with the ACW next.

Definition 2.6.6 (The Localised Autocovariance Function)**.** The LACV function is defined as

$$C(z,\tau) = \sum_{j=0}^{\infty} S_j(z)\Psi_j(\tau), \qquad \text{for } \tau \in \mathbb{Z},\ z \in (0,1)\,, \tag{2.56}$$

where $\{S_j(z)\}_{j=0}^{\infty}$ is the EWS from definition 2.6.2.

In practice, the LACV function cannot actually be obtained using equation (2.56) because an infinite sequence of the estimated EWS ($\{\hat{S}_j(z)\}_{j=0}^{\infty}$) is not available (Eckley, 2001, chapter 6). Therefore, the LACV function can be re-written as the *curtailed local autocovariance*,

$$C_J(z,\tau) = \sum_{j=0}^{J-1} S_j(z)\Psi_j(\tau), \tag{2.57}$$

for $\tau \in \mathbb{Z}$, $z \in [0,1)$ and $J \in \mathbb{N}$.

In Nason et al. (2000, Appendix, page 18) it was proved that the error in the estimate of the LACV compared to the true autocovariance was $\mathcal{O}\left(T^{-1}\right)$. The representation in definition 2.6.6 links together the EWS and the time varying autocovariance of $X_{t;T}$. This is analogous to the Fourier transform of the autocovariance function of a stationary time series, which is known as the stationary spectrum.

Uniqueness of the Autocovariance Representation

In stationary time series analysis the spectrum is the discrete Fourier transform of the autocovariance function (acv.f). Therefore, the acv.f can be obtained by applying the

inverse Fourier transform to the spectrum. Similarly, the EWS can also be obtained from the LACV using the invertible autocorrelation wavelet inner product matrix in definition 2.6.7.

Definition 2.6.7 (The Inner Product Matrix (Nason et al., 2000))**.** The operator $A = (A_{j,l})_{j,l\geq 0}$ is defined by

$$A_{j,l} = \langle \Psi_j, \Psi_l \rangle = \sum_{\tau} \Psi_j(\tau)\ \Psi_l(\tau). \tag{2.58}$$

and the J-dimensional matrix is $A_J = (A_{j,l})_{j,\, l=0,\ldots,J-1}$.

An important theorem (Nason et al., 2000, Theorem 2.15) that allows for the representation of the EWS in terms of the LACV is described below. The proof of this theorem can be found on pages 18–20 of Nason et al. (2000).

Theorem 2.6.1 (Autocorrelation Wavelet)**.** *The family* $\{\Psi_j(\tau)\}_{j=0}^{\infty}$ *are linearly independent. Hence,*

(a) Given a LSW process the corresponding EWS is uniquely defined

(b) All of the eigenvalues of A are positive, which means that A is invertible and for each J the norm $\left\|A_J^{-1}\right\|$ *is bounded above by some* C_J*.*

Part (b) of theorem 2.6.1 allows a representation of the EWS in terms of the LACV, as described next.

Proposition 2.6.2. *For* $j \in \mathbb{Z}^+$ *and* $\tau \in \mathbb{Z}$

$$S_j(z) = \sum_{l=0}^{\infty} A_{j,l}^{-1} \sum_{\tau=-\infty}^{\infty} C(z,\tau)\Psi_l(\tau). \tag{2.59}$$

The EWS from definition 2.6.2 is not unique, unlike the LACV. If we replace $C(z,\tau)$ in equation (2.59) by the curtailed LACV, then as $T \to \infty$ this value will converge to the EWS (see Nason et al. (2000, page 11) for details).

2.6.2 Estimation of the Evolutionary Wavelet Spectrum

To estimate the evolutionary wavelet spectrum we require an estimate of $W_j(z)$, where $z = {}^{k}/_{T}$. We know that for large T, $w_{j,k} \approx W_j(z)$, and from definition 2.6.2, $S_j(z) = |W_j(z)|^2$. The definition of the LSW process in equation (2.53), implies the wavelet transformation of $w_{j,k}$ is approximately $X_{t,T}$ (with error). Therefore, a logical starting point in the estimation of the EWS is the non-decimated wavelet transformation of the process $X_{t,T}$.

Definition 2.6.8 (Empirical Wavelet Coefficients of the LSW Process)**.** Given a data set $X_0, \ldots, X_{T-1}$, the empirical wavelet coefficients are the non-decimated wavelet transformation of the LSW process, defined as

$$\begin{aligned} d_{j,k;T} &= \left\langle X_{t,T}, \tilde{\psi}_{j,k-t} \right\rangle, \\ &= \sum_{t=0}^{T-1} X_{t,T}\, \tilde{\psi}_{j,k-t}, \\ &= \sum_{t=-\infty}^{\infty} X_{t,T}\, \tilde{\psi}_{j,k-t}, \end{aligned} \tag{2.60}$$

where $X_{t,T} = 0$ for $t \neq 0, \ldots, T-1$, $j = 0, \ldots, J-1$, $k = 0, \ldots, T-1$ and $J = \log_2(T)$.

This will produce a matrix of observations with $T \times J$ dimensions. Further exploration of the properties of the LSW process wavelet coefficients can be found in section 3.1.

Using the empirical wavelet coefficients of the LSW process, we can obtain an estimate of the EWS as defined in definition 2.6.9 next.

Definition 2.6.9 (Raw Wavelet Periodogram)**.** The raw wavelet periodogram is defined as

$$I_{j,k;T} = \left| d_{j,k;T} \right|^2. \tag{2.61}$$

where $j = 0, \ldots, J-1$, $k = 0, \ldots, T-1$ and $J = \log_2(T)$.

The raw wavelet periodogram, $I_{j,k}$, is the "power" of the series at location k and scale j.

Asymptotics of the Wavelet Periodogram

The wavelet periodogram $I_{j,k}$ is not an asymptotically unbiased estimator of the evolutionary wavelet spectrum, Nason et al. (2000, proof of proposition 3.3, page 24) show that

$$\mathbb{E}\left[I_{j,k}\right] = \sum_{l} A_{j,l}\, S_l(z) + \mathcal{O}\left(T^{-1}\right), \quad \forall\, z \in [0,1). \tag{2.62}$$

It is simpler to think of $I_{j,k}$ in terms of a vector of raw wavelet periodograms, $\mathbf{I}_k = [I_{0,k}, I_{1,k}, \ldots, I_{J-1,k}]^T$ and similarly $\mathbf{S}_k = [S_{0,k}, S_{1,k}, \ldots, S_{J-1,k}]^T$, for fixed J, $T = 2^J$ and $z = {}^{k}/_{T}$.

This proposition implies that

$$\mathbb{E}[\mathbf{I}_k] = A_J\, \mathbf{S}_k + \mathcal{O}(T^{-1}),$$

where A_J is the restriction of A to the first J rows and columns. Therefore, an unbiased estimator of $\mathbf{S}_k$ is

$$\mathbf{L}_k = A_J^{-1}\mathbf{I}_k, \tag{2.63}$$

and $\mathbb{E}[\mathbf{L}_k] = \mathbf{S}_k + \mathcal{O}(T^{-1})$.

In Nason et al. (2000, page 25) the variance of the wavelet periodogram was calculated as

$$\text{var}[I_{j,k;T}] = 2\left\{\sum_l A_{j,l}\, S_l(z)\right\}^2 + \mathcal{O}\left(\frac{2^j}{T}\right). \tag{2.64}$$

This means that as the sample size increases ($T \to \infty$) the variance does not vanish. In other words, increasing the sample size does not decrease the variance of the raw wavelet periodogram, $I_{j,k;T}$.

In stationary time series analysis a similar problem occurs as the periodogram is not a consistent estimator of the stationary spectrum. There are various parametric and non-parametric techniques which can be used to produce a consistent estimate of the stationary spectrum. For example, autoregressive spectral estimation (Parzen, 1983) is an example of a parametric approach.

Various techniques have already been developed to smooth the wavelet periodogram, such as those by Nason et al. (2000); Fryzlewicz and Nason (2006); Van Bellegem and von Sachs (2008). Although the approach of Van Bellegem and von Sachs (2008) works well in theory, it is difficult to apply this method in practice. This is due to the complicated process required to estimate the parameters. Therefore, we will not discuss this method in further detail. However, we do note their suggestion of using quantiles and the benefits of using this to quantify the estimate.

2.6.3 Translation-Invariant De-Noising EWS Estimator

In Nason et al. (2000) each level, j, of the raw wavelet periodogram is smoothed as a function of z using translation-invariant (TI) de-noising (Coifman and Donoho, 1995). Non-linear wavelet shrinkage is performed on the approximately χ_1^2 distributed raw wavelet periodogram then bias corrected by the inner product matrix (A^{-1}). An appropriate threshold for the shrinkage was determined in Nason et al. (2000, Theorem 3.4). The technique raises a number of questions, such as what is an appropriate wavelet? Nason et al. (2000) believe that smoother wavelets, such as Daubechies extremal phase with 10 vanishing moments, help to avoid 'leakage' of power into the surrounding scales because

of their short support in the Fourier domain. They also produce less spiky and variable estimates in their example.

The function `ewspec` from the package `wavethresh` (Nason, 2012) in `R` was developed to perform this technique for estimating the EWS. This method is very flexible and the user is given many options to define how the smoothing of the raw wavelet periodogram should proceed. For example, should a soft or hard thresholding technique be applied and should the variance of raw wavelet periodogram be calculated by scale or globally.

This method is computationally efficient. However, there are limitations in what type of EWS this method is able to successfully recover. In our definition of the LSW process in 2.6.1, we also include LSW processes with piecewise constant EWS. TI de-noising was developed to recover the EWS from a LSW process with a slowly evolving, time varying, local variance. Alternatively the next method was developed purely for piecewise constant EWS.

2.6.4 Haar-Fisz EWS Estimation

The methodology developed in Fryzlewicz and Nason (2006) produced an estimator which was mean-square consistent, rapidly computable, easy to implement and performs well in practice. However, the technique is restricted to locally stationary processes with a time varying, piecewise constant, local variance.

We shall begin our description of this method by considering a special case of the LSW process from definition 2.6.1. For each $j \geq 0$, $W_j(z)$ is a real valued, piecewise constant function, with a finite, but unknown, number of change points. Let M_j denote the maximum change in $W_j(z)$ at scale j. Suppose the function $W_j(z)$, satisfies:

- $\sum_{j=1}^{\infty} W_j^2(z) < \infty$ uniformly in z,
- $\sum_{j=1}^{\infty} M_j < \infty$.

Assume the constants M_j satisfy $M_j = \mathcal{O}(a^{-j})$, for some $a > 2$. This is a technical assumption, to ensure the rate of approximation of the EWS. Also, if

$$\rho_j(\tau) := \text{supp}_k |\text{Corr}(d_{j,k}, d_{j,k+\tau})|,$$

where $\{d_{j,k}\}$ are the wavelet coefficients from equation (2.60), then assume $\sum_\tau \rho_j(\tau)^2 \leq \mathcal{C}2^{-j}$.

An estimate of the evolutionary wavelet spectrum can be produced by performing the variance stabilising Haar-Fisz transformation on the raw wavelet periodogram, and smoothing the transformed data. Consider the following model:

$$I_{j,k} = \sigma^2_{j,k} Z^2_{j,k}, \qquad \text{for } j = 1, \ldots, \log_2(T), \text{ and } k = 0, \ldots, T-1, \tag{2.65}$$

where

- $I_{j,k}$ is the raw wavelet periodogram from definition 2.6.9.
- $\sigma^2_{j,k}$ is deterministic and close to a piecewise constant function greater than $\sigma^2(z)$ where $z = k/T$, such that

$$T^{-1} \sum_{k=0}^{T-1} \left| \sigma^2_{j,k} - \sigma^2(z) \right|^2 := a_T = o_T(\log^{-1}(T)).$$

  Furthermore, $\sigma^2(z)$ is a piecewise constant function greater than zero, with a finite, but unknown number of change points, denoted by B.

- $\{Z_{j,k}\}$ is multivariate normal with zero mean and variance one. Asymptotically its sequence is absolutely summable, i.e.

$$\rho(\tau) := \sup_{i,T} |\text{Corr}(Z_{j,i}, Z_{j,i+\tau})|,$$

  satisfies $\rho^1_\infty < \infty$, where $\rho^p_\infty := \sum_\tau \rho^p(\tau)$.

The Haar-Fisz transformation of the raw wavelet periodogram for scale j follows the following algorithm.

1. Let $\tilde{c}_{J,m} := I_{j,k}$ for $m = 0, \ldots, T-1$, where $T = 2^J$.
2. For $l = (J-1), \ldots, 0$, recursively define

$$\tilde{d}_{l,m} = \frac{\tilde{c}_{l+1,2m} - \tilde{c}_{l+1,2m+1}}{\sqrt{2}},$$
$$\tilde{c}_{l,m} = \frac{\tilde{c}_{l+1,2m} + \tilde{c}_{l+1,2m+1}}{\sqrt{2}},$$

   where $m = 1, \ldots, 2^l - 1$, and $\tilde{d}_{l,m}$and $\tilde{c}_{l,m}$ are the Haar wavelet and scaling coefficients of the raw wavelet periodogram at scale j respectively.

3. Divide the wavelet coefficients by the scaling coefficients to produce the Haar-Fisz coefficients

$$\tilde{h}_{l,m} = \frac{\tilde{d}_{l,m}}{\tilde{c}_{l,m}}. \tag{2.66}$$

4. For $l = J-1, \ldots, 0$, recursively modify the vectors $\tilde{c}_{l-1}$:

$$\tilde{c}_{l+1,2m} = \tilde{c}_{l,m} + \tilde{h}_{l,m}$$
$$\tilde{c}_{l+1,2m-1} = \tilde{c}_{l,m} - \tilde{h}_{l,m}$$

where $m = 1, \ldots, 2^l$.

5. Let $\mathcal{H}_{j,k} = \tilde{c}_{J,m}$, $m = 1, \ldots, 2^J$.

Let $\mathcal{F}$ denote the non-linear invertible Haar-Fisz operator, hence $\mathcal{H}_{j,k} = \mathcal{F}I_{j,k}$.

If we now consider a special case of equation (2.65),

$$I_{j,k} = \sigma^2(z)Z_{j,k}^2, \qquad \text{for } j = \log_2(T) - 1, \ldots, 0,\ z = \frac{k}{T} \text{ and } k = 0, \ldots, T-1, \tag{2.67}$$

where $\sigma_{j,k}^2 = \sigma^2(z)$, a piecewise constant function greater than zero with a finite but unknown number of change points, and $Z_{j,k} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$, then proposition 6.1 in (Fryzlewicz and Nason, 2006) holds.

Proposition 2.6.3 (Prop. 6.1 in Fryzlewicz and Nason (2006))**.** *If $I_{j,k}$ follows the model in* (2.67)*, then*

(a) Beta distribution: $\tilde{h}_{l,m} \sim 2\beta(2^{J-l-2}, 2^{J-l-2}) - 1$, with mean 0 and variance $(2^{J-l-1}+1)^{-1}$, therefore

$$\tilde{h}_{l,m}\sqrt{2^{J-l-1}+1} \longrightarrow \mathcal{N}(0,1), \qquad \text{as } J \longrightarrow \infty,\ l \leq J(1-\delta),\ \delta \in (0,1).$$

(b) Log-like property of $\mathcal{F}$: we have

$$\frac{1}{T}\sum_{k=0}^{T-1} \mathbb{E}\big[(\mathcal{F}I_{j,k} - \mathbb{E}[I_j]) - (\mathcal{F}\sigma^2(z) - \mathbb{E}[\sigma^2]) - (\mathcal{F}Z_{j,k}^2 - \mathbb{E}[Z_j^2])\big]^2 = \mathcal{O}\left(\frac{\log_2(T)}{T}\right)$$

(c) Variance stabilisation: for all k we have

$$Var\big[\mathcal{F}Z_{j,k}^2\big] = \sum_{l=0}^{J-1}(2^l+1)^{-1} + 2^{1-J} \longrightarrow \sum_{l=0}^{J-1}(2^l+1)^{-1}.$$

(d) Asymptotic Gaussianity: for any k

$$\left(\sum_{l=J-J^*}^{J-1}(2^l+1)^{-1}+2^{1-J}\right)^{-1/2}\mathcal{F}^{(M^*)}Z_{j,k}^2\xrightarrow{d}\mathcal{N}(0,1).$$

(e) Lack of spurious correlation: for any k_1, k_2,

$$\left(\sum_{l=J-J^*}^{J-1}(2^l+1)^{-1}+2^{1-J}\right)^{-1}Cov\Big[\mathcal{F}^{(M^*)}Z_{j,k_1}^2,\mathcal{F}^{(M^*)}Z_{j,k_2}^2\Big]\to 0.$$

The Haar-Fisz operator approximately transforms the model in (2.65), from a multiplicative to an additive model, so that the model in property (b) of proposition 2.6.3 can be interpreted as

$$\mathcal{F}I_{j,k}^2-\mathbb{E}[I_j^2]\approx(\mathcal{F}\sigma^2(z)-\mathbb{E}[\sigma^2])+(\mathcal{F}Z_{j,k}^2-\mathbb{E}[Z_j^2]),$$

where we can think of $\mathcal{F}\sigma^2(z)$ as the signal and $\mathcal{F}Z_{j,k}^2$ as noise.

As the scales become coarser, property 2.6.3(d) implies the distribution of the noise is asymptotically standard Gaussian. However, although $\mathcal{F}Z_{j,k}^2$ is a symmetric random variable which is close to Gaussian, it is not true that $\mathcal{F}Z_{j,k}^2\to$ Gaussian as $J\to\infty$. The distribution of $\mathcal{F}Z_{j,k}^2$ is actually decided by $\tilde{h}_{l,m}$ with the largest variance. This is usually the finest scale, which is far from Gaussian and possesses a bimodal distribution.

The variance stabilisation of property 2.6.3(c) means that the variance of $\mathcal{F}Z_{j,k}^2$ is roughly constant over time. The property in 2.6.3(e), means that the Haar-Fisz transformation does not introduce any spurious correlation (note $Z_{j,k}^2$ was previously assumed to be uncorrelated). However, if the correlation structure of $Z_{j,k}^2$ changes over time, the variance of $\mathcal{F}Z_{j,k}^2$ will not be stabilised exactly, but empirical investigations suggest it will be greatly improved.

One can apply any smoothing technique to the Haar-Fisz transformed data which is suitable for homoscedastic Gaussian data. In Fryzlewicz and Nason (2006), they apply a classical wavelet thresholding estimator based on the Haar wavelets and the soft shrinkage rule with a type of universal threshold (Donoho and Johnstone, 1994*b*). The thresholding can be applied before the inverse wavelet transform (step 4), in which the Haar-Fisz

coefficients are updated using

$$\mu_{l,m} = \begin{cases} \tilde{c}_{l,m}\,\mathrm{sign}(\tilde{h}_{l,m})(|\tilde{h}_{l,m}| - t_l)_+ & (l,m) \in I_{j,k}, \\ 0 & (l,m) \notin I_{j,k}, \end{cases} \tag{2.68}$$

where

$$t_l = \kappa 2^{-(J-l-1)/2}\sqrt{2\log(T)}. \tag{2.69}$$

The constant κ is selected by the user.

The inverse Haar wavelet transform is applied to the coefficients $\mu_{l,m}$ to obtain the smoothed estimate, giving

$$\hat{R}_j\left(\frac{k}{T}\right) = \phi_0\mu_0 + \sum_{l=1}^{J}\sum_{m=1}^{2^{J-l}} \psi_l\left(\frac{m-k}{T}\right)\mu_{l,m}, \tag{2.70}$$

and to obtain the evolutionary wavelet spectrum we correct by the inverse inner product matrix

$$\hat{S}_j\left(\frac{k}{T}\right) = \sum_l A^{-1}_{j,l}\hat{R}_j\left(\frac{k}{T}\right). \tag{2.71}$$

This method works well in practice and is computationally efficient. However, similarly to TI de-noising there are limitations in the structure of the EWS which it can successfully recover. Choosing the right method to estimate the EWS depends on prior knowledge of the EWS, which may not be immediately obvious.

Chapter 3

Foundation Work

To investigate methods of estimating the evolutionary wavelet spectrum (EWS) with associated confidence intervals, we need to understand the statistical properties of the data of interest. As discussed in section 2.6.2, an estimate of the EWS is obtained by smoothing the raw wavelet periodogram (see definition 2.6.9), then correcting bias with the inner product matrix (definition 2.6.7). However, as the raw wavelet periodogram is the squared modulus of the empirical wavelet coefficients (definition 2.6.8), we theorised that by researching the properties of these wavelet coefficients, the findings could determine equivalent properties of the raw wavelet periodogram.

In the first section of this chapter we shall investigate the asymptotic distribution, mean and covariance structure of the empirical wavelet coefficients. In the following section we use these findings to develop estimates for the asymptotic distribution, mean and covariance of the raw wavelet periodogram. Section 3.3 will consider the potential asymptotic independence between wavelet coefficients, and discusses how this will aid our research.

Finally, in section 3.4 we numerically verify our analytic findings by simulating LSW processes with three types of innovations (Gaussian, Student's t and χ^2). We also assess the performance of the translation-invariant method of Nason et al. (2000) for recovering the true EWS using one realisation of each EWS with Gaussian innovations.

3.1 Wavelet Coefficients of the LSW Process

The empirical wavelet coefficients are obtained from the non-decimated wavelet transform of a time series. The same observation will be used to calculate coefficients at different scales and locations, which results in a covariance structure between the scales and lo-

cations. Hence, the empirical wavelet coefficients can be modelled with a multivariate distribution.

Theorem 3.1.1 (Asymptotic Statistical Properties of the Wavelet Coefficients)**.** *The distribution of the empirical wavelet coefficients of the LSW process is asymptotically multivariate Gaussian with probability density function*

$$f_d(d_{1,1},\ldots,d_{J,T}) = \frac{1}{(2\pi)^{JT/2}\,|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}\mathbf{d}'\Sigma^{-1}\mathbf{d}\right\}, \tag{3.1}$$

where $\mathbf{d} = d_{0,0},\ldots,d_{J-1,T-1}$ *and* $j = 0,\ldots,J-1$, *and* $k = 0,\ldots,T-1$. *The mean is* $\mathbb{E}[d_{j,k}] = 0\ \forall\ j,k$, *and the covariance matrix is* Σ. *Each element of the covariance matrix can be calculated from*

$$Cov[d_{j,k}, d_{j',k'}] = \sum_{l=0}^{\infty}\sum_{m=-\infty}^{\infty} \Psi_{j,l}(k-m)\,\Psi_{j',l}(k'-m)\,S_l\Big(\frac{m}{T}\Big) \;+\; \mathcal{O}(T^{-1}), \tag{3.2}$$

for $j,\ j' = 0,\ldots,J-1$, *and* $k,\ k' = 0,\ldots,T-1$.

Brillinger (1994) proves the wavelet coefficients of an additive model (such as equation (2.30)), where the error distribution is stationary but not necessarily Gaussian, are asymptotically Gaussian as the support of the wavelet increases. In Neumann and von Sachs (1995) this proof is extended to include a non-stationary, non-Gaussian error term. To prove the non-decimated wavelet coefficients of a LSW process is asymptotically Gaussian we use similar methods to Brillinger (1994) and Neumann and von Sachs (1995), but making the required adjustments for the LSW process.

Proof (Theorem 3.1.1)**.** Recall the LSW process from (2.53). For a fixed sample size $T = 2^J$, the LSW process $\{X_t\}_{t=0}^{T-1}$ is defined as

$$X_t = \sum_{l=0}^{\infty}\sum_{m=-\infty}^{\infty} w_{l,m}\,\tilde{\psi}_{l,m-t}\,\xi_{l,m},$$

with $\mathbb{E}[\xi_{l,m}] = 0$ and $\mathrm{Cov}[\xi_{l,m}, \xi_{l',m'}] = \delta_{l,l'}\ \delta_{m,m'}$, therefore $\mathbb{E}[X_t] = 0$.

To prove the asymptotic distribution of the wavelet coefficients, consider X_t for $t = 0,\ldots,T-1$, and assume:

(1) X_t has been observed at equally spaced intervals;

(2) $\mathrm{Var}[X_t] \in (0,\infty)$;

(3) ϕ and ψ (wavelet and scaling function) are of bounded total variation on $[0,1]$.

(4) $\mathbb{E}[|X_t|^p] \leq C^p(p!)^\kappa \ \forall\ p \geq 2$ holds for appropriate constant $C < \infty$ and $\kappa \geq 0$;

(5) X_t is α-mixing with $\alpha(s) \leq Ce^{-b|s|}$.

Many text book distributions satisfy assumption (4) with a suitable choice of κ. For the exponential, gamma, inverse-Gaussian and F-distribution $\kappa = 1$ (Johnson et al., 1994, 1995). By Theorem 3, part 2 of Statulevicius and Jakimavicius (1988),

$$\sup_{0\leq t_1\leq\infty}\left\{\sum_{t_1,\ldots,t_p=1}|\mathrm{cum}(X_{t_1},\ldots,X_{t_p})|\right\}\leq \tilde{C}^p(p!)^{3+\kappa}$$

holds for all $p = 2, 3, \ldots$ and appropriate $\tilde{C}$ (see Neumann (1996, Remark 3.1)). Let

$$\tilde{d}_{j,k} = \sum_t u_{j,k}(t)X_t,$$

where $u_{j,k}(t) = \int_{t_{i-1}}^{t_i}\tilde{\psi}_{j,k}(x)\,\mathrm{d}x$. From assumption (3) we have $\int\tilde{\psi}_{j,k}(x)\mathrm{d}x = \mathcal{O}(2^{j/2})$, which implies $u_{j,k}(t) = \mathcal{O}(T^{-1}2^{j/2})$. Then, we can obtain

$$\begin{aligned}\mathrm{cum}_p(d_{j,k}) &= \mathrm{cum}_p\left(\sum_t u_{j,k}(t)X_t\right)\\ &= \sum_{t_1}\sum_{t_2,\ldots,t_p} u_{j,k}(t_1)\ldots u_{j,k}(t_p)\mathrm{cum}(X_{t_1},\ldots,X_{t_p})\\ &= \mathcal{O}\Big(T^{-1}(T^{-1}2^{j/2})^{p-2}\tilde{C}^p(p!)^{3+\kappa}\Big),\end{aligned}$$

holds uniformly in $p \geq 2$ as $T \to \infty$ and $j \to 0$ (Neumann and von Sachs, 1995, equation preceding 20).

Let $\sigma^2_{j,k} = \mathrm{Var}(\tilde{d}_{j,k})$. Now consider two cases:

- $\sigma^2_{j,k} \geq CT^{-1}$, for some constant $C > 0$. Then we have

$$\left|\mathrm{cum}_p\Big(\sigma^{-1}_{j,k}d_{j,k}\Big)\right| = \mathcal{O}\Big((T^{-1}2^{j/2})^{(p-2)/2}\tilde{C}^p(p!)^{3+\kappa}\Big),$$

  uniformly in $p \geq 2$. Then by Lemma 1 from Rudzkis et al. (1978) we have

$$\mathbb{P}\Big(\pm\sigma^{-1}_{j,k}(\tilde{d}_{j,k} - \mathbb{E}[d_{j,k}]) \geq x\Big) = (1 - \Phi(x))(1 + o(1)) \tag{3.3}$$

  holds uniformly for $j = 0, \ldots J-1$ and $k = 0, \ldots, 2^j - 1$, and $-\infty < x \leq \Delta_T,\ \Delta_T = T^\mu$ for $\mu > 0$.

- $\sigma_{j,k}^2 < T^{-1/2}$, consider

$$\tilde{\tilde{d}}_{j,k} = \tilde{d}_{j,k} + \varsigma_{j,k}$$

where $\varsigma_{j,k} \sim \mathcal{N}(0, T^{-1} - \sigma_{j,k}^2)$ is independent of $\tilde{d}_{j,k}$. The random variable $\tilde{\tilde{d}}_{j,k}$ possess a variance of T^{-1} and its higher order cumulants coincide with the cumulants of $\tilde{d}_{j,k}$. Therefore (3.3) for the case of $\tilde{\tilde{d}}_{j,k}$ can be established. Hence, asymptotic normality has been derived.

From equations (2.53) and (2.60), we have the form of the LSW process and the corresponding empirical wavelet coefficients. Therefore the wavelet coefficients obtained from the LSW process can be written as

$$\begin{aligned} d_{j,k} &= \sum_{t=-\infty}^{\infty} \tilde{\psi}_{j,k-t} \sum_{l=0}^{\infty} \sum_{m=-\infty}^{\infty} w_{l,m}\, \tilde{\psi}_{l,m-t}\, \xi_{l,m} \\ &= \sum_{l=0}^{\infty} \sum_{m=-\infty}^{\infty} \Psi_{j,l}(k-m)\, w_{l,m}\, \xi_{l,m}, \end{aligned} \tag{3.4}$$

where $\Psi_{j,l}(k-m)$ is the cross autocorrelation wavelet from definition 2.6.5 between scales j and l. Therefore

$$\mathbb{E}[d_{j,k}] \;=\; \sum_{l=0}^{\infty} \sum_{m=-\infty}^{\infty} \Psi_{j,l}(k-m)\, w_{l,m}\, \mathbb{E}[\xi_{l,m}]$$

As ξ is the orthonormal increments sequence from definition 2.6.1, then property (a) defines $\mathbb{E}[\xi_{l,m}] = 0$. Therefore, we can conclude that $\mathbb{E}[d_{j,k}] = 0$.

The covariance between scales and locations can be expressed as

$$\mathrm{Cov}[d_{j,k}, d_{j',k'}] \;=\; \sum_{l,\,p=0}^{\infty} \sum_{m,q=-\infty}^{\infty} \Psi_{j,l}(k-m)\, \Psi_{j',p}(k'-q)\, w_{l,m}\, w_{p,q}\, \mathbb{E}[\xi_{l,m}\, \xi_{p,q}].$$

Using definition 2.6.1, property (b) states $\mathbb{E}[\xi_{l,m}\, \xi_{p,q}] = \delta_{l,p}\, \delta_{m,q}$ and from (c)ii. $w_{l,m} = W_l(m/T) + \mathcal{O}(T^{-1})$. Therefore

$$\begin{aligned} &\mathrm{Cov}[d_{j,k}, d_{j',k'}] \\ &= \sum_{l=0}^{\infty} \sum_{m=-\infty}^{\infty} \Psi_{j,l}(k-m)\, \Psi_{j',l}(k'-m) \Big[W_l\Big(\frac{m}{T}\Big) + \mathcal{O}(T^{-1}) \Big]^2 \\ &= \sum_{l=0}^{\infty} \sum_{m=-\infty}^{\infty} \Psi_{j,l}(k-m)\, \Psi_{j',l}(k'-m) \Big[W_l\Big(\frac{m}{T}\Big) \Big]^2 \;+\; \mathcal{O}(T^{-1}). \end{aligned}$$

From section 2.6.1 $|W_l(^m/_T)|^2 = S_l(^m/_T)$, so as the covariance of the wavelet coefficients is

$$\mathrm{Cov}[d_{j,k}, d_{j',k'}] \;=\; \sum_{l=0}^{J-1} \sum_{m=0}^{2^l-1} \Psi_{j,l}(k-m)\, \Psi_{j',l}(k'-m)\, S_l\Big(\frac{m}{T}\Big) \;+\; \mathcal{O}(T^{-1}).$$

□

The dimension of the covariance matrix Σ is $JT \times JT$, because potentially, there could be covariance between every single observation of the raw wavelet periodogram. However, in practice often there is little or no covariance between observations which are far apart, as shown in 3.4.

3.2 Raw Wavelet Periodogram

The raw wavelet periodogram is obtained by squaring the empirical wavelet coefficients. By assuming the innovations which generate the LSW process are Gaussian, Theorem 3.1.1 holds for all scales. Using this assumption and results from section 3.1 we have developed the following theorem.

Theorem 3.2.1 (Statistical Properties of the Raw Wavelet Periodogram)**.** *Assume X_t is the LSW process in equation* (2.53) *with innovations $\xi \sim \mathcal{N}(0,1)$. Then, the probability density function of the raw wavelet periodogram is of the form*

$$f_I(I_{1,1}, \ldots, I_{J,T}) = \frac{\prod_{j=0}^{J-1} \prod_{k=0}^{T-1} I_{j,k}^{-1/2}}{(2\pi)^{JT/2}\, |\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}\mathbf{I}^{1/2\prime}\Sigma^{-1}\mathbf{I}^{1/2}\right\}, \tag{3.5}$$

where $\mathbf{I}^{1/2} = (I_{0,0}^{1/2}, \ldots, I_{J-1,T-1}^{1/2})$, and Σ is the wavelet coefficient covariance matrix defined in theorem 3.1.1. The expectation of the raw wavelet periodogram multivariate distribution is

$$\mathbb{E}[I_{j,k}] = \sum_{l=0}^{\infty} \sum_{m=-\infty}^{\infty} \Psi_{j,l}^2(k-m)\, S_l\Big(\frac{m}{T}\Big) \;+\; \mathcal{O}(T^{-1}), \tag{3.6}$$

with a covariance given by

$$Cov[I_{j,k}\, I_{j',k'}] = 2\left(\sum_{l=0}^{\infty} \sum_{m=-\infty}^{\infty} \Psi_{j,l}(k-m)\, \Psi_{j',l}(k'-m)\, S_l\Big(\frac{m}{T}\Big)\right)^2 \;+\; \mathcal{O}(T^{-1}) \tag{3.7}$$

for $j, j' = 0, \ldots, J-1$ and $k, k' = 0, \ldots, T-1$.

To prove theorem 3.2.1 we will use standard probability theory from Ross (2010, Chapter 6)[1].

Lemma 3.2.2. *If a set of random variables* $\mathbf{X} = X_1, \ldots, X_n$ *each have a zero mean, with a symmetric density about its mean, the joint distribution of* $\mathbf{X}^2 = X_1^2, \ldots, X_n^2$ *can be written as*

$$f_{\mathbf{X}^2}(x_1^2, \ldots, x_n^2) = \frac{f_{\mathbf{X}}(|x_1|, \ldots, |x_n|)}{|x_1||x_2| \ldots |x_n|}.$$

Proof (Theorem 3.2.1)**.** As we have assumed that the distribution of the wavelet coefficients, $d_{j,k}$, are asymptotically multivariate Gaussian with zero mean, we can apply Lemma 3.2.2. Therefore, we find

$$\begin{aligned} f_I(I_{1,1}, \ldots, I_{J,T}) = & \frac{f_d(I_{1,1}^{1/2}, \ldots, I_{J,T}^{1/2})}{\prod_{j=0}^{J-1} \prod_{k=0}^{T-1} I_{j,k}^{1/2}} \\ = & \frac{\prod_{j=0}^{J-1} \prod_{k=0}^{T-1} I_{j,k}^{-1/2}}{(2\pi)^{JT/2} |\Sigma|^{1/2}} \exp\left\{ -\frac{1}{2} \mathbf{I}^{1/2\prime} \Sigma^{-1} \mathbf{I}^{1/2} \right\}, \end{aligned}$$

for scales $j = 0, \ldots, J-1$ and locations $k = 0, \ldots, T-1$, where Σ is the covariance matrix of the wavelet coefficients defined in theorem 3.1.1.

To determine the mean and covariance of the raw wavelet periodogram distribution, assume the the second and fourth moments of the multivariate Gaussian distribution exist and are finite. Let $\mathbf{X} \sim \mathcal{N}_J(\mathbf{0}, \Sigma)$, where Σ is the covariance matrix of $\mathbf{X}$, then from Isserlis's Theorem (Isserlis, 1918)

- $\mathbb{E}[X_i] = 0$
- $\text{Var}[X_i] = \mathbb{E}\left[X_i^2\right] = \sigma_{i,i}$
- $\mathbb{E}\left[X_i^4\right] = 3\sigma_{i,i}^4$
- $\mathbb{E}\left[X_i^2 X_{i'}^2\right] = \sigma_{i,i}\sigma_{i',i'} + 2\sigma_{i,i'}^2$,

where $\sigma_{i,i'}$ is the $(i,i')^{\text{th}}$ element of the covariance matrix Σ. Therefore if $Y_i = X_i^2$, we can

[1] Thanks to Rebecca Killick for bringing this result to my attention

calculate

$$\mathbb{E}\left[Y_i\right] = \mathbb{E}\left[X_i^2\right] = \sigma_{i,i}^2$$
$$\mathrm{Var}\left[Y_i\right] = \mathbb{E}\left[X_i^4\right] - \left(\mathbb{E}\left[X_i^2\right]\right) = 3\sigma_{i,i}^2 - \sigma_{i,i}^2 = 2\sigma_{i,i}^2$$
$$\mathrm{Cov}\left[Y_i, Y_{i'}\right] = \mathbb{E}\left[X_i^2 X_{i'}^2\right] - \mathbb{E}\left[X_i^2\right]\mathbb{E}\left[X_{i'}^2\right] = \sigma_{i,i}\sigma_{i',i'} + 2\sigma_{i,i'}^2 - \sigma_{i,i}\sigma_{i',i'} = 2\sigma_{i,i'}^2.$$

With Theorem 3.1.1, this implies that mean of $\mathbf{Y}$ is $\mu_y = \mathrm{diag}(\Sigma)$, and if we define V as the covariance matrix of $\mathbf{Y}$, then $v_{i,i'} = 2\sigma_{i,i'}^2$.

Let $X_i = d_{j,k}$, where $i = (j,k)$ and $i' = (j',k')$ and let $\sigma_{i,i} = \mathrm{Var}[d_{j,k}]$ and $\sigma_{i,i'} = \mathrm{Cov}[d_{j,k}, d_{j',k'}]$. This implies that

$$\mathbb{E}[I_{j,k}] = \sum_{l=0}^{\infty}\sum_{m=-\infty}^{\infty} \Psi_{j,l}^2(k-m)\, S_l\Big(\frac{m}{T}\Big) \;+\; \mathcal{O}(T^{-1}),$$

and

$$\begin{aligned}\mathrm{Cov}[I_{j,k}\, I_{j',k'}] \;&=\; 2\left(\sum_{l=0}^{\infty}\sum_{m=-\infty}^{\infty} \Psi_{j,l}(k-m)\,\Psi_{j',l}(k'-m)\, S_l\Big(\frac{m}{T}\Big) \;+\; \mathcal{O}(T^{-1})\right)^2 \\ &=\; 2\left(\sum_{l=0}^{\infty}\sum_{m=-\infty}^{\infty} \Psi_{j,l}(k-m)\,\Psi_{j',l}(k'-m)\, S_l\Big(\frac{m}{T}\Big)\right)^2 \;+\; \mathcal{O}(T^{-1}).\end{aligned}$$

□

Note. In Nason et al. (2000, page 24) the expectation of the raw wavelet periodogram was calculated as

$$\begin{aligned}\mathbb{E}\left[I_{j,k}\right] =&\; \sum_{l=0}^{J-1}\sum_{n=-\infty}^{\infty} \Psi_{j,l}^2(n)\,\left[S_l\Big(\frac{k}{T}\Big) + \mathcal{O}\left(nT^{-1}\right)\right] +\; \mathcal{O}(T^{-1}) \\ =&\; \sum_{l=0}^{J-1} A_{j,l}\, S_l\Big(\frac{k}{T}\Big) + \mathcal{O}(T^{-1}).\end{aligned}$$

as $A_{j,l} = \sum_n \Psi_{j,l}^2(n)$ (Fryzlewicz and Nason, 2006). Similarly, for the variance

$$\begin{aligned}\mathrm{Var}\left[I_{j,k}\right] =&\; 2\left(\sum_{l=0}^{J-1}\sum_{n=-\infty}^{\infty} \Psi_{j,l}^2(n)\,\left[S_l\Big(\frac{k}{T}\Big) + \mathcal{O}\left(nT^{-1}\right)\right] +\; \mathcal{O}(T^{-1})\right)^2 \\ =&\; 2\left(\sum_{l=0}^{J-1} A_{j,l}\, S_l\Big(\frac{k}{T}\Big)\right)^2 + \mathcal{O}\Big(\frac{2^j}{T}\Big).\end{aligned}$$

These are equivilent to our representations of the mean and variance in (3.6) and (3.7) with the substitution of the inner product matrix A.

3.3 Asymptotic Independence

Let $\beta_{l,m}^{j}$ be the DWT of the raw wavelet periodogram, calculated by

$$\beta_{l,m}^{j} = \sum_{k=0}^{T-1} \psi_{l,m-k} \, I_{j,k}. \tag{3.8}$$

By utilising rescaled time, this wavelet coefficient can be represented as $\beta_l(z) = \beta_{l,[zT]}$, where $z = mT^{-1}$. Take the wavelet coefficient at two different rescaled time points:

$$\begin{aligned} z_1 = \frac{m_1}{T_1} \qquad &\text{and} \qquad z_2 = \frac{m_2}{T_1}, \\ \Rightarrow \quad m_1 = z_1 T_1 \qquad &\text{and} \qquad m_2 = z_2 T_1, \end{aligned}$$

for z_1 and $z_2 \in [0, 1]$, where T_1 is the sample length and $J_1 = \log_2(T_1)$. Suppose another sample is taken of length $T_2 = 2^r T_1$ with $r \geq 1$ and $J_2 = \log(T_2) = J_1 + r$, then the sample size of T_2 is at least twice as big as T_1. The two locations now become

$$\begin{aligned} m_1^* = z_1 T_2 \qquad &\text{and} \qquad m_2^* = z_2 T_2 \\ \Rightarrow \quad m_1^* = z_1 2^r T_1 \qquad &\text{and} \qquad m_2^* = z_2 2^r T_1, \end{aligned}$$

Note that the distance between locations m_1^* and m_2^* is 2^r times the distance between m_1 and m_2.

For a fixed scale, eventually for compact wavelets, there will be a sample size, T_2 large enough such that m_1^* and m_2^* never lie under the same wavelet. For example let $l \leq J_1$, then two wavelet coefficients at l, $\beta_l(z_1)$ and $\beta_l(z_2)$, are asymptotically independent as $T_2 \to \infty$. Similarly, two wavelet coefficients at different scales, $\beta_l(z_1)$ and $\beta_{l'}(z_2)$, where $l \neq l' \leq J_1$, are also asymptotically independent as $T_2 \to \infty$.

This is a simple demonstration of how using the DWT has resulting wavelet coefficients which are not as strongly affected by the covariance structure of the original data.

3.4 Numerical Investigations

To validate the analytical findings in the previous sections we defined two different evolutionary wavelet spectra's (EWS) and simulated LSW processes of length $T = 2^{10} = 1024$ with these pre-defined EWS and innovations from three different distributions. To do this in `R` we used functions from the `wavethresh` package (Nason, 2012). The innovation

distributions were:

(i) Gaussian distribution with mean zero and variance one, $\mathcal{N}(0,1)$;

(ii) Student's t distribution with five degrees of freedom;

(iii) chi-square distribution with one degree of freedom.

The innovations from Student's t and chi-squared distribution were standardised to conform with the LSW process stipulations (i.e mean $= 0$ and variance $= 1$), for the rest of this section we will simply refer to these standardised distributions as t_5 and χ^2_1, respectively.

One thousand LSW processes with zero trend were simulated for each of the Gaussian, t_5 and χ^2_1 distributions. For each process, the empirical wavelet coefficients and corresponding raw wavelet periodogram were calculated. We compared the analytic and numerical mean of the empirical wavelet coefficients and raw wavelet periodogram at three different scales ($j = 9, 6, 0$) and locations ($k = 150, 490, 770$). The scales were selected because they represented the finest, a mid and the coarsest scale, and each location possessed a very different value of the EWS (low, medium and high). As the χ^2_1 distribution was skewed, we suspected the biggest difference between the numerical and analytic means would be for the χ^2_1 innovations.

Due to theorem 3.1.1, we expect to see the biggest difference between the numerical and analytical covariance estimates of the raw wavelet periodogram at the finest scales because the small support of the wavelet would mean these observations are influenced the most by the innovations. Although the innovations were standardised, the fourth moments of these distributions would vary. For example, if we have observations from a Student's t_ν distribution the squared observations will posses a F distribution with 1 and ν degrees of freedom. If $\nu = 5$ (as in our simulations) and the observations are standardised (i.e. $\text{Var}[X] = 1$), then the variance of the squared values is 8. Whereas the squared values of observations simulated from $\mathcal{N}(0,1)$ will have a χ^2_1 distribution, with a variance of 2. Therefore, we suspected the variance of the t_5 periodogram will be approximately four times bigger than the Gaussian.

In section 2.6.3, we described the translation-invariant (TI) de-noising method used by Nason et al. (2000). This is one of the most commonly used methods for estimating the EWS, and we shall use the TI de-noising method (obtained using the `ewspec` function in `R`) to produce an estimate of our defined spectra. This will give us some insight into the performance of current methods when dealing with LSW processes with non-Gaussian

innovations. It also provides a bench mark to compare our new techniques.

For these simulations, we have assumed that the synthesis and hence the analytical wavelet is known (i.e. the Haar wavelet). Therefore, all of the empirical wavelet coefficients have been obtained using the Haar wavelet. Due to the small support of the Haar wavelet the subsequent empirical wavelet coefficients should be the slowest to tend to our asymptotic findings. The only unknown wavelet will be the smoothing wavelet (SW). However, as we know the true value of the EWS we shall select the smoothing wavelet based on the result with the smallest mean squared error.

3.4.1 Piecewise Spectrum

Let $\{X_t\}_{t=0}^{T-1}$ be a simulated LSW process from the piecewise constant EWS, which is defined as

$$S_j(z) = \begin{cases} 1/2 & \text{if } j = J-1 \text{ and for } z \in \left[0, 1/3\right), \\ 1 & \text{if } j = J-1 \text{ and for } z \in \left[1/3, 8/15\right), \\ 1/4 & \text{if } j = J-1 \text{ and for } z \in \left[8/15, 1\right), \\ 0 & \text{otherwise.} \end{cases} \tag{3.9}$$

All of the power of this spectrum is present at the finest scale. The change points in the spectrum were located after a third, then a further fifth of the time series. These locations were selected because the nature of the Haar wavelet would make dyadic break points very easy to detect. A plot of this piecewise EWS can be observed in figure 3.1(a).

A simulated LSW process with Gaussian innovations, defined as X_t for $t = 0, \ldots, 2^{10} - 1 = 1023$, generated from the piecewise spectrum can be observed in figure 3.1(b). As the power is at the finest scale of the EWS, the simulated time series oscillates very quickly. From the plot of X_t, there appears to be a change in the magnitude of the oscillation at approximately $t = 300$ and 550. The difference is more noticeable when you compare the observations before and after $t \approx 550$. The true changes in spectral power were actually at $t = 340$ and 544.

The time series plot indicates that the process is at least non-stationary because of the change in variance. Various transformations, such as differencing and logarithms, were applied to the data to try to stabilise the variance without success.

The plots in figure 3.2 are histograms of the empirical wavelet coefficients obtained from LSW processes with non-Gaussian innovations. As the scale decreases and the support of the wavelet increases, there is evidence to suggest the distribution is tending to

(a) Piecewise Constant EWS

(b) Simulated Time Series (X_t)

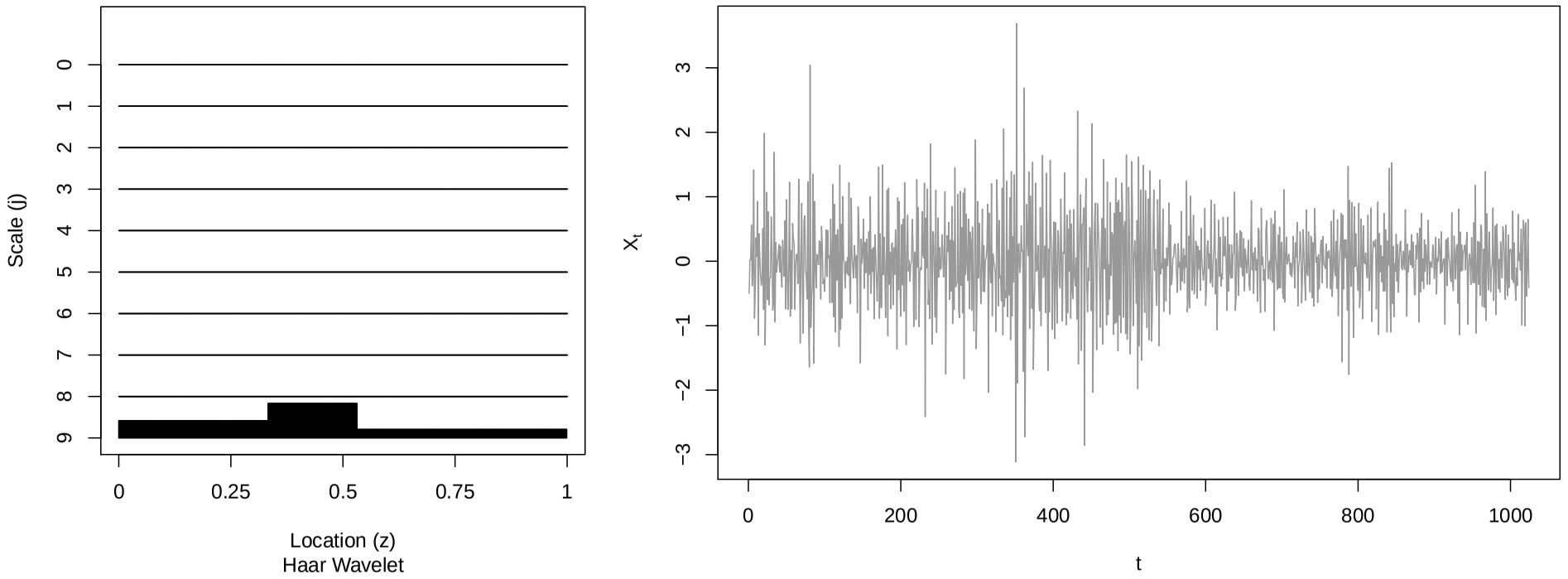


Figure 3.1: Plots of the true piecewise constant spectrum defined in equation (3.9) created using the Haar synthesis wavelet, and a realised LSW process $\{X_t\}_{t=0}^{1023}$ generated with Gaussian innovations.

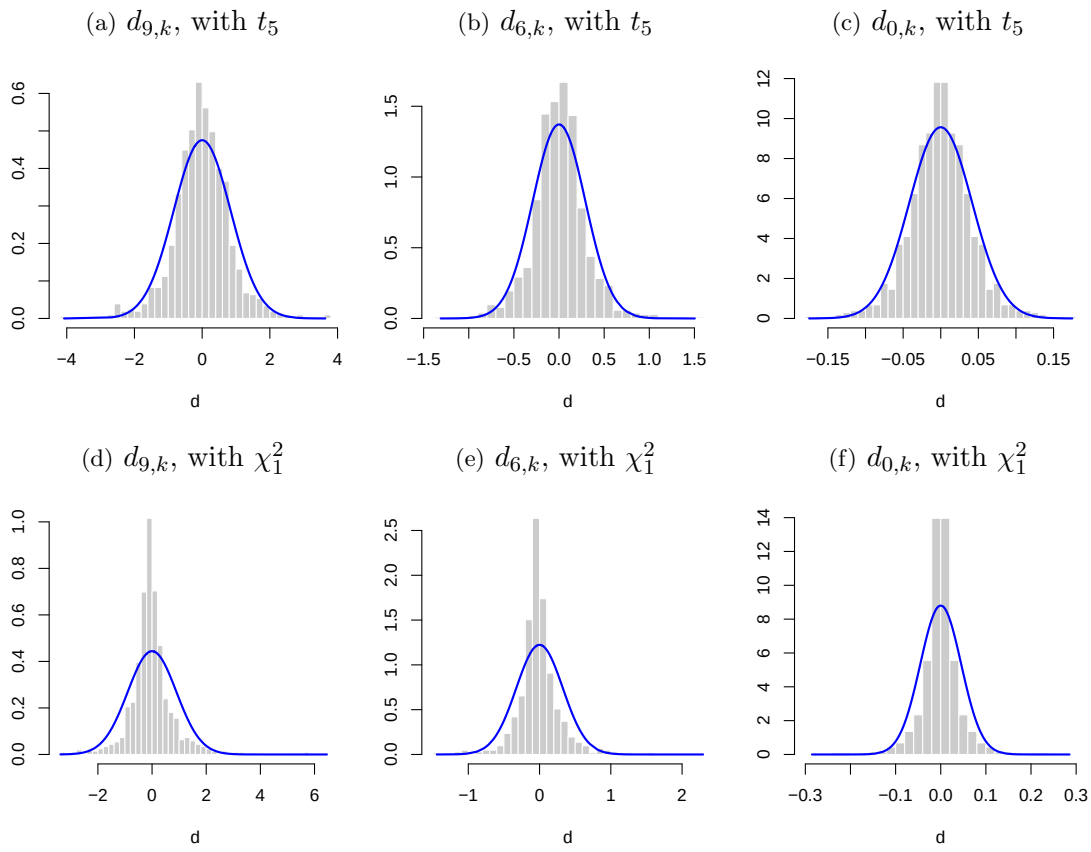


Figure 3.2: Density histograms of the empirical wavelet coefficients at scales $j = 9, 6, 0$, obtained from the piecewise LSW process (X_t) with t_5 (in a, b and c) and χ^2_1 (in d, e and f) innovations and Gaussian pdf (blue solid line).

Gaussianity. Due to the similarities between the t_5 and Gaussian distributions, the histograms of the t_5 empirical wavelet coefficients in figures 3.2 (a) to (c) are quite similar to the Gaussian pdf, even at the finer scales. In contrast, the positive skew of the chi-square distribution means the empirical wavelet coefficients in figures 3.2 (d) and (e) are also positively skewed. However, at the coarsest scale there is no empirical evidence the wavelet coefficients in 3.2(f) are skewed, but there is a higher concentration of values around zero.

Table 3.1 contains the empirical mean and standard error of the piecewise constant empirical wavelet coefficients, at the three selected scales and locations. The scales were selected to represent a fine, mid and coarse scale of the NDWT, and the locations were selected because they were positioned at the three different levels of power. All of the mean estimates were small and close to zero, and the means for the observations at the coarsest scale were the closest to zero for all of the different types of innovations. This is caused by the expectation of the innovations (ξ) being set to zero and the generation of an LSW process with no trend. The standard errors for all the different types of innovations at each scale and location were very similar, as a result of all the innovations possessing

$\mathbb{E}$ (SE)		k	
j	**150**	**490**	**770**
9	0.00986 (0.027)	-0.0292 (0.039)	-0.0157 (0.019)
	0.0051 (0.028)	*0.0051 (0.037)*	*-0.018 (0.020)*
	*0.0081 (0.029)	*0.056 (0.039)	*0.0058 (0.019)
6	0.00046 (0.0096)	-0.0026 (0.014)	-0.0012 (0.0069)
	-0.0054 (0.0092)	*-0.0098 (0.013)*	*0.012 (0.0070)*
	*0.0011 (0.010)	*0.0020 (0.014)	*-0.0017 (0.0071)
0	0.0012 (0.0012)	-0.00047 (0.0016)	0.0018 (0.0012)
	-0.00010 (0.0012)	*-0.00052 (0.0015)*	*0.00087 (0.0012)*
	0.00058 (0.0013)	*0.00039 (0.0016)	*0.00013 (0.0012)

Table 3.1: A table of the mean and standard error of the empirical wavelet coefficients for locations $k = 150, 490, 770$ and scales $j = 9, 6, 0$. The table consists of the numerically estimated mean from $\{X_t\}$ with: Gaussian, Student's t_5 (in italics) and χ^2_1 (denoted by *) innovations and their corresponding standard errors in brackets to 2 significant figures.

the same variance. Also, the standard errors all indicated the empirical mean was not significantly different from zero.

Next, we examined the distribution of the raw wavelet periodogram obtained from a simulated X_t. From theorem 3.2.1 we determined that asymptotically the distribution of the raw wavelet periodogram would be similar to a multivariate chi-squared distribution, therefore for a particular scale and location, $I_{j,k} \approx (AS)_{j,k}\chi^2_1$. The plots in figure 3.3 are histograms of the mean corrected raw wavelet periodogram defined as

$$\mathcal{I}_{j,k} = \frac{I_{j,k}}{(AS)_{j,k}}.$$

By dividing by the expected value of the raw wavelet periodogram we could compare the distribution of the raw wavelet periodogram with the χ^2_1 distribution. The histograms indicated that for coarser scales the distribution is tending to a chi-square distribution with one degree of freedom, regardless of the distribution of the innovations.

Using equations (3.6) and (3.7), we determined the analytic mean and covariance of the raw wavelet periodogram for the piecewise constant EWS, and a sample size of $T = 2^{10} = 1024$, is

$$\mathbb{E}[I_{j,k}] = \frac{1}{2}\sum_{m=0}^{340}\Psi^2_{j,1}(k-m) + \sum_{m=341}^{544}\Psi^2_{j,1}(k-m) + \frac{1}{4}\sum_{m=545}^{1023}\Psi^2_{j,1}(k-m), \tag{3.10}$$

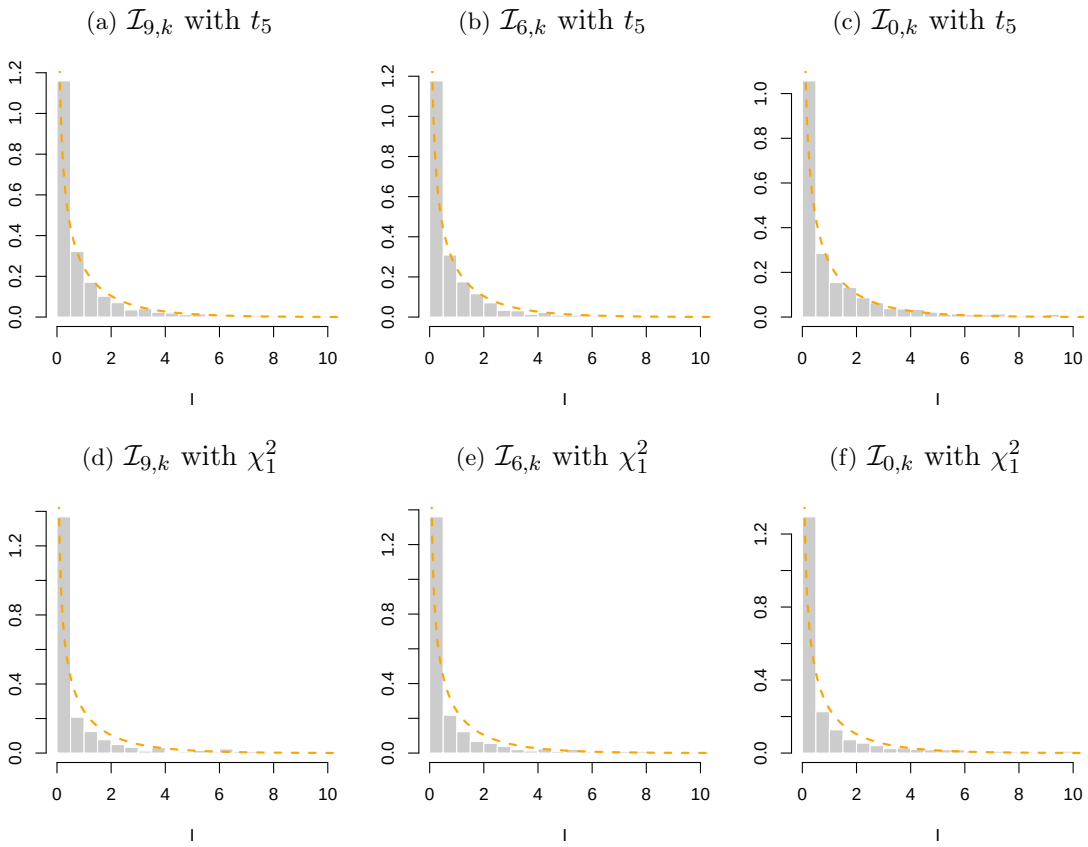


Figure 3.3: Density histograms of the mean corrected raw wavelet periodogram at scales $j = 9, 5, 0$, between 0 and 10, of the piecewise constant LSW process $\{X_t\}$, with t_5 (a, b and c) and χ^2_1 (d, e and f) innovations and chi-square pdf (orange dashed line).

$\mathbb{E}[I_{j,k}]$		k	
j	**150**	**490**	**770**
9	(0.75)	(1.5)	(0.375)
	0.729	1.50	0.357
	0.758	*1.35*	*0.394*
	*0.856	*1.50	*0.362
6	(0.0938)	(0.188)	(0.0469)
	0.0921	0.183	0.0480
	0.0853	*0.182*	*0.0486*
	*0.105)	*0.187)	*0.0508
0	(0.00146)	(0.00293)	(0.00073)
	0.00145	0.00244	0.00153
	0.00134	*0.00234*	*0.00147*
	*0.00157	*0.00246	*0.00147

Table 3.2: A table of the mean value of the raw wavelet periodogram for locations $k = 150, 490, 770$ and scales $j = 9, 6, 0$. The table consists of analytical mean (in brackets), and the numerically estimated mean from $\{X_t\}$ with: Gaussian, Student's t_5 (in italics) and chi-square (denoted by *) innovations (3 s.f.).

	$I_{9,150}$	$I_{9,490}$	$I_{6,490}$	$I_{6,770}$	$I_{0,150}$	$I_{0,770}$
$I_{9,150}$	(1.13)	(0)	(0)	(0)	(6.1×10^{-5})	(0)
	1.10	0.140	0.00485	-9.9×10^{-4}	9.9×10^{-4}	-2.8×10^{-6}
	3.53	*0.220*	*-0.00517*	*-0.00256*	*0.00138*	4.0×10^{-5}
	*4.60	*0.456	*-0.00617	*-0.00131	*0.00742	$*6.3 \times 10^{-5}$
$I_{9,490}$		(4.5)	(0.0156)	(0)	(0)	(0)
		4.29	0.206	-0.00122	2.7×10^{-4}	-2.0×10^{-4}
		5.10	*0.338*	*0.00761*	7.9×10^{-5}	-6.7×10^{-5}
		*23.1	*2.30	*-0.0178	$*8.2 \times 10^{-4}$	$*-4.4 \times 10^{-4}$
$I_{6,490}$			(0.0703)	(0)	(0)	(0)
			0.0628	-1.5×10^{-4}	-6.9×10^{-6}	1.4×10^{-6}
			0.0976	-5.0×10^{-4}	2.4×10^{-5}	-6.4×10^{-5}
			*0.319	*-0.00136)	$*-1.6 \times 10^{-5}$	$*3.6 \times 10^{-5}$
$I_{6,770}$				(0.00439)	(0)	(1.9×10^{-6})
				0.00458	-1.3×10^{-6}	3.4×10^{-5}
				0.00839	-1.0×10^{-5}	8.6×10^{-5}
				*0.0155	$*6.15 \times 10^{-7}$	$*1.8 \times 10^{-4}$
$I_{0,150}$					(4.30×10^{-6})	(7.4×10^{-42})
					4.0×10^{-6}	-2.6×10^{-7}
					4.7×10^{-6}	-2.5×10^{-7}
					$*1.9 \times 10^{-5}$	$*2.8 \times 10^{-8}$
$I_{0,770}$						(1.1×10^{-6})
						4.4×10^{-6}
						7.6×10^{-6}
						$*1.5 \times 10^{-5}$

Table 3.3: A covariance matrix of the raw wavelet periodogram. The table consists of asymptotic covariance (in brackets), and the numerically estimated covariance from $\{X_t\}$ with: Gaussian, Student's t_5 (in italics) and chi-square (denoted by *) innovations (3 s.f.).

and

$$\mathrm{Cov}[I_{j,k}\, I_{j',k'}] \;=\; 2\left(\frac{1}{2}\sum_{m=0}^{340}\Psi_{j,1}(k-m)\,\Psi_{j',1}(k'-m) \;+\; \sum_{m=341}^{544}\Psi_{j,1}(k-m)\,\Psi_{j',1}(k'-m) + \frac{1}{4}\sum_{m=545}^{1023}\Psi_{j,1}(k-m)\,\Psi_{j',1}(k'-m)\right)^2. \tag{3.11}$$

Table 3.2 shows the analytical and numerical mean of the raw wavelet periodogram. Notice that the mean is $\sqrt{1000}$ times the standard error of $d_{j,k}$ in table 3.1, which provides empirical evidence $\mathbb{E}[I_{j,k}] = \mathrm{Var}[d_{j,k}]$. We observed that the expectations of $I_{j,k}$ are quite close to the analytic values.

The estimates of the covariance at the coarsest scales in table 3.3 are the closest to the

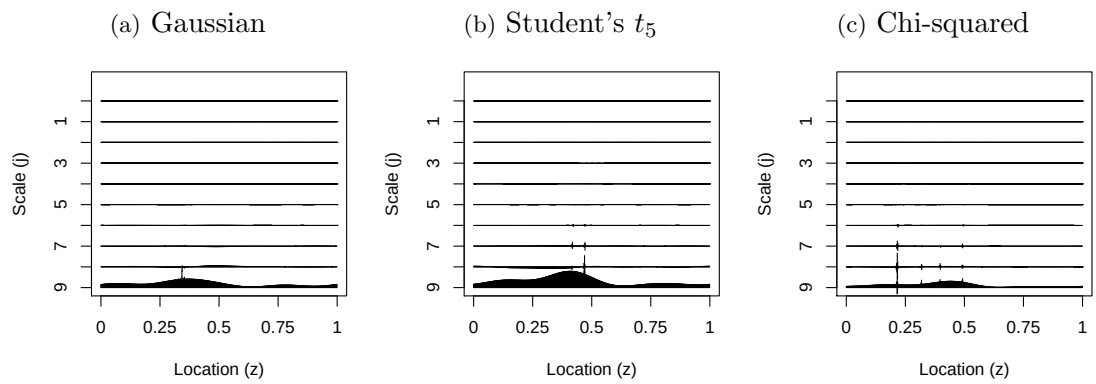


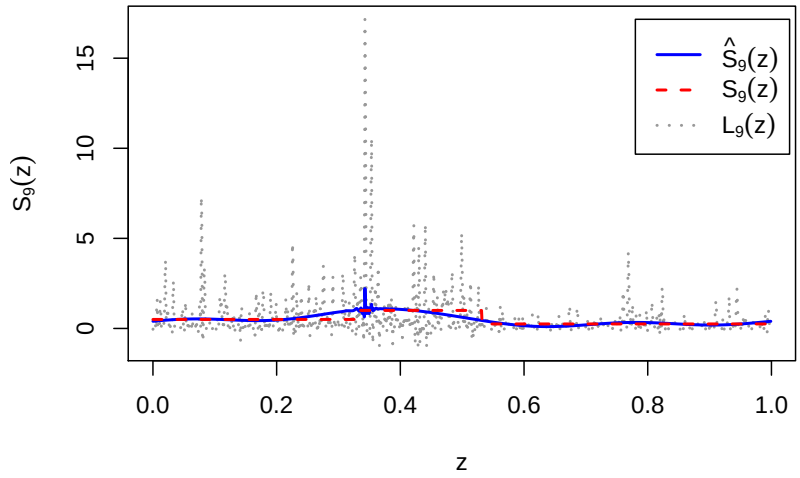
Figure 3.4: Plots of the TI de-noised estimate of the EWS from the same simulated LSW process, using the optimal SW: D_9 in (a) & (b), and D_{10} in (c), for LSW process $\{X_t\}$ simulated from the piecewise constant spectrum with Gaussian, t_5 and χ^2_1 innovations

true values. In this table we have omitted some of the combinations of the selected scales and locations for simplicity, but also because of the decay in the covariance structure, many of the values were close approximately zero. Therefore, the combinations of the covariances calculated were chosen to given a representative sample. The biggest difference can be observed at the finest scale. The variance of $I_{9,150}$ with Student's t_5 and χ^2_1 innovations, are approximately three and four times bigger than the analytic variance. As the distance between locations increase, the analytic covariance is negligible if they were far enough apart. Even though the analytic covariance for the raw wavelet periodogram at the same location but different scales is small, it is larger than different locations at the same scale. There is also a decrease in the variance for coarser scales. However, all the numerical estimates are larger than the analytic covariance.

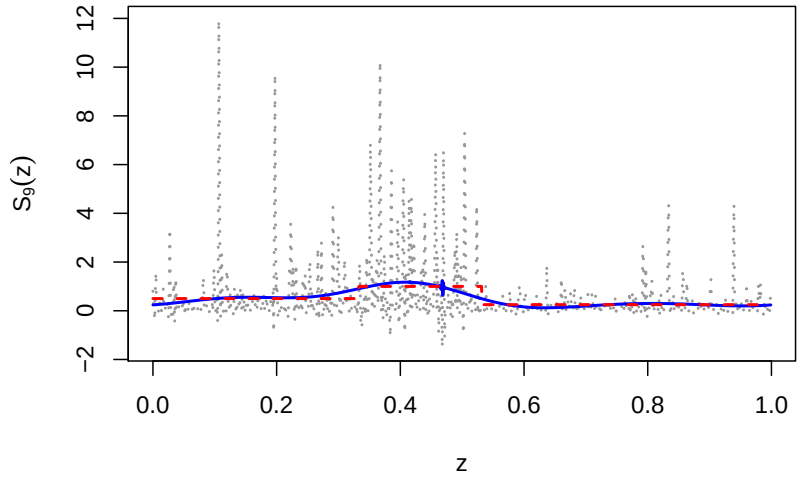
Figure 3.4 shows plots of the EWS using TI de-noising obtained from the LSW process X_t with Gaussian (a), t_5 (b) and χ^2_1 (c) innovations. Each plot is the estimate produced using the optimal smoothing wavelet. In each plot there appears to be very little leakage from the finest scale into the neighbouring coarser scale. Although, there are a few locations where there are Gibbs phenomena at approximately the same location and different scales.

For the finest scale, where all the power of the piecewise constant EWS for this example is contained, we have plotted the true EWS, $S_9(z)$, the TI de-noised estimate, $\hat{S}_9(z)$, and the corrected raw wavelet periodogram, $L_9(z)$, with Gaussian (figure 3.5(a)), t_5 (figure 3.5(b)) and χ^2_1 (figure 3.5(c)) innovations. The plot of $L_9(z)$ depicts how noisy the data is and how well the method works at recovering the signal. The negative values of $L_9(z)$ are a side effect of the leakage between scales. The Gibbs phenomena are present at large

(a) Gaussian Innovations, SW = D_9



(b) Student's t_5 Innovations, SW = D_9



(c) Chi-squared Innovations, SW = D_{10}

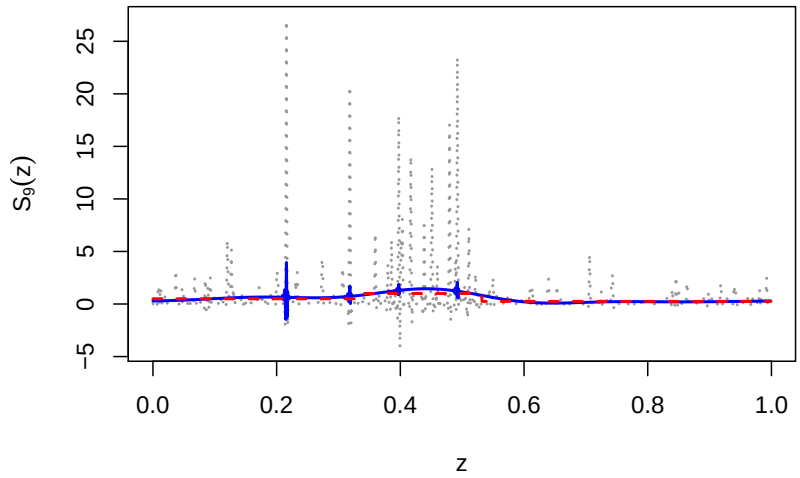


Figure 3.5: Plots of the true piecewise constant EWS ($S_9(z)$: dashed red line), TI denoised estimate ($\hat{S}_9(z)$: solid blue line) using the optimal smoothing wavelet (SW) and bias corrected raw wavelet periodogram ($L_9(z)$: grey dotted line), obtained from the LSW process $\{X_t\}$ simulated from the piecewise constant spectrum with Gaussian, t_5 and χ^2_1 innovations.

errors and negative values of $L_1(z)$. For example, Gibbs phenomena can be observed at $z \approx 1/3$ in figure 3.5(a), $z \approx 0.48$ in figure 3.5(b) and $z \approx 0.2, 0.3, 0.4$ and 0.5 in figure 3.5(c).

One problem with the TI de-noising method is its inability to detect the piecewise constant nature of the signal. This is a consequence of the cycle spinning employed to alleviate the Gibbs phenomena, which would be present at the change points where there would be a discontinuity in the wavelet transformation.

3.4.2 Slowly Evolving Spectrum

Let $\{Y_t\}_{t=0}^{T-1}$ be the simulated time series from the slowly evolving EWS, which is defined as

$$S_j(z) = \begin{cases} \sin(2\pi z) + 0.1 & \text{if } j = J-4 \text{ and for } z \in [0,1), \\ 0 & \text{otherwise.} \end{cases} \tag{3.12}$$

This spectrum has been defined with power only at the fourth finest scale. Figure 3.6(a) is a plot of the slowly evolving spectrum generated, once again with the Haar wavelet. The change in power is a lot slower than the piecewise constant EWS, and because the power is set at a coarser scale the subsequent LSW process in figure 3.6(b) does not oscillate as quickly as the X_t. Also, magnitude of the oscillation changes slowly, displaying an increase, decrease, increase then decrease in the process's variance. The plot of Y_t clearly demonstrates the non-stationary nature of the process.

Similarly to the piecewise example, the slowly evolving coefficients tend to Gaussianity as the scale decreases and the support of the wavelet increases, see the plots in figure 3.7. They behave in a similar manner, where t_5 coefficients are more concentrated around zero and the χ_1^2 coefficients are positively skewed at finer scales.

The mean and standard error of the slowly evolving coefficients for all three types of innovations are displayed in table 3.4. Again, the expectations are all close to zero with a standard error that provided evidence to suggest the means were not significantly different to zero. However, we observed the largest empirical mean was for $d_{6,770}$ with χ_1^2 innovations possessed weak evidence the mean was not zero. The increase in mean was possibly caused by the location of this coefficient near one of the highest two peaks in the EWS and the skewed error distributions. However, as the 95% CI was $(0.01, 0.31)$, we felt the weak evidence could be ignored. Similarly to the piecewise constant wavelet coefficients, we also noticed that the standard errors for the different innovations were

(a) Slowly evolving EWS

(b) Simulated Time Series (Y_t)

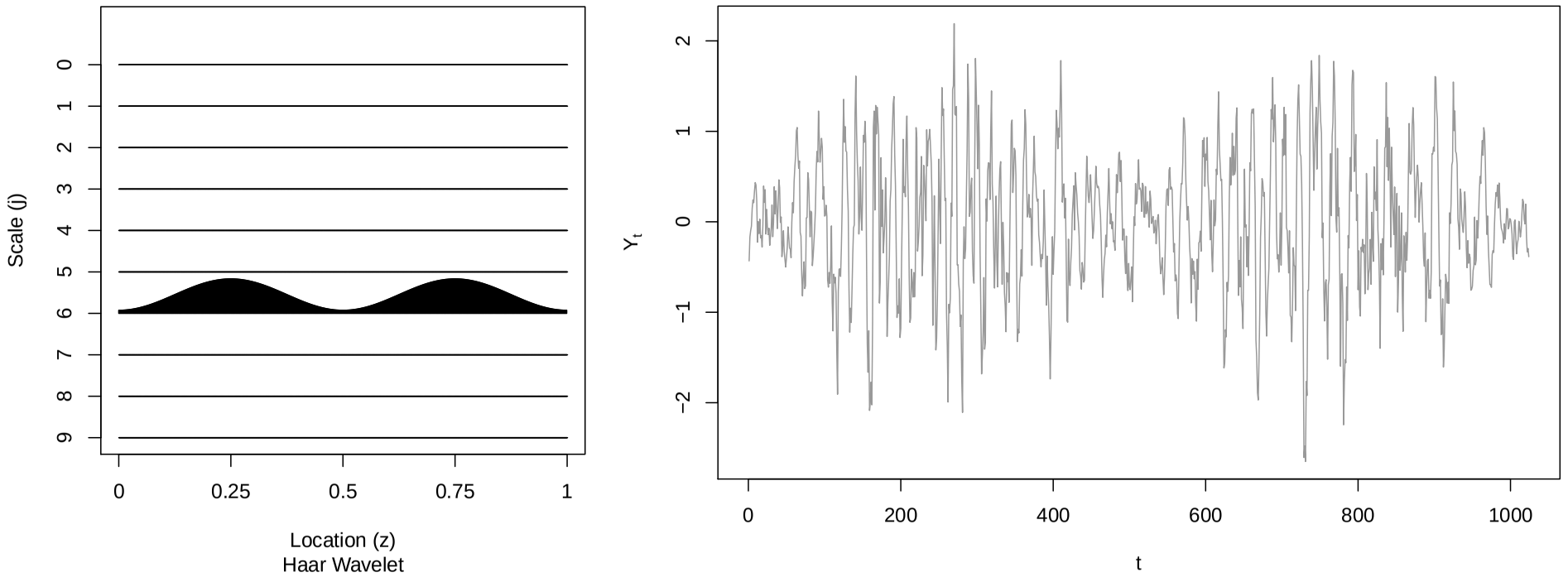


Figure 3.6: Plots of the true slowly evolving spectrum defined in equation (3.12) created using the Haar synthesis wavelet, and a realised LSW process $\{Y_t\}$ generated with Gaussian innovations.

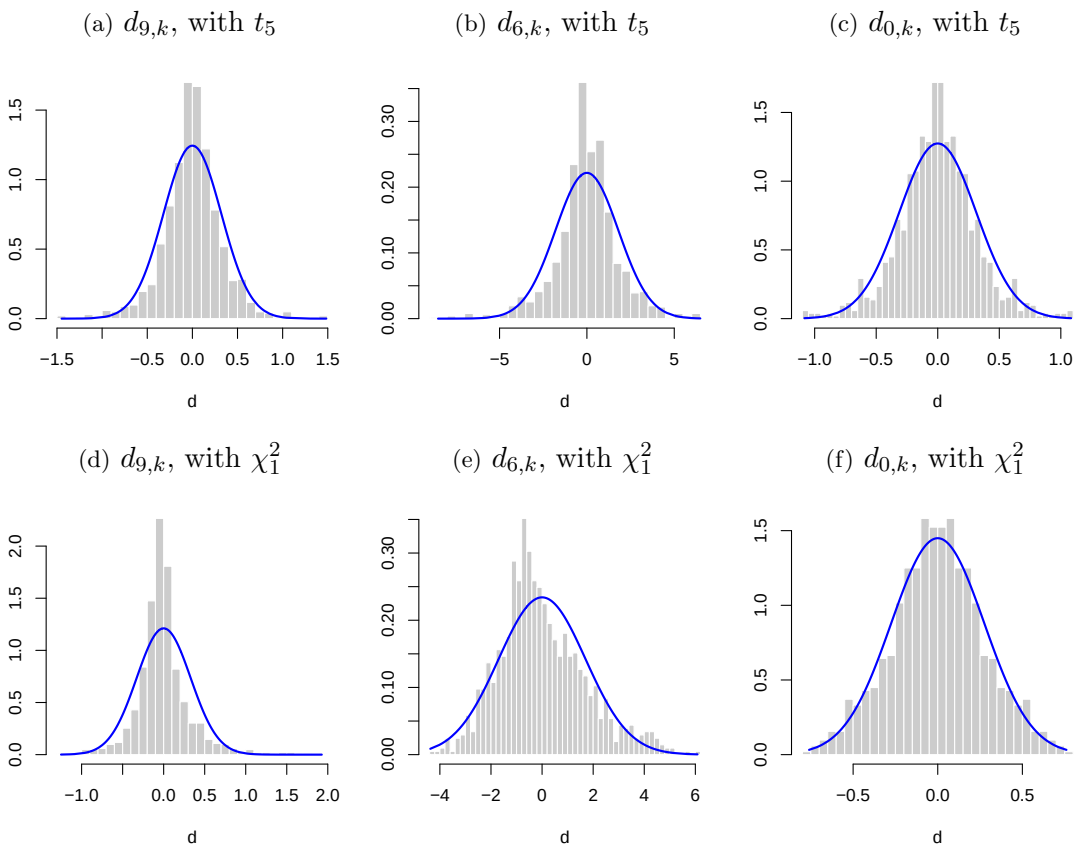


Figure 3.7: Density histograms of the empirical wavelet coefficients at scales $j = 9, 6, 0$, obtained from the slowly evolving LSW process (Y_t) with Student's t_5 (in plots a, b and c) and χ^2_1 (in plots d, e and f) innovations and Gaussian probability density plot (red dashed line).

$\mathbb{E}$ (SE)		k	
j	**150**	**490**	**770**
9	0.0076 (0.011)	-0.0048 (0.0047)	0.0018 (0.015)
	-0.0083 (0.012)	*-0.00057 (0.0048)*	*0.018 (0.014)*
	-0.0051* (0.11)	-0.0026* (0.0051)	0.022* (0.014)
6	0.015 (0.065)	0.026 (0.025)	-0.040 (0.078)
	-0.099 (0.065)	*0.042 (0.025)*	*-0.015 (0.078)*
	-0.040* (0.064)	-0.013* (0.026)	0.16* (0.076)
0	0.0041 (0.011)	0.0071 (0.0045)	-0.032 (0.014)
	-0.013 (0.011)	*-0.0067 (0.0046)*	*-0.013 (0.014)*
	0.0059* (0.011)	0.0033* (0.0045)	-0.016* (0.014)

Table 3.4: A table of the mean and standard error of the empirical wavelet coefficients for locations $k = 150, 490, 770$ and scales $j = 9, 6, 0$. The table consists of the numerically estimated mean from $\{Y_t\}$ with: Gaussian, Student's t_5 (in italics) and χ^2_1 (denoted by *) innovations and their corresponding standard error in brackets to 2 significant figures.

very similar for each scale and location.

The histogram of the mean corrected raw wavelet periodogram ($\mathcal{I}_{j,k}$) for LSW Y_t with χ^2_1 and t_5 innovations are plotted in figure 3.8. Similarly to the plots of $\mathcal{I}_{j,k}$ calculated from X_t, the distribution is tending to a χ^2_1 distribution as the the support of the wavelet increases.

For a sample size of $T = 2^{10} = 1024$, the analytic mean and covariance of the raw wavelet periodogram for the slowly evolving EWS is

$$\mathbb{E}[I_{j,k}] = \sum_{m=0}^{1023} \Psi^2_{j,4}(k-m) \left[\sin\left(\frac{2\pi m}{1024}\right) + 0.1 \right], \tag{3.13}$$

and

$$\mathrm{Cov}[I_{j,k}\, I_{j',k'}] = 2 \left(\sum_{m=0}^{1023} \Psi_{j,4}(k-m)\, \Psi_{j',4}(k'-m) \left[\sin\left(\frac{2\pi m}{1024}\right) + 0.1 \right] \right)^2. \tag{3.14}$$

Table 3.5 shows the true and empirical mean for Gaussian, t_5 and χ^2_1 innovations. As previously observed the expected value of the raw wavelet periodogram is very close to the analytical mean and the standard error of $d_{j,k}$ is approximately $\sqrt{1000}$ times smaller than $\mathrm{Var}[I_{j,k}]$. This evidence gives additional support to our findings in 3.2.

We observe similar results in the covariance matrix for the slowly evolving EWS (table 3.6) as for the piecewise constant EWS, where the covariance values tend to be higher for the non-Gaussian simulations. All of the Gaussian estimates are very close to the analytic

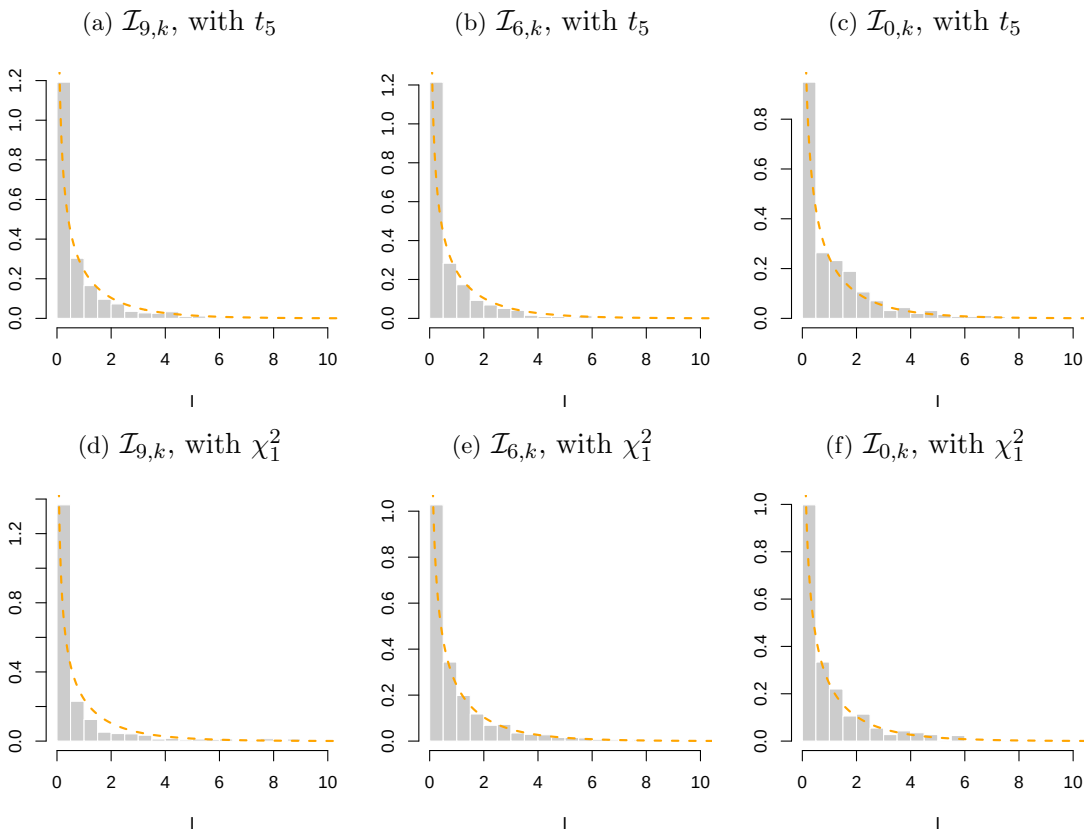


Figure 3.8: Density histograms of the mean corrected raw wavelet periodogram at scales $j = 9, 6, 0$, between 0 and 10, of the slowly evolving LSW process $\{Y_t\}$, with Student's t_5 (in a, b, and c) and χ^2_1 (in d, e and f) innovations, with χ^2_1 pdf (orange dashed line).

$I_{j,k}$	k		
j	**150**	**490**	**770**
9	(0.138)	(0.0222)	(0.206)
	0.132	0.0222	0.213
	0.143	*0.0223*	*0.193*
	*0.130	*0.0261	*0.202
6	(3.99)	(0.642)	(5.98)
	4.21	0.636	6.05
	4.24	*0.627*	*6.10*
	*4.08	*0.687	*5.84
0	(0.0924)	(0.0149)	(0.139)
	0.121	0.0203	0.189
	0.124	*0.0212*	*0.198*
	*0.123	*0.0198	*0.182

Table 3.5: A table of the mean value of the raw wavelet periodogram for locations $k = 150, 490, 770$ and scales $j = 9, 6, 0$. The table consists of analytical mean (in brackets), and the numerically estimated mean from $\{Y_t\}$ with: Gaussian, Student's t_5 (in italics) and chi-square (denoted by *) innovations (3 s.f.).

	$I_{9,150}$	$I_{9,490}$	$I_{6,490}$	$I_{6,770}$	$I_{0,150}$	$I_{0,770}$
$I_{9,150}$	(0.0378)	(0)	(0)	(0)	(2.1×10^{-6})	(1.3×10^{-43})
	0.0359	0.00006	-0.00302	-0.00302	0.00140	-0.00080
	0.196	*-0.00054*	*-0.00188*	*-0.0594*	*0.00432*	*-0.00252*
	*0.0835	*-0.00040	*-0.0169	*-0.0130	*0.00689	*0.00221
$I_{9,490}$		(0.00098)	(0.00180)	(0)	(0)	(0)
		0.00088	0.00654	0.00040	-0.00016	0.00004
		0.00247	*0.0189*	*-0.00913*	*0.00025*	*0.00038*
		*0.00942	*0.0697	*-0.00637	*0.00001	*-0.00073
$I_{6,490}$			(0.825)	(0)	(0)	(2.1×10^{-39})
			0.796	0.0744	-0.0112	-0.00432
			0.977	*-0.434*	*0.00141*	*-0.0192*
			*1.58	*-0.0294	*-0.00671	* -0.00569
$I_{6,770}$				(71.5)	(0)	(0.01770)
				71.2	0.06021	0.470
				79.5	*0.0339*	*0.840*
				*112	*-0.0353	*1.31
$I_{0,150}$					(0.171)	(9×10^{-37})
					0.0277	-0.00122
					0.0532	*0.00450*
					*0.0372	*0.00047
$I_{0,770}$						(0.0384)
						0.0764
						0.0838
						*0.0762

Table 3.6: A covariance matrix of the raw wavelet periodogram. The table consists of asymptotic covariance (in brackets), and the numerically estimated covariance from $\{Y_t\}$ with: Gaussian, Student's t_5 (in italics) and chi-square (denoted by *) innovations (3 s.f.).

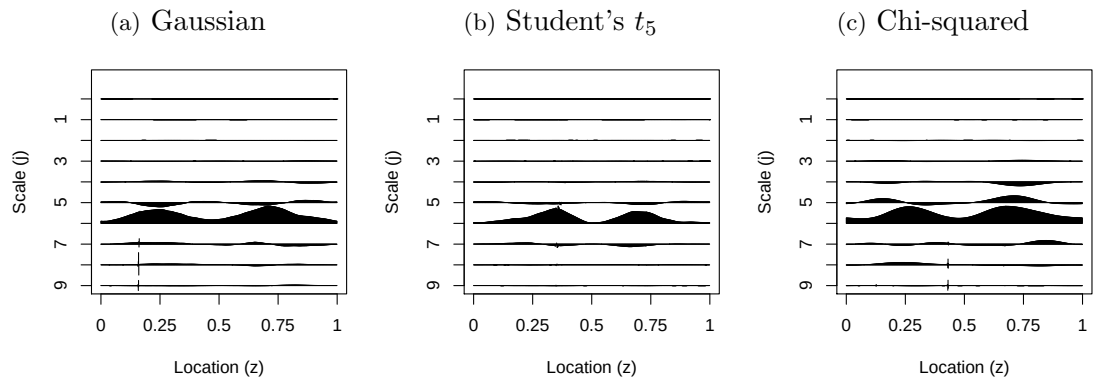


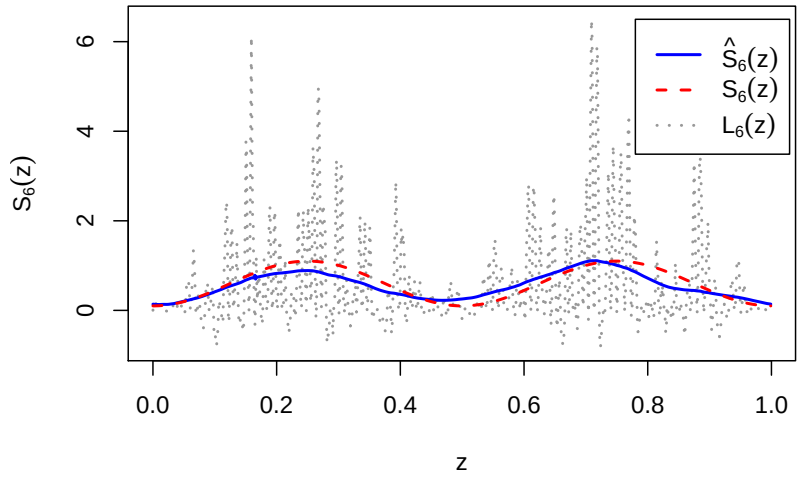
Figure 3.9: Plots of the TI de-noised estimate of the EWS using the optimal smoothing wavelet: Haar in (a) & (b) and D_4 in (c), for LSW process $\{Y_t\}$ simulated from the piecewise constant spectrum using Gaussian, t_5 and χ^2_1 innovations.

covariance except for $\mathrm{Var}[I_{0,770}]$, which is approximately double the analytic variance. This is probably caused by the leakage between scales, as this location where the power is at its greatest on scale $J - 4$.

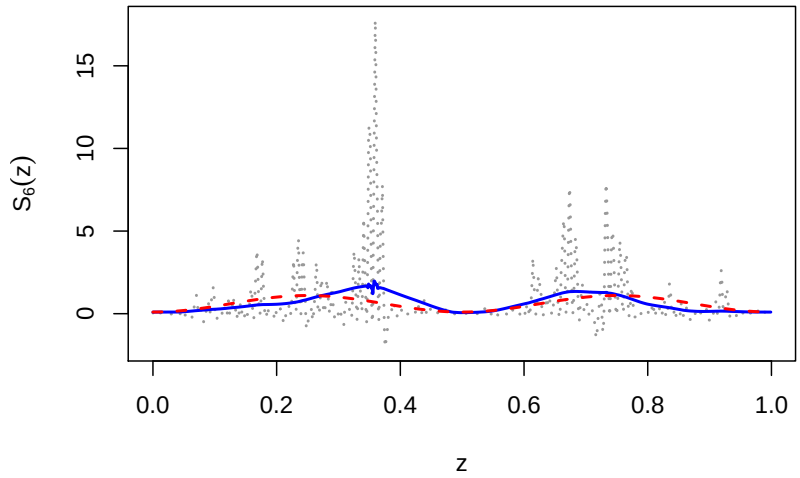
Figure 3.9 shows the plots of the TI de-noising estimates of the EWS from the LSW process Y_t with Gaussian, t_5 and χ^2_1 innovations. Each plot is the estimate produced using the optimal smoothing wavelet. These plots demonstrate how well the TI de-noising method works at recovering the spectra which are smooth. For Y_t with Gaussian or t_5, we have the same optimal smoothing wavelet. There appears to be less leakage from $l = 4$ into adjacent scales for the t_5 EWS estimate in 3.9(b), and the greatest leakage in the χ^2_1 estimate 3.9(c).

As all of the power of the EWS was present at scale $j = 4$, we plotted the bias corrected raw wavelet periodogram, TI de-noising estimator and true EWS in figure 3.10 for $j = 4$ and each type of innovation. The demonstrate how noisy the raw wavelet periodogram is and how well the TI de-noising estimator performs at recovering the true EWS. The estimates for each type of innovation perform reasonably well at recovering the true slowly evolving EWS. Only the estimator for the t_5 innovations suffers from the presence of Gibbs phenomena.

(a) Gaussian Innovations, SW = Haar



(b) Student's t_5 Innovations, SW = Haar



(c) Chi-squared Innovations, SW = D_4

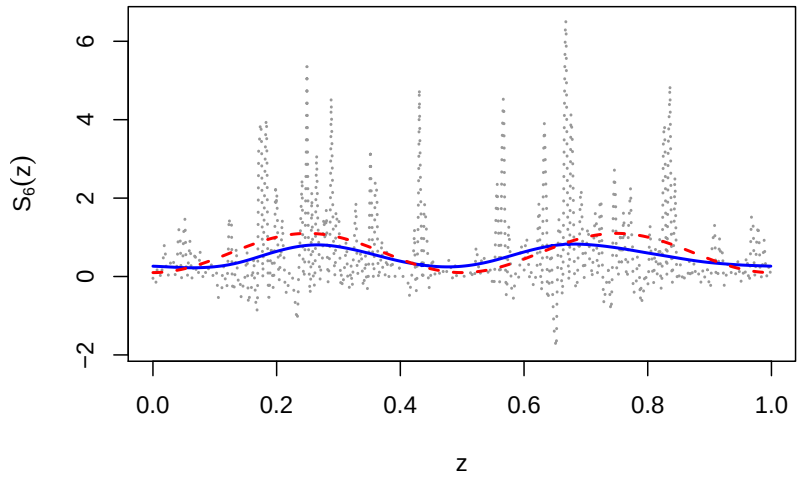


Figure 3.10: Plots of the true slowly evolving EWS ($S_6(z)$: dashed red line), TI denoised estimate ($\hat{S}_6(z)$: solid blue line) using the optimal smoothing wavelet (SW) and bias corrected raw wavelet periodogram ($L_6(z)$: grey dotted line), obtained from the LSW process $\{Y_t\}$ simulated from the slowly evolving spectrum with Gaussian, t_5 and χ^2_1 innovations.

3.4.3 Conclusion

We used empirical estimates to demonstrate theorems 3.1.1 and 3.2.1. The empirical wavelet coefficients and raw wavelet periodogram were obtained from the LSW processes using the Haar wavelet. The Haar wavelet has the smallest support of the wavelets, which led us to believe these results would produce the slowest asymptotic convergence to the distributions in theorems 3.1.1 and 3.2.1.

The results imply the distribution of the empirical wavelet coefficients, and hence the raw wavelet periodogram, at the finest scale are influenced the most by the distribution of the innovations. The histograms suggested the empirical coefficients and periodogram were tending to the asymptotic distributions by mid scales, especially if the distributions of the innovations were reasonably close to a Gaussian. However, even highly skewed distributions, such as the χ^2_1, demonstrated evidence of tending to Gaussianity even at the fourth finest scale.

The means calculated for the empirical wavelet coefficients and raw wavelet periodogram were all reasonably close to the analytic means, although the periodogram variance varied the most. We suspected the variation was caused by the differences between the fourth moments of the Gaussian, t_5 and χ^2_1 distributions. Differences were mainly evident at the finest scale, and we theorised the variance of the t_5 periodogram would be approximately four times bigger than the Gaussian periodogram. In table 3.3 $\text{Var}[I_{9,150}]$ of the t_5 periodogram is approximately 3.5 times bigger than the Gaussian periodogram, whereas in table 3.6 $\text{Var}[I_{9,150}]$ is approximately five times bigger than the t_5 periodogram

TI de-noising estimation works under the assumption of a χ^2_1 distribution, so given the differences between the distribution og the innovations and the subsequent effect this had upon the distribution of the periodogram, the smoothed estimates were quite reasonable. It appeared to cope well with the different innovation distributions, especially considering the assumed error variances at the finer scales were very different to the actual empirical error variance.

As TI de-noising estimation is a wavelet shrinkage method, our estimates in figures 3.5 and 3.10 indicates the benefits of the decimated wavelet transformation (Gaussianity and reducing the covariance) in recovering the true function from noisy data. The addition of a confidence interval may prove beneficial in determining the accuracy of the estimate.

Chapter 4

Naive Spectral Confidence Interval using the Central Moving Average

An unbiased estimate, L, of the evolutionary wavelet spectrum (EWS) can be obtained by multiplying the raw wavelet periodogram by the inner product matrix, as shown in section 2.6.2 in equation (2.63). However, although this estimate is unbiased, it is not consistent. Applying a smoothing technique to the raw wavelet periodogram, before correcting by inner product matrix, can reduce the variance of the estimator.

One simple method of smoothing is to use a weighted linear combination of the wavelet periodogram ordinates such as a running mean. The statistical properties of the running mean are easy to calculate analytically and computationally. Due to its simplicity it provides a good foundation for investigating the statistical behaviour of the EWS estimate.

A useful statistical measure for uncertainty is the confidence interval (CI). It can be used to establish the efficacy of a new technique via simulation to see if the true value is contained within the CI. In practice it provides a measure of accuracy for the results. This chapter investigates methods for producing confidence intervals, when the central moving average of the raw wavelet periodogram, is used as the estimate of the EWS

4.1 Estimation of the Evolutionary Wavelet Spectrum

The central moving average estimate of the EWS will depend on two wavelets. The first is the *synthesis* wavelet, this is the wavelet which we assume has generated the LSW process and in practice, this is usually unknown. To obtain the raw wavelet periodogram we use an *analysis* wavelet. This should be the same as the synthesis wavelet, but as the synthesis

wavelet is unknown it is usually up to the user to determine the best analysis wavelet to use. However, we shall assume that the synthesis wavelet is known and therefore the analysis and synthesis wavelet are equivalent.

4.1.1 Smoothed Wavelet Periodogram

We now define our central moving average (CMA) estimate.

Definition 4.1.1. Let $\tilde{I}_{j,k}$ be the CMA of the raw wavelet periodogram, $\{I_{j,k}\}$, of a LSW process of length T, defined by

$$\tilde{I}_{j,k} = \frac{1}{2n_j + 1} \sum_{r=-n_j}^{n_j} I_{j,k+r}, \qquad \text{for } j \in \mathbb{N} \text{ and } k, n_j \in [0, T-1]. \tag{4.1}$$

We call the quantity n_j the bin width, and it is the number of observations at scale j, before and after location k, we average the raw wavelet periodogram. Note the bin width can be different for each scale.

Using the exact expectation from equation (3.6), we find the expectation of the CMA by

$$\begin{aligned} \mathbb{E}[\tilde{I}_{j,k}] &= \frac{1}{2n_j + 1} \sum_{r=-n_j}^{n_j} \mathbb{E}[I_{j,k+r}] \\ &= \frac{1}{2n_j + 1} \sum_{r=-n_j}^{n_j} \left(\sum_{l=1}^{\infty} \sum_{m=-\infty}^{\infty} \Psi_{j,l}^2(k + r - m)\, S_{l,m} \;+\; \mathcal{O}(T^{-1}) \right). \end{aligned} \tag{4.2}$$

Equation (4.2) could be simplified by using the inner product matrix, to obtain

$$\mathbb{E}[\tilde{I}_{j,k}] = \frac{1}{2n_j + 1} \sum_{r=-n_j}^{n_j} \sum_{l=0}^{J-1} A_{j,l}\, S_{j,k+r} \;+\; \mathcal{O}(T^{-1}). \tag{4.3}$$

We shall use equation (4.3), as it simplifies future analytic and computational calculations.

We shall assume:

- the raw wavelet periodogram is independent between scales;
- the statistical properties of the raw wavelet periodogram in theorem 3.2.1 can be applied to all scales.

These assumptions will also help to simplify future calculations and provide the basis for the confidence interval estimation.

Therefore, with our assumptions the covariance of the CMA estimate is

$$\begin{aligned}
\mathrm{Cov}[\tilde{I}_{j,k}, \tilde{I}_{j,k'}] &= \mathrm{Cov}\left[\frac{1}{2n_j+1}\sum_{r=-n_j}^{n_j} I_{j,k+r}, \frac{1}{2n_j+1}\sum_{u=-n_j}^{n_j} I_{j,k'+u}\right], \\
&= \frac{1}{(2n_j+1)^2}\sum_{r=-n_j}^{n_j}\sum_{u=-n_j}^{n_j} \mathrm{Cov}[I_{j,k+r}, I_{j,k'+u}], \\
&= \frac{2}{(2n_j+1)^2}\sum_{r=-n_j}^{n_j}\sum_{u=-n_j}^{n_j}\left(\sum_{l=0}^{\infty}\sum_{m=-\infty}^{\infty}\Psi_{j,l}(k+r-m)\,\Psi_{j,l}(k'+u-m)\,S_{l,m}\right)^2 + \mathcal{O}(T^{-1}).
\end{aligned}$$

In practice we have no means of estimating the EWS for $m \notin [0, T-1)$ and $l > J-1$. Therefore, we shall assume these values are negliable, and approximate the covariance of the CMA estimate by

$$\mathrm{Cov}[\tilde{I}_{j,k}, \tilde{I}_{j,k'}] \approx \frac{2}{(2n_j+1)^2}\sum_{r=-n_j}^{n_j}\sum_{u=-n_j}^{n_j}\left(\sum_{l=0}^{J-1}\sum_{m=0}^{2^J-1}\Psi_{j,l}(k+r-m)\,\Psi_{j,l}(k'+u-m)\,S_{l,m}\right)^2. \tag{4.4}$$

As $\mathrm{Cov}[\tilde{I}_{j,k}, \tilde{I}_{j,k'}] \neq 0$, the variance of the CMA is the sum of the variance and covariance of the raw wavelet periodogram used to calculate the estimate. Therefore, the variance of the CMA is

$$\begin{aligned}
\mathrm{Var}[\tilde{I}_{j,k}] = &\frac{1}{(2n_j+1)^2}\sum_{r=-n_j}^{n_j}\left[\mathrm{Var}[I_{j,k+r}] + 2\sum_{u=r+1}^{n_j}\mathrm{Cov}[I_{j,k+r}, I_{j,k+u}]\right], \\
= &\frac{2}{(2n_j+1)^2}\sum_{r=-n_j}^{n_j}\left[\left(\sum_{l=0}^{J-1}A_{j,l}S_{l,k}\right)^2 + \mathcal{O}\left(\frac{2^j}{T}\right)\right. \\
&\left. + 2\sum_{v=u+1}^{n_j}\left\{\left(\sum_{l=1}^{\infty}\sum_{m=-\infty}^{\infty}\Psi_{j,l}(k+r-m)\,\Psi_{j,l}(k+u-m)\,S_{l,m}\right)^2 + \mathcal{O}(T^{-1})\right\}\right].
\end{aligned}$$

Then using the same approximations for the variance and covariance described previously, we have

$$\begin{aligned}
\mathrm{Var}[\tilde{I}_{j,k}] \approx \frac{2}{(2n_j+1)^2}\sum_{r=-n_j}^{n_j}&\left[\left(\sum_{l=0}^{J-1}A_{j,l}\,S_{l,k}\right)^2\right. \\
&\left. +2\sum_{u=r+1}^{n_j}\left(\sum_{l=0}^{J-1}\sum_{m=1}^{2^J}\Psi_{j,l}(k+r-m)\,\Psi_{j,l}(k+u-m)\,S_{l,m}\right)^2\right].
\end{aligned} \tag{4.5}$$

Assuming theorem 3.2.1, and independence between scales an approximate distribution of the raw wavelet periodogram is $I_{j,k} \sim (AS)_{j,k}\,\chi^2_1$. As our estimator is a linear combi-

nation of assumed iid observations, we can apply Cochran's Theorem. From Cochran's Theorem we know that the sum of $i = 1, \dots N$ independent chi-squared observations each with ν_i degrees of freedom also has a chi-squared distributions with $\sum_i \nu_i$ degrees of freedom. Therefore, we can approximate the distribution of the CMA estimate as

$$\tilde{I}_{j,k} \sim \frac{1}{2n_j+1} \sum_{r=-n_j}^{n_j} (AS)_{j,k+r}\, \chi^2_1 = \left(\frac{1}{2n_j+1} \sum_{r=-n_j}^{n_j} (AS)_{j,k+r} \right) \chi^2_{2n_j+1},$$

where $\chi^2_{2n_j+1}$ is the chi-squared distribution with $2n_j+1$ degrees of freedom, with a mean of $(2n_j+1)^{-1} \sum_{r=-n_j}^{n_j} (AS)_{j,k+r}$. As $n_j \to \infty$ this distribution is asymptotically Gaussian and this will happen quickly. Therefore, we shall assume for large n_j the distribution of the $\tilde{I}_{j,k}$ is approximately Gaussian with mean and variance as defined in equations (4.3) and (4.5).

4.1.2 Smoothed EWS Estimate

We can now use the information we obtained about the smoothed wavelet periodogram to produce an estimate of the EWS. This procedure is very simple as we now substitute $\tilde{I}_{j,k}$ in equations (2.63) for $I_{j,k}$.

Definition 4.1.2 (CMA estimate of the EWS)**.** Let $\tilde{S}_{j,k}$ be the CMA estimate of the EWS, defined by

$$\tilde{S}_{j,k} = \sum_{l=0}^{J-1} A^{-1}_{j,l}\, \tilde{I}_{l,k} = \frac{1}{2n_j+1} \sum_{l=0}^{J-1} A^{-1}_{j,l} \sum_{r=-n_j}^{n_j} I_{l,k+r}, \qquad \text{for } j \in \mathbb{N},\ k \in \mathbb{Z}, \tag{4.6}$$

where A is the inner product matrix from definition 2.6.7.

To find the expectation of the CMA smoothed EWS, we correct the CMA of the raw wavelet periodogram from equation (4.3) by the inverse inner product matrix, A.

Therefore, the expectation of $\tilde{S}_{j,k}$ is

$$\begin{aligned}
\mathbb{E}[\tilde{S}_{j,k}] &= \frac{1}{2n_j+1}\sum_{l=0}^{J-1} A_{j,l}^{-1} \sum_{r=-n_j}^{n_j} \mathbb{E}[I_{l,k+r}] \\
&= \frac{1}{2n_j+1}\sum_{r=-n_j}^{n_j}\sum_{m=0}^{J-1}\sum_{l=0}^{J-1} A_{j,l}^{-1}\, A_{l,m}\, S_{m,k+r}, \\
&= \frac{1}{2n_j+1}\sum_{r=-n_j}^{n_j}\sum_{m=0}^{J-1} \delta_{j,m}\, S_{m,k+r}, \\
&= \frac{1}{2n_j+1}\sum_{r=-n_j}^{n_j} S_{j,k+r}, \qquad (4.7)
\end{aligned}$$

which is a biased estimate.

Similarly, the variance of the CMA smoothed EWS is simply the variance of the CMA corrected by the squared inverse inner product matrix, which results in

$$\begin{aligned}
&\mathrm{Var}[\tilde{S}_{j,k}] \\
&= \left(\sum_{l=0}^{J-1} A_{j,l}^{-1}\right)^2 \mathrm{Var}\left[\tilde{I}_{l,k}\right] \\
&= \frac{2}{(2n_j+1)^2}\left(\sum_{l=0}^{J-1} A_{j,l}^{-1}\right)^2 \sum_{r=-n_j}^{n_j}\left[\mathrm{Var}[I_{l,k+r}] + 2\sum_{u=r+1}^{n_j}\mathrm{Cov}[I_{l,k+r}, I_{l,k+u}]\right] \\
&= \frac{2}{(2n_j+1)^2}\left(\sum_{l=0}^{J-1} A_{j,l}^{-1}\right)^2 \sum_{r=-n_j}^{n_j}\left[\left(\sum_{m=0}^{J-1} A_{j,m}\, S_{m,k+r}\right)^2\right. \\
&\qquad\qquad \left. + 2\sum_{u=r+1}^{n_j}\left(\sum_{m=0}^{J-1}\sum_{n=1}^{2^J}\Psi_{j,m}(k+r-n)\,\Psi_{j,m}(k+u-n)\,S_{m,n}\right)^2\right] \\
&= \frac{2}{(2n_j+1)^2}\sum_{r=-s}^{s}\left[S_{j,k+r}^2\right. \\
&\qquad\qquad \left. + 2\sum_{u=r+1}^{s}\left(\sum_{l=0}^{J-1}\sum_{m=0}^{J-1}\sum_{n=1}^{2^J} A_{j,l}^{-1}\Psi_{j,m}(k+r-n)\,\Psi_{j,m}(k+u-n)\,S_{m,n}\right)^2\right]. \qquad (4.8)
\end{aligned}$$

The calculations of the expectation and variance demonstrates how smoothing can often reduce the variance but can also introduce bias into our estimate.

As the distribution of the CMA wavelet periodogram is approximately Gaussian and because our estimate is a linear calculation, we can infer that the distribution of the bias corrected CMA wavelet periodogram is also approximately Gaussian with mean and variance from equations (4.7) and (4.8).

4.1.3 Confidence Interval

Assuming theorem 3.2.1 is true for all scales and independence between scales, a $(1-\alpha)100\%$ confidence interval for the EWS ($S_{j,k}$) can be calculated using the same method applied to the confidence interval estimation of the empirical mean for a sample of Gaussian observations. This results in a lower and upper bound of the form

$$\text{Lower} = \tilde{S}_{j,k} - 1.96\sqrt{\text{Var}[\tilde{S}_{j,k}]}, \tag{4.9}$$

$$\text{Upper} = \tilde{S}_{j,k} + 1.96\sqrt{\text{Var}[\tilde{S}_{j,k}]}. \tag{4.10}$$

As $n_j \to \infty$, the critical value will tend towards the critical value of the Gaussian distribution. If α is 0.05 this value will be approximately 1.96.

4.2 Computational Details

We created a function in the statistical package `R` (`R` Development Core Team, 2012) to compute the central moving average estimator of the EWS with confidence intervals. To produce these estimates our program used functions from the `wavethresh` (Nason, 2012) and `igraph` (Csardi, 2010) packages. The user must supply five components, which are:

- The LSW process for which an EWS estimate is to be produced.
- Analysis wavelet (AW), which is defined as a number between 1 and 10 and will be a wavelet from the Daubechies Extremal Phase family.
- Smoothing wavelet (SW), with the same parameters of the analysis wavelet and is used to define the `ewspec` estimate we use to select our bin width.
- p-value, $\alpha \in [0, 1]$, the default value is 0.05 to calculate the 95% confidence interval.
- Whether the bin width, n_j, should be selected separately for each scale.

Although the CMA estimator does not require a SW for its calculations, selection of the bin width (n_j) was based on the CMA estimator with the smallest average mean squared error (AMSE) compared with the translation-invariant de-noised estimator of Nason et al. (2000), which does require a SW. In absence of the true EWS, this is the next best approximation, can be easily obtained using the `ewspec` function and was also the method used by Nason (2013). If there is no prior knowledge available to determine AW and SW, we advise using Daubechies Extremal Phase wavelet with 10 vanishing moments (D_{10}).

Currently, this is the smoothest wavelet available in this function, and should produce a TI de-noising estimate which is less susceptible to Gibbs phenomena and large errors in the data. Also, if there is no prior knowledge, we suggest the user selects the option to calculate the bin width for each scale separately because varying the bin width will increase the probability that features of the EWS at a particular scale are more likely to be identified.

The first step is to obtain the raw wavelet periodogram from the LSW process with the user defined AW, which is obtained using the `ewspec` function from `wavethresh` package. Then for each scale a CMA estimate is produced using the function `running.mean` from the `igraph` package for values of $n_j = 1, \ldots, T-1$. To ensure the sample size of our CMA estimate remained the same as the raw wavelet periodogram the first and last n_j observations were reflected about their ends (similar to many functions in `wavethresh`). An alternate method would be to link the values at the start and end of the process, but this is inappropriate as there is no evidence to suggest that the statistical properties of the initial observations will be the equivalent to the end. As our LSW assumptions restricts the second order structure to evolve slowly, it is not unreasonable to assume that the values preceding and following a particular location will possess similar statistical properties.

Code was written in `R` to calculate the variance of the raw wavelet periodogram based on equation (3.7). We used the CMA estimator of the EWS as proxy for the truth. Firstly, the cross-auto correlation wavelet (definition 2.6.5) is calculated using the given AW, which is then used to calculate the covariance of the raw wavelet periodogram between different locations at the same scale. For each scale it returns a $T \times T$ matrix, with the requested covariance values. This does not mean there is no covariance, only that the calculation has not been made. To increase the speed of the computation the maximum lag evaluated can be decreased. As the covariance decreases to zero as the lag increases it is not unreasonable to set these to zero.

The variance of the CMA estimator of the wavelet periodogram ($\tilde{I}$) is simply the summation of the variance and covariance of the raw wavelet periodogram divided by $(2n_j + 1)^2$. Finally, to obtain the variance of our CMA estimator of the EWS ($\tilde{S}$), we correct $\text{Var}[\tilde{I}]$ by the inverse inner product matrix squared. Using the previously define p-value we calculate the confidence interval for the CMA estimate. The function returns the values of the bin width n_j for $j = (J-1), \ldots 0$, CMA estimator, variance and $(1-\alpha)100\%$ CI for the $\tilde{S}$ and $\tilde{I}$.

n_j	j									
	9	**8**	**7**	**6**	**5**	**4**	**3**	**2**	**1**	**0**
$\mathbf{X_t}$	75	74	77	77	71	68	55	64	42	41
$\mathbf{Y_t}$	56	66	74	78	84	74	77	60	71	63

Table 4.1: Bin width (n_j) used to perform the central moving average on the raw wavelet periodogram for X_t and Y_t at all scales.

4.2.1 Simulations

To test the performance of the CMA, an estimate of the EWS using the test data sets described in section 3.4 with Gaussian innovations was compared to the original spectrum from equations (3.9) and (3.12). The TI de-noising estimate was produced using Daubechies Extremal Phase wavelet with ten vanishing moments, and different bin widths were selected for each scale.

The bin width, n_j, for both LSW processes is listed in table 4.1. The bin width tended to be biggest at scales closest to the scale which contained all the power, i.e. scale $j = 9$ for X_t and $j = 6$ for Y_t. The size of the chosen bin width (generally) decays as the scales become coarser and finer than the power scale.

The plots in figure 4.1(a) and 4.1(b) are the estimated spectra for all scales of the piecewise and slowly evolving spectra, respectively. Some of the leakage between scales has resulted in some negative values in the CMA estimate of the EWS. In figure 4.1(a), the negative values are mainly contained at scales $j = 8, 7$. As all of the power of the piecewise spectrum is at the finest scale the power can only leak into coarser scales. In figure 4.1(b) of the CMA estimate of the slowly evolving EWS there is evidence that power has leaked into scales $j = 8, 7, 5$ and 4, indicating that power located at a mid scale will leak into finer and coarser scales. At scale $j = 5$, there appears to be a large amount of negative power. To evaluate the extent of leakage between scales we examine in detail scales $j = 9, 8$ and 7 of the CMA estimate of the piecewise EWS and $j = 7, 6$ and 5 of the slowly evolving EWS.

The plots in figure 4.2 (a) and (b) are the true, estimated and 95% CI of the piecewise EWS at scales $j = 7, 8$, which are the second and third finest scale. Even though there appears to be some degree of power present at these scales, as the 95% CI contains zero there is no significant difference between these values and zero.

Figure 4.2(c) shows the plots of the true, estimated and 95% CI of the piecewise EWS at scale $j = 9$, which is the scale that contains all the power. The CMA estimator struggles

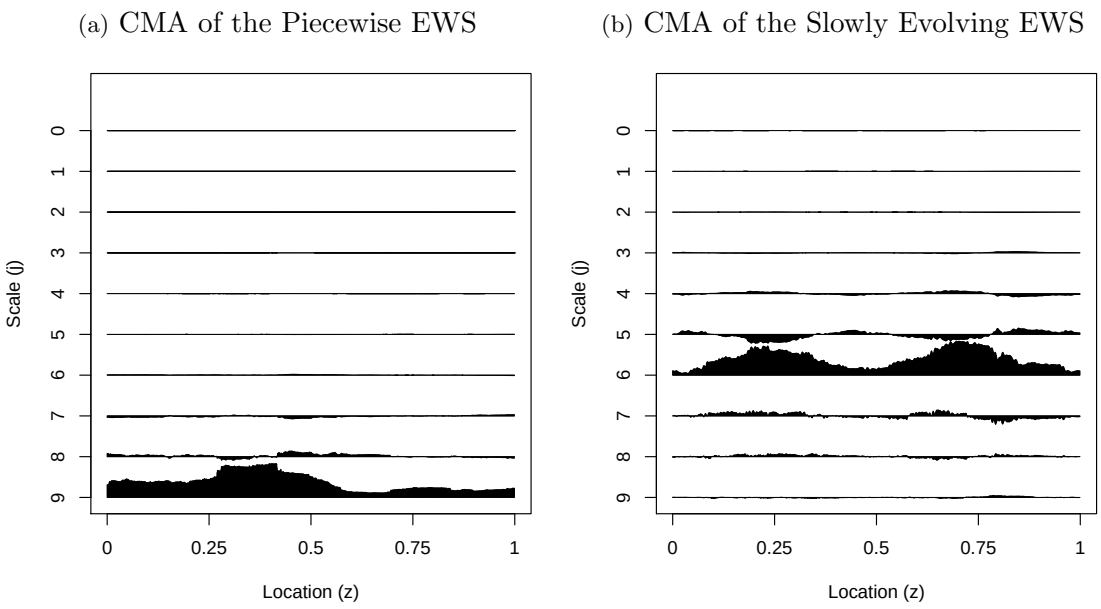


Figure 4.1: Plots of the central moving average estimate of the piecewise (a) and slowly evolving (b) spectra.

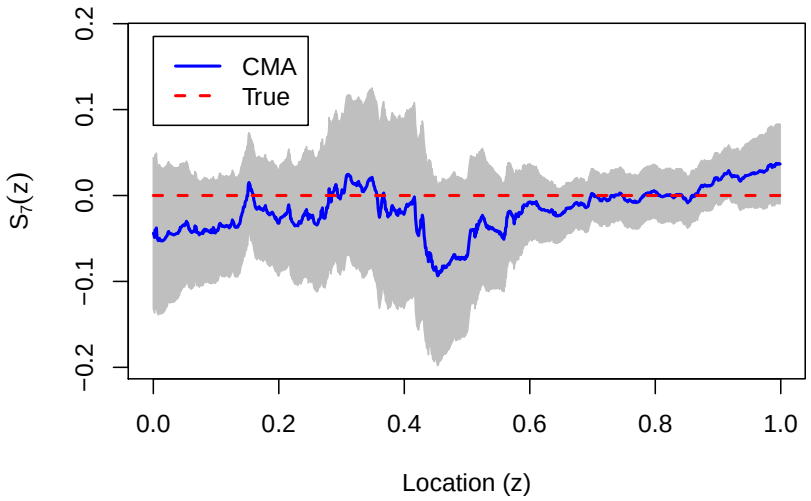
to identify the change points at the correct locations and CMA estimate implies the first change point in the EWS at scale $j = 9$ is $z \approx 0.25$ rather than $z = {}^1\!/_3$. The 95% CI does not always capture the true signal. Approximately 17.09% of the true signal fell outside the CI. However, the CMA estimator is not susceptible to Gibbs phenomena as the TI de-noised estimate is.

As D_{10} is one of the smoother wavelets, selecting of SW = D_{10} will mean the TI de-noising estimator will struggle with piecewise constant signals, and due to the method of selecting the bin widths this also affects how well the CMA estimate performs. However, we believe the shift in the location of the change point observed in figure 4.2(c) is caused by a very large raw wavelet periodogram value after the change point at $z \approx 0.35$ (see figure 3.5(a) in section 3.4.1). As this value enters into the CMA window it greatly increases the estimator, and as it exits there is a sharp reduction. The poor CI indicated that either our assumptions of the covariance or the Gaussian distribution of the CMA were not appropriate. However, as our estimate relies so heavily upon the estimate of TI de-noising this could also contribute to the poor performance of the CI.

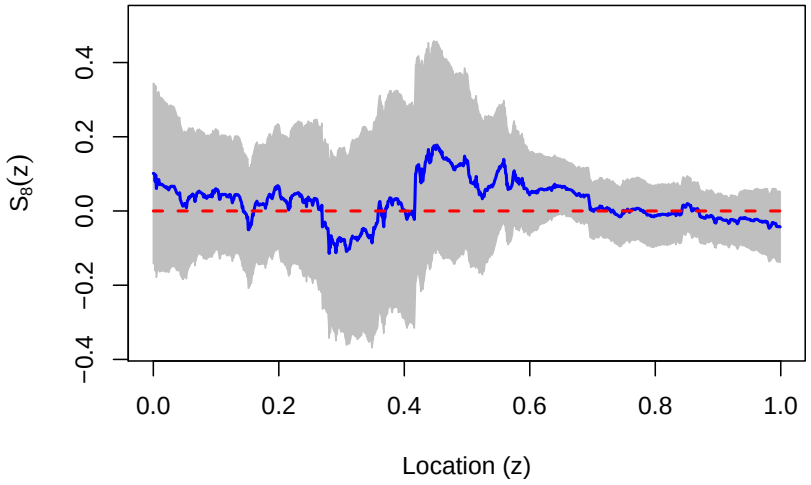
All three of the CMA estimates in figure 4.3 of the slowly evolving EWS for scales $j = 7, 6$ and 5 are very similar to the true value. This demonstrates how well the TI de-noising works at estimating smooth functions, even though smoothing is based on the Haar wavelet (see figure 3.10 in section 3.4.2).

However, the 95% CI for scales $j = 7, 6$ and 5 all include zero $\forall\ z$. The CI imply there

(a) CMA and 95% CI at $j = 7$, with bin width $n_7 = 77$



(b) CMA and 95% CI at $j = 8$, with bin width $n_8 = 74$



(c) CMA and 95% CI at $j = 9$, with bin width $n_9 = 75$

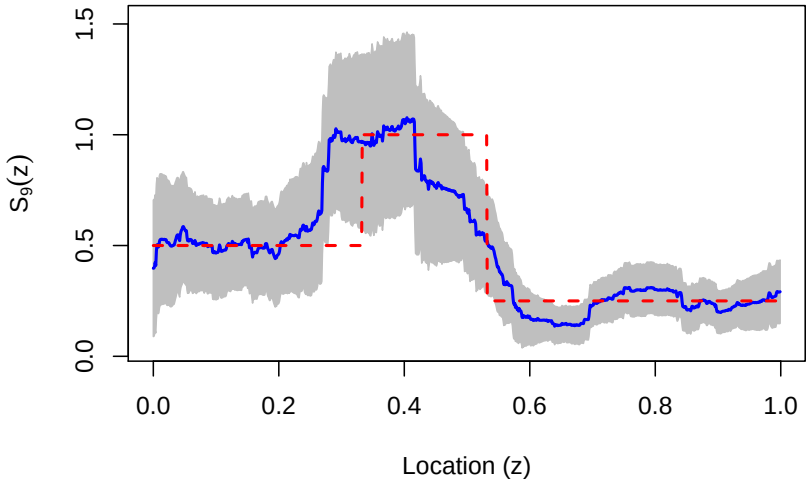
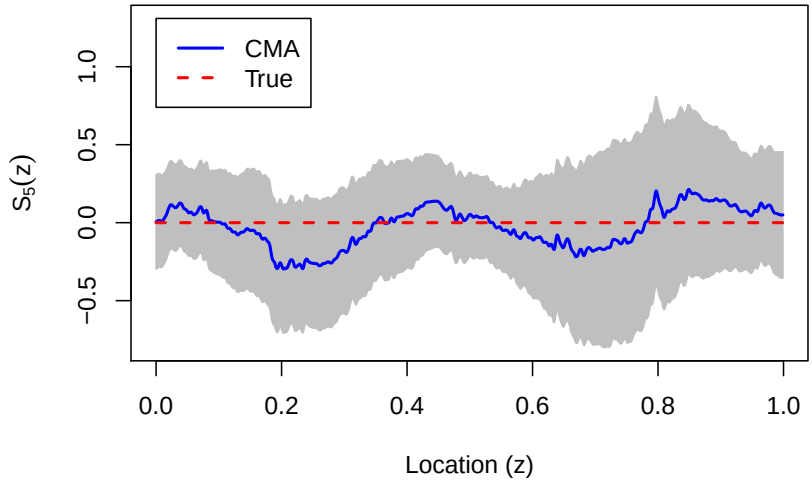
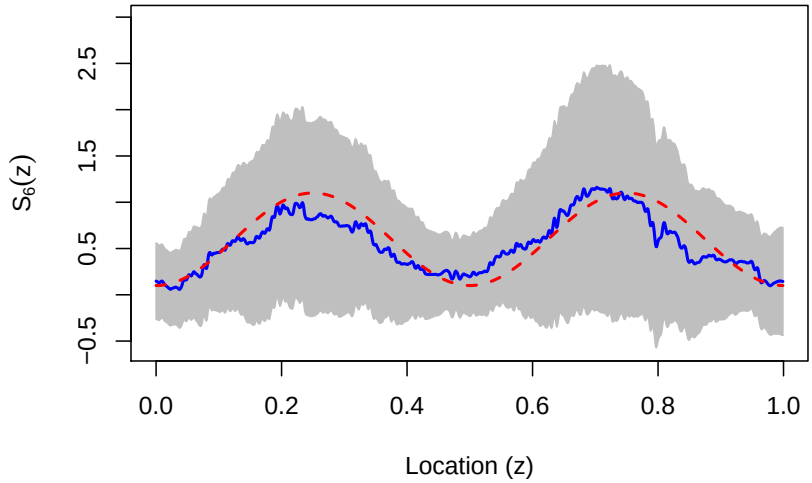


Figure 4.2: Plots of the CMA estimate (blue solid line), with a 95% CI (grey area) and true EWS (red dashed line) of the piecewise EWS at scales $j = 9, 8, 7$. CMA estimate obtained using the Haar analysis wavelet, and SW = D_{10}.

(a) CMA and 95% CI at $j = 5$, with bin width $n_5 = 77$



(b) CMA and 95% CI at $j = 6$, with bin width $n_6 = 74$



(c) CMA and 95% CI at $j = 7$, with bin width $n_7 = 75$

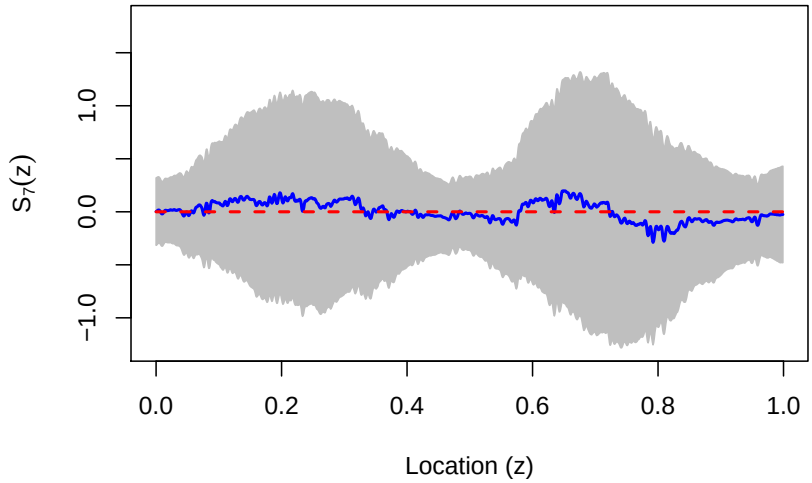


Figure 4.3: Plots of the CMA estimate (blue solid line), with a 95% CI (grey area) and true EWS (red dashed line) of the slowly evolving EWS at scales $j = 7, 6, 5$. CMA estimate obtained using the Haar analysis wavelet, and SW = D_{10}.

is no significant evidence to suggest there is any power at these scales. This is further evidence to suggest the variance of the CMA estimate of the EWS may not be accurate, possibly a result of assuming independence by scale.

4.2.2 Conclusion

The CMA estimator work reasonably well at recovering the true EWS where the underlying function is smooth, but struggles to accurately identify the change points of the piecewise constant EWS because of the over influence of large observations. This suggested that a more appropriate weighting of the observations within the bin width, which puts more importance on observations closer to the location, may help to solve this problem.

As the bin width is the parameter which determines how smooth the CMA estimate, using the TI de-noising estimate influences the performance of the CMA estimate. However, the results of the X_t and Y_t EWS estimates indicate that this may not be the best method for selecting bin widths as it appears to restrict the range of functions which it can adequately model.

Further consideration should be given to the estimation of the variance. In section 3.4 the evidence suggested that there was a covariance between observations at different scales at the same location or close by. Therefore, by considering the covariance of the raw wavelet periodogram at different scales and at the same location or close by, we may be able to improve the accuracy. This may improve the CI.

To improve our method we could include a larger variety of wavelets. This would increase the range of functions which could be recovered. The current method was based upon wavelets which we were confident conformed to our assumptions.

Chapter 5

Bayesian Modelling of the Log Transformed EWS

This chapter describes the theoretical calculations required to produce estimates and corresponding confidence intervals for the evolutionary wavelet spectrum (EWS) using Bayesian wavelet shrinkage of the log transformed raw wavelet periodogram. We shall assume that we wish to model the EWS of an LSW process over a fixed time period.

We use work from chapter 3 to define our regression model, and our method was developed following the smoothing technique of Pensky et al. (2007) for a stationary spectrum. The main difference between the two methods for estimating the spectrum stems from the progression from a stationary to a locally stationary spectrum. Also, we are using the calculation of the posterior distribution to obtain a confidence interval for our estimate.

We shall begin by developing the regression model for our problem in section 5.1. In particular, an expression for the distribution of the error term (section 5.1.1) and discrete wavelet transformation (DWT) of the regression model (section 5.1.2).

We then develop a Bayesian shrinkage rule in section 5.2 to 'shrink' the wavelt coefficient. Section 5.1.1) and knowledge of the DWT provides the foundations for the mixture-likelihood in section 5.2.1. We investigate using the Berger-Müller mixture prior which combines the Dirac delta function and a symmetric unimodal distribution to convey known information about the sparsity of wavelet coefficients in section 5.2.2. These naturally lead to the formation of a posterior distribution (section 5.2.3), allowing calculation of the posterior mean (section 5.2.4) and variance (section 5.2.5). In section 5.2.6 we discuss determining the hyperparameters. Firstly the likelihood weights, followed by

the prior precision and weights.

Sections 5.3.1 and 5.3.2 consider two possible distributions as part of the mixture prior: the Gaussian and Laplace distributions. For each of these distributions we produce the analytic calculations using both priors for the posterior distribution, mean and variance. We also calculate the marginal maximum likelihood to use in hyperparameter estimation.

5.1 Regression Model for the Log-EWS

Suppose we have a LSW process observed at $2^J = T$ uniformly spaced points over a fixed time period. Using a pre-determined analysis wavelet, which we shall assume is the same as the synthesis wavelet used to generate the LSW process, we obtain its raw wavelet periodogram, $I_{j,k}$, for $j = 0, \ldots, J-1$ and $k = 0, \ldots T-1$. Nason et al. (2000) determined that the expectation of the raw wavelet periodogram is $R_j(z) = \sum_l A_{j,l} S_l(z)$, where A is the inner product matrix from definition 2.6.7 and S is the EWS in definition 2.6.2. If we assume the distribution of the innovations ($\xi_{j,k}$) in (2.53) are independent, identically distributed (iid) Gaussian, and the periodogram is independent between scales (j) and locations (k), theorem 3.2.1 suggests we could model the relationship between the EWS and periodogram as

$$I_{j,k} \approx R_j\left(\frac{k}{T}\right) Z_{j,k}^2, \tag{5.1}$$

where $Z_{j,k}$ are iid $\mathcal{N}(0,1)$, such that $Z_{j,k}^2 \overset{iid}{\sim} \chi_1^2$. Direct estimation using (5.1) would be difficult because of the highly skewed error and multiplicative relationship between the expectation and error. However, by using a log transformation equation on (5.1) this multiplicative relationship becomes additive and we can stabilise the error and pull the distribution closer to Gaussianity. This results in a new regression model of the form

$$\log(I_{j,k}) = \log\left\{ R_j\left(\frac{k}{T}\right)\right\} + \varepsilon_{j,k}, \tag{5.2}$$

where we assume $\varepsilon_{j,k}$ is iid with a $\log(\chi_1^2)$ distribution.

The distribution of $\varepsilon_{j,k}$ does not possess a zero mean. However, by subtracting the expectation of $\varepsilon_{j,k}$ from both sides of equation (5.2), we obtain the following model:

$$H_{j,k} = g_j\left(\frac{k}{T}\right) + e_{j,k}, \tag{5.3}$$

for $j = 0, \ldots, J-1$ and $k = 0, \ldots T-1$, where

- $H_{j,k} = \log(I_{j,k}) - \mathbb{E}[\varepsilon_{j,k}]$,
- $g_j\left(k/T\right) = \log\left\{R_j\left(k/T\right)\right\}$,
- $e_{j,k} = \varepsilon_{j,k} - \mathbb{E}[\varepsilon_{j,k}]$.

We shall call $H_{j,k}$ the log-wavelet periodogram. For simplicity, we assume the raw wavelet periodogram ($I_{j,k}$) is independent by scale and hence the log-wavelet periodogram is also independent by scale. Therefore each scale j can be smoothed separately. Hence, we shall drop the scale subscript j, and assume that each scale of the raw wavelet periodogram will be addressed in the same manner.

Equations (5.1) and (5.3) are very similar to equations (2.1) and (2.2) in Pensky et al. (2007) for the stationary periodogram and spectrum. The key difference is between the error distribution for the stationary and LSW cases. The distribution of the stationary peridogram error is approximately χ^2_2 in Pensky et al. (2007, equation (2.1)), whereas the error term in our case is approximately χ^2_1 in (5.1).

Next we shall develope the distribution of the log error term from equation (5.3).

5.1.1 Distribution of the Log-Error

Assume that we have independence between scales, then for a particular scale $Z_k \overset{iid}{\sim} \mathcal{N}(0,1)$, which implies $Z_k^2 \sim \chi^2_1$. We can deduce that the distribution of $\log(Z_k^2)$ is $\log(\chi^2_1)$.

Lemma 5.1.1 (Distribution of the Error Term)**.** *Assuming we have independence by scale and location, the random variables e_k for $k = 0, \ldots, T-1$ are approximately iid with probability density function*

$$f_E(x) = \sqrt{\frac{\gamma^*}{\pi}} \exp\left\{\frac{x}{2} - \gamma^* e^x\right\}, \qquad -\infty \geq x \geq \infty, \tag{5.4}$$

where $\gamma^* = \frac{e^{-\gamma}}{4} \approx 0.1403649$ and $\gamma \approx 0.577126$ is the Euler-Mascheroni constant. This distribution has a mean of 0 and variance is $\pi^2/2$. The cumulants for this distribution are $\kappa_1 = 0$ and $\kappa_r = \psi^{(r-1)}(1/2)$ for $r > 1$, where $\psi^{(m)}(\cdot)$ is the polygamma function of order m.

The probability an observation is less than or equal to x is

$$\mathbb{P}(X \leq x) = \frac{1}{\sqrt{\pi}}\, \gamma\left(\frac{1}{2}, \gamma^* e^x\right), \tag{5.5}$$

where $\gamma(s,x)$ is the lower incomplete gamma function. The probability an observation is greater than or equal to x is

$$\mathbb{P}(X \geq x) = \frac{1}{\sqrt{\pi}}\,\Gamma\left(\frac{1}{2}, \gamma^* e^x\right), \tag{5.6}$$

where $\Gamma(s,x)$ is the upper incomplete gamma function.

Proof (Distribution of the Error Term)**.** Show that $\varepsilon_k - \log\{2\} - \gamma \sim \log\{\chi_1^2\}$. Firstly, we need to find the density function of the $\log\left(\chi_1^2\right)$ distribution. Let $Z \overset{iid}{\sim} \chi_1^2$, then

$$f_Z(z) = \frac{2^{-1/2}}{\Gamma(1/2)} z^{-1/2} e^{-z/2}, \qquad z > 0,$$

where $\Gamma(\cdot)$ is the Gamma function. The moments of the chi-square distribution with one degree of freedom can be calculated from

$$\mathbb{E}[Z^m] = 2^m \frac{\Gamma\left(m + 1/2\right)}{\Gamma(1/2)}, \qquad \text{for } m \in \mathbb{N}.$$

Let $Y = \log(Z)$, which implies that

$$Z = e^Y \qquad \text{and} \qquad \left|\frac{\mathrm{d}z}{\mathrm{d}y}\right| = e^y.$$

The density function of Y can be found using the formula

$$f_Y(y) = f_Z(z(y))\left|\frac{\mathrm{d}z}{\mathrm{d}y}\right| = \frac{2^{-1/2}}{\Gamma(1/2)}\ \exp\left\{\frac{y}{2} - \frac{e^y}{2}\right\},$$

for $y \in \mathbb{R}$. Using the moments of the chi-square distribution with one degree of freedom, we can write the characteristic function of Y as

$$\varphi_Y(t) = \mathbb{E}\left[e^{\mathrm{i}tY}\right] = \mathbb{E}\left[e^{\mathrm{i}t\log(Z)}\right] = \mathbb{E}\left[Z^{\mathrm{i}t}\right] = \frac{2^{\mathrm{i}t}\Gamma\left(\mathrm{i}t + 1/2\right)}{\Gamma(1/2)}.$$

As the n^{th} cumulant is calculated by the n^{th} difference of $\log\{\mathbb{E}\left[e^{\mathrm{i}tY}\right]\}$, and the polygamma function of order $n-1$ is defined by

$$\psi^{(n-1)}(z) = \frac{\mathrm{d}^n}{\mathrm{d}z^n}\log\Gamma(z),$$

we can conclude the cumulants of Y are

$$\kappa_{Y,1} = \log(2) + \psi^{(0)}\left(1/2\right),$$
$$\kappa_{Y,r} = \psi^{(r-1)}(1/2).$$

This results in

$$\mathbb{E}[Y] = \kappa_{Y,1} = -\log(2) - \gamma,$$
$$\mathrm{Var}[Y] = \kappa_{Y,2} = \psi^{(1)}\left(1/2\right) = \frac{\pi^2}{2},$$

where $\gamma \approx 0.577126$ is the Euler-Mascheroni constant.

To find the distribution of the error term e_k, we need to centralise the previous distribution. This will not effect any of the cumulants except the mean. Let $X = Y + \log(2) + \gamma$, which implies that

$$y = x - \log(2) - \gamma \qquad \text{and} \qquad \left|\frac{\mathrm{d}y}{\mathrm{d}x}\right| = 1.$$

Therefore the probability density function of X is

$$\begin{aligned} f_X(x) = f_Y(y(x))\left|\frac{\mathrm{d}y}{\mathrm{d}x}\right| &= \frac{2^{-1/2}}{\Gamma(1/2)}\ \exp\left\{\frac{1}{2}(x - \log(2) - \gamma) - \frac{1}{2}e^{x-\log(2)-\gamma}\right\} \\ &= \sqrt{\frac{\gamma^*}{\pi}}\ \exp\left\{\frac{x}{2} - \gamma^* e^x\right\}, \qquad\qquad \text{for } x \in \mathbb{R}, \end{aligned}$$

where $\sqrt{\pi} = \Gamma\left(1/2\right)$ and $\gamma^* = e^{-\gamma}/4$. Shifting the random variable Y by its mean $\log(2)+\gamma$ will only affect the first cumulant of Y, such that $\kappa_{X,1} = 0$. All other cumulants for X will equal those of $\kappa_{Y,r}$ for $r > 1$.

The probability an observation from this distribution is less than or equal to x is

$$\mathbb{P}(X \leq x) = \int_{-\infty}^{x} \sqrt{\frac{\gamma^*}{\pi}}\ \exp\left\{\frac{u}{2} - \gamma^* e^u\right\} \mathrm{d}u.$$

Let $z = \gamma^* e^u$, then $\mathrm{d}u = z^{-1}$, therefore

$$\mathbb{P}(X \leq x) = \int_{-\infty}^{\gamma^* e^x} \frac{1}{\sqrt{\pi}} z^{-1/2} e^{-z} \mathrm{d}z = \frac{1}{\sqrt{\pi}}\ \gamma\left(\frac{1}{2}, \gamma^* e^x\right),$$

where $\gamma(s, x)$ is the lower incomplete gamma function for s, $x \geq 0$.

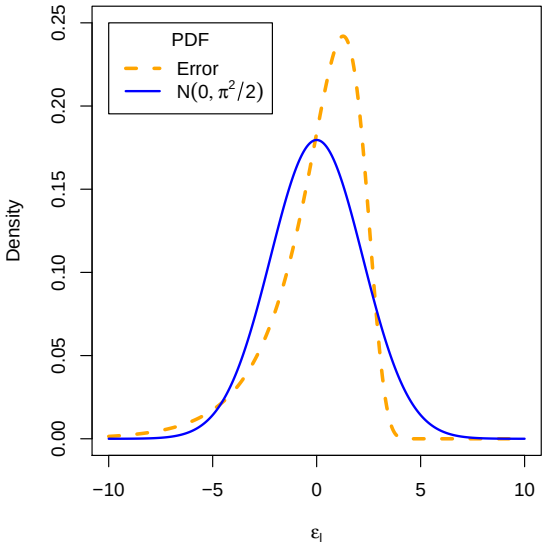


Figure 5.1: Density plots of the error pdf (a.k.a. the centralised $\log(\chi^2_1)$) defined in equation (5.4) (orange dashed line) and the $\mathcal{N}(0, \pi^2/2)$ pdf (solid blue line).

Similarly, the probability an observation is greater than or to x is

$$\mathbb{P}(X \geq x) = \int_x^{-\infty} \sqrt{\frac{\gamma^*}{\pi}} \exp\left\{\frac{u}{2} - \gamma^* e^u\right\} \mathrm{d}u.$$

using the same substitution as before we find

$$\mathbb{P}(X \geq x) = \int_{\gamma^* e^x}^{\infty} \frac{1}{\sqrt{\pi}} z^{-1/2} e^{-z} \mathrm{d}z = \frac{1}{\sqrt{\pi}} \Gamma\left(\frac{1}{2}, \gamma^* e^x\right),$$

where $\Gamma(s, x)$ is the upper incomplete gamma function for s, $x \geq 0$. □

A plot of the error probability density function (pdf) defined in equation (5.4) and the Gaussian pdf with zero mean and variance $\pi^2/2$ can be seen in figure 5.1. The plot demonstrates the difference between the Gaussian and error pdf's with the same variance.

5.1.2 DWT of the Regression Model

By performing the discrete wavelet transform (DWT) on (5.3), we obtain the wavelet coefficients to which we apply the Bayes rule. Let the $T \times T$ matrix W be the orthogonal matrix associated with a discrete wavelet transforms DWT with J scales, where $J = \log_2(T)$. Then, in the wavelet domain equation (5.3) becomes

$$\mathbf{h} = \boldsymbol{\beta} + \boldsymbol{\delta}, \tag{5.7}$$

where

$$\begin{aligned} \mathbf{h} &= W\mathbf{H}; & \mathbf{H} &= [H_0, \ldots, H_{T-1}]^{\mathrm{T}}; \\ \boldsymbol{\beta} &= W\mathbf{g}; & \mathbf{g} &= \left[g(0), \ldots, g\left({}^{(T-1)}/_{T}\right)\right]^{\mathrm{T}}; \\ \boldsymbol{\delta} &= W\boldsymbol{\varepsilon}; & \boldsymbol{\varepsilon} &= [\varepsilon_0, \ldots, \varepsilon_{T-1}]^{\mathrm{T}}. \end{aligned}$$

The vectors $\mathbf{h}$, $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ are of length T, consisting of sub vectors of each individual wavelet transform. For example,

$$\boldsymbol{\delta} = [\delta_0, \boldsymbol{\delta}_{J-1}, \ldots, \boldsymbol{\delta}_1],$$

where $\boldsymbol{\delta}_l = \left[\delta_{l,0}, \delta_{l,1}, \ldots, \delta_{l,2^l-1}\right]$ is the l^{th} sub-vector which corresponds to the l^{th} scale discrete wavelet transform. Here $l = J-1$ is the finest scale and $l = 0$ is the coarsest scale. Each location within a scale sub-vector shall be denoted by m, where $m = 0, \ldots, 2^l - 1$.

Since the exact form of W is known this means that an approximation of the distribution of $\boldsymbol{\delta}$ can be derived. For fine scales the density function of $\delta_{l,m}$ will be strongly affected by the density in Lemma 5.1.1, $f_E(\cdot)$. However, as the support of the wavelet increases and for coarser scales, the distribution of $\boldsymbol{\delta}_l$ tends to Gaussianity as the support of the wavelet increases. Moulin (1994, section III, part B) proved that for variables which are iid, but not necessarily Gaussian, the DWT of these variables tend to Gaussianity asymptotically.

5.2 Bayesian Shrinkage Rule

We shall now develop a Bayesian shrinkage rule for the DWT of the log wavelet periodogram. We shall begin with the likelihood.

5.2.1 Likelihood

The likelihood distribution used in the upcoming Bayesian analysis is determined by the distribution of the error. As $\delta_{l,m} = \sum_t \psi_{l,m-t}\varepsilon_t$ and the wavelets have unit norm, we shall assume the variance of $\delta_{l,m}$ is $\sigma^2 = {}^{\pi^2}/_{2}$. Then following Pensky et al. (2007), we approximate the density of $\boldsymbol{\delta}$ by a mixture model

$$p\left(\delta_{l,m}\,|\beta_{l,m}\right) = \zeta_l\left(\delta_{l,m}\right) = \left(1-\lambda_l\right)\varphi_\sigma(\delta_{l,m}) + \lambda_l\; f_E\left(\delta_{l,m}\right), \qquad \text{for } \lambda_l \in [0,1], \quad (5.8)$$

where $\varphi_\sigma(\cdot)$ is the Gaussian pdf with variance σ^2 and mean zero. The variables λ_l are known as the likelihood weights. The weights determine the mixture between the centralised log chi-square and Gaussian distribution for scale l of the likelihood. We shall now empirically calculate the likelihood weights for Daubechies Extremal Phase wavelets.

Likelihood Weights

As the support of the wavelet increases, the distribution of the wavelet coefficients are asymptotically Gaussian (see theorem 3.1.1). Therefore, we know that for large l, $\lambda_l \approx 0$. Usually the distribution is approximately Gaussian by the middle scale.

Pensky et al. (2007) determined the likelihood weights (λ_l) by matching the skewness of the likelihood mixture function and the empirical distribution of the wavelet transformed errors. As the support of the wavelet increases the weight, λ_l, decreased.

To calculate the likelihood weights we simulated $e_k \sim \log(\chi^2_1)$ for $k = 0, \ldots, 2^{16} - 1 = 65535$ and subtracted $\log\{2\} + \gamma$ from each realisation. We then performed a DWT with different wavelets and compared the empirical skewness of the wavelet coefficients (δ_l) with the skewness of the error and Gaussian distributions. The skewness for the error distribution can be calculated from the cumulants using

$$\frac{\kappa_{e,3}}{(\kappa_{e,2})^{3/2}} = \frac{\psi^{(2)}\left(1/2\right)}{\left[\psi^{(1)}(1/2)\right]^{3/2}} = -1.535142,$$

where $\kappa_{e,2}$ and $\kappa_{e,3}$ are the second and third cumulants of the error distribution. The empirical likelihood weight was estimated by choosing λ_l in equation (5.8) so the skewness of $p(\delta_{l,m}|\beta_{\ l,m})$ matched the empirically derived skewness of our simulations. This was repeated a thousand times for each scale and the final estimate of λ_l was the mean of these replications.

For each scale and replication we also performed a Kolmogorov-Smirnov (KS) one-sided test for Gaussianity on the empirical wavelet coefficients. We compared the distribution of the wavelet coefficients with a $\mathcal{N}(0, \pi^2/2)$ distribution. The final p-value was the mean for the one thousand replications for each wavelet and the first ten finest scales, l.

Figure 5.2 shows the density plots of the six finest scale Haar wavelet coefficients δ_l for $l = 16 \ldots 10$, of the DWT from $T = 2^{16}$ samples of the error pdf, e_k. The KS test statistic p-value was < 0.0001 for the first two finest scales, but > 0.05 for each subsequently coarser scale. The largest likelihood weight $\lambda_l < 0.008$, even at the finest scale.

The Haar wavelet has the smallest support of the all wavelets, so we expected the

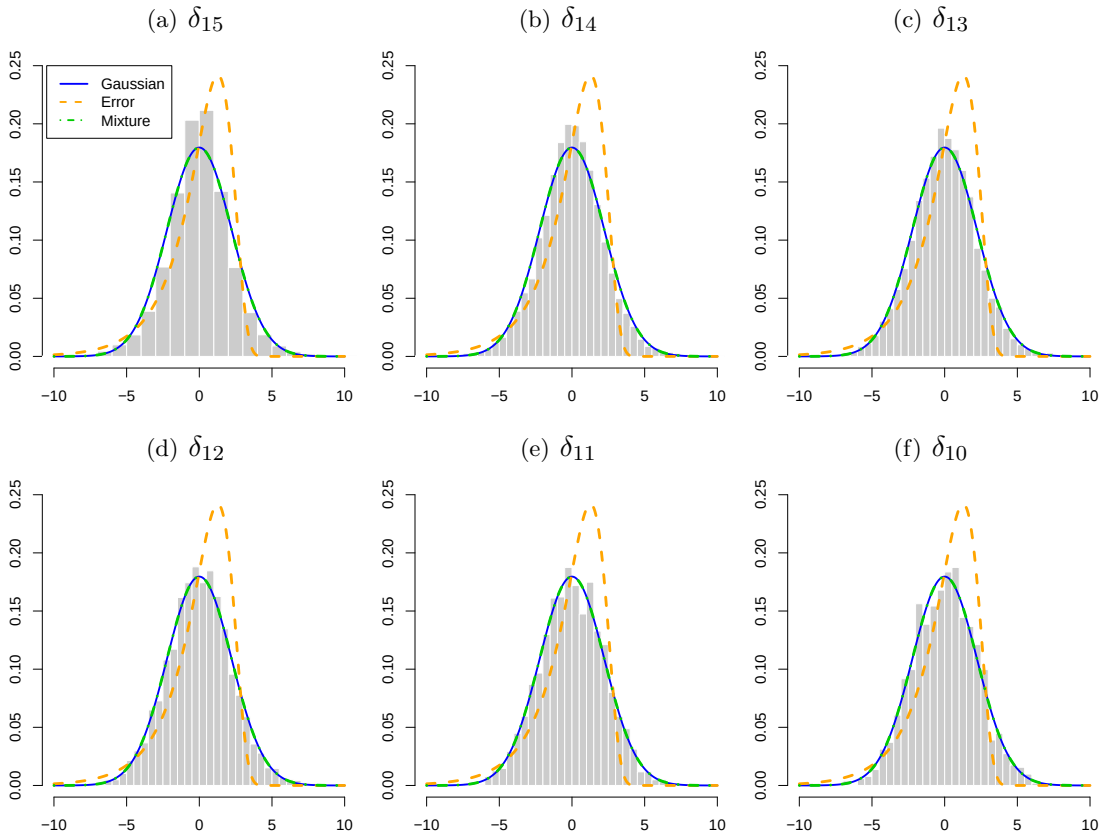


Figure 5.2: Density plots of δ_l for the six finest scales ($l = 15, \ldots, 11$), from a discrete Haar wavelet transform with plots of the Gaussian pdf (solid blue line), error pdf (orange dashed line) and likelihood (green dashed and dotted line) function calculated using empirical likelihood weights.

Scale (l)	$\mathbf{D_2}$ $\boldsymbol{\lambda_l}$	**p-value**	$\mathbf{D_3}$ $\boldsymbol{\lambda_l}$	**p-value**
$J-1$	0.4594	< 0.001	0.3895	< 0.001
$J-2$	0.2850	< 0.001	0.2541	< 0.001
$J-3$	0.1922	0.001	0.1763	0.004
$J-4$	0.1329	0.112	0.1232	0.160
$J-5$	0.0934	0.364	0.0866	0.385
$J-6$	0.0652	0.479	0.0629	0.487
$J-7$	0.0478	0.492	0.0448	0.491
$J-8$	0.0317	0.522	0.0261	0.500
$J-9$	0.0253	0.523	0.0205	0.511
$J-10$	0.0047	0.506	0.0134	0.492
$> J-10$	0		0	

Table 5.1: A table of the empirical weights and p-value from the Kolmogorov-Smirnov test for Gaussianity for a Daubechies Extremal Phase wavelet with two (D_2) and three (D_3) vanishing moments, where $J-1$ is the finest scale.

p-value from the KS test to be the smallest and the weights (λ_l) to be the largest for all scales compared with smoother wavelets. However, as only the first two scales possessed significant p-values and the corresponding weights were so small, the error distribution would have little effect on the likelihood mixture distribution. Therefore, we concluded it would be best to set $\lambda_l = 0$ for all scales. The plots in figure 5.2 verify our estimates of λ_l and demonstrates how little difference there is visually between the Gaussian pdf and density histogram of δ_l.

Table 5.1 shows the empirical likelihood weights, λ_l, and the p-value from the KS test for Daubechies Extremal Phase wavelet with two (D_2) and three (D_3) vanishing moments and the first ten finest scales. The KS p-value indicates that for scales grater than $J-3$, there is no evidence to suggest the distribution of δ is significantly different from $\mathcal{N}(0, \pi^2/2)$. If $\lambda_l < 0.08$ and the p-value > 0.05, we felt it was not unreasonable to set $\lambda_l = 0$ because the error distribution was contributing little to the mixture likelihood and there was no evidence to suggest the distribution was not $\mathcal{N}(0, \pi^2/2)$.

The density plots in figures 5.3 are the five finest scale wavelet coefficient (δ_l for $l = 16, \ldots, 11$) of the DWT using D_2 (in (a) to (e)) and D_3 (in (f) to (j)) wavelets of the error distributed e_k for $k = 0, \ldots, 2^{16} - 1$. The likelihood mixture function was calculated using the empirical likelihood weights λ_l in table 5.1.

We also performed a KS Gaussianity test and calculated weights for D_n where $n > 3$. Even though for the first two scales generally resulted in a KS p-value less than 0.001, the empirical likelihood weights were 0.1 or less. Therefore, we decided to set $\lambda_l = 0, \ \forall\ l$ for the Daubechies Extremal Phase wavelet with four or more vanishing moments.

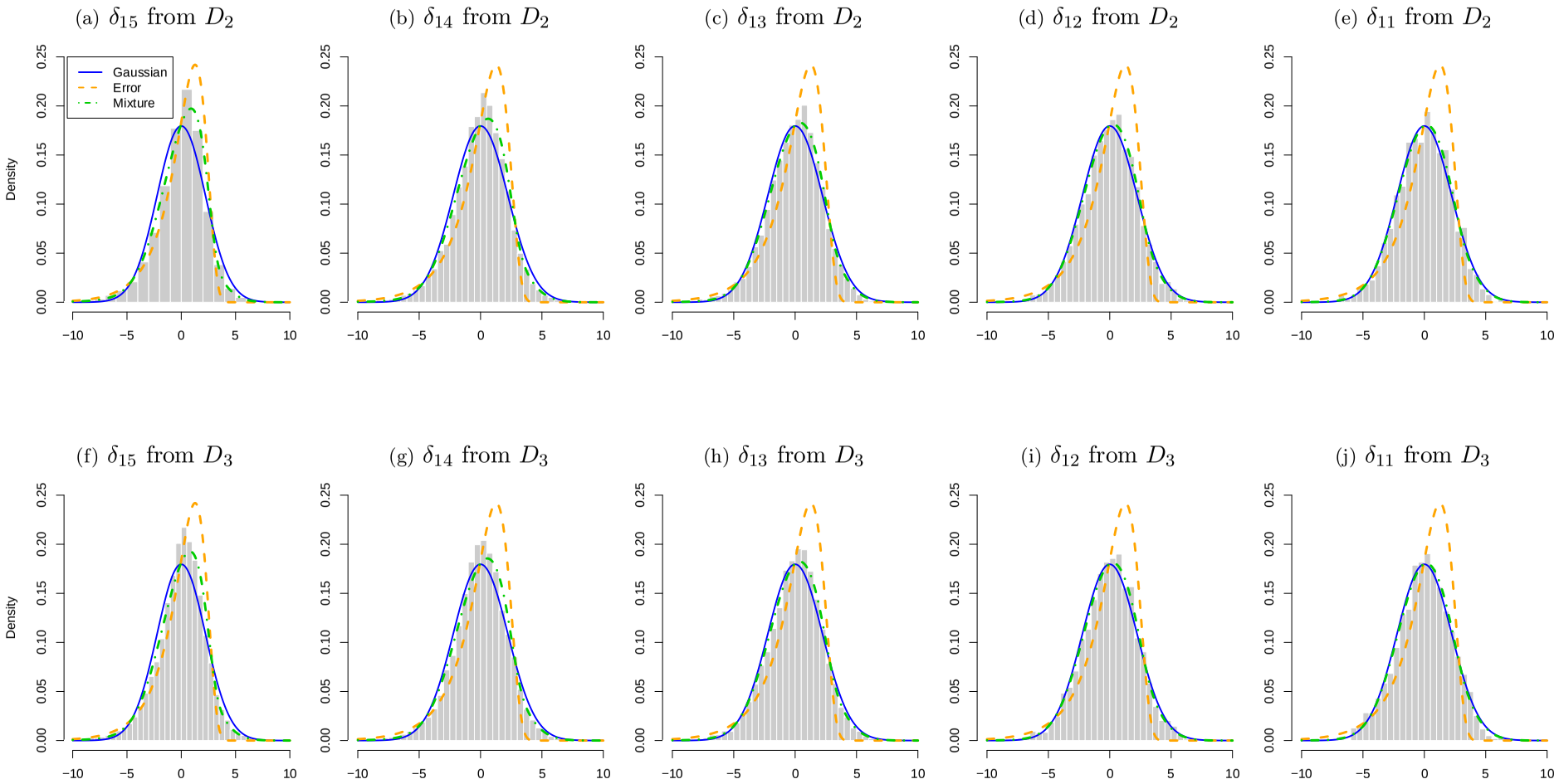


Figure 5.3: Density plots of δ_l for the five finest scales ($l = 15, \ldots, 11$), from the discrete D_2 (a-e) and D_3 (f-j) wavelet transform with plots of the Gaussian pdf (solid blue line), error pdf (orange dashed line) and likelihood (green dashed and dotted line) function calculated using empirical likelihood weights

5.2.2 Prior

As suggested by Pensky et al. (2007) the Berger-Müller prior from equation (2.36) was used on the discrete wavelet coefficients $\beta_{l,m}$. The DWT of a function in the Besov ball $B^r_{p,q}(\mathbb{R})$, for $p \geq 1$ and $q \leq \infty$ will yield a set of discrete wavelet coefficients which are sparse in the wavelet domain, as discussed in 2.4.2. Therefore, we choose the prior to be

$$p(\beta_{l,m}) = \alpha_l \, \delta(\beta_{l,m}) \; + \; (1 - \alpha_l) \, \nu_l \, \xi(\nu_l \; \beta_{l,m}), \tag{5.9}$$

where

- $\delta(\cdot)$ is the Dirac delta function,
- $\xi(\cdot)$ is a symmetric, uni-modal pdf,
- $\alpha_l \in [0, 1]$ is known as the *prior weight*, which is the probability that $\beta_{l,m}$ is zero at level l.

We assume that the $\{\beta_{l,m}\}$ are independent.

Let θ_l be the odds ratio at scale l, defined as

$$\theta_l = \frac{\alpha_l}{1 - \alpha_l}, \tag{5.10}$$

where α_l is the prior weight from equation (5.9).

5.2.3 Posterior Distribution

We perform Bayesian shrinkage upon the β's. By rearranging the regression model in equation (5.7), it can be shown that

$$\delta_{l,m} = h_{l,m} - \beta_{l,m}. \tag{5.11}$$

By substituting (5.11) into the likelihood equation (5.8) we can determine the posterior distribution of $\beta_{l,m}$. The form of the prior will mean our posterior distribution will consist of two parts, the dirac delta function and the tail density. The following calculations will only concern the tail density as this is the part of the distribution which will result in the shrinkage rule. How much of the posterior distribution is contained within the tail density will depend on α.

For notational simplicity, we shall drop the l, m subscripts on the β's and h's. Then the posterior tail distribution of β is

$$p(\beta|h) = \frac{\zeta_l(h-\beta)\,(1-\alpha_l)\,\nu_l\,\xi(\nu_l\beta)}{\int\,\zeta_l(h-x)\,\{\alpha_l\,\delta(x)\,+\,(1-\alpha_l)\,\nu_l\,\xi(\nu_l x)\}\,\mathrm{d}x}$$

$$= \frac{\zeta_l(h-\beta)\,\nu_l\,\xi(\nu_l\beta)}{\theta_l\,\int\zeta_l(h-x)\,\delta(x)\,\mathrm{d}x\,+\,\int\nu_l\,\zeta_l(h-x)\,\xi(\nu_l x)\,\mathrm{d}x}$$

$$= \frac{\zeta_l(h-\beta)\,\nu_l\,\xi(\nu_l\beta)}{\theta_l\,\zeta_l(h)\,+\,\nu_l\,\int\zeta_l(h-x)\,\xi(\nu_l x)\,\mathrm{d}x}, \tag{5.12}$$

for $h,\ \beta \in (-\infty, \infty)$. By combining the posterior tail density with the dirac delta function, we yield a posterior distribution which integrates to one.

5.2.4 Posterior Mean

Using (5.12) we can derive the posterior mean of β given h, denoted by $\hat{\beta}$, by

$$\hat{\beta} = \mathbb{E}[\beta|h] = \int y\ p(y|h)\mathrm{d}y$$

$$= \int \frac{y\,\zeta_l(h-y)\,\nu_l\,\xi(\nu_l y)}{\theta_l\,\zeta_l(h)\,+\,\nu_l\,\int\zeta_l(h-x)\,\xi(\nu_l x)\,\mathrm{d}x}\,\mathrm{d}y$$

$$= \int \frac{(1-\lambda_l)\,y\,\nu_l\,\xi(\nu_l y)\,\varphi_\sigma(h-y)\,+\,\lambda_l\,y\,\nu_l\,\xi(\nu_l y)\,f_E(h-y)}{\theta_l\,\zeta_l(h)\,+\,\nu_l\,\int\zeta_l(h-x)\,\xi(\nu_l x)\,\mathrm{d}x}\,\mathrm{d}y, \tag{5.13}$$

where $\sigma^2 = {}^{\pi^2}\!/_2$ (variance of the error distribution). Using the formula for the likelihood in (5.8) and by letting

$$Q_i(h) = \nu_l\,\int_{-\infty}^{\infty} y^i\,\xi(\nu_l y)\,\varphi_\sigma(h-y)\,\mathrm{d}y, \qquad i\ \in\ \mathbb{Z}^+; \tag{5.14}$$

$$Q_i^*(h) = \nu_l\,\int_{-\infty}^{\infty} y^i\,\xi(\nu_l y)\,f_E(h-y)\,\mathrm{d}y, \qquad i\ \in\ \mathbb{Z}^+; \tag{5.15}$$

the estimated posterior mean in equation (5.13) becomes

$$\hat{\beta} = \frac{(1-\lambda_l)\,Q_1(h)\,+\,\lambda_l\,Q_1^*(h)}{\theta_l\,\zeta_l(h)\,+\,(1-\lambda_l)\,Q_0(h)\,+\,\lambda_l\,Q_0^*(h)}. \tag{5.16}$$

This may appear complicated but depending on the form of the probability density function $\xi(\cdot)$, an analytical solution can be obtained for many cases.

5.2.5 Posterior Variance

To obtain the confidence intervals we also calculate the posterior variance of β. Although this can not be directly used to produce confidence intervals for the EWS, it could be used for importance sampling. Further details of the computational method for calculating the confidence interval can be found in section 6.1.

The posterior conditional variance of $\beta|h$ can be calculated using first and second moments with the formula

$$\begin{aligned} \text{Var}[\beta|h] &= \mathbb{E}[\beta^2|h] - (\mathbb{E}[\beta|h])^2 \\ &= \int_{-\infty}^{\infty} y^2 \; p(y|h) \; \mathrm{d}y \; - \left[\int_{-\infty}^{\infty} y \; p(y|h) \; \mathrm{d}y\right]^2. \end{aligned} \tag{5.17}$$

Section 5.2.4 already obtained the first moment ($\mathbb{E}[\beta|h] = \hat{\beta}$). The second moment can be obtained using equations (5.14) and (5.15) as

$$\mathbb{E}[\beta^2|h] = \frac{(1-\lambda_l)\, Q_2(h) \; + \; \lambda_l \, Q_2(h)}{\theta_l \, \zeta_l(h) \; + \; (1-\lambda_l)\, Q_0(h) \; + \; \lambda_l \, Q_0(h)}. \tag{5.18}$$

5.2.6 Hyperparameters

The hyperparameters in our model are the prior weights ($\alpha_l \in [0,1]$) and prior precision ($0 < \nu_l < \infty$ for $l = 0, \ldots, J-1$). In Pensky et al. (2007) the prior precision and weight values were fixed before analysis, and changed only with the sample size. The odds ratios from equation (5.10) were predefined as

$$\theta_l = 0.1 + \frac{0.8l}{J-1}, \tag{5.19}$$

where $J = \log_2(T)$ is the total number of scales, and $l = 0, \ldots, J-1$ is a particular scale, where $J-1$ is the finest scale and 0 being the coarsest. They also suggested that

$$\nu_l = (1-\lambda_l) \; (l+2), \tag{5.20}$$

where ν_l^{-2} is the prior variance. The form of this value ensured that the variance decreased as the scale increased and the wavelet coefficients became coarser. However, it was noted

that the variance of the coefficients changed greatly depending on the data. Therefore an alternative method was considered.

An alternative method is to use empirical Bayes methods which rely on determining the hyperparameters from the data. We shall use the method of Johnstone and Silverman (2005), described in section 2.4.2. The hyperparameters are calculated through numerical maximisation of the marginal likelihood (ML). Using equation (2.38), we substitute in our values for the error pdf (the likelihood function in (5.8)) and prior (the mixture prior from equation (5.9)), for our method, to produce our ML defined as

$$\sum_k \log\left\{\alpha_l \left[\frac{(1-\lambda_l)}{\sigma}\varphi\left(\frac{h_{l,m}}{\sigma}\right) + \lambda_l f_E(h_{l,m})\right] \; + \; (1-\alpha_l)\,\gamma(h_{l,m}|\nu_l)\right\}, \tag{5.21}$$

where $f_E(\cdot)$ is the error pdf, $\varphi(\cdot)$ is the standard Gaussian pdf,

$$\gamma(y|\nu_l) = \int_{-\infty}^{\infty} \nu_l\,\xi(\nu_l\,x)\left[\frac{(1-\lambda_l)}{\sigma}\varphi\left(\frac{\sqrt{T}[y-x]}{\sigma}\right) + \lambda_l f_E\Big(\sqrt{T}[y-x]\Big)\right]\,\mathrm{d}x, \tag{5.22}$$

and $\xi(\cdot)$ is the symmetric pdf chosen for the mixture prior. The only component of the ML not readily available is $\gamma(y|\nu_l)$, which will have to be calculated once a suitable unimodal symmetric pdf has been selected.

5.2.7 Bayesian Log-EWS Estimator

Recall for a particular scale j in equation (5.3) we defined $g(z) = \log(R(z))$, where $R(z)$ is the EWS multiplied by the inner product matrix A at location $z \in [0,1)$. By applying a DWT to our regression model in (5.3), the reconstruction of g is given by

$$g(z) = \beta_0\;\phi(z) \; + \; \sum_{l=0}^{\infty}\sum_{m=0}^{\infty} \beta_{l,m}\;\psi_{l,m}(z), \qquad z \in [0,1), \tag{5.23}$$

where $z = {}^{k}\!/_{T}$ and

$$\begin{aligned}
\psi_{l,m}(z) &= 2^{l/2}\,\psi(2^l k - m);\\
\beta_0 &= \int_0^1 \phi(z)\,g(z)\;\mathrm{d}z;\\
\beta_{l,m} &= \int_0^1 \psi_{l,m}(z)\,g(z)\;\mathrm{d}z;
\end{aligned}$$

and $\phi(\cdot)$ denotes the associated scaling function and $\psi(\cdot)$ is the associated wavelet function periodised on $[0,1]$.

Given the posterior mean estimate of the wavelet coefficients $\beta_{l,m}$ from section 5.2.4, let $\hat{\beta}_{l,m} = \beta_{l,m}$ in equation (5.23) which gives

$$\hat{g}(z) = \hat{\beta}_0\,\phi(z) + \sum_{l=0}^{J-1}\sum_{m=0}^{2^l-1}\hat{\beta}_{l,m}\,\psi_{l,m}(z), \tag{5.24}$$

where $\hat{\beta}_0$ is the posterior mean estimate of $g(z)$. We estimate $\hat{\beta}_0$ by h_0 as this is a weighted mean of all the observations H.

However, we wish to produce an estimate of the EWS, $S_j(z)$. Therefore, as $R_j(z) = \sum_l A_{j,l}S_l(z)$, our log Bayesian wavelet shrinkage estimator of the EWS at scale j is

$$\hat{S}_j(z) = \sum_l A_{j,l}^{-1}\exp\{\hat{g}(z)\}\,, \tag{5.25}$$

for $z \in [0,1)$ and $j = 0,\ldots,J-1$.

Once we obtained the estimator in (5.25), we noticed that due to the bias correction and log transformation required to obtain $\hat{S}_j(z)$, it was difficult to determine the variance of the EWS estimator analytically. Therefore, we concluded the simplest method to produce the credible intervals would be to numerically simulate observations from the posterior distribution via importance sampling. Further details on this methodology are given in chapter 6.

5.3 Choice of Tail Density Prior

We shall now discuss two possible choices of tail density priors: the Gaussian (in section 5.3.1) and Laplace (in section 5.3.2) distributions.

5.3.1 Gaussian Mixture Prior

The Gaussian distribution is one of the most well studied distributions in statistics, therefore we wondered if using the Gaussian distribution as part of the prior for the β's, the posterior quantities could be obtained analytically. The analytical solution could provide the means to understand how the Bayesian shrinkage was working. Pensky et al. (2007) do not use the Gaussian distribution for the prior because the Gaussian distribution does not possess heavy tails, although they mention that it is a possible candidate. Therefore, there is an extremely low prior probability that the β's will take very large or small values (Johnstone and Silverman, 2005).

The Gaussian part of the mixture prior has a distribution $\mathcal{N}(0,\, \nu_l^{-2})$. Let ξ equal to this Gaussian pdf in equations (5.14) and (5.15) to yield the following lemmas. The proofs of all the lemmas associated with the Gaussian mixture prior are available in appendix A.1.1.

Lemma 5.3.1. *The component $Q_i(h)$ with a Gaussian mixture prior is*

$$Q_i(h) = \frac{\nu_l}{\sqrt{2\pi\kappa_l}} \exp\left\{\frac{h^2}{\pi^2}\left(\frac{1}{\kappa_l} - 1\right)\right\} \int_{-\infty}^{\infty} y^i\, \varphi_{\sigma_g}(y - \mu_g)dy,$$

where

- $\sigma^2 = {}^{\pi^2}\!/_2$,
- $\kappa_l = 1 + (\nu_l\sigma)^2$,
- $\mu_g = h\kappa_l^{-1}$,
- $\sigma_g^2 = \sigma^2\kappa_l^{-1}$
- $\varphi_{\sigma_g}(\cdot)$ *is the Gaussian pdf with variance* σ_g^2.

Using the definition of f_E from (5.4), we can determine $Q_i^*(h)$ in the following.

Lemma 5.3.2. *The quantity $Q_i^*(h)$ for the Gaussian mixture prior is*

$$Q_i^*(h) = \frac{\nu_l}{\pi}\left(\frac{1}{2}\right)^{1/2} \int_{-\infty}^{\infty} y^i \exp\left\{-\frac{\nu_l^2\, y^2}{2}\right\} \sqrt{\gamma^*} \exp\left\{\frac{1}{2}(h - y) - \gamma^* e^{(h-y)}\right\}\, dy.$$

To solve the integral in lemma 5.3.2 for $i = 0, 1$, we require its Fourier transform, $\hat{Q}_i^*(h)$, given in lemma 5.3.3.

Lemma 5.3.3. *The Fourier transformation of $Q_i^*(h)$ is*

$$\hat{Q}_i^*(\omega) \;=\; (\gamma^*)^{i2\pi\omega} \frac{e^{-2(\pi\omega/\nu_l)^2}}{\sqrt{\pi}}\, \Gamma\left(\frac{1}{2} - i2\pi\omega\right) \int_{-\infty}^{\infty} y^i\, \nu_l\, \varphi\left(\nu_l\left[y + \frac{i2\pi\omega}{\nu_l^2}\right]\right)\, dy,$$

for $\omega \in [-\pi, \pi]$, where $\Gamma(\cdot)$ is the gamma function.

Gaussian Posterior Distribution, Mean and Variance

To calculate the posterior distribution, mean and variance obtained when using the Gaussian mixture prior we require the calculation of $Q_i(h)$ and $\hat{Q}_i^*(h)$ for $i = 0, 1, 2$. These we have calculated in the following two lemmas.

Lemma 5.3.4. *The component $Q_i(h)$ for the Gaussian mixture prior*

(a) for $i = 0$ is

$$Q_0(h) = \frac{\nu_l}{\sqrt{2\pi\kappa_l}} \exp\left\{\frac{h^2}{\pi^2}\left(\frac{1}{\kappa_l} - 1\right)\right\}$$

(b) for $i = 1$ is

$$Q_1(h) = \mu_g \frac{\nu_l}{\sqrt{2\pi\kappa_l}} \exp\left\{\frac{h^2}{\pi^2}\left(\frac{1}{\kappa_l} - 1\right)\right\},$$

(c) for $i = 2$ is

$$Q_2(h) = \frac{\nu_l}{\sqrt{2\pi\kappa_l}} \exp\left\{\frac{h^2}{\pi^2}\left(\frac{1}{\kappa_l} - 1\right)\right\}\left(\mu_g^2 + \sigma_g^2\right).$$

For each case κ_l, μ_g and σ_g^2 are as defined in lemma 5.3.1.

Lemma 5.3.5. *The component $\hat{Q}_i^*(\omega)$ for the Gaussian mixture prior*

(a) for $i = 0$ is

$$\hat{Q}_0^*(\omega) = (\gamma^*)^{i2\pi\omega} \frac{e^{-2(\pi\omega/\nu_l)^2}}{\sqrt{\pi}} \Gamma\left(\frac{1}{2} - i2\pi\omega\right)$$

(b) for $i = 1$ is

$$\hat{Q}_1^*(\omega) = -\frac{i2\omega\sqrt{\pi}}{\nu_l^2} (\gamma^*)^{i2\pi\omega} e^{-2(\pi\omega/\nu_l)^2} \Gamma\left(\frac{1}{2} - i2\pi\omega\right).$$

(c) for $i = 2$ is

$$\hat{Q}_2^*(\omega) = (\gamma^*)^{i2\pi\omega} \frac{e^{-2(\pi\omega/\nu_l)^2}}{\sqrt{\pi}} \Gamma\left(\frac{1}{2} - i2\pi\omega\right)\left(\frac{4\pi^2\omega^2}{\nu_l^4} + \frac{1}{\nu_l^2}\right).$$

For each case $\omega \in [-\pi, \pi]$ and $\Gamma(\cdot)$ is the gamma function.

Although an explicit representation for the $\hat{Q}_i^*(\omega)$ can be found, determining the correct values for ω proved to be more challenging. So to make progress, we determine $Q_i^*(h)$ for $i = 0, 1, 2$ through numerical integration.

Gaussian Hyperparameters

To determine the hyperparameters, the prior weight (α_l) and prior precision (ν_l), we numerically maximised the marginal likelihood from section 5.2.6 for the Gaussian mixture prior. This required the calculation of $\gamma(y|\nu_l)$ in (5.22). The first half of this integral could

be calculated by

$$\begin{aligned}
&\int \frac{\nu_l}{\sigma}\,\varphi(\nu_l\,x)\,\varphi\Big(\frac{x-y}{\sigma}\Big)\,\mathrm{d}x \\
&\quad= \frac{\nu_l\sigma_g}{\sigma}\left(\frac{1}{2\pi}\right)^{1/2}\int_{-\infty}^{\infty}\frac{1}{\sigma_g\sqrt{2\pi}}\exp\Big\{-\frac{1}{2\sigma_g^2}\left(x^2-2x\frac{y}{\kappa_l}+\frac{y^2}{\kappa_l}\right)\Big\}\,\mathrm{d}x. \\
&\quad= \frac{\nu_l\sigma_g}{\sigma}\left(\frac{T}{2\pi}\right)^{1/2}\exp\Big\{\frac{y^2}{\pi^2}\left(\frac{1}{\kappa_l}-1\right)\Big\}\int_{-\infty}^{\infty}\frac{1}{\sigma_g}\varphi\Big(\frac{1}{\sigma_g}\left(x-\frac{y}{\kappa_l}\right)\Big)\,\mathrm{d}x \\
&\quad= \frac{\nu_l\sigma_g}{\sigma}\left(\frac{T}{2\pi}\right)^{1/2}\exp\Big\{\frac{y^2}{\pi^2}\left(\frac{1}{\kappa_l}-1\right)\Big\}\,\Phi\Big(\frac{1}{\sigma_g}\left(x-\frac{y}{\kappa_l}\right)\Big).
\end{aligned}$$

where $\sigma^2 = {}^{\pi^2}\!/_2$, $\kappa_l = 1+(\nu_l\sigma)^2$ and $\sigma_g^2 = \sigma^2\kappa_l^{-1}$. However, the second half required numerical integration to evaluate the integral $f_E\Big(\sqrt{T}[y-x]\Big)$.

5.3.2 Laplace Mixture Prior

The literature suggests that using heavy-tailed mixture priors might achieve superior results, (Johnstone and Silverman, 2005). The pdf of the Laplace, $Lap(a,b)$, distribution is defined as

$$f_L(x|a,b) = \frac{b}{2}\exp\{-b|x-a|\} \qquad x \in (-\infty,\infty), \tag{5.26}$$

where a is the location and b is the scale parameter. This distribution has a mean of a and a variance of $2b^{-2}$.

If we let $\xi(\cdot)$ in the Berger-Müller prior defined in (5.9) equal the Laplace pdf from (5.26) with $a=0$ and $b=1$, then we can determine the values of $Q_i(h)$ from (5.14) and $Q_i^*(h)$ from (5.15) in the following lemmas. Proofs for all the lemmas associated with the Laplace mixture prior are available in appendix A.1.2.

Lemma 5.3.6. *The quantity $Q_i(h)$ for the Laplace mixture prior is*

$$Q_i(h) = \frac{\nu_l}{2}\,e^{-h^2/2\sigma^2}\left[e^{\mu_1^2/2\sigma^2}\int_{-\infty}^{0} y^i\,\varphi_\sigma(y-\mu_1)\;dy + e^{\mu_2^2/2\sigma^2}\int_{-\infty}^{0}(-y)^i\,\varphi_\sigma(y+\mu_2)\;dy\right],$$

where

- $\varphi_\sigma(\cdot)$ *is the Gaussian pdf with variance* σ^2,
- $\sigma^2 = {}^{\pi^2}\!/_2$,
- $\mu_1 = h+\nu_l\sigma^2$,
- $\mu_2 = h-\nu_l\sigma^2$.

Lemma 5.3.7. *The quantity $Q_i^*(h)$ for the Laplace mixture prior is*

$$Q_i^*(h) = \frac{\nu_l}{2\sqrt{\pi}}\left[\kappa_l \int_{\gamma^* e^h}^{\infty} (h - [\log(x) - \log(\gamma^*)])^i \, x^{-1/2-\nu_l} e^{-x} dx \right. \\ \left. + \kappa_l^{-1} \int_0^{\gamma^* e^h} (h - [\log(x) - \log(\gamma^*)])^i \, x^{-1/2+\nu_l} e^{-x} dx \right].$$

where $\kappa_l = (\gamma^*)^{\nu_l} e^{\nu_l h}$.

Laplace Posterior Distribution, Mean and Variance

To calculate the posterior distribution, mean and variance of the Laplace mixture prior we require $Q_i(h)$ and $Q_i^*(h)$ for $i = 0, 1, 2$. These quantities are given in lemmas 5.3.8 and 5.3.9, next.

Lemma 5.3.8. *The component $Q_i(h)$ of the Laplace mixture prior*

(a) for $i = 0$ is

$$Q_0(h) = \frac{\nu_l}{2} \, e^{-h^2/2\sigma^2} \left[e^{\mu_1^2/2\sigma^2} \, \Phi\Big(\frac{-\mu_1}{\sigma}\Big) + e^{\mu_2^2/2\sigma^2} \, \Phi\Big(\frac{\mu_2}{\sigma}\Big) \right].$$

(b) for $i = 1$ is

$$Q_1(h) = \frac{\nu_l}{2} \, e^{-h^2/2\sigma^2} \left[e^{\mu_1^2/2\sigma^2} \, \mu_1 \, \Phi\Big(-\frac{\mu_1}{\sigma}\Big) + e^{\mu_2^2/2\sigma^2} \, \mu_2 \, \Phi\Big(\frac{\mu_2}{\sigma}\Big) \right].$$

(c) for $i = 2$ is

$$Q_2(h) = \\ \frac{\nu_l}{2} e^{-h^2/2\sigma^2} \left[e^{\mu_1^2/2\sigma^2} [\sigma^2 + \mu_1^2] \Phi\Big(-\frac{\mu_1}{\sigma}\Big) + e^{\mu_2^2/2\sigma^2} [\sigma^2 + \mu_2^2] \Phi\Big(\frac{\mu_2}{\sigma}\Big) + \frac{3\sigma}{\sqrt{2\pi}} (\mu_2 - \mu_1) \right].$$

For all cases $\varphi_\sigma(\cdot)$, σ^2, μ_1 and μ_2 are as defined in lemma 5.3.6, and $\Phi(\cdot)$ is the standard Gaussian cumulative distribution function.

Lemma 5.3.9. *The evaluation of $Q_i^*(h)$*

(a) for $i = 0$ is

$$Q_0^*(h) \; = \; \frac{\nu_l}{2\sqrt{\pi}} \left[\kappa_l \, \Gamma\Big(\frac{1}{2} - \nu_l \gamma^* e^h\Big) + \; \kappa_l^{-1} \gamma\Big(\frac{1}{2} + \nu_l, \gamma^* e^h\Big) \right].$$

(b) for $i = 1$ is

$$Q_1^*(h) \;=\; Q_0^*(h)\,[h + \log(\gamma^*)] - \frac{\nu_l}{2\sqrt{\pi}}\left[\kappa_l \int_{\gamma^* e^h}^{\infty} \log(x)\, x^{-1/2-\nu_l}\, e^{-x}\, dy \right. \\ \left. + \;\kappa_l^{-1} \int_0^{\gamma^* e^h} \log(x)\, x^{-1/2+\nu_l}\, e^{-x}\, dx\right].$$

(c) for $i = 2$ is

$$Q_2^*(h) \;=\; \left(h + \frac{\log(\gamma^*)}{}\right)^2 Q_0^*(h) - 2\left(h + \frac{\log(\gamma^*)}{}\right) Q_1^*(h) \\ + \frac{\nu_l}{2\sqrt{\pi}}\left[\kappa_l \int_{\gamma^* e^h}^{0} \log(x)^2 x^{-1/2-\nu_l} e^{-x}\; dx \right. \\ \left. + \;\kappa_l^{-1} \int_{\gamma^* e^h}^{\infty} \log(x)^2 x^{-1/2+\nu_l} e^{-x}\; dy\right].$$

For all cases $\gamma(s,x)$ and $\Gamma(s,x)$ are the lower and upper incomplete gamma functions respectively.

The analytical answer of $Q_1^*(h)$ and $Q_2^*(h)$ consisted of Meijer's G function which is a path integral in the complex plane, and proved more difficult to evaluate than the original integral. Therefore, we used numerical integration to calculate these components.

Laplace Hyperparameters

To determine the hyperparameters α_l and ν_l (prior weight and precision) by maximising the ML for the Laplace mixture prior, let $\nu_l \xi(\nu_l\, x) = f_L(x|\nu_l)$ in equation (5.22). Then we can evaluate this integral as shown in lemma 5.3.10.

Lemma 5.3.10. *The function $\gamma(y|\nu_l)$ from equation* (5.22) *for the Laplace mixture prior is*

$$\gamma(y|\nu_l) \;= (1-\lambda_l)\;\frac{\nu_l}{2}\;e^{-y^2/2\sigma^2}\left[e^{\mu_3^2/2\sigma^2}\Phi\left(\frac{-\mu_3}{\sigma}\right) + e^{\mu_3^2/2\sigma^2}\Phi\left(\frac{\mu_4}{\sigma}\right)\right] \\ + \lambda_l\;\frac{\nu_l}{2\sqrt{\pi}}\left[\kappa_l\,\Gamma\left(\frac{1}{2} - \nu_l, \gamma^* e^y\right) + \frac{1}{\kappa_l}\,\gamma\left(\frac{1}{2} + \nu_l, \gamma^* e^y\right)\right],$$

where $\sigma^2 = {}^{\pi^2}\!/_2$, $\mu_3 = y + \nu_l\sigma^2$, $\mu_4 = y - \nu_l\sigma^2$ and $\kappa_l^* = e^{\nu_l y}(\gamma^*)^{\nu_l}$.

Using the results in lemma 5.3.10 we can obtain the ML and hence determine the Laplace hyperparameters via maximisation.

5.4 Asymptotics for the Bayesian Log-EWS Estimator

This section demonstrates the asymptotic convergence of our Bayesian wavelet shrinkage estimate of the log transformed raw wavelet periodogram. Our proof builds on the methods of Pensky et al. (2007), although our error distribution is different and our setting is non-stationary.

Section 5.4.1 refines our model in terms of $k = \lfloor zT \rfloor$, so we can demonstrate how increasing the frequency of observations can improve our EWS estimate. Section 5.4.2 details the assumptions made and finally in section 5.4.3 we describe our asymptotic results. Many of the proofs will consider the convergence of the scaled wavelet coefficients for fine, middle and coarse scales.

5.4.1 Asymptotic Model

Equation (5.23) in section 5.2.7, defines the expected log periodogram $g(z)$. If we let $z = {}^{k}\!/_{T}$, then

$$g(k) = \frac{1}{\sqrt{T}} b_0 \, \phi\left(\frac{k}{T}\right) \; + \; \frac{1}{\sqrt{T}} \sum_{l=-\infty}^{\infty} \sum_{m=0}^{\infty} b_{l,m} \, \psi_{l,m}\left(\frac{k}{T}\right), \qquad k = 0, \ldots, T-1, \quad (5.27)$$

where

- $\psi_{l,m}\left({}^{k}\!/_{T}\right) = 2^{l/2} \, \psi(2^l k - m)$;
- $b_0 = T^{-1/2} \int_0^T \phi\left({}^{k}\!/_{T}\right) \, g(k) \, \mathrm{d}k$;
- $b_{l,m} = T^{-1/2} \int_0^T \psi_{l,m}\left({}^{k}\!/_{T}\right) \, g(k) \, \mathrm{d}k$;

and as before, $\phi(\cdot)$ and $\psi(\cdot)$ are the scaling and wavelet function periodised on the interval $[0, 1]$. The coefficients $\beta_{l,m}$ from (5.7) and $b_{l,m}$ are related by $\beta_{l,m} \approx \sqrt{T} b_{l,m}$. Let $h^*_{l,m} = h_{l,m}/\sqrt{T}$ and $\upsilon_l = \sqrt{T}\nu_l$ (where ν_l is the prior precision from (5.9)), then we find

$$\mathcal{Q}_i(h^*_{l,m}) = \upsilon_l \sqrt{T} \int_{-\infty}^{\infty} y^i \, \xi(\upsilon_l^* y) \, \varphi_\sigma(h^*_{l,m} - y) \, \mathrm{d}y, \qquad i \in \mathbb{Z}^+, \quad (5.28)$$

$$\mathcal{Q}^*_i(h^*_{l,m}) = \upsilon_l \sqrt{T} \int_{-\infty}^{\infty} y^i \, \xi(\upsilon_l^* y) \, f_E(h^*_{l,m} - y) \, \mathrm{d}y, \qquad i \in \mathbb{Z}^+, \quad (5.29)$$

and

$$\hat{b}_{l,m} = \frac{(1 - \lambda_l) \, \mathcal{Q}_1(h^*_{l,m}) \; + \; \lambda_l \, \mathcal{Q}^*_1(h^*_{l,m})}{\theta_l \, \zeta_l(h^*_{l,m}) \; + \; (1 - \lambda_l) \, \mathcal{Q}_0(h^*_{l,m}) \; + \; \lambda_l \, \mathcal{Q}^*_0(h^*_{l,m})}. \quad (5.30)$$

We will prove the asymptotic convergence of $\hat{b}_{l,m}$, and due to the relationship between $b_{l,m}$ and $\beta_{l,m}$, we can infer the asymptotic convergence of $\hat{\beta}_{l,m}$.

5.4.2 Assumptions

As is conventional, we consider functions belonging to a restricted set of 'smoothness spaces' for a non-trivial theory. By assuming the function g, from equation (5.23), belongs to a ball, $B^r_{p,q}(A)$, in the Besov space, $B^r_{p,q}$ for $p \geq 1,\ q \leq \infty$ and non integer r, the Besov ball can be characterised in terms of wavelet coefficients

$$g \ \in \ B^r_{p,q}(A) \iff b_0^2 + \left\{ \sum_{l=1}^{\infty} 2^{l\left(r+\frac{1}{2}-\frac{1}{p}\right)q} \left(\sum_{m=0}^{2^l-1} |b_{l,m}|^p \right)^{q/p} \right\}^{1/q} \ \leq \ A, \tag{5.31}$$

as shown in Vidakovic (1999, section 6.4.1 on page 187).

When $g \in B^r_{p,q}(A)$ with $r > 1/p$ Pensky et al. (2007, equation (3.2)) shows that the wavelet coefficients, $b_{l,m}$, can be expressed as

$$\sum_{m=0}^{2^l-1} b_{l,m}^2 = \begin{cases} B_1 2^{-2rl} & \text{if } p \ \geq \ 2, \\ B_1 2^{-2l(r+\frac{1}{2}-\frac{1}{p})} & \text{if } 1 \leq \ p \ < \ 2, \end{cases} \tag{5.32}$$

for some $B_1 > 0$. When $p \geq 2$ the functions are spatially homogeneous, and when $1 \leq p < 2$ the functions are spatially inhomogeneous.

Let the risk of the estimator $\hat{g}$ of g, with T observations, over the set $\mathcal{F}$, be defined as

$$R(T, \hat{g}, \mathcal{F}) = \sup_{g \ \in \ \mathcal{F}} \mathbb{E}\Big[\|g - \hat{g}\|^2_{L^2[0,1]} \Big]. \tag{5.33}$$

We shall assume the scaling and wavelet functions we use are s-regular with $s \geq r$. Also, assume that the $\xi(\cdot)$ component in the Berger-Müller prior is three times differentiable (in at least a piecewise sense), has a finite fourth moment and satisfies the conditions

$$\left| \frac{\xi^i(x)}{\xi(x)} \right| \leq \mathcal{C}_{\xi,1}(1 + |x|^\kappa)^i, \quad i = 1, 2, 3, \kappa \geq 0, \tag{5.34}$$

$$\mathcal{C}_{\xi,2} \exp\left\{ -\frac{x^2}{2\sigma^2} \right\} \leq \xi(x) \leq \xi_0, \tag{5.35}$$

for positive constants $\mathcal{C}_{\xi,1}$ and $\mathcal{C}_{\xi,2}$. Let the integrals $\mathcal{Q}_i(h^*)$ for $i = 1, 2$ from equation

(5.28) be such that, if $\upsilon_l/\sqrt{T} \to 0$ and $\upsilon_l|h^*| \to \infty$, then assume

$$\left|\frac{\mathcal{Q}_1(h^*)}{\mathcal{Q}_0(h^*)} - h^*\right| \;=\; \mathcal{O}\left(\frac{\upsilon_l^2|h^*|}{T}\right), \tag{5.36}$$

and the distribution of $\mathcal{Q}_0(h^*)$ is

$$\mathcal{Q}_0(h^*) \;\sim\; \upsilon_l \xi(\upsilon_l h^*), \tag{5.37}$$

where $\xi(\cdot)$ is the pdf of the tail density prior. Also, if $\upsilon_l/\sqrt{T} \to \infty$, then assume

$$\left|\frac{\mathcal{Q}_1(h^*)}{\mathcal{Q}_0(h^*)}\right| = \mathcal{O}\left(\frac{T|h^*|}{\upsilon_l^2}\right). \tag{5.38}$$

Let $p^* = \min(p, 2)$ then if $J = \log_2(T)$ and

$$J_0 = J(2r+1)^{-1}, \tag{5.39}$$

$$J_1 = 2rJ\left[(2r+1)\left(2r+1-\frac{2}{p^*}\right)\right]^{-1}, \tag{5.40}$$

$$J_2 = J\left(\frac{1}{2} + r\left[(2r+1)\left(2r+1-\frac{2}{p^*}\right)\right]^{-1}\right). \tag{5.41}$$

These are used to determine coarse scales ($0 \le l \le J_0$), intermediate scales ($J_0 < l \le J_1$), fine scales ($J_1 < l \le J_2$) and finest scales ($l > J_2$). Note that if $p = 2$ the intermediate scales vanish.

We also define

$$r^* = \frac{1}{2}\left[\left(\frac{1}{p^*} - \frac{1}{2}\right) + \sqrt{\left(\frac{1}{p^*} - \frac{1}{2}\right)^2 + 2\left(\frac{1}{p^*} - \frac{1}{2}\right)}\right]. \tag{5.42}$$

Assume that the prior precision for different scales can be expressed as

$$\upsilon_l = 2^{\mu l} \qquad \text{with} \quad \mu = \begin{cases} \mu_1 = r + \frac{1}{2} + \frac{1}{4} - \frac{1}{2p^*}, & 0 \le l \le J_0, \\ \mu_2 = (r + \frac{1}{2} - \frac{1}{p^*}) + \frac{1}{r}(\frac{r}{2} + \frac{1}{2} - \frac{1}{p^*}), & J_0 < l \le J_1, \\ \mu_3 = r + \frac{1}{2}, & J_1 < l. \end{cases} \tag{5.43}$$

We also assume

$$\theta_l^2 = \mathcal{O}\left(2^{\left(4r+1+3\left[\frac{1}{2}-\frac{1}{p^*}\right]\right)(J-l)} T^{-(4r+1)/(2r+1)}\right), \qquad \text{if } l \leq J_0, \tag{5.44}$$

$$\theta_l^{-2} = \mathcal{O}\left(T^{-\left(\frac{1}{p^*}-\frac{1}{2}+\epsilon\right)/\left(\left[r+\frac{1}{2}\right]\left[r+\frac{1}{2}-\frac{1}{p^*}\right]\right)}\right), \qquad \text{if } J_0 < l \leq J_1, \tag{5.45}$$

for some $\epsilon > 0$.

5.4.3 Asymptotic Results

To prove the asymptotic convergence of $\hat{\beta}_{l,m}$, we want to show that as we increase T the risk decreases.

Theorem 5.4.1. *Let $r > r^*$ (5.42), (5.34)-(5.37) and (5.43)-(5.45) be valid. Then*

$$R(T, \hat{g}, \mathcal{F}) = \mathcal{O}(\log(T)^{\alpha} T^{-2r/(2r+1)}), \qquad \text{as } T \to \infty, \tag{5.46}$$

where $\alpha = (2r+1)^{-1}$ if $p \geq 2$ and $\alpha = 2r(2r+1)^{-1}$ if $1 \leq p < 2$. If equation (5.38) holds, and $p \geq 2$ then $\alpha = 0$.

The proof of theorem 5.4.1 can be found in appendix A.1.3.

Remark 5.4.1. As $\upsilon_l = \sqrt{T}\nu_l$, the assumption in (5.43) can be translated to a restrictions on ν_l, such that $\nu_l = 2^{\mu l} T^{-1}$.

Remark 5.4.2. The condition that $\xi(\cdot)$ has a finite fourth moment is a technical condition so that we can obtain an asymptotic expansion for $\mathcal{Q}_i(h)$ for $i = 0, 1$. However this condition can be dropped and replaced with lemmas A.1.1 and A.1.2.

Remark 5.4.3. Condition (5.41) is quite realistic, and agrees with the Central Limit Theorem. If $r > r^*$, we have an infinite number of scales,

$$J - J_2 = \log_2(T)\left[4r^2 + 2r\left(1\frac{2}{p^*}\right) + 1\frac{2}{p^*}\right]\left[(2r+1)\left(2r+1\frac{2}{p^*}\right)\right]$$

until the Central Limit Theorem takes over. In practice, the assumption of Gaussianity can be verified via scale-by-scale testing.

Chapter 6

Simulation Results of the Bayesian Log-EWS Estimator

In this chapter we describe and test the computational performance of our method. We use the analytical calculations from sections 5.3.1 and 5.3.2 in chapter 5 to calculate the key components required to perform Bayesian wavelet shrinkage of the log transformed raw wavelet periodogram.

We begin in section 6.1 by describing the computational process developed in the statistical package `R` to produce the EWS estimate and corresponding confidence intervals. We compare our methods using the Gaussian or Laplace prior. The best prior was selected to test our method on the simulated data from chapter 3. In section 6.3 we examine the results and discuss its performance.

6.1 Programming

The code to perform Bayesian wavelet shrinkage of the log transformed raw wavelet periodogram was written in `R`. It required the use of programs from the libraries `wavethresh`, `TSA`, `zipfR` and `NORMT3`. Figure 6.1 is a flow diagram of the operations required to produce an estimate of the EWS via Bayesian wavelet shrinkage of the log transformed wavelet periodogram and confidence intervals. We shall explain in detail next, how each operation described in 6.1 is performed computationally.

Assume that the user will provide the appropriate analysis wavelet (AW) which is used to obtain the raw wavelet periodogram from the user supplied LSW process of dyadic length. For a particular scale, log and centralise the raw wavelet periodogram as in (5.3),

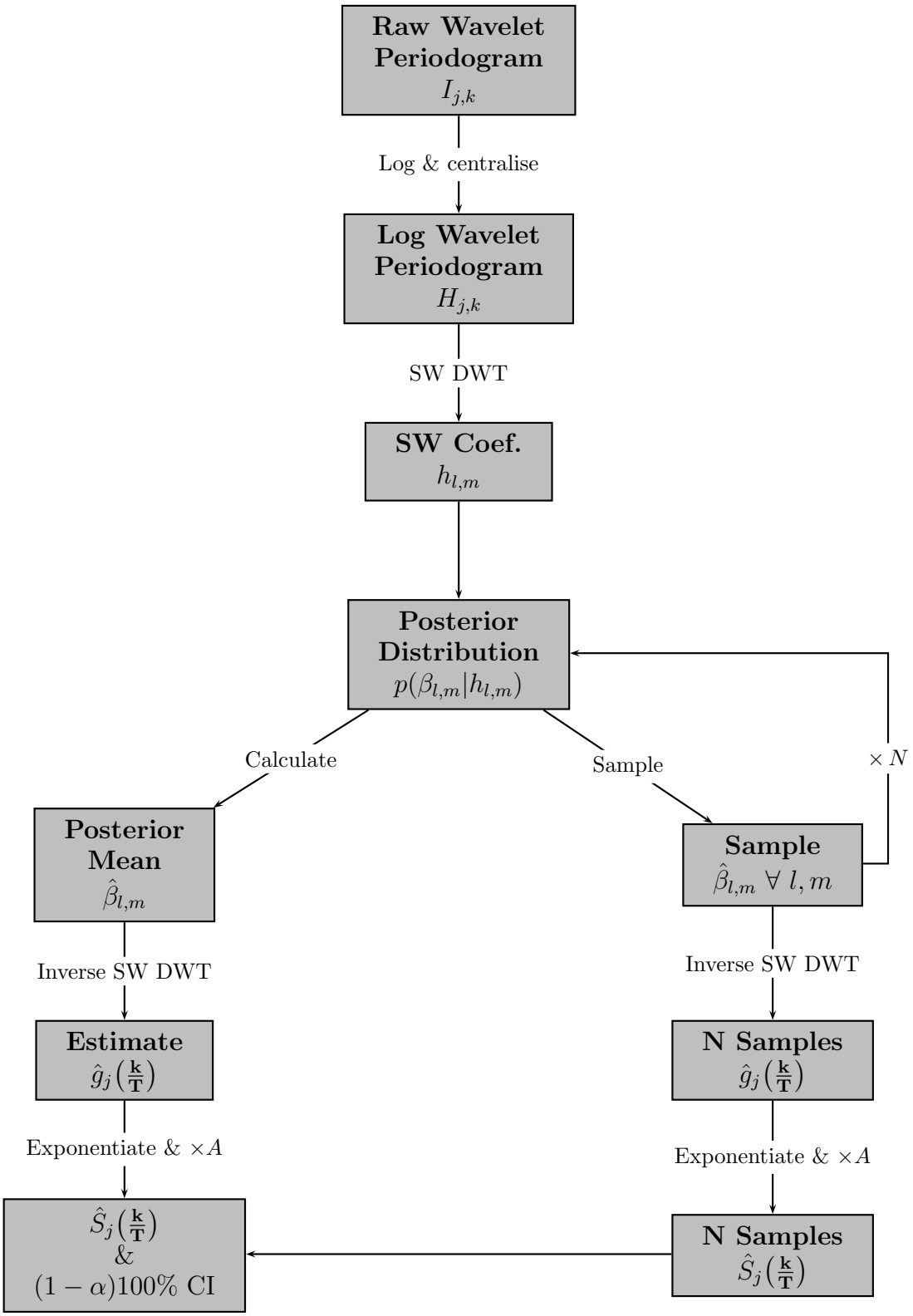


Figure 6.1: A flow diagram of how Bayesian wavelet shrinkage and sampling can estimate the EWS from the log raw wavelet periodogram with confidence intervals for a particular smoothing wavelet (SW).

then perform the DWT with the smoothing wavelet (SW), also defined by the user.

The next stage stage is to determine the hyperparameters. The $\gamma(\cdot)$ component of the marginal log-likelihood(MLL) in (5.21) required numerical integration in the case of the Gaussian mixture prior, but could be evaluated for the Laplace prior as shown in lemma 5.3.10. Then the MLL was maximised using the program `optim` and the `L-BFGS-B` method (Byrd et al., 1995). This method was used as it allows for lower and upper constraints. There were a few issues to consider when determining the hyperparameters:

1. if we used the Laplace mixture prior, $\lambda_l > 0 \Rightarrow \nu_l$ must be less than $1/2$;
2. for coarse scales ($l < 4$) there are very few coefficients, resulting in an insufficient number of wavelet coefficients to estimate hyperparameters;
3. small sample sizes and sparse data (many coefficients are close to zero) resulted in non-numeric values when maximising over the entire log likelihood.

If $\lambda > 0$, we needed to calculate $Q_i^*(h)$. For the Laplace mixture prior, lemma 5.3.9 shows the calculation of $Q_i^*(h)$ consists of the incomplete gamma function, $\Gamma(s, x)$. As $s = 1/2 - \nu_l$, if $\nu_l > 1/2$ this results in negative s. Hence, $Q_0^*(h)$ is unsolvable, and consequently $Q_1^*(h)$ and $Q_2^*(h)$, as they also consist of $Q_0^*(h)$. Therefore if $\lambda > 0$, the search area for ν_l was limited to $(0, 1/2]$ and if $\lambda = 0$, $\nu_l \in (0, \infty)$ for the Laplace mixture prior.

The four coarsest scales, $l = 0, 1, 2, 3$, consist of $1, 2, 4$ and 8 wavelet coefficients, respectively. Empirical investigations showed that maximising the MLL for each scale resulted in biased hyperparameter estimates. Therefore, instead of maximising the log-likelihood for the four coarsest scales separately, the coefficients were grouped together and maximisation was performed over all the four scales. To distinguish between scales, the hyperparameter estimates were scaled appropriately, such that as the scale decreased α_l decreased and ν_l increased by a factor of two.

For some scales the numerical maximisation of the MLL would not converge because wavelet coefficient are sparse, i.e. there was insufficient information in the data. To solve this problem instead of maximising over the entire MLL, we maximised over equation (2.39).

Once we have appropriate hyperparameters, we use these to obtain the 'shrunk' wavelet coefficients, $\hat{\beta}_{l,m}$, (5.16). If we use the Gaussian mixture prior, $Q_i(h)$ for $i = 0, 1$ is calculated using lemma 5.3.4 parts (a) and (b), and $Q_i^*(h)$ for $i = 0, 1$ is determined via

numerical integration of lemma 5.3.2. When using the Laplace mixture prior, $Q_i(h)$ for $i = 0, 1$ is solved via lemma 5.3.8 parts (a) and (b), and $Q_0^*(h)$ is calculated using lemma 5.3.9(a), whereas $Q_1^*(h)$ requires numerical maximisation to solve lemma 5.3.9(b).

The posterior variance (5.17) of $\beta_{l,m}$ is calculated using lemmas 5.3.2 and 5.3.4 or 5.3.8 and 5.3.9 (depending on the mixture prior). This could then be could be used to calculate the variance of our estimate $\hat{g}(z)$ from equation (5.24). However, to calculate $\text{Var}[\hat{g}(z)]$ requires the squared wavelets ($\psi_{l,m}^2$). Barber et al. (2001) demonstrated that $\psi_{l,m}^2$ could be estimated through the scaling function ($\phi_{l,m}$). This can be easily done for the Haar wavelet as $\psi_{l,m}^2 = 2^{l/2}\,\phi_{l,m}$, however for other wavelets (such as Daubechies Extremal phase wavelets) this is more challenging and computationally expensive.

Also, although we could produce the variance for the wavelet coefficients and consequently $\hat{g}(z)$, our goal is to obtain a confidence interval for the EWS $S_j(z)$. This proved to be challenging as $\hat{g}_j(z) = \log(\hat{R}_j(z))$ and $\hat{S}_j(z) = (A^{-1}\hat{R})_j(z)$, the log transformation and inner product matrix correction makes the variance and confidence interval calculations difficult. Hence, as an alternative, we use importance sampling to produce N possible estimates of the EWS. As we "knew" the posterior distribution (5.12) of the wavelet coefficients, we sampled these coefficients to obtain the sampled EWS.

To improve our estimates and to attempt to make them less susceptible to Gibbs phenomena we used cycle spinning (Coifman and Donoho, 1995). Before performing the discrete wavelet transform (DWT) using the smoothing wavelet, the raw wavelet periodogram was shifted n places. Then, Bayesian wavelet shrinkage was performed on the wavelet coefficients obtained from using the SW DWT of the cycle spun data. After applying the SW inverse DWT, the estimated wavelet periodogram was rcycle spun to return the estimated data to their original location.

6.2 Comparison of the Gaussian and Laplace Prior

In this section we compare using the Gaussian and Laplace Berger-Müller prior to perform Bayesian wavelet shrinkage on the log-transformed raw wavelet periodogram. We begin by comparing the difference between the Gaussian and Laplace distribution. Figure 6.2 shows a plot of the Gaussian ($\mathcal{N}(0,1)$) and Laplace, $Lap(0, \sqrt{2})$ pdf, both with variance 1. From the plot we can see the Laplace distribution has slightly heavier tails. This implies there will be a slightly higher acceptance rate of potential wavelet coefficients (β), which is important for numerical calculations of Bayesian statistics (i.e. MCMC).

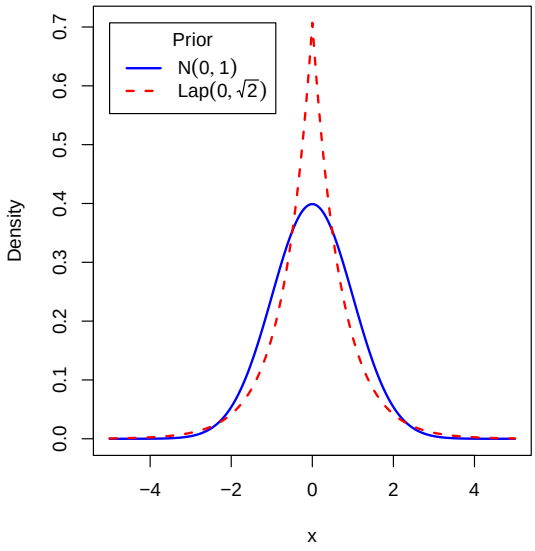


Figure 6.2: Plot of the standard Gaussian and Laplace pdf both with variance 1.

The plots in figures 6.3 (a), (c) and (e) are the Gaussian and Laplace posterior tail distribution (5.12) with

- data $h = {}^1\!/_2$, from (5.7)
- odds ratio $\theta_l = {}^1\!/_5$, from (5.10)
- prior weights $\lambda_l = 0, {}^1\!/_2, 1$
- Gaussian prior precision $\nu_l = {}^1\!/_4$ and Laplace prior precision $\nu_l = {}^{\sqrt{2}}\!/_4$ (this ensure both priors have equal variance).

The value for θ corresponds to $\alpha = \theta(1+\theta)^{-1} = {}^1\!/_6$, implying most of the weight of the posterior distribution is present in the posterior tail distribution. As θ increases, we find α decreases and most of the weight of the distribution will be concentrated at zero via the dirac delta function. As $\alpha \to 0$ more wavelet coefficients will be 'shrunk' to zero as demonstrated by the shrinkage function in figures 6.3 (b), (d) and (f). The shrinkage function is the posterior mean from equation (5.16), as we vary the value of h, with fixed $\theta_l = 5$, Laplace $\nu_l = {}^{\sqrt{2}}\!/_{100}$ and Gaussian $\nu_l = {}^1\!/_{100}$. If ν_l increases, this also increases the amount of shrinkage performed on the h, i.e. $\nu_l \uparrow \Rightarrow \ \beta_{l,m} \downarrow$.

As the likelihood weight (λ) increases, the likelihood function becomes more skewed with an increase in influence from the error distribution (centralised $\log(\chi^2_1)$). This affects the skewness of the posterior distribution and the symmetry of the shrinkage function. When $\lambda = 0$, the posterior distribution and shrinkage function are symmetric. In figure 6.3 (b) values of h between approximately -5 and 5 result in a posterior mean estimate

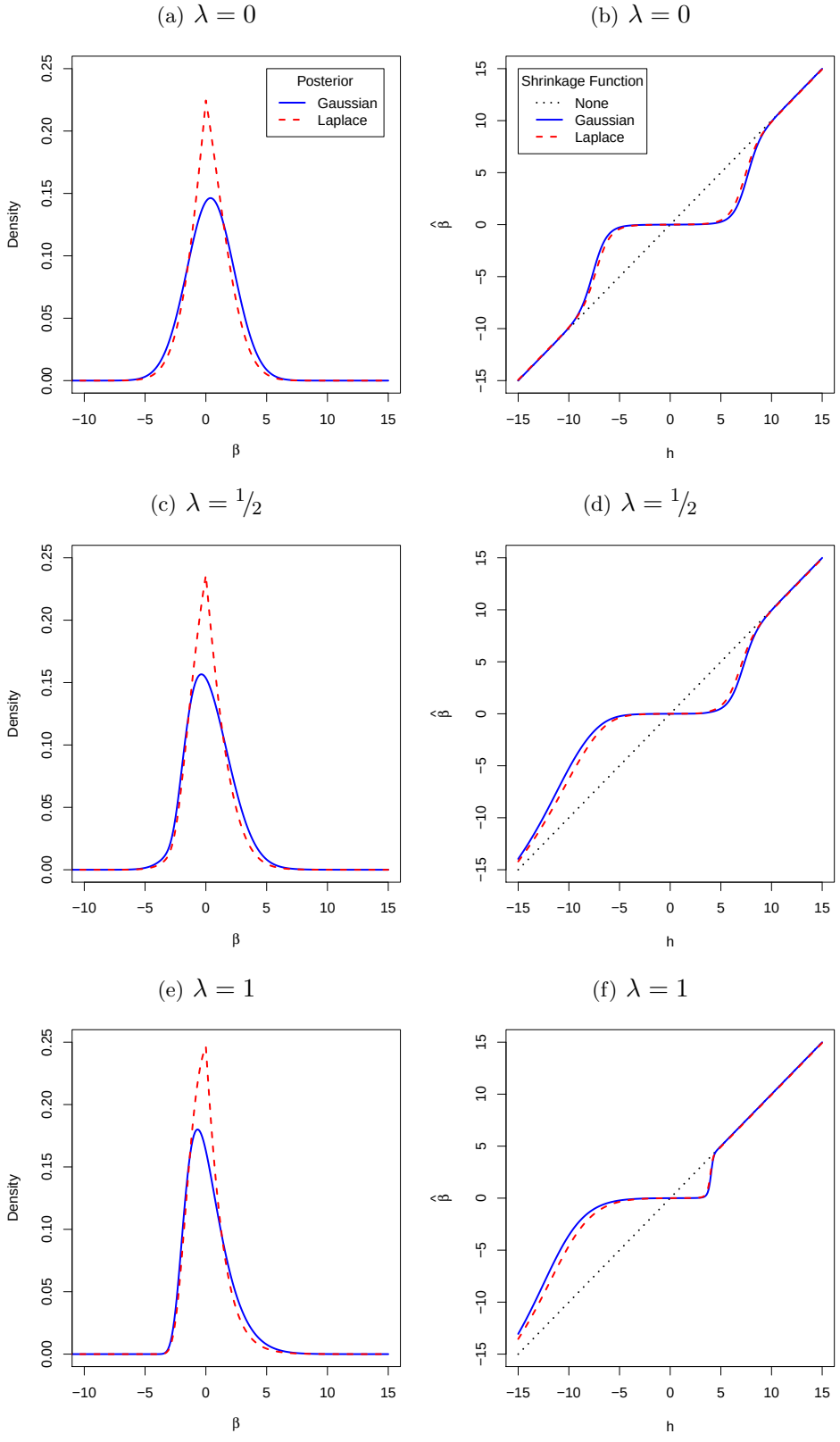


Figure 6.3: (a), (c) and (e) are of the Gaussian and Laplace posterior tail distribution for $h = 1/2$, $\theta = 1/5$ for both, Gaussian $\nu = 1/4$ and Laplace $\nu = \sqrt{2}/4$ for different values of λ. Plots (b), (d) and (f) are the Gaussian and Laplace shrinkage functions for $\theta = 5$ for both, and Gaussian $\nu = 1/100$ and Laplace $\nu = \sqrt{2}/100$ and different λ.

$\hat{\beta} \approx 0$. This implies that values of $|h| < 5$ are likely due to noise in the data. If $|h| \in [5, 10)$, the estimate of $|\hat{\beta}|$ are smaller than $|h|$. In figures 6.3 (d) and (f) of the shrinkage function for $\lambda > 0$, the range of values of $h < 0$ which results in $\hat{\beta} = 0$ increases. There is also a decrease in the range of values of $h > 0$ which yield a $\hat{\beta}$ value equal to zero.

Although the posterior distributions appear to be very different between the Laplace and Gaussian prior, there is very little difference between the shrinkage functions. Each plot of the shrinkage functions suggest that using the Gaussian prior will result in a slightly higher proportion of values being reduced to zero.

Calculation of $Q_i^*(\cdot)$ for the Gaussian prior required numerical integration. An explicit representation could be produced in the frequency domain (as shown in lemma 5.3.5), but how to obtain ω from h was not clear. Although numerical integration was also required to calculate $Q_1^*(\cdot)$ for the Laplace prior (lemma 5.3.9(b)), it was only needed for one component. We conclude the computational cost of using the Gaussian prior is greater than the Laplace. Hence from now on, only the Laplace prior is used to evaluate the performance of Bayesian wavelet shrinkage of the log transformed raw wavelet periodogram to estimate the EWS.

6.3 Simulation Example

To evaluate the performance of our method, we used one simulated locally stationary wavelet (LSW) processes with Gaussian innovations, X_t and Y_t, described in section 3.4 simulated from the piecewise constant and slowly evolving spectra of (3.9) and (3.12). We assumed the Haar synthesis wavelet was known, and all of our analysis was performed on the raw wavelet periodogram obtained using the Haar AW.

Firstly, for both LSW processes, we empirically verified the distribution of our model's error term from equation (5.3) was approximately the central $\log(\chi_1^2)$ distribution (error distribution from lemma 5.1.1). To obtain our error, the true value of $\log(R_j(z))$ was calculated using equations (3.9) and (3.12), and the inverse Haar inner product matrix A. This was subtracted from $H_{j,k}$ (the error centralised log transformed raw wavelet periodogram) to obtain $e_{j,k}$. The pdf of the error distribution was imposed over a histogram of $e_{j,k}$ of the finest ($j = 9$), mid ($j = 6$) and coarsest ($j = 0$) scale.

Estimation of the piecewise constant and slowly evolving EWS from the simulated LSW process X_t and Y_t, was produced using the smoothing wavelets Haar and Daubechies Extremal Phase wavelets with $2-10$ vanishing moments. For each estimator, we calculated

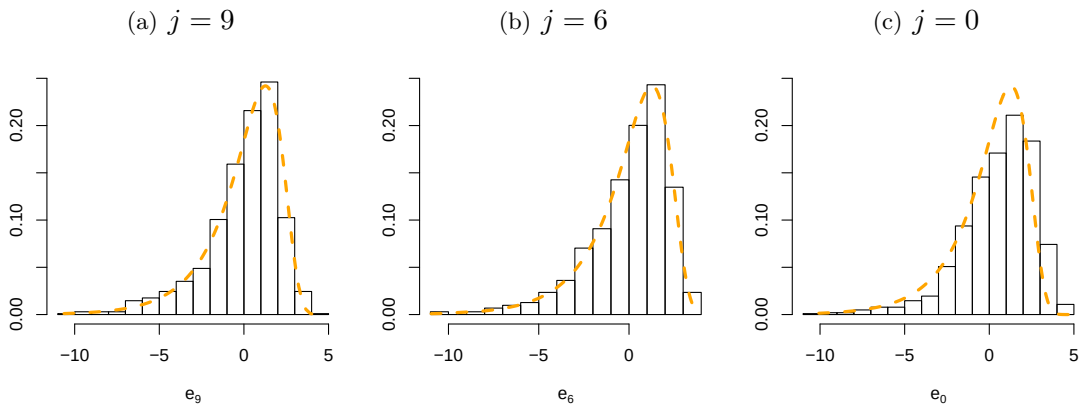


Figure 6.4: Histogram plots of $e_{j,k} = H_{j,k} - \log(R_{j,k})$ obtained from the LSW process X_t with Gaussian innovations, simulated from the piecewise constant spectrum for scales $j = 9, 4, 0$, and probability density plot of the error distribution (yellow dashed line).

the average mean squared error (AMSE), defined as

$$\text{AMSE} = \frac{1}{JT} \sum_{j=0}^{J-1} \sum_{k=0}^{T-1} \left(S_{j,k} - \hat{S}_{j,k} \right)^2,$$

without cycle spinning.

The SW which produced the estimator with the smallest AMSE was then used to produce the estimator which was cycle spun twenty and forty times. As visually there was little difference between the results for the different number of cycle spins, we concluded for computational efficiency, twenty cycle spins would be sufficient (appendix B contains the results for forty cycle spins). For each EWS estimate, two hundred Monte-Carlo samples were generated to use in calculating the 50% and 90% quartiles.

6.3.1 Piecewise Constant EWS

We begin our investigation by empirically assessing the distribution of the error term. Figure 6.4 shows the histogram plots of the model error of the finest (a), mid (b) and coarsest (c) scale of the log transformed wavelet periodogram. For $j = 9$ and 6 the error pdf fits the data reasonably well. The histogram of the error at the coarsest scale does not fit the centralised $\log(\chi^2_1)$ as well, but the departure is not large enough to suggest our error assumption is unreasonable.

The SW which produced an estimate of the piecewise constant EWS with the smallest AMSE was the Daubechies Extremal Phase wavelet with 3 vanishing moments (D_3). We wish to produce an EWS estimate where the majority of the power is contained within the finest scale $j = 9$ and there is evidence of the piecewise constant nature. The plots of

(a) No Cycle Spinning $\hat{S}(z)$ (b) 20 Cycle Spins $\hat{S}(z)$

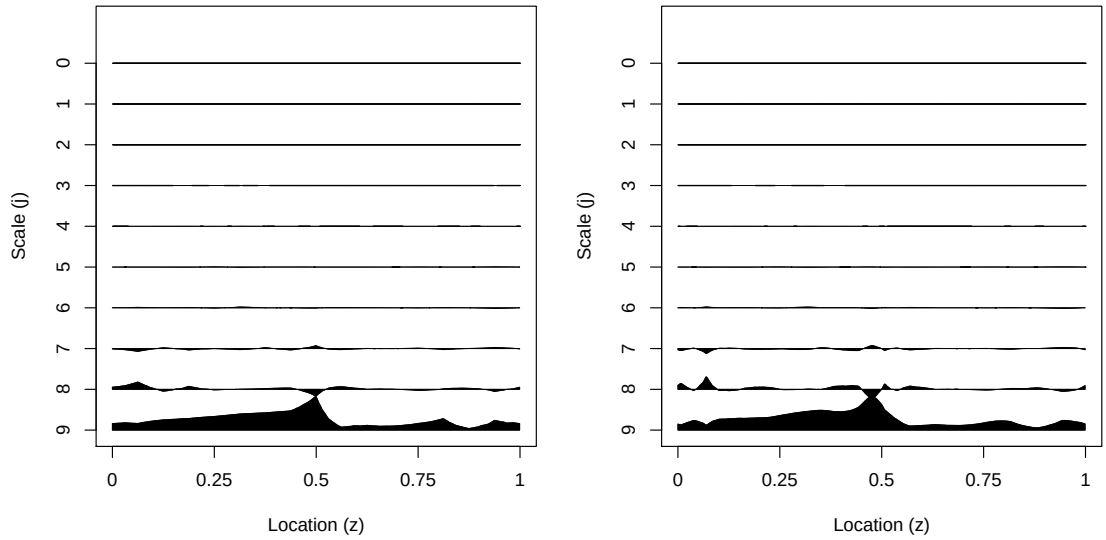


Figure 6.5: Plots of the estimated piecewise constant EWS using Bayesian wavelet shrinkage estimate of the log-wavelet periodogram obtained from X_t.

$\hat{S}(z)$ in figure 6.5 do not capture the piecewise constant nature of the spectrum and for the estimate without cycle spinning (a) and with 20 spins (b), there is evidence power has leaked from the finest ($j = 9$) scale into the next two finest scales ($j = 8, 7$). Comparing these plots with figure 3.4(a), which is the TI de-noised estimator of the same data, there appears to be no improvement. In the Bayesian log-EWS estimator there is more leakage from the finest into the adjacent coarser scales than the TI de-noised estimator.

We plotted the EWS estimate, confidence intervals and true EWS for scales $j = 9, 8, 7$ in figure 6.6. Figures (a), (c) and (e) are the estimates without cycle spinning, whereas (b), (d) and (f) are the cycle spun estimates of $S_j(z)$. There is a slight improvement in each scale estimate after cycle spinning the data, and an increase in the coverage of the CIs.

In figures 6.6 (e) and (f), perform reasonably well at detecting the second change point at $z = {}^{8}/_{15}$, possibly because it is a larger change in power than the first and the long period where the power remains constant. Both estimates fail to detect the first change point at $z = {}^{1}/_{3}$, instead implying a gradual increase in power until a sudden peak around $z \approx {}^{1}/_{2}$.

The 90% CI for the plots in figures 6.6 (a), (b), (c) and (d), where there should be no power, all appear to capture the true spectral value of zero. This information could be used to suggest any evidence of power at these scales is a result of leakage and potentially could be ignored.

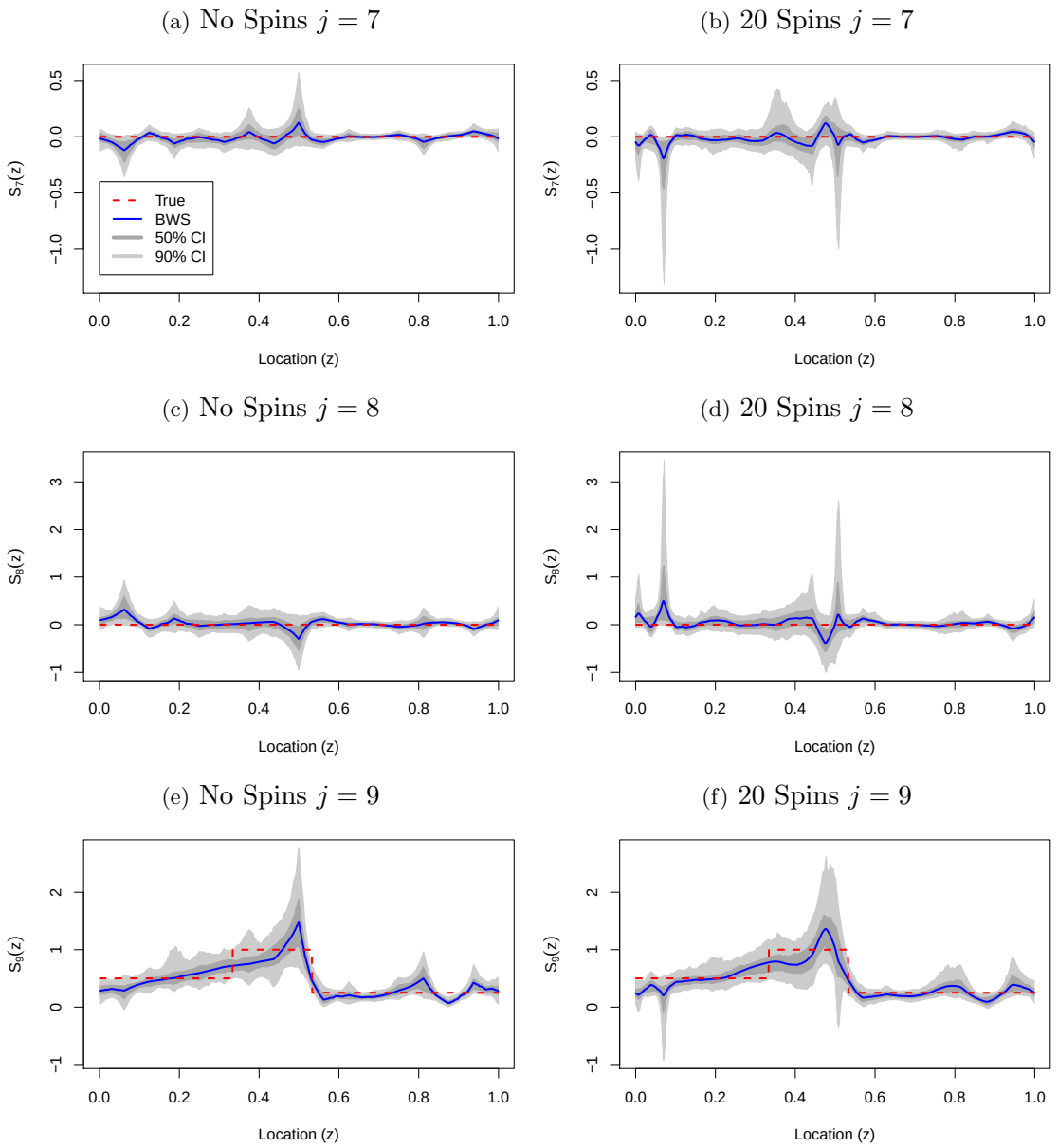


Figure 6.6: Plots of the piecewise constant EWS estimator for scales $j = 9, 8, 7$. Figures (a), (c) and (e) are estimated without cycle spinning, and figures (b), (d) and (f) are with 20 cycle spins. Analysis wavelet = Haar and smoothing wavelet = D_3. The true piecewise constant EWS is the dashed red line and our log-EWS estimator is the blue solid line. The dark grey shaded area is the 50% CI and the light grey is the 90% CI.

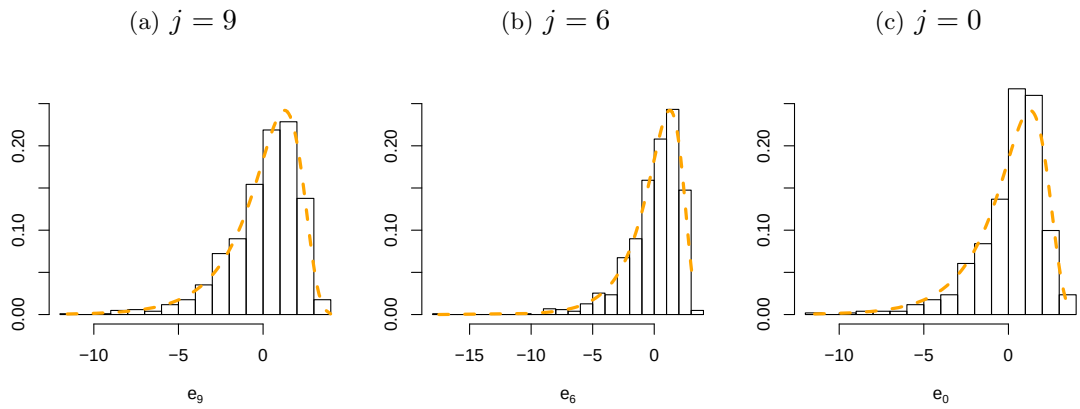


Figure 6.7: Histogram plots of $e_{j,k} = H_{j,k} - \log(R_{j,k})$ obtained from the Y_t with Gaussian innovations, simulated from the slowly evolving spectrum for scales $j = 9, 4, 0$.

As we were aware of the piecewise constant nature of the spectrum we performed the analysis again using the Haar smoothing wavelet. Without cycle spinning the estimate was also piecewise constant. However, the location of the change points were completely inaccurate, and cycle spinning the data produced results very similar to the D_3 estimates, (see appendix B)

6.3.2 Slowly Evolving EWS

Similarly to the piecewise constant spectrum, for scales $j = 9, 6, 0$ we plot the histograms of $e_{j,k}$ and the error pdf in figure 6.7. We observed a very similar pattern as before in which there was a slight departure from the centralised $\log(\chi^2_1)$ distribution in the coarsest scale. As the difference was only marginal, we concluded that there was not enough evidence to suggest our model assumptions were invalid.

The plots of the estimated slowly evolving EWS in figure 6.8 are poor, especially if we compare the Bayesian log-EWS estimator with the TI de-noised estimator in figure 3.9(a). There is a large amount of leakage from the power scale ($j = 6$) into all the other scales, especially into the coarser scale $j = 7, 8$ and the finer scales $j = 5, 4$. Comparing the plots of the cycle spun estimate in (b) with the estimate without cycle spinning in (a), there appears to be very little difference between them.

By plotting our estimate, CI and true of the slowly evolving spectrum for scales $j = 7, 6, 5$ in figure 6.9 we can examine in detail how poorly this method is working. On closer inspection cycle spinning does slightly improve the estimate and there is a reduction in the size of the CI. The CI for all the scales and estimates appear to capture the true EWS. For $j = 6$, the CI often are negative. It would imply there are periods when the power detected in our estimate is a side effect of leakage between scales. However, as this is not

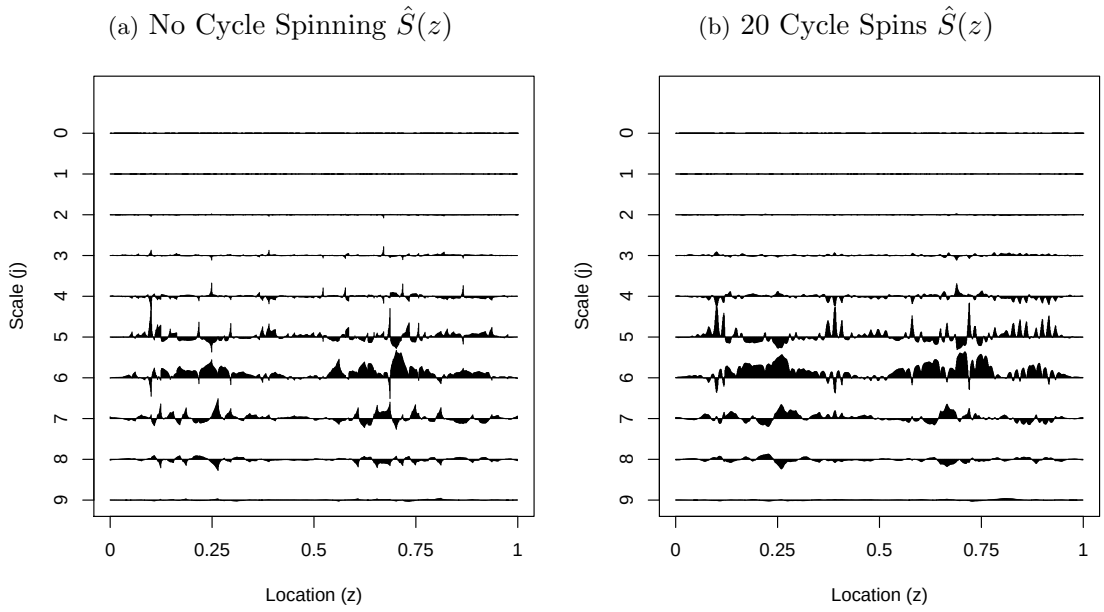


Figure 6.8: Plots of the slowly evolving EWS estimator using Bayesian wavelet shrinkage estimate of the log-wavelet periodogram obtained from Y_t.

the case, this provides further evidence of our method's poor performance.

As the form of the spectrum was known, it was suspected a smoother wavelet might improve the results. We performed the analysis using the D_{10} wavelet (see appendix B). The results appeared to be very similar to those in figures 6.8 and 6.9.

6.3.3 Conclusion

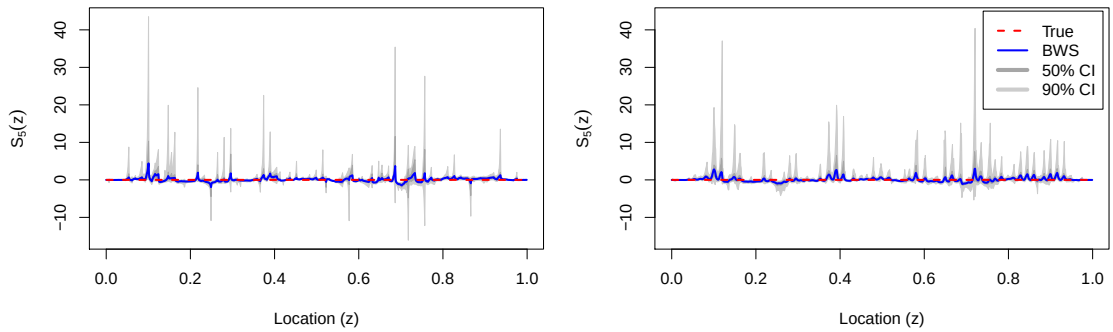
Although our model assumptions appeared to be reasonable, our subsequent estimates of the EWS are poor. The leakage and noise in the data led to imprecise estimates and large confidence intervals. Changing the smoothing wavelet had little effect on improving our estimates. Increasing the sample size might improve the problem, however increasing the sample size would require $T = 2^{11} = 2048$ observations. This may be difficult to obtain in practice.

Some investigations two possible sources of error:

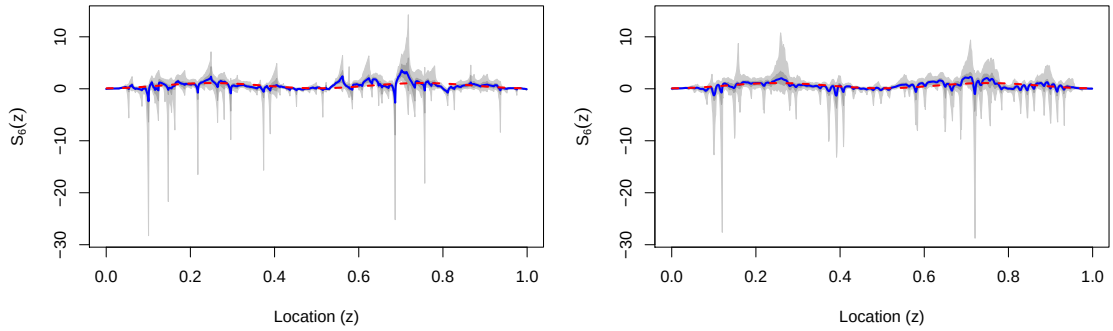
- hyperparameter estimation,
- assumption of independence.

Further research in Clyde and George (1998) suggested that using equation (2.39) to produce hyperparameter estimates could lead to biased results. We also used the method of Abramovich et al. (1998) to determine the hyperparameters. As we were able to tune the hyper-hyperparameters, a and b in (2.37), we produced very accurate results. However,

(a) No Spins $j = 5$ (b) 20 Spins $j = 5$



(c) No Spins $j = 6$ (d) 20 Spins $j = 6$



(e) No Spins $j = 7$ (f) 20 Spins $j = 7$

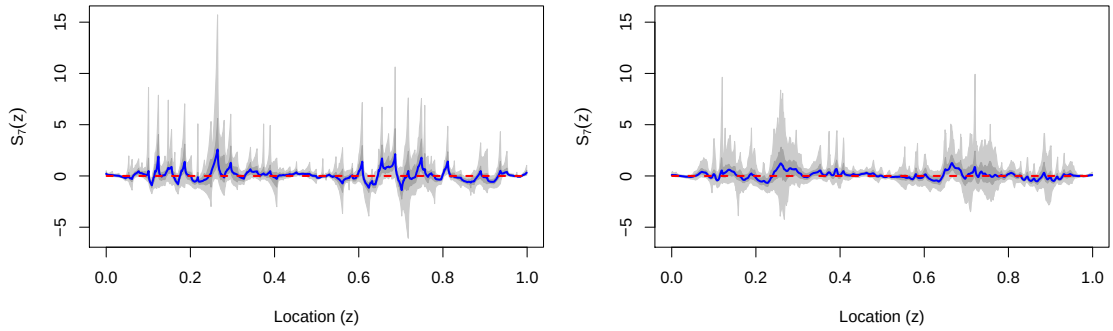


Figure 6.9: Plots of the piecewise constant EWS estimator for scales $j = 7, 6, 5$. Figures (a), (c) and (e) are estimated without cycle spinning, and figures (b), (d) and (f) are with 20 cycle spins. Analysis wavelet = Haar and smoothing wavelet = D_2. The true slowly evolving EWS is the dashed red line and our Bayesian log-EWS estimator is the blue solid line. The dark grey shaded area is the 50% CI and the light grey is the 90% CI.

this was also a consequence of knowing the true form of the EWS. Another possible method would be to use the EM algorithm (Clyde and George, 1998), however this would required further development due to the complicated form of the likelihood.

We assumed the raw wavelet periodogram was independent between scales and locations. The log transformation helped to stabilise the variance and induce a theoretically additive model, however, it did not affect the correlation structure of the periodogram. Previous research did suggest that an assumption of independence between locations which were 'far enough' apart was not unreasonable (section 3.2). However, with the leakage between scales, by addressing the possible covariance between the raw wavelet periodogram at the same location and different scales may yield more accurate results.

Chapter 7

Bayesian Wavelet Shrinkage of the Haar-Fisz EWS

In this chapter we shall describe how estimates and corresponding confidence intervals can be calculated for the evolutionary wavelet spectrum (EWS) after smoothing, via wavelet shrinkage, the Haar-Fisz transformed raw wavelet periodogram. We calculate confidence interval estimates for the soft shrinkage and Bayesian wavelet shrinkage of the Haar-Fisz wavelet periodogram.

We shall begin by describing the Haar-Fisz transformation of the raw wavelet periodogram in section 7.1 and investigating the correlation structure before and after the transformation in section 7.2. We develop a method for performing Bayesian wavelet shrinkage upon the full Haar-Fisz transformed raw wavelet periodogram, with corresponding credible intervals (CI) in 7.3, using a Uniform (7.4.1) and Laplace (7.4.2) prior. Finally in section 7.5 we evaluate our Bayesian wavelet shrinkage method using the simulated data from section 3.4.

7.1 Haar-Fisz Transformation

Applying the Fisz transformation helps to stabilise the variance and introduce Gaussianity to the data (Fisz, 1955). The Haar-Fisz transformation (described in section 2.6.4) of a non-Gaussian data set implies that any smoothing technique can be applied to the Haar-Fisz transformed data with the assumption of a Gaussian error distribution.

We shall begin by considering the full Haar-Fisz transformation of the raw wavelet periodogram. Firstly, assume the raw wavelet periodogram can be modelled as equation

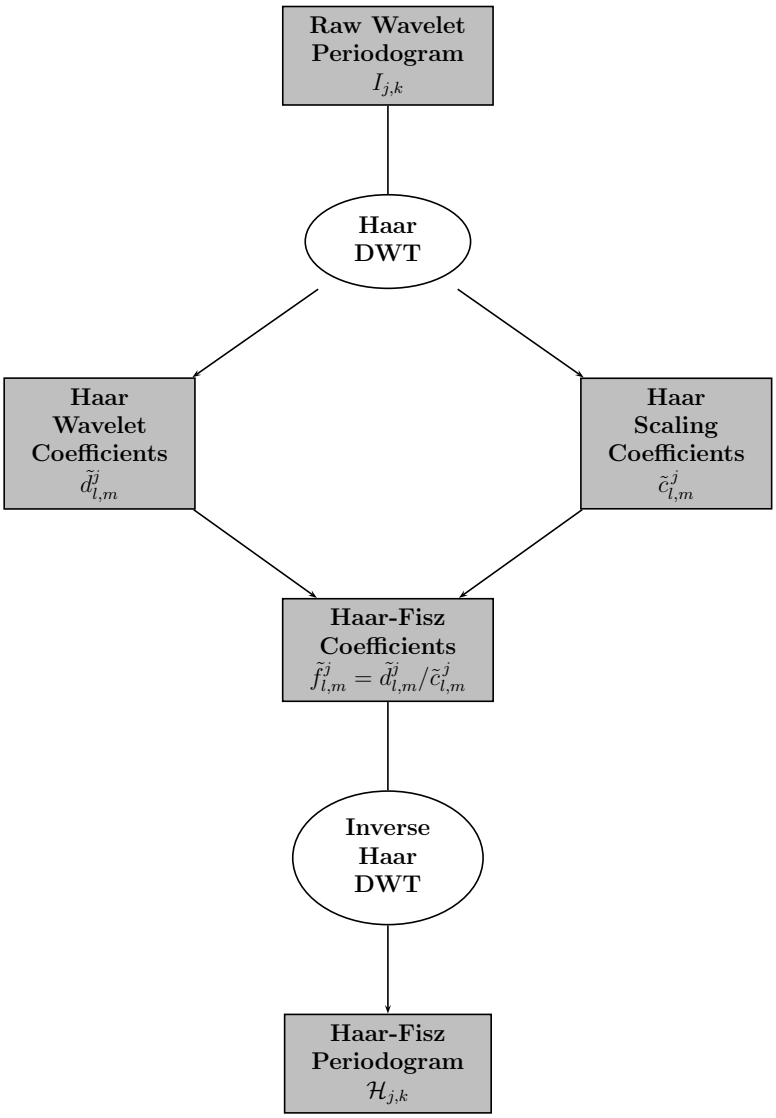


Figure 7.1: Flow diagram of the full Haar-Fisz transformation of the raw wavelet periodogram

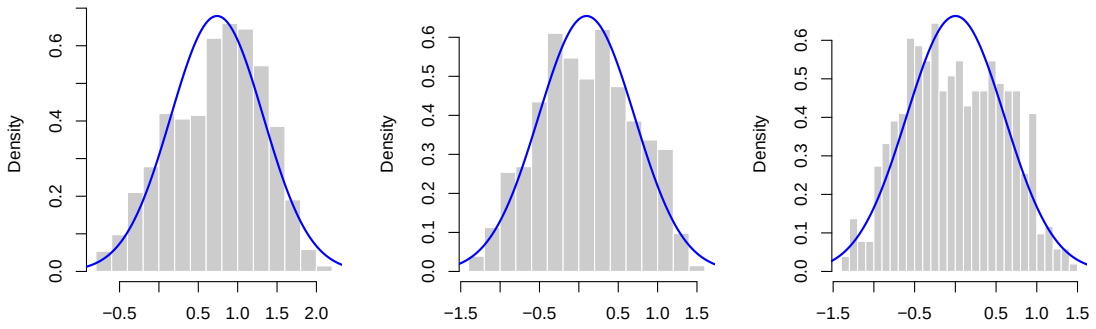
(5.1),

$$I_{j,k} \;\approx\; R_j\left(\frac{k}{T}\right) Z^2_{j,k},$$

where $R_j\left({}^{k}/_{T}\right) = (AS)_j\left({}^{k}/_{T}\right)$ and $Z^2_{j,k} \sim \chi^2_1$, for $j = 0, \ldots, J-1,\ k = 0, \ldots, 2^J - 1$ and a sample size of $T = 2^J$.

Let the full Haar-Fisz transformation be denoted by $\mathcal{F}$, and the full Haar-Fisz transformation of the raw wavelet periodogram be $\mathcal{F}I_{j,k} = \mathcal{H}_{j,k}$. Figure 7.1 shows a flow diagram of the steps required to perform a full Haar-Fisz transformation. From proposition 2.6.3(b) in section 2.6.4, the Haar-Fisz transformation possesses the log-like property.

(a) Histogram of X_t's $\mathcal{H}_{9,k}$ (b) Histogram of X_t's $\mathcal{H}_{6,k}$ (c) Histogram of X_t's $\mathcal{H}_{0,k}$



(d) Histogram of Y_t's $\mathcal{H}_{9,k}$ (e) Histogram of Y_t's $\mathcal{H}_{6,k}$ (f) Histogram of Y_t's $\mathcal{H}_{0,k}$

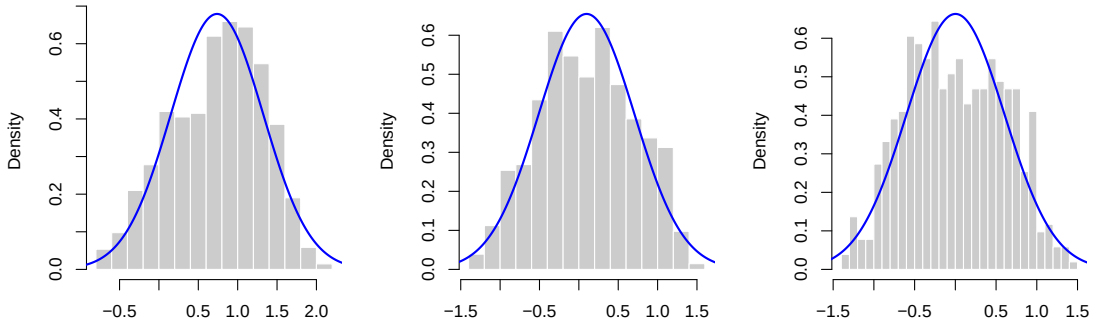


Figure 7.2: Density plots of the Haar-Fisz wavelet periodogram ($\mathcal{H}_{j,k}$) of X_t (in a, b and c) and Y_t (in d, e and f) at: the finest scale in (a); a mid scale in (b); the coarsest scale in (c)). For a sample size of $T = 2^{10} = 1024$. The solid blue line is the $\mathcal{N}(\mathbb{E}[\mathcal{H}_j], \mathrm{Var}[\mathcal{H}_j])$ pdf.

This suggests a possible model for the Haar-Fisz transformed wavelet periodogram is

$$\mathcal{H}_{j,k} = \mathcal{B}_j\left(\frac{k}{T}\right) + e_{j,k} \tag{7.1}$$

for $j = 1, \ldots, J$ and $k = 1, \ldots, 2^J$, where $\mathcal{B}_j\left({}^{k}/_{T}\right) = \mathcal{F}R_j\left({}^{k}/_{T}\right)$ and $e_{j,k} = \mathcal{F}Z^2_{j,k}$. The distribution of $\mathcal{H}_{j,k}$ is approximately $\mathcal{N}(\mathcal{B}_j\left({}^{k}/_{T}\right), \sigma_j^2)$ and therefore $e_{j,k} \approx \mathcal{N}(0, \sigma_j^2)$.

7.2 The Distribution and Correlation of the Haar-Fisz Periodogram

We determined in theorem 3.2.1 that the distribution of the raw wavelet periodogram with Gaussian innovations was approximately multivariate chi-square. To assess whether the Haar-Fisz wavelet periodogram ($\mathcal{H}$) was closer to Gaussianity we used the simulated data from section 3.4 to numerically assess this assumption. We used the LSW processes X_t and Y_t with Gaussian innovations and extracted the raw wavelet periodogram using the

Haar analysis wavelet, then we obtained the Haar-Fisz wavelet periodogram. We produced density plots of the Haar-Fisz wavelet periodogram and compared their distribution to the Gaussian pdf with mean $\mathbb{E}[\mathcal{H}_j]$ and variance $\text{Var}[\mathcal{H}_j]$, see figures 7.2.

Although we observed that the distribution of the Haar-Fisz wavelet periodogram was not exactly Gaussian, it was closer to Gaussianity than the raw wavelet periodogram. Hence, we felt this helped validated the assumption that the distribution of the Haar-Fisz periodogram was approximately Gaussian.

Using the DWT helps to break down the covariance structure within a data set. However, we were unsure how the Haar-Fisz transformation may effect the correlation structure (Johnstone and Silverman, 1998). To quantify the possible effect, we calculated the autocorrelation (acf) and partial autocorrelation (p.acf) function for lags $\tau = 0, \ldots, 500$ of the raw wavelet periodogram and Haar-Fisz wavelet periodogram of X_t and Y_t and compared the results.

We found the p.acf revealed similar information as acf (see appendix B for p.acf empirical evidence). Therefore we only displayed the acf plots in figures 7.3 and 7.4 for scales $j = 9, 6$ and 0 of the finest, mid and coarse scales. A 5% significance level was also plotted, usually values of the acf which are less than this line provide insufficient evidence they are significantly different from zero and can be ignored.

Figure 7.3 is the empirical acf of the full raw wavelet periodogram (a, c, e) and Haar-Fisz wavelet periodogram (b, d, and f) of the LSW process X_t. At the finest and mid scale, the acf of the raw wavelet periodogram is positive for approximately $\tau = 0, \ldots, 180$ and mainly negative after. Using the Haar-Fisz transformation appears to remove this pattern, so that the acf is randomly positive or negative. Also, the acf value is less for the Haar-Fisz wavelet periodogram compared to the raw wavelet periodogram. At the coarsest scale, except for $\tau = 0$, there appears to be a reduction in the acf value.

The acf of the raw wavelet periodogram (a, c, e) and Haar-Fisz wavelet periodogram (b, d, and f) calculated using the Haar analysis wavelet from the LSW process Y_t can be observed in figure 7.4. For all scales, we observe a similar pattern in the acf as the EWS at the fourth finest scale ($j = 6$). The Haar-Fisz transformation has removed some of the pattern in the acf, and there appears to be an overall reduction in the values of the acf.

The evidence suggests that using the Haar-Fisz transformation, will not totally remove all of the correlation in the raw wavelet periodogram, but there is certainly a reduction. This will help to ensure the validity of assumptions of independence, which proved prob-

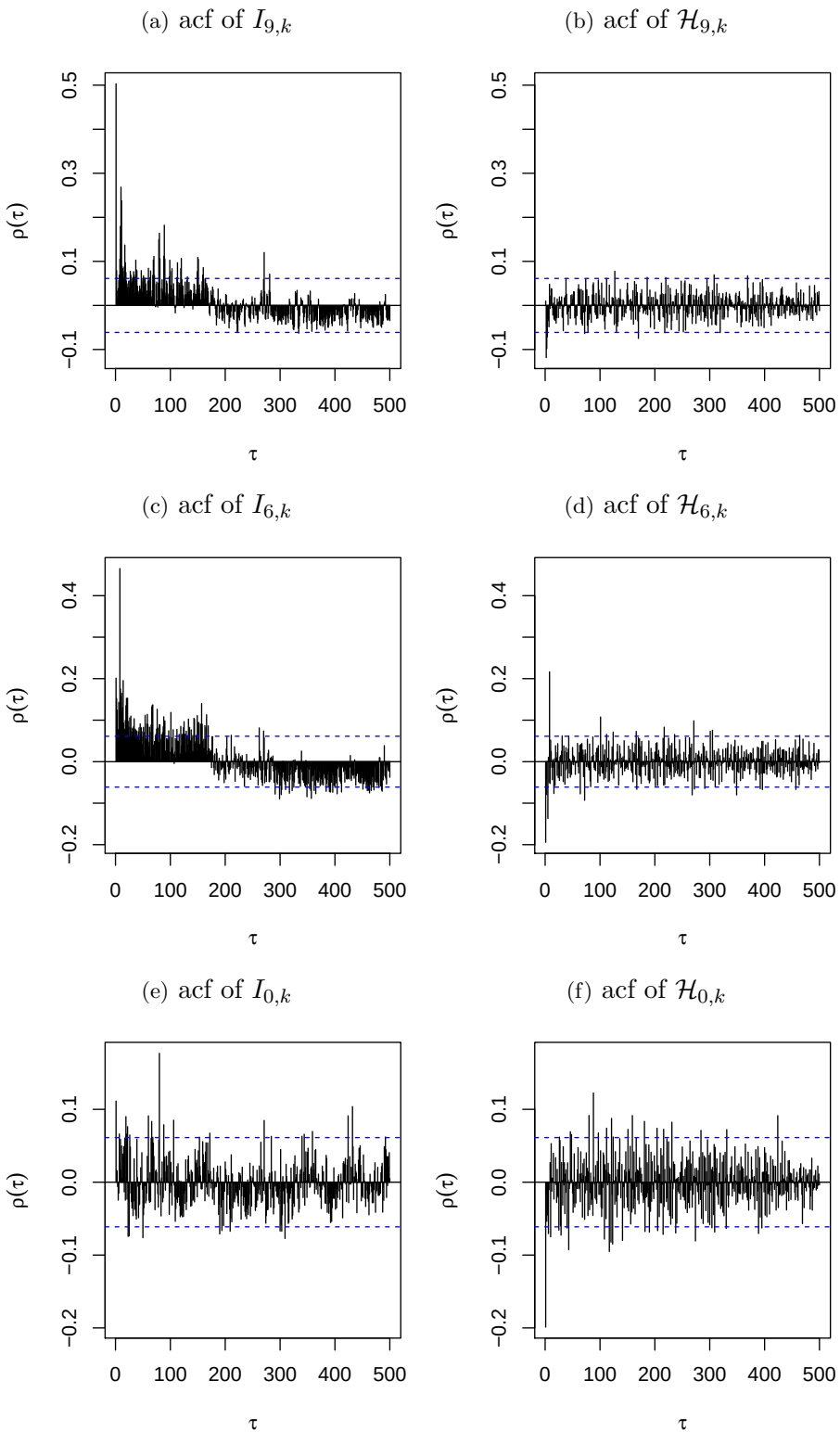


Figure 7.3: Plots of the empirical acf over lag $\tau = 0, \ldots, 500$ of the raw wavelet periodogram ($I_{j,k}$) and Haar-Fisz wavelet periodogram ($\mathcal{H}_{j,k}$) of X_t simulated from the piecewise constant EWS at: the finest scale in (a) and (b); a mid scale in (c) and (d); the coarsest scale in (e) and (f). For a sample size of $T = 2^{10} = 1024$. The blue dashed line in each plot is the 5% significance level.

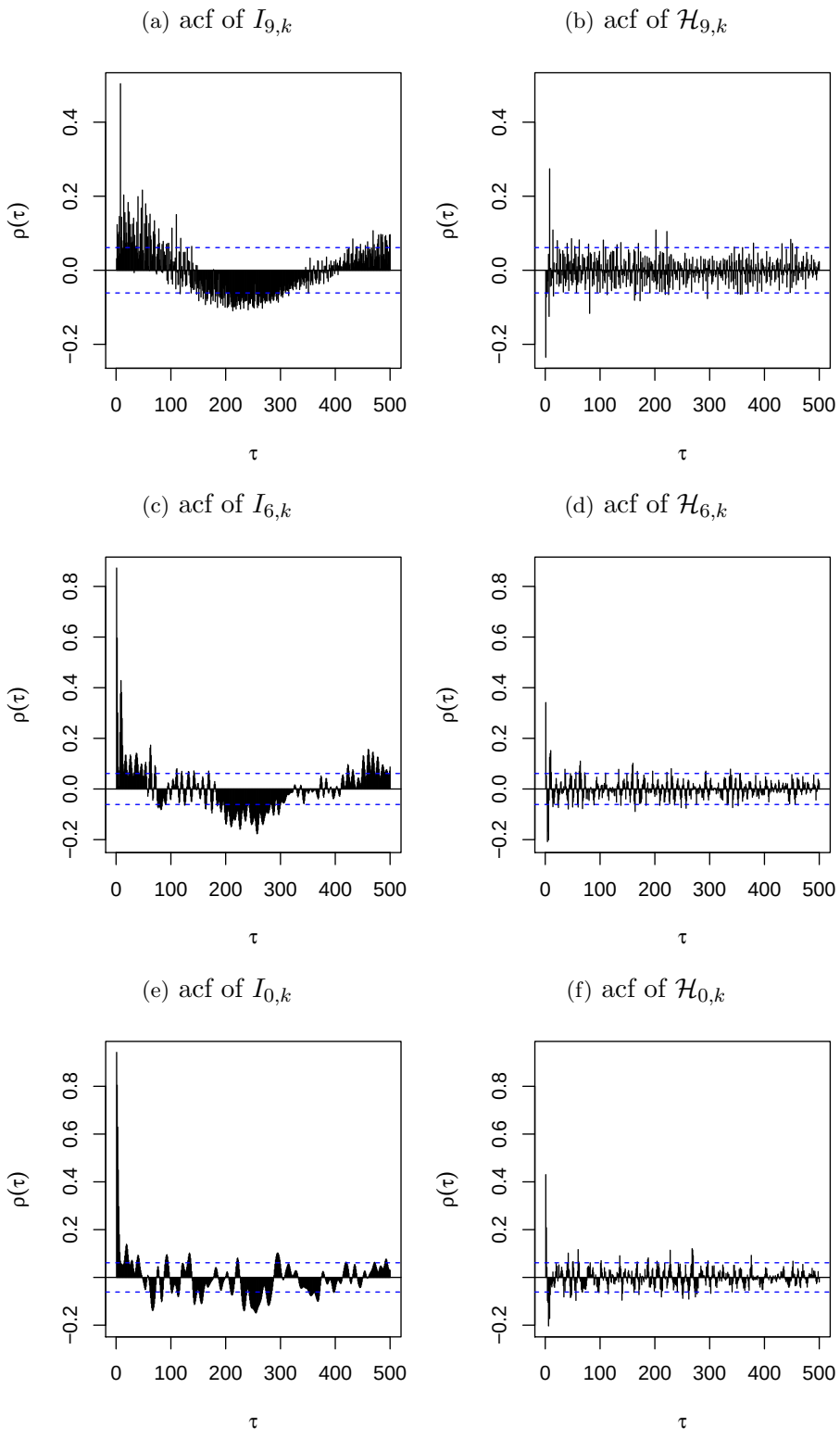


Figure 7.4: Plots of the empirical acf over lag $\tau = 0, \ldots, 500$ of the raw wavelet periodogram ($I_{j,k}$) and Haar-Fisz wavelet periodogram ($\mathcal{H}_{j,k}$) of Y_t simulated from the slowly evolving EWS at: the finest scale in (a) and (b); a mid scale in (c) and (d); the coarsest scale in (e) and (f). For a sample size of $T = 2^{10} = 1024$. The blue dashed line in each plot is the 5% significance level.

lematic in the previous Bayesian wavelet shrinkage methodology.

7.3 Bayesian Wavelet Shrinkage of the Haar-Fisz Periodogram

Once a full Haar-Fisz transformation has been performed upon the raw wavelet periodogram we have demonstrated through simulations that the resulting data appear to possess a Gaussian distribution. In this section we suggest using a Bayesian wavelet shrinkage approach, but to improve the flexibility of this method, we shall assume that this shrinkage can be performed on the wavelet coefficients of the Haar-Fisz transformed data using any smoothing wavelet (SW).

7.3.1 Model

Take the DWT of (7.1), for a particular scale j, we obtain

$$h_{l,m} = \beta_{l,m} + \varepsilon_{l,m},$$

where $h_{l,m} = (W\mathcal{H}_j)_{l,m}$, $\beta_{l,m} = (W\mathcal{B}_j)_{l,m}$, $\varepsilon_{l,m} = (We_j)_{l,m}$ for scales $l = 0, \ldots, J-1$ and locations $m = 1, \ldots, 2^l$, and W is the $T \times T$ wavelet transformation matrix. Due to the properties of the wavelet transformation, the error distribution will be approximately $\varepsilon_{l,m} \sim \mathcal{N}(0, \nu_l^2)$ where $\nu_l^2 = 2^{J-l}\sigma_j^2$. For future notation we shall assume j is fixed and drop this notation.

7.3.2 Prior

We propose using the Berger-Müller mixture prior, as used in Bayesian wavelet shrinkage of the coefficients from the log-transformed raw wavelet periodogram from equation (5.9),

$$p(\beta_{l,m}) = \alpha_l \delta(\beta_{l,m}) + (1 - \alpha_l)\xi(\beta_{l,m}). \tag{7.2}$$

The appropriate choice of the function $\xi(\cdot)$ will depend upon the chosen smoothing wavelet (SW).

7.3.3 Likelihood

Due to the normalisation properties of the Haar-Fisz transformation, the likelihood will be of the form

$$p(\varepsilon_{l,m}|\beta_{l,m}) = \varphi_{\nu_l}(\varepsilon_{l,m}),$$

where $\varphi_{\nu_l}(\cdot)$ is the probability density function of the Gaussian distribution with variance ν_l^2. Let $\varepsilon_{l,m} = h_{l,m} - \beta_{l,m}$, then our likelihood becomes

$$\begin{aligned} p(h_{l,m}|\beta_{l,m}) &= \varphi_{\nu_l}(\beta_{l,m} - h_{l,m}) \\ &= \frac{1}{\nu_l\sqrt{2\pi}} \exp\left\{-\frac{1}{2\nu_l^2}(h_{l,m} - \beta_{l,m})^2\right\}. \end{aligned} \tag{7.3}$$

7.3.4 Posterior Distribution

Using the Berger-Müller mixture prior, yields a posterior distribution of two parts: the dirac delta function and tail density. We shall only calculate the posterior tail density, because this will eventually result in the shrinkage rule. The proportion of the distribution which is contained within the tail density is controlled by α.

For simplicity, in what follows we shall ignore the l, m subscripts for the β's and h's. Then by combining the tail density prior and the likelihood, the posterior distribution is calculated from

$$\begin{aligned} p(\beta|h) &= \frac{p(\beta)p(h|\beta)}{\int p(y)p(h|y)\mathrm{d}y} \\ &= \frac{(1-\alpha_l)\,\xi(\beta)\,\varphi_{\nu_l}(\beta - h)}{\int[\alpha_l\delta(y) + (1-\alpha_l)\,\xi(y)]\,\varphi_{\nu_l}(h-y)\mathrm{d}y} \\ &= \frac{\xi(\beta)\,\varphi_{\nu_l}(\beta - h)}{\theta_l\,\varphi_{\nu_l}(h) + \int \xi(y)\,\varphi_{\nu_l}(y-h)\mathrm{d}y}, \end{aligned} \tag{7.4}$$

where θ_l is the odds ratio, $\theta_l = \alpha_l(1-\alpha_l)^{-1}$.

7.3.5 Posterior Mean

To produce an estimate of the wavelet coefficients β, we shall find the posterior mean. This is calculated from the integral

$$\hat{\beta} = \int x \, \frac{p(x)p(h|x)}{\int p(y)p(h|y)\mathrm{d}y} \, \mathrm{d}x$$

$$= \frac{\int x \, \xi(x) \, \varphi_{\nu_l}(x-h)\mathrm{d}x}{\theta_l \, \varphi_{\nu_l}(h) + \int \xi(y) \, \varphi_{\nu_l}(y-h)\mathrm{d}y}. \tag{7.5}$$

7.3.6 Posterior Variance

The posterior variance is calculated from the integral

$$\mathrm{Var}[\beta|h] = \mathbb{E}[\beta^2|h] - (\mathbb{E}[\beta|h])^2$$

$$= \int x^2 \, \frac{p(x)p(h|x)}{\int p(y)p(h|y)\mathrm{d}y} \, \mathrm{d}x \; - \; \left(\int x \, \frac{p(x)p(h|x)}{\int p(y)p(h|y)\mathrm{d}y} \, \mathrm{d}x \right)^2$$

$$= \frac{\int x^2 \, \xi(x) \, \varphi_{\nu_l}(x-h)\mathrm{d}x}{\theta_l \, \varphi_{\nu_l}(h) + \int \xi(y) \, \varphi_{\nu_l}(y-h)\mathrm{d}y} \; - \; \hat{\beta}^2. \tag{7.6}$$

7.3.7 Hyperparameter Determination

We will use marginal maximum likelihood estimation (MLE)to determine the hyperparameters ν_l and α_l. Using the same methodology of Johnstone and Silverman (2005), whereby we maximise the hyperparameters over the log-likelihood of the error distribution multiplied by the prior,

$$\mathcal{L}(\alpha_l, \nu_l|\mathbf{h}_l) \; = \; \sum_{m=0}^{2^l-1} \log\{\alpha_l \varphi_{\nu_l}(h_{l,m}) + (1-\alpha_l)\gamma(h_{l,m}|\nu_l)\}\,, \tag{7.7}$$

where $\mathbf{h}_l = [h_{l,0}, \ldots, h_{l,2^l-1}]$ and

$$\gamma(y|\nu_l) = \int_{-\infty}^{\infty} \varphi_{\nu_l}(y-x)\xi(x)\mathrm{d}x. \tag{7.8}$$

This will require a computational estimation using an appropriate optimisation program.

7.3.8 Credible Intervals

To calculate credible intervals we will use importance sampling similarly to the Bayesian wavelet shrinkage method described in 6.1. For a particular wavelet coefficient, $\hat{\beta}_{l,m}$, we simulate 100 observations from the posterior distribution. We then calculate the mean of these observations, this is now our simulated value of $\hat{\beta}_{l,m}$. This is repeated for all of the wavelet coefficients, which are then applied to the inverse wavelet transform to produce a simulated smoothed wavelet periodogram. This process is repeated 100 times and the variance of these simulations for each location is an estimate for the variance of our smoothed estimate.

7.4 Choice of Tail Density Prior

We shall now discuss two choices of tail density priors: the uniform (section 7.4.1) and Laplace (section 7.4.2) distributions. The appropriate choice of mixture prior will depend on our selected of smoothing wavelet (SW).

7.4.1 Uniform Mixture Prior

Suppose we use the Haar SW, then the DWT of $\mathcal{H}$ will return the Haar-Fisz coefficients (i.e. $h_{l,m} = \tilde{f}_{l,m}$ from the flow diagram 7.1). As a result of the Fisz transformation, we find $\tilde{f}_{l,m} \in [-1, 1]$, therefore $\xi(\cdot)$ in (7.2) should be selected with a support of $[-1, 1]$ to convey the appropriate prior information. One such example is the Uniform$[-1, 1]$, which is fairly uninformative, however, it can easily reflect the constraints of the wavelet coefficients.

For ease of notation we shall drop the l, m subscripts during our calculations. Let the function $\xi(\cdot)$, in the prior distribution, be the uniform probability density function, therefore

$$p(\beta) = \alpha_l \delta(0) \; + \; (1 - \alpha_l) \, \frac{1}{2} \, \mathbb{I}_{[-1,1]}(\beta). \tag{7.9}$$

Then to calculate the posterior distribution, mean and variance, consider the integral

$$\begin{aligned} Q_u^i(h) &= \int_{-\infty}^{\infty} y^i \, \varphi_{\nu_l}(y - h) \, \xi(y) \, \mathrm{d}y \\ &= \int_{-\infty}^{\infty} y^i \, \varphi_{\nu_l}(y - h) \, \frac{1}{2} \, \mathbb{I}_{[-1,1]}(y) \, \mathrm{d}y \\ &= \frac{1}{2} \int_{-1}^{1} y^i \, \varphi_{\nu_l}(y - h) \, \mathrm{d}y. \end{aligned} \tag{7.10}$$

By calculating $Q_u^i(h)$ for $i = 0, 1, 2$ with the uniform mixture prior, we can determine the posterior mean and variance, using the equations

$$\hat{\beta}_{l,m} = \frac{Q_u^1(h_{l,m})}{\theta_l \, \varphi_{\nu_l}(h_{l,m}) + Q_u^0(h_{l,m})} \tag{7.11}$$

$$\text{Var}[\beta_{l,m}|h_{l,m}] = \frac{Q_u^2(h_{l,m})}{\theta_l \, \varphi_{\nu_l}(h_{l,m}) + Q_u^0(h_{l,m})} - \hat{\beta}_{l,m}^2. \tag{7.12}$$

Let $i = 0$ then we find

$$\begin{aligned} Q_u^0(h) &= \frac{1}{2}\int_{-1}^{1} \varphi_{\nu_l}(y-h)\,\mathrm{d}y \\ &= \frac{1}{2}\left[\int_{-\infty}^{1} \varphi_{\nu_l}(y-h)\;\mathrm{d}y - \int_{-\infty}^{-1} \frac{1}{\nu_l}\varphi_{\nu_l}(y-h)\;\mathrm{d}y\right] \\ &= \frac{1}{2}\left[\Phi\left(\frac{1-h}{\nu_l}\right) - \Phi\left(\frac{-1-h}{\nu_l}\right)\right]. \end{aligned} \tag{7.13}$$

If $i = 1$ and by using the results in Appendix A.1.2, equation (A.11), this is

$$\begin{aligned} Q_u^1(h) &= \frac{1}{2}\int_{-1}^{1} y\;\varphi_{\nu_l}(y-h)\,\mathrm{d}y, \\ &= \frac{1}{2}\left[\int_{-\infty}^{1} y\;\varphi_{\nu_l}(y-h)\,\mathrm{d}y - \int_{-\infty}^{-1} y\;\varphi_{\nu_l}(y-h)\,\mathrm{d}y\right], \\ &= \frac{1}{2}\left\{h\left[\Phi\left(\frac{1-h}{\nu_l}\right) - \Phi\left(\frac{-1-h}{\nu_l}\right)\right] - \nu_l\left[\varphi\left(\frac{1-h}{\nu_l}\right) - \varphi\left(\frac{-1-h}{\nu_l}\right)\right]\right\}. \end{aligned} \tag{7.14}$$

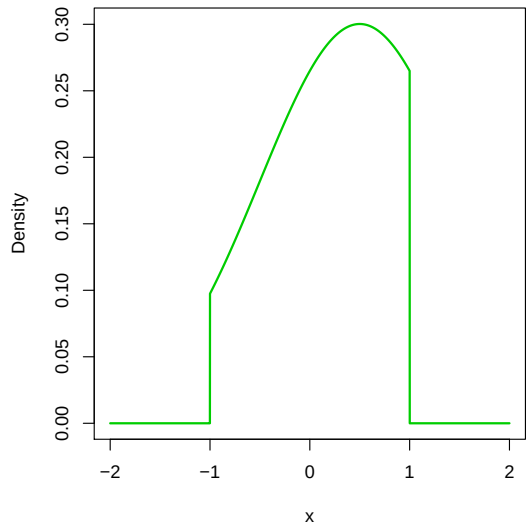
When $i = 2$, using the results in equation (A.16) we find

$$\begin{aligned} Q_u^2(h) &= \frac{1}{2}\int_{-1}^{1} y^2\;\varphi_{\nu_l}(y-h)\,\mathrm{d}y \\ &= \frac{1}{2}\left[\int_{-\infty}^{1} y^2\;\varphi_{\nu_l}(y-h)\,\mathrm{d}y - \int_{-\infty}^{-1} y^2\;\varphi_{\nu_l}(y-h)\,\mathrm{d}y\right], \\ &= \frac{1}{2}\left\{(\nu_l^2+h^2)\left[\Phi\left(\frac{1-h}{\nu_l}\right) - \Phi\left(\frac{-1-h}{\nu_l}\right)\right] - h\nu_l\left[\varphi\left(\frac{1-h}{\nu_l}\right) - \varphi\left(\frac{-1-h}{\nu_l}\right)\right]\right\}. \end{aligned} \tag{7.15}$$

To determine the hyperparameters via MLE we must first calculate the convolution of the error and uniform distribution to produce $\gamma(\cdot)$,

$$\begin{aligned} \gamma(y|\nu_l) &= \int_{-\infty}^{\infty} \frac{1}{\nu_l\sqrt{2\pi}} \exp\left\{-\frac{(y-x)^2}{2\nu_l^2}\right\} \frac{1}{2}\mathbb{I}_{[-1,1]}(x)\mathrm{d}x \\ &= \frac{1}{2}\int_{-1}^{1} \frac{1}{\nu_l\sqrt{2\pi}} \exp\left\{-\frac{(x-y)^2}{2\nu_l^2}\right\} \mathrm{d}x \\ &= \frac{1}{2}\left[\Phi\left(\frac{1-y}{\nu_l}\right) - \Phi\left(\frac{-1-y}{\nu_l}\right)\right]. \end{aligned} \tag{7.16}$$

(a) Posterior Tail Distribution



(b) Shrinkage Function

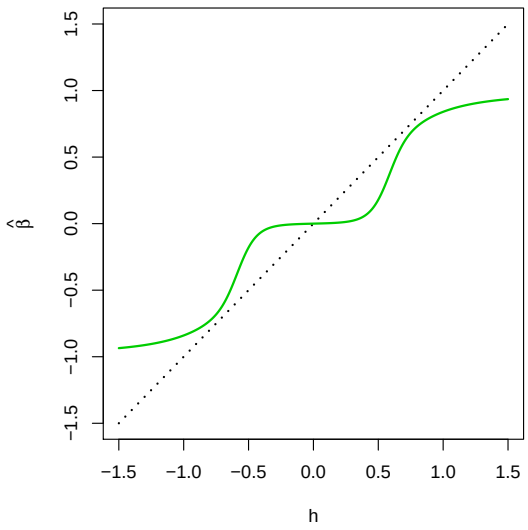


Figure 7.5: Plots of the posterior tail distribution and shrinkage function obtained using the uniform mixture prior. Posterior tail distribution for parameters $h = {}^1\!/_2$, $\nu = 1$ and $\theta = {}^1\!/_3$. Shrinkage function for parameters $\nu = {}^1\!/_5$ and $\theta = 50$

Therefore, the log-likelihood function for this step which we need to maximise over is

$$\mathcal{L}(\alpha_l, \nu_l | \mathbf{h}_l) \;=\; \sum_{m=0}^{2^l-1} \log \left\{ \alpha_l \varphi_{\nu_l}(h_{l,m}) + \frac{1-\alpha_l}{2} \left[\Phi\left(\frac{1-h_{l,m}}{\nu_l} \right) - \Phi\left(\frac{-1-h_{l,m}}{\nu_l} \right) \right] \right\}. \tag{7.17}$$

The plot in figure 7.5(a) shows the posterior tail distribution using our uniform mixture prior for the Haar smoothing wavelet for data $h = {}^1\!/_2$, prior precision $\nu = 1$ and odds ratio $\theta_l = {}^1\!/_3$. The value of h causes the symmetry of the posterior tail distribution to be centred around h, and as the support of our mixture prior is $[-1, 1]$ the posterior pdf is zero for values which are not within the interval $[-1, 1]$. The plot in figure 7.5(b) shows the shrinkage function for $h \in [-1.5, 1.5]$ with prior precision $\nu = {}^1\!/_5$ and $\theta = 50$. The plot demonstrates how values of $|h| > 1$ are are tending to sign(h). Due to the Haar-Fisz transformation and the use of the Haar smoothing wavelet, no value would be outside the interval $[-1, 1]$. Only values of $|h|$ which are less than and close to 1 remain relatively unchanged, whereas $|h| < {}^1\!/_2$ appear to be zero.

7.4.2 Laplace Mixture Prior

Due to the short support of the Haar wavelet, we believe that using the Haar SW can limit the type of EWS which can be successfully recovered, i.e. as defined in (2.65). Therefore, to increase the flexibility of our method we shall consider performing Bayesian wavelet shrinkage on the wavelet coefficients obtained from the DWT of the Haar-Fisz

periodogram when SW $\neq$ Haar.

For non-Haar smoothing wavelets the support of the corresponding wavelet coefficients obtained from the full Haar-Fisz transformed raw wavelet periodogram is not between $[-1, 1]$. Therefore the uniform distribution is not an appropriate candidate for $\xi(\cdot)$ in the Berger-Müller mixture prior. We suggest using the Laplace distribution, because of its heavy tails, but also due to the performance of the Laplace distribution when applied to the Bayesian wavelet shrinkage of the log-transformed raw wavelet periodogram in section 6.2.

As before, we shall drop the l, m subscripts for ease of notation. Let the function $\xi(\cdot)$, in the prior distribution, be the Laplace probability density function, therefore

$$p(\beta) = \alpha_l \delta(0) + (1 - \alpha_l)\ \frac{\tau_l}{2}\ \exp\{-\tau_l\,|\beta|\} \tag{7.18}$$

where τ_l is the scale parameter and $2\tau_l^{-2}$ is the prior variance for scale $l = 1, \ldots, J$.

To determine the posterior tail distribution, mean and variance consider we require the following lemmas.

Lemma 7.4.1. *The integral required to calculate the posterior distribution, mean and variance using the Laplace mixture prior is*

$$Q_l^i(h) = \frac{\tau_l}{2} e^{-h^2/2\nu_l^2} \left[e^{\mu_1^2/2\nu_l^2} \int_{-\infty}^{0} y^i\, \varphi_{\nu_l}(y - \mu_1)\, dy \ +\ e^{\mu_2^2/2\nu_l^2} \int_{-\infty}^{0} (-y)^i\, \varphi_{\nu_l}(y + \mu_2)\, dy \right],$$

where

- $\varphi_\nu(\cdot)$ *is the zero mean Gaussian pdf with variance* ν^2,
- $\mu_1 = h + \nu_l^2\, \tau_l$,
- $\mu_2 = h - \nu_l^2\, \tau_l$.

The posterior mean for the Laplace mixture prior is

$$\hat{\beta}_{l,m} = \frac{Q_l^1(h_{l,m})}{\theta_l\, \varphi_{\nu_l}(h_{l,m}) + Q_l^0(h_{l,m})},$$

and the posterior variance is

$$\mathrm{Var}[\beta_{l,m}|h_{l,m}] = \frac{Q_l^2(h_{l,m})}{\theta_l\, \varphi_{\nu_l}(h_{l,m}) + Q_l^0(h_{l,m})} - \hat{\beta}_{l,m}^2.$$

The solution to these quantities can be obtained through lemma 7.4.2.

Lemma 7.4.2. *The solutions for lemma 7.4.1 for $i = 0, 1, 2$ are*

(a) $i = 0$

$$Q_l^0(h) = \frac{\tau_l}{2} e^{-h^2/2\nu_l^2} \left[e^{\mu_1^2/2\nu_l^2} \Phi\left(\frac{-\mu_1}{\nu_l}\right) \; + \; e^{\mu_2^2/2\nu_l^2} \Phi\left(\frac{\mu_2}{\nu_l}\right) \right].$$

(b) $i = 1$

$$Q_l^1(h) = \frac{\tau_l}{2} e^{-h^2/2\nu_l^2} \left\{ e^{\mu_1^2/2\nu_l^2} \left[\mu_1 \, \Phi\left(-\frac{\mu_1}{\nu_l}\right) - \nu_l \, \varphi\left(-\frac{\mu_1}{\nu_l}\right) \right] + \; e^{\mu_2^2/2\nu_l^2} \left[\mu_2 \, \Phi\left(\frac{\mu_2}{\nu_l}\right) + \nu_l \, \varphi\left(\frac{\mu_2}{\nu_l}\right) \right] \right\}.$$

(c) $i = 2$

$$Q_l^2(h) = \frac{\tau_l}{2} e^{-h^2/2\nu_l^2} \left\{ e^{\mu_1^2/2\nu_l^2} \left[(\nu_l^2 + \mu_1^2) \Phi\left(-\frac{\mu_1}{\nu_1}\right) - \mu_1 \nu_l \, \varphi\left(-\frac{\mu_1}{\nu_1}\right) \right] + \; e^{\mu_2^2/2\nu_l^2} \left[(\nu_l^2 + \mu_2^2) \Phi\left(\frac{\mu_2}{\nu_l}\right) + \mu_2 \nu_l \, \varphi\left(\frac{\mu_2}{\nu_l}\right) \right] \right\}.$$

To determine the hyperparameters by MLE, we require the log-likelihood function of our Bayesian model from (7.7) for the Laplace mixture prior. this can be found in lemma 7.4.3.

Lemma 7.4.3. *The log-likelihood function for the Laplace mixture prior is*

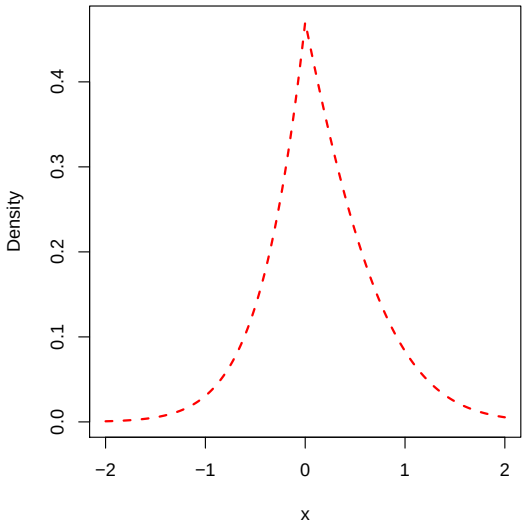
$$\mathcal{L}(\alpha_l, \nu_l, \tau_l | \mathbf{h}_l) = \sum_{m=0}^{2^l - 1} \log \left\{ \alpha_l \varphi_{\nu_l}(h_{l,m}) + \frac{\tau_l (1 - \alpha_l)}{2} e^{-y^2/2\nu_l^2} \left[e^{\mu_3^2/2\nu_l^2} \Phi\left(\frac{-\mu_3}{\nu_l}\right) + e^{\mu_4^2/2\nu_l^2} \Phi\left(\frac{\mu_4}{\nu_l}\right) \right] \right\},$$

where

- $\varphi_\nu(\cdot)$ *is the zero mean Gaussian pdf with variance* ν^2,
- $\Phi(\cdot)$ *is the Gaussian cdf,*
- $\mu_3 = y + \nu_l^2 \, \tau_l$,
- $\mu_4 = y - \nu_l^2 \, \tau_l$.

The proofs for lemmas 7.4.1 to 7.4.3 are available in appendix A.2.

(a) Posterior Tail Distribution



(b) Shrinkage Function

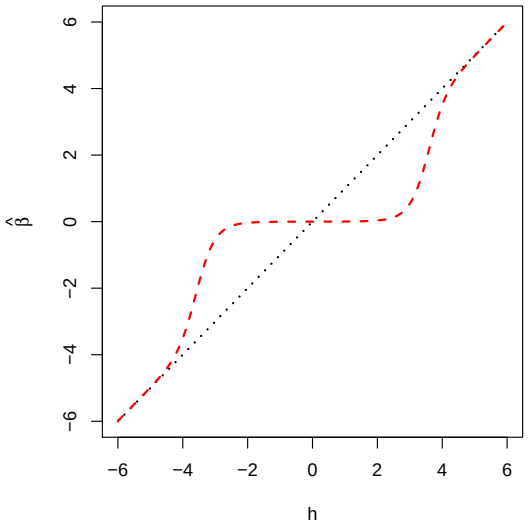


Figure 7.6: Plots of the posterior tail distribution and shrinkage function obtained using the Laplace mixture prior. Posterior tail distribution for parameters $h = {}^1\!/_2,\ \nu = 1,\ \theta = {}^1\!/_3$ and $\tau = \sqrt{3}$. Shrinkage function for parameters $\nu = 1,\ \theta = 5$ and $\tau = 1/100$

The plot in figure 7.6(a) is the posterior tail distribution for the Laplace prior and figure 7.6(b) is the Laplace shrinkage function. The shrinkage function demonstrates how values of h are 'shrunk' to produce the posterior mean estimate, $\hat{\beta}$ with parameters $\nu = 1,\ \theta = 5$ and $\tau = 1/100$. Values of $|h| \geq 4$ remain unchanged, whereas absolute values which are less than four are reduced in magnitude.

7.5 Simulation Examples

The coding required for the Bayesian wavelet shrinkage of the Haar-Fisz wavelet periodogram is very similar to the code described in chapter 6. However, as the likelihood function consists only of the Gaussian probability density function, half the number of calculations are required. With the Laplace prior many components required to calculate the Laplace posterior mean are the same as $Q_0(h)$ and $Q_1(h)$ in section 5.3.2.

We determined the hyperparameters by maximising the marginal log likelihood using the `optim` function and `L-BFGS-B` method in `R`. However, we also found that there were problems with insufficient data and sparsity when we tried to numerically maximise the log likelihood function for both the uniform and Laplace prior distributions. In particular we found that with the Laplace prior as we wished to estimate three hyperparameters this became particularly problematic. Therefore, we had to simplify the maximisation in two ways. Firstly, we estimate ν_l directly from the data.

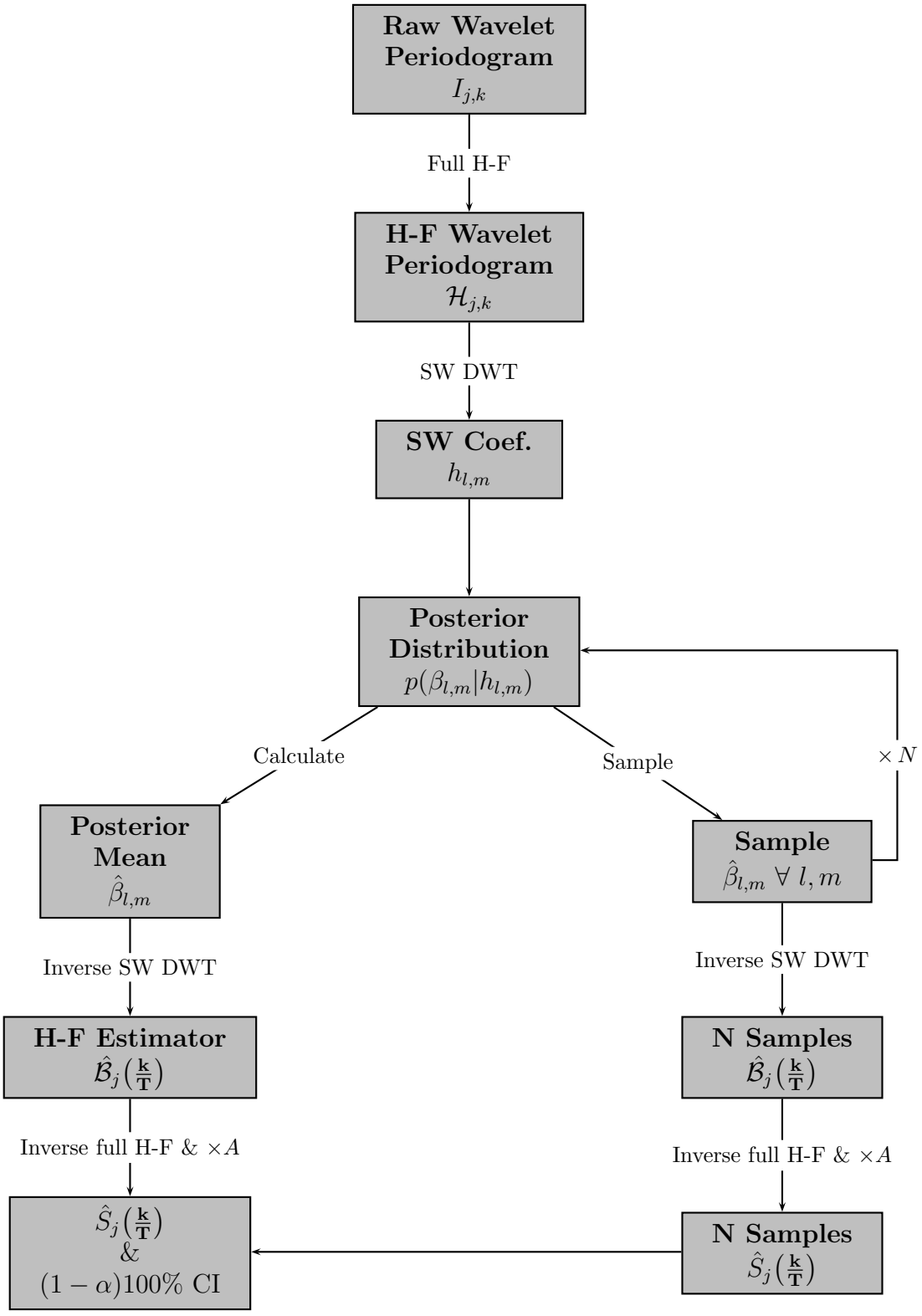


Figure 7.7: A flow diagram of using Bayesian wavelet shrinkage to estimate the EWS from the Haar-Fisz (H-F) wavelet periodogram. Note: If the smoothing wavelet is Haar, the smoothing wavelet (SW) discrete wavelet transform (DWT) is not required.

The sampled EWS estimates are also produced using importance sampling as described in section 6.1. For the uniform prior, the sample wavelet coefficients, x, were sampled from a Gaussian distribution with mean $\hat{\beta}_{l,m}$ and variance $\mathrm{Var}[\beta_{l,m}|h_{l,m}]$ truncated on the interval $[-1, 1]$. When the smoothing was performed using the Laplace prior we also sampled from the Gaussian distribution with the same mean and variance but without truncation. The flow diagram in figure 7.7, depicts the process required to produce the Bayesian wavelet shrinkage estimate of the Haar-Fisz transformed raw wavelet periodogram.

To assess the performance of the Bayesian wavelet shrinkage of the Haar-Fisz wavelet periodogram we applied the method to two simulated LSW processes. These were the same LSW processes X_t and Y_t with Gaussian innovations described in section 3.4, previously analysed. As we knew the true EWS we could compare our Bayesian Haar-Fisz EWS estimator with the truth and obtain the AMSE.

For each LSW process, we compared the results of using the uniform and the Laplace prior. We produced spectral estimates with the Laplace prior using Daubechies Extremal Phase smoothing wavelet with two to ten vanishing moments without cycle spinning. Plots of the estimates without cycle spinning are available in Appendix B. The estimate which produced the smallest AMSE was then cycle spun twenty times, and we used this Laplace prior spectral estimate to compare with the uniform prior spectral estimate.

To calculate the CI for the cycle spun data, we sampled an estimate of the cycle spun N (which is defined by the user). Each of these samples are then returned to the original location. The CI are calculated from the quantiles of all the samples.

We plotted the EWS for all scales to determine which scales warranted closer examination. Then we plotted the EWS at these scales with their corresponding 50% and 90% quantiles.

7.5.1 Piecewise Constant EWS

The plots in figure 7.8 are the estimates of the EWS for the LSW process X_t obtained using the Haar smoothing wavelet with the uniform prior in figure (a) and Daubechies Extremal phase wavelet with 3 vanishing moments with the Laplace prior in figure (b). For both estimates there is a very small amount of leakage of power from $j = 9$ into $j = 8, 7, 6$.

Both estimates detect the second change point in $S_9(z)$ reasonably well. However, it seems the first change point is more difficult to detect and locate accurately. The Haar

(a) Uniform Mixture Prior EWS Estimate (b) Laplace Mixture Prior EWS Estimate

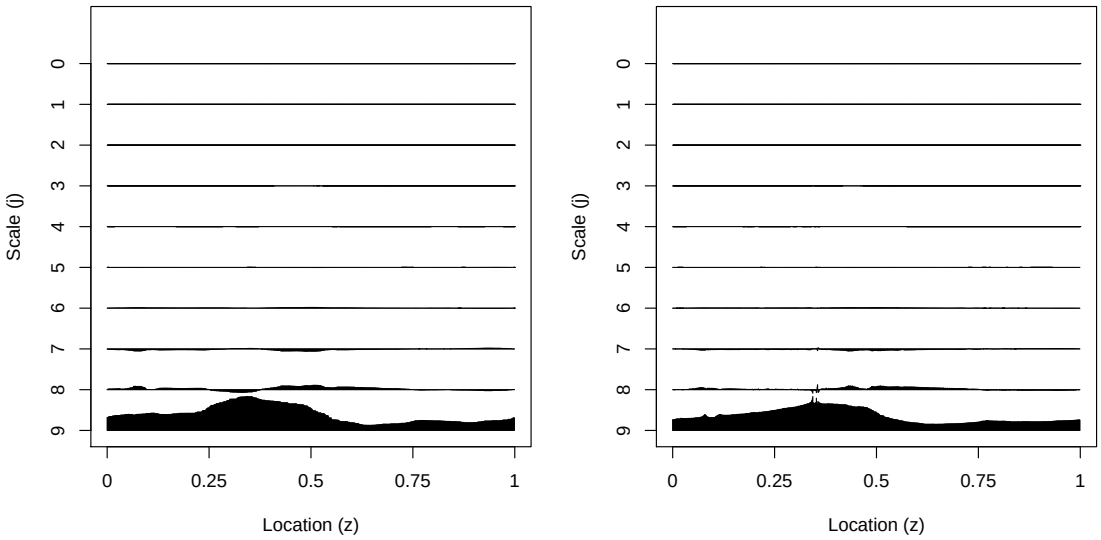


Figure 7.8: Plots of the estimated EWS of X_t, with 20 cycle spins. (a) is the estimated EWS using the Haar wavelet and uniform prior. (b) is the estimated EWS using D_3 and Laplace prior.

wavelet appears to perform slightly better at locating the changing point and determining the correct magnitude. The Laplace estimate gives the impression there is a gradual increase in power over the first third of the spectrum rather than a sudden change.

The plots of the EWS of scales $j = 9, 8, 7$ with 50% and 90% confidence intervals appear in figure 7.9. For all three scales in these plots, there is evidence of a Gibbs phenomenon for all of the D_3 estimates close to the first change point, $z \approx {}^1/_3$, as shown in (b), (d) and (f). This results in a very large CI at this point, conveying the estimator's uncertainty. This could be used as evidence to suggest that the Haar wavelet is maybe a more appropriate smoothing wavelet as there might be a discontinuity at that point.

The CI for all of the uniform estimates are smaller than the Laplace estimates. This could be caused by using the truncated Gaussian distribution to generate observations from the posterior distribution. The smaller confidence intervals result in a larger proportion of the true EWS falling outside of the CI of the uniform estimate compared to the Laplace estimate, especially for the 50% CI.

Neither the Laplace or uniform 90% CI capture the change points at $z = {}^1/_3$ or $z = {}^8/_{15}$. However, this could be a result of cycle spinning which can smooth estimates over discontinuities.

The 90% CI for the uniform $S_7(z)$ does not capture the true EWS for a small period around $z \approx {}^1/_2$, however as we are plotting the 90% CI there is a 10% chance the true EWS will not be within the CI. As it is for such a small period of time, and the CI almost

captures the actual value, the slight deviation is negligible.

7.5.2 Slowly Evolving EWS

The plots in figure 7.10 are the Uniform and Laplace estimates of the slowly evolving EWS. The wavelet with the smallest AMSE for the Laplace prior was a Daubechies extremal phase wavelet with 8 vanishing moments. Both wavelets did well at recovering the true signal of the EWS, especially given the piecewise constant nature of the Haar wavelet, even the estimate without cycle spinning was more accurate than expected (see appendix B).

There is a small amount of leakage between scales, particularly from $j = 6$ into $j = 5$. However, this is considerably less than the previously developed methods.

On closer inspection of the estimates in figure 7.11, it appears that the uniform prior and Haar wavelet are better at estimating the $S_6(z)$ in figure (c) at the start and end of the interval compared to the Laplace estimates (see figure (d)).

By examining the 50% and 90% confidence intervals of $S_j(z)$ for $j = 7, 6, 5$, we gain a sense of how much power has leaked into scales $j = 5$ and 7 from 6 and the accuracy of the estimates. The power leakage from $\hat{S}_6(z)$ to $\hat{S}_5(z)$ is very little and there is strong evidence from the CI for both prior distributions that there is no significant difference between our estimated values and zero.

The true EWS value for $S_7(z)$ for all z should be zero. However, the uniform estimate suggests that at $z \approx 0.2$ there is negative power. As a spectrum cannot contain a negative amount of power this does not make any practical sense, but should indicate that further investigations should be made into estimating the spectrum at this scale and adjacent scales.

7.5.3 Conclusion

Using the Haar-Fisz transformation helps to stabilise the variance, bring the distribution closer to Gaussianity and reduces the correlation structure between locations. All of these benefits help to simplify the Bayesian calculations and improve the computational expense. Whereas our assumption of independence for Bayesian wavelet shrinkage of the log periodogram appears to be invalid and results in inaccurate estimates, it appears the Haar-Fisz transformation aids convergence.

With larger confidence intervals there is a higher probability the true EWS is within the

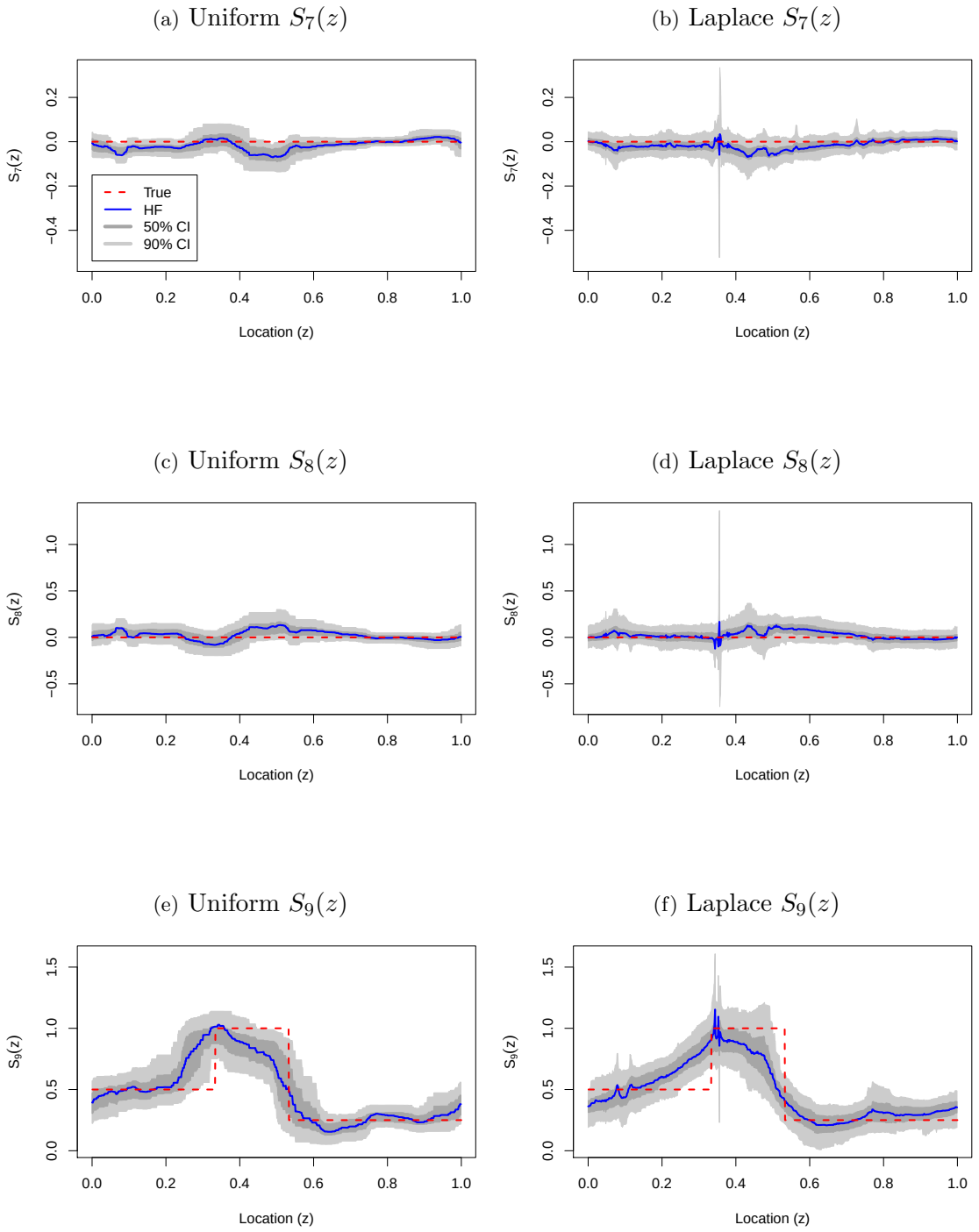


Figure 7.9: Plots of the estimated EWS with 50% and 90% CI, at scale $j = 9, 8, 7$ using Bayesian wavelet shrinkage on X_t's Haar-Fisz transformed wavelet periodogram with 20 cycle spins. Figures (a), (c) and (e) have been smoothed using the Haar wavelet and Uniform prior. Figures (b), (d) and (f) have been smoothed using the Laplace prior and D_3 wavelet.

(a) Uniform Mixture Prior EWS Estimate (b) Laplace Mixture Prior EWS Estimate

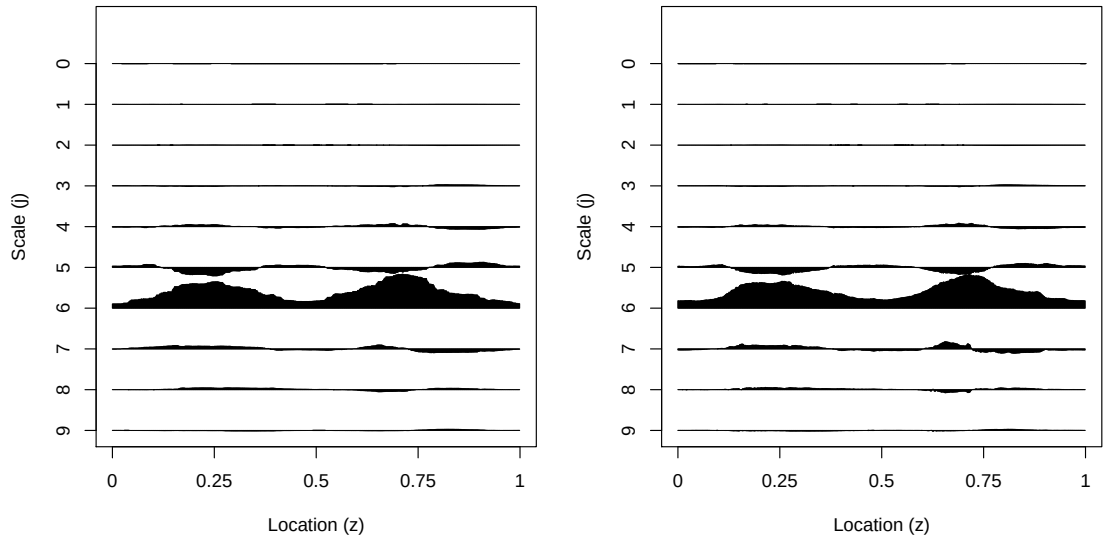


Figure 7.10: Plots of the estimated EWS of Y_t, with 20 cycle spins. (a) is the estimated EWS using the Haar wavelet and uniform mixture prior. (b) is the estimated EWS using D_8 and Laplace mixture prior.

CI values. However, it also suggests the estimate is inaccurate and possibly an alternate wavelet would be more appropriate. If the nature of the EWS varied greatly between scales it maybe more appropriate to considered different smoothing wavelets for each scale.

There is evidence to suggest varying the smoothing wavelets between scales may improve our estimates. Especially if there is the possibility that the nature of the power is different between scales, i.e. piecewise constant or slowly evolving. This may also help to further reduce the effect the leakage of power has upon the recovered EWS. However, considering how well the uniform prior and Haar wavelet performed at recovering the slowly evolving EWS, this does imply the method is fairly robust to the selection of smoothing wavelet.

Overall this method works well at estimating the EWS and producing meaningful CI to assess the validity of our estimate.

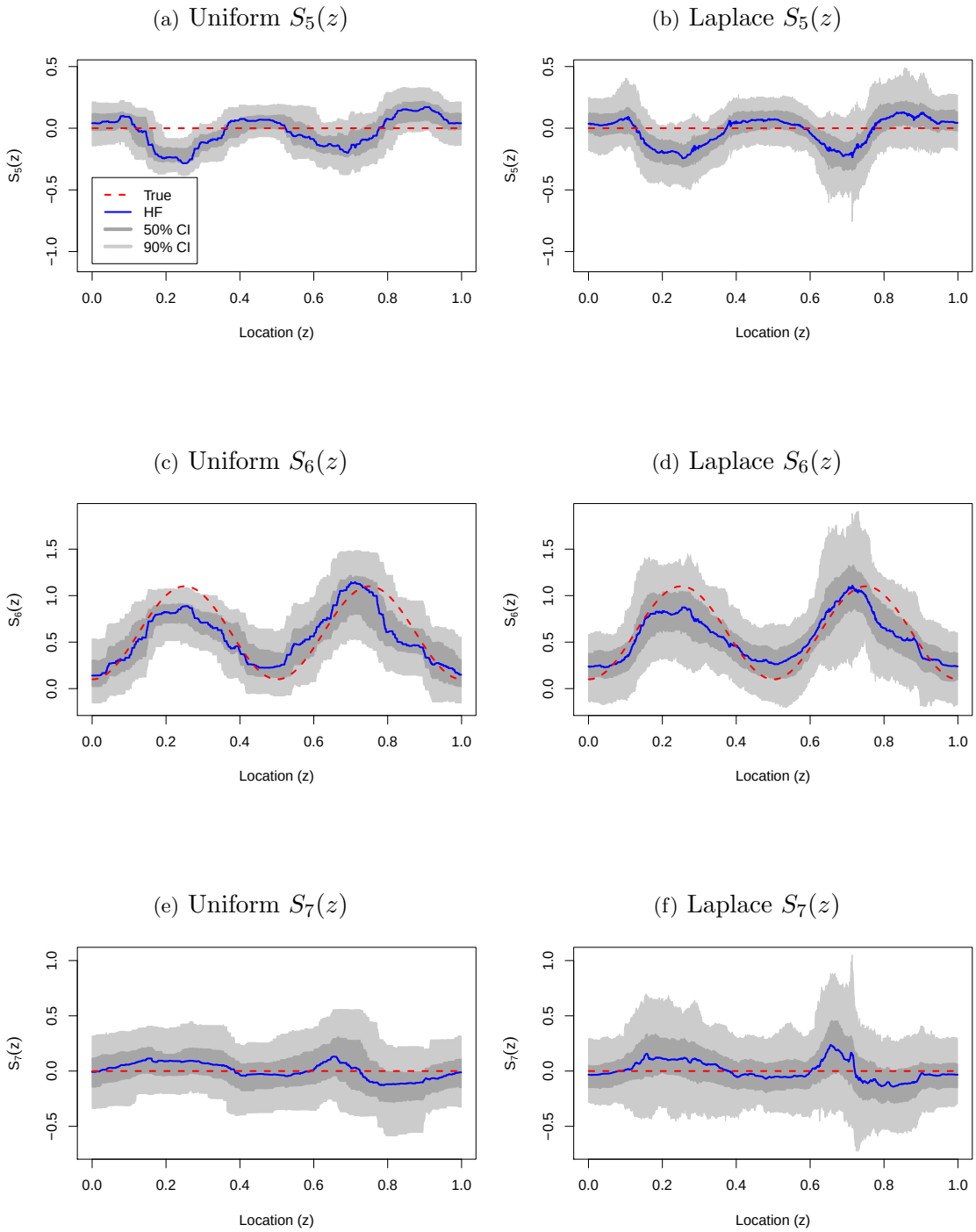


Figure 7.11: Plots of the estimated EWS with 50% and 90% CI, at scale $j = 7, 6, 5$ using Bayesian wavelet shrinkage on Y_t's Haar-Fisz transformed wavelet periodogram with 20 cycle spins. Figures (a), (c) and (e) have been smoothed using the Haar wavelet and uniform mixture prior, whereas (b), (d) and (f) have been smoothed using the Laplace mixture prior and D_8 wavelet.

Chapter 8

Conclusion

The aim of this thesis was to produce methods for smoothing the evolutionary wavelet spectrum (EWS) of the locally stationary wavelet (LSW) process with confidence intervals. In conclusion we shall first highlight the important aspects of each chapter, finishing with a a paragraph summarising the main findings of this thesis.

Chapter 3: Foundation Work

By investigating some of the statistical properties of the empirical wavelet coefficients and raw wavelet periodogram, we:

- determined the asymptotic mean, variance and distribution of the coefficients and periodogram;
- numerically verified our findings for a range of LSW processes with different innovations.

Our findings suggested:

- we could assume asymptotic independence between scales and locations, if the locations were not 'close';
- as the support of the wavelet increased the wavelet coefficients were asymptotically Gaussian and hence the raw wavelet periodogram was approximately multivariate chi-square.

Chapter 4: Naive Spectral Confidence Interval using the Central Moving Average

In this chapter

- we analytically developed a smoothing technique and 95% confidence interval based on the central moving average,
- tested our method on the simulated LSW processes from chapter 3.

We found that with this simple method

- the estimate and CI were easy to compute,
- more consideration should be given to the selection of bin width,
- we should consider calculating the variance using different close locations at scales to improve accuracy.

Chapter 5: Bayesian Modelling of the Log Transformed EWS

In this chapter we

- theoretically established a model based on the findings in chapter 3 of the log wavelet periodogram,
- developed a Bayesian wavelet shrinkage method for the log transformed raw wavelet periodogram,
- calculated the required components for the Gaussian and Laplace prior.
- developed results for the asymptotic convergence of Bayesian wavelet shrinkage of the log transformed raw wavelet periodogram, assuming
  (a) our estimate was a function in the Besov ball, $B^r_{p,q}(A)$,
  (b) our data set was independent and identically distributed.

Although many of the components could be calculated exactly analytically, some components required numerical integration because of the complex error distribution of the log transformed raw wavelet periodogram

Chapter 6: Simulation Results of the Bayesian Log-EWS Estimator

The method developed in chapter 5 was programmed and tested using the simulated time series in chapter 3. We compared the two prior distributions and found:

- the Laplace pdf has fatter tails than the Gaussian pdf;
- the posterior distribution using the different prior distributions with the same variance were very different, but the shrinkage functions were similar;
- using the Laplace prior required less numerical approximations.

Using the Laplace prior, we tested our method on the LSW process, our results indicated that

- the error distribution was reasonably close to our theory;
- the CI conveyed how inaccurate the estimates were;
- more consideration should be given to the hyperparameter estimates;
- the assumption of independence may be invalid and possibly a multivariate approach would be more appropriate.

Chapter 7: Bayesian Wavelet Shrinkage of the Haar-Fisz EWS

In this chapter we:

- numerically investigated the distribution and correlation structure of the Haar-Fisz wavelet periodogram;
- created CI for the soft shrinkage estimates;
- developed Bayesian wavelet shrinkage of the Haar-Fisz wavelet periodogram.

From our investigations, we found:

- using the Haar-Fisz transformation of the raw wavelet periodogram:
  (a) stabilises the variance;
  (b) brings the data closer to Gaussianity;
  (c) reduces the correlation structure between locations.
- using Bayesian wavelet shrinkage of the Haar-Fisz wavelet periodogram:

(a) was simple to calculate and numerically efficient;

(b) produced an accurate EWS estimate;

(c) was reasonably robust to the choice of smoothing wavelet;

(d) produced associated CI which were insightful.

Further work into Bayesian modelling of the Haar-Fisz periodogram would be

- to improve the hyperparameter estimation to make it more efficient;
- to test how well this method performed on LSW processes with non-Gaussian innovations.

In this thesis we developed three methods with confidence intervals to estimate the EWS. With the development of each method we gained a greater understanding of the statistical properties of the raw wavelet periodogram and which techniques were more successful than others. Further work might include using the posterior median rather than the posterior mean for estimation, and improving the analytical calculation of the estimator's variance and hyperparameter determination. Also, the methods described in this thesis are also limited to regularly spaced time series, in practice a time series of this nature may not be available. Therefore, we could also consider looking into producing EWS estimates with confidence intervals of irregularly-spaced LSW processes.

Bibliography

Abramovich, F., Bailey, T. C. and Sapatinas, T. (2000), 'Wavelet analysis and its statistical applications', *Journal of the Royal Statistical Society: Series D* **49**, 1–29.

Abramovich, F., Sapatinas, T. and Silverman, B. W. (1998), 'Wavelet thresholding via a Bayesian approach', *Journal of the Royal Statistical Society: Series B* **60**, 725–749.

Antoniadis, A. (2007), 'Wavelet methods in statistics: Some recent developments and their applications', *Statistics Surveys* **1**, 16.

Barber, S., Nason, G. P. and Silverman, B. W. (2001), 'Posterior probability intervals for wavelet thresholding', *Journal of the Royal Statistical Society: Series B* **64**, 189–205.

Bartlett, M. S. (1955), *An Introduction to Stochastic Processes with Special Reference to Methods and Applications*, Cambridge University Press, Cambridge.

Besov, O. V. (1959), 'On a family of function spaces: Embedding and extension theorems.', *Doklady Akademii Nauk SSSR (in Russian)* **126**, 1163–1165.

Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (2008), *Time Series Analysis: Forecasting and Control*, Vol. 734 of *Wiley Series in Probability and Statistics*, eighth edn, John Wiley & Sons, New York.

Brillinger, D. R. (1994), Some asymptotics of wavelet fits in the stationary error case, Technical report, Department of Statistics, University of California, Berkeley.

Bruce, A. G. and Gao, H.-Y. (1996), 'Understanding waveshrink: Variance and bias estimation', *Biometrika* **83**, 727–745.

Burrus, C. S., Gopinath, R. A. and Guo, H. (1998), *Introduction to Wavelets and Wavelet Transforms: A Primer*, Prentice Hall, London.

Byrd, R. H., Lu, P., Nocedal, J. and Zhu, C. (1995), 'A limited memory algorithm for bound constrained optimization', *SIAM Journal of Scientific Computing* **16**, 1190–1208.

Chatfield, C. (2004), *The Analysis of Time Series: An Introduction*, sixth edn, Chapman and Hall/CRC, London.

Chipman, H. A., Kolaczyk, E. D. and McCulloch, R. E. (1997), 'Adaptive Bayesian wavelet shrinkage', *Journal of the American Statistical Association* **92**, 1413–1421.

Chui, C. K. (1997), *Wavelets: A Mathematical Tool for Signal Analysis*, SIAM Monographs on Mathematical Modelling and Computation, SIAM, Philadelphia.

Clyde, M. and George, E. I. (1998), 'Robust empirical Bayes estimation in wavelets', *ISDS Discussion Paper* pp. 98–21.

Clyde, M., Parmigiani, G. and Vidakovic, B. (1998), 'Multiple shrinkage and subset selection in wavelets', *Biometrika* **85**, 394–402.

Cohen, A., Daubechies, I. and Vial, P. (1993), 'Wavelets on the interval and fast wavelet transforms', *Applied and Computational Harmonic Analysis* **1**, 54–81.

Coifman, R. R. and Donoho, D. L. (1995), Translation-invariant de-noising, *in* A. Antoniadis and G. Oppenheim, eds, 'Wavelets and Statistics', Vol. 103 of *Lecture Notes in Statistics*, Springer-Verlag, Berlin, pp. 125–15.

Csardi, G. (2010), '**igraph**: Network analysis and visualization', `R` package, version 0.6-3. **URL:** *http://cran.r-project.org/web/packages/igraph/index.html*

Dahlhaus, R. (1997), 'Fitting time series models to nonstationary processes', *Annals of Statistics* **25**, 1–37.

Daubechies, I. (1992), *Ten Lectures on Wavelets*, CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM, Philadelphia.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977), 'Maximum likelihood from incomplete data via the EM algorithm', *Journal of the Royal Statistical Society: Series B* **39**, 1–38.

DeVore, R. A., Jawerth, B. and Popov, V. (1992), 'Compression of wavelet decompositions', *American Journal of Mathematics* **114**, 737–785.

DeVore, R. A. and Popov, V. (1988), 'Interpolation of Besov spaces', *Transactions of the American Mathematical Society* **305**, 397–414.

Donoho, D. L. and Johnstone, I. M. (1994*a*), 'Ideal denoising in an orthonormal basis chosen from a library of bases', *Comptes Rendus Academic Science, Series I* **319**, 1317–1322.

Donoho, D. L. and Johnstone, I. M. (1994*b*), 'Ideal spatial adaption via wavelet shrinkage', *Biometrika* **81**, 425–455.

Donoho, D. L. and Johnstone, I. M. (1995), 'Adapting to unknown smoothness via wavelet shrinkage', *Journal of the American Statistical Association* **90**, 1200–1224.

Donoho, D. L., Johnstone, I. M. and Picard, G. K. D. (1995), 'Wavelet shrinkage: Asymptopia?', *Journal of the Royal Statistical Society: Series B* **52**, 301–369.

Donoho, D. L., Johnstone, I. M. and Picard, G. K. D. (1996), 'Density estimation by wavelet thresholding', *The Annals of Statistics* **24**, 508–539.

Eckley, I. (2001), Wavelet Methods for Time Series and Spatial Data, PhD thesis, University of Bristol.

Fisz, M. (1955), 'The limiting distribution of a function of two independent random variables and its statistical application', *Colloquium Mathematicum* **3**, 138–146.

Folland, G. B. (2003), *Fourier Analysis and its Applications*, Pure and Applied Undergraduate Texts, American Mathematical Society, Rhode Island.

Fryzlewicz, P. and Nason, G. P. (2006), 'A Haar-Fisz estimation of the evolutionary wavelet spectra', *Journal of the Royal Statistical Society: Series B* **68**, 611–634.

Gao, H.-Y. and Bruce, A. G. (1997), 'Waveshrink with firm shrinkage', *Statistica Sinica* **7**, 855–874.

George, E. I. and Foster, D. P. (1997), Empirical Bayes variable selection, Technical report, University of Texas, Austin.

Goswami, J. C. and Chan, A. K. (2011), *Fundamentals of Wavelets: Theory, Algorithms, and Applications*, Wiley Series in Microwave and Optical Engineering, second edn, John Wiley & Sons, New York.

Gradshteyn, I. S. and Ryzhik, I. M. (1980), *Table of Integrals, Series and Products*, second edn, Academic Press Inc, San Diego.

Green, P. J. and Silverman, B. W. (1993), *Nonparametric Regression and Generalized Linear Models: A roughness penalty approach*, Chapman and Hall/CRC Texts in Statistical Science, Chapman and Hall/CRC, London.

Grenander, U. and Rosenblatt, M. (1957), *Statistical Analysis of Stationary Time Series*, John Wiley & Sons, New York.

Grenier, Y. (1983), 'Time-dependent ARMA modeling of nonstationary signals', *IEEE Transactions on Acoustics, Speech, and Signal Processing* **31**, 899–911.

Haar, A. (1911), 'Zur theorie der orthogonalen funktionensysteme', *Mathematische Annalen* **71**, 38–53.

Hall, P. and Turlach, B. A. (1997), 'Interpolation methods for nonlinear wavelet regression with irregularly spaced design', *The Annals of Statistics* **25**, 1912–1925.

Isserlis, L. (1918), 'On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables', *Biometrika* **12**, 134–139.

Johnson, N. L. (1949), 'Systems of frequency curves generated by methods of translation', *Biometrika* **36**, 149–176.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994), *Continuous Univariate Distributions, Volume 1*, 2 edn, John Wiley & Sons, New York.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995), *Continuous Univariate Distributions, Volume 2*, 2 edn, John Wiley & Sons, New York.

Johnstone, I. M. and Silverman, B. W. (1998), Empirical Bayes approaches to mixture problems and wavelet regression, Technical report, Statistics Department, Stanford University.

Johnstone, I. M. and Silverman, B. W. (2005), 'Empirical Bayes selection of wavelet thresholds', *The Annals of Statistics* **33**, 1700–1752.

Kirchgaẗssner, G., Wolters, J. and Hassler, U. (2013), *Introduction to Modern Time Series Analysis*, Springer Texts in Business and Economics, second edn, Springer-Verlag, Berlin.

Lu, H. H.-S., Huang, S.-Y. and Lin, F.-J. (2003), 'Generalized cross-validation for wavelet shrinkage in nonparametric mixed effects models', *Journal of Computational and Graphical Statistics* **12**, 714–730.

Mallat, S. G. (1989), 'A theory for multiresolution signal decomposition: The wavelet representation', *IEEE Transactions on Analysis and Machine Intelligence* **11**, 674–693.

Meyer, Y. (1992), *Wavelets and Operators*, Cambridge University Press, Cambridge.

Moulin, P. (1994), 'Wavelet thresholding techniques for power spectrum estimation', *IEE Transactions on Signal Processing* **42**, 3126–3136.

Nadaraya, E. A. (1964), 'On estimating regression', *Theory of Probability and its Applications* **9**, 157–159.

Nason, G. P. (1995), Choice of threshold parameter in wavelet function estimation, *in* A. Antoniadis and G. Oppenheim, eds, 'Wavelets and Statistics', Vol. 103 of *Lecture Notes in Statistics*, Springer-Verlag, Berlin.

Nason, G. P. (2001), 'Choice of wavelet smoothness, primary resolution and threshold in wavelet shrinkage', *Statistics and Computing* **12**, 219–227.

Nason, G. P. (2008), *Wavelet Methods in Statistics with R*, Springer, New York.

Nason, G. P. (2012), '**wavethresh**: Wavelets Statistics and Transforms', `R` package, version 4.6.1.
URL: *http://cran.r-project.org/web/packages/wavethresh/index.html*

Nason, G. P. (2013), 'A test for second-order stationarity and approximate confidence intervals for localized autocovariances for locally stationary time series', *Journal of the Royal Statistical Society: Series B* **75**, to appear.

Nason, G. P. and Silverman, B. W. (1994), 'The discrete wavelet transform in S', *Journal of Computational and Graphical Statistics* **3**, 163–191.

Nason, G. P. and Silverman, B. W. (1995), The stationary wavelet transform and some applications, *in* A. Antoniadis and G. Oppenheim, eds, 'Wavelets and Statistics', Vol. 103 of *Lecture Notes in Statistics*, Springer-Verlag, Berlin.

Nason, G. P., von Sachs, R. and Kroisandt, G. (2000), 'Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum', *Journal of the Royal Statistical Society: Series B* **62**, 271–292.

Neumann, M. H. (1996), 'Spectral density estimation via nonlinear wavelet methods for stationary non-Gaussian time series', *Journal of Time Series Analysis* **17**, 601–633.

Neumann, M. H. and von Sachs, R. (1995), Wavelet Thresholding: Beyond the Gaussian I.I.D. Situation, *in* A. Antoniadis and G. Oppenheim, eds, 'Wavelets and Statistics', Vol. 103 of *Lecture Notes in Statistics*, Springer-Verlag, Berlin, pp. 301–329.

Park, C. G., Oh, H.-S. and Lee, H. (2008), 'Bayesian selection of primary resolution and wavelet basis functions for wavelet regression', *Comput. Stat.* **23**, 291–302.

Parzen, E. (1983), Autoregressive spectral estimation, *in* D. Brillinger and P. Krishnaiah, eds, 'Time Series in the Frequency Domain', Vol. 3 of *Handbook of Statistics*, Elsevier, Amsterdam, pp. 221–247.

Pensky, M., Vidakovic, B. and De Cannditiis, D. (2007), 'Bayesian decision theoretic scale-adaptive estimation of a log-spectral density', *Statistica Sinica* **17**, 635–666.

Percival, D. B. and Walden, A. T. (2000), *Wavelet Methods for Time Series Analysis*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, Cambridge.

Priestley, M. B. (1965), 'Evolutionary spectra and non-stationary processes', *Journal of the Royal Statistical Society: Series B* **27**, 204–237.

Priestley, M. B. (1994), *Spectral Analysis and Time Series*, Probability and Mathematical Statistics, Academic Press, London.

Ramsay, J. O. and Silverman, B. W. (1997), *Functional Data Analysis*, Springer-Verlag, Berlin.

R Development Core Team (2012), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. **URL:** *http://www.R-project.org*

Ross, S. (2010), *A First Course in Probability*, Pearson Education, New York.

Rudzkis, R., Saulis, L. and Staulevicus, V. (1978), 'A general lemma on probabilities of large deviations', *Lituanian Mathemtics Journal* **18**, 226–238.

Ruggeri, F. and Vidakovic, B. (2005), Bayesian modeling in the wavelet domain, *in* D. Dey and C. Rao, eds, 'Bayesian Thinking Modeling and Computation', Vol. 25 of *Handbook of Statistics*, Elsevier, Amsterdam, pp. 315–338.

Shannon, C. E. (1949), 'Communication in the presence of noise', *Proc. Institute of Radio Engineers* **37**, 10–21.

Silverman, B. W. (1985), 'Some aspects of the spline smoothing approach to non-parametric regression curve fitting', *Journal of the Royal Statistical Society: Series B* **47**, 1–52.

Simonoff, J. S. (1996), *Smoothing Methods in Statistics*, Springer Series in Statistics, Springer-Verlag, Berlin.

Statulevicius, V. and Jakimavicius, D. (1988), 'Estimates of semiinvariants and centered moments of stochastic processes with mixing. I', *Lithuanian Mathematical Journal* **28**, 67–80.

Stein, C. M. (1981), 'Estimating the Mean of a Multivariate Normal Distribution', *Annals of Statistics* **9**, 1135–1151.

Stein, E. M. and Shakarchi, R. (2003), *Fourier Analysis: An Introduction*, Princeton Lectures in Analysis, Princeton University Press, Princeton.

Subba Rao, T. (1970), 'The fitting of nonstationary time series models with time dependent parameters', *Journal of the Royal Statistical Society: Series B* **32**, 312–322.

Sweldens, W. (1996), 'Wavelets and the lifting scheme: A 5 minute tour', *Z. Angew. Math. Mech.* **76 (Suppl. 2)**, 41–44.

Triebel, H. (2000), *Theory of Function Spaces II*, Monographs in Mathematics, second edn, Birkhaüser Verlag, Basel.

Triebel, H. (2006), *Theory of Function Spaces III*, Monographs in Mathematics, Birkhaüser Verlag, Basel.

Van Bellegem, S. and von Sachs, R. (2008), 'Locally adaptive estimation of evolutionary wavelet spectra', *Annals of Statistics* **36**, 1879–1924.

Vidakovic, B. (1998*a*), 'Nonlinear wavelet shrinkage with Bayes Rule and Bayes factors', *Journal of the American Statistical Association* **93**, 173–179.

Vidakovic, B. (1998*b*), Wavelet-Based Nonparametric Bayes Methods, *in* 'Practical Nonparametric and Semiparametric Bayesian Statistics', Vol. 133 of *Lecture notes in Statistics*, Springer-Verlag, Berlin, pp. 133–155.

Vidakovic, B. (1999), *Statistical Modeling by Wavelets*, Wiley Series in Probability and Statistics, John Wiley & Sons, New York.

Wahba, G. (1990), *Spline Models for Observational Data*, SIAM, Philadelphia.

Watson, G. S. (1964), 'Smooth regression analysis', *The Indian Journal of Statistics, Series A (1961-2002)* **26**, 359–372.

Appendix A

Appendix of Proofs

A.1 Proofs for the Bayesian Log-Periodogram

A.1.1 Gaussian Mixture Prior

Proof (Lemma 5.3.1)**.** Let the Gaussian part of the mixture prior ξ equal to the pdf of $\mathcal{N}(0, \nu_l^{-2})$, then (5.14) becomes

$$\begin{aligned} Q_i(h) = & \int_{-\infty}^{\infty} \frac{\nu_l y^i}{\sqrt{2\pi}} \exp\left\{-\frac{\nu_l^2 y^2}{2}\right\} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(h-y)^2}{2\sigma^2}\right\} \mathrm{d}y \\ = & \ \nu_l \left(\frac{1}{2\pi}\right)^{1/2} \int_{-\infty}^{\infty} \frac{y^i}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}\left(y^2\left[1+(\nu_l\sigma)^2\right] - 2hy + h^2\right)\right\} \mathrm{d}y. \end{aligned} \tag{A.1}$$

Let $\kappa_l = 1 + (\nu_l\sigma)^2$, where $\sigma^2 = \pi^2/2$ (the variance of the error distribution), which means that equation (A.1) can be written as

$$Q_i(h) = \nu_l \left(\frac{1}{2\pi}\right)^{1/2} \int_{-\infty}^{\infty} \frac{y^i}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{\kappa_l}{2\sigma^2}\left[y^2 - 2y\frac{h}{\kappa_l} + \frac{h^2}{\kappa_l}\right]\right\} \mathrm{d}y. \tag{A.2}$$

By completing the square it can be shown that the exponential part of equation (A.2) can be re-written as

$$\begin{aligned} -\frac{\kappa_l}{2\sigma^2}\left[y^2 - \frac{2hy}{\kappa_l} + \frac{h^2}{\kappa_l}\right] &= -\frac{\kappa_l}{2\sigma^2}\left[\left(y - \frac{h}{\kappa_l}\right)^2 + \frac{h^2}{\kappa_l} - \frac{h^2}{\kappa_l^2}\right] \\ &= -\frac{\kappa_l}{2\sigma^2}\left(y - \frac{h}{\kappa_l}\right)^2 + \frac{h^2}{2\sigma^2}\left(\frac{1}{\kappa_l} - 1\right). \end{aligned} \tag{A.3}$$

Then, if we also then let $\mu_g = h\kappa_l^{-1}$ and $\sigma_g^2 = \sigma^2\kappa_l^{-1}$ in equation (A.3), then $Q_i(h)$ can be written as

$$Q_i(h) = \frac{\nu_l}{\sqrt{2\pi\kappa_l}} \exp\left\{\frac{h^2}{\pi^2}\left(\frac{1}{\kappa_l} - 1\right)\right\} \int_{-\infty}^{\infty} y^i \, \varphi_{\sigma_g}(y - \mu_g)\mathrm{d}y,$$

where $\varphi_{\sigma_g}(\cdot)$ is the Gaussian distribution with variance σ_g^2. □

Proof (Lemma 5.3.2)**.** Let ξ the pdf of $\mathcal{N}(0, \nu_l^{-2})$, then $Q_i^*(h)$ is

$$\begin{aligned} Q_i^*(h) &= \nu_l \int_{-\infty}^{\infty} \frac{y^i}{\sqrt{2\,\pi}} \exp\left\{-\frac{\nu_l^2\, y^2}{2}\right\} \left(\frac{\gamma^*}{\pi}\right)^{1/2} \exp\left\{\frac{1}{2}(h-y) - \gamma^* e^{(h-y)}\right\} \mathrm{d}y \\ &= \frac{\nu_l}{\pi}\left(\frac{1}{2}\right)^{1/2} \int_{-\infty}^{\infty} y^i \exp\left\{-\frac{\nu_l^2\, y^2}{2}\right\} \sqrt{\gamma^*} \exp\left\{\frac{1}{2}(h-y) - \gamma^* e^{(h-y)}\right\} \mathrm{d}y, \end{aligned}$$

using the definition of f_E from (5.4). □

Proof (Lemma 5.3.3)**.** Apply the Fourier transform to (5.3.2) with respect to h, giving

$$\begin{aligned} \hat{Q}_i^*(\omega) &= \frac{\nu_l}{\pi}\left(\frac{1}{2}\right)^{1/2} \int_{-\infty}^{\infty} y^i \exp\left\{-\frac{\nu_l^2\, y^2}{2}\right\} \\ &\quad \times \int_{-\infty}^{\infty} \sqrt{\gamma^*} \exp\left\{\frac{1}{2}(h-y) - \gamma^* e^{(h-y)}\right\} e^{-\mathrm{i}2\pi h\omega} \,\mathrm{d}h \,\mathrm{d}y. \end{aligned}$$

We now perform a substitution of variable for h, let

$$x = \gamma^* e^{(h-y)} \quad \text{and} \quad h = y + [\log(x) - \log(\gamma^*)], \qquad \Rightarrow \qquad \frac{\mathrm{d}h}{\mathrm{d}x} = \frac{1}{x},$$

which gives

$$\hat{Q}_i^*(\omega) = (\gamma^*)^{\mathrm{i}2\pi\omega} \frac{\nu_l}{\pi\sqrt{2}} \int_{-\infty}^{\infty} y^i \exp\left\{-\frac{\nu_l^2\, y^2}{2} - \mathrm{i}2\pi\omega y\right\} \int_0^{\infty} x^{-1/2-\mathrm{i}2\pi\omega} e^{-x} \,\mathrm{d}x \,\mathrm{d}y, \quad \text{(A.4)}$$

for $-\pi \leq \omega \leq \pi$. The integral over x can be re-written as

$$\int_0^{\infty} x^{-1/2-\mathrm{i}2\pi\omega} e^{-x} \mathrm{d}x = \Gamma\left(\frac{1}{2} - \mathrm{i}2\pi\omega\right),$$

where $\Gamma(\cdot)$ is the gamma function. Taking the y exponential part of equation (A.4) it can

be rewritten as

$$\begin{aligned} -\frac{\nu_l^2\, y^2}{2} - \mathrm{i}2\pi\omega y = & -\frac{\nu_l^2}{2}\left(y^2 + 4\frac{\mathrm{i}\pi\omega}{\nu_l^2}y\right) \\ = & -\frac{\nu_l^2}{2}\left(y + \frac{\mathrm{i}2\pi\omega}{\nu_l^2}\right)^2 - 2\left(\frac{\pi\omega}{\nu_l}\right)^2, \end{aligned}$$

using completing the square technique. Substituting this back into equation (A.4) results in

$$\begin{aligned} \hat{Q}_i^*(\omega) = & (\gamma^*)^{\mathrm{i}2\pi\omega}\, \frac{e^{-2(\pi\omega/\nu_l)^2}}{\sqrt{\pi}}\, \Gamma\left(\frac{1}{2} - \mathrm{i}2\pi\omega\right) \\ & \times \int_{-\infty}^{\infty} y^i\, \frac{\nu_l}{\sqrt{2\pi}}\, \exp\left\{-\frac{\nu_l^2}{2}\left(y + \frac{\mathrm{i}2\pi\omega}{\nu_l^2}\right)^2\right\}\, \mathrm{d}y \\ = & (\gamma^*)^{\mathrm{i}2\pi\omega}\, \frac{e^{-2(\pi\omega/\nu_l)^2}}{\sqrt{\pi}}\, \Gamma\left(\frac{1}{2} - \mathrm{i}2\pi\omega\right) \int_{-\infty}^{\infty} y^i\, \nu_l\, \varphi\left(\nu_l\left[y + \frac{\mathrm{i}2\pi\omega}{\nu_l^2}\right]\right)\, \mathrm{d}y \end{aligned}$$

□

Proof (Lemma 5.3.4)**.** (a) Let $i = 0$, lemma 5.3.1. As the integral of a probability density function over the whole support of the variable is 1 we obtain the desired result.

(b) Let $i = 1$, lemma 5.3.1. As μ_g is the mean of $\varphi_{\sigma_g}(y - \mu_g)$, which is obtained from the integral we wish to evaluate, we obtain the desired result.

(c) Let $i = 2$ in lemma 5.3.1, then as the second moment of a Gaussian distribution is

$$\int_{-\infty}^{\infty} \frac{x^2}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)\mathrm{d}x = \mu^2 + \sigma^2.$$

we find

$$\begin{aligned} Q_2(h) &= \frac{\nu_l}{\sqrt{2\pi\kappa_l}}\, \exp\left\{\frac{h^2}{\pi^2}\left(\frac{1}{\kappa_l} - 1\right)\right\} \int_{-\infty}^{\infty} y^2\, \varphi_{\sigma_g}(y - \mu_g)\mathrm{d}y, \\ &= \frac{\nu_l}{\sqrt{2\pi\kappa_l}}\, \exp\left\{\frac{h^2}{\pi^2}\left(\frac{1}{\kappa_l} - 1\right)\right\} \left(\mu_g^2 + \sigma_g^2\right). \end{aligned}$$

□

Proof (Lemma 5.3.5)**.** (a) We can determine $\hat{Q}_0^*(\omega)$ using the knowledge that the integral of any Gaussian pdf over $-\infty$ to ∞ is 1.

(b) The quantity $\hat{Q}_1^*(\omega)$ is determined using the property that the mean, μ, of a Gaussian pdf is obtained by evaluating the integral $\int x\varphi_\sigma(x - \mu)\mathrm{d}x$ for $x \in (-\infty, \infty)$.

(c) Using the second moment of the Gaussian pdf from the previous proof, $\hat{Q}_2^*(\omega)$ can be shown to equal

$$\begin{aligned}\hat{Q}_2^*(\omega) &= (\gamma^*)^{\mathrm{i}2\pi\omega}\,\frac{e^{-2(\pi\omega/\nu_l)^2}}{\sqrt{\pi}}\,\Gamma\!\left(\frac{1}{2}-\mathrm{i}2\pi\omega\right)\int_{-\infty}^{\infty} y^2\,\nu_l\,\varphi\!\left(\nu_l\left[y+\frac{\mathrm{i}2\pi\omega}{\nu_l^2}\right]\right)\,\mathrm{d}y \\ &= (\gamma^*)^{\mathrm{i}2\pi\omega}\,\frac{e^{-2(\pi\omega/\nu_l)^2}}{\sqrt{\pi}}\,\Gamma\!\left(\frac{1}{2}-\mathrm{i}2\pi\omega\right)\left(\frac{4\pi^2\omega^2}{\nu_l^4}+\frac{1}{\nu_l^2}\right).\end{aligned}$$

□

A.1.2 Laplace Mixture Prior

Proof (Lemma 5.3.6)**.** Let $\xi(\cdot)$ in the Berger-Müller prior equal the Laplace pdf in (5.26) with $a = 0$ and $b = 1$, then we find equation (5.14) is

$$\begin{aligned}Q_i(h) = {}& \nu_l\int_{-\infty}^{\infty} y^i\,\frac{e^{-|\nu_l y|}}{2}\,\frac{1}{\sigma\sqrt{2\pi}}\,\exp\!\left\{-\frac{(h-y)^2}{2\sigma^2}\right\}\,\mathrm{d}y \\ = {}& \frac{\nu_l}{2}\int_{-\infty}^{0}\frac{y^i}{\sigma\sqrt{2\pi}}\,\exp\!\left\{-\frac{1}{2\sigma^2}\left(h^2-2hy+y^2-2\nu_l\sigma^2 y\right)\right\}\,\mathrm{d}y \qquad (A.5)\\ &+\frac{\nu_l}{2}\int_{0}^{\infty}\frac{y^i}{\sigma\sqrt{2\pi}}\,\exp\!\left\{-\frac{1}{2\sigma^2}\left(h^2-2hy+y^2+2\nu_l\sigma^2 y\right)\right\}\,\mathrm{d}y, \qquad (A.6)\end{aligned}$$

where $\sigma^2 = {}^{\pi^2}\!/_2$. First, consider the exponential part of equation (A.5), which can be written as

$$-\frac{1}{2\,\sigma^2}\left(y^2-2y\left[h+\nu_l\,\sigma^2\right]+h^2\right).$$

Let $\mu_1 = h + \nu_l\sigma^2$, then the previous equation can be written as

$$-\frac{1}{2\,\sigma^2}(y-\mu_1)^2+\frac{1}{2\,\sigma^2}(\mu_1^2-h^2). \qquad (A.7)$$

Similarly, the exponential part in equation (A.6) can also be expressed as

$$-\frac{1}{2\,\sigma^2}(y-\mu_2)^2+\frac{1}{2\,\sigma^2}(\mu_2^2-h^2), \qquad (A.8)$$

where $\mu_2 = h - \nu_l \sigma^2$. Therefore using equations (A.7) and (A.7), $Q_i(h)$ can be represented as

$$\begin{aligned} Q_i(h) = \; & \frac{\nu_l}{2} \; \exp\left\{-\frac{h^2}{2\,\sigma^2}\right\} \times \\ & \left(\exp\left\{\frac{\mu_1^2}{2\,\sigma^2}\right\} \int_{-\infty}^{0} \frac{y^i}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\,\sigma^2}(y-\mu_1)^2\right\} \, \mathrm{d}y \right. \\ & \left. + \; \exp\left\{\frac{\mu_2^2}{2\,\sigma^2}\right\} \int_{0}^{\infty} \frac{y^i}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\,\sigma^2}(y-\mu_2)^2\right\} \, \mathrm{d}y\right) \end{aligned}$$

As the integrals have the same form of the Gaussian pdf and by a change of variable, $Q_i(h)$ can be written in terms of $\varphi_\sigma(\cdot)$ as

$$Q_i(h) = \; \frac{\nu_l}{2} \; e^{-h^2/2\sigma^2} \left[e^{\mu_1^2/2\sigma^2} \int_{-\infty}^{0} y^i \, \varphi_\sigma(y-\mu_1) \; \mathrm{d}y + e^{\mu_2^2/2\sigma^2} \int_{-\infty}^{0} (-y)^i \, \varphi_\sigma(y+\mu_2) \; \mathrm{d}y\right].$$

□

Proof (Lemma 5.3.7)**.** Quantity $Q_i^*(h)$, from equation (5.15), with $\xi(\cdot)$ equivalent to the Laplace pdf is

$$\begin{aligned} Q_i^*(h) = \; & \nu_l \; \int_{-\infty}^{\infty} y^i \, \frac{1}{2} \, e^{-|\nu_l y|} \left(\frac{\gamma^*}{\pi}\right)^{1/2} \exp\left\{\frac{1}{2}(h-y) - \gamma^* e^{(h-y)}\right\} \\ = \; & \frac{\nu_l}{2} \left(\frac{T}{\pi}\right)^{1/2} \left[\int_{-\infty}^{0} y^i \left(\gamma^* e^{(h-y)}\right)^{1/2} \exp\left\{-\gamma^* e^{(h-y)}\right\} e^{\nu_l y} \mathrm{d}y \right. \\ & \left. + \int_{0}^{\infty} y^i \left(\gamma^* e^{(h-y)}\right)^{1/2} \exp\left\{-\gamma^* e^{(h-y)}\right\} e^{-\nu_l y} \, \mathrm{d}y\right]. \end{aligned}$$

By making the substitution of

$$x = \gamma^* e^{(h-y)} \qquad \Rightarrow \qquad y = h - [\log(x) - \log(\gamma^*)] \quad \text{and} \quad \mathrm{d}y = -\frac{\mathrm{d}x}{x}$$

in the previous equation, and if we let $\kappa_l = (\gamma^*)^{\nu_l} e^{\nu_l h}$, we obtain

$$\begin{aligned} Q_i^*(h) = \; & \frac{\nu_l}{2\sqrt{\pi}} \left[\kappa_l \; \int_{\gamma^* e^h}^{\infty} (h - [\log(x) - \log(\gamma^*)])^i \, x^{-1/2-\nu_l} e^{-x} \mathrm{d}x \right. \\ & \left. + \kappa_l^{-1} \int_{0}^{\gamma^* e^h} (h - [\log(x) - \log(\gamma^*)])^i \, x^{-1/2+\nu_l} e^{-x} \mathrm{d}x\right]. \end{aligned}$$

□

Proof (Lemma 5.3.8)**.** (a) Let $i = 0$ in lemma 5.3.6, and using properties of the Gaussian distribution we can determine $Q_0(h)$ for the Laplace mixture prior.

(b) To evaluate $Q_1(h)$, we require the use of integration by parts. Consider

$$\int_{-\infty}^{y} x\,\varphi(x)\,\mathrm{d}x,$$

where $\varphi(\cdot)$ is the standard Gaussian distribution. Let $u = x$ and $\mathrm{d}v = \varphi(\cdot)$, then $\mathrm{d}u = 1$ and $v = \Phi(x)$, where $\Phi(x)$ is the Gaussian cumulative distribution function. Therefore,

$$\int_{-\infty}^{y} x\,\varphi(x)\,\mathrm{d}x = \left[x\Phi(x) - \int \Phi(x)\,\mathrm{d}x\right]_{x=-\infty}^{y}$$

Using the mathematical package Mathematica it can be found that

$$\int \Phi(x)\ \mathrm{d}x \ = \ x\,\Phi(x) + \varphi(x), \tag{A.9}$$

implying

$$\int_{-\infty}^{y} x\,\varphi(x)\,\mathrm{d}x = -\varphi(y) \tag{A.10}$$

Therefore, if we have

$$\int_{-\infty}^{y} x\,\varphi\left(\frac{x-\mu}{\sigma}\right)\,\mathrm{d}x,$$

let $u = \sigma^{-1}(x-\mu) \ \Rightarrow \ x = \sigma u + \mu$ and $\mathrm{d}x = \sigma$, giving

$$\begin{aligned}\int_{-\infty}^{y} x\,\varphi\left(\frac{x-\mu}{\sigma}\right)\,\mathrm{d}x \ &= \ \int_{-\infty}^{\sigma^{-1}(y-\mu)} (\sigma u + \mu)\,\varphi(u)\,\mathrm{d}u, \\ &= \mu\,\Phi\left(\frac{y-\mu}{\sigma}\right) - \sigma\,\varphi\left(\frac{y-\mu}{\sigma}\right).\end{aligned} \tag{A.11}$$

Similarly, if we let $\mu = -\mu$ in (A.11) we find

$$\int_{-\infty}^{y} x\,\varphi\left(\frac{x+\mu}{\sigma}\right)\,\mathrm{d}x \ = \ -\mu\,\Phi\left(\frac{y+\mu}{\sigma}\right) - \sigma\,\varphi\left(\frac{y+\mu}{\sigma}\right). \tag{A.12}$$

This implies

$$\int_{-\infty}^{0} y\,\varphi\left(\frac{y-\mu_1}{\sigma}\right)\mathrm{d}y = \ \mu_1\,\Phi\left(-\frac{\mu_1}{\sigma}\right) + \sigma\,\varphi\left(-\frac{\mu_1}{\sigma}\right), \tag{A.13}$$

and

$$\int_{-\infty}^{0} y\varphi\left(\frac{y+\mu_2}{\sigma}\right)\mathrm{d}y = \ -\mu_2\,\Phi\left(\frac{\mu_2}{\sigma}\right) - \sigma\,\varphi\left(\frac{\mu_2}{\sigma}\right). \tag{A.14}$$

Therefore, we find

$$\begin{aligned}Q_1(h) =& \frac{\nu_l}{2}\, e^{-h^2/2\sigma^2}\left[e^{\mu_1^2/2\sigma^2}\,\mu_1\,\Phi\Big(-\frac{\mu_1}{\sigma}\Big) - \sigma\, e^{\mu_1^2/2\sigma^2}\,\varphi\Big(-\frac{\mu_1}{\sigma}\Big)\right.\\ &\left.+e^{\mu_2^2/2\sigma^2}\,\mu_2\,\Phi\Big(\frac{\mu_2}{\sigma}\Big) + \sigma\, e^{\mu_2^2/2\sigma^2}\,\varphi\Big(\frac{\mu_2}{\sigma}\Big)\right]\\ =& \frac{\nu_l}{2}\, e^{-h^2/2\sigma^2}\left[e^{\mu_1^2/2\sigma^2}\,\mu_1\,\Phi\Big(-\frac{\mu_1}{\sigma}\Big) + e^{\mu_2^2/2\sigma^2}\,\mu_2\,\Phi\Big(\frac{\mu_2}{\sigma}\Big)\right].\end{aligned}$$

(c) From Mathematica we found

$$\int_{-\infty}^{y} x^2\,\varphi(x)\,\mathrm{d}x = \Phi(y) - y\varphi(y), \tag{A.15}$$

using a substitution of variable we found

$$\begin{aligned}&\int_{-\infty}^{0} \frac{x^2}{\sigma}\,\varphi\Big(\frac{x-\mu}{\sigma}\Big)\,\mathrm{d}x\\ &\quad= \int_{-\infty}^{-\mu\sigma^{-1}} (\sigma y+\mu)^2\,\varphi(y)\,\mathrm{d}y,\\ &\quad= \sigma^2\int_{-\infty}^{-\mu\sigma^{-1}} y^2\,\varphi(y)\,\mathrm{d}y + 2\sigma\mu\int_{-\infty}^{-\mu\sigma^{-1}} y\,\varphi(y)\,\mathrm{d}y + \mu^2\int_{-\infty}^{-\mu\sigma^{-1}} \varphi(y)\,\mathrm{d}y,\\ &\quad= \sigma^2\left[\Phi\Big(-\frac{\mu}{\sigma}\Big) + \frac{\mu}{\sigma}\,\varphi\Big(-\frac{\mu}{\sigma}\Big)\right] - 2\sigma\mu\,\varphi\Big(-\frac{\mu}{\sigma}\Big) + \mu^2\,\Phi\Big(-\frac{\mu}{\sigma}\Big),\\ &\quad= [\sigma^2+\mu^2]\Phi\Big(-\frac{\mu}{\sigma}\Big) - \mu\sigma\,\varphi\Big(-\frac{\mu}{\sigma}\Big).\end{aligned} \tag{A.16}$$

Similarly, let $\mu = -\mu$ to find

$$\int_{-\infty}^{0} \frac{x^2}{\sigma}\,\varphi\Big(\frac{x+\mu}{\sigma}\Big)\,\mathrm{d}x = [\sigma^2+\mu^2]\Phi\Big(\frac{\mu}{\sigma}\Big) + \mu\sigma\,\varphi\Big(\frac{\mu}{\sigma}\Big). \tag{A.17}$$

Using equation (A.16) and (A.17), we calculated $Q_2(h)$ as

$$\begin{aligned}&Q_2(h)\\ &= \frac{\nu_l}{2}e^{-h^2/2\sigma^2}\left[e^{\mu_1^2/2\sigma^2}[\sigma^2+\mu_1^2]\Phi\Big(-\frac{\mu_1}{\sigma}\Big) + e^{\mu_2^2/2\sigma^2}[\sigma^2+\mu_2^2]\Phi\Big(\frac{\mu_2}{\sigma}\Big) + \frac{3\sigma}{\sqrt{2\pi}}(\mu_2-\mu_1)\right],\end{aligned}$$

where $\sigma^2 = {}^{\pi^2}\!/_2$, $\mu_1 = h + \nu_l\sigma^2$ and $\mu_2 = h - \nu_l\sigma^2$.

□

Proof (Lemma 5.3.9)**.** (a) To obtain $Q_0^*(h)$ we simply replace i by 0 in lemma 5.3.7 and rearranged the equations to the final format.

(b) To obtain $Q_1^*(h)$ replace i by 1 in lemma 5.3.7 and rearranged the equations to the

final format.

(c) To calculate $Q_2^*(h)$ set $i = 2$ in lemma 5.3.7

$$\begin{aligned}
Q_2^*(h) =& \frac{\nu_l}{2}\left(\frac{1}{\pi}\right)^{1/2}\left[\kappa_l \int_{\gamma^* e^h}^{\infty} \left(h - [\log(x) - \log(\gamma^*)]\right)^2 x^{-1/2-\nu_l/} e^{-x} \mathrm{d}y \right. \\
& \left. + \kappa_l^{-1} \int_0^{\gamma^* e^h} \left(h - [\log(x) - \log(\gamma^*)]\right)^2 x^{-1/2+\nu_l/} e^{-x} \mathrm{d}y \right] \\
=& \left(h + \log(\gamma^*)\right)^2 Q_0^*(h) - 2\left(h + \log(\gamma^*)\right) Q_1^*(h) \\
& + \frac{\nu_l}{2\sqrt{\pi}}\left[\kappa_l \int_{\gamma^* e^h}^{0} \log(x)^2 x^{-1/2-\nu_l} e^{-x}\, \mathrm{d}x \right. \\
& \left. + \kappa_l^{-1} \int_{\gamma^* e^h}^{\infty} \log(x)^2 x^{-1/2+\nu_l} e^{-x}\, \mathrm{d}y \right].
\end{aligned}$$

□

Proof (Lemma 5.3.10)**.** To calculate (5.22) for the Laplace mixture prior, consider this integral in two parts. Firstly

$$\int_{-\infty}^{\infty} f_L(x|\nu_l)\ \frac{1}{\sigma}\,\varphi_\sigma(y-x)\mathrm{d}x \tag{A.18}$$

and secondly

$$\int_{-\infty}^{\infty} f_L(x|\nu_l)\ f_E(y-x)\mathrm{d}x. \tag{A.19}$$

Both integrals will be separated into the negative and positive values of x.

Firstly, lets consider equation (A.18), which is essentially the same as lemma 5.3.6, therefore

$$\begin{aligned}
&\int_{-\infty}^{\infty} f_L(x|\nu_l)\ \frac{}{\sigma}\,\varphi_\sigma(y-x)\mathrm{d}x \\
&\qquad = \frac{\nu_l}{2} e^{-y^2/2\sigma^2}\left[e^{\mu_3^2/2\sigma^2}\Phi\left(\frac{-\mu_3}{\sigma}\right) + e^{\mu_3^2/2\sigma^2}\Phi\left(\frac{\mu_4}{\sigma}\right)\right],
\end{aligned} \tag{A.20}$$

where $\sigma^2 = {}^{\pi^2}\!/_2$, $\mu_3 = y + \nu_l\sigma^2$ and $\mu_4 = y - \nu_l\sigma^2$.

Similarly, by making the same substitution as equation (A.19) is

$$\begin{aligned}
&\int_{-\infty}^{\infty} f_L(x|\nu_l)\ f_E(y-x)\mathrm{d}x \\
&= \frac{\nu_l}{2\sqrt{\pi}}\left(\kappa_l \int_{\gamma^* e^y}^{\infty} u^{-1/2-\nu_l/} e^{-u}\mathrm{d}u + \frac{1}{\kappa_l^*}\int_0^{\gamma^* e^y} u^{-1/2+\nu_l/} e^{-u}\mathrm{d}u\right) \\
&= \frac{\nu_l}{2\sqrt{\pi}}\left[\kappa_l\,\Gamma\left(\frac{1}{2} - \frac{\nu_l}{}, \gamma^* e^y\right) + \frac{1}{\kappa_l}\,\gamma\left(\frac{1}{2} + \frac{\nu_l}{}, \gamma^* e^y\right)\right]
\end{aligned} \tag{A.21}$$

where $\kappa_l^* = e^{\nu_l y}(\gamma^*)^{\nu_l/}$. □

A.1.3 Asymptotic Convergence

To prove Theorem 5.4.1, we require lemmas A.1.1-A.1.5, which describe certain asymptotic properties of our estimator. For simplicity in these lemmas, let

$$\omega = r + \frac{1}{2} - \frac{1}{p^*}, \tag{A.22}$$

therefore we can write $\mu_2 = \omega\left(1 + {}^1\!/_r\right) - {}^1\!/_2$, and if $1 \leq p < 2$, then (5.32) can be expressed as

$$\sum_{m=0}^{2^{(J-l)-1}} b_{l,m}^2 = B_1 2^{-2\omega l} = \mathcal{O}(2^{-2\omega l}).$$

Also, let $\mathbb{P}$ denote the probability, $\mathbb{I}$ be the indicator function and $\varphi(x)$ be the Gaussian pdf with mean zero and variance σ^2.

Lemma A.1.1.

(i) If $|\upsilon_l h^*|$ is bounded or $\upsilon_l|\upsilon_l h^*|^\lambda/\sqrt{T} \to 0$, then as $\upsilon_l/\sqrt{T} \to 0$,

a. $\mathcal{Q}_0(h^*) \;=\; \upsilon_l \xi(\upsilon_l h^*)\left[1 + \mathcal{O}\big(\upsilon_l^2 |\upsilon_l h^*|^{2\kappa}/T\big)\right]$,

b. $\mathcal{Q}_1(h^*)/\mathcal{Q}_0(h^*) \;=\; h^* + \mathcal{O}(\upsilon_l|\upsilon_l h^*|^\kappa/T)$,

for $\kappa \geq 0$.

(ii) If $\sqrt{T}|h^*|$ is bounded or $T|h^*|\upsilon_l^{-1} \to 0$, then as $\sqrt{T}\upsilon_l^{-1} \to 0$,

a. $\mathcal{Q}_0(h^*) \;\sim\; \sqrt{T}\,\varphi(\sqrt{T}h^*)\left\{1 + \mathcal{O}\Big(\big[Th^*/\upsilon_l\big]^2\Big)\right\}$,

b. $\mathcal{Q}_1(h^*)/\mathcal{Q}_0(h^*) \;\sim\; \mathcal{O}\big(Th^*/\upsilon_l^2\big)$.

Proof (Lemma A.1.1). (i) a. Take equation (5.28) and substitute in $x = \sqrt{T}\,[h^* - y]$, therefore $y = h^* - x/\sqrt{T}$ and $\mathrm{d}y = -1/\sqrt{T}\mathrm{d}x$,

$$\mathcal{Q}_i(h^*) \;=\; \upsilon_l \int_{-\infty}^{\infty} \left[h^* - \frac{x}{\sqrt{T}}\right]^i \xi\left(\upsilon_l\left[h^* - \frac{x}{\sqrt{T}}\right]\right) \varphi_\sigma(x)\,\mathrm{d}x,$$

where $\varphi_\sigma(x)$ is the Gaussian pdf with mean zero and variance $\sigma^2 = {}^{\pi^2}\!/_2$. Now

perform a Taylor expansion upon $\xi(\cdot)$ about x in the previous equation,

$$\begin{aligned}&\mathcal{Q}_i(h^*) = \\ &\upsilon_l \int_{-\infty}^{\infty} \left[h^* - \frac{x}{\sqrt{T}}\right]^i \varphi_\sigma(x) \\ &\times \xi(\upsilon_l h^*) \left\{1 - x\frac{\upsilon_l}{\sqrt{T}}\frac{\xi'(\upsilon_l h^*)}{\xi(\upsilon_l h^*)} + x^2\frac{\upsilon_l^2}{2T}\frac{\xi''(\upsilon_l h^*)}{\xi(\upsilon_l h^*)} - x^3\frac{\upsilon_l^3}{6T^{3/2}}\frac{\xi'''(\upsilon_l h^*)}{\xi(\upsilon_l h^*)} + \ldots\right\} \mathrm{d}x\end{aligned} \tag{A.23}$$

The moments of a Gaussian distribution, $X \sim \mathcal{N}(\mu, \sigma^2)$ are defined as

$$\mathbb{E}_N\left[(X-\mu)^i\right] = \int_{-\infty}^{\infty} x^i \varphi_\sigma(x-\mu)\mathrm{d}x = \begin{cases} 0 & \text{if } i \text{ is odd} \\ \sigma^i(i-1)!! & \text{if } i \text{ is even} \end{cases}, \tag{A.24}$$

where !! is the double factorial Isserlis (1918). If $\upsilon_l/\sqrt{T} \to 0$, using condition (5.34) and equation (A.24), we find (A.23) is

$$\mathcal{Q}_0(h^*) = \upsilon_l\xi(\upsilon_l h^*) + \frac{\upsilon_l^2\sigma^2}{2T}\xi''(\upsilon_l h^*) + \ldots = \upsilon_l\xi(\upsilon_l h^*)\left[1 + \mathcal{O}\left(\frac{\upsilon_l^2}{T}|\upsilon_l h^*|^{2\kappa}\right)\right]. \tag{A.25}$$

b. For $i = 1$, by splitting (A.23) in two, we find that

$$\begin{aligned}\mathcal{Q}_1(h^*) &= \upsilon_l h^*\left[\xi(\upsilon_l h^*) + \frac{\upsilon_l^2\sigma^2}{2T}\xi''(\upsilon_l h^*)\right] + \\ &\quad \frac{\upsilon_l}{\sqrt{T}}\left[\frac{\sigma^2\upsilon_l}{\sqrt{T}}\xi'(\upsilon_l h^*) + \frac{\sigma^4\upsilon_l^3}{6T^{3/2}}\xi'''(\upsilon_l h^*)\right] + \ldots \\ &= \upsilon_l h^*\xi(\upsilon_l h^*) + \frac{\sigma^2\upsilon_l^2}{T}\xi'(\upsilon_l h^*) + \frac{\upsilon_l^3\sigma^2 h^*}{2T}\xi''(\upsilon_l h^*) + \ldots\end{aligned} \tag{A.26}$$

Therefore, from the condition from equation (5.34), we find that

$$\begin{aligned}\frac{\mathcal{Q}_0(h^*)}{\mathcal{Q}_1(h^*)} &= h^* + \frac{\sigma^2\upsilon_l}{T}\frac{\xi'(\upsilon_l h^*)}{\xi(\upsilon_l h^*)}\left[1 + \mathcal{O}\left(\frac{\upsilon_l^2}{T}|\upsilon_l h^*|^{2\kappa}\right)\right] \\ &= h^* + \mathcal{O}\left(\frac{\upsilon_l}{T}|\upsilon_l h^*|^{\kappa}\right).\end{aligned} \tag{A.27}$$

(ii) a. For ease of notation, let $\varphi(x) = \varphi_\sigma(x)$ for $\sigma^2 = \pi^2/2$. To calculate the distribution of $\mathcal{Q}_i(h^*)$ for $i = 0$ or 1, firstly let $x = \upsilon_l y$ in equation (5.28), which implies

$y = x/\upsilon_l$ and $\mathrm{d}y = \upsilon_l^{-1}\mathrm{d}x$. Then Taylor expand $\varphi(\cdot)$ instead of $\xi(\cdot)$. Therefore,

$$\begin{aligned}\mathcal{Q}_i(h^*) &\sim \sqrt{T}\int_{-\infty}^{\infty}(x/\upsilon_l)^i\,\xi(x)\,\varphi\Big(\sqrt{T}\,[h - x/\upsilon_l]\Big)\,\mathrm{d}x \\ &= \sqrt{T}\int_{-\infty}^{\infty}(x/\upsilon_l)^i\,\xi(x)\left\{\varphi(\sqrt{T}h^*) - x\frac{\sqrt{T}}{\upsilon_l}\varphi'(\sqrt{T}h^*)\right. \\ &\qquad\left. + x^2\frac{T}{2\upsilon_l^2}\varphi''(\sqrt{T}h^*) - \; x^3\frac{T^{3/2}}{6\upsilon_l^3}\varphi'''(\sqrt{T}h^*) + \ldots\right\}\mathrm{d}x\end{aligned} \tag{A.28}$$

As $\varphi(x)$ is a probability density function we know that the integral over $x \in \mathbb{R}$ equals 1, also from equation (A.24), we can determine the moments of a zero mean Gaussian pdf. The zero mean Gaussian pdf with variance σ^2 and the first three derivatives are below

$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{x^2}{2\sigma^2}\right\}, \tag{A.29}$$

$$\varphi'(x) = \frac{-x}{\sigma^3\sqrt{2\pi}}\exp\left\{-\frac{x^2}{2\sigma^2}\right\} = -\frac{x}{\sigma^2}\,\varphi(x), \tag{A.30}$$

$$\varphi''(x) = -\frac{x}{\sigma^2}\left[-\frac{x}{\sigma^2}\,\varphi(x)\right] - \varphi(x)\frac{1}{\sigma^2} = \left[\frac{x^2}{\sigma^4} - \frac{1}{\sigma^2}\right]\varphi(x) \tag{A.31}$$

$$\varphi'''(x) = \left[\frac{x^2}{\sigma^4} - \frac{1}{\sigma^2}\right]\left[-\frac{x}{\sigma^2}\,\varphi(x)\right] + \varphi(x)\frac{2x}{\sigma^4} = \left[\frac{3x}{\sigma^4} - \frac{x^3}{\sigma^6}\right]\varphi(x) \tag{A.32}$$

for $x \in \mathbb{R}$. Let $\xi(x)$ be a unimodal, symmetric, zero mean pdf for $x \in [-a, a]$, then for odd n

$$\begin{aligned}\int_{-a}^{a} x^n\xi(x)\mathrm{d}x &= \int_0^a x^n\xi(x)\mathrm{d}x - \int_0^{-a} x^n\xi(x)\mathrm{d}x \\ &= \int_0^a x^n\xi(x)\mathrm{d}x + \int_0^a (-x)^n\xi(-x)\mathrm{d}x.\end{aligned}$$

As n is odd this implies that $(-x)^n = -x^n$, and due to the properties of the pdf $\xi(x) = \xi(-x)$, therefore

$$\int_{-a}^{a} x^n\xi(x)\mathrm{d}x = \int_0^a x^n\xi(x)\mathrm{d}x - \int_0^a x^n\xi(x)\mathrm{d}x = 0.$$

Then, if we set $i = 0$, and by substituting (A.29) and (A.31) into (A.28), we find

that

$$\begin{aligned}\mathcal{Q}_0(h^*) &\sim \sqrt{T}\varphi(\sqrt{T}h^*) + \frac{T^{3/2}}{2\upsilon_l^2}\,\varphi''(\sqrt{T}h^*)\,\mathcal{M}_2 + \dots \\ &= \sqrt{T}\,\varphi(\sqrt{T}h^*) + \frac{T^{3/2}}{2\upsilon_l^2}\left[\frac{Th^2}{\sigma^4} - \frac{1}{\sigma^2}\right]\varphi(\sqrt{T}h^*)\,\mathcal{M}_2 + \dots \\ &= \sqrt{T}\,\varphi(\sqrt{T}h^*)\left\{1 + \mathcal{O}\left(\left[\frac{Th^*}{\upsilon_l}\right]^2\right)\right\}, \end{aligned} \tag{A.33}$$

where $\mathcal{M}_i$, for i is an even integer, denote the even moments of $\xi(x)$.

b. For $i = 1$

$$\begin{aligned}\mathcal{Q}_1(h^*) &\sim \frac{\sqrt{T}}{\upsilon_l}\int x^2\xi(x)\left[-\frac{\sqrt{T}}{\upsilon_l}\varphi'(\sqrt{T}h^*) - x\frac{T^{3/2}}{6\upsilon_l^3}\varphi'''(\sqrt{T}h^*) + \dots\right]\mathrm{d}x \\ &= \frac{T}{\upsilon_l^2}\varphi(\sqrt{T}h^*)\int x^2\xi(x)\left[\frac{\sqrt{T}h^*}{\sigma^2} - x\frac{T}{6\upsilon_l^2}\left[\frac{3\sqrt{T}h^*}{\sigma^4} - \frac{(\sqrt{T}h^*)^3}{\sigma^6}\right] + \dots\right]\mathrm{d}x \\ &= \frac{T^{3/2}h^*}{(\upsilon_l\sigma)^2}\varphi(\sqrt{T}h^*)\int x^2\xi(x)\left\{1 + \mathcal{O}\left(\left[\frac{Th^*}{\upsilon_l}\right]^2\right)\right\}\mathrm{d}x \end{aligned} \tag{A.34}$$

Dividing (A.34) by (A.33), we find

$$\begin{aligned}\frac{\mathcal{Q}_1(h^*)}{\mathcal{Q}_0(h^*)} &\sim \frac{Th^*}{\upsilon_l^2\sigma^2}\int x^2\xi(x)\left\{1 + \mathcal{O}\left(\left[\frac{Th^*}{\upsilon_l}\right]^2\right)\right\}\mathrm{d}x \\ &= \mathcal{O}\left(\frac{Th^*}{\upsilon_l^2}\right). \end{aligned} \tag{A.35}$$

□

Lemma A.1.2. *The distribution of $\mathcal{Q}_i^*(h^*)$ for $i = 0, 1$ is*

(i) $\mathcal{Q}_0^*(h^*) \sim \sqrt{T}f_E(\sqrt{T}h^*)\left\{1 + \mathcal{O}\left(T\upsilon_l^{-2}\left[1 - e^{\sqrt{T}h^*}/2\right]\right)\right\}$,

(ii) $\mathcal{Q}_1^*(h^*) \sim T\upsilon_l^{-2}f_E(\sqrt{T}h^*)(1/2 - \gamma^* e^{\sqrt{T}h^*})$,
$\times \int_{-\infty}^{\infty} x^2\xi(x)dx\left\{1 + \mathcal{O}\left(T\upsilon_l^{-2}\left[1 - e^{\sqrt{T}h^*}\right]\right)\right\}$.

Proof (Lemma A.1.2)**.** First we shall prove the distribution of $\mathcal{Q}_i^*(h^*)$ in part (i), using the definition of $\mathcal{Q}_i^*(h^*)$ from equation (5.29) let $x = \upsilon_l y$ and $\mathrm{d}y = \upsilon_l^{-1}\mathrm{d}x$ and use the

Taylor expansion, similar to the proof of Lemma A.1.3,

$$\begin{aligned}\mathcal{Q}_i^*(h^*) &\sim \sqrt{T}\int_{-\infty}^{\infty}\left(\frac{x}{\upsilon_l}\right)^i \xi(x)\, f_E(\sqrt{T}[h-x/\upsilon_l])\,\mathrm{d}x \\ &= \sqrt{T}\int_{-\infty}^{\infty}\left(\frac{x}{\upsilon_l}\right)^i \xi(x)\left\{f_E(\sqrt{T}h^*) - x\frac{\sqrt{T}}{\upsilon_l}f_E'(\sqrt{T}h^*) + x^2\frac{T}{2\upsilon_l^2}f_E''(\sqrt{T}h^*)\right. \\ &\qquad \left. -x^3\frac{T^{3/2}}{6\upsilon_l^3}f_E'''(\sqrt{T}h^*) + \ldots\right\}\mathrm{d}x.\end{aligned} \tag{A.36}$$

In lemma 5.1.1, equation (5.4), we defined the error distribution which we can use to calculate the first two differences using the chain rule

$$\begin{aligned} f_E(x) &= \sqrt{\frac{\gamma^*}{\pi}}\exp\left\{\frac{x}{2}-\gamma^* e^x\right\}, \\ f_E'(x) &= \sqrt{\frac{\gamma^*}{\pi}}\exp\left\{\frac{x}{2}-\gamma^* e^x\right\}\left[\frac{1}{2}-\gamma^* e^x\right] = \left[\frac{1}{2}-\gamma^* e^x\right] f_E(x) \qquad (A.37) \\ f_E''(x) &= \left[\frac{1}{2}-\gamma^* e^x\right]^2 f_E(x) + \left[-\gamma^* e^x\right] f_E(x) \\ &= \left[\frac{1}{4} - 2\gamma^* e^x + (\gamma^* e^x)^2\right] f_E(x), \qquad (A.38) \\ f_E'''(x) &= \left[\frac{1}{4} - 2\gamma^* e^x + (\gamma^* e^x)^2\right]\left[\frac{1}{2}-\gamma^* e^x\right] f_E(x) + \left[2(\gamma^* e^x)^2 - 2\gamma^* e^x\right] f_E(x) \\ &= \left[\frac{1}{8} - \frac{13}{4}\gamma^* e^x + \frac{9}{2}(\gamma^* e^x)^2 - (\gamma^* e^x)^3\right] f_E(x), \qquad (A.39) \end{aligned}$$

$x \in \mathbb{R}$. Using equation (A.24), (5.4) and (A.37) to (A.38) implies that equation (A.36) for $i = 0$ is

$$\begin{aligned}\mathcal{Q}_0^*(h^*) &\sim \sqrt{T}\int_{-\infty}^{\infty}\xi(x)\left\{f_E(\sqrt{T}h^*) - \; x\frac{\sqrt{T}}{\upsilon_l}f_E'(\sqrt{T}h^*) + \; x^2\frac{T}{2\upsilon_l^2}f_E''(\sqrt{T}h^*)\right. \\ &\qquad \left. -\; x^3\frac{T^{3/2}}{6\upsilon_l^3}f_E'''(\sqrt{T}h^*) + \ldots\right\}\mathrm{d}x \\ &= \sqrt{T}f_E(\sqrt{T}h^*) + \frac{T^{3/2}}{2\upsilon_l^2}f_E''(\sqrt{T}h^*) + \ldots \\ &= \sqrt{T}f_E(\sqrt{T}h^*)\left\{1 + \frac{T}{2\upsilon_l^2}\left[\frac{1}{4} - 2\gamma^* e^{\sqrt{T}h^*} + (\gamma^* e^{\sqrt{T}h^*})^2\right] + \ldots\right\} \\ &= \sqrt{T}f_E(\sqrt{T}h^*)\left\{1 + \mathcal{O}\left(\frac{T}{8\upsilon_l^2}\left[1 - \frac{\gamma^*}{2}e^{\sqrt{T}h^*} + \frac{1}{4}(\gamma^* e^{\sqrt{T}h^*})^2\right]\right)\right\} \end{aligned} \tag{A.40}$$

Using a similar technique the distribution for $\mathcal{Q}_1^*(h^*)$ of part (ii) can be found to be

$$\begin{aligned}
\mathcal{Q}_1^*(h^*) \sim \frac{\sqrt{T}}{\upsilon_l}\int_{-\infty}^{\infty} x\xi(x)\Bigg\{ & f_E(\sqrt{T}h^*) - \; x\frac{\sqrt{T}}{\upsilon_l}f_E'(\sqrt{T}h^*) + \; x^2\frac{T}{2\upsilon_l^2}f_E''(\sqrt{T}h^*) \\
& - x^3\frac{T^{3/2}}{6\upsilon_l^3}f_E'''(\sqrt{T}h^*) + \ldots\Bigg\}\,\mathrm{d}x \\
= \frac{T}{\upsilon_l^2}\int_{-\infty}^{\infty} x^2\xi(x)\Bigg\{ & f_E'(\sqrt{T}h^*) + \; x^2\frac{T}{6\upsilon_l^2}f_E'''(\sqrt{T}h^*) + \ldots\Bigg\}\,\mathrm{d}x \\
= \frac{T}{\upsilon_l^2}f_E(\sqrt{T}h^*)\int_{-\infty}^{\infty} x^2\xi(x)\Bigg\{ & \frac{1}{2} - \gamma^* e^{\sqrt{T}h^*} \\
& + x^2\frac{T}{6\upsilon_l^2}\left[\frac{1}{8} - \frac{13}{4}\gamma^* e^{\sqrt{T}h^*} + \frac{9}{2}(\gamma^* e^{\sqrt{T}h^*})^2 - (\gamma^* e^{x})^3\right] + \ldots\Bigg\}\,\mathrm{d}x \\
= \frac{T}{\upsilon_l^2}f_E(\sqrt{T}h^*)\left(\frac{1}{2} - \gamma^* e^{\sqrt{T}h^*}\right)\int_{-\infty}^{\infty} x^2\xi(x)\mathrm{d}x & \left\{1 + \mathcal{O}\left(\frac{T}{\upsilon_l^2}\left[1 - e^{\sqrt{T}h^*}\right]\right)\right\} \qquad \text{(A.41)}
\end{aligned}$$

□

Lemma A.1.3. *Let the probability density function of $h_{l,m}^*$ given $b_{l,m}$ be of the form $\sqrt{T}\zeta_l(\sqrt{T}[h_{l,m}^* - b_{l,m}])$ where $\zeta_l(\cdot)$ is defined in equation* (5.8) *and J_2 be as defined in* (5.41)*. Then as $T \to \infty$ and for any constants $a > 0$*

(i) $\mathbb{E}\Big[(h_{l,m}^* - b_{l,m})^{2i}\Big] = \mathcal{O}(T^{-i})$, *for* $i = 1, 2$,

(ii) $\mathbb{P}\Big(\sqrt{T}|h_{l,m}^* - b_{l,m}| > a\sqrt{\log(T)}\Big) = o(T^{-a^2/2\sigma^2})$, *for* $l > J_2$,

(iii) $\mathbb{P}\Big(|h_{l,m}^* - b_{l,m}| > a\log(T)\Big) = (1 - \lambda_l)\mathcal{O}(T^{-a}) + \lambda_l\, o\,(T^{-a})$,

(iv) $\mathbb{P}\Big(\sqrt{T}(h_{l,m}^* - b_{l,m}) > a\log(T)\Big) = o(T^{-a})$.

Proof (Lemma A.1.3)**.** The validation of Lemma A.1.3 follows directly from equation (5.8) where

$$\begin{aligned}
\sqrt{T}[h_{l,m}^* - b_{l,m}] \;\sim\; & (1 - \lambda_l)\,\varphi(\sqrt{T}[h_{l,m}^* - b_{l,m}]|\mu = 0, \sigma^2 = \pi^2/2) \\
& + \;\lambda_l\, f_E(\sqrt{T}[h_{l,m}^* - b_{l,m}]).
\end{aligned}$$

(i) From Lemma 5.1.1, the cumulants for the error distribution, $f_E(x)$, are $\psi^{(r-1)}(1/2)$. The second moment is equal to the second cumulant, but the fourth moment can be calculated from the cumulants using the formula $\kappa_4 + 3\kappa_2^2$. Therefore, the fourth moment of the error distribution is

$$\mathbb{E}_E\big[X^4\big] = \psi^{(3)}\big(1/2\big) + 3\psi^{(1)}\big(1/2\big) = \pi^4 + 3\frac{\pi^2}{2} = \frac{\pi^2}{2}(2\pi^2 + 3) \qquad \text{(A.42)}$$

Equation (A.24) defined the moments for a Gaussian distribution, therefore

$$\mathbb{E}_N\big[(X-\mu)^2\big] = \sigma^2 \qquad \text{and} \qquad \mathbb{E}_N\big[(X-\mu)^4\big] = 3\sigma^4.$$

By substituting in the moments for the Gaussian distribution with mean 0 and variance $\pi^2/2$, and the error distribution it can be shown that

$$\begin{aligned}
\mathbb{E}\big[(h^*_{l,m} - b_{l,m})^2\big] &= \frac{(1-\lambda_l)}{T}\,\mathbb{E}_N\Big[\{\sqrt{T}(h^*_{l,m} - b_{l,m})\}^2\Big] + \frac{\lambda_l}{T}\,\mathbb{E}_E\Big[\{\sqrt{T}(h^*_{l,m} - b_{l,m})\}^2\Big] \\
&= \frac{(1-\lambda_l)}{T}\frac{\pi^2}{2} + \frac{\lambda_l}{T}\frac{\pi^2}{2} \\
&= \frac{\pi^2}{2T} \\
&= \mathcal{O}(T^{-1})
\end{aligned} \tag{A.43}$$

and similarly

$$\begin{aligned}
\mathbb{E}\big[(h^*_{l,m} - b_{l,m})^4\big] &= \frac{(1-\lambda_l)}{T^2}\,\mathbb{E}_N\Big[\{\sqrt{T}(h^*_{l,m} - b_{l,m})\}^4\Big] + \frac{\lambda_l}{T^2}\,\mathbb{E}_E\Big[\{\sqrt{T}(h^*_{l,m} - b_{l,m})\}^4\Big] \\
&= \frac{(1-\lambda_l)}{T^2}\frac{3\pi^4}{4} + \frac{\lambda_l}{T^2}\frac{\pi^2}{2}(2\pi^2+3) \\
&= \frac{\pi^2}{2T^2}\left[\frac{3\pi^2}{2} + \lambda_l\left(\frac{\pi^2}{2}+3\right)\right] \\
&= \mathcal{O}(T^{-2})
\end{aligned} \tag{A.44}$$

(ii) Firstly, let $X = \sqrt{T}(h^*_{l,m} - b_{l,m})$. If we are only looking at values of $l > J_2$, this implies that $\lambda_l = 0$, therefore we only need to consider the Gaussian part of the mixture distribution. As the Gaussian distribution is symmetric about zero, we can evaluate the probability bound by using

$$\begin{aligned}
\mathbb{P}\Big(|X| > a\sqrt{\log(T)}\Big) &= 2\mathbb{P}\Big(X < -a\sqrt{\log(T)}\Big) \\
&= \int_{-\infty}^{-a\sqrt{\log(T)}} \varphi(x|\mu=0,\sigma^2=\pi^2/2)\;\mathrm{d}x \\
&= 2\Phi\Big(-\frac{a}{\sigma}\sqrt{\log(T)}\Big)
\end{aligned}$$

The cumulative Gaussian distribution function can be written in terms of the error function using the equation

$$\Phi(z) = \frac{1}{2}\left[1 + \mathrm{Erf}\Big(\frac{z}{\sqrt{2}}\Big)\right], \qquad -\infty \geq z \geq \infty, \tag{A.45}$$

and because the error function is odd, this implies that $\text{Erf}(-x) = -\text{Erf}(x)$, therefore we find that

$$\mathbb{P}\Big(|X| > a\sqrt{\log(T)}\Big) = 1 - \text{Erf}\left(\frac{a}{\sigma}\sqrt{\frac{\log(T)}{2}}\right) \tag{A.46}$$

A Taylor expansion of the error function is given by

$$\text{Erf}(z) = \frac{2}{\sqrt{\pi}}\left(z - \frac{z^3}{3} + \frac{z^5}{10} - \frac{z^7}{42} + \frac{z^9}{216} - \dots\right), \tag{A.47}$$

where the integrand of e^{-z^2} is expanded and each term integrated individually. Or for values much greater or smaller than z a series approximation of the error function can be given by

$$\text{Erf}\,(z) = \begin{cases} 1 - \frac{e^{-z^2}}{z\sqrt{\pi}}\left(1 - \frac{1}{2z^2} + \frac{1\cdot 3}{(2z^2)^2} - \frac{1\cdot 3\cdot 5}{(2z^2)^3} + \dots\right) & \text{for } z \gg 1 \\ \frac{2ze^{-z^2}}{\sqrt{\pi}}\left(1 + \frac{z^2}{3\cdot 1!} + \frac{z^4}{5\cdot 2!} + \frac{z^6}{7\cdot 3!} + \dots\right) & \text{for } z \ll 1 \end{cases}. \tag{A.48}$$

As $\frac{a}{\sigma}\sqrt{\frac{\log(T)}{2}}$ tends to infinity we apply the expansions defined in equation (A.48), for values much greater than 1, to the error function of this quantity. This yields

$$\text{Erf}\left(\frac{a}{\sigma}\sqrt{\frac{\log(T)}{2}}\right) = 1 - T^{-a^2/2\sigma^2}\frac{\sigma}{a}\sqrt{\frac{2}{\pi\log(T)}}\left(1 - \frac{\sigma^2}{a^2\log(T)} + \frac{3\sigma^4}{a^4\log^2(T)} - \dots\right). \tag{A.49}$$

We find that equation (A.49) will tend to one at a rate of $T^{-a^2/2\sigma^2}[\log(T)]^{-1}$ as $T \to \infty$, and because $T^{-a^2/2\sigma^2}[\log(T)]^{-1} \leq T^{-a^2/2\sigma^2}$, this implies that

$$\mathbb{P}\Big(\sqrt{T}|h^*_{l,m} - b_{l,m}| > a\sqrt{\log(T)}\Big) = o\left(T^{-a^2/2\sigma^2}\right) \qquad \text{for } l > J_2.$$

(iii) This probability can be re-written as

$$\mathbb{P}(|h^*_{l,m} - b_{l,m}| > a\log(T)) = \mathbb{P}(\sqrt{T}|h^*_{l,m} - b_{l,m}| > a\sqrt{T}\log(T)).$$

As there is no restrictions on l, we cannot assume that $\lambda_l = 0$. This implies we have to consider both parts of the mixture distribution. As the error distribution is not symmetric we have to consider

- $\mathbb{P}(X > a\sqrt{T}\log(T))$
- $\mathbb{P}(X < -a\sqrt{T}\log(T))$

where $X = \sqrt{T}(h^*_{l,m} - b_{l,m})$. Then we just need to sum these two probabilities together. Starting with the first probability, we find that

$$\begin{aligned}&\mathbb{P}(X > a\sqrt{T}\log(T)) \\ &\quad = (1-\lambda_l)\int_{a\sqrt{T}\log(T)}^{\infty} \varphi(x)\ \mathrm{d}x + \lambda_l \int_{a\sqrt{T}\log(T)}^{\infty} f_E(x)\ \mathrm{d}x\end{aligned}$$

where $\varphi(x)$ is the Gaussian pdf with variance $\pi^2/2$ and $f_E(x)$ is the error pdf. The Gaussian part of this probability can be calculated using the Gaussian cdf, as

$$(1-\lambda_l)\int_{a\sqrt{T}\log(T)}^{\infty} \varphi(x)\ \mathrm{d}x = (1-\lambda_l)\Phi\Big(-\frac{a}{\sigma}\sqrt{T}\log(T)\Big).$$

For the second integral we use (5.6) from lemma 5.1.1, giving

$$\int_{a\sqrt{T}\log(T)}^{\infty} f_E(x)\ \mathrm{d}x \ = \ \frac{1}{\sqrt{\pi}}\Gamma\left(\frac{1}{2}, \gamma^* T^{a\sqrt{T}}\right),$$

where $\Gamma(\cdot,\cdot)$ is the upper incomplete gamma function. For the case where we examine $\mathbb{P}(X < -a\sqrt{T}\log(T))$, the Gaussian part is also solved in a similar manner, and for the error distribution we use (5.5) from lemma 5.1.1, giving

$$\int_{-\infty}^{-a\sqrt{T}\log(T)} f_E(x)\ \mathrm{d}x = \ \frac{1}{\sqrt{\pi}}\gamma\left(\frac{1}{2}, \gamma^* T^{-a\sqrt{T}}\right),$$

where $\gamma(\cdot,\cdot)$ is the lower incomplete gamma function. Therefore, we find that

$$\begin{aligned}\mathbb{P}\Big(|X| > a\sqrt{T}\log(T)\Big) = 2(1-\lambda_l)\,\Phi\left(-\frac{a\sqrt{T}}{\sigma}\log(T)\right) \\ + \frac{\lambda_l}{\sqrt{\pi}}\left[\Gamma\left(\frac{1}{2}, \gamma^* T^{a\sqrt{T}}\right) + \gamma\left(\frac{1}{2}, \gamma^* T^{-a\sqrt{T}}\right)\right].\end{aligned}$$

The upper and lower incomplete gamma function can be written in terms of the error function using the equations

$$\gamma\left(\frac{1}{2}, z\right) = \sqrt{\pi}\,\mathrm{Erf}(\sqrt{z}) \qquad z > 0, \tag{A.50}$$

$$\Gamma\left(\frac{1}{2}, z\right) = \sqrt{\pi}\left[1 - \mathrm{Erf}(\sqrt{z})\right] \qquad z > 0. \tag{A.51}$$

Using equations (A.50), (A.51) and (A.45) we have

$$\begin{aligned}\mathbb{P}\Big(|X| > a\sqrt{T}\log(T)\Big) =& 1 - (1-\lambda_l)\mathrm{Erf}\bigg(\frac{a}{\sigma\sqrt{2T}}\log(T)\bigg) \\ &- \lambda_l\left[\mathrm{Erf}\left(\sqrt{\frac{\gamma^*}{2}}T^{a\sqrt{T}}\right) - \mathrm{Erf}\left(\sqrt{\frac{\gamma^*}{2}}T^{-a\sqrt{T}}\right)\right].\end{aligned}$$

These error functions can be written as series expansions using the representations in equations (A.47) and (A.48),

$$\begin{aligned}\mathrm{Erf}\left(\frac{a\sqrt{T}}{\sigma\sqrt{2}}\log(T)\right) &= 1 - \exp\bigg\{-\frac{a^2T}{2\sigma^2}\log^2(T)\bigg\}\frac{\sigma}{a\log(T)}\sqrt{\frac{2}{\pi T}} \\ &\quad\times\left(1 - \frac{\sigma^2}{a^2T\log^2(T)} + 3\frac{\sigma^4}{a^4T^2\log^4(T)} - \dots\right) \qquad \text{(A.52)} \\ \mathrm{Erf}\left(\sqrt{\frac{\gamma^*}{2}}T^{a\sqrt{T}}\right) &= 1 - T^{-a\sqrt{T}}\exp\left\{-\frac{\gamma^*}{2}T^{2a\sqrt{T}}\right\}\sqrt{\frac{2}{\gamma^*\pi}} \\ &\quad\times\left(1 - \frac{1}{\gamma^*}T^{-2a\sqrt{T}} + \frac{3}{(\gamma^*)^4}T^{-4a\sqrt{T}} - \dots\right) \qquad \text{(A.53)} \\ \mathrm{Erf}\left(\sqrt{\frac{\gamma^*}{2}}T^{-a\sqrt{T}}\right) &= \frac{2}{\sqrt{\pi}}\left(\sqrt{\frac{\gamma^*}{2}}T^{-a\sqrt{T}} - \frac{1}{3}\left(\frac{\gamma^*}{2}\right)^{3/2}T^{-3a\sqrt{T}} + \dots\right) \qquad \text{(A.54)}\end{aligned}$$

Equation (A.52) will tend to 1 at a rate of

$$\begin{aligned}\exp\bigg\{-\frac{a^2T}{2\sigma^2}\log^2(T)\bigg\}\frac{1}{\sqrt{T}\log(T)} &= T^{-a^2T\log(T)(2\sigma^2)^{-1}}\frac{1}{\sqrt{T}\log(T)} \\ &\leq T^{-a}\end{aligned}$$

While equation (A.53) will tend to one at a rate of

$$\begin{aligned}T^{-a\sqrt{T}}\exp\left\{-\frac{\gamma^*}{2}T^{2a\sqrt{T}}\right\} &\leq T^{-a\sqrt{T}} \\ &\leq T^{-a}\end{aligned}$$

and (A.54) will tend to zero at a rate of $T^{-a/\sqrt{(T)}} \leq T^{-a}$. Therefore, we find that

$$\mathbb{P}\big(|h^*_{l,m} - b_{l,m}| > a\log(T)\big) = (1-\lambda_l)\mathcal{O}(T^{-a}) + \lambda_l\, o(T^{-a}).$$

(iv) Using equations (5.6) and (A.46) from part (ii), the probability that $X > a\log(T)$ is

$$\begin{aligned}\mathbb{P}(X > a\log(T)) =&(1-\lambda_l)\int_{a\log(T)}^{\infty}(1-\lambda_l)\ \varphi(x|\mu=0,\sigma^2=\pi^2/2)\ \mathrm{d}x\\ &+\lambda_l\int_{a\log(T)}^{\infty}\lambda_l\ f_E(x)\ \mathrm{d}x\\ =&(1-\lambda_l)\Phi\Big(-\frac{a}{\sigma}\log(T)\Big)-\frac{\lambda_l}{\sqrt{\pi}}\gamma\left(\frac{1}{2},\frac{\gamma^*}{2}T^{2a}\right)\\ =&\frac{(1-\lambda_l)}{2}\left[1-\mathrm{Erf}\left(\frac{a}{\sigma\sqrt{2}}\log(T)\right)\right]-\lambda_l\,\mathrm{Erf}\left(\sqrt{\frac{\gamma^*}{2}}T^a\right)\end{aligned}$$

Then writing the error functions in terms of equation (A.48) yields

$$\begin{aligned}\mathrm{Erf}\left(\frac{a}{\sigma\sqrt{2}}\log(T)\right) =&\ 1-\exp\left\{-\frac{a^2}{2\sigma^2}\log^2(T)\right\}\frac{\sigma}{a\log(T)}\sqrt{\frac{2}{\pi}}\\ &\times\left(1-\frac{\sigma^2}{a^2\log^2(T)}+3\frac{\sigma^4}{a^4\log^4(T)}-\dots\right) \qquad (A.55)\\ \mathrm{Erf}\left(\sqrt{\frac{\gamma^*}{2}}T^a\right) =&\ 1-\exp\left\{-\frac{\gamma^*}{2}T^{2a}\ \right\}\sqrt{\frac{2}{\gamma^*\pi}}T^{-a}\left(1-\frac{T^{-2a}}{\gamma^*}+\frac{3T^{-4a}}{(\gamma^*)^2}-\dots\right)\end{aligned}$$

(A.56)

Both of these functions will tend to 1, and the speed of the decay will both be less than T^{-a}. So we can determine that

$$\mathbb{P}(X > a\log(T)) = o(T^{-a}).$$

□

Lemma A.1.4. *If $\xi(x)$ is a unimodal probability density function, then if $\sqrt{T}|h^*| \to \infty$ and $v_l/\sqrt{T} \to \infty$*

(i) $\left|\frac{\mathcal{Q}_1(h^*)}{\mathcal{Q}_0(h^*)}\right| = \mathcal{O}(|h^*|)$,

(ii) $\left|\frac{\mathcal{Q}_1^*(h^*)}{\mathcal{Q}_0^*(h^*)}\right| = \mathcal{O}(|h^*|)$.

Proof (Lemma A.1.4)**.** Assume that we have $h^* > 0$.

(i) To prove that $\left|\frac{\mathcal{Q}_1(h^*)}{\mathcal{Q}_0(h^*)}\right|$ is $\mathcal{O}(|h^*|)$ we use the fact that $\mathcal{Q}_0(h^*)$ is an even function of h^* and $\mathcal{Q}_1(h^*)$ is an odd function. We firstly partition the integral for $\mathcal{Q}_1(h^*)$ into three parts: $(-\infty, -h^*/2)$, $(-h^*/2, 3h^*/2)$ and $(3h^*/2, \infty)$ and let these three integrals be defined as $\mathcal{Q}_{1,1}(h^*)$, $\mathcal{Q}_{1,2}(h^*)$ and $\mathcal{Q}_{1,3}(h^*)$, respectively. Firstly, let us address the

first partition,

$$|\mathcal{Q}_{1,1}(h^*)| \leq \int_{-\infty}^{-h^*/2} \frac{|y|\sqrt{T}}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{T}{2\sigma^2}[h^* - y]^2\right\} \upsilon_l\, \xi(\upsilon_l y)\, \mathrm{d}y.$$

Let $y = x - h^*/2$ and consider $\xi(\upsilon_l[x - h^*/2]) \leq \xi(\upsilon_l x)$ for $x < 0$, therefore

$$|\mathcal{Q}_{1,1}(h^*)| \leq \int_{-\infty}^{0} \left|x - \frac{h^*}{2}\right| \frac{\sqrt{T}}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{T}{2\sigma^2}\left(\frac{3h^*}{2} - x\right)^2\right\} \upsilon_l\, \xi(\upsilon_l x)\, \mathrm{d}x$$

Consider that

$$\left(\frac{3h^*}{2} - x\right)^2 - (h^* - x)^2 = \frac{5h^{*2}}{4} - h^* x,$$

then

$$\begin{aligned}
&|\mathcal{Q}_{1,1}(h^*)| \\
&\leq \int_{-\infty}^{0} \left|x - \frac{h^*}{2}\right| \frac{\sqrt{T}}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{T}{2\sigma^2}\left(\frac{5h^{*2}}{4} - h^* x\right) - \frac{T}{2\sigma^2}(h^* - x)^2\right\} \upsilon_l\, \xi(\upsilon_l x)\, \mathrm{d}x \\
&= \exp\left\{-\frac{5h^{*2}T}{8\sigma^2}\right\} \int_{-\infty}^{0} \left|x - \frac{h^*}{2}\right| \frac{\sqrt{T}}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{Th^* x}{2\sigma^2} - \frac{T(h^* - x)^2}{2\sigma^2}\right\} \upsilon_l\, \xi(\upsilon_l x)\, \mathrm{d}x
\end{aligned}$$

For $x < 0$, as $\sqrt{T}h^* \to \infty$, using Taylor expansions we find

$$\begin{aligned}
\left|x - \frac{h^*}{2}\right| \exp\left\{\frac{h^* T}{2\sigma^2} x\right\} &= \left|x - \frac{h^*}{2}\right| \left[1 + \frac{h^* T}{2\sigma^2} x + \frac{1}{2}\left(\frac{h^* T}{2\sigma^2} x\right)^2 + \frac{1}{3!}\left(\frac{h^* T}{2\sigma^2} x\right)^3 \ldots\right]^{-1} \\
&= \mathcal{O}(|h^*|) + \mathcal{O}(|Th^*|^{-1}) \\
&= \mathcal{O}(|h^*|),
\end{aligned}$$

Then, by putting these components back together we have

$$\begin{aligned}
|\mathcal{Q}_{1,1}(h^*)| \leq\ & \mathcal{O}\left(|h^*| \exp\left\{-\frac{5h^{*2}T}{8\sigma^2}\right\}\right) \\
& \times \int_{-\infty}^{0} \frac{\sqrt{T}}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{T}{2\sigma^2}(h^* - x)^2\right\} \upsilon_l\, \xi(\upsilon_l x)\, \mathrm{d}x. \qquad \text{(A.57)}
\end{aligned}$$

This implies that if $\mathcal{Q}_0(h^*)$ is as defined in (5.28)

$$\left|\frac{\mathcal{Q}_{1,1}(h^*)}{\mathcal{Q}_0(h^*)}\right| = \mathcal{O}(|h^*|). \qquad \text{(A.58)}$$

The second integral is defined as

$$\mathcal{Q}_{1,2}(h^*) = \int_{-h^*/2}^{3h^*/2} y\,\sqrt{T}\varphi_\sigma\Big(\sqrt{T}[h^* - y]\Big)\ \upsilon_l\,\xi(\upsilon_l y)\,\mathrm{d}y.$$

Using integration by parts, where

$$\begin{aligned} u &= y\xi(\upsilon_l y) & \mathrm{d}v &= \varphi_\sigma\Big(\sqrt{T}[h^* - y]\Big) \\ \mathrm{d}u &= y\upsilon_l\xi'(\upsilon_l y) + \xi(\upsilon_l y) & v &= \Phi\left(\frac{\sqrt{T}}{\sigma}[h^* - y]\right) \end{aligned}$$

we can write $\mathcal{Q}_{1,2}(h^*)$ as

$$\begin{aligned} \mathcal{Q}_{1,2}(h^*) = &\left[y\xi(\upsilon_l y)\Phi\left(\frac{\sqrt{T}}{\sigma}[h^* - y]\right)\right]_{-h^*/2}^{3h^*/2} \\ &- \int_{-h^*/2}^{3h^*/2} \Phi\left(\frac{\sqrt{T}}{\sigma}[h^* - y]\right)\left[y\upsilon_l\xi'(\upsilon_l y) + \xi(\upsilon_l y)\right]\mathrm{d}y. \end{aligned}$$

Then, using integration by parts on $\mathcal{Q}_0(h^*)$, however letting

$$u = \xi(\upsilon_l y) \qquad\qquad \mathrm{d}u = \upsilon_l\xi'(\upsilon_l y)$$

instead, we find that

$$\mathcal{Q}_0(h^*) = \left[\xi(\upsilon_l y)\Phi\left(\frac{\sqrt{T}}{\sigma}[h^* - y]\right)\right]_{-h^*/2}^{3h^*/2} - \int_{-h^*/2}^{3h^*/2} \Phi\left(\frac{\sqrt{T}}{\sigma}[h^* - y]\right)\upsilon_l\xi'(\upsilon_l y)\mathrm{d}y.$$

Through the cancellation of terms, we find

$$\left|\frac{\mathcal{Q}_{1,2}(h^*)}{\mathcal{Q}_0(h^*)}\right| = \mathcal{O}(|h^*|). \tag{A.59}$$

The third integral, $\mathcal{Q}_{1,3}(h^*)$, is similar to the first, $\mathcal{Q}_{1,1}(h^*)$,

$$|\mathcal{Q}_{1,3}(h^*)| = \int_{3h^*/2}^{\infty} \frac{|y|\,\sqrt{T}}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{T}{2\sigma^2}[h^* - y]^2\right\}\ \upsilon_l\,\xi(\upsilon_l y)\,\mathrm{d}y.$$

Let $y = x + 3h^*/2$ and considering $\xi(\upsilon_l[x + 3h^*/2]) \leq \xi(\upsilon_l x)$ for $x > 0$, therefore $\mathcal{Q}_{1,3}(h^*)$ can be written as

$$|\mathcal{Q}_{1,3}(h^*)| \leq \int_0^{\infty} \left|x + \frac{3h^*}{2}\right| \frac{\sqrt{T}}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{T}{2\sigma^2}\left(-\frac{3h^*}{2} - x\right)^2\right\}\ \upsilon_l\,\xi(\upsilon_l x)\,\mathrm{d}x.$$

The quadratic equation in the exponential can be written as

$$\left(-\frac{3h^*}{2}-x\right)^2-(h^*-x)^2=-\frac{3h^{*2}}{4}+3h^*x.$$

The Taylor expansion of $\exp\left\{-\frac{3h^*T}{2\sigma^2}x\right\}$ is

$$\exp\left\{-\frac{3h^*T}{2\sigma^2}x\right\}=\left[1+\frac{3h^*T}{2\sigma^2}x+\frac{1}{2}\left(\frac{3h^*T}{2\sigma^2}x\right)^2+\frac{1}{3!}\left(\frac{3h^*T}{2\sigma^2}x\right)^3+\ldots\right]^{-1}$$

This implies that

$$\begin{aligned}
&\left|\frac{3h^*}{2}+x\right|\exp\left\{-\frac{3h^*T}{2\sigma^2}x\right\}\\
&\qquad=\left|\frac{3h^*}{2}+x\right|\left[1+\frac{3h^*T}{2\sigma^2}x+\frac{1}{2}\left(\frac{3h^*T}{2\sigma^2}x\right)^2+\frac{1}{3!}\left(\frac{3h^*T}{2\sigma^2}x\right)^3+\ldots\right]^{-1}\\
&\qquad=\mathcal{O}(|h^*|)+\mathcal{O}(|Th^*|^{-1})\\
&\qquad=\mathcal{O}(|h^*|).
\end{aligned}$$

So we can show that

$$\begin{aligned}
|\mathcal{Q}_{1,3}(h^*)|\leq\mathcal{O}&\left(\exp\left\{\frac{3h^{*2}}{4}\right\}h^*\right)\\
&\times\int_0^\infty\frac{\sqrt{T}}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{T}{2\sigma^2}\left(h^*-x\right)^2\right\}\upsilon_l\,\xi(\upsilon_l x)\,\mathrm{d}x.
\end{aligned}\tag{A.60}$$

Similarly to the previous two integrals, dividing $\mathcal{Q}_{1,3}(h^*)$ by $\mathcal{Q}_0(h^*)$ as defined in (5.28) results in

$$\left|\frac{\mathcal{Q}_{1,3}(h^*)}{\mathcal{Q}_0(h^*)}\right|=\mathcal{O}(|h^*|).\tag{A.61}$$

Therefore, considering all three sections

$$\left|\frac{\mathcal{Q}_1(h^*)}{\mathcal{Q}_0(h^*)}\right|=\mathcal{O}(|h^*|).\tag{A.62}$$

(ii) To show that $|\mathcal{Q}_1^*(h^*)/\mathcal{Q}_1^*(h^*)|=\mathcal{O}(|h^*|)$, also requires splitting $\mathcal{Q}_1^*(h^*)$ into three sections: $\mathcal{Q}_{1,1}^*(h^*)$ for values over $(-\infty,-h^*)$; $\mathcal{Q}_{1,2}^*(h^*)$ is the integral over $(-h^*,h^*)$; $\mathcal{Q}_{1,3}^*(h^*)$ the integral over (h^*,∞). The first integral is

$$\mathcal{Q}_{1,1}^*(h^*)=\int_{-\infty}^{-h^*}y\sqrt{T}\,f_E\Big(\sqrt{T}[h^*-y]\Big)\,\upsilon_l\,\xi(\upsilon_l y)\mathrm{d}y.$$

In equation (A.37), we calculated $f'_E(x)$, therefore $f'_E(x)/f_E(x)$ is $1/2 - \gamma^* e^x$. Let $y = -(x + h^*)$, as $\xi(x)$ is a zero mean unimodal pdf we know that $\xi(x) = \xi(-x)$

$$\mathcal{Q}^*_{1,1}(h^*) = \int_0^\infty (x + h^*)\sqrt{T}\, f_E\Big(\sqrt{T}[x + 2h^*]\Big)\ \upsilon_l\, \xi(\upsilon_l[x + h^*])\mathrm{d}x,$$

As ξ is a symmetric unimodal probability density function this implies that $\xi(\upsilon_l[x + h^*]) \leq \xi(\upsilon_l x)$. We wish to re-write the function $f_E(\sqrt{T}[x + 2h^*])$ as $f_E(\sqrt{T}[x + h^*])$. Taking the first exponent of equation (5.4) let

$$\begin{aligned} y_1(x) &= \sqrt{T}[x + 2h^*] - \frac{\gamma^*}{2}\exp\Big\{2\sqrt{T}[x + 2h^*]\Big\} \\ y_2(x) &= \sqrt{T}[x + h^*] - \frac{\gamma^*}{2}\exp\Big\{2\sqrt{T}[x + 2h^*]\Big\}. \end{aligned}$$

Then,

$$y_1(x) - y_2(x) = \sqrt{T}h^* - \frac{\gamma^*}{2}\left(\exp\Big\{2\sqrt{T}[x + 2h^*]\Big\} - \exp\Big\{2\sqrt{T}[x + h^*]\Big\}\right).$$

The Taylor expansion of $\exp\{\sqrt{T}[x + 2h^*]\}$ is

$$\exp\Big\{\sqrt{T}[x + 2h^*]\Big\} = 1 + 2\sqrt{T}(x + 2h^*) + T(x + 2h^*)^2 + \frac{2}{3!}T^{3/2}(x + 2h^*)^3 + \ldots$$

and the Taylor expansion of $\exp\Big(\sqrt{T}[x + h^*]\Big)$ is

$$\exp\Big\{\sqrt{T}[x + h^*]\Big\} = 1 + 2\sqrt{T}(x + h^*) + T(x + h^*)^2 + \frac{2}{3!}T^{3/2}(x + h^*)^3 + \ldots$$

Therefore, we find that

$$\begin{aligned} y_1(x) - y_2(x) &= \sqrt{T}h^* - \gamma^*\left(\sqrt{T}h^* + Thx + 3Th^{*2} + \mathcal{O}(T^{3/2}h^*)\right) \\ &\geq -Th^*x - 3Th^{*2} - \mathcal{O}(T^{3/2}h^*). \end{aligned}$$

Therefore $\exp\{y_1(x)\} \leq \exp\{y_2(x) - Th^*(x + 3h^*)\}$, and

$$\begin{aligned} \left|\frac{\mathcal{Q}^*_{1,1}(h^*)}{\mathcal{Q}^*_0(h^*)}\right| &= \frac{\int_0^\infty (x + h^*)\sqrt{T}\, f_E\Big(\sqrt{T}[x + h^*]\Big)\ e^{-Th(x+3h^*)}\ \upsilon_l\, \xi(\upsilon_l x)\mathrm{d}x}{\int_0^\infty (x + h^*)\sqrt{T}\, f_E\Big(\sqrt{T}[x + h^*]\Big)\ \upsilon_l\, \xi(\upsilon_l x)\mathrm{d}x} \\ &= \mathcal{O}(|Th|^{-1}) + \mathcal{O}(|h^*|) \\ &= \mathcal{O}(|h^*|) \end{aligned} \tag{A.63}$$

For the second integral over $x \in (-h^*, h^*)$, $\mathcal{Q}_{1,1}(h^*)$, we follow the same procedure used to solve $\mathcal{Q}_{1,1}(h^*)$, which also results in

$$\left| \frac{\mathcal{Q}^*_{1,2}(h^*)}{\mathcal{Q}^*_0(h^*)} \right| = \mathcal{O}(|h^*|). \tag{A.64}$$

In the third integral, let $y = h^* - x$, which implies

$$\begin{aligned}
\mathcal{Q}^*_{1,3}(h^*) &= \int_{-\infty}^{0} (h^* - x)\sqrt{T}\, f_E\Big(\sqrt{T}x\Big)\ \upsilon_l\, \xi(\upsilon_l[h^* - x])\mathrm{d}x \\
&= h^* \int_{-\infty}^{0} \sqrt{T}\, f_E\Big(\sqrt{T}x\Big)\ \upsilon_l\, \xi(\upsilon_l[h^* - x])\mathrm{d}x \qquad \text{(A.65)} \\
&\quad + \int_{-\infty}^{0} x\sqrt{T}\, f_E\Big(\sqrt{T}x\Big)\ \upsilon_l\, \xi(\upsilon_l[h^* - x])\mathrm{d}x. \qquad \text{(A.66)}
\end{aligned}$$

It is easy to see that when divided by $Q^*_0(h^*)$ equation (A.65) will be $\mathcal{O}(h^*)$, however it is not as clear for equation (A.66). We know that $f_E(x) \leq \sqrt{2\gamma^*\pi^{-1}}e^x$, so substituting this into equation (A.66) results in

$$-\int_{-\infty}^{0} x\sqrt{T}\, f_E\Big(\sqrt{T}x\Big)\ \upsilon_l\, \xi(\upsilon_l[h^* - x])\mathrm{d}x \leq -\upsilon_l \sqrt{\frac{2\,\gamma^*\,T}{\pi}} \int_{-\infty}^{0} x\, e^{\sqrt{T}x}\, \xi(\upsilon_l[h^* - x])\mathrm{d}x.$$

By just considering the integral and using integration by parts, let

$$\begin{aligned}
u = -x\,\xi(\upsilon_l[h^* - x]) \qquad &\Rightarrow \qquad \mathrm{d}u = x\,\upsilon_l\,\xi'(\upsilon_l[h^* - x]) - \xi(\upsilon_l[h^* - x]), \\
\mathrm{d}v = e^{\sqrt{T}x} \qquad &\Rightarrow \qquad v = \frac{1}{\sqrt{T}} e^{\sqrt{T}x}.
\end{aligned}$$

Therefore

$$\begin{aligned}
&\int_{-\infty}^{0} x\, e^{\sqrt{T}x}\, \xi(\upsilon_l[h^* - x])\mathrm{d}x \\
&\qquad = \underbrace{\left[-\frac{x}{\sqrt{T}}\, \xi(\upsilon_l[h^* - x])\, e^{\sqrt{T}x} \right]_{-\infty}^{0}}_{\to 0} \\
&\qquad\quad - \frac{1}{\sqrt{T}} \int_{-\infty}^{0} x\,\upsilon_l\,\xi'(\upsilon_l[h^* - x])\, e^{\sqrt{T}x}\mathrm{d}x + \frac{1}{\sqrt{T}} \int_{-\infty}^{0} e^{\sqrt{T}x}\, \xi(\upsilon_l[h^* - x])\mathrm{d}x
\end{aligned} \tag{A.67}$$

As both of the integrals in equation (A.67) will be positive this implies that

$$\int_{-\infty}^{0} x\, e^{\sqrt{T}x}\, \xi(\upsilon_l[h^* - x])\mathrm{d}x \leq \frac{1}{\sqrt{T}} \int_{-\infty}^{0} e^{\sqrt{T}x}\, \xi(\upsilon_l[h^* - x])\mathrm{d}x,$$

dividing this integral by $\mathcal{Q}_0^*(h^*)$ will result $\mathcal{O}(T^{-1/2})$. Therefore, $|\mathcal{Q}_{1,3}^*(h^*)/\mathcal{Q}_0^*(h^*)| = \mathcal{O}(|h^*|)$ and

$$\left|\frac{\mathcal{Q}_1^*(h^*)}{\mathcal{Q}_0^*(h^*)}\right| = \mathcal{O}(|h^*|).$$

□

Lemma A.1.5. *If the function $g \in B_{p,q}^r(A)$, then for any constant $a > 0$,*

(i) $\sum_l \sum_{m=1}^{2^{J-l}} b_{l,m}^2 \mathbb{I}\Big(\sqrt{T}|b_{l,m}| \le a\sqrt{\log(T)}\Big) = \mathcal{O}\bigg(\Big[\frac{\log(T)}{T}\Big]^{2r/(2r+1)}\bigg)$,

(ii) $\sum_l \sum_{m=1}^{2^{J-l}} T^{-1} \mathbb{I}\Big(\sqrt{T}|b_{l,m}| > a\sqrt{\log(T)}\Big) = \mathcal{O}\big(T^{-2r/(2r+1)}[\log(T)]^{-p/2}\big)$.

Proof (Lemma A.1.5)**.** The proof of this can be found in Donoho et al. (1996) □

Proof (Theorem 5.4.1)**.** As the wavelet basis is orthonormal, the mean integrated squared error (MISE) can be expressed as

$$R(T, \hat{g}, \mathcal{F}) = \mathbb{E}\Big[(\hat{b}_0 - b_0)^2\Big] + \sum_{l=1}^{J} \sum_{m=1}^{2^{J-l}} \mathbb{E}\Big[(\hat{b}_{l,m} - b_{l,m})^2\Big] + \sum_{l=1}^{\infty} \sum_{m=1}^{\infty} b_{l,m}^2. \tag{A.68}$$

Given the data h^*, we obtain $b_0 = h_0^*$ and $\hat{b_0}$ is simply b_0. The first part in equation (A.68) can be re-written as

$$\begin{aligned}
\mathbb{E}\Big[(\hat{b}_0 - b_0)^2\Big] &= \text{Var}\left[\hat{b}_0\right] + \text{Var}\left[b_0\right] + 2\text{Cov}\left[\hat{b}_0, b_0\right] \\
&= \text{Var}[h_0^*] + \text{Var}[h_0^*] + 2\text{Var}[h_0^*] \\
&= \frac{\pi^2}{2T} + \frac{\pi^2}{2T} + \frac{\pi^2}{T} \\
&= \mathcal{O}(T^{-1})
\end{aligned} \tag{A.69}$$

From equation (5.32), the last part of the MISE is bounded by $2^{-2rJ}A = \mathcal{O}(T^{-2r})$. Therefore, the second term provides the main contribution to the MISE.

Separate the scales, $l = 0, \ldots, J$, into: coarse $0 \le l \le J_0$; intermediate $J_0 < l \le J_1$, fine $J_1 < l \le J_2$; and the finest scales $l > J_2$. Then separate the second term into four sums over these partitions, denoting them as the risks: R_1, R_2, R_3 and R_4 over the coarse, intermediate, fine and finest scales respectively.

Coarse Scales: $0 \le l \le J_0$

We shall begin by examining the coarsest scales. As the scales are coarse, asymptotically the wavelet transform of the wavelet coefficients will be asymptotically Gaussian, as shown

in (Moulin, 1994, section III, part B). If we let

$$\mathcal{A}_{l;T}(h^*) = \frac{\theta_l\sqrt{T}}{\mathcal{Q}_0(h^*)}\varphi_\sigma(\sqrt{T}h^*), \tag{A.70}$$

where $\mathcal{Q}_i(\cdot)$ is as defined in (5.28) and $\varphi_\sigma(\cdot)$ is the Gaussian pdf with mean zero and variance $\sigma^2 = {}^{\pi^2}\!/_2$. Assuming the distribution of the wavelet coefficients is approximately Gaussian, implies we can set $\lambda_l = 0$ in what follows. Therefore

$$\begin{aligned}
\hat{b}_{l,m} &= \frac{\mathcal{Q}_1(h^*_{l,m})}{\theta_l\sqrt{T}\varphi(\sqrt{T}h^*_{l,m}) + \mathcal{Q}_0(h^*_{l,m})} \\
&= \frac{\mathcal{Q}_1(h^*_{l,m})}{\mathcal{Q}_0(h^*_{l,m})}\left(\frac{\theta_l\sqrt{T}}{\mathcal{Q}_0(h^*_{l,m})}\varphi(\sqrt{T}h^*_{l,m}) + 1\right)^{-1} \\
&= \frac{\mathcal{Q}_1(h^*_{l,m})}{\mathcal{Q}_0(h^*_{l,m})}\left(\mathcal{A}_{l;T}(h^*_{l,m}) + 1\right)^{-1}. \qquad \text{(A.71)}
\end{aligned}$$

Consider that

$$\begin{aligned}
\mathbb{E}\Big[(\hat{b}_{l,m} - b_{l,m})^2\Big] &= \mathbb{E}\left[\left(\hat{b}_{l,m} - b_{l,m} - \frac{\mathcal{Q}_1(h^*_{l,m})}{\mathcal{Q}_0(h^*_{l,m})} + \frac{\mathcal{Q}_1(h^*_{l,m})}{\mathcal{Q}_0(h^*_{l,m})}\right)^2\right] && \text{(A.72)} \\
&= \mathbb{E}\left[\left(\frac{\mathcal{Q}_1(h^*_{l,m})}{\mathcal{Q}_0(h^*_{l,m})} - b_{l,m}\right)^2\right] + \mathbb{E}\left[\left(\frac{\mathcal{Q}_1(h^*_{l,m})}{\mathcal{Q}_0(h^*_{l,m})} - \hat{b}_{l,m}\right)^2\right] && \text{(A.73)} \\
&\quad - 2\mathbb{E}\left[\left(\frac{\mathcal{Q}_1(h^*_{l,m})}{\mathcal{Q}_0(h^*_{l,m})} - b_{l,m}\right)\left(\frac{\mathcal{Q}_1(h^*_{l,m})}{\mathcal{Q}_0(h^*_{l,m})} - \hat{b}_{l,m}\right)\right] && \text{(A.74)} \\
&\leq 2\mathbb{E}\left[\left(\frac{\mathcal{Q}_1(h^*_{l,m})}{\mathcal{Q}_0(h^*_{l,m})} - \hat{b}_{l,m}\right)^2 + \left(\frac{\mathcal{Q}_1(h^*_{l,m})}{\mathcal{Q}_0(h^*_{l,m})} - b_{l,m}\right)^2\right]. && \text{(A.75)}
\end{aligned}$$

If we then substitute in eqaution (A.71) for $\hat{b}_{l,m}$, we can write

$$\begin{aligned}
\frac{\mathcal{Q}_1(h^*_{l,m})}{\mathcal{Q}_0(h^*_{l,m})} - \hat{b}_{l,m} &= \frac{\mathcal{Q}_1(h^*_{l,m})}{\mathcal{Q}_0(h^*_{l,m})} - \frac{\mathcal{Q}_1(h^*_{l,m})}{\mathcal{Q}_0(h^*_{l,m})}\left(\mathcal{A}_{l;T}(h^*_{l,m}) + 1\right)^{-1} \\
&= \frac{\mathcal{Q}_1(h^*_{l,m})}{\mathcal{Q}_0(h^*_{l,m})}\left(1 - \frac{1}{\mathcal{A}_{l;T}(h^*_{l,m}) + 1}\right) \\
&= \frac{\mathcal{Q}_1(h^*_{l,m})}{\mathcal{Q}_0(h^*_{l,m})}\left(\frac{\mathcal{A}_{l;T}(h^*_{l,m})}{\mathcal{A}_{l;T}(h^*_{l,m}) + 1}\right)
\end{aligned}$$

Therefore, $R_1 \leq 2(R_{1,1} + R_{1,2})$, where

$$R_{1,1} = \sum_{l=0}^{J_0} \sum_{m=0}^{2^l-1} \mathbb{E}\left[\left(\frac{\mathcal{Q}_1(h^*_{l,m})}{\mathcal{Q}_0(h^*_{l,m})} - b_{l,m}\right)^2\right] \tag{A.76}$$

$$R_{1,2} = \sum_{l=0}^{J_0} \sum_{m=0}^{2^l-1} \mathbb{E}\left[\left(\frac{\mathcal{Q}_1(h^*_{l,m})\mathcal{A}_{l;T}(h^*_{l,m})}{\mathcal{Q}_0(h^*_{l,m})[1+\mathcal{A}_{l;T}(h^*_{l,m})]}\right)^2\right] \tag{A.77}$$

An important feature of $\mathcal{A}_{l;T}(h^*_{l,m})$ is that

$$\frac{\mathcal{A}_{l;T}(h^*_{l,m})}{1+\mathcal{A}_{l;T}(h^*_{l,m})} \approx \min(1, \mathcal{A}_{l;T}(h^*_{l,m})), \tag{A.78}$$

For $l \leq J_0$, using a combination of Lemma A.1.1, equation (5.36) and the same proceedure as (A.75), $R_{1,1}$ can be re-written as

$$\mathbb{E}\left[\left(\frac{\mathcal{Q}_1(h^*_{l,m})}{\mathcal{Q}_0(h^*_{l,m})} - b_{l,m}\right)^2\right] \leq 2\left\{\mathbb{E}\left[\left(\frac{\mathcal{Q}_1(h^*_{l,m})}{\mathcal{Q}_0(h^*_{l,m})} - h^*_{l,m}\right)^2\right] + \mathbb{E}\Big[\big(h^*_{l,m} - b_{l,m}\big)^2\Big]\right\}$$

Using lemmas A.1.1 and A.1.3 (i), this is equal to

$$\begin{aligned} 2\left\{\mathbb{E}\left[\left(\frac{\mathcal{Q}_1(h^*_{l,m})}{\mathcal{Q}_0(h^*_{l,m})} - h^*_{l,m}\right)^2\right] + \mathbb{E}\Big[\big(h^*_{l,m} - b_{l,m}\big)^2\Big]\right\} &= \mathcal{O}\left(\mathbb{E}\left[\frac{\upsilon_l^4 {h^*_{l,m}}^2}{T^2}\right] + \frac{\upsilon_l^2}{T^2} + \frac{\sigma^2}{T}\right) \\ &= \mathcal{O}\left(\frac{\upsilon_l^4 b_{l,m}^2}{T^2} + \frac{\upsilon_l^2}{T^2} + \frac{\sigma^2}{T}\right), \end{aligned} \tag{A.79}$$

Replace the summation of $b^2_{l,m}$ by (5.32), and substite in (A.76) into (A.79), we have

$$R_{1,1} = \mathcal{O}\left(\sum_{l=0}^{J_0} 2^{-2\omega l}\frac{\upsilon_l^4}{T^2} + 2^l\frac{\upsilon_l^2}{T^2} + \frac{2^l}{T}\right). \tag{A.80}$$

Then by re-writing the summation using the representation of a geometric progression, this implies that equation (A.80) is

$$\begin{aligned} &\sum_{l=0}^{J(2r+1)^{-1}} 2^{-2\omega l}\frac{2^{4\mu_1 l}}{T^2} + 2^l\frac{2^{2\mu_1 l}}{T^2} + \frac{2^l}{T} \\ &\quad= \sum_{l=0}^{J(2r+1)^{-1}} \frac{(2^{4\mu_1-2\omega})^l}{T^2} + \frac{(2^{1+2\mu_1})^l}{T^2} + \frac{2^l}{T} \\ &\quad= \frac{1-(2^{4\mu_1-2\omega})^{J(2r+1)^{-1}+1}}{T^2(1-2^{4\mu_1-2\omega})} + \frac{1-(2^{1+2\mu_1})^{J(2r+1)^{-1}+1}}{T^2(1-2^{1+2\mu_1})} \\ &\qquad\qquad - \frac{1-2^{J(2r+1)^{-1}+1}}{T}. \end{aligned} \tag{A.81}$$

As $J = \log_2(T)$ and from assumption (5.43), $\upsilon_l = 2^{\mu l}$ with $\mu_1 = r + 3/4 - (2p^*)^{-1}$, then by substituting in these values and considering the value of 2ω from equation (A.22), this implies that equation (A.81) can be written as

$$\begin{aligned}
&\frac{T^{-2} - 2^{4\mu_1 - 2\omega} T^{2(2r+1)^{-1}(2\mu_1 - \omega) - 2}}{(1 - 2^{2(\mu_1 - \omega)})} + \frac{T^{-2} - 2^{1+2\mu_1} T^{(2r+1)^{-1}(1+2\mu_1)-2}}{(1 - 2^{1+2\mu_1})} \\
&\qquad - T^{-1} + 2T^{(2r+1)^{-1}-1} \\
&= \frac{T^{-2} - T^{-2r/(2r+1)}}{1 - 2^{1/2+1/p^*}} + \frac{T^{-2} - T^{-(2r+1/2-1/p^*)/(2r+1)}}{1 - 2^{2r+5/3-1/p^*}} - T^{-1} + T^{-2r/(2r+1)}.
\end{aligned}$$

Therefore, as $1 \geq p^* \geq 2$ and as r is greater than $r^* \geq 0$, as $r \to \infty$ then $2r(2r+1)^{-1} \to 1$ this implies that

$$R_{1,1} = \mathcal{O}\Big(T^{-2r/(2r+1)}\Big). \tag{A.82}$$

To calculate $R_{1,2}$, using (A.78) as $T \to \infty$ we find

$$\left| \frac{\mathcal{Q}_1(h^*_{l,m}) \mathcal{A}_{l;T}(h^*_{l,m})}{\mathcal{Q}_0(h^*_{l,m})[1 + \mathcal{A}_{l;T}(h^*_{l,m})]} \right| = \left| \frac{\mathcal{Q}_1(h^*_{l,m})}{\mathcal{Q}_0(h^*_{l,m})} \mathcal{A}_{l;T}(h^*_{l,m}) \right|.$$

Using the definition of $\mathcal{A}_{l;T}(h^*_{l,m})$ from (A.70), and using lemmas A.1.1(ii)a. and A.1.4(i) this becomes

$$\begin{aligned}
\left| \frac{\mathcal{Q}_1(h^*_{l,m}) \mathcal{A}_{l;T}(h^*_{l,m})}{\mathcal{Q}_0(h^*_{l,m})[1 + \mathcal{A}_{l;T}(h^*_{l,m})]} \right| &= \mathcal{O}\left(|h^*_{l,m}| \frac{\theta_l \sqrt{T}}{\mathcal{Q}_0(h^*)} \varphi_\sigma(\sqrt{T} h^*) \right) \\
&= \mathcal{O}\left(\frac{\theta_l \sqrt{T} |h^*_{l,m}|}{\upsilon_l} \right)
\end{aligned} \tag{A.83}$$

Therefore, by substituting in equation (A.83) into $R_{1,2}$ from equation (A.77), this implies that

$$R_{1,2} = \mathcal{O}\left(\sum_{l=0}^{J_0} \frac{\theta_l^2 T}{\upsilon_l^2} \sum_{m=0}^{2^l - 1} b_{l,m}^2 + \frac{1}{T} \right). \tag{A.84}$$

Next we substitute in the summation of $b_{l,m}^2$ from equation (5.32) and $\upsilon_l^2 = 2^{2\mu_1 l}$, the summation from equation (A.84) can be written as

$$R_{1,2} = \mathcal{O}\left(\sum_{l=0}^{J_0} 2^{-2\mu_1 l} \theta_l^2 \left\{ T 2^{-2\omega l} + 2^l - 1 \right\} \right).$$

Then we substitute in θ_l^2 from equation (5.44), this yields

$$R_{1,2} = \mathcal{O}\Bigg(\sum_{l=0}^{J_0} 2^{l(4r+1+3/2-3/p^*)}\, T^{-(4r+1)/(2r+1)} \times \Big\{ T 2^{-2l(r+1/2-1/p^*+2\mu_1)} + 2^{l(1+2\mu_1)} - 2^{-2\mu_1 l} \Big\} \Bigg).$$

If we substitute in J_0 from (5.39), this results in

$$R_{1,2} = \mathcal{O}\left(T^{-(4r+1)/(2r+1)} \sum_{l=0}^{J(2r+1)^{-1}} T 2^{l(2r+3/2-2/p^*-2\mu_1)} + 2^{l(4r+7/2-3/p^*-2\mu_1)} - 2^{2l(2r+5/4-3/2p^*-\mu_1)} \right).$$

Finally, substitute in $\mu_1 = r + 3/4 - (2p^*)^{-1}$ which results in

$$\begin{aligned} R_{1,2} &= \mathcal{O}\left(T^{-(4r+1)/(2r+1)} \sum_{l=0}^{J(2r+1)^{-1}} T + 2^{l(2r+2-1/p^*)} \right) \\ &= \mathcal{O}\left(T^{-2r/(2r+1)} \frac{\log_2(T)}{2r+1} + T^{-(4r+1)/(2r+1)} \sum_{l=0}^{J(2r+1)^{-1}} 2^{l(2r+2-1/p^*)} \right) \end{aligned}$$

Therefore, as the summation is a geometric series and $T^{-2r/(2r+1)}$ tends to zero faster than $\log_2(T)$ tends to infinity, we find that

$$\begin{aligned} R_{1,2} &= \mathcal{O}\left(T^{-2r/(2r+1)} \frac{\log_2(T)}{2r+1} + T^{-(4r+1)/(2r+1)} \frac{1 - 2^{J(2r+2-1/p^*)/(2r+1)}}{1 - 2^{2r+2-1/p^*}} \right) \\ &= \mathcal{O}\Big(T^{-2r/(2r+1)} \Big). \end{aligned} \tag{A.85}$$

Intermediate Scales: $J_0 < l \leq J_1$

For intermediate scales let $\lambda_l = 0$, because we found in section 5.2.1, that the value of λ swifty converged to zero as the support of the wavelet increased. Also, let $1 \leq p < 2$. These imply that $\hat{b}_{l,m}$ will be the same as equation (A.71). Partition R_2 into

$$\begin{aligned} R_{2,1} &= \sum_{l=J_0+1}^{J_1} \sum_{m=0}^{2^l-1} \mathbb{E}\Big[(\hat{b}_{l,m} - b_{l,m})^2 \Big] \mathbb{I}\Big(|b_{l,m}| > \sqrt{T^{-1}\log(T)} \Big), \\ R_{2,2} &= \sum_{l=J_0+1}^{J_1} \sum_{m=0}^{2^l-1} \mathbb{E}\Big[(\hat{b}_{l,m} - b_{l,m})^2 \Big] \mathbb{I}\Big(|b_{l,m}| \leq \sqrt{T^{-1}\log(T)} \Big), \end{aligned}$$

where we consider the values of $b_{l,m}$ greater than and less than $\sqrt{T^{-1}\log(T)}$ separately. Further partition $R_{2,1}$ into

$$R_{2,11} = \sum_{l=J_0+1}^{J_1} \sum_{m=0}^{2^l-1} \mathbb{E}\left[\left(\frac{\mathcal{Q}_1(h^*_{l,m})}{\mathcal{Q}_0(h^*_l, m)} - b_{l,m}\right)^2\right] \mathbb{I}\Big(|b_{l,m}| > \sqrt{T^{-1}\log(T)}\Big),$$
$$R_{2,12} = \sum_{l=J_0+1}^{J_1} \sum_{m=0}^{2^l-1} \mathbb{E}\left[\left(\frac{\mathcal{Q}_1(h^*_{l,m})A_{l,T}(h^*_{l,m})}{\mathcal{Q}_0(h^*_{l,m})[1+A_{l,T}(h^*_{l,m})]}\right)^2\right] \mathbb{I}\Big(|b_{l,m}| > \sqrt{T^{-1}\log(T)}\Big),$$

as we separated R_1. To determine the order of convergence of $R_{2,11}$ repeat equation (A.79) and chose $\mu = \mu_2$. Then we find that

$$\sum_{l=J_0+1}^{J_1} \sum_{m=0}^{2^l-1} \frac{\upsilon_l^2}{T^2} + \frac{b_{l,m}^2\upsilon_l^4}{T^2} = \sum_{l=J_0+1}^{J_1} \sum_{m=0}^{2^l-1} \frac{2^{2\mu_2 l}}{T^2} + \frac{b_{l,m}^2 2^{4\mu_2 l}}{T^2}$$
$$= \mathcal{O}\left(\sum_{l=J_0+1}^{J_1} 2^{(2\mu_2+1)l}T^{-2} + 2^{4\mu_2 l}2^{-2(r+1/2-1/p^*)l}T^{-2}\right).$$

From the definition of 2ω in equation (A.22), which implies

$$\mathcal{O}\left(\sum_{l=J_0+1}^{J_1} 2^{(2\mu_2+1)l}T^{-2} + 2^{4\mu_2 l}2^{-2\omega l}T^{-2}\right)$$
$$= \mathcal{O}\left(\sum_{l=J_0+1}^{J_1} 2^{(2\mu_2+1)l}T^{-2} \; + \; 2^{(4\mu_2-2\omega)l}T^{-2}\right)$$
$$= \mathcal{O}\left(\frac{2^{(2\mu_2+1)(J_0+1)} - 2^{(2\mu_2+1)J_1}}{T^2(1-2^{2\mu_2+1})} \; + \; \frac{2^{(4\mu_2-2\omega)(J_0+1)} \; - \; 2^{(4\mu_2-2\omega)J_1}}{T^2(1-2^{4\mu_2-2\omega})}\right)$$

By substituting in J_0 from (5.39) and J_1 from (5.40)this implies that

$$\mathcal{O}\left(\frac{2^{(2\mu_2+1)(J_0+1)} - 2^{(2\mu_2+1)J_1}}{T^2(1-2^{2\mu_2+1})} \; + \; \frac{2^{(4\mu_2-2\omega)(J_0+1)} - 2^{(4\mu_2-2\omega)J_1}}{T^2(1-2^{4\mu_2-2\omega})}\right)$$
$$= \mathcal{O}\left(\frac{2^{(2\mu_2+1)(J/(2r+1)+1)} - 2^{(2\mu_2+1)(2rJ/2\omega(2r+1))}}{T^2(1-2^{2\mu_2+1})}\right.$$
$$\left. + \frac{2^{(4\mu_2-2\omega)(J/(2r+1)+1)} - 2^{(4\mu_2-2\omega)(2rJ/2\omega(2r+1))}}{T^2(1-2^{4\mu_2-2\omega})}\right) \tag{A.86}$$

Substituting in μ_2 written in terms of ω into equation (A.86) yields

$$\mathcal{O}\left(\frac{2^{2\omega(1+1/r)(J/(2r+1)+1)} - 2^{(1+1/r)(2rJ/(2r+1))}}{T^2(1-2^{2\omega(1+1/r)})}\right.$$
$$\left. + \frac{2^{(2\omega[1+2/r]-2)(J/(2r+1)+1)} - 2^{2rJ(1+2/r-1/\omega)/(2r+1)}}{T^2(1-2^{2\omega[1+2/r]-1})}\right)$$

As $J = \log_2(T)$ this implies that

$$\mathcal{O}\Bigg(\frac{2^{2\omega(1+1/r)}T^{2\omega(1+1/r)/(2r+1)-2} - T^{(2r+2)/(2r+1)-2}}{(1-2^{2\omega(1+1/r)})} + \frac{2^{2\omega[1+2/r]-2}T^{(2\omega[1+2/r]-2)/(2r+1)-2} - T^{2r(1+2/r-1/\omega)/(2r+1)-2}}{(1-2^{2\omega[1+2/r]-1})}\Bigg)$$
$$= \mathcal{O}\Bigg(\frac{2^{2\omega(1+1/r)}T^{(-2r+[1+1/r][1-2/p^*])/(2r+1)} - T^{-2r/(2r+1)}}{(1-2^{2\omega(1+1/r)})} + \frac{2^{2\omega[1+2/r]-2}T^{(-2r-[2/p^*+1/p^*r-1/2r])/(2r+1)} - T^{-2r(1+1/2\omega)/(2r+1)}}{(1-2^{2\omega[1+2/r]-1})}\Bigg).$$

Therefore, the most slowly decaying term will be of order $T^{-2r/(2r+1)}$, which implies

$$\sum_{l=J_0+1}^{J_1}\sum_{m=0}^{2^l-1}\frac{\upsilon_l^2}{T^2} + \frac{b_{l,m}^2\upsilon_l^4}{T^2} = \mathcal{O}\Big(T^{-2r/(2r+1)}\Big). \tag{A.87}$$

Then, by lemma A.1.5 part (ii),

$$R_{2,11} = \mathcal{O}\Bigg(\sum_{l=J_0+1}^{J_1}\sum_{m=0}^{2^l-1}\Bigg[\frac{\upsilon_l^2}{T} + \frac{b_{l,m}^2\upsilon_l^4}{T^2} + \frac{\sigma^2}{T}\mathbb{I}\Big(|b_{l,m}| > \sqrt{T^{-1}\log(T)}\Big)\Bigg]\Bigg)$$
$$= \mathcal{O}\Big(T^{-2r/(2r+1)}\Big).$$

To determine the convergence of $R_{2,12}$ let

$$\gamma_l = \frac{2\sigma^2}{T}\log\left(\frac{\sqrt{T}\theta_l}{\upsilon_l\xi_0}\right), \tag{A.88}$$

as $T \to \infty \;\Rightarrow\; \gamma_l \to 0$. We have partition $R_{2,12}$ into two parts depending on whether ${h_{l,m}^*}^2$ is less than or greater than γ_l. If ${h_{l,m}^*}^2 < \gamma_l$ we find that as $T \to \infty$ then

$$\frac{\mathcal{A}_{l;T}(h_{l,m}^*)}{1+\mathcal{A}_{l;T}(h_{l,m}^*)} = 1.$$

Therefore, using (A.78), we can express $R_{2,121}$ is

$$R_{2,121} \;=\; \mathcal{O}\Bigg(\sum_{l=J_0+1}^{J_1}\sum_{m=0}^{2^l-1}\mathbb{E}\Big[{h_{l,m}^*}^2\mathbb{I}\Big({h_{l,m}^*}^2 < \gamma_l\Big)\Big]\mathbb{I}\Big(|b_{l,m}| > \sqrt{T^{-1}\log(T)}\Big)\Bigg)$$

As $T \to \infty$, we find$\mathbb{E}[{h^*_{l,m}}^2 \mathbb{I}({h^*_{l,m}}^2 < \gamma_l)] \to 0$ at a rate of $T^{-1}\log(T)$. Therefore,

$$R_{2,121} \;=\; \mathcal{O}\left(\sum_{l=J_0+1}^{J_1}\sum_{m=0}^{2^l-1}\frac{\log(T)}{T}\,\mathbb{I}\Big(|b_{l,m}| > \sqrt{T^{-1}\log(T)}\Big)\right).$$

Using Lemma A.1.5 part (ii) this can be expressed as

$$R_{2,121} = \mathcal{O}\Big(T^{-2r/(2r+1)}\log(T)^{1-p/2}\Big).$$

If ${h^*_{l,m}}^2 > \gamma_l$, then as $T \to \infty$.

$$\frac{\mathcal{A}_{l;T}(h^*_{l,m})}{1+\mathcal{A}_{l;T}(h^*_{l,m})} = \mathcal{A}_{l;T}(h^*_{l,m}).$$

Therefore, using A.1.1(i)a. and substituting in (A.70) for $\mathcal{A}_{l;T}(h^*_{l,m})$, implies $R_{2,122}$ is

$$R_{2,122} = \mathcal{O}\left(\sum_{l=J_0+1}^{J_1}\sum_{m=0}^{2^l-1}\mathbb{E}\left[{h^*_{l,m}}^2\frac{\theta_l^2 T}{\upsilon_l^2}\frac{e^{-T{h^*_{l,m}}^2/\sigma^2}}{\xi^2(\upsilon_l h^*_{l,m})}\mathbb{I}\Big({h^*_{l,m}}^2 > \gamma_l\Big)\right]\mathbb{I}\Big(|b_{l,m}| > \sqrt{T^{-1}\log(T)}\Big)\right).$$

By the condition in equation (5.35)

$$\frac{1}{\xi^2(\upsilon_l h^*_{l,m})}\exp\left\{-\frac{T{h^*_{l,m}}^2}{\sigma^2}\right\} = \mathcal{O}\left(\exp\left\{-\frac{{h^*_{l,m}}^2(T-\upsilon_l^2)}{\sigma^2}\right\}\right).$$

Also consider

$$\max_x\left[x^2\exp\left\{-\frac{x^2(T-\upsilon_l^2)}{\sigma^2}\right\}\mathbb{I}\big(x^2>\gamma_l\big)\right] = \gamma_l\exp\left\{-\frac{\gamma_l(T-\upsilon_l^2)}{\sigma^2}\right\}, \tag{A.89}$$

then by substituting in γ_l from equation (A.88), using the $\upsilon_l = 2^{\mu_2}$ from (5.43) and (5.45), we can simplify

- $\log(\upsilon_l^{-1}\theta_l\sqrt{T}) \sim \log(T)$ as $T \to \infty$
- $T^{-1}\upsilon_l^2\log(T) = o(1)$ as $T \to \infty$ as previous.
- $\Big(\upsilon_l^{-1}\theta_l\sqrt{T}\Big)^{\upsilon_l^2/T} = \exp\big\{T^{-1}\mathcal{C}\log(T)\upsilon_l^2\big\} \sim 1$ by condition (5.45)

equation (A.89) can be written as

$$\max_x \left[x^2 \exp\left\{ -\frac{x^2(T-\upsilon_l^2)}{\sigma^2} \right\} \mathbb{I}\big(x^2 > \gamma_l\big) \right] = \frac{2\sigma^2}{T} \log\left(\frac{\theta_l \sqrt{T}}{\upsilon_l \xi_0} \right) \left(\frac{\theta_l \sqrt{T}}{\upsilon_l \xi_0} \right)^{-2(T-\upsilon_l^2)/T}$$
$$= \mathcal{O}\left(\left[\frac{\upsilon_l}{\theta_l T} \right]^2 \log(T) \right). \tag{A.90}$$

So, using (A.90) and lemma A.1.5 (ii), we find

$$R_{2,122} = \mathcal{O}\left(\sum_{l=J_0+1}^{J_1} \sum_{m=0}^{2^l-1} \frac{\theta_l^2 T}{\upsilon_l^2} \left(\frac{\upsilon_l}{\theta_l T} \right)^2 \log(T) \mathbb{I}\Big(|b_{l,m}| > \sqrt{T^{-1}\log(T)} \Big) \right)$$
$$= \mathcal{O}\left(\sum_{l=J_0+1}^{J_1} \sum_{m=0}^{2^l-1} \frac{\log(T)}{T} \mathbb{I}\Big(|b_{l,m}| > \sqrt{T^{-1}\log(T)} \Big) \right)$$
$$= \mathcal{O}\left(T^{-2r/(2r+1)} \log(T)^{1-p/2} \right). \tag{A.91}$$

Note that for $r > r^*$ we have $2r(2r+1)^{-1} > 1 - p/2$ if $p \geq 2$, hence

$$R_{2,1} = \mathcal{O}\left([T^{-1}\log(T)]^{1-p/2} \right).$$

To show the convergence of the second component of R_2, by using the triangle inequality, consider

$$|\hat{b}_{l,m} - b_{l,m}| \leq \frac{1}{1 + \mathcal{A}_{l;T}(h^*_{l,m})} \left[\left| \frac{\mathcal{Q}_1(h^*_{l,m})}{\mathcal{Q}_0(h^*_{l,m})} - h^*_{l,m} \right| + |h^*_{l,m} - b_{l,m}| + |b_{l,m}| \right] + |b_{l,m}|$$
$$= \mathcal{O}\left(\frac{\upsilon_l}{T} + \frac{\upsilon_l^2 |h^*_{l,m}|}{T[1 + \mathcal{A}_{l;T}(h^*_{l,m})]} + \frac{|h^*_{l,m} - b_{l,m}|}{1 + \mathcal{A}_{l;T}(h^*_{l,m})} + |b_{l,m}| \right)$$
$$= \mathcal{O}\left(|b_{l,m}| + \frac{\upsilon_l}{T} + \frac{|h^*_{l,m} - b_{l,m}|}{1 + \mathcal{A}_{l;T}(h^*_{l,m})} \right).$$

Partition $R_{2,2}$ into three parts, then using lemma A.1.5 and the result of equation (A.87) the first two parts are

$$R_{2,21} = \mathcal{O}\left(\sum_{l=J_0+1}^{J_1} \sum_{m=0}^{2^l-1} b_{l,m}^2 \mathbb{I}\Big(\sqrt{T} |b_{l,m}| \leq \sqrt{\log(T)} \Big) \right) = \mathcal{O}\Big([T^{-1}\log(T)]^{2r/(2r+1)} \Big),$$
$$R_{2,22} = \mathcal{O}\left(\sum_{l=J_0+1}^{J_1} \sum_{m=0}^{2^l-1} \left[\frac{\upsilon_l}{T} \right]^2 \right) = \mathcal{O}\Big(T^{-2r/(2r+1)} \Big).$$

We define the thirs part as

$$R_{2,23} = \mathcal{O}\left(\sum_{l=J_0+1}^{J_1} \sum_{m=0}^{2^l-1} \mathbb{E}\left[\frac{(h^*_{l,m} - b_{l,m})^2}{(1+\mathcal{A}_{l;T}(h^*_{l,m}))^2} \right] \mathbb{I}\Big(\sqrt{T}|b_{l,m}| \leq \sqrt{\log(T)}\Big) \right),$$

and requires partitioning again into a further two parts

$$R_{2,231} = \mathcal{O}\left(\sum_{l=J_0+1}^{J_1} \sum_{m=0}^{2^l-1} \mathbb{E}\left[\frac{(h^*_{l,m} - b_{l,m})^2}{(1+\mathcal{A}_{l;T}(h^*_{l,m}))^2} \right] \mathbb{I}\Big(1 \leq \sqrt{T}|b_{l,m}| \leq \sqrt{\log(T)}\Big) \right)$$
$$R_{2,232} = \mathcal{O}\left(\sum_{l=J_0+1}^{J_1} \sum_{m=0}^{2^l-1} \mathbb{E}\left[\frac{(h^*_{l,m} - b_{l,m})^2}{(1+\mathcal{A}_{l;T}(h^*_{l,m}))^2} \right] \mathbb{I}\Big(\sqrt{T}|b_{l,m}| \leq 1\Big) \right).$$

First consider $R_{2,231}$, by lemma A.1.5 this can be written as

$$\begin{aligned} R_{2,231} &= \mathcal{O}\left(\sum_{l=J_0+1}^{J_1} \sum_{m=0}^{2^l-1} \frac{1}{T} \mathbb{I}\big(Tb^2_{l,m} > 1\big)\, \mathbb{I}\big(Tb^2_{l,m} \leq \log(T)\big) \right) \\ &= \mathcal{O}\left(\left[\frac{\log(T)}{T} \right]^{2r/(2r+1)} \right) \end{aligned}$$

by lemma A.1.5. For $R_{2,232}$, we further partition this into two sections, depending on whether $\sqrt{T}|h^*_{l,m} - b_{l,m}|$ is greater than $a\sigma\sqrt{2\log(T)}$ or not. it is useful to observe that

$$\begin{aligned} &\mathbb{E}\Big[(h^*_{l,m} - b_{l,m})^2\, \mathbb{I}\Big(\sqrt{T}|h^*_{l,m} - b_{l,m}| > a\sigma\sqrt{2\log(T)}\Big) \Big] \\ &\qquad = \frac{2\sigma^2}{T} \int_{a^2\log(T)}^{\infty} \sqrt{x} e^{-x} \mathrm{d}x \;\sim\; \frac{\sqrt{\log(T)}}{T^{1+a^2}} \end{aligned}$$

(Gradshteyn and Ryzhik, 1980, equation 8.357). Therefore, if $a^2 > (r-\omega)/\omega(2r+1)$ we can write the first part as

$$\begin{aligned} R_{2,2321} &= \mathcal{O}\left(\sum_{l=J_0+1}^{J_1} \sum_{m=0}^{2^l-1} \frac{\sqrt{\log(T)}}{T^{1+a^2}} \right) \\ &= \mathcal{O}\left(\sum_{l=J_0+1}^{J_1} 2^l\, \frac{\sqrt{\log(T)}}{T^{1+a^2}} \right) \\ &= \mathcal{O}\left(-(2T^{1/(2r+1)} - T^{r/\omega(2r+1)})\, \frac{\sqrt{\log(T)}}{T^{1+a^2}} \right) \\ &= \mathcal{O}\Big(T^{-2r/(2r+1)}\Big). \end{aligned}$$

Using the definition of (A.70), the last term is defined as

$$
\begin{aligned}
&R_{2,2322} \\
&= \mathcal{O}\left(\sum_{l=J_0+1}^{J_1} \sum_{m=0}^{2^l-1} \mathbb{E}\left[\frac{(h^*_{l,m} - b_{l,m})^2}{(1+\mathcal{A}_{l;T}(h^*_{l,m}))^2} \mathbb{I}\Big(\sqrt{T}|h^*_{l,m} - b_{l,m}| \leq a\sigma\sqrt{2\log(T)}\Big)\right] \right. \\
&\qquad \left. \mathbb{I}\Big(\sqrt{T}|b_{l,m}| \leq 1\Big)\right) \\
&= \mathcal{O}\left(\sum_{l=J_0+1}^{J_1} \sum_{m=0}^{2^l-1} \mathbb{E}\left[\frac{(h^*_{l,m} - b_{l,m})^2 v_l^2}{\theta_l^2 T} \exp\left\{ \frac{T {h^*_{l,m}}^2}{\sigma^2} \right\} \mathbb{I}\Big(|h^*_{l,m} - b_{l,m}| \leq \frac{a\sigma}{\sqrt{T}}\sqrt{2\log(T)}\Big)\right] \right. \\
&\qquad \left. \mathbb{I}\Big(\sqrt{T}|b_{l,m}| \leq 1\Big)\right).
\end{aligned}
$$

As $\sqrt{T}h^*_{l,m} \leq a\sigma\sqrt{2\log(T)} + 1$, we find that this expectation is bounded by

$$
\frac{\sigma^2}{T}\exp\left\{ \frac{(a\sigma\sqrt{2\log(T)}+1)^2}{\sigma^2} \right\} \int_0^{a^2\log(T)} e^{-x}\sqrt{x}\mathrm{d}x = \mathcal{O}\left(T^{2a^2-1}\exp\left\{ 2a\sqrt{\frac{2\log(T)}{\sigma}} \right\} \right)
$$

Therefore, provided that

$$
\frac{1}{\theta_l}\exp\left\{ a\sqrt{2\log(T)} + \frac{1}{\sigma} \right\} \;=\; \mathcal{O}(1),
$$

which is true when

$$
a^2 = \left(\frac{1}{p} - \frac{1}{2} + \varepsilon\right)\left(\omega\left[r + \frac{1}{2} - \frac{1}{p}\right]\right)^{-1}
$$

by the condition in equation (5.45).

Fine Scales: $J_1 < l \leq J_2$

For fine scales we also find that $\lambda_l = 0$ because of the swift decrease in the likelihood weights as the wavelet increases, and since $|\hat{b}_{l,m}| \leq |\mathcal{Q}_1(h^*_{l,m})/\mathcal{Q}_0(h^*_{l,m})|$ we find that

$$
R_3 = \mathcal{O}\left(\sum_{l=J_1+1}^{J_2} \sum_{m=0}^{2^l-1} \mathbb{E}\left[\left(b^2_{l,m} + \frac{\mathcal{Q}_1(h^*_{l,m})}{\mathcal{Q}_0(h^*_{l,m})} \right)^2 \right] \right) = R_{3,1} + R_{3,2},
$$

where

$$
\begin{aligned}
R_{3,1} &= \sum_{l=J_1+1}^{J_2} \sum_{m=0}^{2^l-1} b^2_{l,m} \\
R_{3,2} &= \mathcal{O}\left(\sum_{l=J_1+1}^{J_2} \sum_{m=0}^{2^l-1} \mathbb{E}\left[\left(\frac{\mathcal{Q}_1(h^*_{l,m})}{\mathcal{Q}_0(h^*_{l,m})} \right)^2 \right] \right)
\end{aligned}
$$

By substituting in equation (5.32), by using 2ω from equation (A.22) and as we have a geometric series, we find that

$$\begin{aligned} R_{3,1} = \sum_{l=J_1+1}^{J_2} 2^{-2\omega l} & \\ &= \frac{2^{-2\omega(2rJ[2\omega(2r+1)]^{-1}+1)} - 2^{-2\omega J(1/2+r[2\omega(2r+1)]^{-1})}}{1-2^{-2\omega}} \\ &= \frac{2^{-2\omega}T^{-2r/(2r+1)} - T^{-(\omega+r/(2r+1))}}{1-2^{-2\omega}} \\ &= \mathcal{O}\Big(T^{-2r/(2r+1)}\Big), \end{aligned} \tag{A.92}$$

because of the choice of J_2.

We need to consider two cases for $R_{3,2}$, when the condition in equation (5.38) is valid and when it is not. Firstly if (5.38) holds, then by Lemma A.1.1 (ii) part b considering whether or not $\sqrt{T}h^*_{l,m}$ is bounded then

$$\left|\frac{\mathcal{Q}_1(h^*_{l,m})}{\mathcal{Q}_0(h^*_{l,m})}\right| = \mathcal{O}\left(\frac{\sqrt{T}}{\upsilon_l^2}\right) + \mathcal{O}\left(\frac{T|h^*_{l,m}|}{\upsilon_l^2}\right), \tag{A.93}$$

Then, by substiting (A.93) into $R_{3,2}$, adding and subtracting $b_{l,m}$ to the second part of (A.93), we find that

$$\begin{aligned} R_{3,2} &= \mathcal{O}\left(\sum_{l=J_1+1}^{J_2}\sum_{m=0}^{2^l-1}\mathbb{E}\left[\left(\frac{\sqrt{T}}{\upsilon_l^2}\right)^2\right] + \mathbb{E}\left[\left(\frac{T(h^*_{l,m}-b_{l,m})}{\upsilon_l^2}\right)^2\right] + \mathbb{E}\left[\left(\frac{Tb_{l,m}}{\upsilon_l^2}\right)^2\right]\right) \\ &= \mathcal{O}\left(\sum_{l=J_1+1}^{J_2}\sum_{m=0}^{2^l-1}\frac{T^2\mathbb{E}\Big[(h^*_{l,m}-b_{l,m})^2\Big]}{\upsilon_l^4} + \left(\frac{b_{l,m}T}{\upsilon_l^2}\right)^2 + \frac{T}{\upsilon_l^4}\right) \end{aligned}$$

From lemma A.1.3 (i) we know that $\mathbb{E}\Big[(h^*_{l,m}-b_{l,m})^2\Big] = \mathcal{O}\big(T^{-1}\big)$. Then by substituting in eqution (5.32) and $\upsilon_l = 2^{\mu_3(J-l)}$, we find that

$$\begin{aligned} R_{3,2} = & \ \mathcal{O}\left(\sum_{l=J_1+1}^{J_2}\sum_{m=0}^{2^l-1}\frac{T}{\upsilon_l^4} + \left(\frac{b_{l,m}T}{\upsilon_l^2}\right)^2 + \frac{T}{\upsilon_l^4}\right) \\ = & \ \mathcal{O}\left(\sum_{l=J_1+1}^{J_2} 2^{1+(1-4\mu_3)(J-l)} + T^2 2^{-(2\omega+4\mu_3)(J-l)}\right). \end{aligned}$$

Using the representation of a geometric series this is

$$R_{3,2} = \mathcal{O}\Bigg(2\frac{2^{(1-4\mu_3)(J_1+1)} - 2^{(1-4\mu_3)J_2}}{1-2^{(1-4\mu_3)}} + T^2\frac{2^{-2(\omega+2\mu_3)(J_1+1)} - 2^{-2(\omega+2\mu_3)J_2}}{1-2^{-2(\omega+2\mu_3)}}\Bigg)$$

Substituting in J_1 and J_2 from equations (5.40) and (5.41), respectively yields

$$\begin{aligned}R_{3,2} &= \mathcal{O}\Bigg(2\frac{2^{(1-4\mu_3)(rJ[(2r+1)\omega]^{-1}+1)} - 2^{J(1-4\mu_3)(1/2+r[(2r+1)2\omega]^{-1})}}{1-2^{1-4\mu_3}} \\ &\qquad + T^2\frac{2^{-2(\omega+2\mu_3)(rJ[(2r+1)\omega]^{-1}+1)} - 2^{-2(\omega+2\mu_3)(J(1/2+r[(2r+1)2\omega]^{-1}))}}{1-2^{-2(\omega+2\mu_3)}}\Bigg) \\ &= \mathcal{O}\Bigg(2\frac{2^{1-4\mu_3}T^{r(1-4\mu_3)/(2r+1)\omega} - T^{(1-4\mu_3)(1/2+r/(2r+1)2\omega)}}{1-2^{1-4\mu_3}} \\ &\qquad + T^2\frac{2^{-2(\omega+2\mu_3)}T^{-2r(\omega+2\mu_3)/(2r+1)\omega} - T^{-(\omega+2\mu_3)(1+r/(2r+1)\omega)}}{1-2^{-2(\omega+2\mu_3)}}\Bigg).\end{aligned}$$

Then by substituting in $\mu_3 = r + 1/2$, we can write this as

$$R_{3,2} = \mathcal{O}\Bigg(2\frac{2^{-(4r+1)}T^{-r(4r+1)/(2r+1)\omega} - T^{-(2r+1/2)(1+r/(2r+1)\omega)}}{1-2^{-(4r+1)}} + \frac{2^{-2(\omega+2r+1)}T^{-2r/(2r+1)+(1-2/p^*)/\omega} - T^{2-(\omega+2r+1)(1+r/(2r+1)\omega)}}{1-2^{-2(\omega+2r+1)}}\Bigg)$$

All of these values are less than $T^{-2r/(2r+1)}$ because $1 \leq p* \leq 2$ and $r > r^*$, this gives

$$R_{3,2} = \mathcal{O}\Big(T^{-2r/(2r+1)}\Big).$$

When equation (5.38) does not hold then separate $R_{3,2}$ into two sections:

$$R_{3,21} = \mathcal{O}\Bigg(\sum_{l=J_1+1}^{J_2}\mathbb{E}\left[\left(\frac{\mathcal{Q}_1(h^*_{l,m})}{\mathcal{Q}_0(h^*_{l,m})}\right)^2\right]\mathbb{I}\Big(\sqrt{T}|b_{l,m}| \to \infty\Big)\Bigg) \tag{A.94}$$

$$R_{3,22} = \mathcal{O}\Bigg(\sum_{l=J_1+1}^{J_2}\mathbb{E}\left[\left(\frac{\mathcal{Q}_1(h^*_{l,m})}{\mathcal{Q}_0(h^*_{l,m})}\right)^2\right]\mathbb{I}\Big(\sqrt{T}b_{l,m} \leq M\Big)\Bigg), \tag{A.95}$$

where M is some constant, implying that $\sqrt{T}|b_{l,m}|$ is bounded by M.

From lemma A.1.1, we know that if $\sqrt{T}|h^*_{l,m}|$ is also bounded, then $|\mathcal{Q}_1(h^*_{l,m})/\mathcal{Q}_0(h^*_{l,m})|$ $= \mathcal{O}(\sqrt{T}/\upsilon_l^2)$, and conversely if $\sqrt{T}|h^*_{l,m}| \to \infty$ then $|\mathcal{Q}_1(h^*_{l,m})/\mathcal{Q}_0(h^*_{l,m})| = \mathcal{O}\Big(h^*_{l,m}\Big)$ (see lemma A.1.4). Recall (5.32) and (5.43) and substitute these in for $\sum b^2_{l,m}$ and υ_l

(respectively), we can write $R_{3,21}$ as

$$
\begin{aligned}
R_{3,21} &= \mathcal{O}\Bigg(\sum_{l=J_1+1}^{J_2} \sum_{m=0}^{2^l-1} \mathbb{E}\big[(h^*_{l,m} - b_{l,m})^2\big]\, \mathbb{I}\Big(\sqrt{T}|b_{l,m}| \to \infty\Big) \;+\; b^2_{l,m} \;+\; \frac{T}{\upsilon_l^4} \Bigg) \\
&= \mathcal{O}\Bigg(\sum_{l=J_1+1}^{J_2} 2^{-2\omega l} \;+\; T2^{(1-4\mu_3)l} \;+\; \sum_{m=0}^{2^l-1} \mathbb{E}\big[(h^*_{l,m} - b_{l,m})^2\big]\, \mathbb{I}\Big(\sqrt{T}|b_{l,m}| \to \infty\Big) \Bigg) \\
&= \mathcal{O}\Bigg(\frac{2^{-2\omega(J_1+1)} - 2^{-2\omega J_2}}{1 - 2^{-2\omega}} \;+\; T\frac{2^{(1-4\mu_3)(J_1+1)} - 2^{(1-4\mu_3)J_2}}{1 - 2^{(1-4\mu_3)}} \\
&\qquad + \sum_{l=J_1+1}^{J_2} \sum_{m=0}^{2^l-1} \mathbb{E}\big[(h^*_{l,m} - b_{l,m})^2\big]\, \mathbb{I}\Big(\sqrt{T}|b_{l,m}| \to \infty\Big) \Bigg)
\end{aligned}
$$

As $\mathbb{E}\Big[(h^*_{l,m} - b_{l,m})^2\Big] = \mathcal{O}(T^{-1})$, we can re-write this as

$$
\mathcal{O}(T^{-1}) = \mathcal{O}\left(T^{-1}\frac{Tb^2_{l,m}}{Tb^2_{l,m}} \right) = \mathcal{O}(b^2_{l,m})
$$

because $\sqrt{T}|b_{l,m}|$ is bounded, this implies that

$$
\begin{aligned}
R_{3,21} &= \mathcal{O}\left(\frac{2^{(1-2\omega)(J_1+1)} - 2^{(1-2\omega)J_2}}{1 - 2^{1-2\omega}} \;+\; T\frac{2^{(1-4\mu_3)(J_1+1)} - 2^{(1-4\mu_3)J_2}}{1 - 2^{1-4\mu_3}} \right) \\
&= \mathcal{O}\Bigg(\frac{2^{-2(r-1/p^*)}T^{-2(r-1/p^*)/(2r+1)\omega} - T^{-2(r-1/p^*)(1+r/(2r+1)\omega)}}{1 - 2^{2/p^*-2r}} \\
&\qquad\qquad + \frac{2^{-4r-1}T^{1-r(4r+1)/(2r+1)\omega} - T^{1-(2r+1/2)(1+r/(2r+1)\omega)}}{1 - 2^{-2r-1}} \Bigg) \\
&= \mathcal{O}\Big(T^{-2r/(2r+1)} \Big),
\end{aligned}
$$

as this is the term which is decaying to zero the slowest. We split $R_{3,22}$ into the summation of $R_{3,221}$ and $R_{3,222}$, where

$$
\begin{aligned}
R_{3,221} &= \sum_{l=J_1+1}^{J_2} \sum_{m=0}^{2^l-1} \mathbb{E}\left[\left(\frac{\mathcal{Q}_1(h^*_{l,m})}{\mathcal{Q}_0(h^*_{l,m})} \mathbb{I}(\sqrt{T}|h^*_{l,m} - b_{l,m}| > a\sqrt{\log(T)}) \right)^2 \right] \mathbb{I}(\sqrt{T}b_{l,m} \leq M), \\
R_{3,222} &= \sum_{l=J_1+1}^{J_2} \sum_{m=0}^{2^l-1} \mathbb{E}\left[\left(\frac{\mathcal{Q}_1(h^*_{l,m})}{\mathcal{Q}_0(h^*_{l,m})} \mathbb{I}(\sqrt{T}|h^*_{l,m} - b_{l,m}| \leq a\sqrt{\log(T)}) \right)^2 \right] \mathbb{I}(\sqrt{T}b_{l,m} \leq M),
\end{aligned}
$$

where $a^2 \geq 4\sigma^2$. Using lemma A.1.3 parts (i) and (ii), equations (5.32) and (5.43) (similarly

to $R_{3,21}$), we can write $R_{3,221}$ as

$$\begin{aligned}
R_{3,221} &= \mathcal{O}\Bigg(\sum_{l=J_1+1}^{J_2} \sum_{m=0}^{2^l-1} b_{l,m}^2 + \frac{T}{\upsilon_l^4} \\
&\qquad + \sqrt{\mathbb{E}\Big[(h_{l,m}^* - b_{l,m})^4\Big]\, \mathbb{P}\Big(\sqrt{T}|h_{l,m}^* - b_{l,m}| > a\sqrt{\log(T)}\Big)} \Bigg) \\
&= \mathcal{O}\left(\sum_{l=J_1+1}^{J_2} \sum_{m=0}^{2^l-1} b_{l,m}^2 + \frac{T}{\upsilon_l^4} + \sqrt{\mathcal{O}(T^{-2}) o(T^{-a^2/2\sigma^2})} \right) \\
&= \mathcal{O}\left(\sum_{l=J_1+1}^{J_2} \sum_{m=0}^{2^l-1} b_{l,m}^2 + \frac{T}{\upsilon_l^4} + \frac{1}{T} \right) \\
&= \mathcal{O}\Bigg(\frac{2^{-2\omega} T^{-2r/(2r+1)} - T^{-(\omega + r/(2r+1))}}{1 - 2^{-2\omega}} + -2T^{r/(2r+1)\omega - 1} - T^{-1/2 + r/2(2r+1)\omega} \\
&\qquad + \frac{2^{-4r-1} T^{-r(4r+1)/(2r+1)\omega + 1} - T^{(-2r-1/2)(1+r/(2r+1)\omega)+1}}{1 - 2^{-4r-1}} \Bigg)
\end{aligned}$$

As $J_2 < J$ from (5.41), we find that $R_{3,221} = \mathcal{O}\big(T^{-2r/(2r+1)}\big)$.

To derive an asymptotic bound for $R_{3,222}$ consider the identity functions of $R_{3,222}$ and that for large T, $\sqrt{T}|b_{l,m}|$ is bounded for some $M_1 > 0$, then

$$\begin{aligned}
&\mathbb{I}(\sqrt{T}|b_{l,m}| \leq M)\ \mathbb{I}(\sqrt{T}|h_{l,m}^* - b_{l,m}| \leq a\sqrt{\log(T)}) \\
&\leq \mathbb{I}(\sqrt{T}|b_{l,m}| \leq M)\ \mathbb{I}(\sqrt{T}|h_{l,m}^*| \leq 2a\sqrt{\log(T)}) \\
&\leq \mathbb{I}(\sqrt{T}|b_{l,m}| \leq M) \left[\mathbb{I}(\sqrt{T}|h_{l,m}^*| \leq 2a\sqrt{\log(T)})\ \mathbb{I}\left(\frac{\sqrt{T\log(T)}}{\upsilon_l} \to 0 \right) \right. \\
&\qquad \left. + \mathbb{I}\left(\frac{\upsilon_l}{\sqrt{T\log(T)}} \leq M_1 \right) \right] \\
&\leq \mathbb{I}(\sqrt{T}|b_{l,m}| \leq M) \left[\mathbb{I}\left(\frac{T|h_{l,m}^*|}{\upsilon_l} \to 0 \right) + \mathbb{I}\left(\frac{2^{(J-l)(2r+1)}}{T\log(T)} \leq M_1 \right) \right]. \qquad \text{(A.96)}
\end{aligned}$$

Using lemma A.1.1 and as $\sqrt{T}\upsilon_l^{-1}(\sqrt{T}|h_{l,m}^*|) \to 0$, this implies that

$$\mathbb{E}\left[\left(\frac{\mathcal{Q}_1(h_{l,m}^*)}{\mathcal{Q}_0(h_{l,m}^*)} \right)^2 \right] = \mathcal{O}\left(\frac{T}{\upsilon_l^4} (\sqrt{T}|h_{l,m}^*|)^2 + b_{l,m}^2 \right) = \mathcal{O}\left(\frac{T}{\upsilon_l^4} + b_{l,m}^2 \right),$$

because $\mathbb{E}[T{h_{l,m}^*}^2] \leq 2T(\mathbb{E}[(h_{l,m}^* - b_{l,m})^2] + b_{l,m}^2) = \mathcal{O}(1)$. Therefore, as shown in previous calculations, this implies that the part of $R_{3,222}$ which corresponds to the first term in equation (A.96) is $\mathcal{O}(T^{-2r/(2r+1)})$.

Using lemma A.1.4 we know that $\mathbb{E}[\mathcal{Q}_1(h_{l,m}^*)/\mathcal{Q}_0(h_{l,m}^*)] = \mathcal{O}(\mathbb{E}[(h_{l,m}^* - b_{l,m})^2] + b_{l,m}^2)$.

Then, by reorganising second term in equation (A.96), the portion of $R_{3,222}$ related to this is

$$\begin{aligned}
&\mathcal{O}\left(\sum_{l=J_1+1}^{J_2}\sum_{m=0}^{2^l-1}(T^{-1}+b_{l,m}^2)\mathbb{I}\Big(2^l=\mathcal{O}([T\log(T)]^{1/(2r+1)})\Big)\right)\\
&=\mathcal{O}\left(\sum_{l=J_1+1}^{J_2}\left(\frac{2^j}{T}+2^{-2\omega l}\right)\mathbb{I}\Big(2^l=\mathcal{O}([T\log(T)]^{1/(2r+1)})\Big)\right)\\
&=\mathcal{O}\Big(T^{-2r/(2r+1)}[\log(T)]^{1/(2r+1)}\Big)\,.
\end{aligned}$$

So, using the same process as before to re-write the double summation,

$$R_{3,222}=\mathcal{O}\Big(T^{-2r/(2r+1)}[\log(T)]^{1/(2r+1)}\Big)\,,$$

which implies that when equation (5.38) is invalid $R_3=\mathcal{O}\big(T^{-2r/(2r+1)}[\log(T)]^{1/(2r+1)}\big)$.

Finest Scales: $J_2<l<J$

For the finest scales we find that $\lambda_l>0$, therefore $\hat{b}\neq$ equation (A.71). If we let

$$\begin{aligned}
\mathcal{B}_{l;T}(h^*)&=\frac{1}{(1-\lambda_l)\,\mathcal{Q}_0(h^*_{l,m})}\left[\theta_l\sqrt{T}\,\zeta_l(\sqrt{T}h^*_{l,m})\;+\;\lambda_l\,\mathcal{Q}_0^*(h^*_{l,m})\right]\\
\mathcal{B}^*_{l;T}(h^*)&=\frac{1}{\lambda_l\,\mathcal{Q}_0^*(h^*_{l,m})}\left[\theta_l\sqrt{T}\,\zeta_l(\sqrt{T}h^*_{l,m})\;+\;(1-\lambda_l)\,\mathcal{Q}_0(h^*_{l,m})\right].
\end{aligned}$$

Therefore

$$\begin{aligned}
\hat{b}_{l,m}&=\frac{(1-\lambda_l)\,\mathcal{Q}_1(h^*_{l,m})\;+\;\lambda_l\,\mathcal{Q}_1^*(h^*_{l,m})}{\theta_l\sqrt{T}\,\zeta_l(\sqrt{T}h^*_{l,m})\;+\;(1-\lambda_l)\,\mathcal{Q}_0(h^*_{l,m})+\lambda_l\,\mathcal{Q}_0^*(h^*_{l,m})}\\
&=\frac{\mathcal{Q}_1(h^*_{l,m})}{\mathcal{Q}_0(h^*_{l,m})}[\mathcal{B}_{l;T}(h^*_{l,m})+1]^{-1}+\frac{\mathcal{Q}_1^*(h^*_{l,m})}{\mathcal{Q}_0^*(h^*_{l,m})}[\mathcal{B}^*_{l;T}(h^*_{l,m})+1]^{-1}. \qquad \text{(A.97)}
\end{aligned}$$

As we know that $\zeta_l(\cdot)$, $\mathcal{Q}_0(\cdot)$ and $\mathcal{Q}_0^*(\cdot)$ are all combinations of pdfs, therefore they are all positive functions. This implies that

$$\frac{\mathcal{B}_{l;T}(h^*)}{\mathcal{B}_{l;T}(h^*)+1}\qquad\text{and}\qquad\frac{\mathcal{B}^*_{l;T}(h^*)}{\mathcal{B}^*_{l;T}(h^*)+1}\qquad\text{are both }\leq 1.$$

Then, using (A.97) we can express

$$\frac{\mathcal{Q}_1(h^*_{l,m})}{\mathcal{Q}_0(h^*_{l,m})} + \frac{\mathcal{Q}^*_1(h^*_{l,m})}{\mathcal{Q}^*_0(h^*_{l,m})} - \hat{b}_{l,m} = \frac{\mathcal{Q}_1(h^*_{l,m})}{\mathcal{Q}_0(h^*_{l,m})}\left(\frac{\mathcal{B}_{l;T}(h^*)}{\mathcal{B}_{l;T}(h^*)+1}\right) + \frac{\mathcal{Q}^*_1(h^*_{l,m})}{\mathcal{Q}^*_0(h^*_{l,m})}\left(\frac{\mathcal{B}^*_{l;T}(h^*)}{\mathcal{B}^*_{l;T}(h^*)+1}\right)$$
$$\leq \frac{\mathcal{Q}_1(h^*_{l,m})}{\mathcal{Q}_0(h^*_{l,m})} + \frac{\mathcal{Q}^*_1(h^*_{l,m})}{\mathcal{Q}^*_0(h^*_{l,m})}.$$

Consider that

$$\mathbb{E}\Big[(\hat{b}_{l,m} - b_{l,m})^2\Big] \leq \mathbb{E}\left[\left(\frac{\mathcal{Q}_1(h^*_{l,m})}{\mathcal{Q}_0(h^*_{l,m})}\right)^2 + \left(\frac{\mathcal{Q}^*_1(h^*_{l,m})}{\mathcal{Q}^*_0(h^*_{l,m})}\right)^2 + b^2_{l,m}\right].$$

Then we split R_4 into six sections and $R_4 = \sum_{i=1}^{6} R_{4,i}$. We shall start with $R_{4,1}$, and from (5.32) this can be expressed as

$$R_{4,1} = \sum_{l=J_2+1}^{J-1}\sum_{m=0}^{2^l-1} b^2_{l,m} = \mathcal{O}\left(\sum_{l=J_1+1}^{J_2} 2^{-2\omega l}\right) = \mathcal{O}\left(\frac{2^{-2\omega(J_2+1)} - 2^{-2\omega(J-1)}}{1-2^{-2\omega}}\right)$$
$$= \mathcal{O}\left(\frac{2^{-2\omega}T^{-(\omega(2r+1)+r)/(2r+1)} - 2^{-2\omega}T^{-2\omega}}{1-2^{-2\omega}}\right) = \mathcal{O}(T^{-2r/(2r+1)})$$

From lemma A.1.4 we know that if $\sqrt{T}|h^*_{l,m}| \to \infty$, we find that $\mathcal{Q}_1(h^*_{l,m})/\mathcal{Q}_0(h^*_{l,m})$ and $\mathcal{Q}^*_1(h^*_{l,m})/\mathcal{Q}^*_0(h^*_{l,m})$ will both equal $\mathcal{O}(|h^*_{l,m}|)$. Hence

$$R_{4,2} = \sum_{l=J_2+1}^{J-1}\sum_{m=0}^{2^l-1}\mathbb{E}\left[\left(\frac{\mathcal{Q}_1(h^*_{l,m})}{\mathcal{Q}_0(h^*_{l,m})}\right)^2 + \left(\frac{\mathcal{Q}^*_1(h^*_{l,m})}{\mathcal{Q}^*_0(h^*_{l,m})}\right)^2\right] \mathbb{I}(\sqrt{T}|b_{l,m}| \to \infty)$$
$$= \mathcal{O}\left(\sum_{l=J_2+1}^{J-1}\sum_{m=0}^{2^l-1}\mathbb{E}[{h^*_{l,m}}^2]\mathbb{I}(\sqrt{T}|b_{l,m}| \to \infty)\right)$$
$$= \mathcal{O}\left(\sum_{l=J_2+1}^{J-1}\sum_{m=0}^{2^l-1}\mathbb{E}[(h^*_{l,m} - b_{l,m})^2]\mathbb{I}(\sqrt{T}|b_{l,m}| \to \infty) + \frac{T}{v_l^4} + b^2_{l,m}\right).$$

This is similar to $R_{3,21}$, so we can re-write this as

$$\begin{aligned}
R_{4,2} &= \mathcal{O}\left(\sum_{l=J_2+1}^{J-1}\sum_{m=0}^{2^l-1}\frac{T}{\upsilon_l^4}+2b_{l,m}^2\right)\\
&= \mathcal{O}\left(\sum_{l=J_2+1}^{J-1} T2^{(1-4\mu_3)l}+2^{1-2\omega l}\right)\\
&= \mathcal{O}\left(T\frac{2^{-(4r+1)(J_2+1)}-2^{-(4r+1)(J-1)}}{1-2^{-4r-1}}+2\frac{2^{-2\omega(J_2+1)}-2^{-2\omega(J-1)}}{1-2^{-2\omega}}\right)\\
&= \mathcal{O}\left(2^{-(4r+1)}\frac{T^{-(4r+1)(1/2+r/2\omega(2r+1))+1}-2^{4r+1}T^{-(4r+1)+1}}{1-2^{-4r-1}}\right.\\
&\qquad\left.+\frac{2^{-2\omega}T^{-(\omega+r/(2r+1))}-2^{2\omega}T^{-2\omega}}{1-2^{-2\omega}}\right)\\
&= \mathcal{O}\Big(T^{-2r/(2r+1)}\Big).
\end{aligned}$$

The next part is defined as

$$R_{4,3}=\sum_{l=J_2+1}^{J-1}\sum_{m=0}^{2^l-1}\mathbb{E}\left[\left(\frac{\mathcal{Q}_1(h_{l,m}^*)}{\mathcal{Q}_0(h_{l,m}^*)}\right)^2\ \mathbb{I}\Big(\sqrt{T}|h_{l,m}^*-b_{l,m}|>2\log(T)\Big)\right].$$

Using lemma A.1.3 part (iii) we can repeat the same process which was used to solve $R_{3,221}$ as

$$\begin{aligned}
&R_{4,3}\\
&= \mathcal{O}\left(\sum_{l=J_2+1}^{J-1}\sum_{m=0}^{2^l-1}\sqrt{\mathbb{E}\left[\Big(h_{l,m}^*-b_{l,m}\Big)^4\right]\mathbb{P}\Big(|h_{l,m}^*-b_{l,m}|>2\log(T)/\sqrt{T}\Big)}+b_{l,m}^2+\frac{T}{\upsilon_l^4}\right)\\
&= \mathcal{O}\left(\sum_{l=J_2+1}^{J-1}\sum_{m=0}^{2^l-1}T^{-1}+b_{l,m}^2+\frac{T}{\upsilon_l^4}\right)\\
&= \mathcal{O}\left(\sum_{l=J_2+1}^{J-1}2^lT^{-1}+2^{-2\omega l}+T2^{(1-4\mu_3)l}\right)\\
&= \mathcal{O}\left(2T^{(r/\omega(2r+1)-1)/2}-2+\frac{2^{-2\omega}T^{-(\omega+r/(2r+1))}-2^{2\omega}T^{-2\omega}}{1-2^{-2\omega}}\right.\\
&\qquad\left.+T\frac{2^{-(4r+1)}T^{-(4r+1)(1/2+r/2\omega(2r+1))}-2^{-(4r+1)}T^{-(4r+1)}}{1-2^{-(4r+1)}}\right)\\
&= \mathcal{O}\Big(T^{-2r/(2r+1)}\Big).
\end{aligned}$$

The fourth part is

$$R_{4,4} = \sum_{l=J_2+1}^{J-1} \sum_{m=0}^{2^l-1} \mathbb{E}\left[\left(\frac{\mathcal{Q}_1(h^*_{l,m})}{\mathcal{Q}_0(h^*_{l,m})} \right)^2 \mathbb{I}\Big(\sqrt{T}|h^*_{l,m} - b_{l,m}| \leq 2\log(T) \Big) \right] \mathbb{I}\Big(\sqrt{T}|b_{l,m}| \leq M \Big).$$

An upper bound can be found by considering

$$\mathbb{I}\Big(\sqrt{T}|h^*_{l,m} - b_{l,m}| < 2\log(T) \Big) \leq \mathbb{I}\Big(\sqrt{T}|h^*_{l,m}| \leq 2\log(T) + \sqrt{T}|b_{l,m}| \Big).$$

Therefore, using lemma A.1.1 part (ii) b, we find that

$$\begin{aligned}
R_{4,4} &= \sum_{l=J_2+1}^{J-1} \sum_{m=0}^{2^l-1} \mathbb{E}\left[\left(\frac{\mathcal{Q}_1(h^*_{l,m})}{\mathcal{Q}_0(h^*_{l,m})} \right)^2 \mathbb{I}\Big(\sqrt{T}|h^*_{l,m}| \leq 2\log(T) + M \Big) \right] \\
&= \mathcal{O}\left(\sum_{l=J_2+1}^{J-1} \sum_{m=0}^{2^l-1} \sqrt{ \mathbb{E}\left[\left(\frac{\mathcal{Q}_1(h^*_{l,m})}{\mathcal{Q}_0(h^*_{l,m})} \right)^4 \mathbb{P}\Big(\sqrt{T}|h^*_{l,m}| \leq 2\log(T) + M \Big)^2 \right] } \right) \\
&= \mathcal{O}\left(\sum_{l=J_2+1}^{J-1} \sum_{m=0}^{2^l-1} \frac{T^2}{\upsilon_l^4} \frac{\log^2(T)+1}{T} \right) \\
&= \mathcal{O}\left(\sum_{l=J_2+1}^{J-1} \sum_{m=0}^{2^l-1} \frac{T\log^2(T)}{\upsilon_l^4} + \frac{T\log^2(T)}{\upsilon_l^4} \right) \\
&= \mathcal{O}(\upsilon_l^{-2}) \\
&= o(T^{-1}).
\end{aligned}$$

The fifth and sixth part of R_4 focus on $\mathcal{Q}^*_i(\cdot)$ and these are defined as

$$R_{4,5} = \sum_{l=J_2+1}^{J-1} \sum_{m=0}^{2^l-1} \mathbb{E}\left[\left(\frac{\mathcal{Q}^*_1(h^*_{l,m})}{\mathcal{Q}^*_0(h^*_{l,m})} \right)^2 \mathbb{I}\Big(\sqrt{T}(h^*_{l,m} - b_{l,m}) > \frac{r}{4}\log(T) \Big) \right]$$

$$R_{4,6} = \sum_{l=J_2+1}^{J-1} \sum_{m=0}^{2^l-1} \mathbb{E}\left[\left(\frac{\mathcal{Q}^*_1(h^*_{l,m})}{\mathcal{Q}^*_0(h^*_{l,m})} \right)^2 \mathbb{I}\Big(\sqrt{T}(h^*_{l,m} - b_{l,m}) \leq \frac{r}{4}\log(T) \Big) \right] \mathbb{I}\Big(\sqrt{T}|b_{l,m}| \leq M \Big).$$

For $R_{4,5}$ we follow a similar procedure to $R_{3,221}$, however $a\log(T)$ is replaced with $\frac{r}{4}\log(T)$ and we use lemma A.1.3 part (iv). This results in

$$\begin{aligned}
R_{4,5} &= \mathcal{O}\left(\sum_{l=J_2+1}^{J-1}\sum_{m=0}^{2^l-1}\sqrt{\mathbb{E}\left[\left(h^*_{l,m}-b_{l,m}\right)^4\right]\ \mathbb{P}\Big(|h^*_{l,m}-b_{l,m}|>\frac{r}{4}\log(T)\Big)}+b^2_{l,m}+\frac{T}{\upsilon_l^4}\right)\\
&= \mathcal{O}\left(\sum_{l=J_2+1}^{J-1}\sum_{m=0}^{2^l-1}\sqrt{\mathcal{O}(T^{-2})\ o(T^{-r/4})}+b^2_{l,m}+\frac{T}{\upsilon_l^4}\right)\\
&= \mathcal{O}\left(\sum_{l=J_2+1}^{J-1}\sum_{m=0}^{2^l-1}T^{-1}+b^2_{l,m}+\frac{T}{\upsilon_l^4}\right).
\end{aligned}$$

This is the same as $R_{4,3}$, which implies that $R_{4,5}=\mathcal{O}\big(T^{-2r/(2r+1)}\big)$.

For the sixth part consider that when

$$\sqrt{T}(h^*_{l,m}-b_{l,m})<\frac{\log(T)}{4r+1}\qquad\Rightarrow\qquad\frac{\sqrt{T}}{\upsilon_l}\exp\{2\sqrt{T}h^*_{l,m}\}=o(1).$$

This implies that we can use lemma A.1.2 and

$$\begin{aligned}
&\left|\frac{\mathcal{Q}^*_1(h^*_{l,m})}{\mathcal{Q}^*_0(h^*_{l,m})}\right|\\
&= \frac{\sqrt{T}f_E(\sqrt{T}h^*_{l,m})(1-\gamma^*e^{2\sqrt{T}h^*_{l,m}})\int_{-\infty}^{\infty}x^2\xi(x)\mathrm{d}x\left\{1+\mathcal{O}\Big(T\upsilon_l^{-2}\left[1-e^{2\sqrt{T}h^*_{l,m}}\right]\Big)\right\}}{\upsilon_l^2 f_E(\sqrt{T}h^*_{l,m})\left\{1+\mathcal{O}\Big(T\upsilon_l^{-2}\left[1-e^{2\sqrt{T}h^*_{l,m}}\right]\Big)\right\}}\\
&= \mathcal{O}\left(\frac{\sqrt{T}}{\upsilon_l^2}[1-\exp\{2\sqrt{T}h^*_{l,m}\}]\right)\\
&= \mathcal{O}(\upsilon_l^{-1}).
\end{aligned}$$

Therefore, we find that

$$\begin{aligned}
R_{4,6} &= \mathcal{O}\left(\sum_{l=J_2+1}^{J-1}\sum_{m=0}^{2^l-1}\frac{1}{\upsilon_l}\mathbb{I}\Big(\sqrt{T}|b_{l,m}|\le M\Big)\right)\\
&= \mathcal{O}\left(\sum_{l=J_2+1}^{J-1}\sum_{m=0}^{2^l-1}\frac{1}{\upsilon_l}\right)\\
&= \mathcal{O}\left(\sum_{l=J_2+1}^{J-1}2^{(1-\mu_3)l}\right)\\
&= \mathcal{O}\left(\frac{2^{-r}T^{-r(1/2+r/2\omega(2r+1))}-2^rT^{-r}}{1-2^{-r}}\right)
\end{aligned}$$

□

A.2 Proofs for the Bayesian Modelling of the Haar-Fisz Periodogram

Proof (Lemma 7.4.1)**.**

$$\begin{aligned}
Q_l^i(h) &= \int_{-\infty}^{\infty} \frac{y^i}{\nu_l}\, \varphi_{\nu_l}(y-h)\, \xi(y)\, \mathrm{d}y \\
&= \int_{-\infty}^{\infty} \frac{y^i}{\nu_l}\, \varphi_{\nu_l}(y-h)\, \frac{\tau_l}{2} \exp\{-\tau_l\,|y|\}\, \mathrm{d}y \\
&= \frac{\tau_l}{2} \int_{-\infty}^{0} \frac{y^i}{\nu_l}\, \varphi_{\nu_l}(y-h)\, \exp\{\tau_l\, y\}\, \mathrm{d}y + \frac{\tau_l}{2} \int_{0}^{\infty} \frac{y^i}{\nu_l}\, \varphi_{\nu_l}(y-h)\, \exp\{-\tau_l\, y\}\, \mathrm{d}y \\
&= \frac{\tau_l}{2} \int_{-\infty}^{0} \frac{y^i}{\nu_l\sqrt{2\pi}}\, \exp\left\{-\frac{y^2-2hy+h^2}{2\nu_l^2}\right\}\, \exp\{\tau_l\, y\}\, \mathrm{d}y \qquad \text{(A.98)} \\
&\qquad + \frac{\tau_l}{2} \int_{0}^{\infty} \frac{y^i}{\nu_l\sqrt{2\pi}}\, \exp\left\{-\frac{y^2-2hy+h^2}{2\nu_l^2}\right\}\, \exp\{-\tau_l\, y\}\, \mathrm{d}y. \qquad \text{(A.99)}
\end{aligned}$$

If we first consider the exponential parts of (A.98), this can be expressed as

$$\begin{aligned}
-\frac{y^2-2yh+h^2}{2\nu_l^2} + \tau_l\, y &= -\frac{1}{2\nu_l^2}\left(y^2-2yh+h^2-2\,y\,\nu_l^2\,\tau_l\right) \\
&= -\frac{1}{2\nu_l^2}\left(y^2-2y\left[h+\nu_l^2\,\tau_l\right]+h^2\right).
\end{aligned}$$

Let $\mu_1 = h + \nu_l^2\,\tau_l$, then

$$-\frac{y^2-2yh+h^2}{2\nu_l^2} + \tau_l\, y = -\frac{1}{2\nu_l^2}\left(y-\mu_1\right)^2 - \frac{1}{2\nu_l^2}\left(h^2-\mu_1^2\right).$$

Similarly, if we let $\mu_2 = h - \nu_l^2\,\tau_l$, equation (A.99) can be expressed as

$$-\frac{1}{2\nu_l^2}\left(y-\mu_2\right)^2 - \frac{1}{2\nu_l^2}\left(h^2-\mu_2^2\right).$$

Therefore, our previous integrals become

$$Q_l^i(h) \;=\; \frac{\tau_l}{2} e^{-h^2/2\nu_l^2}\left[e^{\mu_1^2/2\nu_l^2}\int_{-\infty}^{0} y^i\,\varphi_{\nu_l}(y-\mu_1)\,\mathrm{d}y \;+\; e^{\mu_2^2/2\nu_l^2}\int_{-\infty}^{0}(-y)^i\,\varphi_{\nu_l}(y+\mu_2)\,\mathrm{d}y\right].$$

□

Proof (Lemma 7.4.2)**.** (a) As $\varphi_{\nu_l}(\cdot)$ is the Gaussian pdf, we can use the Gaussian cdf $\Phi(\cdot)$ to solve $Q_l^0(h)$.

(b) The results from equations (A.11) and (A.12) were used to determine $Q_l^1(h)$.

(c) Equations (A.16) and (A.17) were used to determine $Q_l^2(h)$.

□

Proof (Lemma 7.4.3)**.** The log-likelihood is obtained by first calculating $\gamma(y|\nu_l)$ from (7.8)

$$\begin{aligned}
\gamma(y|\nu_l) &= \int_{-\infty}^{\infty} \frac{1}{\nu_l}\varphi\left(\frac{y-x}{\nu_l}\right)\frac{\tau_l}{2}\ \exp\{-\tau_l\,|x|\}\,\mathrm{d}x \\
&= \frac{\tau_l}{2}\int_{-\infty}^{0}\frac{1}{\nu_l\sqrt{2\pi}}\exp\left\{-\frac{y^2-2xy+x^2}{2\nu_l^2}\right\}\exp\{\tau_l\,x\}\mathrm{d}x \\
&\quad + \frac{\tau_l}{2}\int_{0}^{\infty}\frac{1}{\nu_l\sqrt{2\pi}}\exp\left\{-\frac{y^2-2xy+x^2}{2\nu_l^2}\right\}\exp\{-\tau_l\,x\}\mathrm{d}x \\
&= \frac{\tau_l}{2}\int_{-\infty}^{0}\frac{1}{\nu_l\sqrt{2\pi}}\exp\left\{-\frac{x^2-2x[y+\nu_l^2\tau_l]+y^2}{2\nu_l^2}\right\}\mathrm{d}x \\
&\quad + \frac{\tau_l}{2}\int_{0}^{\infty}\frac{1}{\nu_l\sqrt{2\pi}}\exp\left\{-\frac{x^2-2x[y-\nu_l^2\tau_l]+y^2}{2\nu_l^2}\right\}\mathrm{d}x
\end{aligned}$$

Let $\mu_3 = y+\nu_l^2\tau_l$ and $\mu_4 = y-\nu_l^2\tau_l$, hence

$$\begin{aligned}
\gamma(y|\nu_l) &= \frac{\tau_l}{2}e^{-y^2/2\nu_l^2}\left[e^{\mu_3^2/2\nu_l^2}\int_{-\infty}^{0}\frac{1}{\nu_l\sqrt{2\pi}}\exp\left\{-\frac{(x-\mu_3)^2}{2\nu_l^2}\right\}\mathrm{d}x\right. \\
&\qquad \left. + e^{\mu_4^2/2\nu_l^2}\int_{0}^{\infty}\frac{1}{\nu_l\sqrt{2\pi}}\exp\left\{-\frac{(x-\mu_4)^2}{2\nu_l^2}\right\}\mathrm{d}x\right] \\
&= \frac{\tau_l}{2}e^{-y^2/2\nu_l^2}\left[e^{\mu_3^2/2\nu_l^2}\int_{-\infty}^{0}\varphi_{\nu_l}(x-\mu_3)\,\mathrm{d}x\ +\ e^{\mu_4^2/2\nu_l^2}\int_{0}^{\infty}\varphi_{\nu_l}(x-\mu_4)\,\mathrm{d}x\right] \\
&= \frac{\tau_l}{2}e^{-y^2/2\nu_l^2}\left[e^{\mu_3^2/2\nu_l^2}\Phi\left(\frac{-\mu_3}{\nu_l}\right)+e^{\mu_4^2/2\nu_l^2}\Phi\left(\frac{\mu_4}{\nu_l}\right)\right]. \qquad \text{(A.100)}
\end{aligned}$$

We then substitute the result of (A.100) into $\gamma(y|\nu_l)$ of (7.7) to obtain the log-likelihood function. □

Appendix B

Appendix of Computational Output

Figures B.1 shows the plots of the Bayesian log-periodogram estimate of the EWS for X_t and Y_t after forty cycle spins. The plots in figure B.2 are the estimates of the Bayesian log-periodogram estimate of the EWS for X_t using the Haar smoothing wavelet, and figure B.3 are the estimates of the Bayesian log-periodogram estimate of the EWS for Y_t using Daubechies extremal phase smoothing wavelet with ten vanishing moments.

The plots in figure B.4 and B.5 are the empirical p.acf over lag $\tau = 0, \ldots, 500$ for a fine, mid and coarse scale ($j = 9, 6, 0$).

Figure B.6 shows the estimate of the EWS of X_t with no cycle spinning using the uniform (a) and Laplace (b) mixture priors. Figure B.8 shows the EWS at the three finest scales. Figure B.7 shows the estimate of the EWS of Y_t with no cycle spinning using the uniform (a) and Laplace (b) mixture priors. Figure B.9 shows the EWS $j = 5, 6, 7$.

(a) $\hat{S}(z)$ for X_t (b) $\hat{S}(z)$ for Y_t

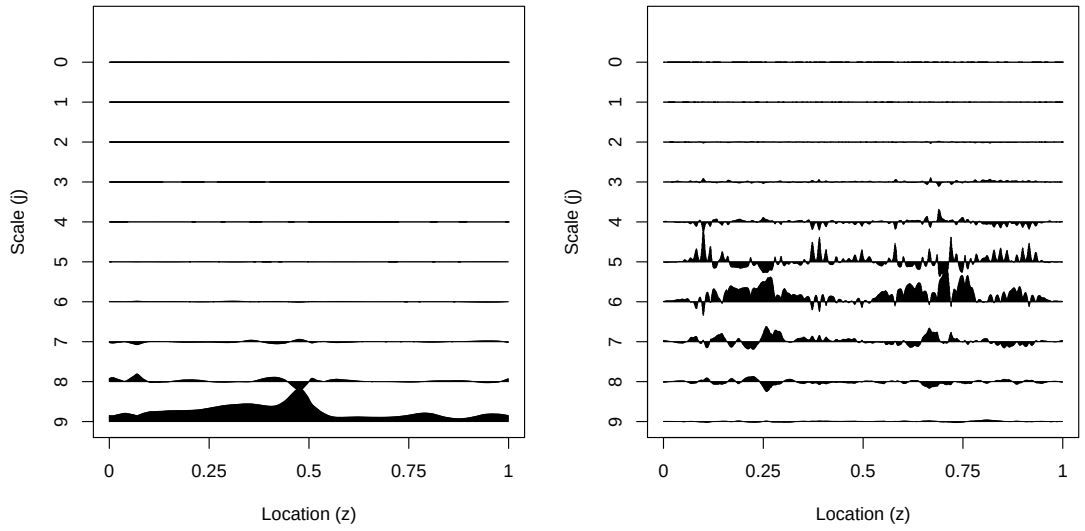
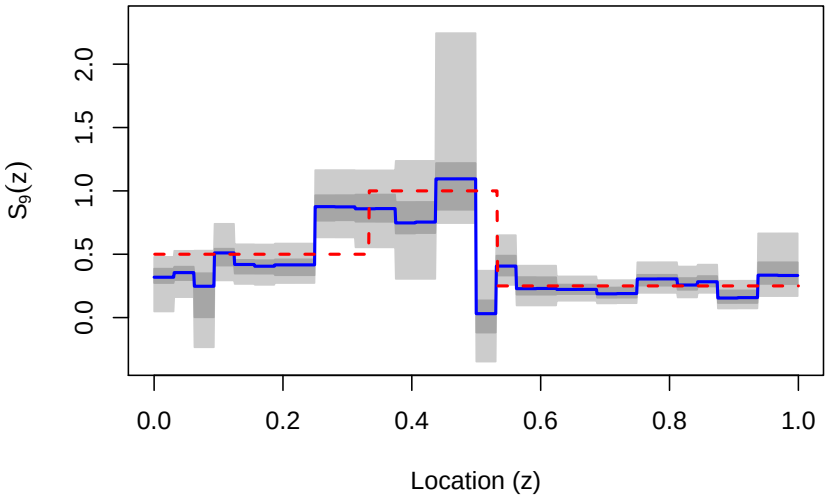
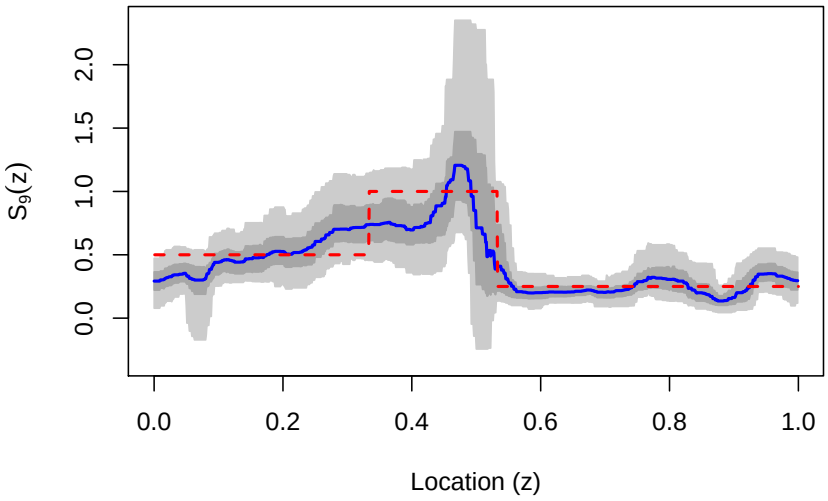


Figure B.1: Plots of the estimated EWS from 40 cycle spins, using Bayesian wavelet shrinkage estimate of the log transformed raw wavelet periodograms obtained from X_t and Y_t.

(a) No cycle spins



(b) 20 cycle spins



(c) 40 cycle spins

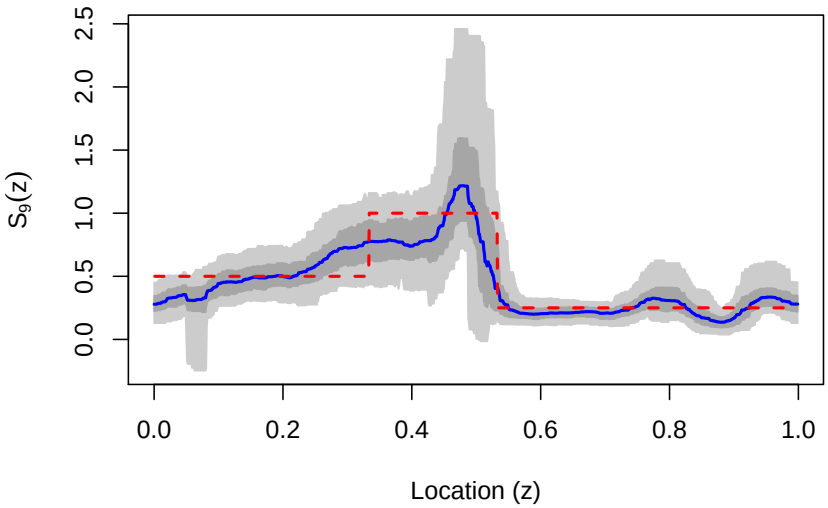
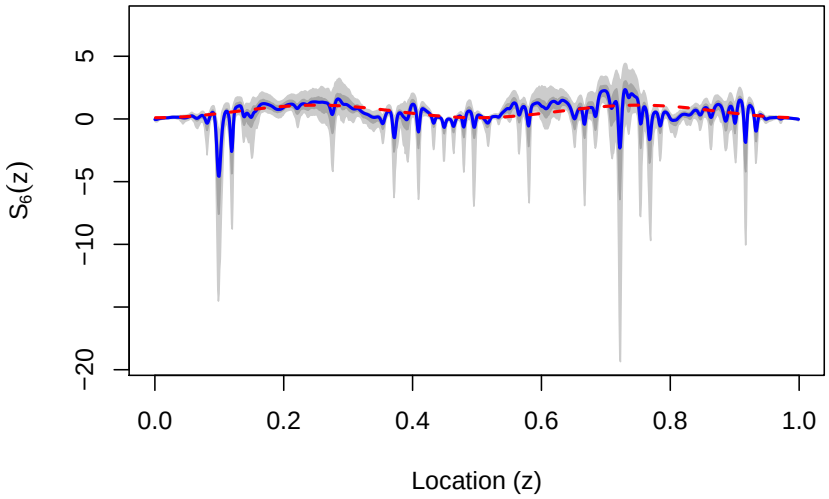
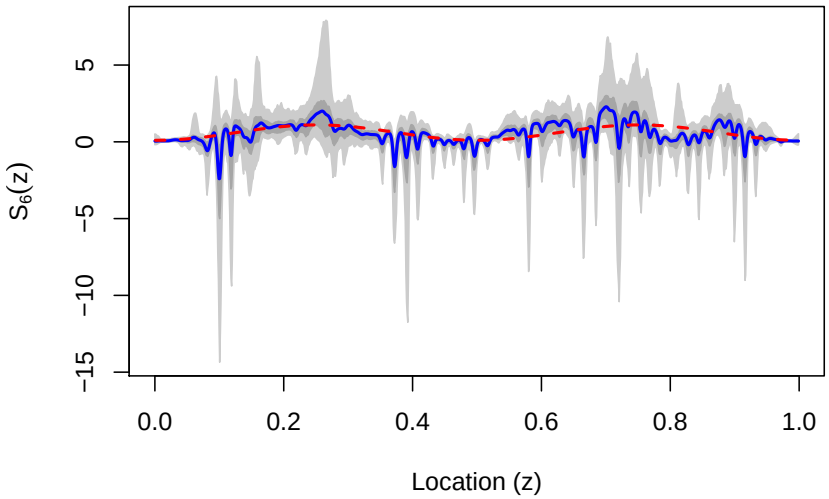


Figure B.2: Plots of the estimated EWS at scale $j = 9$ from 0, 20, 40 cycle spins, using Bayesian wavelet shrinkage estimate of the log transformed raw wavelet periodograms obtained from X_t using the Haar smoothing and analysis wavelet.

(a) No cycle spins



(b) 20 cycle spins



(c) 40 cycle spins

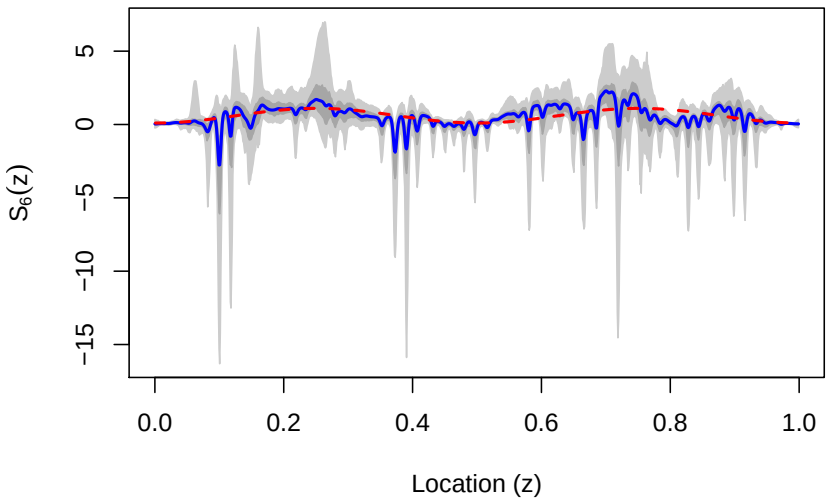


Figure B.3: Plots of the estimated EWS at scale $j = 6$ from 0, 20, 40 cycle spins, using Bayesian wavelet shrinkage estimate of the log transformed raw wavelet periodograms obtained from Y_t using the D_{10} smoothing and analysis wavelet.

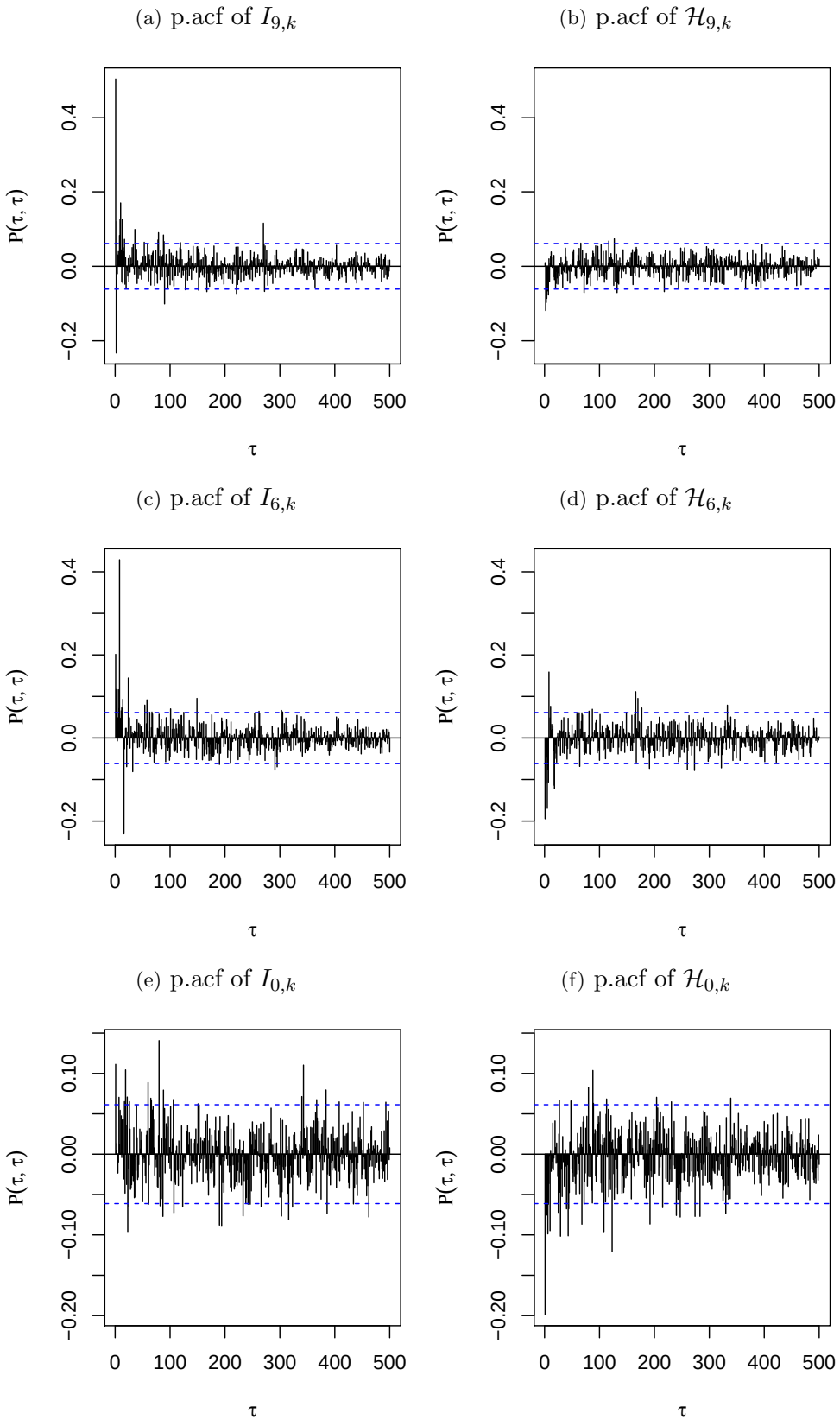
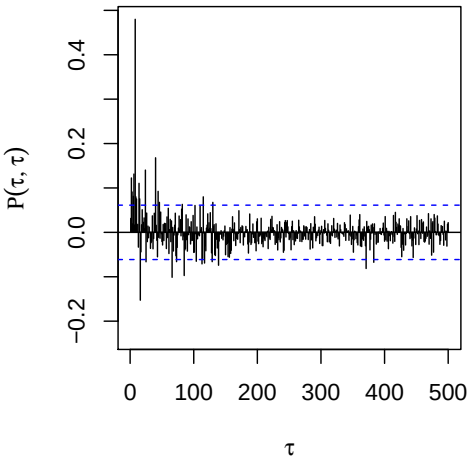
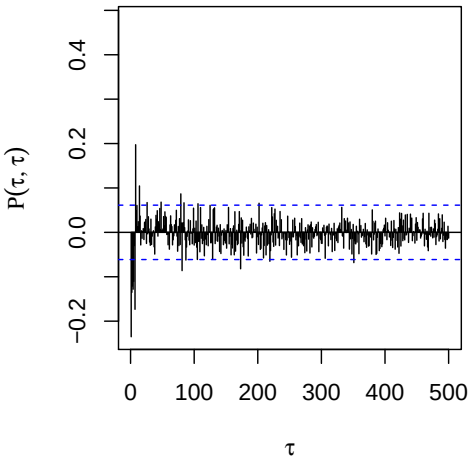


Figure B.4: Plots of the empirical p.acf over lag $\tau = 0, \ldots, 500$ of the raw wavelet periodogram ($I_{j,k}$) and Haar-Fisz wavelet periodogram ($\mathcal{H}_{j,k}$) of X_t simulated from the piecewise constant EWS at: the finest scale in (a) and (b); a mid scale in (c) and (d); the coarsest scale in (e) and (f). For a sample size of $T = 2^{10} = 1024$. The blue dashed line in each plot is the 5% significance level.

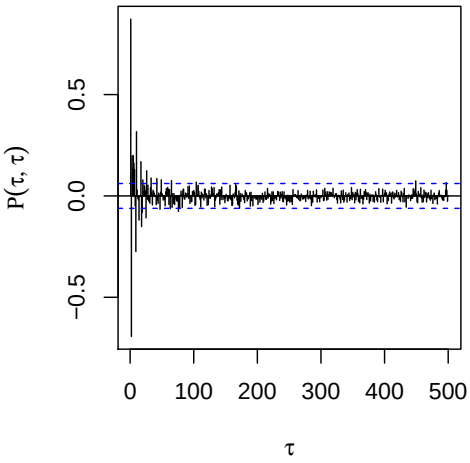
(a) p.acf of $I_{9,k}$



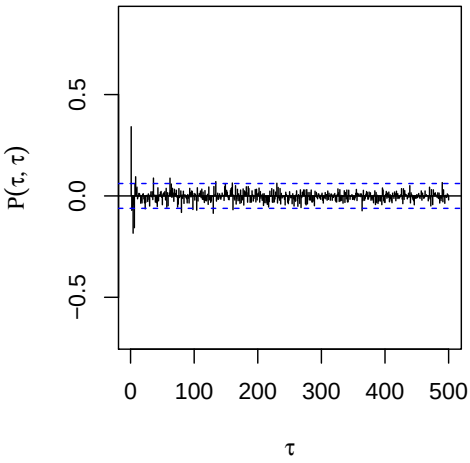
(b) p.acf of $\mathcal{H}_{9,k}$



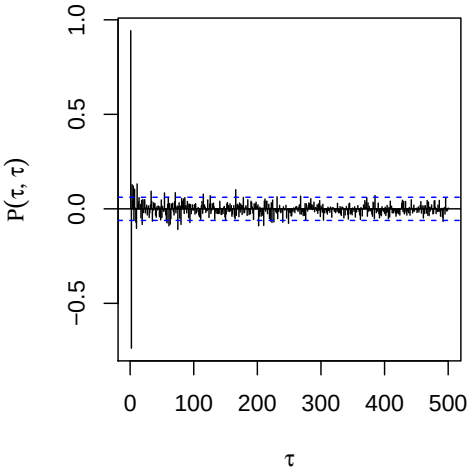
(c) p.acf of $I_{6,k}$



(d) p.acf of $\mathcal{H}_{6,k}$



(e) p.acf of $I_{0,k}$



(f) p.acf of $\mathcal{H}_{0,k}$

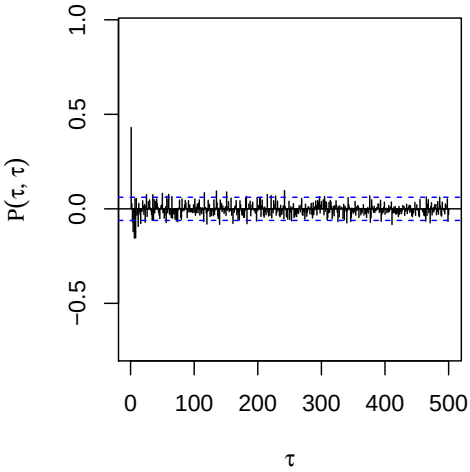


Figure B.5: Plots of the empirical p.acf over lag $\tau = 0, \ldots, 500$ of the raw wavelet periodogram ($I_{j,k}$) and Haar-Fisz wavelet periodogram ($\mathcal{H}_{j,k}$) of Y_t simulated from the slowly evolving EWS at: the finest scale in (a) and (b); a mid scale in (c) and (d); the coarsest scale in (e) and (f). For a sample size of $T = 2^{10} = 1024$. The blue dashed line in each plot is the 5% significance level.

(a) Uniform Mixture Prior EWS Estimate (b) Laplace Mixture Prior EWS Estimate

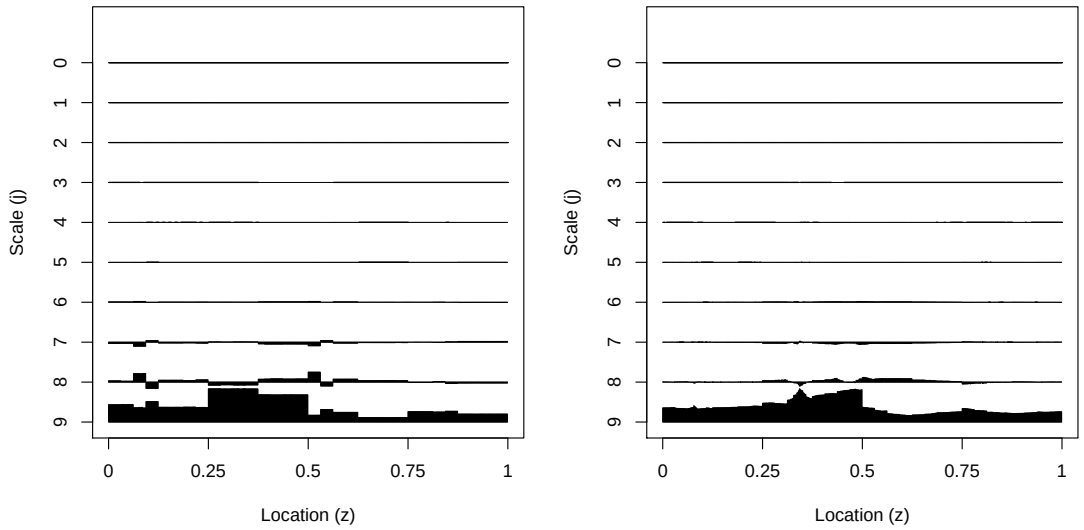


Figure B.6: Plots of the estimated EWS of X_t, with no cycle spinning. (a) is the estimated EWS using the Haar wavelet and uniform mixture prior. (b) is the estimated EWS using D_3 and Laplace mixture prior.

(a) Uniform Mixture Prior EWS Estimate (b) Laplace Mixture Prior EWS Estimate

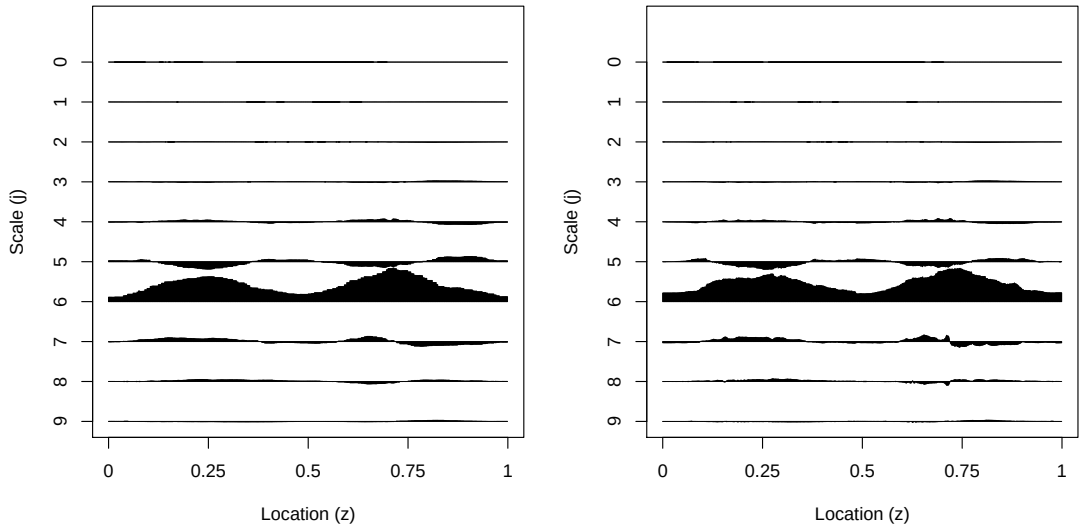


Figure B.7: Plots of the estimated EWS of Y_t, with no cycle spinning. (a) is the estimated EWS using the Haar wavelet and uniform mixture prior. (b) is the estimated EWS using D_8 and Laplace mixture prior.

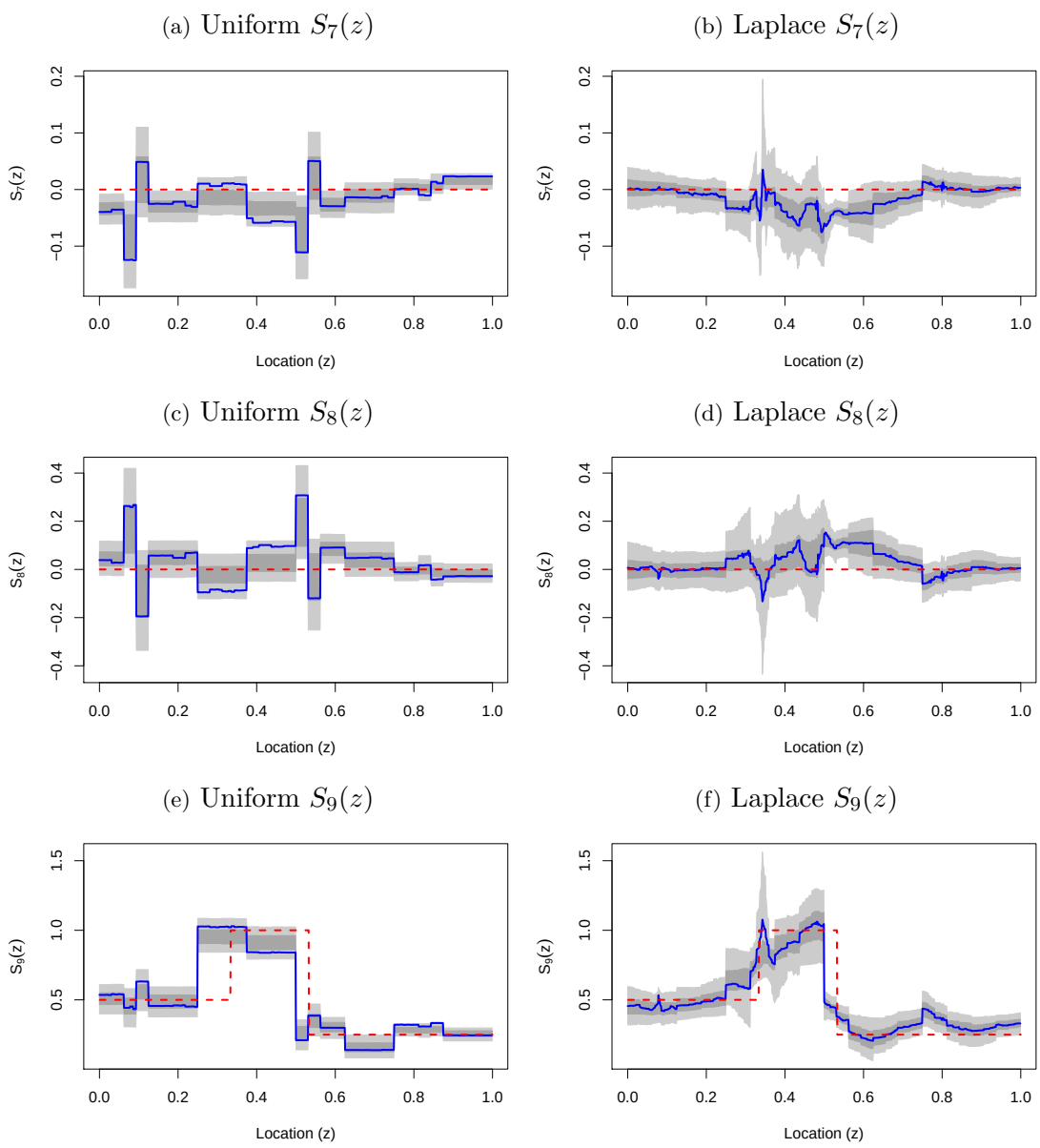


Figure B.8: Plots of the estimated EWS with 50% and 90% CI, at scale $j = 9, 8, 7$ using Bayesian wavelet shrinkage on X_t's Haar-Fisz transformed wavelet periodogram with no cycle spinning. Figures (a), (c) and (e) have been smoothed using the Haar wavelet and uniform mixture prior, whereas (b), (d) and (f) have been smoothed using the Laplace mixture prior and D_3 wavelet.

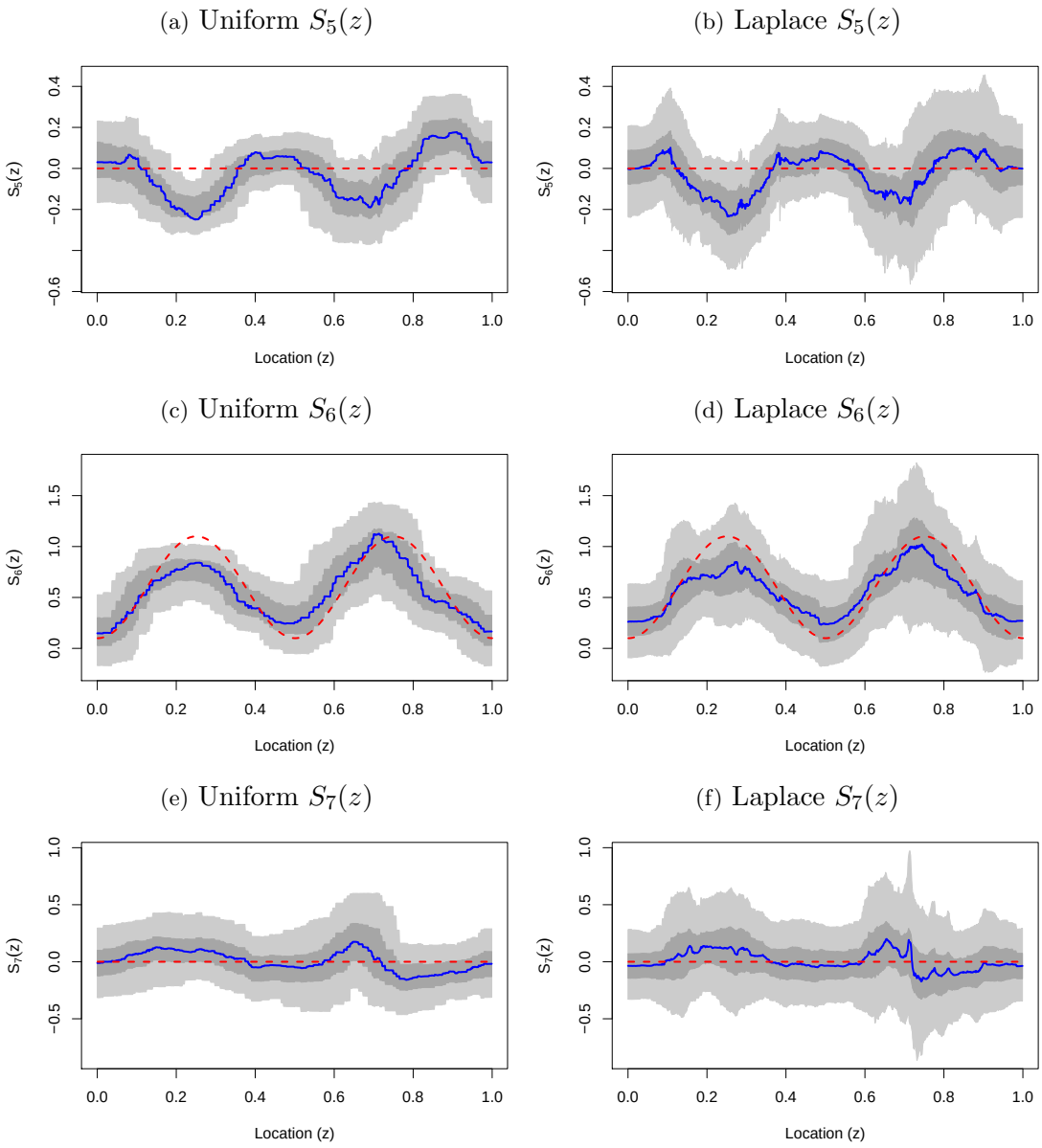


Figure B.9: Plots of the estimated EWS with 50% and 90% CI, at scale $j = 7, 6, 5$ using Bayesian wavelet shrinkage on Y_t's Haar-Fisz transformed wavelet periodogram with no cycle spinning. Figures (a), (c) and (e) have been smoothed using the Haar wavelet and uniform mixture prior, whereas (b), (d) and (f) have been smoothed using the Laplace mixture prior and D_8 wavelet.