Wavelet Methods for Locally Stationary Time Series

Euan Thomas McGonigle, M.Sci (Hons.), M.Res

Submitted for the degree of Doctor of Philosophy

at Lancaster University.

July 2020

STOR-i

excellence with impact

Abstract

Time series data can often possess complex and dynamic characteristics. Two key statistical properties of time series – the mean (first-order) and autocovariance (second-order) – commonly change over time. Modelling this evolution of so-called nonstationary time series is crucial to making informed inference on the data. This thesis focuses on wavelet-based methodology for the simultaneous modelling of first and second-order nonstationary time series, for which we provide three main contributions.

First, we propose a method using differencing to jointly estimate the time-varying trend and second-order structure of a time series, within the locally stationary wavelet processes framework. We discuss a wavelet-based estimator of the second-order structure of the original time series by employing differencing, and show how this can be incorporated into the estimation of the trend of the time series.

Second, we propose a framework for modelling series with simultaneous time-varying first and second-order structure by removing the restrictive zero-mean assumption of locally stationary wavelet (LSW) processes and extending the applicability of the locally stationary wavelet model to include a trend component. We develop associated estimation theory for both first and second-order time series quantities and show that our estimators achieve good properties in isolation of each other by making appropriate assumptions on the series trend.

Last, we consider simultaneous modelling of first and second-order structure in the scenario where the mean function is piecewise constant. We propose a

likelihood-based method using wavelets to detect changes in mean in time series that exhibit time-varying autocovariance. This allows for a more flexible model for mean changepoint detection, since commonly the second-order structure is assumed to be independent and identically distributed. The performance of the method is investigated via simulation, and is shown to perform well in a variety of time series scenarios.

Acknowledgements

Firstly, I would like to thank my academic supervisors, Rebecca Killick and Matt Nunes, for their guidance and mentorship throughout my PhD. Your support has helped me develop as a researcher and I'm truly grateful for the opportunities that I've had as a result of your supervision. Thanks also to my industrial supervisor Martyn Byng at the Numerical Algorithms Group, for his encouragement and easy-going approach to the PhD.

I would also like to express my gratitude for my time spent at the STOR-i Centre for Doctoral Training. It has been a real privilege to spend the last four years in such a vibrant, friendly environment. Many thanks to all those involved at STOR-i, especially the directors and administrative team, for making the department such a success. In particular, I'd like to thank Professor Jonathan Tawn for his tireless work in the running of STOR-i and always being available whenever I needed advice – cheers Jon.

I'd like to thank my housemate Sam, whose kindness and generosity is matched only by his inventiveness with the vacuum cleaner. Thanks also to Jake for taking over from Sam as my housemate, providing many entertaining nights and amusing discussions.

A special thanks to my year group, in particular Edwin, Georgia, Rob, Séan, and Zak. It's been a pleasure to share this experience with you all, thanks for making my time in Lancaster such an enjoyable one. Thanks also for the regular video calls which have helped make these strange times seem a bit more normal.

A very big thank you also to my family, for their continuing support throughout

my education. Thanks for your invaluable help and advice over the years. I'm very thankful for the opportunities that I've received with your help.

Lastly, thank you especially to Hankui. I am beyond fortunate to have had your love, support and encouragement over these past few years. Thanks for your help checking for typos, liberal uses of the hyphen, and attempts to shorten some of my very long sentences. Thank you for your positivity and patience, especially with our "sightseeing" in Liverpool. My time in Lancaster would not have been the same without you.

Declaration

I declare that the work in this thesis has been done by myself and has not been submitted elsewhere for the award of any other degree.

A version of Chapter 3 has been submitted for publication as McGonigle, E. T., Killick, R. and Nunes, M. A. (2020). Modelling Time-Varying First and Second-Order Structure of Time Series via Wavelets and Differencing.

A version of Chapter 4 has been submitted for publication as McGonigle, E. T., Killick, R. and Nunes, M. A. (2020). Trend Locally Stationary Wavelet (T-LSW) Processes.

A version of Chapter 5 has been submitted for publication as McGonigle, E. T., Killick, R. and Nunes, M. A. (2020). Detecting Changes in Mean in the Presence of Time-Varying Autocovariance.

Euan Thomas McGonigle

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Wavelets can be informally described as oscillatory basis functions that decay rapidly, constructed in such a manner so as to possess several attractive properties. Wavelets are multiscale in nature, can often provide sparse representations of functions, and enjoy localisation in both time and frequency. Due to their many advantageous properties, wavelets have found prominent use in many areas of statistics, and in particular have enjoyed popularity for many years in time series analysis. Wavelets can provide an alternative approach to the classical Fourier representation of a time series using trigonometric functions.

In this thesis, we focus on developing wavelet-based methodology to model the evolution of time series data, which can often possess complex and dynamic characteristics. Most classical time series models are built upon the assumption of stationarity, for example autoregressive moving average (ARMA) processes. However, in most practical scenarios, two key statistical properties of a time series – the mean and autocovariance – vary over time. Appropriately modelling this time-varying behaviour is crucial in making informed inference on the data. We refer to a time series whose mean and/or autocovariance is time-varying as first, respectively second-order nonstationary. Areas of research that feature time series that exhibit nonstationarity in both first and second-order structure are numerous. Recent examples include biomedical time series (Hargreaves et al., 2018), climatology (Das and Politis, 2020), and financial time series (Roueff and von Sachs, 2019).

In recent years, there has been much research effort in the area of nonstationary time series analysis, including so called "locally stationary" time series. As such, there are many time series models that can cope with nonstationarity in first *or* second-order structure. However, less attention in the time series literature is given to the joint consideration of time-varying first *and* second-order behaviour, due to the highly challenging nature of the problem.

In this thesis, we consider the extension and adaptation of the modelling framework of locally stationary wavelet (LSW) processes. The locally stationary wavelet (LSW) model, introduced by Nason et al. (2000), provides a way to capture nonstationary second-order behaviour. The LSW model has gained attention and much research focus in the time series literature, due to its ability to represent the series at different scales, enabling the second-order structure to be more readily discovered and estimated. There have been many successful applications of the LSW framework to prominent problems in time series analysis. For example, LSW modelling has been utilised in forecasting (Fryzlewicz et al., 2003), handling missing observations (Knight et al., 2012), changepoint detection (Killick et al., 2013), and classification (Krzemieniewska et al., 2014).

The LSW framework is, however, only suitable for the modelling of second-order nonstationary time series. It cannot account for time-varying first-order behaviour, as the methodology is restricted to zero-mean time series. This means that LSW modelling is not immediately applicable to most time series in practice, due to this restrictive assumption. Commonly, a time series is preprocessed to be converted to a zero-mean process. This adds an extra level of data analysis which will have an effect on the results of the task in question, as well as additional computational complexity. Furthermore, the task of removing the first-order structure is itself a highly challenging problem, made difficult due to the presence of time-varying second-order behaviour.

The main contribution of this thesis is to expand the LSW modelling framework to incorporate a nonstationary first-order component. In doing so, we ensure

wider applicability of the methodology as well as a more realistic set of modelling assumptions for most practical time series scenarios. Using the well-established theory of wavelets, we propose new tools for modelling nonstationary time series, expanding upon the rigorous theory developed in Nason et al. (2000). We provide theoretical results that enable consistent estimation of the various quantities of interest, discuss the necessary computational aspects of our work, and describe the practical considerations for implementation of our methodology.

1.1 Contributions and Thesis Outline

In this thesis, we provide three main contributions to the nonstationary time series literature. Each main chapter of the thesis is summarised below. The contributions are described in Chapters 3 – 5, and can be broadly described as contributions to the problem of jointly modelling first and second-order nonstationary time series. The thesis consists of an introductory literature review in Chapter 2, the main contributions are given in Chapters 3 – 5, while concluding remarks and areas for future research are given in Chapter 6. All proofs of theoretical results presented in the main body are given in the appropriate appendices.

Chapter 2: Literature Review

This chapter reviews the basics of wavelet theory, and provides a survey of wavelet-based techniques in time series analysis. In particular, we discuss the locally stationary wavelet (LSW) time series model of Nason et al. (2000). This model forms the basis for the work carried out in Chapters 3 – 5, in which we extend the LSW model, capable only of modelling second-order time series behaviour, to include a first-order component. We also provide a survey of trend estimation techniques, both wavelet-based and otherwise. We conclude with a short survey of methods applicable to first and second-order nonstationary time series.

Chapter 3: Modelling Time-Varying First and Second-Order Structure of Time Series via Wavelets and Differencing

In this chapter, we consider the problem of incorporating a time-varying first-order component in a second-order nonstationary time series. We propose a method for joint estimation of nonstationary first and second-order structure of a time series using both wavelets and differencing. Differencing is a commonly used technique to remove a trend from a times series without the need to estimate it. We investigate the effect that differencing has on the second-order properties of a locally stationary time series. We propose methodology that enables the consistent estimation of the second-order structure of the original time series using the differenced series, by taking into account the effect of differencing. This information is then used in a wavelet thresholding-based estimator of the trend of the series. We demonstrate the performance of the method through a simulation study, and show the potential uses of the methodology by analysing two data examples. The second example highlights the ability of the method to make features of the original time series more apparent, when compared with analysis of the differenced data.

Chapter 4: Trend Locally Stationary Wavelet Processes with Applications to Environmental Data

This work can be viewed as a companion to the work in Chapter 3, as the same general problem is addressed. However, a different solution to that problem is presented here in Chapter 4, and some key modelling assumptions differ between the chapters. This chapter can be viewed as more of a complete modelling framework compared with Chapter 3, which investigated how to utilise the technique of differencing in a locally stationary time series setting. In this chapter, we develop a wavelet-based method that expands upon the work on locally stationary wavelet process modelling of Nason et al. (2000). We remove the restrictive zero-mean assumption of LSW processes, and extend the applicability of the LSW model by incorporating a trend component. We show that under certain assumptions we

can estimate the first and second-order properties of the time series in isolation of one another. We provide theoretical results for the consistency properties of these estimators. We illustrate the utility of the methodology by analysing the much studied global mean sea temperature time series. Furthermore, we describe a new way to perform boundary handling of the discrete wavelet transform in the case where a trend is present. Finally, we provide an important theoretical contribution to the LSW literature, by proving a result relating to the correction matrix employed in LSW estimation theory. We adapt an erroneous result in Cardinali and Nason (2017), which enables theoretical results on the LSW model to apply to a wider family of wavelets than originally described in Nason et al. (2000).

Chapter 5: Detecting Changes in Mean in the Presence of Time-Varying Autocovariance

In this chapter, we consider the scenario where the mean function of the time series is given by a piecewise constant function. This setting is within the area of changepoint detection. Commonly, changepoint detection methods assume that the error structure of the time series is independent, identically distributed (IID) noise. We place an LSW process assumption on the error structure, allowing for a more flexible modelling framework. We show that we can estimate the second-order structure in the presence of a piecewise constant mean function by using a running-median smoothed wavelet estimator. We propose a likelihood-based method using wavelets to detect changes in mean in time series that can display a much more general error structure than simply IID errors. Our proposed technique is shown to work well through the use of a simulation study. The method is applied to two financial time series data sets, which illustrates the method's ability to work well in the presence of significant autocorrelation.

Chapter 6: Conclusions and Future Work

In this chapter we summarise the conclusions of the thesis and describe several possible directions for future work, stemming from the research proposed in Chapters 3 – 5.

Note that Chapters 3, 4 and 5 are presented as a series of papers. Therefore, they are written to be read in isolation of one another. Due to this, there may be expository detail and background information that is repeated in these chapters.

Chapter 2

Literature Review

This chapter provides an overview of the background wavelet theory necessary for the work in the thesis. In addition, we review the relevant literature in the statistical areas that are the focus of this thesis: time series analysis and trend estimation/nonparametric regression. In particular, we concentrate on locally stationary time series and wavelet-based trend estimation.

2.1 Wavelet Theory

Loosely speaking, wavelets are localised, oscillatory basis functions that possess several useful properties not typically enjoyed by Fourier trigonometric basis functions. A desirable property of basis functions is efficiency, in the sense that only a few of the coefficients in a basis representation are non-zero. Although Fourier trigonometric functions are localised in frequency, they are not localised in time. Consequently, many (Fourier) basis functions contribute to the basis representation at any one point. Therefore, Fourier trigonometric functions struggle to represent functions with discontinuities or "sharp" changes efficiently, due to their infinite support. By contrast, wavelets enjoy localisation in both time and frequency. This allows wavelets to represent discontinuities well, and provide efficient representations of functions.

The study of wavelets began with the work of Haar (1910), who developed the

theory of the Haar basis – an orthonormal basis of $L^2(\mathbb{R})$, the space of square-integrable functions. The true advent of wavelets did not occur until the late 1980s, when Daubechies, Mallat, and many others further developed the underlying theory. Since then, wavelets have proved to be an important tool of analysis in statistics, as well as areas such as engineering (Williams and Amaratunga, 1994) and image processing (Chan and Shen, 2005).

Following Daubechies (1992), a wavelet is formally defined as a function $\psi \in L^2(\mathbb{R})$ (often referred to as the *mother wavelet*) that satisfies the following admissibility condition:

$$C_\psi = \int_{-\infty}^{\infty} \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega < \infty, \tag{2.1.1}$$

where $\hat{\psi}(\omega)$ is the Fourier transform of $\psi(x)$. In this thesis, we utilise wavelet functions whose dyadic dilations and translations, given by

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

form an orthonormal basis of $L^2(\mathbb{R})$. The indices j and k are known as scale/dilation and location/translation parameters, respectively. Condition (2.1.1) implies that

$$0 = \hat{\psi}(0) = \int_{\mathbb{R}} \psi(x) dx. \tag{2.1.2}$$

Equation (2.1.2) highlights the oscillatory nature of a wavelet, and motivates the name itself. The “-let" suffix comes from the fact that ψ is oscillatory and well localised in both time and frequency. Equation (2.1.2) also implies time localisation (since it implies that $\psi \in L^1(\mathbb{R})$). Equation (2.1.1) can be interpreted as localisation in frequency. Many wavelets have compact support (only non-zero on a compact set), and all wavelets decay rapidly to zero. The coefficients of a wavelet basis function expansion of a function $f \in L^2(\mathbb{R})$ are given in the usual way by the inner product $d_{j,k} := \langle f, \psi_{j,k} \rangle = \int f(x)\psi_{j,k}(x)dx$. The coefficients $d_{j,k}$ can be interpreted as scaled, localised differences of the function f.

2.1.1 Multiresolution Analysis (MRA)

In this thesis we are concerned with discretely sampled time series, and hence discrete wavelets. We therefore are interested in wavelet decompositions of vectors, as opposed to continuous functions as in the previous section. The key idea of constructing the necessary wavelet bases lies in the concept of a multiresolution analysis (MRA), which is a collection of subspaces $\{V_j\}_{j\in\mathbb{Z}}$ of $L^2(\mathbb{R})$. Informally, we can think of each space V_j as the space of functions that have detail up to some defined level of resolution. The formal definition of an MRA, as given by Mallat (1989a), is as follows.

Definition 2.1.1. A *multiresolution analysis* (MRA) is a sequence of closed subspaces $\{V_j\}_{j\in\mathbb{Z}}$ of $L^2(\mathbb{R})$ such that they lie in containment hierarchy

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

such that their union is dense in $L^2(\mathbb{R})$ and their intersection is trivial:

$$\overline{\bigcup_j V_j} = L^2(\mathbb{R}), \qquad \bigcap_j V_j = \{0\}.$$

Furthermore, the V_j are constructed such that (i) V-spaces are self similar, i.e.

$$f(2^j x) \in V_j \iff f(x) \in V_0,$$

and (ii) there exists a *scaling function*, or *father wavelet*, $\phi \in V_0$ whose integer translates span V_0, that is

$$V_0 = \left\{ f \in L^2(\mathbb{R}) \mid f(x) = \sum_k c_k \phi(x-k) \right\},$$

and for which the set $\{\phi(\cdot - k), k \in \mathbb{Z}\}$ is an orthonormal basis.

Furthermore, we assume that $\int \phi(x)dx \neq 0$. Given that $\{\phi(x-k)\}_{k\in\mathbb{Z}}$ is an

orthonormal basis for V_0, self similarity implies that

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

for fixed j forms an orthonormal basis of V_j. Now, since $V_0 \subset V_1$, the function $\phi(x) \in V_0$ can be written as a linear combination of functions that belong to V_1, that is

$$\phi(x) = \sum_k h_k \phi_{1k}(x) = \sum_k h_k \sqrt{2}\phi(2x - k), \tag{2.1.3}$$

for some coefficients h_k, $k \in \mathbb{Z}$. Equation (2.1.3) is known as the *scaling equation* or *dilation equation*, and is one of the most fundamental results in wavelet analysis. The solution of this equation enables the construction of a general MRA. It describes how the scaling functions relate to one another for two consecutive scales. The (possibly infinite) vector $\{h_k, k \in \mathbb{Z}\}$ is known as a *wavelet filter*, and is a low-pass (averaging) filter. It is often useful to work in the Fourier domain to understand properties of wavelets. To this end, we require the following definition.

Definition 2.1.2. The *transfer function* m_0 is defined by

$$m_0(\omega) = \frac{1}{\sqrt{2}} \sum_{k \in \mathbb{Z}} h_k e^{-ik\omega} = \frac{1}{\sqrt{2}} H(\omega).$$

The function m_0 describes the behaviour of the filter $\{h_k\}$ in the Fourier domain. The function is 2π-periodic and the filter taps $\{h_k, k \in \mathbb{Z}\}$ are the Fourier coefficients of the function $H(\omega) = \sqrt{2}m_0(\omega)$. In the Fourier domain, the scaling equation (2.1.3) becomes

$$\hat{\phi}(\omega) = m_0\left(\frac{\omega}{2}\right)\hat{\phi}\left(\frac{\omega}{2}\right),$$

where $\hat{\phi}(\omega)$ is the Fourier transform of $\phi(x)$. By iterating this equation, we obtain

$$\hat{\phi}(\omega) = \prod_{n=1}^{\infty} m_0\left(\frac{\omega}{2^n}\right),$$

an infinite product which is convergent under very mild conditions on the rate of decay of the father wavelet ϕ (Vidakovic, 2009, Section 3.3). The transfer function is important as it forms part of the construction of generalised wavelets. Wavelet filters associated with an MRA satisfy the normalisation property $\sum_{k\in\mathbb{Z}} h_k = \sqrt{2}$ and the orthogonality property $\sum_k h_k h_{k-2l} = \delta_l$, for all $l \in \mathbb{Z}$.

To answer the question of how to construct $\phi(x)$ such that it satisfies the conditions for an MRA, as well as possibly satisfying other useful properties, Daubechies (1988) introduced a projection operator P_j that projects a function onto the space V_j. Now, since $\{\phi_{j,k}(x)\}_{k\in\mathbb{Z}}$ is a basis for V_j, this projection can be written as

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

for some coefficients $\{c_{j,k}\}_{k\in\mathbb{Z}}$. Informally, $P_j f$ can be thought of as a sort of approximation of f using just the father wavelets at level j. Orthogonality of the basis means that the coefficients can be computed in the standard way as

$$c_{j,k} = \langle f, \phi_{j,k}\rangle = \int_{-\infty}^{\infty} f(x)\phi_{j,k}(x)dx.$$

The following result, due to Daubechies (1992), establishes both existence, and construction, of a wavelet basis of $L^2(\mathbb{R})$.

Theorem 2.1.1. If $\{V_j\}_{j\in\mathbb{Z}}$ with ϕ form a multiresolution analysis of $L^2(\mathbb{R})$, then there exists an orthonormal wavelet basis $\{\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k) : j, k \in \mathbb{Z}\}$ of $L^2(\mathbb{R})$ such that for $j \in \mathbb{Z}$,

$$P_{j+1}f = P_j f + \sum_k \langle f, \psi_{j,k}\rangle\psi_{j,k}(x). \tag{2.1.4}$$

One possibility for the construction of the wavelet $\psi(x)$ is

$$\hat{\psi}(\omega) = e^{i\omega/2}\overline{m_0(\omega/2+\pi)}\hat{\phi}(\omega/2),$$

where $\hat{\psi}$ and $\hat{\phi}$ are the Fourier transforms of ψ and ϕ respectively and m_0 is the transfer function as defined in Definition 2.1.2. Equivalently,

$$\psi(x) = \sum_k (-1)^k h_{1-k} \phi_{1,k}(x). \tag{2.1.5}$$

The function $\psi(x) = \psi_{0,0}(x)$ is the *mother wavelet*. The set $\{\psi_{j,k}(x), k \in \mathbb{Z}\}$ for fixed j is an orthonormal basis of the "difference space" W_j, where $V_{j+1} = W_j \bigoplus V_j$. The coefficient that appears in Equation (2.1.5) describes how the wavelet is constructed in terms of the next finer scale father wavelet coefficients, and as such has its own notation:

$$g_n = (-1)^n h_{1-n}. \tag{2.1.6}$$

The filters $\{h_k\}$ and $\{g_k\}$ are known as *quadrature mirror filters*. Equation (2.1.4) can be telescoped to produce a fine-scale representation of a function f:

$$f(x) = \sum_{k \in \mathbb{Z}} c_{j_0,k} \phi_{j_0,k}(x) + \sum_{j=j_0}^{\infty} \sum_{k \in \mathbb{Z}} d_{j,k} \psi_{j,k}(x). \tag{2.1.7}$$

Equation (2.1.7) can be interpreted that a function $f \in L^2(\mathbb{R})$ can be represented as a 'smooth' part involving the $\phi_{j_0,k}$ and a set of detail representations $\sum_{k \in \mathbb{Z}} d_{j,k} \psi_{j,k}(x)$ that accumulate information at a set of scales j going from j_0 to infinity. The first set of terms represents the 'average' level of function while the second set represents the 'detail', or changes in average.

2.1.2 The Discrete Wavelet Transform

Any function $f \in L^2(\mathbb{R})$ can be represented via its wavelet basis expansion as

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

where $d_{j,k} = \langle f, \psi_{j,k} \rangle$. This representation is unique and corresponds to an MRA decomposition of $L^2(\mathbb{R}) = \bigoplus_{j=-\infty}^{\infty} W_j$. For any fixed j_0, the decomposition

$L^2(\mathbb{R}) = V_{j_0} \oplus \bigoplus_{j=j_0}^{\infty} W_j$ corresponds to the representation given in Equation (2.1.7).

Given $n = 2^J$ observations $\boldsymbol{y} = (y_1, \ldots, y_n)^\mathsf{T}$, the discrete wavelet transform (DWT) computes the coarsest father wavelet coefficient $c_{0,0}$ and the wavelet coefficients $d_{j,k}$ for $j = 0, 1, \ldots, J-1$, $k = 0, 1, \ldots, 2^j - 1$. The DWT is initiated from a set of father wavelet coefficients $c_{J,0}, \ldots, c_{J,n-1}$, and correspond to the finest resolution, which is at level J.

In time series analysis, it is normally the case that we simply use the data itself as the finest scale father wavelet coefficients, that is set $c_J = \boldsymbol{y}$, where c_J is the vector $(c_{J,0}, \ldots, c_{J,n-1})$. Father wavelet coefficients at level $j-1$ can be calculated using the level j coefficients by the relation

$$c_{j-1,k} = \sum_n h_{n-2k} c_{j,n}. \tag{2.1.8}$$

Similarly for the wavelet coefficients

$$d_{j-1,k} = \sum_n g_{n-2k} c_{j,n}. \tag{2.1.9}$$

These equations apply to any scale $j = 1, \ldots, J$. The coefficients $c_{j,k}$ are often referred to as the *smoothing* or *scaling* coefficients, while the $d_{j,k}$ are known as the *detail* or *wavelet* coefficients. We can interpret these two equations in another way. Equation (2.1.8) can be obtained by first applying the filter $\{h_n\}$ to the sequence $\{c_{j,n}\}$, giving

$$c^*_{j-1,k} = \sum_n h_{n-k} c_{j,n}. \tag{2.1.10}$$

We can then perform *dyadic decimation* on this sequence to obtain $c_{j-1,k} = c^*_{j-1,2k}$, i.e. we only keep the even elements of the sequence $\{c^*_{j-1,k}\}$. In signal processing this is also known as *downsampling* by a factor of 2. Define the (even) dyadic decimation operator $\mathcal{D}_0$ by

$$(\mathcal{D}_0 x)_l = x_{2l},$$

for a sequence $\{x_i\}$. Thus, we can express Equations (2.1.8) and (2.1.9) in the form

$$c_{j-1} = \mathcal{D}_0 \mathcal{H} c_j, \quad d_{j-1} = \mathcal{D}_0 \mathcal{G} c_j,$$

where $\mathcal{H}$ and $\mathcal{G}$ refer to the filtering operation given in Equation (2.1.10). Note that we have used vector notation here for c_j etc., rather than sequences. Nason and Silverman (1995) show that the entire set of DWT transform coefficients can be written as

$$d_j = \mathcal{D}_0 \mathcal{G} (\mathcal{D}_0 \mathcal{H})^{J-j-1} c_J,$$

for $j = 0, \ldots, J-1$. Similarly, for the father wavelet coefficients we have that

$$c_j = (\mathcal{D}_0 \mathcal{H})^{J-j} c_J,$$

for $j = 0, \ldots, J-1$, and where c_j and d_j are vectors of length 2^j (for periodised wavelet transforms).

Vanishing moments. Wavelet functions can be constructed to possess a number of *vanishing moments*, a key property which has several important consequences. A wavelet ψ is said to have m vanishing moments if

$$\int x^l \psi(x) dx = 0, \qquad \text{for } l = 0, \ldots, m-1.$$

Vanishing moments are useful in wavelet analysis since if a wavelet has m vanishing moments, then all wavelet coefficients of any polynomial of degree $m-1$ or less will be zero. Hence, if we have a relatively smooth function with only a few discontinuities, then the wavelet coefficients located at the smooth areas will be near zero or exactly zero if the behaviour of the function at that location is polynomial of degree $m-1$ or less. This means that wavelets are often capable of representing a signal in a sparse manner. This property is also attractive from a computational standpoint as it enables the compression of a signal into its wavelet transform.

Matrix notation. The DWT can be represented in matrix notation, which lends itself to efficient computational usage. The transform can be written as $\boldsymbol{d} = W\boldsymbol{y}$, where W is (usually) an orthogonal matrix. When W is orthogonal, we have that

$$||\boldsymbol{d}||^2 = \boldsymbol{d}^\mathsf{T}\boldsymbol{d} = (W\boldsymbol{y})^\mathsf{T}W\boldsymbol{y} = \boldsymbol{y}^\mathsf{T}W^\mathsf{T}W\boldsymbol{y} = \boldsymbol{y}^\mathsf{T}\boldsymbol{y} = ||\boldsymbol{y}||^2, \tag{2.1.11}$$

i.e. the transformation preserves energy, with (2.1.11) following from Parseval's relation.

Boundary issues. When performing a wavelet transform, a problem arises when deciding what to do when the support of the wavelet filter extends beyond the support of the input vector. Several approaches have been suggested, such as symmetric reflection of the input vector at the boundaries, polynomial extrapolation, and periodising the vector. See Section 2.8 of Nason (2008) or Section 4.11 of Percival and Walden (2006) for more discussion.

Computational speed. If the size of the input vector is n, the DWT requires $\mathcal{O}(n)$ operations, which is faster than the $\mathcal{O}(n \log n)$ operations needed for the fast Fourier transform.

Inverse discrete wavelet transform. Mallat (1989b) proved that for general wavelets, the inverse relation for the DWT is given by

$$c_{j,n} = \sum_k h_{n-2k} c_{j-1,k} + \sum_k g_{n-2k} d_{j-1,k},$$

where $\{h_k\}$ and $\{g_k\}$ are the quadrature mirror filters given in Equations (2.1.3) and (2.1.6). If the wavelet transform is orthogonal, then the matrix associated to the inverse transform is simply W^T.

2.1.3 Examples of Wavelets

In this section we describe several key examples of wavelets, many of which will be utilised in the remainder of the thesis. There are a vast array of wavelets

that can be found in both the mathematical and statistical literature; for a more complete discussion of common wavelets, see Vidakovic (2009, Chapter 3.4) and Nason (2008, Chapter 2.5 – 2.6).

The Haar wavelet. The Haar wavelet dates back to the work of Haar (1910), and is perhaps the most well-known example of a wavelet. The Haar father wavelet is defined by

$$\phi(x) = \begin{cases} 1 & x \in [0,1], \\ 0 & \text{otherwise.} \end{cases}$$

The Haar mother wavelet is given by $\psi(x) = \phi(2x) - \phi(2x-1)$, or equivalently

$$\psi(x) = \begin{cases} 1 & x \in [0, \frac{1}{2}), \\ -1 & x \in [\frac{1}{2}, 1), \\ 0 & \text{otherwise.} \end{cases}$$

The Haar basis is not always an appropriate choice of wavelet basis for several reasons. Firstly, the building blocks of the Haar basis are discontinuous functions, and so there will be difficulty in attempting to approximate smooth functions. However, this means it can be useful for modelling rapidly changing functions. Although the Haar wavelets are well localised in the time domain, they decay at a rate of $\mathcal{O}(1/n)$ in the frequency domain.

Daubechies' compactly supported wavelets. A landmark result in wavelet theory was the construction of orthogonal wavelets that were compactly supported, but smoother than Haar wavelets. Daubechies (1988) constructed wavelets using a solution of the dilation equation, giving several families of orthonormal wavelets. Each member of each family is indexed by N, which represents the number of vanishing moments. As the value of N increases, the wavelets have increasing degrees of smoothness.

Two of the most commonly used wavelet families are called the *Least Asym-*

metric (LA) and *Extremal Phase* (EP) wavelets. Only the Haar wavelet can be compact, orthonormal and symmetric (or antisymmetric) simultaneously. The least asymmetric family minimises the degree of asymmetry, whilst maintaining compactness and orthogonality. The Haar wavelet, which possesses one vanishing moment, is a member of the EP family.

The Shannon wavelet. The Shannon wavelet can be thought of as the Fourier equivalent of the Haar wavelet. It is a time-scale mirror image of the Haar wavelet. The Shannon wavelet is the limiting wavelet in the Daubechies' compactly supported series as the number of vanishing moments $N \to \infty$. The scaling function is defined in the Fourier domain as

$$\hat{\phi}(\omega) = \begin{cases} 1 & \omega \in [-\pi, \pi), \\ 0 & \text{otherwise.} \end{cases}$$

In the time domain, we have the following representation:

$$\phi(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\omega x} d\omega = \frac{\sin(\pi x)}{\pi x}.$$

In the Fourier domain, the mother wavelet is given by

$$\hat{\psi}(\omega) = -e^{-i\omega/2} \mathbb{I}(\pi < |\omega| \leq 2\pi),$$

where $\mathbb{I}$ is the indicator function. The corresponding time domain function $\psi(x)$ is

$$\psi(x) = \frac{\sin(2\pi x) - \cos(\pi x)}{\pi(x - 1/2)}.$$

Shannon's filter has poor time localisation properties: it corresponds to an infinite impulse response filter with slowly decaying coefficients, making it non-localised in time.

2.1.4 Non-Decimated Wavelets

Previously we have described how the discrete wavelet transform can be viewed as a filter operation $\mathcal{H}$ followed by a dyadic decimation step $\mathcal{D}_0$, in which we only kept the even elements. It is also possible to remove the even elements while keeping the odd ones, using odd dyadic decimation, given by the operator $\mathcal{D}_1$ where $(\mathcal{D}_1 x)_l = x_{2l+1}$. Then, the j-th mother and father wavelet coefficients would be found using the same method as before. We could, in fact, use $\mathcal{D}_0$ or $\mathcal{D}_1$ at each level of the wavelet transform to produce an orthonormal basis. Conventionally, this basis can be represented as a binary number $\epsilon = \epsilon_{J-1}\epsilon_{J-2}\cdots\epsilon_0$, where ϵ_j is equal to 1 if $\mathcal{D}_1$ was used to produce level j and 0 if $\mathcal{D}_0$ was used. This transform is known as the *ϵ-decimated wavelet transform.*

Now, note that the operator $\mathcal{D}_1$ is the same as first shifting the sequence by one position and then applying $\mathcal{D}_0$, that is $\mathcal{D}_1 = \mathcal{D}_0\mathcal{S}$, where $\mathcal{S}$ is the shift operator defined by $(\mathcal{S}x)_j = x_{j+1}$. Using an extension of this idea, along with another similar relation $\mathcal{S}\mathcal{D}_0 = \mathcal{D}_0\mathcal{S}^2$, and finally the fact that $\mathcal{S}$ commutes with the filter operations $\mathcal{H}$ and $\mathcal{G}$, Nason and Silverman (1995) show that the basis vectors of the ϵ-decimated wavelet transform can be found from those of the standard DWT by applying a particular shift operator. This leads Nason and Silverman to observe that the choice of ϵ corresponds to a particular choice of 'origin' with respect to which the basis functions are defined.

Thus, since the standard DWT is dependent on a choice of origin, a shift in position of the input data could produce a different set of wavelet coefficients compared to the original data. For uses like regression, it would be desirable for the method to not be dependent on the choice of origin, i.e. *translation invariant.*

In the non-decimated wavelet transform (NDWT), introduced by Pesquet et al. (1996), we retain *both* the odd and even decimations at each scale and proceed in this manner for each subsequent scale. Starting with the input vector $\boldsymbol{y} = (y_1, \ldots, y_n)$ where $n = 2^J$, we apply and retain both $\mathcal{D}_0\mathcal{G}\boldsymbol{y}$ and $\mathcal{D}_1\mathcal{G}\boldsymbol{y}$. Each of these sequences is of length $n/2$, so in total we have n wavelet coefficients at the

finest scale.

Similarly, we compute $\mathcal{D}_0\mathcal{H}\boldsymbol{y}$ and $\mathcal{D}_1\mathcal{H}\boldsymbol{y}$ to obtain the finest scale father wavelet coefficients. Again, we have n coefficients in total in this step. Now, for the next level wavelet coefficients we apply both $\mathcal{D}_0\mathcal{G}$ and $\mathcal{D}_1\mathcal{G}$ to both of $\mathcal{D}_0\mathcal{H}\boldsymbol{y}$ and $\mathcal{D}_1\mathcal{H}\boldsymbol{y}$. The result of each of these 4 computations is $n/4$ wavelet coefficients at scale $J-2$, once again giving n coefficients in total.

This process is repeated at each scale: at scale $J-j$ we will have 2^j sets of coefficients, each of length $2^{-j}n$ for $j = 1, \ldots, J$. Thus, the number of wavelet coefficients at each scale is always $2^j \times 2^{-j}n = n$. There are J scales, so in total the NDWT produces Jn coefficients. The computational speed of the NDWT is $\mathcal{O}(n \log n)$. For further information, see also Nason and Silverman (1995) and Coifman and Donoho (1995).

2.1.5 Other Extensions of Wavelets

Since the inception of wavelets, there have been many extensions to the classical wavelet techniques we have described so far. In this section we provide a brief discussion of some of the most well-known extensions. Further information and a broader discussion can be found, for example, in Nason (2008) and Vidakovic (2009).

Mallat (1989a) constructs *multidimensional wavelets*, capable of performing multiscale analysis in spaces of higher dimensions. The DWT is extended to a multivariate version, by constructing wavelets $\psi \in L^2(\mathbb{R}^d)$. Another extension to classical wavelets is the *biorthogonal* wavelets of Cohen et al. (1992). Biorthogonal wavelets use different non-orthogonal wavelets for the decomposition and reconstruction steps. These wavelets are however, in some sense, mutually orthogonal.

Coifman and Wickerhauser (1992) introduced the concept of wavelet *packets*: redundant collections of linear combinations of wavelet functions that provide a generalisation of standard wavelet bases. The entire collection is referred to as a *library* of packet functions. Wavelet packets allow for more flexibility than a classi-

cal wavelet and hence can provide more efficient decompositions. When performing a wavelet packet transform, both the low and high pass filters are applied at each stage of the decomposition, to both the scaling and wavelet coefficients. Coifman and Wickerhauser (1992) also develop methodology for searching to find the *best basis* representation within a given library. Lawton (1993) derived *complex-valued wavelets*, while Lina and Mayrand (1995) provides a comprehensive discussion of complex-valued Daubechies' wavelets. Complex-valued wavelet analysis can extract useful information that is potentially missed when using traditional real-valued wavelet techniques.

The *lifting scheme*, proposed by Sweldens (1995), addresses a limitation of classical wavelets to allow for multiscale transforms of irregularly sampled data. This transform is composed of two steps, involving prediction and updating. For particular filters, the lifting scheme can be seen as a generalisation of the (bi)orthogonal DWT or the wavelet packet transform. Wavelet functions obtained via lifting are known as second generation wavelets; unlike standard wavelets they are not necessarily translates and dilates of the same function. Wavelet lifting has also been applied to the problem of missing observations of time series; see for example Knight et al. (2012).

2.2 Time Series Analysis

In this thesis, wavelets are to be used as tools in time series analysis, and so in this section we describe some of the basic concepts of the subject. Time series are a class of stochastic processes that consist of observations made sequentially over time. A discrete time series is a collection of random variables $\{X_t, t \in T\}$, which has been measured at discrete, successive time points. Often, the time series is observed on a uniformly spaced interval. A key feature of time series data is that the observations are, in general, not independent.

2.2.1 Basic Concepts in Time Series Analysis

Commonly, the first and second moments are used to describe a time series. The *mean function* is given by $\mu_t = \mathbb{E}(X_t)$ provided it exists, where $\mathbb{E}$ is the expected value operator. The degree of dependence between two values in the time series can be found, as in classical statistics, using the notions of covariance and correlation. Without loss of generality, assuming a zero-mean time series and finite variance of X_t, the autocovariance function is defined as

$$\gamma_t(\tau) = \mathrm{Cov}(X_t, X_{t+\tau}) = \mathbb{E}\left(X_t X_{t+\tau}\right),$$

The autocorrelation function (ACF) is given by

$$\rho_t(\tau) = \frac{\gamma_t(\tau)}{\mathrm{Var}(X_t)}.$$

The autocovariance measures the linear dependence between two points on a time series observed at different times. If a time series is very smooth, then the autocovariance will stay large even for large values of τ. On the other hand, if a time series is choppy then the autocovariance will usually be close to zero for far apart observations. Even if $\gamma_t(\tau) = 0$, there may still be some dependence between X_t and $X_{t+\tau}$, although not linear. When $\tau = 0$, the autocovariance reduces to the variance function.

2.2.2 Stationary Time Series

It is necessary to impose assumptions on the evolution of a time series in order to facilitate inference on its behaviour. A large proportion of the literature is concerned with *stationary* time series. Informally a stationary times series is one whose statistical properties do not change over time. More formally, a *strictly stationary* time series is one where the joint distribution of $(X_{t_1}, \ldots, X_{t_n})$ is the same as $(X_{t_1+\tau}, \ldots, X_{t_n+\tau})$ for all t_i, n and τ. Strict stationarity is usually too

strong an assumption in practice, which gives rise to a weaker form of stationarity. A time series is said to be *second-order* or *weakly stationary* if $\mathbb{E}(X_t) = \mu$ and the autocovariance $\gamma(\tau) = \text{cov}(X_t, X_{t+\tau})$ is a function of τ only. Employing a slight abuse of language, henceforth we will refer to a weakly stationary time series as stationary.

Any weakly stationary zero-mean discrete time series can be decomposed via the well-known Cramér representation

$$X_t = \int_{-\pi}^{\pi} A(\omega) \exp(i\omega t) dZ(\omega), \tag{2.2.1}$$

where $A(\omega)$ is the amplitude, and $Z(\omega)$ is an orthonormal increment process, i.e. $\mathbb{E}(dZ(\omega_1)\overline{dZ(\omega_2)}) = \delta_{\omega_1,\omega_2}$. The parameter ω can be interpreted as frequency, in that X_t is given by a weighted linear combination of Fourier trigonometric functions that oscillate at different frequencies. The spectrum, or spectral density function $f(\omega) := |A(\omega)|^2$, is a measure of the 'amount' of oscillation at different frequencies $\omega \in (-\pi, \pi)$. In particular, the term $f(\omega)d\omega$ is the contribution to the total variance of X_t for frequencies in the range $(\omega, \omega + d\omega)$. The autocovariance function of X_t is related to the spectrum by the following Fourier formula

$$\gamma(\tau) = \int_{-\pi}^{\pi} f(\omega) \exp(i\omega\tau) d\omega.$$

There is a vast array of time series models in the literature. We briefly describe ARMA (Autoregressive Moving Average) processes, which are arguably the most ubiquitous time series models. An ARMA(p, q) process is defined as

$$X_t = \sum_{j=1}^{p} \alpha_j X_{t-j} + \sum_{i=1}^{q} \beta_i \epsilon_{t-i} + \epsilon_t,$$

where the ϵ_t are independent, identically distributed zero-mean random variables, often assumed to be Gaussian. An ARMA(p, q) process is stationary provided that

the polynomial

$$\alpha(z) = 1 - \alpha_1 z - \ldots - \alpha_p z^p$$

has no roots in the closed unit disk. ARMA models are examples of linear models, in which X_t and its random innovations ϵ_t are related through a linear transformation. For a detailed discussion on common time series models, see for example Brockwell et al. (1991) and Andersen et al. (2009).

2.2.3 Nonstationary Time Series

In the Cramér representation of Equation (2.2.1), the amplitude $A(\omega)$ does not depend on time, and so the amplitude of the process remains constant across all time. This means that the frequency behaviour of the time series is the same for all time. For many time series in practical applications, this will not be true, and so a stationary time series model will not suffice. One solution is to only analyse small intervals of the observed data, upon which the assumption of stationarity may hold. However, this raises questions such as the length of interval used, and can compromise estimation accuracy. Another common approach is to induce stationarity by differencing the data. This is the case in ARIMA (Autoregressive Integrated Moving Average) modelling, where "Integrated" refers to the degree of differencing necessary to induce stationarity.

There are several ways in which model (2.2.1) can be adapted to introduce time dependence. For example, the amplitudes, $A(\omega)$, can be replaced with a time-dependent version $A_t(\omega)$, as in Priestley (1965) and Dahlhaus (1997). A method of this form results in a time-frequency model. Dahlhaus (1997) introduces the class of locally stationary Fourier (LSF) processes, in which the process X_t is modelled as a triangular stochastic array $\{X_{t,T}\}_{t=0}^{T-1}$, such that

$$X_t = \int_{-\pi}^{\pi} A^0_{t,T}(\omega) \exp(i\omega t) dZ(\omega). \tag{2.2.2}$$

Furthermore, there exists a transfer function $A : [0,1] \times \mathbb{R} \to \mathbb{C}$, continuous in the

first argument, and 2π-periodic in the second, such that

$$\sup_{t,\omega} \left| A^0_{t,T}(\omega) - A(t/T, \omega) \right| \leq \frac{K}{T}, \tag{2.2.3}$$

where K is a positive constant. Asymptotic analysis is made difficult in a nonstationary setting since future observations of the process do not provide information on its current structure. Dahlhaus (1997) addresses this problem through the notion of *rescaled time*, $z = t/T \in (0, 1)$. Modelling the amplitudes as a function of rescaled time facilitates meaningful asymptotic theory. As the length of the time series increases, more information about the local structure of $A(z, \omega)$ is revealed.

There are a number of other approaches in the literature for modelling nonstationary time series. Van Bellegem and Dahlhaus (2006) discuss nonstationary autoregressive models, where the parameters are allowed to vary with time. Ombao et al. (2002) describe an alternative representation for nonstationary time series in terms of SLEX (Smoothed Localised Complex Exponential) functions, that can be viewed as localised versions of the Fourier exponential functions. The time series is segmented into approximately stationary intervals in a data-driven fashion, via the Auto-SLEX method (Ombao et al., 2001). This allows for estimation of the time-varying spectra. For an overview of nonstationary time series modelling, see Dahlhaus (2012).

2.3 Locally Stationary Wavelet Processes

The time series models considered in Chapters 3 – 5 of this thesis are based upon the locally stationary wavelet (LSW) model, introduced in Nason et al. (2000). In this section, we describe the LSW model and discuss various related quantities. The LSW model follows Dahlhaus (1997) in that it adopts the use of rescaled time, but differs by replacing the Fourier basis representation with a non-decimated wavelet basis representation. First, we describe discrete non-decimated wavelets which will form the building blocks for LSW processes.

2.3.1 Discrete Non-Decimated Wavelets

Let $\{h_k\}$ and $\{g_k\}$ be the low and high-pass quadrature mirror filters that are used in the construction of Daubechies compactly supported wavelets. Nason et al. (2000) recursively construct the compactly supported discrete wavelets $\psi_j = (\psi_{j,0}, \ldots, \psi_{j,N_j-1})$ of length N_j for scale $j < 0$ by using the following equations:

$$\psi_{-1,n} = \sum_{k\in\mathbb{Z}} g_{n-2k}\delta_{0,k} = g_n, \text{ for } n = 0, \ldots, N_{-1} - 1,$$

$$\psi_{j-1,n} = \sum_{k\in\mathbb{Z}} h_{n-2k}\psi_{j,k}, \text{ for } n = 0, \ldots, N_{j-1} - 1,$$

$$N_j = (2^{-j} - 1)(N_h - 1) + 1,$$

where N_h is the number of non-zero elements of $\{h_k\}$. Note here that we adopt the convention whereby the scale j takes negative values. In this setting, the data can be viewed as being at scale 0, while the finest level of detail is given by scale -1, and coarser scales given by smaller values of j. For example, the discrete Haar wavelets at scales -1 and -2 are given by

$$\psi_{-1} = (g_0, g_1) = \frac{1}{\sqrt{2}}(1, -1), \qquad \psi_{-2} = (h_0g_0, h_1g_0, h_1g_0, h_1g_1) = \frac{1}{2}(1, 1, -1, -1),$$

and so on for coarser scales:

$$\psi_j = 2^{j/2}(\underbrace{1, 1, \ldots, 1}_{2^{-j-1} \text{ times}}, \underbrace{-1, -1, \ldots, -1}_{2^{-j-1} \text{ times}}).$$

The discrete wavelets are the same as the vectors produced by Daubechies' cascade algorithm used for producing discrete approximations to continuous-time wavelets at successively finer scales. We further define from Nason et al. (2000) the autocorrelation wavelets, which appear in the representation of the autocovariance

functions of LSW processes. The *autocorrelation wavelets* (ACW) are defined by

$$\Psi_j(\tau) := \sum_{k\in\mathbb{Z}} \psi_{j,k}\psi_{j,k-\tau},\ j < 0, \tau \in \mathbb{Z}.$$

The autocorrelation wavelets average the discrete wavelets over all locations within one scale j, and form a family of symmetric, compactly supported and positive definite functions on $\tau \in \mathbb{Z}$. The autocorrelation wavelets, although not orthogonal, satisfy the following three important properties: $\Psi_j(0) = 1$, $\sum_\tau \Psi_j(\tau) = 0$ for all j, and $\sum_j 2^j \Psi_j(\tau) = \delta_{\tau 0}$. The discrete autocorrelation wavelets are related to their continuous counterparts by the formula

$$\Psi_j(\tau) = \Psi(2^j|\tau|), \tag{2.3.1}$$

where the continuous autocorrelation wavelet $\Psi(u)$ is defined by

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

where $u \in \mathbb{R}$ and the support of $\Psi(u)$ is $\mathbb{R}$. Equation (2.3.1) holds for all Daubechies compactly supported wavelets. Lastly, we define the autocorrelation wavelet inner product matrix A, which will be used later in the estimation of LSW process quantities. Define the operator $A = (A_{jl})_{j,l<0}$ by

$$A_{jl} = \langle \Psi_j, \Psi_l \rangle = \sum_\tau \Psi_j(\tau)\Psi_l(\tau),$$

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

2.3.2 The LSW Model

The LSW model utilises wavelets to decompose a stochastic process with respect to an orthonormal increment process in the time-scale domain. Following the original LSW model in Nason et al. (2000) and the adaptations described in Fryzlewicz

(2003), the non-decimated wavelets described above are used to define the following time series model.

Definition 2.3.1. An *LSW process* $\{X_{t,T}\}$, $t = 0, \ldots, T-1$, and $T = 2^J \geq 1$ is a doubly-indexed stochastic process with the following representation in the mean-square sense:

$$X_{t,T} = \sum_{j=-\infty}^{-1} \sum_{k \in \mathbb{Z}} w_{j,k;T} \psi_{j,k-t} \xi_{j,k}, \tag{2.3.2}$$

where $\{\xi_{j,k}\}$ is a random orthonormal increment sequence, $\{\psi_{j,k-t}\}_{j,k}$ is a set of discrete non-decimated wavelets, and $\{w_{j,k;T}\}_{j,k}$ is a set of amplitudes. The quantities in representation (2.3.2) possess the following properties:

1. $\mathbb{E}(\xi_{j,k}) = 0$ for all j, k.

2. $\mathrm{Cov}(\xi_{jk}, \xi_{lm}) = \delta_{jl}\delta_{km}$.

3. There exists, for each $j \leq -1$, a Lipschitz continuous function $W_j(z)$ for $z \in (0,1)$ which satisfies the following properties:

$$\sum_{j=-\infty}^{-1} |W_j(z)|^2 < \infty \text{ uniformly in } z \in (0,1). \tag{2.3.3}$$

The Lipschitz constants L_j are uniformly bounded in j and

$$\sum_{j=-\infty}^{-1} 2^{-j} L_j < \infty.$$

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

$$\sup_k \left| w_{j,k;T} - W_j\left(\frac{k}{T}\right) \right| \leq \frac{C_j}{T}, \tag{2.3.4}$$

where for each j the supremum is over $k = 0, \ldots, T-1$, and where the sequence $\{C_j\}$ satisfies

$$\sum_{j=-\infty}^{-1} C_j < \infty.$$

This representation can be thought of as building a time series model $X_{t,T}$ out of a linear combination of oscillatory functions $\psi_{j,k}$ with random amplitudes $w_{j,k;T}\xi_{j,k}$, which can be viewed as a multiscale version of the stationary time series definition.

The key assumption of the model lies in requiring that $\mathbb{E}(\xi_{j,k}) = 0$, which in doing so implies that $\mathbb{E}(X_{t,T}) = 0$. In most practical situations it is simply not the case that the time series is zero-mean, and so the data must be preprocessed in some way to convert it to a zero-mean time series in order to make it amenable to the LSW process representation. This can be achieved, for example, by detrending the time series using wavelet-based techniques, such as in von Sachs and MacGibbon (2000). However, this is far from straightforward as it requires a pre-estimate of the nonstationary second-order structure. We will discuss in Chapters 3 – 5 our proposed methodology of simultaneously considering first and second-order structure within the same methodological framework, building upon the LSW model.

2.3.3 The Evolutionary Wavelet Spectrum

In the classical stationary setting, the spectrum is the usual summary statistic used for describing oscillatory behaviour. Its wavelet analogue, the *evolutionary wavelet spectrum* (EWS), measures the local power (contribution to variance) in an LSW process at a particular rescaled time $z = k/T$ and scale j. It is defined as

$$S_j(z) = |W_j(z)|^2.$$

Since the W_j are assumed to be Lipschitz continuous, the spectrum S_j is also Lipschitz continuous, which ensures it evolves slowly over time. The local autocovariance (LACV) function for an LSW process provides information about the autocovariance at a rescaled location $z = k/T \in (0, 1)$. In stationary time series analysis the spectrum is the Fourier transform of the autocovariance function; in

LSW analysis we obtain an analogous relationship between the EWS and LACV. The *local autocovariance* (LACV), $c(z, \tau)$, is defined as

$$c(z, \tau) = \sum_{j=-\infty}^{-1} S_j(z)\Psi_j(\tau), \ \tau \in \mathbb{Z}, z \in (0, 1). \tag{2.3.5}$$

Equation (2.3.5) is a decomposition of the autocovariance of a process over scales and rescaled time locations. In Nason et al. (2000), it was shown that the error in approximating the true autocovariance of the process with the LACV was $\mathcal{O}(T^{-1})$. Furthermore, the representation in Equation (2.3.5) is invertible:

$$S_j(z) = \sum_l A_{jl}^{-1} \sum_\tau c(z, \tau)\Psi_l(\tau),$$

provided that the operator A is invertible and possesses a bounded inverse. In the original work of Nason et al. (2000), it was shown that the operator A possesses a bounded inverse in the case where A is constructed using either Haar or Shannon wavelets. The authors also conjectured that the general result holds for all Daubechies compactly supported wavelets. In Chapter 4, we affirmatively answer the conjecture, by adapting techniques from Cardinali and Nason (2017), in which the authors state that bounded invertibility holds for all wavelet packet-based operators. We correct a mistake in the authors' proof and make a crucial alteration in order to show bounded invertibility for all Daubechies compactly supported wavelets.

2.3.4 Example

Next, we briefly describe an example to illustrate some properties of the LSW model. We define the following evolutionary wavelet spectrum:

$$S_j(z) = \begin{cases} \sin^2(4\pi z) & \text{for } j = -5, z \in (0,1), \\ 1 & \text{for } j = -1, z \in (800/1024, 900/1024), \\ 0 & \text{otherwise.} \end{cases} \tag{2.3.6}$$

In Figure 2.3.1, we see the spectrum plot of the EWS in Equation 2.3.6. The way to interpret the plot is that from the bottom to the top, we go from fine scales to coarser scales. The finest scale is given by scale -1 at the bottom of the plot, with coarser scales above. This spectrum contains a burst of power at the finest scale of the spectrum, and slowly-evolving, oscillatory behaviour at a coarser scale.

We simulate a realisation from this spectrum in order to highlight the behaviour of the time series. The simulation is achieved by constructing the coefficients $w_{j,k}\xi_{j,k}$, where $w_{j,k} = S_j(k/T)^{1/2}$ and the $\xi_{j,k}$ are Gaussian with mean zero and variance one. We see in Figure 2.3.2 a realisation from this process. The sinusoidal nature of the spectrum at a fairly coarse scale can be seen in the plot and the very high frequency behaviour due to the burst between 800 and 900 is also apparent.

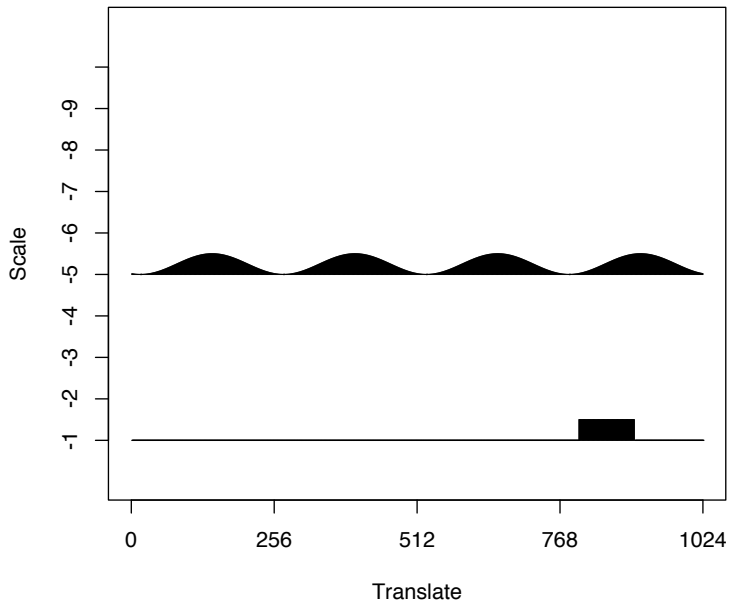

Figure 2.3.1: Square sine at scale -5, with a burst at 800 at scale -1, as defined in Equation (2.3.6).

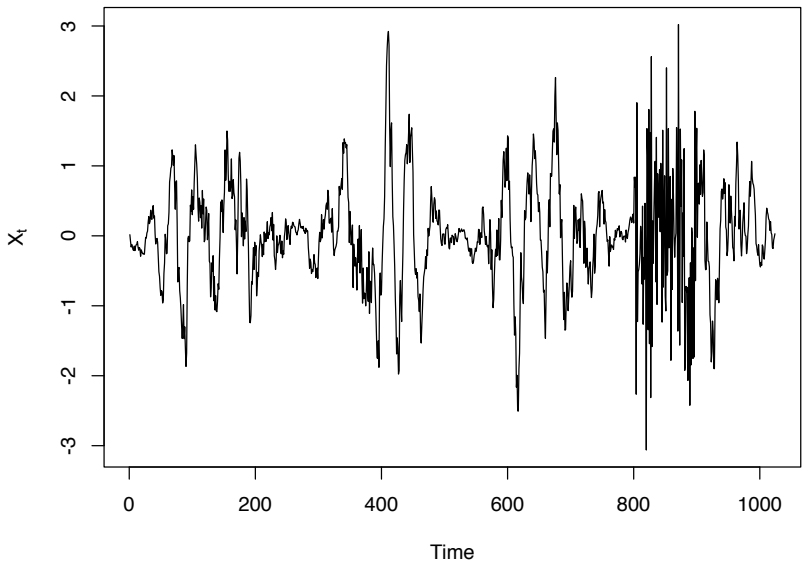

Figure 2.3.2: A realisation from the square sine and burst at 800 spectrum, shown in Figure 2.3.1.

2.3.5 Estimation of the Evolutionary Wavelet Spectrum

In order to estimate the EWS, we begin by taking a non-decimated wavelet transform of the LSW process $X_{t,T}$. This is motivated by the fact that the definition of an LSW process suggests that the wavelet transform of the amplitude $w_{j,k}$ is approximately $X_{t,T}$ (with error). The *empirical wavelet coefficients* of an LSW process $X_{t,T}$ are given by

$$d_{j,k;T} := \langle X_{t,T}, \psi_{j,k-t} \rangle = \sum_t X_{t,T} \psi_{j,k-t}. \tag{2.3.7}$$

This transform produces a matrix of coefficients with dimension $J \times T$. We can then estimate the EWS of an LSW process using the *raw wavelet periodogram*, which is given by

$$I^j_{k,T} := |d_{j,k;T}|^2,$$

where $j = -1, \ldots, -J$, $k = 0, \ldots, T-1$. The raw wavelet periodogram of an LSW process $X_{t,T}$ is the squared modulus of the non-decimated empirical wavelet coefficients, and measures the "power" of the time series at location k and scale j. It is a biased, inconsistent estimate of the EWS:

Proposition 2.3.1.

$$\mathbb{E}\left(I^j_{\lfloor zT\rfloor,T}\right) = \sum_l A_{jl} S_l(z) + \mathcal{O}(T^{-1}). \tag{2.3.8}$$

If $X_{t,T}$ is Gaussian, then

$$\mathrm{Var}(I^j_{\lfloor zT\rfloor,T}) = 2\left(\sum_l A_{jl} S_l(z)\right)^2 + \mathcal{O}(2^{-j}T^{-1}). \tag{2.3.9}$$

Hence, for the vector of periodograms $\mathbf{I}(z) := \{I^l_{\lfloor zT\rfloor,T}\}_{l=-1,\ldots,-J}$, and the vector of corrected periodograms $\mathbf{L}(z) := \{L^j_{\lfloor zT\rfloor,T}\}_{j=-1,\ldots,-J}$ with $\mathbf{L}(z) = A_J^{-1}\mathbf{I}(z)$,

$$\mathbb{E}\left(\mathbf{L}(z)\right) = \mathbb{E}\left(A_J^{-1}\mathbf{I}(z)\right) = \mathbf{S}(z) + \mathcal{O}(T^{-1}) \qquad \forall\ z \in (0,1), \tag{2.3.10}$$

where $\mathbf{S}(z) := \{S_j(z)\}_{j=-1,\ldots,-J}$.

The bias is due to the redundancy of the non-decimated wavelet transform, and is corrected by premultiplication by A_J^{-1}. The variance does not vanish as the sample size T tends to infinity, and hence smoothing must be performed in order to obtain a consistent estimate of the spectrum. Nason et al. (2000) use a wavelet thresholding approach for smoothing, while a running mean smoother can also be employed; see Nason (2013). In practice, the raw wavelet periodogram is first smoothed, then corrected, to obtain the final estimate of the EWS.

There are several associated techniques described in the literature for estimation of evolutionary wavelet spectra. Fryzlewicz and Nason (2006) propose an estimation strategy involving the Haar-Fisz transform, for LSW processes with piecewise constant EWS. Van Bellegem and von Sachs (2008) discuss an adaptive procedure for estimation of the EWS of LSW processes whose EWS is subject to a total variation assumption. Nason and Stevens (2015) use Bayesian wavelet shrinkage to estimate the EWS under the same assumptions as Van Bellegem and von Sachs (2008). The LACV function can be estimated by substituting the consistent spectral estimate into Equation (2.3.5), and replacing the infinite sum with

a finite one with lower limit $-J_0$, where $2^{J_0} = o(T)$. The resulting estimator is shown to be mean square consistent in Nason et al. (2000).

Lastly, we note that the LSW model is but one possible representation for a locally stationary time series. Depending on the application at hand, the appropriate type of locally stationary model to use will vary. For example, in situations where the second-order structure evolves rapidly, LSW processes will work well due to the wavelet being compactly supported. If, however, the data application is known to posses certain cyclic behaviour, such as in the case of some environmental time series where certain physical processes drive the variability, then a locally stationary Fourier approach may be more appropriate. It could also be the case that a locally stationary model is too complicated for the problem, and an approach assuming stationary is more appropriate. Indeed, principled methodology for model selection in locally stationary time series remains an open problem.

2.4 Nonparametric Regression/Trend Estimation

As discussed in the introduction, the LSW framework does not incorporate first-order behaviour within the model. To that end, it is necessary to review some ideas from nonparametric regression and trend estimation; most importantly wavelet-based techniques. We aim to estimate a function $g : [0, 1] \to \mathbb{R}$, given noisy observations y_i, observed on an equally spaced grid:

$$y_i = g(i/n) + \epsilon_i, \quad i = 1, \ldots, n, \tag{2.4.1}$$

where the ϵ_t are noise with $\mathbb{E}(\epsilon_t) = 0$. There are a number of subclasses of this problem, depending on the nature of the noise ϵ_t and the smoothness of the function g. Often, the area of trend estimation is concerned with estimation of smooth functions, while in general in nonparametric regression the function is not necessarily constrained to be smooth.

2.4.1 Wavelet-Based Methods

Donoho (1993, 1995), Donoho and Johnstone (1994, 1995) and Donoho et al. (1995) introduced the concept of wavelet shrinkage in their seminal papers. The main idea of their work is that we can apply the DWT to the model (2.4.1), to obtain $d_{j,k} = \theta_{j,k} + \epsilon_{j,k}$, where $d_{j,k}$, $(\theta_{j,k}, \epsilon_{j,k})$ is the DWT of y_i, $(g(i/n), \epsilon_i)$.

The motivation behind this approach is threefold. Firstly, when the function g is relatively smooth, the wavelet coefficients will be close to zero. Secondly, by Parseval's relation, the energy in the function domain is equal to the sum of squares of wavelet coefficients. Combining this fact with the sparsity of the wavelet coefficients, this means that the energy of the original signal g is often strongly concentrated into fewer wavelet coefficients. Thus, relative to the noise variance the values of the coefficients are often larger. Finally, since the wavelet transform is orthogonal, the wavelet transform of white noise is also white noise.

Donoho and Johnstone (1994) use these ideas in their wavelet shrinkage method to estimate g. The idea of the method is that the large values of the empirical wavelet coefficients are the ones that are most likely to contain true signal and noise, whereas the small coefficients are only the result of noise. Donoho and Johnstone (1994) define the following hard and soft thresholding functions

$$\hat{d}^H_{j,k} = \eta_H(d_{j,k}, \lambda) = d_{j,k}\mathbb{I}(|d_{j,k}| > \lambda),$$
$$\hat{d}^S_{j,k} = \eta_S(d_{j,k}, \lambda) = \text{sgn}(d_{j,k})(|d_{j,k}| - \lambda)\mathbb{I}(|d_{j,k}| > \lambda),$$

where $\mathbb{I}$ is the indicator function, $d_{j,k}$ is the empirical coefficient to be thresholded, and λ is the threshold. Donoho and Johnstone (1994) define the *universal threshold*, one of the most widely used thresholds, by $\lambda^u = \sigma\sqrt{2\log n}$. In practical situations, the noise level σ must be estimated by $\hat{\sigma}$, an estimate of the common standard deviation of the noise ϵ_i. Donoho and Johnstone (1994) suggest estimating σ using the median absolute deviation (MAD) of the finest-scale wavelet coefficients, which ensures robustness.

There are many other threshold functions in the literature. For example, in Donoho (1995), the authors introduce the Stein's Unbiased Risk Estimator (SURE) threshold, considering a soft thresholding procedure based on Stein's shrinkage method for estimating the mean of multivariate normal random variables. Gao and Bruce (1997) use firm shrinkage, while it is also possible to take a Bayesian approach to thresholding: see for example Johnstone and Silverman (2005). Nason (1996) employs cross-validation in order to select the threshold. Cai and Silverman (2001), among others, consider a block thresholding approach in which wavelet coefficients are considered in overlapping blocks. Cai and Brown (1998) and Sardy et al. (1999), for example, consider wavelet shrinkage in the case of unequally spaced data.

Wavelet thresholding has been applied successfully in a variety of other scenarios. For example, Neumann and von Sachs (1995) consider thresholding in the case where the error structure is non-Gaussian and non-IID. Johnstone and Silverman (1997) propose a variety of threshold choices for data with correlated noise. Neumann and von Sachs (1997) and von Sachs and Schneider (1996) propose thresholding estimators of time-varying spectra where the errors are assumed to follow the locally stationary time series model of Dahlhaus (1997).

Coifman and Donoho (1995) introduce the concept of translation invariant (TI) denoising, in which the non-decimated wavelet transform is used in place of the standard decimated DWT. The non-decimated wavelet coefficients are then shrunk via the universal threshold, and an inverse NDWT is performed to obtain the estimate of the function g. Since the NDWT is redundant, there are many possible methods of performing the inverse NDWT transform. Coifman and Donoho (1995) propose a method that is equivalent to taking the average over all possible DWTs contained in the NDWT. This corresponds to the DWT of all possible circular shifts of the data, which inspires the name "translation invariant". The inverse transform can also be performed using a single basis, selected using, for example, the best-basis algorithm of Coifman and Wickerhauser (1992).

2.4.2 Other Estimation Methods

There is a vast array of non wavelet-based methods in the statistical literature concerning the well studied problem of nonparametric regression. A common technique is to use kernel smoothing to estimate the unknown function g. Examples of well-known kernel functions include the Nadaraya-Watson kernel estimator (Nadaraya, 1964; Watson, 1964), the Epanechnikov kernel (see for example Benedetti (1977)), and the Preistley-Chao kernel estimator (Priestley and Chao, 1972).

Another well studied technique is local polynomial regression, in which polynomials of some prescribed order that are localised are fit to the data – see for example Fan and Gijbels (1996). For generality assume that the data y_i are observed at not necessarily equally spaced values x_i. Local polynomial estimators use weight functions W to fit polynomials of order p, according to the weighted least squares criterion

$$\sum_{i=1}^{n} \left(y_i - \beta_0 - \ldots - \beta_p (x - x_i)\right)^2 W\left(\frac{x - x_i}{h}\right). \tag{2.4.2}$$

The estimator $\hat{g}$ is the minimiser of Equation (2.4.2). The bandwidth parameter h specifies the smoothing window and affects the smoothness of the resulting estimate of g.

Another common method is spline smoothing, which trades off fidelity to the data with the smoothness of the estimate in order to estimate the unknown function g. A cubic smoothing spline estimator $\hat{g}$ is defined as the function that minimises the equation

$$\frac{1}{n} \sum_{i=1}^{n} \left(y_i - g(x_i)\right)^2 + \alpha \int g''(x)^2 dx.$$

Spline functions are piecewise polynomials that are smooth and satisfy continuity constraints at the knots $\{x_i\}$ joining their pieces; for detailed discussion see for example Green and Silverman (1993).

At their inception, all of the aforementioned methods were based on the assumption that the error processes are independent, which is unlikely to be true when considering time series data. This correlation in the errors will have an effect on the asymptotic properties of the estimators. The methods have since been extended to accommodate more general error structures. For example, for spline smoothing, Wang (1998) considers Gaussian autocorrelated errors, while Xiao et al. (2003) perform local polynomial estimation by first whitening the data to remove the autocorrelation. For detailed discussion of nonparametric smoothing methods, see for example Härdle (1991) or Simonoff (2012).

The trend of a time series may also be modelled as a stochastic component, as opposed to a deterministic one. One of the most extensively studied methods which can be used to incorporate a stochastic trend component is the use of dynamic linear models (DLMs). The idea is that the observations depend on an underlying unobserved state, which is assumed to follow a well-defined stochastic process. The Kalman filter approach utilises this idea and can be used to model a trend by assuming a certain form for the unobserved state. For a detailed discussion of DLMs, see for example Chapter 6 of Shumway and Stoffer (2010).

2.4.3 Joint Consideration of First and Second-Order Nonstationarity

Up until now, the literature that has been cited has been concerned with nonstationary first *or* second-order structure estimation. There is, however, an emerging literature related to inference in the case that *both* first and second-order quantities are assumed to be nonstationary in some manner. We devote the rest of the section to briefly surveying these methods.

von Sachs and MacGibbon (2000) construct a wavelet thresholding estimator of the first-order structure of a time series that possesses locally stationary, not necessarily Gaussian errors. The authors derive the asymptotic properties of the empirical wavelet coefficients under mild, general assumptions, showing that at ap-

propriate scales the wavelet coefficients are asymptotically Gaussian. This permits the use of a time-dependent universal threshold of the form $\lambda_{j,k} = \sigma_{j,k}\sqrt{2\log(2^j)}$, where $\sigma_{j,k}$ is the noise level of the empirical wavelet coefficient $d_{j,k}$.

Therefore, a pre-estimate of the variance of the wavelet coefficients is required in order to perform the thresholding procedure. This can potentially hamper the practical performance of the method, since the locally stationary second-order structure of the time series is unknown. The authors propose to use a hybrid local MAD type method in order to estimate the unknown variance. First, a linear Haar wavelet estimator is fitted to the squared wavelet coefficients $d_{j,k}^2$. Then, the MAD of the squared wavelet coefficients is calculated in each region of the Haar wavelet estimator that is constant, giving the final estimate. All coefficients within the same constant block with respect to the Haar estimator have the same variance estimate, and so the threshold can be viewed as a locally universal threshold.

Vogt (2012) considers the problem of nonparametric regression in the presence of locally stationary errors. The trend, assumed to be relatively smooth, is estimated using kernel-based methods, while the error structure can be nonlinear. Krampe et al. (2015) use a bootstrap approach to estimate a trend function that satisfies Hölder continuity type conditions, in the presence of linear locally stationary errors.

Vogt and Dette (2015) propose nonparametric methodology to detect gradual changes in statistical properties of locally stationary processes, including the mean, autocovariance, and higher moments. Dette and Wu (2019) test for relevant changes in the mean of nonstationary processes, while Wu and Zhou (2020) test for abrupt changes in smooth mean functions in the presence of locally stationary errors. Das and Politis (2020) and Dette and Wu (2020) consider the problem of prediction in time series that can exhibit trend and locally stationary errors.

Beran and Feng (2002) propose parallel first and second-order estimation using a semiparametric approach, in which the second-order structure follows a parametric model and is assumed to be difference-stationary. Tunyavetchakit (2010)

examines mean estimation for the case of time-varying autoregressive processes. Ferreira et al. (2013) perform Kalman filtering for models that display time-varying trends and errors belonging to a class of locally stationary processes.

In this thesis, we consider time series that display a non-trivial mean function and LSW process errors, a problem that has yet to be studied in the literature. Further discussion on the topic of joint first and second-order nonstationary time series can be found in the introductory sections of Chapters 3 – 5, and can also be found within Dahlhaus (2012).

Chapter 3

Modelling Time-Varying First and Second-Order Structure of Time Series via Wavelets and Differencing

3.1 Introduction

Time series data can often possess complex and dynamic characteristics. Most commonly encountered time series in practice are nonstationary – the mean and autocovariance of the series vary over time. Modelling how these properties change over time is crucial for making inference on the data. Nonstationary time series arise in a variety of applications, for example in environmental sciences (Hu et al., 2019), climatology (Das and Politis, 2020), and financial time series (Roueff and von Sachs, 2019). In this chapter, we consider a time series model of the form

$$X_{t,T} = \mu\left(\frac{t}{T}\right) + \epsilon_{t,T}, \quad 0 \leq t < T, \tag{3.1.1}$$

where $\mu : [0,1] \to \mathbb{R}$ is a non-parametric deterministic trend function, and $\epsilon_{t,T}$ is a locally stationary wavelet (LSW) process with $\mathbb{E}(\epsilon_{t,T}) = 0$. This model accounts for nonstationarity in *both* the first and second-order structure of the time series: the time-varying mean function is encapsulated in the trend term, while the time-

varying second-order behaviour is described by the LSW process term. A thorough explanation of Model (3.1.1), including necessary background on LSW processes, is given in Section 3.2.

First and second-order estimation of a time series are most commonly performed in isolation, rather than in parallel within the one common methodological framework. However, there is an emerging literature where the mean function is estimated as well as parameters responsible for second-order nonstationarity. For example Tunyavetchakit (2010) consider time varying AR(p)-processes where the mean curve is estimated in parallel. Other methods, such as Dahlhaus and Neumann (2001) which focuses on the semi-parametric setting, consider results where the mean function is time-varying and/or estimated. Khismatullina and Vogt (2020) test for increases and decreases in trend in the presence of stationary time series errors, while Dette and Wu (2020) consider the problem of prediction in locally stationary time series.

The problem of performing inference on the mean function in the presence of nonstationary second-order structure is a highly challenging one. In von Sachs and MacGibbon (2000), for example, the authors describe a method using wavelet thresholding, however the threshold used is data-driven and depends on the nonstationary second-order structure, which is ultimately unknown. Vogt (2012) employs kernel-based methods to estimate a smooth non-parametric trend function in the presence of locally stationary errors. Dette and Wu (2019) test for relevant changes in the mean of nonstationary processes. Similarly, there is less attention in the literature on nonstationary second-order estimation in the presence of a non-trivial mean function.

In order to estimate a nonstationary second-order structure a zero-mean process is often assumed. One of the most well-known methods for removing the trend in a time series is differencing: see for example Chan et al. (1977) and Shumway and Stoffer (2010). The time series $\{X_t\}$ can be, for example, first-differenced to obtain a new time series, $\{\nabla X_t = X_t - X_{t-1}\}$. Differencing aims to remove a trend

without the need to estimate any parameters (whose estimation often includes an assumption of stationary errors). For example, Hart (1989) utilises differencing to estimate the stationary covariance function, while Dai et al. (2015) propose a difference-based variance estimator, both within the context of nonparametric regression. In the time series literature, it is often the case that inference is made on the properties of the differenced time series. In contrast, this chapter proposes a method to jointly estimate the time-varying trend and second-order structure of the original time series, by employing the commonly used strategy of differencing a time series in order to remove the trend.

Our approach to this problem can be summarised as follows. By differencing the time series to remove the trend, we can estimate the appropriate second-order quantities of interest of the locally stationary wavelet part of Model (3.1.1). This is achieved by considering the effect of differencing on the second-order properties of the series in order to develop an appropriate estimation procedure. Expanding upon the rigorous theory developed in Nason et al. (2000), we obtain results on the consistent estimation of the second-order structure using our modified estimation strategy for the original time series. Using this estimate, we discuss a wavelet thresholding technique to estimate the trend function $\mu(t/T)$ in a principled manner, by taking into account the time-varying second-order behaviour. Our methodology thus enables the joint estimation of the first and second-order structure of nonstationary time series. The data applications analysed in this chapter demonstrate the added utility that estimating the second order structure of the original time series brings.

The rest of the chapter is organised as follows. In Section 3.2 we introduce the time series model which we focus on in this chapter, describe key assumptions, and discuss necessary background to LSW processes. In Section 3.3, we analyse the effect that differencing has on the spectral structure of a time series, and explain the intuition behind our methodology. Furthermore, we describe the methodology for consistent estimation of the second-order structure in the presence of trend,

and in Section 3.4 we use this estimate in order to estimate the trend of the series. Simulation studies assessing the method's performance are given in Section 3.5. In Section 3.6, we apply our framework to two data examples, demonstrating the utility of the method, while concluding remarks are given in Section 3.7. All proofs of stated results are contained within Appendix A.

3.2 Model Formulation

In this section we introduce the modelling paradigm that we will use, as well as explaining the necessary background concepts. Our *Trend* Locally Stationary Wavelet (T-LSW) model, developing on the theory of locally stationary wavelet processes of Nason et al. (2000), allows for simultaneous inference on the time-varying mean and autocovariance of a time series. Wavelets are useful in estimating time-varying quantities, especially nonstationary characteristics, as they are compactly supported oscillatory functions that can be translated and dilated to provide location-scale decompositions. For an overview of wavelet techniques, see for example Nason (2008) or Vidakovic (2009).

3.2.1 Model Definition

Below, we define the T-LSW model, which is composed of a deterministic Lipschitz continuous trend component and a locally stationary wavelet component.

Definition 3.2.1. A *trend locally stationary wavelet (T-LSW) process* $\{X_{t,T}\}$, $t = 0, \ldots, T-1$, and $T = 2^J \geq 1$ for $J \in \mathbb{N}$ is a doubly-indexed stochastic process with the following representation in the mean-square sense:

$$X_{t,T} = \mu\left(\frac{t}{T}\right) + \sum_{j=-\infty}^{-1} \sum_{k \in \mathbb{Z}} w_{j,k;T} \psi_{j,k-t} \xi_{j,k}, \tag{3.2.1}$$

where $\{\xi_{j,k}\}$ is a random zero-mean orthonormal increment sequence, $\{\psi_{j,k-t}\}_{j,k}$ is a set of discrete non-decimated wavelets, and $\{w_{j,k;T}\}$ is a set of amplitudes. The

quantities in representation (3.2.1) possess the following properties:

1. The function $\mu : [0,1] \in \mathbb{R}$ is Lipschitz continuous with constant $K > 0$, i.e

$$\left|\mu\left(\frac{t}{T}\right) - \mu\left(\frac{s}{T}\right)\right| \leq \frac{K}{T}, \ \forall \ s,t \in [0,T].$$

2. There exists, for each $j \leq -1$, a Lipschitz continuous function $W_j(z)$ for $z \in (0,1)$ which satisfies the following properties:

$$\sum_{j=-\infty}^{-1} |W_j(z)|^2 < \infty \text{ uniformly in } z \in (0,1);$$

the Lipschitz constants L_j are uniformly bounded in j and $\sum_{j=-\infty}^{-1} 2^{-j} L_j < \infty$. There exists a sequence of constants C_j such that for each T

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

where for each $j \leq -1$ the supremum is over $k = 0, \ldots, T-1$, and where the sequence $\{C_j\}$ satisfies $\sum_{j=-\infty}^{-1} C_j < \infty$.

The model imposes the same assumptions on the LSW component as in Nason et al. (2000), allowing for locally stationary second-order structure, while also permitting nonstationary first-order behaviour by incorporating a smooth mean function μ. Imposing a Lipschitz continuous trend assumption is not overly restrictive, given that trend functions are generally assumed to be smooth and slowly-evolving. In particular, polynomials (restricted to the interval $[0,1]$) are Lipschitz continuous, as are sinusoids. Such an assumption is commonly made in the literature, see for example Vogt (2012) and Khismatullina and Vogt (2020). Furthermore, in Section 3.3.6 we will discuss the case when the trend is not Lipschitz.

3.2.2 Background to LSW Processes

In the original work of Nason et al. (2000), the trend function $\mu(t/T)$ in Equation (3.2.1) is assumed to be everywhere zero, which forces the process mean $\mathbb{E}(X_{t,T})$ to be equal to zero for all t. Consequently, within the original LSW framework it is only possible to estimate time-varying second-order structure when the time series does not exhibit a trend. In order to discuss our proposed methodology in the setting where a trend *is* present, we dedicate the rest of the section to recalling a number of definitions and results from Nason et al. (2000), which we will expand upon and adapt to the setting where a trend is present.

The second-order structure of an LSW process is encoded in the spectrum. The evolutionary wavelet spectrum (EWS) of an LSW process is defined as $S_j(z) := |W_j(z)|^2$ for rescaled time $z = k/T \in (0,1)$, and measures the contribution to variance at a particular rescaled time z and scale j. Since the W_j are assumed to be Lipschitz continuous, the spectrum S_j is also Lipschitz continuous, which ensures it evolves slowly over time. Alterations to the LSW model that use different assumptions on the EWS can be found in Fryzlewicz and Nason (2006) and Van Bellegem and von Sachs (2008), which assume respectively that the EWS is piecewise constant and of bounded total variation.

For ease of notation we now drop the dependence on T in the subscripts of the model quantities. The EWS is estimated via the empirical wavelet coefficients of the time series, given by $d_{j,k} := \langle X_t, \psi_{j,k-t}\rangle = \sum_t X_t \psi_{j,k-t}$. As with Fourier approaches, the raw wavelet periodogram $I_k^j := |d_{j,k}|^2$ is a biased, inconsistent estimator of the EWS (Nason et al. (2000), Proposition 4):

$$\mathbb{E}\left(I_k^j\right) = \sum_l A_{jl} S_l(k/T) + \mathcal{O}(T^{-1}), \tag{3.2.2}$$

$$\mathrm{Var}(I_k^j) = 2\left(\sum_l A_{jl} S_l(k/T)\right)^2 + \mathcal{O}(2^{-j}T^{-1}), \tag{3.2.3}$$

where the operator $A = (A_{jl})_{j,l<0}$ is given by $A_{jl} := \langle \Psi_j, \Psi_l\rangle = \sum_\tau \Psi_j(\tau)\Psi_l(\tau)$, and

the autocorrelation wavelets are defined by $\Psi_j(\tau) := \sum_{k\in\mathbb{Z}} \psi_{j,k}\psi_{j,k-\tau},\ j < 0, \tau \in \mathbb{Z}$. Hence, for the vector of periodograms $\mathbf{I}(z) := \{I^l_{\lfloor zT \rfloor}\}_{l=-1,\ldots,-J}$, and the vector of corrected periodograms $\mathbf{L}(z) := \{L^j_{\lfloor zT \rfloor}\}_{j=-1,\ldots,-J}$ with $\mathbf{L}(z) = A_J^{-1}\mathbf{I}(z)$, where the J-dimensional matrix $A_J := (A_{jl})_{j,l=-1,\ldots,-J}$,

$$\mathbb{E}\left(\mathbf{L}(z)\right) = \mathbb{E}\left(A_J^{-1}\mathbf{I}(z)\right) = \mathbf{S}(z) + \mathcal{O}(T^{-1}) \qquad \forall\ z \in (0,1), \tag{3.2.4}$$

where $\mathbf{S}(z) := \{S_j(z)\}_{j=-1,\ldots,-J}$. The raw wavelet periodogram is first smoothed and then corrected to produce an asymptotically unbiased, consistent estimator. There are several approaches to smoothing for consistency, for example via a running mean as in Nason (2013) or wavelet thresholding as in Nason et al. (2000). The correction is performed by premultiplying the (smoothed) raw wavelet periodogram by A_J^{-1}, as motivated by Equation (3.2.4). The operator A is shown to have bounded inverse for all Daubechies compactly supported wavelets in Chapter 4.

The local autocovariance (LACV) function for an LSW process provides information about the covariance at a rescaled location $z = k/T \in (0,1)$. The LACV, $c(z,\tau)$, of an LSW process with EWS $\{S_j(z)\}$ is defined as $c(z,\tau) = \sum_{j=-\infty}^{-1} S_j(z)\Psi_j(\tau)$, for $\tau \in \mathbb{Z}$, $z \in (0,1)$. The LACV can be thought of as a decomposition of the autocovariance of a process over scales and rescaled time locations. The LACV is estimated by plugging in the smoothed, corrected estimate for the EWS into the definition of the LACV. Using wavelet thresholding of the EWS estimator, it is shown in Nason et al. (2000) that the LACV estimator is consistent. The next section addresses the consistent estimation of these second-order quantities in the presence of first-order nonstationarity.

3.3 LSW Estimation in the Presence of Trend via Differencing

In this section, we discuss methodology for estimation of the evolutionary wavelet spectrum and local autocovariance function of a trend-LSW process. In order to estimate these quantities, we employ the ubiquitous practice of differencing to remove the trend, but crucially correct for the effect of this on the spectrum.

3.3.1 Using Differencing to Detrend a Time Series

One of the most common methods for removing the trend in a time series is differencing, see for example Chan et al. (1977) and Shumway and Stoffer (2010). The time series $\{X_t\}$ can be, for example, first-differenced to obtain a new time series, $\{\nabla X_t = X_t - X_{t-1}\}$, upon which inference is then performed. One advantage of differencing is that no parameters are estimated in the differencing operation, which is not always the case for general detrending. An n-th order difference is capable of removing a polynomial trend of degree n from the data.

One of the consequences of differencing is that the second-order statistical properties of the time series in Model (3.2.1) will change, sometimes quite drastically. Therefore, it is potentially problematic to directly use the differenced process for inference on the original process. In the context of ARIMA modelling, differencing is performed in order to induce stationarity, and estimation is then performed on the differenced series. However, this reasoning does not hold within our setting: if we difference a nonstationary LSW process, the result will still be a nonstationary process. Due to the potentially complex structure of the LSW process, properties of the differenced series are not necessarily representative of the original series. However, in order to estimate the trend component in (3.2.1), we require an estimate of the second-order structure of the original time series, not the differenced series.

It is straightforward to produce an example where spectral behaviour can

be completely altered as a consequence of detrending using, for example, first-differencing. Consider the zero-mean LSW process of length $T = 2^{10} = 1024$, defined by the spectrum $S_j(z) = 1$ for $j = -1$, 0 otherwise. This time series is referred to as the scale -1 Haar moving average process, and can be written as $X_t = (\epsilon_t - \epsilon_{t-1})/\sqrt{2}$, where the ϵ_t are independent identically distributed (IID) random variables. Taking the first-difference, we obtain $\nabla X_t = X_t - X_{t-1} = (\epsilon_t - 2\epsilon_{t-1} + \epsilon_{t-2})/\sqrt{2}$. Computing the expectation of the raw wavelet periodogram of the differenced time series, we find that for $j = -1$, $\mathbb{E}(I_k^{-1}) = 5$, and for $j < -1$, $\mathbb{E}(I_k^j) = 3 \times 2^{j+1}$. Therefore, for all $z \in (0, 1)$, the expected value of the LSW estimator at time z is $\mathbb{E}(\mathbf{L}(z)) = \mathbb{E}(A_{10}^{-1}\mathbf{I}(z))$, where A_{10}^{-1} is the inverse of the 10-dimensional A-matrix.

Having differenced the time series, a problem arises since $\mathbb{E}(I_k^j) \neq \sum_l A_{jl} S_l(k/T) + \mathcal{O}(T^{-1})$. In particular, the expectation of the EWS estimate at scale -2 at any time point is given by $\mathbb{E}\left(\sum_l A_{-2,l}^{-1} I_k^l\right) = -0.79 < 0$. Therefore, the expectation of our estimate of the spectrum at level -2 is -0.79, while in the original time series we had $S_{-2}(z) = 0$ for all $z \in (0, 1)$. In Figure 3.3.1 left, we see a plot of the original spectrum, while on the right, we see the expectation of the corrected periodogram estimate, showing a clear discrepancy.

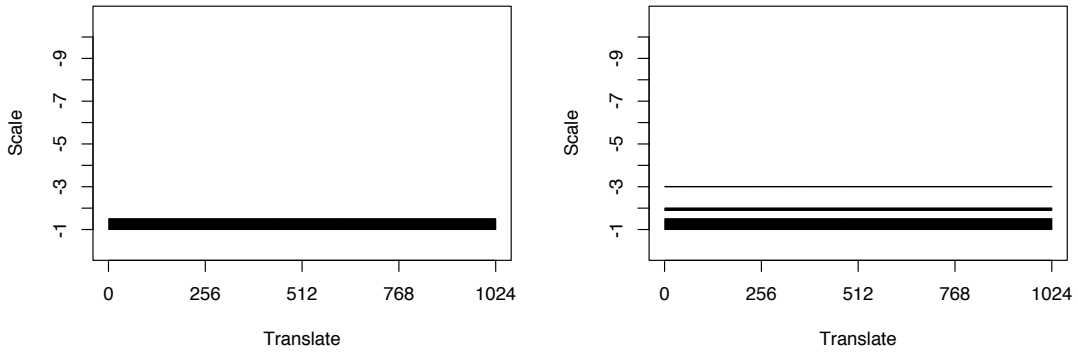

Figure 3.3.1: Left: original spectrum of the scale -1 Haar moving average process. Right: expectation of spectral estimate of differenced time series.

Differencing can also cause a very non-sparse spectral structure to be altered into a very sparse one. The white noise process $X_t = \epsilon_t$, where the ϵ_t are IID random variables, can be represented as an LSW process using Haar wavelets and

a spectrum defined by $S_j(z) = 2^j$ for all $z \in (0,1)$. Differencing this time series gives $\nabla X_t = X_t - X_{t-1} = \epsilon_t - \epsilon_{t-1}$, which is a Haar LSW process with spectrum $S_j(z) = 2$ for $j = -1$, $S_j(z) = 0$ otherwise. This induces autocorrelation in the time series at lag 1, which is similar to what is observed when over-differencing in classical stationary time series. Therefore, it can be seen that we must take into account the differencing of the time series if we wish to say something meaningful about the original series.

3.3.2 Intuition Behind Estimation Procedure

Given that the trend component of the Model (3.2.1) is Lipschitz continuous, first-differencing the time series yields

$$\begin{aligned}\nabla X_t = X_t - X_{t-1} &= \mu\left(\frac{t}{T}\right) + \epsilon_t - \mu\left(\frac{t-1}{T}\right) - \epsilon_{t-1} \\ &= \epsilon_t - \epsilon_{t-1} + \mathcal{O}(T^{-1}).\end{aligned}$$

Hence, differencing the original series results in a differenced locally stationary wavelet process that is asymptotically zero-mean. We wish to estimate the evolutionary wavelet spectrum of the original time series $\{X_t\}$ using the differenced series $\{\nabla X_t\}$. Proceeding by using the standard estimation procedure of Nason et al. (2000) by taking the squared wavelet coefficients and premultiplying by the inverse of the A matrix, as in Equation (3.2.4), is not appropriate, as we have seen in Section 3.3.1. Denote the empirical non-decimated wavelet coefficients of the first-differenced series by

$$\tilde{d}_{j,k} := \sum_t \nabla X_t \psi_{j,k-t}.$$

We can compute the expectation of the raw wavelet periodogram $\tilde{I}_k^j := |\tilde{d}_{j,k}|^2$, which will yield an analogous result to Equations (3.2.2) and (3.2.3). Hence, our estimation strategy will follow that of Nason et al. (2000). However, the correction of the raw wavelet periodogram to achieve unbiasedness will require a different

correction matrix, to account for the fact that the raw wavelet periodogram is calculated on the differenced series.

Remark 1. It is not immediately obvious that if we first-difference an LSW process, we obtain another LSW process. We can write the differenced process, in the mean-square sense, as

$$\epsilon_t - \epsilon_{t-1} = \sum_j \sum_k w_{j,k} \psi_{j,k-t} \tilde{\xi}_{j,k},$$

where $\tilde{\xi}_{j,k} = \xi_{j,k} - \xi_{j,k-1}$. This process satisfies all of the required properties of a standard LSW process exept one: the increments are no longer orthonormal. Instead, we have that

$$\mathrm{Cov}(\tilde{\xi}_{j,k}, \tilde{\xi}_{l,m}) = \begin{cases} 2 & \text{for } j = l, k = m, \\ -1 & \text{for } j = l \text{ and } k = m+1 \text{ or } k+1 = m, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, one must be careful to make a distinction between the observed time series, which we assume to have LSW errors, and the differenced series, which we do not. This is in contrast to existing approaches utilising differencing for second-order estimation, such as ARIMA models, which instead assume that the differenced series follows the model form rather than the original series.

3.3.3 Asymptotic Behaviour of the Differenced Raw Wavelet Periodogram

As motivated by the discussion in the previous section, we estimate the spectrum by calculating the raw wavelet periodogram of the first-differenced time series. We return to the case of general n-th differencing in Section 3.3.6. For the purpose of theoretical results, we assume that the $\{\xi_{j,k}\}$ in Model (3.2.1) are Gaussian. In practice, this assumption is not required, and in the simulation study in Section

3.5 we also investigate the case where the $\{\xi_{j,k}\}$ are exponential random variables. Before obtaining the result of the expectation of the raw wavelet periodogram, we define the following two operators, which involve the inner product of the autocorrelation wavelets at lag 1.

Definition 3.3.1. Define the operator $A^1 = (A^1_{jl})_{j,l<0}$ by

$$A^1_{jl} := \langle \Psi_j(\tau), \Psi_l(\tau - 1) \rangle = \sum_\tau \Psi_j(\tau)\Psi_l(\tau - 1),$$

and the operator $D^1 = (D^1_{jl})_{j,l<0}$ by

$$D^1_{jl} := \langle \Psi_j(\tau) - \Psi_j(\tau - 1), \Psi_l(\tau) - \Psi_l(\tau - 1) \rangle = 2A_{jl} - 2A^1_{jl}$$

$$= 2\sum_\tau \Psi_j(\tau) \left(\Psi_l(\tau) - \Psi_l(\tau - 1)\right).$$

Denote the J-dimensional matrices $A^1_J := (A^1_{jl})_{j,l=-1,\ldots,-J}$ and $D^1_J := (D^1_{jl})_{j,l=-1,\ldots,-J}$.

Proposition 3.3.1. The matrix D^1_J is invertible.

Intuitively, it is not surprising that the quantity D^1 will appear when we calculate the expectation of the squared wavelet coefficients of the first-differenced series. Indeed, we have the following result for the asymptotic behaviour of the raw wavelet periodogram of the first-differenced time series.

Proposition 3.3.2. Let $\tilde{I}^j_k = |\tilde{d}_{j,k}|^2$ be the wavelet periodogram of the first-differenced time series. Under the assumptions of Model (3.2.1),

$$\mathbb{E}(\tilde{I}^j_k) = \sum_l D^1_{jl} S_l \left(\frac{k}{T}\right) + \mathcal{O}(T^{-1}),$$

$$\mathrm{Var}(\tilde{I}^j_k) = 2\left(\sum_l D^1_{jl} S_l \left(\frac{k}{T}\right)\right)^2 + \mathcal{O}(2^{-j}T^{-1}).$$

Therefore, for the vector of periodograms $\mathbf{I}(z) := \{I^l_{\lfloor zT \rfloor}\}_{l=-1,\ldots,-J}$, and the vec-

tor of corrected periodograms $\mathbf{L}(z) := \{L^j_{\lfloor zT \rfloor}\}_{j=-1,\ldots,-J}$ with $\mathbf{L}(z) = (D^1_J)^{-1}\mathbf{I}(z)$,

$$\mathbb{E}\left(\mathbf{L}(z)\right) = \mathbb{E}\left((D^1_J)^{-1}\mathbf{I}(z)\right) = \mathbf{S}(z) + \mathcal{O}(T^{-1}) \qquad \forall\ z \in (0,1),$$

where $\mathbf{S}(z) := \{S_j(z)\}_{j=-1,\ldots,-J}$ is the EWS of the original process. Thus we can bias correct the raw wavelet periodogram of the first-differenced time series by using the inverse of D^1_J, instead of the inverse of A_J from Nason et al. (2000).

3.3.4 Bounded Invertibility of Haar and Shannon Operators

In order to achieve an *asymptotically* unbiased estimator of the EWS, we require boundedness of the inverse operator used when performing the bias correction step in the estimation procedure. In the original LSW work of Nason et al. (2000), boundedness of the inverse of the operator A was shown in the case of the Haar and Shannon wavelets. As noted in von Sachs et al. (1997), the Haar and Shannon wavelets can be viewed as the lower and upper "extremes" of the family of Daubechies compactly support wavelets, and so proving bounded invertibility in these two cases intuitively suggests that bounded invertibility should hold in the case of all Daubechies' compactly supported wavelets. It is for these two wavelets that we prove bounded invertibility here; extensions to other wavelets are left for future work.

Proposition 3.3.1 shows that the matrix D^1_J is invertible with bounded inverse. Asymptotically, however, the inverse operator is unbounded. This is due to the decay properties of D^1_J: the diagonal entries decay to zero, and thus so do the eigenvalues. Intuitively this is expected, since the differencing operator itself is asymptotically non-invertible. We can interpret the expectation result in Proposition 3.3.2 as a quantification of the "information lost" in the second-order structure of the time series as a result of differencing.

We note that the problem of an unbounded inverse operator mirrors a result in the setting of locally stationary Fourier time series (Roueff and von Sachs, 2011,

Equation 5). Loosely speaking, the original spectrum at frequency ω is related to the differenced one through multiplication of the term $|1 - e^{-i\omega}|^{-2}$. As $\omega \to 0$ (corresponding to low frequencies), $|1 - e^{-i\omega}|^{-2} \to \infty$. This is similar to our scenario, where, as the correction matrix grows in size – and we consider coarse-scale (low frequency) behaviour – the inverse matrix norm becomes larger.

We can account for this theoretical issue by proving bounded invertibility for a related, rescaled operator P, where the entries of P are given by $P_{jl} = 2^{-j/2} D^1_{jl} 2^{-l/2}$. Showing that P possesses a bounded inverse enables us to show theoretical consistency properties of the wavelet periodogram-based estimator. Hence, we show for the Haar and Shannon wavelet families that P possesses a bounded inverse, which enables consistent estimation of the EWS for Model (3.2.1). Note that, in the practical implementation of the methodology, we still use the inverse of the D^1_J matrix for correcting the raw wavelet periodogram as it is an invertible matrix (Proposition 3.3.1). To illustrate the above discussion, we compute the entries of the Haar D^1 and P matrices for $J = 4$, shown below.

$$D^1 = \begin{bmatrix} 5 & 1.5 & 0.75 & 0.375 \\ 1.5 & 2.5 & 0.75 & 0.375 \\ 0.75 & 0.750 & 1.25 & 0.375 \\ 0.375 & 0.375 & 0.375 & 0.625 \end{bmatrix}, \quad P = \begin{bmatrix} 10 & 4.24 & 3 & 2.12 \\ 4.24 & 10 & 4.24 & 3 \\ 3 & 4.24 & 10 & 4.24 \\ 2.12 & 3 & 4.24 & 10 \end{bmatrix}.$$

We see that for D^1, the diagonal entries behave as $\mathcal{O}(2^j)$, while for P the diagonal entries are constant. Working with the operator P, which has more attractive properties than D^1, allows us to prove the necessary consistency results for the EWS estimator. Theorem 3.3.3 below shows that for Haar and Shannon wavelets, P possesses a bounded inverse.

Theorem 3.3.3. Let $\lambda_{\min}(P)$ denote the smallest eigenvalue of P, where the entries of P are given by $P_{jl} = 2^{-j/2} D^1_{jl} 2^{-l/2}$. Then, for the Haar and Shannon wavelet families, there exists $\delta > 0$ such that $\lambda_{\min}(P) \geq \delta$ and hence $||P^{-1}|| < \infty$.

That is, P is positive-definite and has a bounded inverse.

3.3.5 Smoothing and Estimation Theory

As in the original LSW model, the wavelet periodogram is not a consistent estimator and must be smoothed. Smoothing to achieve consistency can be performed, for example, via wavelet thresholding or a running mean. For brevity, we provide theoretical results for wavelet thresholding, building on known results in the literature. Note that the results presented here require boundedness of the inverse of the operator P, so are only applicable to the Haar and Shannon wavelets. We conjecture that Theorem 3.3.3 can be extended to all Daubechies compactly supported wavelets, and therefore conjecture that the results in this section can also be extended to all Daubechies compactly supported wavelets.

In order to utilise the result on boundedness of the operator P^{-1} in Theorem 3.3.3, we rewrite the formula for the expectation of the wavelet periodogram. We can express the expectation in terms of P and a scaled version of S_j, given by $\tilde{S}_j = 2^{j/2} S_j$, by rescaling the periodogram appropriately. Concretely, consider the auxiliary process $\epsilon_t = \sum_{j,k} \tilde{w}_{jk} \tilde{\psi}_{j,k-t} \xi_{lm}$, where $\tilde{w}_{jk} = 2^{j/4} w_{jk}$ and $\tilde{\psi}_{j,k-t} = 2^{-j/4} \psi_{j,k-t}$. Then, the expectation of the raw wavelet periodogram (with respect to the rescaled wavelet $\tilde{\psi}_{j,k-t}$) is given by

$$\mathbb{E}(\tilde{I}_k^j) = \sum_l P_{jl} \tilde{S}_l(k/T) + \mathcal{O}(T^{-1}),$$

where $\tilde{S}_j(k/T) = 2^{j/2} S_j(k/T)$ and $P_{jl} = 2^{-j/2} D^1_{jl} 2^{-l/2}$.

Then, to achieve consistency we take a similar approach to Nason et al. (2000). For each fixed scale j, the rescaled periodogram $\tilde{I}_k^j$ of a Gaussian LSW process (which is scaled χ^2-distributed) is smoothed as a function of $z = k/T$ using, for example, discrete wavelet transform (DWT) shrinkage or translation invariant (TI) denoising of Coifman and Donoho (1995). Using the DWT, smoothing is performed with respect to an orthonormal wavelet basis $\{\phi'_{r_0,s}(z), \psi'_{rs}(z)\}$ of $L^2([0,1])$. Here,

$\psi'_{rs}(z) = 2^{r/2}\psi'(2^r z - s)$, where r_0 is the coarsest scale analysed and $s = 0, \ldots, 2^r - 1$. Smoothing is achieved using non-linear thresholding of the empirical wavelet coefficients $\hat{v}^j_{rs}$ of $\tilde{I}^j_k$.

Explicitly, as described in von Sachs et al. (1997), for levels $j = -1, \ldots, -J$, the wavelet expansion of the scaled periodogram can be written as

$$\tilde{I}^j_{\lfloor zT \rfloor} = \sum_r \sum_s v^j_{rs} \psi'_{rs}(z),$$

where the "true" wavelet coefficients are given by $v^j_{rs} = \int_0^1 \tilde{I}^j_{\lfloor zT \rfloor} \psi'_{rs}(z) dz$. As in von Sachs et al. (1997), we employ a slight abuse of notation, with $\psi'_{r_0-1,s} = \phi'_{r_0,s}$, in order to include the scaling coefficient at the coarsest scale r_0 of the second wavelet scheme. The empirical analogues of the wavelet coefficients are given by

$$\hat{v}^j_{rs} = T^{-1} \sum_{n=0}^{T-1} \tilde{I}^j_{n,T} \psi'_{rs}(n/T), \quad \text{for } r = r_0, \ldots, \log_2(T), \quad s = 0, \ldots, 2^r - 1.$$

Analogously, we can build non-decimated wavelet coefficients for TI denoising. Let $\psi'_r(z) = 2^{r/2}\psi'(2^r z)$, and let

$$\hat{v}^j_{rs} = T^{-1} \sum_{n=0}^{T-1} \tilde{I}^j_{n,T} \psi'_{rs}((n-s)/T), \quad \text{for } r = r_0, \ldots, \log_2(T), \quad s = 0, \ldots, T - 1.$$

Then, denoising is achieved by applying non-linear hard wavelet thresholding to the wavelet coefficients $\hat{v}^j_{rs}$. The resulting estimator is obtained by inverting the wavelet transform using only the coefficients which remain after thresholding:

$$\hat{I}^j_{\lfloor zT \rfloor} = \sum_r \sum_s \tilde{v}^j_{rs} \psi'_{rs}(z), \quad z \in (0, 1),$$

where $\tilde{v}^j_{rs} = \hat{v}^j_{rs} \mathbb{I}(|\hat{v}^j_{rs}| > \lambda)$ are the hard thresholded wavelet coefficients with threshold λ. The particular threshold is specified in Theorem 3.3.4, as is the appropriate set of indices r over which to perform the summation.

The theoretical argument to consistently estimate $S_j(k/T)$ is thus as follows.

First, smooth the rescaled wavelet periodogram at each scale j using wavelet thresholding. Next, use the (bounded) P-inverse matrix to correct the smoothed, rescaled periodogram. Finally, multiply the estimate at each scale by $2^{-j/2}$, since $S_j(k/T) = 2^{-j/2}\tilde{S}_j(k/T)$, yielding the final estimator, $\widehat{S}(k/T)$, of $S_j(k/T)$. Using the hard threshold $\lambda(j,r,s,T)^2 = \mathrm{Var}(\hat{v}^j_{rs})\log^2(T)$ when smoothing the periodogram via wavelet thresholding, we can show that the smoothed, corrected estimate $\widehat{S}_j(z)$ is mean square consistent in the L_2 sense.

Theorem 3.3.4. Denote by ψ' the wavelet used in the thresholding procedure, where ψ' is of bounded variation. For the wavelet coefficients $\hat{v}^j_{rs}$ of the rescaled periodogram $\tilde{I}^j_k$, assume that $2^r = o(T)$. For a Gaussian trend-LSW process and using the hard threshold $\lambda^2(j,r,s,T) = \mathrm{Var}(\hat{v}^j_{rs})\log^2(T)$, for each fixed j,

$$\mathbb{E}\left[\int_0^1 \left(\widehat{S}_j(z) - S_j(z)\right)^2 dz\right] = \mathcal{O}\left(2^{-j}T^{-2/3}\log^2(T)\right). \tag{3.3.1}$$

The appropriate threshold is obtained using the expressions for the asymptotic expectation and variance of the wavelet coefficients, which is given in Lemma A.1.2 in the appendix. Roughly speaking, we would expect the variance of the wavelet coefficients to decrease as j decreases, due to the effect of differencing. The rate of error obtained in Equation (3.3.1) is a consequence of known results on wavelet thresholding estimators, utilised in Neumann and von Sachs (1995), and the multiplication of $2^{-j/2}$ that occurs in the estimation procedure. The rate highlights the fact that differencing the time series has resulted in an information loss, with spectral estimation in coarser scale resulting in slower rates of convergence. As a consequence, we can only consistently estimate the wavelet spectrum for a proportion of the finest scales j. Next, we tackle estimation of the local autocovariance via the EWS estimate.

Proposition 3.3.5. Define $\hat{c}(z,\tau)$ by replacing $S_j(z)$ by $\widehat{S}_j(z)$ in the equation for the local autocovariance and replacing the lower limit in the sum from $j = -\infty$ to

$j = -J_0$, i.e.

$$\hat{c}(z,\tau) = \sum_{j=-J_0}^{-1} \widehat{S}_j(z)\Psi_j(\tau).$$

Let $T \to \infty$ and let $J_0 = \alpha \log_2 T$ for $\alpha \in (0,1)$. Assume that $S_j(z) \leq D2^{\gamma j}$ for some positive constant D, where $\gamma = (3\alpha)^{-1} - 1/2$. Then,

$$\mathbb{E}\left[\int_0^1 \left(\hat{c}(z,\tau) - c(z,\tau)\right)^2 dz\right] = \mathcal{O}\left(T^{\alpha-2/3}\log^2(T)\right),$$

i.e. $\hat{c}(z,\tau)$ is a consistent estimator of $c(z,\tau)$ for each fixed $\tau \in \mathbb{Z}$, provided that $T^{\alpha-2/3}\log^2(T) \to 0$.

We reiterate that the rescaling argument used to achieve consistency in Theorem 3.3.4 and Proposition 3.3.5 is performed purely for theoretical reasons; practical considerations for estimation are discussed at the end of the section. The results in Theorem 3.3.4 and Proposition 3.3.5 show that in the case where the trend is Lipschitz continuous, we can consistently estimate both the EWS and LACV of the original process using first-order differences and a modified bias correction.

The assumption placed on the decay rate of the EWS in Proposition 3.3.5 is a purely technical one, utilised in order to ensure mean square consistency of the estimator. The assumption controls for the fact that the local autocovariance is estimated using the finest J_0 scales, instead of across infinite scales which are not available in practice. The specific form of the decay rate is calculated in order to balance with the error rate of the wavelet thresholding procedure.

3.3.6 n-th Order Differencing

In some cases, a first-difference may not be enough to remove a trend. Further differencing can be performed, although it is usually only necessary to at most second-difference a time series (Brockwell et al., 1991). If we assume that the $(n-1)$-th derivative of μ is Lipschitz, then the n-th difference of the time series will be (asymptotically) free of trend. We denote the n-th difference of a time series as $\{\nabla^n X_t\}$.

To calculate the expectation of the squared non-decimated wavelet coefficients of the n-th differenced series, we can argue in a similar fashion to the case of first-differencing. Denote the operator $A^n = (A^n_{jl})_{j,l<0}$ by $A^n_{jl} := \langle \Psi_j(\tau), \Psi_l(\tau - n) \rangle = \sum_\tau \Psi_j(\tau)\Psi_l(\tau - n)$, and the J-dimensional matrix $A^n_J := (A^n_{jl})_{j,l=-1,\ldots,-J}$. The entries of the matrix A^n_J are given by the inner product of the autocorrelation wavelet with the autocorrelation wavelet at lag n. If we difference a time series n times, then the expectation of the squared wavelet coefficients will involve linear combinations of inner product autocorrelation wavelet matrices from lag 0 (i.e. the standard A-matrix) up to lag n (the matrix A^n). The result for the expectation of the wavelet periodogram of the n-th differenced time series mirrors that of the first-differenced series, and is given by the following proposition.

Proposition 3.3.6. Let $\tilde{d}_{j,k} = \sum_t \nabla^n X_t \psi_{j,k-t}$ be the non-decimated wavelet coefficients of ΔX_t, and let $\tilde{I}^j_k := |\tilde{d}_{j,k}|^2$. If the $(n-1)$-th derivative of μ is Lipschitz, then

$$\mathbb{E}(\tilde{I}^j_k) = \sum_l D^n_{jl} S_l(k/T) + \mathcal{O}(T^{-1}),$$

where

$$D^n_{jl} = \binom{2n}{n} A_{jl} + 2 \sum_{\tau=1}^{n} (-1)^\tau \binom{2n}{n+\tau} A^\tau_{jl}.$$

Corollary 3.3.7. For second differences,

$$\mathbb{E}(\tilde{I}^j_k) = \sum_l \left(6A_{jl} - 8A^1_{jl} + 2A^2_{jl}\right) S_l\left(\frac{k}{T}\right) + \mathcal{O}(T^{-1}).$$

Note that Proposition 3.3.6 generalises Proposition 4 in Nason et al. (2000), in which $n = 0$. As in the case of first-differences, the bias operator D^n does not possess a bounded inverse. Intuitively, for higher order differences, the eigenvalues of D^n decay to 0 at increasingly faster rates. As such, correcting the estimation in a similar fashion as was described in Theorem 3.3.4 yields much slower rates of convergence.

For the remainder of the section, we focus on the situation where a second-difference suffices to remove the trend. We can show that using first-differences

suffices to obtain a consistent spectral estimate, even though the trend has not been fully removed. The first-differences of the trend are Lipschitz continuous, and as such the magnitude of the wavelet coefficients can be bounded using the wavelet characterisation of Hölder spaces (Daubechies, 1992). Using this bound, we can bound the error of the first-differenced estimator in terms of the error due to estimation and error due to the squared wavelet coefficients of the first-differenced trend. Using this argument we are able to show consistency of both the spectrum and local autocovariance estimator as follows.

Theorem 3.3.8. Assume that the first derivative of μ is Lipschitz, and let $J_1 = \beta \log_2 T$ for $\beta \in (0,1)$. Further assume that the smoothed raw wavelet periodogram is corrected across the finest J_1 scales only. Under the same conditions as Theorem 3.3.4, for each fixed j, $\widehat{S}_j(z)$ is a consistent estimator of $S_j(z)$, provided that $T^{7\beta-4} \to 0$ as $T \to \infty$, since

$$\mathbb{E}\left[\int_0^1 \left(\widehat{S}_j(z) - S_j(z)\right)^2 dz\right] = \mathcal{O}\left(2^{-j}T^{7\beta-4}\right) + \mathcal{O}\left(2^{-j}T^{-2/3}\log^2(T)\right) + \mathcal{O}\left(2^{-j}T^{-\beta}\right).$$

Define $\hat{c}(z,\tau)$ by replacing $S_j(z)$ by $\widehat{S}_j(z)$ in the equation for the local autocovariance and replacing the lower limit in the sum from $j = -\infty$ to $j = -J_0$, i.e.

$$\hat{c}(z,\tau) = \sum_{j=-J_0}^{-1} \widehat{S}_j(z)\Psi_j(\tau),$$

where $J_0 = \alpha \log_2 T$ for $\alpha < \beta \in (0,1)$. Under the assumptions of Proposition 3.3.5, provided that $T^{\alpha+7\beta-4} \to 0$ and $T^{\alpha-2/3}\log^2(T) \to 0$, $\hat{c}(z,\tau)$ is a consistent estimator of $c(z,\tau)$, since for each fixed $\tau \in \mathbb{Z}$,

$$\mathbb{E}\left[\int_0^1 \left(\hat{c}(z,\tau) - c(z,\tau)\right)^2 dz\right] = \mathcal{O}\left(T^{\alpha+7\beta-4}\right) + \mathcal{O}\left(T^{\alpha-2/3}\log^2(T)\right) + \mathcal{O}(T^{\alpha-\beta}).$$

Hence, we can still consistently estimate the EWS and LACV via first-order differences when the trend of the time series has a Lipschitz continuous first derivative. Therefore, we argue that in most practical scenarios, it is sufficient to only

perform one difference in order to estimate the evolutionary wavelet spectrum and local autocovariance in the presence of a trend.

In practice, the EWS is estimated by analysing the time series at the finest J_0 scales. We smooth the wavelet periodogram at these scales and then correct the estimate using the J_0-dimensional matrix $(D^1_{J_0})^{-1}$. We take the same approach for estimation of the LACV, using only the finest J_0 scales. Performing estimation in this fashion ensures that the spectral estimate is well-behaved, and is an approach commonly adopted in the LSW literature. We suggest using $J_0 = \lfloor \alpha \log_2(T) \rfloor$, with $\alpha = 2/3$, in agreement with other discussion in the literature; see for example Sanderson et al. (2010). (Note that although this value for α does not satisfy the required conditions of Proposition 3.3.5, in practice it performs well).

This choice of J_0 is further justified by observing that, at coarser scales, the wavelet coefficients are composed entirely of boundary values, and so estimates at coarser scales are inherently artificial. For a concrete example, consider a time series of length $T = 1024$. For the Daubechies least asymmetric wavelet with 4 vanishing moments, the length of the wavelet filter used to calculate the wavelet coefficients at scale j is given by $7 \times (2^{-j} - 1) + 1$. The length of the filter at scale -8 is $1786 > 1024$, so all wavelet coefficients at this, and coarser scales, will be calculated via boundary extensions of the time series. Therefore, we ague that we are not neglecting informative scales by only using the finest J_0 scales.

3.4 Trend Estimation Using the Spectral Estimate

In this section, we discuss a wavelet thresholding approach for the estimation of the trend component of Model (3.2.1). If a first (or second) difference is capable of removing the trend from a time series, we have shown that we can consistently estimate the time-varying evolutionary wavelet spectrum using the smoothed, corrected raw wavelet periodogram of the differenced time series. We now wish to use this estimate in order to estimate the trend of the time series.

3.4.1 Wavelet Thresholding Estimator

The approach we take is to use the spectral estimate directly within a wavelet thresholding estimation procedure. In von Sachs and MacGibbon (2000), the authors describe a wavelet thresholding methodology for consistent curve estimation in the presence of locally stationary errors, subject to mild regularity conditions. The authors propose to use a local median absolute deviation pre-estimate for the variance of the wavelet coefficients, which is used in the threshold. We instead incorporate the consistent spectral estimate into the thresholding procedure, which enables a more systematic method of thresholding. This yields an analogous version of Theorem 1 in von Sachs and MacGibbon (2000), for the specific case of Lipschitz continuous trend functions.

With a slight abuse of notation, denote the estimated wavelet coefficients of the time series by $\hat{v}_{rs}$. The results in this section apply to the commonly used soft and hard thresholding rules, given respectively by

$$\hat{v}^S_{rs} = \operatorname{sgn}(d_{r,s})\left(|d_{r,s}| - \lambda\right)\mathbb{I}(|d_{r,s}| > \lambda)\,,$$
$$\hat{v}^H_{rs} = d_{r,s}\mathbb{I}(|d_{r,s}| > \lambda),$$

where $\lambda = \lambda(r, s, T)$ is the threshold, and $\mathbb{I}$ is the indicator function. Asymptotic normality of the empirical wavelet coefficients, established in von Sachs and MacGibbon (2000), permits the use of a coefficient-dependent universal threshold $\lambda(r, s, T) = \sigma_{rs}\sqrt{2\log(T)}$, where σ^2_{rs} is the variance of the wavelet coefficients. This yields the following result for the wavelet thresholding estimator $\hat{\mu}$ obtained using the threshold $\lambda(r, s, T)$, and either soft or hard thresholding.

Proposition 3.4.1. Let $\tilde{\psi}$ be a wavelet of bounded variation, with $2^r = o(T)$ for wavelet coefficients $\hat{v}_{rs}$. For a trend-LSW process with Lipschitz continuous trend, and using the threshold $\lambda(r, s, T) = \sigma_{rs}\sqrt{2\log(T)}$, the estimator $\hat{\mu}$ satisfies

$$\mathbb{E}\left[\int_0^1 \left(\hat{\mu}(z) - \mu(z)\right)^2 dz\right] = \mathcal{O}\left(\left(\frac{\log(T)}{T}\right)^{2/3}\right).$$

Note that this result is subject to mild regularity assumptions on the LSW component of the model, all of which are satisfied. The innovations $\{\xi_{j,k}\}$ are not restricted to be Gaussian, and can for example be exponential, gamma, or inverse Gaussian distributed. To estimate the variance σ^2_{rs}, which is necessary to choose the threshold $\lambda(r, s, T)$, we use an estimate of the variance of the empirical wavelet coefficients. If the LSW process is generated by wavelet ψ^0, and the wavelet used for thresholding is denoted ψ^1, the variance of the empirical wavelet coefficients $d_{j,k} := \sum_t X_t \psi^1_{j,k-t}$ is given by

$$\mathrm{Var}(d_{j,k}) = \sum_l C^{1,0}_{jl} S_l(k/T) + \mathcal{O}(T^{-1}), \tag{3.4.1}$$

where $C^{(1,0)}_{jl} = \sum_\tau \Psi^0_j(\tau)\Psi^1_l(\tau)$, and where $\Psi^0_j(\tau)$ and $\Psi^1_j(\tau)$ are autocorrelation wavelets with respect to wavelets ψ^0 and ψ^1. By plugging in the estimate $\widehat{S}_j(z)$, obtained in Section 3.3.5, into the expression (3.4.1), this yields the universal-type threshold $\lambda(r, s, T) = \hat{\sigma}_{r,s}\sqrt{2\log(T)}$, where $\hat{\sigma}^2_{r,s} = \sum_l C^{1,0}_{rl}\widehat{S}_l(s/T)$.

3.4.2 Practical Considerations

In alignment with discussion in von Sachs and MacGibbon (2000), in practice we analyse approximately the finest 2/3 scales of the time series, the same as in the spectral estimation procedure. In practice, we have found that applying hard thresholding yields better performance. We recommend the use of translation invariant thresholding over a standard discrete wavelet transform. We have found that it offers stronger practical performance, in terms of the mean squared error of the estimator. As noted in Nason (2008), use of a non-decimated transform ensures that there is "more chance" of the wavelet coefficients picking up the signal of the time series. Furthermore, note that it is possible to obtain negative estimates of the variance of the wavelet coefficients. In this case we replace the negative values by the nearest neighbouring positive value, which was found to have no discernible impact on the trend estimation procedure. We recommend the use

of the Daubechies Least Asymmetric wavelet with 4 vanishing moments, as we have found empirically that it works well for estimation purposes and also helps to minimise the number of negative variance estimates. Note that Proposition 3.4.1 holds for non-Gaussian time series, while theoretical results concerning the second-order estimation require an assumption of normality. In practice, our approach still performs well in the presence of non-Gaussian noise, as we show in the simulation study.

3.5 Simulation Study

In this section we illustrate the ability of our proposed methodology to jointly estimate the mean and EWS of a trend-LSW process by performing a simulation study. For each set of simulations, we use the three EWS shown in Figure 3.5.1, which represent spectra with distinct characteristics. Spectrum S^1, studied in Nason (2008), displays coarse-scale, slowly-evolving sinusoidal behaviour with a fine-scale burst in power at time point 800. Spectrum S^2 is a concatenation of moving average processes and contains power moving from fine to coarser scales, and was examined in Nason et al. (2000). Spectrum S^3 contains slowly-evolving power at fine scales. The three evolutionary wavelet spectra are defined as

$$S_j^1(z) = \begin{cases} \sin^2(4\pi z) & \text{for } j = -5, z \in (0,1), \\ 1 & \text{for } j = -1, z \in (800/1024, 900/1024), \\ 0 & \text{otherwise,} \end{cases}$$

$$S_j^2(z) = \begin{cases} 1 & \text{for } j = -1, z \in (0/1024, 256/1024), \\ 1 & \text{for } j = -2, z \in (256/1024, 512/1024), \\ 1 & \text{for } j = -3, z \in (512/1024, 768/1024), \\ 1 & \text{for } j = -4, z \in (768/1024, 1), \\ 0 & \text{otherwise,} \end{cases}$$

$$S_j^3(z) = \begin{cases} \frac{1}{2} + \frac{1}{4}\sin(\pi z) - \frac{1}{2}\cos(3\pi z/2) & \text{for } j = -1, z \in (0,1), \\ \frac{1}{2} - \frac{1}{8}\sin(2\pi z) - \frac{1}{4}\cos(\pi z/2) & \text{for } j = -3, z \in (0,1), \\ 0 & \text{otherwise.} \end{cases}$$

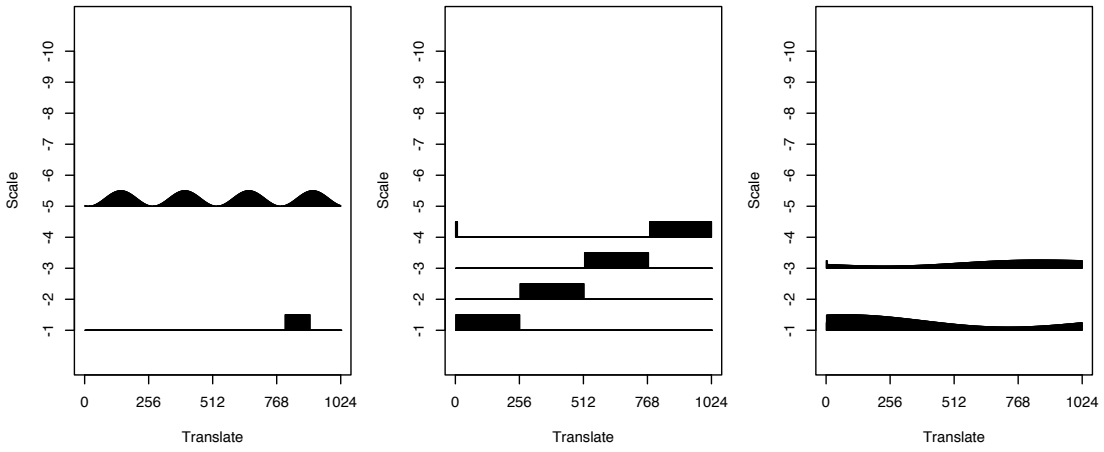

Figure 3.5.1: Left: spectrum S^1, sinusoid with "burst". Centre: S^2, concatenated moving average process. Right: S^3, slowly-evolving fine-scale power.

We simulate 100 realisations of time series $\{X_t\}_{t=0}^{T-1}$ of length $T = 2^{10} = 1024$ from LSW processes with those spectra, with different additive trend functions. The trends used in the simulation study are a linear, sinusoidal, logistic and piecewise quadratic trend, denoted in Figure 3.5.2 by μ_{li}, μ_s, μ_{lo} and μ_q respectively. These functions are Lipschitz continuous with varying degrees of smoothness, with μ_q also being non-differentiable at two time-points. The four trend functions used in the simulation study are defined as

$$\begin{aligned} \mu_{li}(z) &= 4z, \\ \mu_s(z) &= -2\sin(2\pi z) - \frac{3}{2}\cos(\pi z), \\ \mu_{lo}(z) &= \frac{4}{1 + \exp(4 - 7\log 4z)}, \end{aligned}$$

$$\mu_q(z) = \begin{cases} 12z^2 + 2z & \text{for } z \in (0, 300/1024), \\ 1.81 - 16z^2 + 4z & \text{for } z \in (300/1024, 800/1024), \\ 4z - 7.94 & \text{for } z \in (800/1024, 1), \end{cases}$$

where $z = t/T$. We report results where the LSW process are simulated using the Daubechies Least Asymmetric wavelet with 10 vanishing moments (LA10), and the Daubechies Extremal Phase wavelet with 4 vanishing moments (EP4). These two wavelets were chosen as they have differing degrees of smoothness; similar results were obtained with other wavelet choices. Example realisations from the simulation study are shown in Figure 3.5.2. All simulations were performed in `R` with estimation procedures implemented using modifications to code in the `locits` (Nason, 2016a) and `wavethresh` (Nason, 2016b) packages.

3.5.1 Spectral Estimation Performance

In this simulation we show that by first-differencing to remove the trend, we can obtain an unbiased EWS estimate, which in turn can be used to obtain a trend estimate that performs well. For each realisation, the un-smoothed estimate of the EWS was calculated, which was then used to obtain an averaged estimate for the EWS across the 100 realisations. In alignment with the discussion in Section 3.3.5, we correct the raw wavelet periodogram across the finest 7 scales. Further justification for using a reduced number of scales in estimation is due to the lengths of the wavelet filters used. For the LA10 wavelet, the support of the wavelet is smaller than the length of the time series at the finest 5 scales, while for the EP4 wavelet it is smaller for the finest 7 scales.

We report the mean squared error of the averaged spectrum compared with the true spectrum in Table 3.5.1 for Daubechies LA10 wavelet, and in Table 3.5.2 for the Daubechies EP4 wavelet. We compare theses values with the mean squared error obtained by using the standard LSW estimation procedure of Nason et al. (2000). In this case, no trend is present, the estimate is performed using the

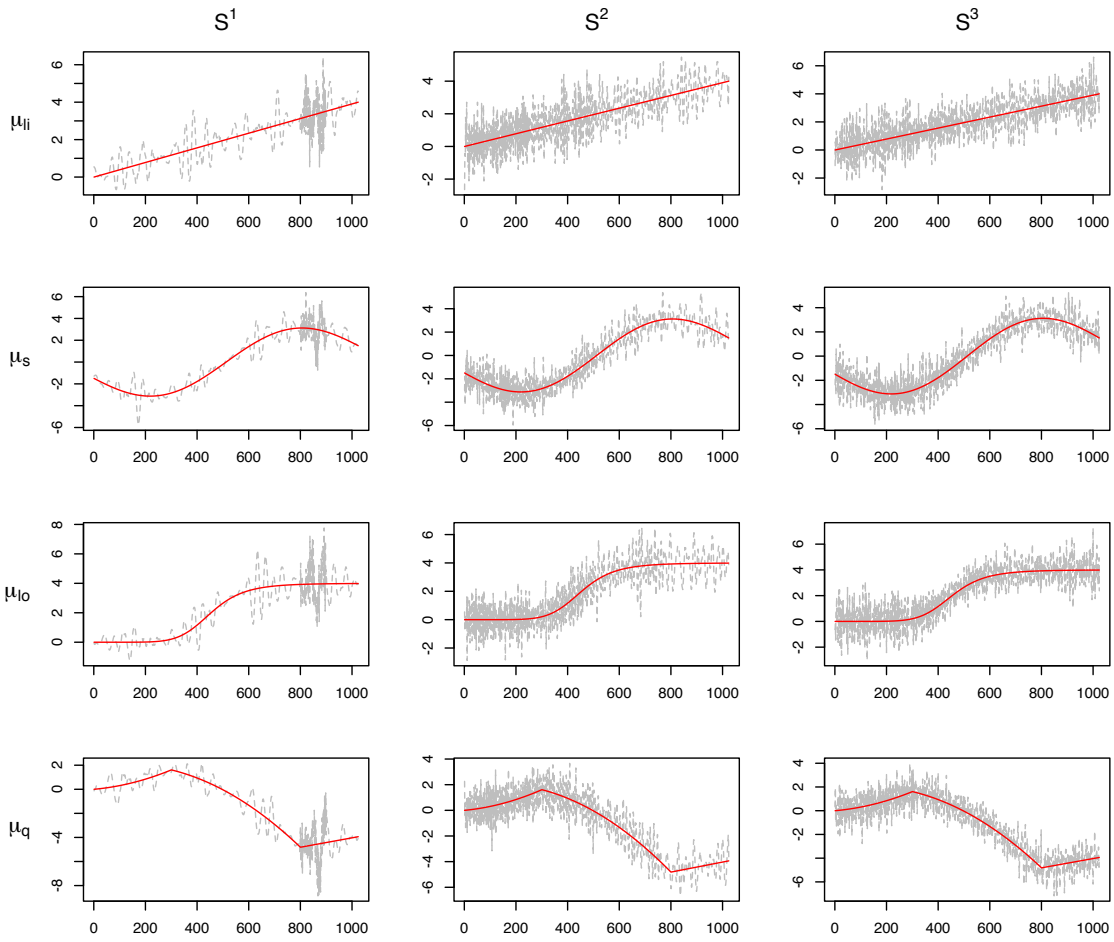

Figure 3.5.2: Example realisations from each trend and spectrum scenario. Dashed line shows time series with true underlying trend shown in solid line. Left: spectrum S^1, sinusoid with "burst" . Centre: S^2, concatenated moving average process. Right: S^3, slowly-evolving fine-scale power.

original A matrix inverse for bias correction, and no differencing is performed. This is calculated using the `ewspec3` command in the `locits R` package. This is reported in the "None" row in the tables, and represents the 'best-case' performance with which to compare.

We see that despite the presence of a trend, differencing enables us to approximately remove it in order to accurately estimate the spectrum. For Table 3.5.1, the estimation error for spectrum 1 is almost identical comparing to "None" using our methodology with a trend present, while for spectrum 2 and 3 there is only a marginal increase. In Table 3.5.2, we see that the estimation error is smaller using our methodology than the "None" case for spectrum 2, while it is marginally worse for the other two spectra. The results show that the method performs well irrespective of the wavelet that generates the LSW process.

Trend	Spectrum 1	Spectrum 2	Spectrum 3
None	1.25	3.32	1.30
Linear	1.29	4.09	2.34
Sine	1.29	4.09	2.34
Logistic	1.29	4.09	2.34
Piece. Quad.	1.29	4.11	2.34

Table 3.5.1: Mean squared error comparison for the averaged spectrum estimate, multiplied by 10^3, across the spectrum and trend scenarios, for the Daubechies LA10 wavelet.

Trend	Spectrum 1	Spectrum 2	Spectrum 3
None	3.13	4.88	1.87
Linear	3.32	4.63	2.78
Sine	3.32	4.63	2.78
Logistic	3.32	4.63	2.78
Piece. Quad.	3.32	4.67	2.81

Table 3.5.2: Mean squared error, multiplied by 10^3, comparing across the spectrum and trend scenarios, for the Daubechies EP4 wavelet.

3.5.2 Trend Estimation Performance

We next assess the trend estimation procedure. The trend estimate is computed using the Daubechies Least Asymmetric wavelet with 4 vanishing moments, analysing the finest 6 scales. The spectrum estimate used in the thresholding procedure is smoothed with a running mean of bin width size 128. We compute the average mean squared error across the 100 realisations, as well as the standard deviation of the errors. For all simulations, a hard threshold of $\hat{\sigma}_{r,s}\sqrt{2\log(1024)}$ is used.

We compare our method to three other trend estimation methods, two of which are wavelet-based in nature. For comparison to the other wavelet-based methods, we use the same hyper-parameters where applicable in order to make a fair comparison. Firstly, we compare to the standard wavelet thresholding method based upon a stationary universal threshold, computed using the `wavethresh` package in `R`. Secondly, we compare to the wavelet-based trend estimation procedure in von Sachs and MacGibbon (2000). No code is publicly available for this method, and so we have implemented the method utilising `wavethresh` and following the description of the computation of the threshold in Section 2.5 of von Sachs and MacGibbon (2000). Lastly, we compare to the estimate computed using smoothing splines, where the smoothing parameter is chosen automatically via cross validation. This is computed using the `smooth.spline` function in `R`. In the tables, the abbreviation LSWT refers to our method of locally stationary wavelet thresholding, SWT refers to stationary wavelet thresholding, MVSWT refers to the wavelet-based method of von Sachs and MacGibbon (2000), and Splines refers to the smoothing splines method.

For each of the two wavelets used to simulate the LSW processes, we repeat this simulation twice; firstly using Gaussian innovations in the LSW process, and secondly using exponentially distributed innovations. Table 3.5.3 reports the average mean squared error and standard deviation for the trend estimates where the 100 LSW processes have been generated using Gaussian innovations and the LA10 wavelet. Table 3.5.4 reports the average mean squared error and standard

Trend	Spectrum	Method			
		LSWT	SWT	VSMWT	Spline
Linear	1	**0.019** (0.013)	0.516 (0.117)	0.275 (0.091)	0.436 (0.123)
	2	**0.038** (0.024)	0.746 (0.083)	0.207 (0.0644)	0.149 (0.091)
	3	0.029 (0.027)	0.494 (0.042)	0.171 (0.039)	$\mathbf{1.82 \times 10^{-5}}$ (2.94×10^{-5})
Sine	1	**0.021** (0.012)	0.547 (0.130)	0.287 (0.111)	0.467 (0.132)
	2	**0.031** (0.016)	0.752 (0.077)	0.214 (0.067)	0.153 (0.090)
	3	0.024 (0.018)	0.485 (0.045)	0.156 (0.040)	**0.002** (5×10^{-4})
Logistic	1	**0.022** (0.013)	0.544 (0.112)	0.306 (0.107)	0.456 (0.113)
	2	**0.042** (0.028)	0.764 (0.088)	0.214 (0.066)	0.165 (0.089)
	3	0.028 (0.025)	0.491 (0.038)	0.170 (0.041)	**0.002** (1×10^{-4})
Piece. quad.	1	**0.025** (0.014)	0.537 (0.127)	0.280 (0.110)	0.452 (0.130)
	2	**0.041** (0.024)	0.736 (0.076)	0.211 (0.080)	0.154 (0.091)
	3	0.034 (0.028)	0.414 (0.043)	0.162 (0.042)	**0.005** (3×10^{-4})

Table 3.5.3: Average mean squared error and standard deviation in brackets of trend estimate over 100 realisations generated using Gaussian innovations.

deviation for the trend estimates where the 100 LSW processes have been generated using exponential innovations and the LA10 wavelet. Values in bold are the lowest values for each trend and spectrum value across the four methods.

From the tables we observe that the performance of our estimator in the presence of exponentially distributed random innovations is comparable to the Gaussian case. The error is consistent across the various time series scenarios, and the standard deviation is low. We see that our method outperforms the two other wavelet-based methods across all trend and spectrum scenarios. We also outperform splines for all spectra 1 and 2 scenarios, while splines outperforms our method for all spectrum 3 scenarios. This is perhaps due to the fact that the process defined by spectrum 3 possesses very weak autocorrelation that is close to stationary.

Table 3.5.5 reports the average mean squared error and standard deviation for the trend estimates where the 100 LSW processes have been generated using Gaussian innovations and the EP4 wavelet. Table 3.5.6 reports the average mean squared error and standard deviation for the trend estimates where the 100 LSW processes have been generated using exponential innovations and the EP4 wavelet. Again, the error for our proposed method is consistent across the various time series scenarios. We see that the trend estimate performs well irrespective of the generating wavelet of the process. We observe similar behaviour to the pre-

Trend	Spectrum	Method			
		LSWT	SWT	VSMWT	Spline
Linear	1	**0.027** (0.016)	0.547(0.140)	0.312 (0.107)	0.465 (0.139)
	2	**0.050** (0.028)	0.793 (0.113)	0.337(0.096)	0.159 (0.121)
	3	0.040 (0.029)	0.529 (0.066)	0.311 (0.068	$\mathbf{1.14 \times 10^{-5}}$ (1.68×10^{-5})
Sine	1	**0.027** (0.014)	0.540 (0.138)	0.311 (0.121)	0.450 (0.136)
	2	**0.041** (0.021)	0.788 (0.118)	0.340 (0.103)	0.148 (0.099)
	3	0.036 (0.019)	0.542 (0.059)	0.308 (0.067)	**0.002** (5×10^{-4})
Logistic	1	**0.026** (0.017)	0.544 (0.143)	0.315 (0.118)	0.459 (0.134)
	2	**0.046** (0.026)	0.801 (0.112)	0.254 (0.091)	0.171 (0.099)
	3	0.045 (0.030)	0.541 (0.053)	0.327 (0.059)	**0.002** (2×10^{-4})
Piece. quad.	1	**0.032** (0.020)	0.515 (0.122)	0.296 (0.102)	0.426 (0.124)
	2	**0.044** (0.028)	0.774 (0.111)	0.347 (0.100)	0.160 (0.100)
	3	0.042 (0.029)	0.473 (0.057)	0.319 (0.061)	**0.005** (4×10^{-4})

Table 3.5.4: Average mean squared error and standard deviation of trend estimate over 100 realisations generated using exponential innovations.

Trend	Spectrum	Method			
		LSWT	SWT	VSMWT	Spline
Linear	1	**0.024** (0.012)	0.519 (0.112)	0.303 (0.100)	0.463 (0.117)
	2	**0.030** (0.019)	0.762 (0.078)	0.225 (0.068)	0.149 (0.095)
	3	0.028 (0.024)	0.441 (0.040)	0.160 (0.038)	$\mathbf{2 \times 10^{-5}}$ (3×10^{-5})
Sine	1	**0.022** (0.010)	0.526 (0.110)	0.286 (0.109)	0.461 (0.113)
	2	**0.026** (0.014)	0.744 (0.075)	0.215 (0.068)	0.132 (0.090)
	3	0.022 (0.017)	0.447 (0.043)	0.156 (0.037)	**0.002** (0.001)
Logistic	1	**0.023** (0.015)	0.494 (0.109)	0.261 (0.093)	0.437 (0.112)
	2	**0.033** (0.019)	0.753 (0.083)	0.226 (0.071)	0.172 (0.083)
	3	0.027 (0.024)	0.449 (0.037)	0.164 (0.035)	**0.002** (1×10^{-4})
Piece. quad.	1	**0.022** (0.012)	0.517 (0.127)	0.290 (0.105)	0.456 (0.129)
	2	**0.032** (0.018)	0.720 (0.087)	0.216 (0.072)	0.148 (0.093)
	3	0.028 (0.024)	0.366 (0.046)	0.167 (0.039)	**0.005** (2×10^{-4})

Table 3.5.5: Average mean squared error and standard deviation in brackets of trend estimate over 100 realisations generated using Gaussian innovations and the Daubechies EP4 wavelet.

vious results, with our method consistently outperforming the two wavelet-based methods. Again, the spline-based method outperforms our method in the case of spectrum 3, while we outperform splines for spectra 1 and 2.

Finally, we also assess the performance of the trend estimation procedure in the presence of non LSW-type error structures to highlight the versatility of our methodology. In particular, we simulate 100 realisations of time series using the previously defined trends, with (Gaussian) errors simulated from the following models:

(A) Time-varying AR(2) model with parameters $\phi_1(z) = 0.8\cos(1.5 - \cos(4\pi z))$, $\phi_2(z) = -0.2 + 0.4z$, $z \in (0, 1)$.

Trend	Spectrum	Method			
		LSWT	SWT	VSMWT	Spline
Linear	1	**0.030** (0.018)	0.521 (0.148)	0.305 (0.127)	0.460 (0.144)
	2	**0.035** (0.024)	0.779 (0.108)	0.328 (0.092)	0.174 (0.095)
	3	0.040 (0.026)	0.497 (0.052)	0.318 (0.057)	$\mathbf{2 \times 10^{-5}}$ (3×10^{-5})
Sine	1	**0.027** (0.025)	0.520 (0.119)	0.308 (0.102)	0.456 (0.120)
	2	**0.033** (0.020)	0.782 (0.110)	0.322 (0.085)	0.162 (0.103)
	3	0.037 (0.022)	0.495 (0.060)	0.309 (0.064)	**0.002** (6×10^{-4})
Logistic	1	**0.030** (0.018)	0.508 (0.126)	0.291 (0.108)	0.447 (0.130)
	2	**0.036** (0.022)	0.788 (0.098)	0.332 (0.085)	0.166 (0.105)
	3	**0.044** (0.031)	0.506 (0.061)	0.327 (0.058)	**0.002** (2×10^{-4})
Piece. quad.	1	**0.031** (0.016)	0.524 (0.116)	0.319 (0.106)	0.463 (0.109)
	2	**0.038** (0.022)	0.748 (0.098)	0.326 (0.100)	0.164 (0.095)
	3	0.045 (0.031)	0.432 (0.055)	0.318 (0.055)	**0.005** (4×10^{-4})

Table 3.5.6: Average mean squared error and standard deviation in brackets of trend estimate over 100 realisations generated using exponential innovations and the Daubechies EP4 wavelet.

(B) AR(1) model with parameter $\phi = 0.6$.

(C) Time-varying AR(1) model with parameter $\phi(z) = 0.7$ for $z \in (0, 600/1024)$, $\phi(z) = -0.3$ for $z \in (600/1024, 1)$.

(D) ARMA(1,3) model with AR parameter $\phi = 0.4$ and MA parameters $\theta_1 = 0.8$, $\theta_2 = -0.3$, and $\theta_3 = 0.4$.

These scenarios represent some common stationary and nonstationary error structures. Model A is an AR(2) process with slowly-evolving autoregressive parameters, and is a variant of a process studied in von Sachs and MacGibbon (2000). Model C represents a piecewise stationary AR(1) process with a single changepoint in the autoregressive parameter. The results of these simulations are given in Table 3.5.7, again with bold values showing the lowest reported mean squared error. We see that our methodology is able to perform well across the four scenarios, working well in both stationary and nonstationary second-order settings. It outperforms the other three methods in all of the trend and error scenarios. Interestingly, the stationary wavelet thresholding method outperforms the nonstationary methodology of von Sachs and MacGibbon (2000). Perhaps most surprisingly, this includes the case where the AR(1) parameter changes abruptly. One possible explanation for this is that the maximum analysing scale may not be optimal, as we have set it

Trend	Model	Method			
		LSWT	SWT	VSMWT	Spline
Linear	A	**0.140** (0.061)	0.376 (0.108)	0.482 (0.139)	0.264 (0.137)
	B	**0.147** (0.040)	0.265 (0.070)	0.500 (0.086)	0.617 (0.083)
	C	**0.129** (0.048)	0.435 (0.080)	0.532 (0.095)	0.564 (0.090)
	D	**0.216** (0.059)	0.349 (0.160)	0.840 (0.129)	1.029 (0.162)
Sine	A	**0.124** (0.069)	0.389 (0.134)	0.494 (0.130)	0.273 (0.151)
	B	**0.134** (0.033)	0.241 (0.091)	0.499 (0.087)	0.625 (0.083)
	C	**0.122** (0.040)	0.421 (0.083)	0.516 (0.100)	0.560 (0.109)
	D	**0.174** (0.045)	0.256 (0.146)	0.806 (0.135)	1.020 (0.154)
Logistic	A	**0.129** (0.048)	0.370 (0.091)	0.474 (0.100)	0.234 (0.106)
	B	**0.150** (0.033)	0.267 (0.077)	0.495 (0.092)	0.614 (0.083)
	C	**0.129** (0.047)	0.431 (0.093)	0.531 (0.098)	0.561 (0.105)
	D	**0.214** (0.051)	0.357 (0.147)	0.837 (0.143)	1.060 (0.145)
Piece. quad.	A	**0.141** (0.051)	0.362 (0.094)	0.467 (0.115)	0.249 (0.112)
	B	**0.141** (0.040)	0.275 (0.073)	0.500 (0.099)	0.624 (0.091)
	C	**0.131** (0.046)	0.420 (0.084)	0.525 (0.098)	0.549 (0.104)
	D	**0.221** (0.057)	0.332 (0.138)	0.834 (0.145)	1.035 (0.130)

Table 3.5.7: Average mean squared error and standard deviation in brackets of trend estimate over 100 realisations generated from Models A – D.

the same in all of the wavelet-based methods in order to make a direct comparison. Overall, we see that our proposed methodology offers strong practical performance across a variety of time series models.

3.6 Data Applications

Next, we discuss two data applications which highlight the benefits of our proposed methodology. In the first example, we analyse a wave height data set collected from a buoy in the Atlantic Ocean that was studied in Killick et al. (2012). In the second, we re-analyse the baby electrocardiogram data set first examined in Nason et al. (2000).

3.6.1 Canadian Wave Height Data

Here we examine publicly available wave height data for a location in the North Atlantic. The data we use are obtained from Fisheries and Oceans Canada, East Scotian Slop buoy ID C44137. The data measures wave heights collected at hourly intervals, from approximately mid-June 2005 to mid-May 2006, giving a time series

of length $2^{13} = 8192$, and is available in the `changepoint` package in `R` (Killick and Eckley, 2014). Killick et al. (2012) also perform nonstationary analysis on the series, using the first-differences of a longer version of the series to detect changepoints in variance only. The data are plotted in Figure 3.6.1 left, from which we see larger wave heights in the winter and smaller wave heights in the summer, as well as increased variability in the winter. Figure 3.6.1 right shows the first-differenced time series. Killick et al. (2012) take first-differences to remove the trend, and we also find this is sufficient.

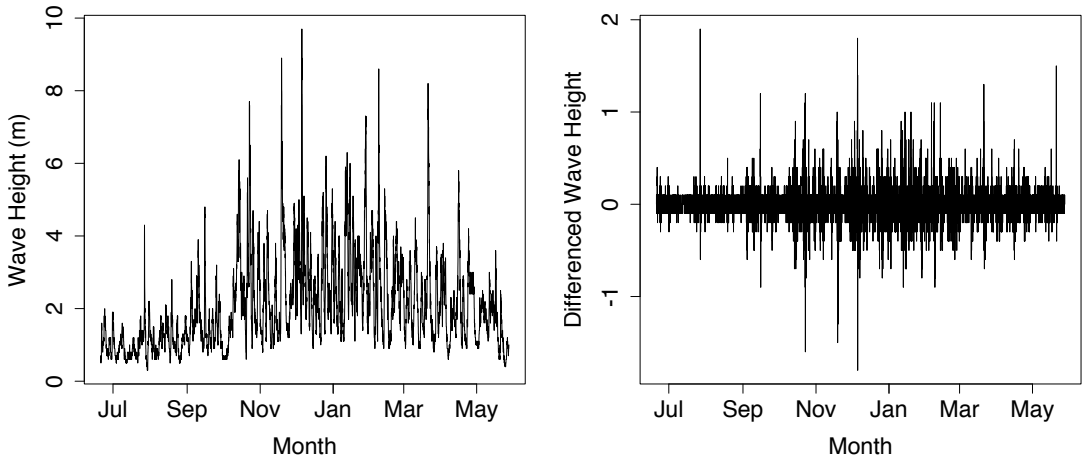

Figure 3.6.1: Left: North Atlantic wave heights recorded hourly between June 2005 and May 2006. Right: first-differenced wave heights.

For the spectral estimate, we use the Daubechies Least Asymmetric wavelet with 10 vanishing moments. For simplicity, the periodogram is smoothed using a running mean with bin width 512, corresponding to a time length of roughly 3 weeks. The estimate is shown in Figure 3.6.2, where each level in the plot is scaled individually for clarity. We can see strong nonstationarity in the spectrum, with more variability during the winter months and less in the summer, as found in Killick et al. (2012) using a changepoint approach.

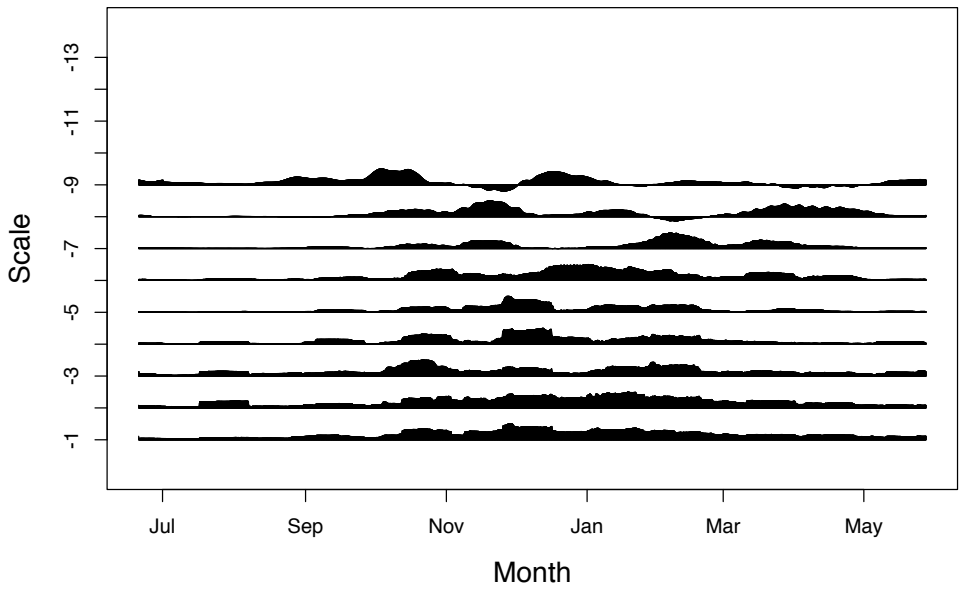

Figure 3.6.2: Smoothed spectral estimate for the wave height data in Figure 3.6.1, performed using first-differencing of the data.

To estimate the trend of the series, we perform TI wavelet thresholding using the Daubechies Least Asymmetric wavelet with 4 vanishing moments. In line with previous discussion, we analyse the finest 9 scales of the series. We use a soft universal threshold of $\hat{\sigma}_{r,s}\sqrt{2\log(8192)}$ to account for possible non-Gaussianity, where $\hat{\sigma}_{r,s}$ is calculated using the spectral estimate in Figure 3.6.2. We see that the estimated trend function is relatively smooth with occasional sharp changes, with the mean wave height larger during the winter and smaller in the summer.

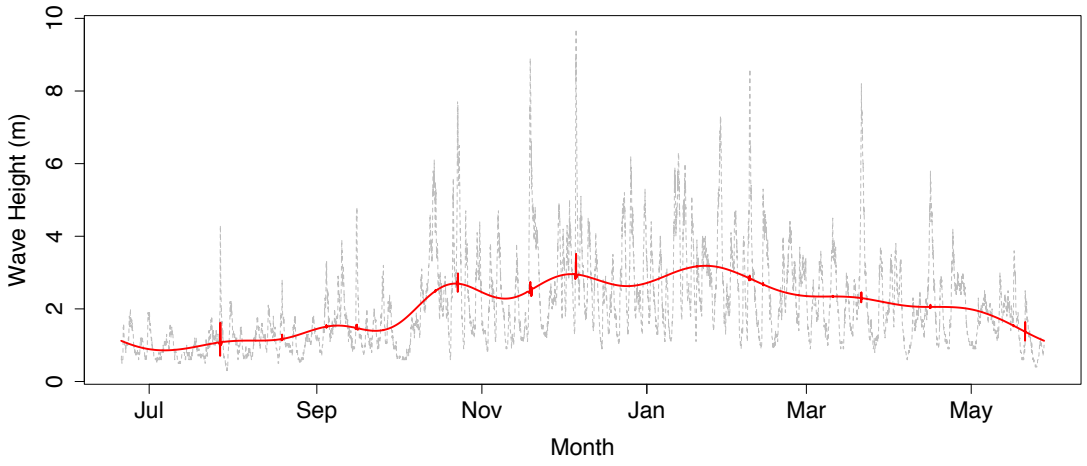

Figure 3.6.3: Trend estimate for the North Atlantic wave data shown in solid line, with data shown in dashed line.

Next we describe additional plots that help to understand the second-order structure of the wave height data. Using the spectral estimate in Figure 3.6.2, we calculated the local autocovariance estimate of the series. In Figure 3.6.4, we see in the solid line the estimate using our methodology for the local variance function of the time series. The nonstationary nature of the variance is clear to see, with the summer months containing low variability, while winter months display much higher variability. As one way of determining the performance of our method, we can compare the variance estimate with the estimate obtained by using the detrended wave height data. The detrended wave height data is obtained by subtracting the wavelet thresholding trend estimate from the data. We then perform standard LSW inference on this series, using the same parameters as in our approach. In dashed line, we see the local variance estimate obtained in this way. We see that the two estimates agree, which is reassuring on two counts: first that our spectral estimate obtained using the differenced data is accurate, and secondly that the trend estimate obtained was accurate.

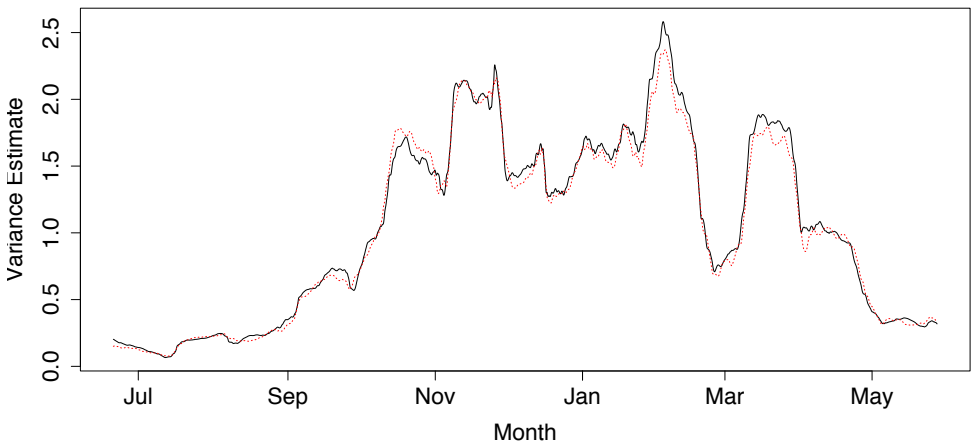

Figure 3.6.4: Local variance estimate for the wave height data. Solid line: obtained using our methodology. Dashed line: obtained from the detrended data.

Finally, we plot the autocorrelation function across 4 time points, which highlights the second-order nonstationary nature of the data, and showing how the structure of the autocorrelation varies considerably over time. In Figure 3.6.5, we see that estimate for the local autocorrelation function for the first observation recorded in the months of July 2005, October 2005, January 2006 and April 2006.

These are shown in the solid, dashed, dotted, and dashed and dotted lines respectively. We see that in general the series exhibits strong autocorrelation, which we may expect due to the observations being recorded at hourly intervals. Furthermore, we note that the shape of the autocorrelation function changes across the four months plotted.

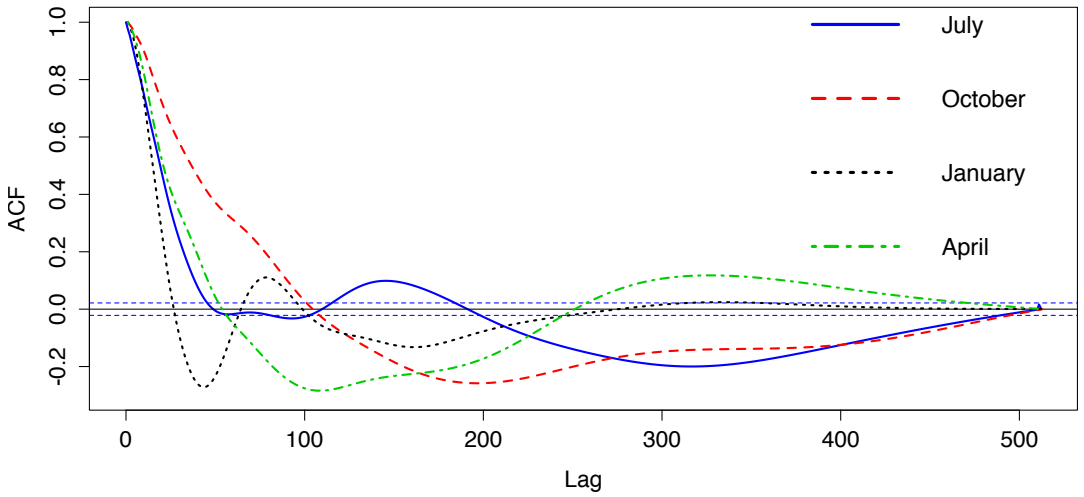

Figure 3.6.5: Local autocorrelation function estimate for the wave height data, at 4 different time points. Solid: July, dashed: October, dotted: January, dotted and dashed: April.

3.6.2 Baby Electrocardiogram Data

In this section, we re-analyse the baby electrocardiogram (ECG) data set that was first analysed in Nason et al. (2000). In Figure 3.6.6 left, we see a time series for the ECG reading of a 66-day-old infant. The series is available in the `wavethresh` package and was collected by the Institute of Child Health at the University of Bristol. The series can be seen to exhibit both nonstationary trend and second-order behaviour. As Nason et al. (2000) note, one reason for the nonstationarity is that the ECG varies considerably over time and changes significantly between periods of sleep and waking. In Nason et al. (2000), the authors first-difference the series to remove the trend, as shown in Figure 3.6.6 right, resulting in an approximately zero-mean time series. Then, LSW analysis is performed on the differenced series.

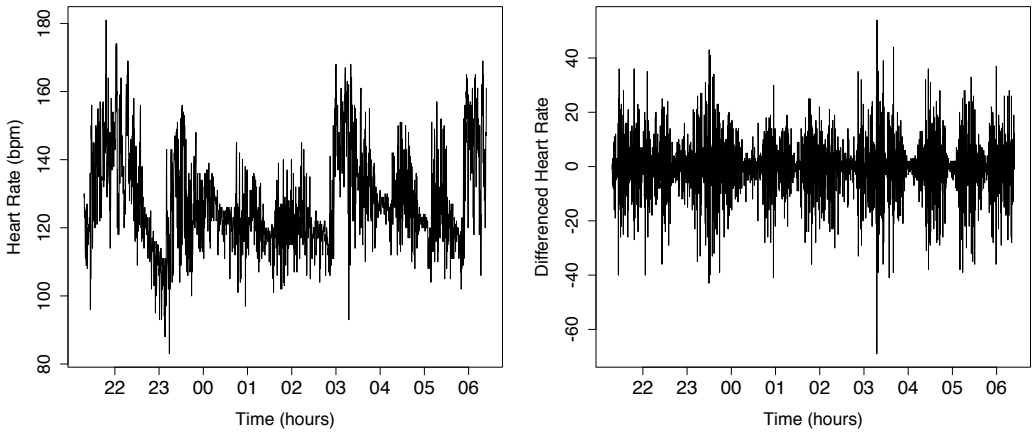

Figure 3.6.6: Left: original ECG data. Right: first-differenced ECG data.

We instead perform analysis on the original series, and note the similarities and differences to the original analysis in Nason et al. (2000). We use the Daubechies Least Asymmetric wavelet with 10 vanishing moments to analyse both series and the same bin width of size 128 for the running mean smoother, but as discussed we use a different matrix to correct for bias. In Figure 3.6.7 left, we see our analysis on the original series. The top plot is not scaled, while the bottom plot is individually scaled at each level. On the right, we see (our version of) the original analysis of Nason et al. (2000), again with top plot unscaled and bottom plot scaled individually. In line with previous discussion, only the finest 7 scales were used for analysis. From the top plots, we can see the effect that differencing has had on the resulting spectral estimate. In our estimate based on the original time series, we see that the power is spread fairly evenly across scales. However, in the spectral estimate of the differenced series, power is concentrated in the finest scales. From the bottom plots, we see that both estimates exhibit similar overall behaviour, with both evolving over time in a similar fashion.

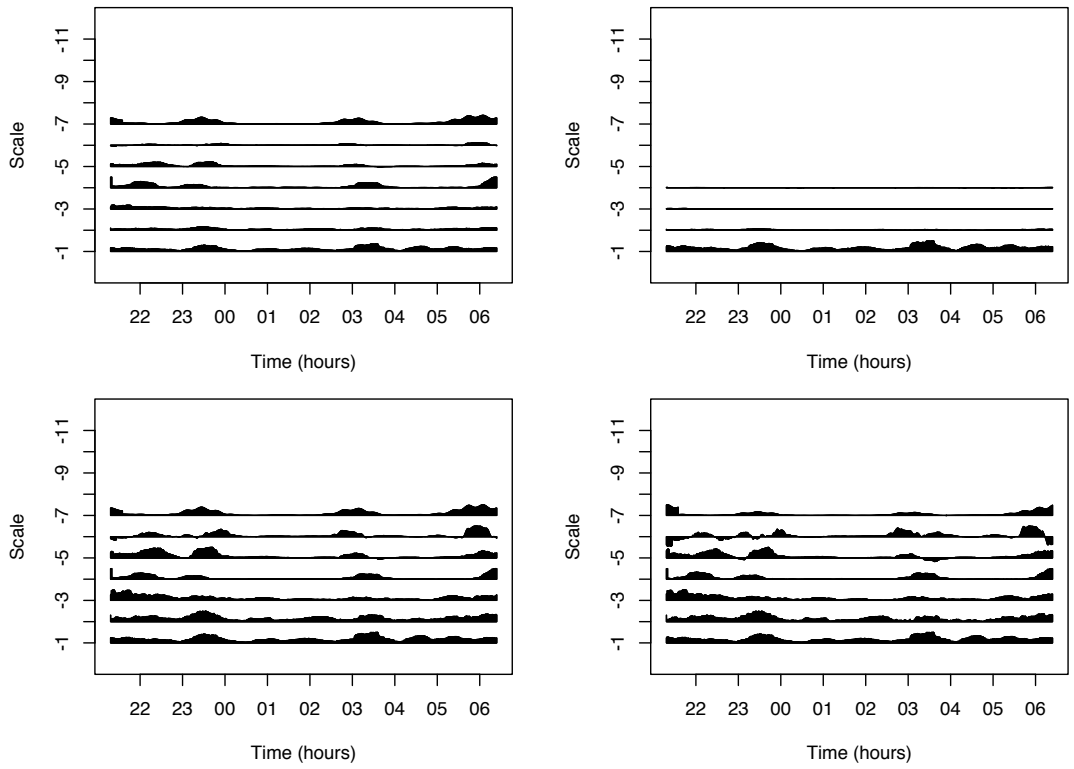

Figure 3.6.7: Comparison of the spectral estimates. Top left: spectrum estimate using our methodology, unscaled, bottom left: scaled individually. Top right: spectrum estimate of the differenced series, unscaled, bottom right: scaled individually.

Due to the difference in magnitude between the estimated spectra at different scales, some features of the original series can be identified more easily in our analysis. In Nason et al. (2000), an accompanying data set, the sleep state of the observed infant, is also used in the analysis. The sleep state is judged by a trained human observer, and is measured as either quiet (1), between quiet and active (2), active (3), or awake (4). A strong association between sleep state and spectral value is observed for scales -1 to -5 in Nason et al. (2000). However, using the original series instead of the differenced series can allow this association to become more apparent: in particular, consider the spectrum estimate at scale -4, enlarged in Figure 3.6.8. In solid line is the spectral estimate of the original series, while the dashed line shows the spectral estimate for the differenced series. The dotted line shows the sleep state of the infant. Our estimate correlates strongly with the sleep state, especially in highlighting periods of being awake (sleep state 4). In general, the "signal" of variability appears stronger in our analysis, which is due to the

differencing lowering the level of autocorrelation within the series. Furthermore, our estimate contains fewer negative spectral values, perhaps suggesting that the original time series is better represented as an LSW process as opposed to the first-differenced series.

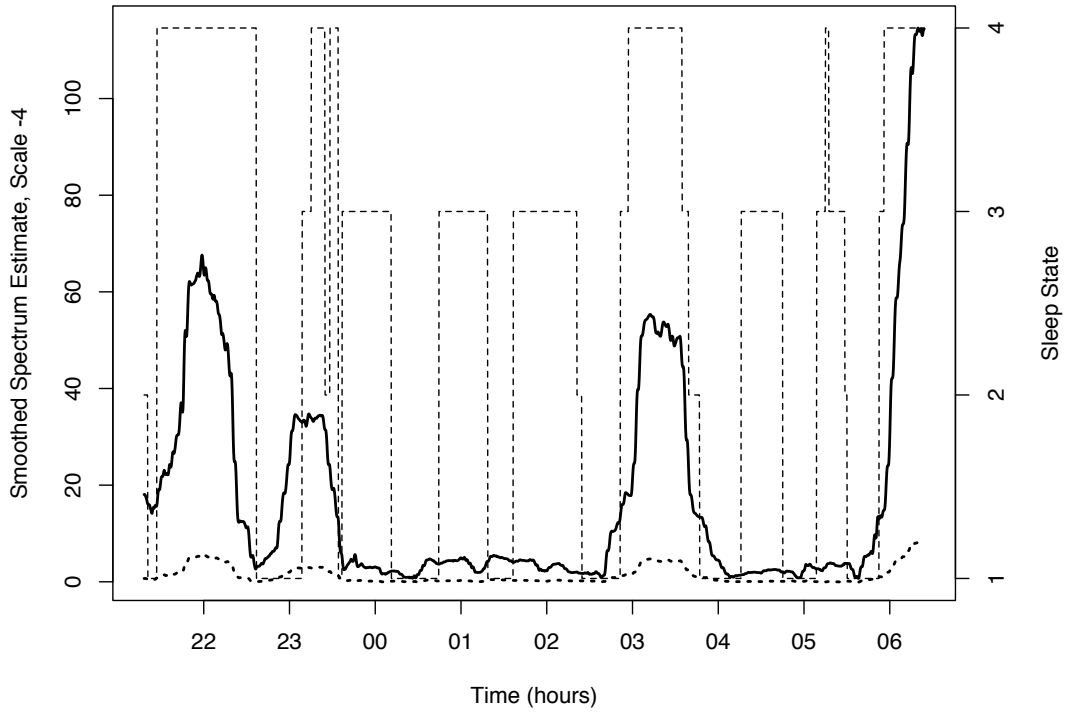

Figure 3.6.8: Estimate of the EWS at scale -4: solid line is for the original series, while the dashed line is for the differenced series. Sleep state shown in dotted line.

Lastly, we perform TI wavelet thresholding using the Daubechies Least Asymmetric wavelet with 4 vanishing moments to estimate the trend of the Baby ECG series. We analyse the finest 7 scales of the series, using a hard universal threshold of $\hat{\sigma}_{r,s}\sqrt{2\log(2048)}$, where $\hat{\sigma}_{r,s}$ is calculated using the spectral estimate in Figure 3.6.7 left. The resulting estimate is shown in the solid line in Figure 3.6.9. We see that in general the estimate is quite smooth, with more rapid changes in mean occurring at approximately 23:00, 03:00, and 06:00, corresponding to changes in sleep state.

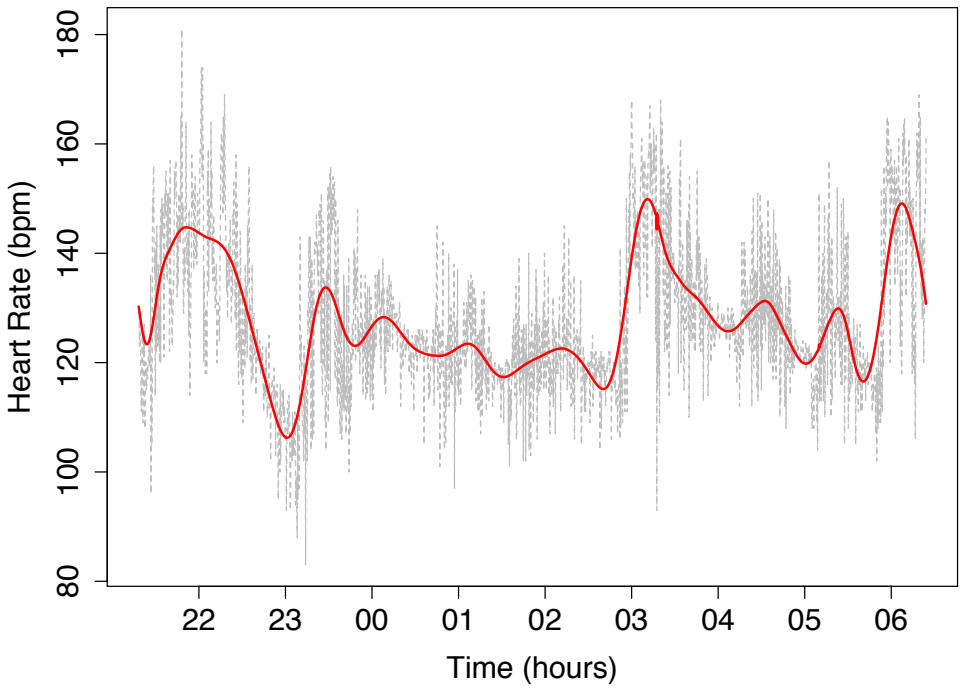

Figure 3.6.9: Trend estimate for the Baby ECG data shown in solid line, with data shown in dashed line.

3.7 Concluding Remarks

In this chapter, we have considered the problem of jointly modelling time-varying first and second-order properties of nonstationary time series. Our model employs the locally stationary wavelet model of Nason et al. (2000), used for modelling second-order nonstationarity, and adapts it to incorporate a nonstationary trend component. Using the common statistical technique of differencing, we have shown that we can consistently estimate the evolutionary wavelet spectrum and local autocovariance of the original series. Using these results, we have proposed a wavelet thresholding approach for the nonparametric estimation of the trend of the time series. The trend estimation methodology benefits from the information provided by the ability to consistently estimate the second-order structure of the time series, despite the presence of the trend component.

The simulation study showed the strong performance of the method, outperforming other wavelet-based methods. Using our methodology, we have analysed

two time series that highlight the strength of the method to provide information about the original time series, which can become masked when only considering the differenced time series. In particular, our analysis of the Baby ECG data set highlights the benefit of modelling the stochastic component of the original series for the purpose of identifying sleep states more readily.

Chapter 4

Trend Locally Stationary Wavelet Processes with Applications to Environmental Data

4.1 Introduction

In many contexts, it is common to encounter time series for which the mean and autocovariance of the series vary over time. Examples of time series that exhibit such nonstationary first and second-order behaviour span a broad range of applications, including financial time series (Fryzlewicz, 2005), climate data (Beaulieu and Killick, 2018), and meteorology (Shen, 2015). Consequently, methodology that can incorporate both first and second-order nonstationarity within the same modelling framework is appealing in numerous practical scenarios. Often, however, (nonstationary) first and second-order properties are estimated in isolation, rather than within a single framework. In this chapter, we propose a wavelet-based framework to *jointly* estimate the trend – the smooth, long-term behaviour – and the time-varying second-order structure of a time series.

Due to their many advantageous properties, wavelet methods have enjoyed popularity for many years in time series analysis. In particular, wavelets' localised

nature can provide sparse location-scale signal decompositions; for an overview of wavelet techniques, see for example Vidakovic (2009) or Nason (2008). Recently, there has been much focus in the area of locally stationary wavelet (LSW) modelling, introduced in the seminal work of Nason et al. (2000). The LSW framework provides a modelling approach for time series whose second-order structure evolves slowly over time. The model has been successfully applied in a number of areas in time series analysis since, including forecasting (Fryzlewicz et al., 2003), change-point analysis (Killick et al., 2013), and stationarity testing (Nason, 2013). A comprehensive overview of second-order nonstationary time series modelling can be found in Dahlhaus (2012).

Under the LSW framework a time series is assumed to have zero mean, and so typically the time series is detrended prior to applying the model. This extra step will have implications for data analysis, for example by introducing extra uncertainty, and potentially a bias in estimation. Furthermore, the task of estimating the mean function in the presence of nonstationary second-order structure is highly challenging. In principle, most standard techniques from nonparametric regression such as kernel or local polynomial estimates may be used. However, due to the nonstationarity of the error structure, theoretical and practical performance suffer.

There exists a body of work for mean estimation in the presence of time-varying dependent error structures. For example, von Sachs and MacGibbon (2000) use wavelet thresholding to estimate the first-order structure only. The threshold used is data-driven and depends on the unknown second-order structure, and hence the method can be difficult to implement in practice. In the semi-parametric setting, Dahlhaus (2000) and Dahlhaus and Neumann (2001) establish results where the mean function is time-varying and/or estimated. For literature on trend estimation by wavelet thresholding, see for example Brillinger (1994), Donoho and Johnstone (1995) or Donoho et al. (1995). For a review of other nonparametric regression methods including kernel smoothing and splines, see Härdle (1990).

Some techniques do consider mean estimation alongside parameters relating to second-order nonstationarity, see for example Tunyavetchakit (2010) for time-varying autoregressive processes. Beran and Feng (2002) propose parallel first and second-order estimation, but the second-order structure follows a parametric model and is assumed to be difference-stationary. However, for locally stationary series, many authors acknowledge that the topic "needs more investigation" (Dahlhaus, 2012, p. 56), and thus a principled approach to this challenge remains open.

In this chapter we propose a wavelet-based framework to model time series that exhibit time-varying first and second-order structure, inspired by the LSW model of Nason et al. (2000). The trend is considered to be a deterministic smooth function representing long-term (low frequency) patterns or systematic variations in the time series. By making modest assumptions on the trend, we develop rigorous theory to achieve consistent estimation of the key time-varying second-order quantities of interest.

Crucially, this includes a proof that a key wavelet matrix, utilised in LSW estimation theory, has a bounded inverse for all Daubechies compactly supported wavelets. Previously, this was only shown for the boundary cases of Haar and Shannon wavelets. Our approach adapts techniques used in Cardinali and Nason (2017), who investigate the properties of wavelet packet-based operators. Importantly, we correct a flaw in the author's arguments by making a key alteration, which enables a finer characterise of the properties of the wavelet matrix.

Furthermore, we propose a wavelet thresholding technique to consistently estimate the trend term in the presence of second-order nonstationarity. In practice, however, our estimators also work well for trends that depart from our assumptions. Finally, we provide new methodology for the boundary handling in the discrete wavelet transform employed during the estimation of first and second-order quantities. This is of particular interest when analysing first and second-order nonstationary time series which are not of dyadic length, as standard techniques can produce spurious artefacts at the boundaries of the time series.

This chapter is organised as follows. In Section 4.2 we give a brief overview of the LSW framework and propose the Trend Locally Stationary Wavelet Process model, discussing the key assumptions. Sections 4.3 and 4.4 introduce the methodology for estimating the second-order structure and trend in the model, and describe a new technique for boundary handling which brings improvements in both the first and second-order estimation procedures. In Section 4.5 we outline comprehensive numerical experiments to assess the performance of our proposed methodology. In Section 4.6, we examine the much-studied global mean sea temperature data set, providing new insights and highlighting the potential uses of our framework. Concluding remarks are given in Section 4.7, while proofs of results are contained in Appendix B.

4.2 The Trend Locally Stationary Wavelet Process Model

4.2.1 Background to LSW Processes

We begin by giving a brief overview of the original LSW modelling paradigm of Nason et al. (2000), which has been widely used for the modelling of second-order nonstationary – but first-order stationary – time series. This will form the basis for our more general trend nonstationary time series model. A locally stationary wavelet (LSW) process is defined as follows.

Definition 4.2.1. A triangular stochastic array $\{X_{t,T}\}_{t=0}^{T-1}$ for $T = 2^J \geq 1$ is in a class of locally stationary wavelet (LSW) processes if there exists a mean square representation

$$X_{t,T} = \sum_{j=-\infty}^{-1} \sum_{k \in \mathbb{Z}} w_{j,k;T} \psi_{j,k-t} \xi_{j,k}, \tag{4.2.1}$$

where j and $k \in \mathbb{Z}$ are scale and location parameters respectively, $\{\xi_{j,k}\}$ is a random orthonormal increment sequence, $\{\psi_{j,k-t}\}_{j,k}$ is a set of discrete non-decimated wavelets, and $\{w_{j,k;T}\}$ is a set of amplitudes. The quantities in representation

(4.2.1) possess the following properties:

1. $\mathbb{E}(\xi_{j,k}) = 0$ for all j, k.

2. $\mathrm{Cov}(\xi_{jk}, \xi_{lm}) = \delta_{jl}\delta_{km}$.

3. There exists, for each $j \leq -1$, a Lipschitz continuous function $W_j(z)$ for $z \in (0,1)$ which satisfies the following properties:

$$\sum_{j=-\infty}^{-1} |W_j(z)|^2 < \infty \text{ uniformly in } z \in (0,1);$$

the Lipschitz constants L_j are uniformly bounded in j and $\sum_{j=-\infty}^{-1} 2^{-j} L_j < \infty$. There exists a sequence of constants C_j such that for each T,

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

and where the sequence $\{C_j\}$ satisfies $\sum_{j=-\infty}^{-1} C_j < \infty$.

Since $\mathbb{E}(\xi_{j,k}) = 0$ for all j and k, the process mean $\mathbb{E}(X_{t,T}) = 0$ for all t. This means the LSW framework is only able to model time-varying *second-order* structure. As with classical time series theory, the second-order structure of an LSW process is encoded in the spectrum. The evolutionary wavelet spectrum (EWS) of an LSW process is defined as $S_j(z) := |W_j(z)|^2$ for rescaled time $z = k/T \in (0,1)$ and measures the contribution to variance at a particular rescaled time z and scale j. Since the W_j are Lipschitz continuous, the spectrum at level j, S_j, is also Lipschitz continuous; however Fryzlewicz and Nason (2006) and Van Bellegem and von Sachs (2008) extend the LSW model to consider piecewise constant spectra and those of bounded total variation respectively.

The EWS is estimated via the empirical wavelet coefficients of the time series, given by $d_{j,k;T} := \langle X_{t,T}, \psi_{j,k-t} \rangle = \sum_t X_{t,T}\psi_{j,k-t}$. The raw wavelet periodogram, defined by $I^j_{k,T} := |d_{j,k;T}|^2$ is a biased, inconsistent estimator of the EWS (Nason

et al. (2000), Proposition 4):

$$\mathbb{E}\left(I^j_{k,T}\right) = \sum_l A_{jl} S_l(k/T) + \mathcal{O}(T^{-1}), \tag{4.2.2}$$

$$\mathrm{Var}(I^j_{k,T}) = 2\left(\sum_l A_{jl} S_l(k/T)\right)^2 + \mathcal{O}(2^{-j}T^{-1}), \tag{4.2.3}$$

where the operator $A = (A_{jl})_{j,l<0}$ is given by $A_{jl} := \langle \Psi_j, \Psi_l \rangle = \sum_\tau \Psi_j(\tau)\Psi_l(\tau)$, and the autocorrelation wavelets are defined by $\Psi_j(\tau) := \sum_{k\in\mathbb{Z}} \psi_{j,k}\psi_{j,k-\tau},\ j < 0, \tau \in \mathbb{Z}$.

The vector of periodograms $\mathbf{I}(z) := \{I^l_{\lfloor zT \rfloor,T}\}_{l=-1,\ldots,-J}$ can be bias-corrected using the J-dimensional matrix $A_J := (A_{jl})_{j,l=-1,\ldots,-J}$ which is invertible. The vector of corrected periodograms $\mathbf{L}(z) := \{L^j_{\lfloor zT \rfloor,T}\}_{j=-1,\ldots,-J}$ with $\mathbf{L}(z) = A_J^{-1}\mathbf{I}(z)$ satisfies

$$\mathbb{E}\left(\mathbf{L}(z)\right) = \mathbb{E}\left(A_J^{-1}\mathbf{I}(z)\right) = \mathbf{S}(z) + \mathcal{O}(T^{-1}) \qquad \forall\ z \in (0,1), \tag{4.2.4}$$

where $\mathbf{S}(z) := \{S_j(z)\}_{j=-1,\ldots,-J}$. The raw wavelet periodogram is first smoothed and then corrected by A_J^{-1} to produce an asymptotically unbiased, consistent estimator. Smoothing can be carried out using a number of techniques, for example via a running mean as in Nason (2013) or using wavelet thresholding as in Nason et al. (2000).

The local autocovariance (LACV) function for an LSW process provides information about the covariance at a rescaled location $z = k/T \in (0,1)$. The LACV, $c(z,\tau)$, of an LSW process with EWS $\{S_j(z)\}$ is defined as $c(z,\tau) = \sum_{j=-\infty}^{-1} S_j(z)\Psi_j(\tau),\ \tau \in \mathbb{Z}, z \in (0,1)$. The LACV can be thought of as a decomposition of the autocovariance of a process over scale and rescaled time. The process autocovariance arises as the asymptotic limit of the LACV. In practice, the LACV is estimated by plugging in the smoothed, corrected estimate for the EWS into the equation for the LACV, which results in a consistent estimator.

4.2.2 Modelling Locally Stationary Time Series with Trend

In this section we introduce our proposed model which relaxes the zero mean assumption inherited from Assumption 1 in Definition 4.2.1 above. We will show that it is possible to incorporate a nonstationary polynomial trend term within the context of a second-order nonstationary time series. Furthermore, we will show that this polynomial assumption can be relaxed, facilitating modelling of time series with more complex first-order structure within our paradigm. This allows for self-contained, joint modelling of both first and second-order nonstationarity of a time series. Hence, it is not necessary to remove the trend before estimating second-order structure, and vice versa, we can obtain an estimate of first-order structure in the presence of the nonstationary second-order behaviour. To this end, we define the polynomial trend locally stationary wavelet process as follows.

Definition 4.2.2. A *polynomial trend LSW process* $\{X_{t,T}\}$, $t = 0, \ldots, T-1$, and $T = 2^J$ is a doubly-indexed stochastic process with the following representation in the mean square sense:

$$X_{t,T} = \mu\left(\frac{t}{T}\right) + \sum_{j=-\infty}^{-1} \sum_{k} w_{j,k;T} \psi_{j,k-t} \xi_{j,k}, \tag{4.2.5}$$

where the quantities in representation (4.2.5) possess the same properties as in Definition 4.2.1, and in addition the function μ is a polynomial of degree p, i.e.

$$\mu\left(\frac{t}{T}\right) = P_p\left(\frac{t}{T}\right) = \sum_{i=0}^{p} a_i \frac{t^i}{T^i}, \qquad a_i \in \mathbb{R}.$$

Our model imposes the same assumptions on the LSW component as in Nason et al. (2000), allowing for locally stationary second-order structure, while also permitting nonstationary first-order behaviour by incorporating a polynomial mean function. Using the model defined in Equation (4.2.5), we build upon the rigorous theory of Nason et al. (2000) in order to estimate the nonstationary first and second-order structure of the time series. Note that the assumption that $T = 2^J$

is a theoretical one; in practice, using the boundary handling method described in Section 4.4.4, we can analyse time series of any length.

4.3 Spectral Estimation Theory

Given the process model definition in Equation (4.2.5), we now turn to the question of estimating the appropriate quantities of the model. In this section, we discuss estimation of the evolutionary wavelet spectrum $S_j(z) = |W_j(z)|^2$ and local autocovariance function. In Nason et al. (2000), rigorous estimation theory is provided for the case where the time series does not display a trend. We extend this work to allow for the inclusion of a trend by examining three separate scenarios: firstly, we examine the case where the trend is a polynomial of degree less than the order of the wavelet used in the LSW process (4.2.5); secondly, the case where the trend is a higher degree than the wavelet; and lastly, the case of a general Hölder continuous trend.

Using wavelets to model the second-order nonstationarity allows the incorporation of smooth trends due to the ability of the wavelet to 'cancel out' this trend. This means we circumvent the problem of detrending the series first, which is challenging due to the second-order nonstationarity. In all three scenarios we provide appropriate theoretical results to enable consistent estimation of both the EWS and local autocovariance of the time series. Given that trends are usually assumed to be slowly-varying, smooth functions, such as low-order polynomials (see e.g. Priestley (1983); Craigmile et al. (2005); Kallache et al. (2005)), our model in (4.2.5) is applicable to a wide range of data scenarios. In practice, this can serve as a good approximation even when we depart from the given assumptions.

4.3.1 Spectral Estimation with Low-Order Polynomial Trends

First, we examine the case of a low-order polynomial trend. We show that we can use the same estimation procedure as in Nason et al. (2000) to obtain an

asymptotically unbiased, consistent estimator of the EWS. This can then be used to obtain a consistent estimator of the LACV.

Recall that a function $\psi \in L^2(\mathbb{R})$ is said to have m vanishing moments if it satisfies $\langle x^l, \psi(x) \rangle = \int x^l \psi(x) dx = 0$, for $l = 0, 1, \ldots, m-1$. If a wavelet has m vanishing moments, then all wavelet coefficients of any polynomial of degree $m-1$ or less will be zero. The commonly used extremal phase (EP) and least asymmetric (LA) Daubechies wavelet families have k vanishing moments for a filter of length $2k$. For example, the Haar wavelet has one vanishing moment and can only annihilate constant behaviour, while the Daubechies EP wavelet with 10 vanishing moments can annihilate polynomial behaviour up to degree 9.

Assume that the trend is a polynomial of degree $p < m$, where m is the number of vanishing moments of the wavelet defined in the LSW process representation (henceforth referred to as the generating wavelet) in Equation (4.2.5). Since second-order estimation is performed using these wavelet coefficients, it will be unaffected by the trend, and hence we can obtain an unbiased estimate of the spectrum in the usual way.

Lemma 4.3.1. When the degree p of $\mu(t/T)$ is less than m, the number of vanishing moments of the wavelet, the expectation and variance of the non-boundary raw wavelet periodogram is unaffected by the trend, and given by Equation (4.2.2) and Equation (4.2.3) respectively.

We can correct the raw wavelet periodogram estimator in the usual way with A_J^{-1} as in Equation (4.2.4) to obtain an asymptotically unbiased estimator, just as in the original LSW model. Hence, as long as the trend of the time series can be represented by a low-order polynomial and we use a high enough order wavelet, we can still apply the standard EWS estimation procedure without incurring any additional bias.

Remark 2. In practice, only the non-boundary wavelet coefficients will be free from trend, while the boundary wavelet coefficients will still contain the trend. This is due to the way in which the non-decimated discrete wavelet transform

(DWT) is computed at the boundary values of the time series, see for example Section 4.11 of Percival and Walden (2006). In Section 4.4.4 we will address this issue by modifying standard boundary handling methods for estimation at the boundaries of the time series.

As in the original LSW model, resulting spectral estimates are inconsistent and must be smoothed. To achieve consistency we can simply follow the smoothing strategy taken in Nason et al. (2000): for each fixed scale j, the periodogram $I^j_{k,T}$ (which is scaled χ^2-distributed) is smoothed as a function of $z = k/T$ using, for example, DWT shrinkage or translation invariant (TI) denoising of Coifman and Donoho (1995). This is analogous to the results described in Section 3.3.5 of Chapter 3. The following results, which hold for Gaussian LSW processes (and in turn polynomial trend Gaussian LSW processes), describe how consistency is achieved through DWT shrinkage. Work in von Sachs et al. (1997) explains how to smooth using an orthonormal second stage wavelet basis of $L_2([0,1])$, denoted $\{\tilde{\psi}_{rs}\}$.

Smoothing is achieved using nonlinear thresholding of the empirical wavelet coefficients, $\hat{v}^j_{rs}$, of $I^j_{\lfloor zT \rfloor}$ and then inverting using the inverse matrix A^{-1} to give the estimate $\hat{S}_j(z)$. Concretely, denote by $\tilde{I}^j_{\lfloor zT \rfloor}$ the wavelet thresholding estimator of the periodogram at scale j and rescaled time z, calculated using the threshold $\lambda^2(j,r,s,T) = \mathrm{Var}(\hat{v}^j_{rs}) \log^2(T)$. Then, the estimator $\hat{S}_j(z)$ of the EWS $S_j(z)$ is given by

$$\hat{S}_j(z) = \sum_{l=-J}^{-1} \tilde{I}^l_{\lfloor zT \rfloor} A^{-1}_{lj}.$$

In order to prove consistency for the estimator, we must investigate the properties of the matrix A. In particular, since we consider the case where $T \to \infty$, we are interested in the properties of A, when viewed as an operator acting on the Hilbert space $\ell^2(\mathbb{N})$. In the original work of Nason et al. (2000), the authors show that A is invertible and possesses a bounded inverse, for the special cases of the Haar and Shannon wavelets. This fact allows us to bound the error that arises when performing the inversion step in the estimation procedure.

In Cardinali and Nason (2017), the authors attempt to extend the result on bounded invertibility of A to all operators arising from inner products of wavelet packets. This more general result includes all Daubechies compactly supported wavelets. Unfortunately, the proof contained in Cardinali and Nason (2017) contains a flaw, which we correct here for the case of all Daubechies compactly supported wavelets. Crucially, we consider the related operator $B_{jl} := 2^{j/2}A_{jl}2^{l/2}$, rather than A directly, which leads to the following proposition.

Proposition 4.3.2. Let A be the autocorrelation wavelet inner product matrix constructed using a compactly supported Daubechies wavelet. The semi-infinite operator B, with entries given by $B_{jl} = 2^{j/2}A_{jl}2^{l/2}$ is a bounded, invertible operator on $\ell^2(\mathbb{N})$, with bounded inverse.

The proof of Proposition 4.3.2 follows the strategy of Cardinali and Nason (2017), with one key modification, which we now outline. First, we show that B is positive definite. Secondly, we show that the operator B possesses certain decay properties, which is the crucial change to the original proof of Cardinali and Nason (2017). This is because the operator A has unbounded entries on its diagonal, which are of order $\mathcal{O}(2^{-j})$, and therefore does not satisfy the necessary decay properties needed for the proof. By considering the related operator $B_{jl} = 2^{j/2}A_{jl}2^{l/2}$ instead, we ensure that B does have bounded diagonal entries.

Next, we utilise the results of Goodman et al. (1995), to show that we can construct a Cholesky factorisation of B which also satisfies decay properties. The inverse of this Cholesky factorisation itself has decay properties, which allows us to show that the inverse of B is bounded. Then, Proposition 4.3.2 immediately gives rise to the following corollary.

Corollary 4.3.3. Let A be the autocorrelation wavelet inner product matrix associated to a Daubechies compactly supported wavelet. Then, for some positive constant C,

$$|A_{jl}^{-1}| \leq C2^{j/2}2^{l/2}.$$

Corollary 4.3.3 enables us to bound the entries of the A-inverse operator, which can be used when bounding the error of the LSW process-based estimators. In fact, it allows for a finer characterisation of the off-diagonal decay of the entries of A^{-1} in terms of the indices j and l. This is a stronger result than simply showing that $||A^{-1}||$ is bounded, allowing for improved rates of convergence in LSW estimators. This leads to an extension of Theorem 4 and Proposition 5 of Nason et al. (2000) – which are only applicable for Haar and Shannon wavelet zero-mean LSW processes – to trend LSW processes generated by any wavelet in the Daubechies compactly supported family.

Proposition 4.3.4. Let $\tilde{\psi}$ be a wavelet of bounded variation, with $2^r = o(T)$ for wavelet coefficients $\hat{v}^j_{rs}$. Suppose that $S_j(k/T) \leq D2^{5j/6}$ for some constant D. For a polynomial trend Gaussian LSW process generated by any Daubechies compactly supported wavelet, with threshold given by $\lambda^2(j,r,s,T) = \mathrm{Var}(\hat{v}^j_{rs}) \log^2(T)$, for each fixed j,

$$\mathbb{E}\left[\int_0^1 \left(\hat{S}_j(z) - S_j(z)\right)^2 dz\right] = \mathcal{O}\left(2^j \log^2(T) T^{-2/3}\right). \tag{4.3.1}$$

Lastly, we can consistently estimate the LACV by using the EWS estimate.

Proposition 4.3.5. Suppose the assumptions of Proposition 4.3.4 hold, and define $\hat{c}(z,\tau)$ by replacing $S_j(z)$ with $\hat{S}_j(z)$ in the equation for the LACV, i.e.

$$\hat{c}(z,\tau) = \sum_{j=-J}^{-1} \hat{S}_j(z)\Psi_j(\tau). \tag{4.3.2}$$

Then, $\hat{c}(z,\tau)$ is a consistent estimator of $c(z,\tau)$, since for each fixed $\tau \in \mathbb{Z}$,

$$\mathbb{E}\left[\int_0^1 \left(\hat{c}(z,\tau) - c(z,\tau)\right)^2 dz\right] = \mathcal{O}\left(\log^2(T) T^{-2/3}\right).$$

With these results, we have shown that in the case of a low-order polynomial trend, we can still consistently estimate the evolutionary wavelet spectrum and local autocovariance on the non-boundary values of the wavelet periodogram. The

assumption on the decay rate of the EWS that $S_j \leq D2^{5j/6}$ is necessary for the proof of Proposition 4.3.4, and hence Proposition 4.3.5 as well. This assumption is, for example, satisfied for the white noise process where $S_j = 2^j$. The assumption is required in order to control the mean squared error in estimating the EWS using the J available scales, as we do not have access to all (infinite) scales in practice. This assumption is the weakest possible assumption that can be used, and is chosen to balance the mean squared error rate (up to the log factor) of $\mathcal{O}(\log^2(T)T^{-2/3})$ that arises from the wavelet thresholding procedure. Note that this is a slightly weaker assumption than that of Fryzlewicz and Nason (2006) and Sanderson et al. (2010), who instead assume that $S_j \leq D2^j$.

The extension of consistency results to all Daubechies compactly supported wavelets is in itself an important contribution to the area of LSW analysis, and is due Proposition 4.3.2. This result can be used to extend much of the current theory of LSW processes – which previously was only valid for Haar and Shannon wavelets – to all compactly supported Daubechies wavelets. This includes, for example, the work on bivariate LSW processes of Sanderson et al. (2010), and the extension of the LSW model to spectra of bounded total variation of Van Bellegem and von Sachs (2008).

4.3.2 Spectral Estimation with High-Order Polynomial Trends

Next, we consider the case of higher-order polynomial trends with respect to the generating wavelet, which allows for the modelling of less smooth trends. The key idea is to use a smoother wavelet to analyse the series, which ensures that the trend is removed from the wavelet coefficients. By calculating the bias that this causes, we can correct the raw wavelet periodogram analogously to Equation (4.2.4) to obtain an unbiased estimate.

If the polynomial trend of the time series is a higher degree than the wavelet generating the LSW process, then the standard non-decimated wavelet transform will not zero out the trend, causing bias in our estimation. For example, if the Haar

wavelet is the generating wavelet, and the time series exhibits a linear trend, then additional bias will be incurred in the computation of the raw wavelet periodogram, since the Haar wavelet is only capable of annihilating constants. If we instead use the Daubechies EP wavelet with 2 vanishing moments, the linear trend would be removed by the wavelet. Then, the bias of the raw wavelet periodogram will be quantified in terms of the mismatch between the generating wavelet and analysing wavelet. The next result shows we can correct for this bias, and extends the work of Gott and Eckley (2013), which examined the effect of wavelet misspecification on EWS estimation.

Lemma 4.3.6. Suppose that the polynomial LSW process is generated by the wavelet ψ^0 with m_0 vanishing moments, and let ψ^1 be a Daubechies compactly supported wavelet with m_1 vanishing moments. Let p be the degree of the polynomial trend, and suppose that $m_1 > p$. Define the non-decimated wavelet coefficients to be $d_{j,k} = \sum_t X_t \psi^1_{j,k-t}$. Let the matrix $C^{1,0}$ be the inner product matrix of autocorrelation wavelets whose j,l-th entry is given by

$$C^{1,0}_{jl} = \sum_{\tau} \Psi^1_j(\tau)\Psi^0_l(\tau),$$

where the superscript denotes the fact that the underlying wavelet is different. Then,

$$\begin{aligned} \mathbb{E}(d^2_{j,k}) &= \sum_l C^{1,0}_{jl} S_l\left(k/T\right) + \mathcal{O}(T^{-1}), \\ \mathrm{Var}(d^2_{j,k}) &= 2\left(\sum_l C^{1,0}_{jl} S_l\left(k/T\right)\right)^2 + \mathcal{O}(2^{-j}T^{-1}). \end{aligned}$$

Provided $C^{1,0}$ is invertible, Lemma 4.3.6 implies that, for the vector of periodograms $\mathbf{I}(z) := \{I^l_{[zT],T}\}_{l=-1,\ldots,-J}$, and the vector of corrected periodograms $\mathbf{L}(z) := \{L^j_{[zT],T}\}_{j=-1,\ldots,-J}$ with $\mathbf{L}(z) = C_J^{1,0^{-1}}\mathbf{I}(z)$,

$$\mathbb{E}\left(\mathbf{L}(z)\right) = \mathbb{E}\left(C_J^{1,0^{-1}}\mathbf{I}(z)\right) = \mathbf{S}(z) + \mathcal{O}(T^{-1}) \qquad \forall\ z \in (0,1), \tag{4.3.3}$$

where $\mathbf{S}(z) = \{S_j(z)\}_{j=-1,\ldots,-J}$. Hence, we do not have to restrict to assuming the trend has a lower order than the wavelet that generates the LSW process. In this scenario, we can perform the following altered estimation of the EWS. First, we apply the non-decimated wavelet transform using a *higher order wavelet*, removing the trend. Next, we form the periodogram and perform smoothing. Finally, we correct the estimate by premultiplying by the appropriate inverse matrix $(C^{1,0})^{-1}$.

Correcting in this fashion can be beneficial even when the time series contains no trend. Smoother wavelets have a shorter support in the Fourier domain which means they experience less spectral leakage, where power can leak into nearby scales. For example, a Haar wavelet estimate will experience more leakage than using a smoother wavelet like the EP10 wavelet.

If the LSW process is generated by a Haar wavelet and we use a Haar wavelet to analyse the series, then the spectral estimate may be prone to leakage and the accuracy of the estimate will suffer as a result. However, if the process is generated by a EP10 wavelet and we use the EP10 wavelet to analyse, the estimate will exhibit less leakage. This is due to the fact that the off-diagonal entries of the Haar A matrix exhibit slower decay than that of the EP10 A matrix. Hence, the off-diagonal entries of the corresponding C matrix will be smaller than those of the less smooth wavelet. Therefore, it can be beneficial, in the case where the generating wavelet is not very smooth, to use a smoother wavelet to analyse, and then correct the estimate with the corresponding C-inverse matrix, rather than using the generating wavelet to analyse and the usual A-inverse matrix. Using the smoother wavelet to analyse helps to reduce leakage, while still achieving an unbiased estimate by correcting in the appropriate way. This leads to a more accurate spectral estimate: in Section 4.5.4 we provide numerical experiments to verify this observation.

To further illustrate this point, we compute the entries of the A and C matrices discussed above for the case where $J = 4$, where the entries are accurate to 3 decimal places. Denote by A_H and A_{EP} the A-matrices constructed using Haar

and EP10 wavelets respectively, and let $C_{EP,H}$ denote the C-matrix computed from a Haar generating and EP10 analysing wavelet. Note that $C^{1,0} = (C^{0,1})^{\mathsf{T}}$ by definition, and so $C_{H,EP}$ is omitted for brevity. From the matrices below, we see that the Haar A matrix exhibits slower decay than the EP10 A matrix, while the corresponding C matrix has a level of decay in between the two A matrices.

$$A_H = \begin{bmatrix} 1.5 & 0.75 & 0.375 & 0.188 \\ 0.75 & 1.75 & 1.125 & 0.563 \\ 0.375 & 1.125 & 2.875 & 2.063 \\ 0.188 & 0.563 & 2.063 & 5.438 \end{bmatrix}, \quad A_{EP} = \begin{bmatrix} 1.839 & 0.322 & 0.000 & 0.000 \\ 0.322 & 3.035 & 0.643 & 0.001 \\ 0.000 & 0.643 & 6.070 & 1.285 \\ 0.000 & 0.001 & 1.285 & 12.141 \end{bmatrix},$$

$$C_{EP,H} = \begin{bmatrix} 1.621 & 0.653 & 0.184 & 0.047 \\ 0.605 & 2.106 & 1.164 & 0.357 \\ 0.226 & 0.967 & 3.828 & 2.264 \\ 0.121 & 0.308 & 1.834 & 7.479 \end{bmatrix}.$$

Intuitively, there is strong evidence to suggest that the matrix $C^{1,0}$ is invertible. In Appendix B, we show that the inverse of the operator A is bounded for all Daubechies compactly supported wavelets. The A-matrix is the Gram matrix of the set $\{\Psi_j(\tau)\}_{j\leq -1}$ of linearly independent autocorrelation wavelets. In our case, the matrix $C^{1,0}$ is the cross-Gram matrix of two sets of autocorrelation wavelets $\{\Psi_j^0\}_{j\leq -1}$ and $\{\Psi_j^1\}_{j\leq -1}$. Intuitively, forming the cross-Gram matrix uses a mix of two wavelet families, and we might expect that the properties of the cross-Gram matrix are a 'mixture' or 'average' of the properties of the two Gram matrices associated to the two wavelet families.

For example, when $J = 10$, the numerically calculated condition number of the Haar A matrix is 860.47, while it is equal to 393.12 for the A matrix constructed using the Daubechies extremal phase wavelet with 10 vanishing moments. For the two C matrices ($C^{1,0}$ and $C^{0,1}$) constructed using a combination of these two wavelets, the condition numbers are 596.97 and 610.97 respectively. These numbers

are close to the average of the two A matrix condition numbers. This suggests that the C matrices should also be invertible, with bounded inverse.

4.3.3 Spectral Estimation with Non-Polynomial Trends

A polynomial trend may not be an appropriate assumption for a given time series. In this section, we investigate how we can modify the spectral estimation procedure when we place a Hölder continuous smoothness assumption on the trend. Often, wavelet methods for curve estimation (distinct from trend estimation where more general mean functions are considered) assume that the mean function lies in some Besov space (see for example von Sachs and MacGibbon (2000) and Neumann and von Sachs (1995)), of which Hölder continuous functions are a special case.

Intuitively, the wavelet coefficients are calculated using difference operations, and so if the trend function is sufficiently smooth, we expect that these wavelet coefficients will be reasonably small. When using these coefficients to estimate the EWS – in some of the finer scales at least – where we use less of the data to calculate the wavelet coefficients, and where local changes are small, we could expect the estimate to remain reasonably accurate.

Formally, consider the class of Hölder continuous functions with exponent β, where $0 < \beta \leq 1$, in place of the polynomial trend assumption in Definition 4.2.2. For example, $\mu(z) = z^{\beta}$ is Hölder continuous with exponent β, and an exponent of $\beta = 1$ corresponds to Lipschitz continuity. Again, this seems a reasonable assumption to make, given that trend functions are generally assumed to be smooth and slowly-evolving – for example, sinusoids are Lipschitz continuous. As the following lemma shows, we accumulate bias in the raw wavelet periodogram proportional to the level of the wavelet transform and the smoothness of the trend, which is governed by β.

Lemma 4.3.7. For any Hölder continuous trend LSW process with exponent β,

the expectation of the raw wavelet periodogram is given by

$$\mathbb{E}\left(I_{k,T}^{j}\right) = \sum_{l} A_{jl} S_l(k/T) + \mathcal{O}(2^{-j(2\beta+1)} T^{-2\beta}) + \mathcal{O}(T^{-1}).$$

Thus we introduce extra bias due to the trend, which we must account for in our estimation procedure, i.e. when smoothing and correcting. To smooth the wavelet periodogram, we propose to use a simple running mean, as in Nason (2013) and Sanderson et al. (2010), with bin width size $2n + 1$. The smoothed wavelet periodogram is thus defined as

$$\tilde{I}_k^j = \frac{1}{2n+1} \sum_{m=-n}^{n} I_{k+m}^j. \tag{4.3.4}$$

For simplicity we use a running mean, but one could just as easily use wavelet thresholding, and achieve similarly consistent estimation. We must now correct the smoothed wavelet periodogram using the inverse matrix A^{-1}. When we do this, we alter the standard estimation procedure by only correcting across a certain number of scales – in relation to T – in order to control for the bias caused by the trend and the smoothing step. One can view this use of a reduced number of scales as akin to a tapered estimator, which we utilise for both EWS and LACV estimation. Based on the above observations, we can now state our results related to the consistency of the smoothed estimators of the EWS and local autocovariance in the presence of a Hölder continuous trend.

Theorem 4.3.8. Suppose that $S_j(k/T) \leq D2^j$ for some constant D, and μ is Hölder continuous with exponent β, and let $J_0 = \alpha \log_2(T)$ for $\alpha \in (0, 1)$. The EWS estimator $\hat{S}_j(k/T)$, defined by

$$\hat{S}_j(k/T) = \sum_{l=-J_0}^{-1} \tilde{I}_k^l A_{lj}^{-1}, \tag{4.3.5}$$

is mean square consistent for each fixed scale j, provided that $n^{-1}T^{\alpha} \to 0$, $nT^{-1} \to 0$ and $T^{4\beta(\alpha-1)+\alpha} \to 0$ as $T \to \infty$ and $n \to \infty$. The mean squared error of $\hat{S}_j(k/T)$

is given by

$$\mathbb{E}\left(\hat{S}_j(k/T) - S_j(k/T)\right)^2 = \mathcal{O}(2^j n^{-1} T^{\alpha}) + \mathcal{O}(2^j n T^{-1}) + \mathcal{O}(2^j T^{4\beta(\alpha-1)+\alpha}).$$

Theorem 4.3.9. Suppose the assumptions of Theorem 4.3.8 hold. Let $\hat{S}_j(k/T)$ be the estimator of the EWS defined in Equation (4.3.5). Then, for each fixed $\tau \in \mathbb{Z}$, the estimator of the local autocovariance $c(k/T, \tau)$, defined by

$$\hat{c}(k/T, \tau) = \sum_{j=-J_0}^{-1} \hat{S}_j(k/T)\Psi_j(\tau) \tag{4.3.6}$$

is mean square consistent, provided that $n^{-1}T^{\alpha} \to 0$, $nT^{-1} \to 0$ and $T^{4\beta(\alpha-1)+\alpha} \to 0$ as $T \to \infty$ and $n \to \infty$.. The mean squared error of $\hat{c}(k/T, \tau)$ is given by

$$\mathbb{E}\left(\hat{c}(k/T, \tau) - c(k/T, \tau)\right)^2 = \mathcal{O}(n^{-1}T^{\alpha}) + \mathcal{O}(nT^{-1}) + T^{4\beta(\alpha-1)+\alpha}. \tag{4.3.7}$$

Note that these results hold for all Daubechies compactly supported wavelets, just as in Propositions 4.3.4 and 4.3.5. As in Propositions 4.3.4 and 4.3.5, we require an assumption on the decay rate of the EWS, which is necessary to ensure consistent estimation. In this case, we assume that $S_j(k/T) \leq D2^j$ in order to utilise results from Sanderson et al. (2010), who make the same assumption.

The third error term in Equation 4.3.7, corresponding to the error caused by the trend, will be asymptotically dominated by the first term, irrespective of the bin width n, for $\beta = 1$ and $\alpha < 3/4$. Therefore, the trend term can be seen to have minimal impact on the spectral estimation, provided it is smooth enough. In the presence of a smooth trend, we can still recover a consistent estimator of the EWS and LACV. We simply alter the standard estimation procedure by only using the finest J_0 scales. Further, one could just as easily use a different analysing wavelet to the generating wavelet as described in Section 4.3.2, and use the appropriate $C^{1,0^{-1}}$ matrix for bias correction.

Choice of contributing scales and bin width. When using the tapered estimator, we must choose a value of α, the proportion of scales over which we correct. Higher values of α ensure a decomposition over a larger number of scales. In practice, not all scales will be informative, and the number of informative scales will depend on both the data and choice of wavelet, as noted in Sanderson et al. (2010). In our experiments we have observed no significant bias incurred by using most scales, and propose using the proportion $\alpha = 2/3$. This is motivated by results in the next section on trend estimation and is in alignment with Sanderson et al. (2010).

The choice of the bin width parameter n in Equation (4.3.4) will also affect the quality of the estimate. We can choose to use level-dependent smoothing of the periodogram, and so a larger bin width is used in coarser scales. Choosing a bin width in this way will still ensure consistent estimation, and further is natural as coarser scales will experience stronger autocorrelation.

The results in this section establish that we can still consistently estimate the nonstationary second-order structure in the presence of a trend, provided that the trend function is smooth. In practice, using a smoother wavelet to analyse the series is recommended, as it is able to better remove the trend of the series, whilst also helping to reduce leakage. Smoothing of the wavelet periodogram can be carried out using wavelet thresholding or a running mean smoother. We advocate the latter due to its simplicity and because only one parameter needs to be specified.

Finally, we note that many existing methods that utilise the LSW framework immediately extend to our new model. For example, the extension of the LSW model to spectra of bounded total variation in Van Bellegem and von Sachs (2008) is valid within our model. Furthermore, the test of second-order stationarity of Nason (2013) will still be applicable, provided the wavelet used is smooth enough to remove the trend of the time series. Similarly, locally stationary wavelet models for multivariate time series, such as Sanderson et al. (2010) and Park et al. (2014)

will be valid for any trend scenario described above.

4.4 Trend Estimation

Having described the necessary methodology for estimating the nonstationary second-order structure of the time series, we now address the question of estimating the trend of the series. If we assume that the trend function is smooth, a natural method for estimating it is to use wavelets, given that polynomial functions can be exactly represented using wavelet scaling functions (Vidakovic (2009)). Using a wavelet-based approach enables us to obtain a simple intuitive estimator from which we can derive appropriate confidence intervals.

4.4.1 Trend Estimation Theory

We follow the same strategy as Craigmile et al. (2004) and Craigmile et al. (2005), who apply wavelet thresholding in the long memory time series setting. Our problem can be posed as the following. The observed data vector $\boldsymbol{X} = (X_0, \ldots, X_{T-1})^{\mathsf{T}}$ is given by

$$\boldsymbol{X} = \boldsymbol{\mu} + \boldsymbol{\epsilon},$$

where $\boldsymbol{\mu}$ is the trend component and $\boldsymbol{\epsilon}$ is the LSW process term. If we use a wavelet that has a high enough order with respect to the trend, the wavelet transform of the time series will decompose it into a trend and noise component. This is because the non-boundary wavelet coefficients of a polynomial are zero due to the vanishing moments property, and so since the increments in an LSW process are zero mean, these coefficients have expectation zero. The boundary wavelet coefficients and the scaling coefficients will contain a contribution due to the trend alone.

Motivated by this observation, we can estimate the trend using a type of wavelet thresholding estimator. We perform the wavelet transform on the time series, choosing the coarsest scale j_0 to which we analyse the series. Then, we set the non-boundary wavelet coefficients to zero. Then, we perform the inverse wavelet

transform in order to obtain the estimator of the trend. This estimator captures the long-term behaviour of the time series (trend), but will also contain short-scale variability at the boundaries due to the inclusion of boundary wavelet coefficients.

Formally, let us write the wavelet transform of the time series in matrix form, from which we can derive the matrix form of the estimator. Given the time series is of length $T = 2^J$, the total number of (decimated) wavelet coefficients at each level j is given by $N_j = T2^j$. Then, the $B_j = \lceil (2m-2)(1-2^j) \rceil$ boundary coefficients are given by the outer coefficients, where m is the number of vanishing moments of the wavelet. If p is the order of the polynomial trend, then we can zero out this trend so long as $p < m$. The number of non-boundary coefficients is denoted by $M_j = N_j - B_j$. Denote by $\text{diag}(\mathbf{x})$ the diagonal matrix with $\mathbf{x}$ along the main diagonal, and let $\mathbf{0}_n$ and $\mathbf{1}_n$ denote a vector of n zeros or ones respectively. Let $A = \text{diag}(\mathbf{1}_{B_1}, \mathbf{0}_{M_1}, \mathbf{1}_{B_2}, \mathbf{0}_{M_2}, \ldots, \mathbf{1}_{B_J}, \mathbf{0}_{M_J}, \mathbf{1}_{N_J})$, and let I_T denote the $T \times T$ identity matrix. Using Daubechies wavelets ensures the DWT is orthogonal, and so we can partition the data vector $\mathbf{x}$ into a trend component and noise component, as

$$\boldsymbol{X} = W^{\mathsf{T}} W \boldsymbol{X} = W^{\mathsf{T}} A W \boldsymbol{X} + W^{\mathsf{T}} (I_T - A) W \boldsymbol{X} := \hat{\boldsymbol{\mu}} + \hat{\boldsymbol{\epsilon}}.$$

Analogously to Proposition 1 of Craigmile et al. (2004), the wavelet-based trend estimator is unbiased, and further we can show that the estimator is mean square consistent, in the case of both polynomial and Hölder continuous trend functions.

Proposition 4.4.1. For a polynomial trend of degree p, and using a wavelet with $m > p$ vanishing moments, $\mathbb{E}(\hat{\boldsymbol{\mu}}) = \boldsymbol{\mu}$.

Proposition 4.4.2. Let j_0 be the coarsest scale of the wavelet decomposition, where $j_0 = \alpha \log_2 T$ for some $\alpha \in (0,1)$. Assume that $S_j(z) \leq D2^j$ for all $z \in (0,1)$ and some constant D. Let m denote the number of vanishing moments of the wavelet used, and let p be the degree of the polynomial trend. Then, provided $p < m$, ignoring boundary effects, the trend estimator is consistent in the mean

square sense, that is:

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\left((\mu(t/T)-\hat{\mu}(t/T))^2\right)=\mathcal{O}(T^{-\alpha}).$$

Proposition 4.4.3. If μ is Hölder continuous with exponent β, then under the same conditions on α and S_j as Proposition 4.4.2, the trend estimator is mean-square consistent:

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\left(\left(\mu\left(t/T\right)-\hat{\mu}\left(t/T\right)\right)^2\right)=\mathcal{O}(T^{-\alpha})+\mathcal{O}(T^{2\beta(\alpha-1)}).$$

For a Lipschitz continuous trend, balancing the rates in Proposition 4.4.3 leads to taking $\alpha = 2/3$. We use this for both trend and second-order estimation in practice for simplicity. Under the conditions of Proposition 4.4.2 and subject to some mild regularity conditions, the trend estimate is multivariate Gaussian, with mean $\boldsymbol{\mu}$ and covariance $R\mathrm{Cov}(\boldsymbol{\epsilon})R^{\mathsf{T}} := R\Sigma R^{\mathsf{T}}$, where $R = W^{\mathsf{T}}AW$, following from asymptotic normality of the wavelet coefficients (Stevens, 2013, Theoerem 3.1.1). Full discussion of the appropriate assumptions can be found in Stevens (2013). LSW processes with random innovations from the exponential, gamma, inverse-Gaussian and F-distribution families all satisfy these assumptions.

4.4.2 Pointwise Confidence Intervals

Using the results of the previous section we can derive pointwise confidence intervals (CIs) for a polynomial trend. Let q_α denote the α-th quantile of the standard normal distribution, then a pointwise $100(1-\alpha)\%$ CI for the trend estimate is given by

$$\hat{\mu}\left(\frac{t}{T}\right)\pm q_{1-\alpha/2}\sqrt{\mathrm{Var}\left(\hat{\mu}\left(\frac{t}{T}\right)\right)},\qquad t=0,\ldots,T-1.$$

In practice we do not know the true variance. Instead, we have access to a consistent estimate of it on the non-boundary values of the trend estimate, since we have a consistent estimate of the local autocovariance of the time series. Thus

we can construct confidence intervals for the trend using this estimate in a rather straightforward manner. This is in contrast to nonlinear wavelet thresholding techniques, for which construction of confidence intervals is far from straightforward. In practice, we may need to regularise the covariance matrix to ensure positive-definiteness. This can be achieved straightforwardly using the method of Rothman (2012), for example, which is explained in more detail in Section 5.3.2 of Chapter 5.

Thus, within our framework we can straightforwardly calculate an estimate and confidence intervals for the trend. In turn this could be used for hypothesis testing of the presence of trend. Although we have assumed a low-order polynomial for the trend, we shall see in the simulation results in Section 4.5 that even if we depart from this assumption, the trend estimator is still able to perform well. This is due to the fact that a polynomial can serve as an accurate approximation to a more complicated trend function, and the ability of smooth wavelets to zero out the trend, as explained in Section 4.3.3.

4.4.3 Simultaneous Confidence Intervals

We can also compute simultaneous $100(1-\alpha)\%$ confidence intervals for the trend estimate. Denote the vector of standard errors of the trend estimate by $\mathbf{v}$, i.e.

$$\mathbf{v} = \left(\sqrt{\mathrm{Var}(\hat{\mu}_0)}, \sqrt{\mathrm{Var}(\hat{\mu}_1)}, \ldots, \sqrt{\mathrm{Var}(\hat{\mu}_{T-1})}\right)^{\mathsf{T}}.$$

Recall that the critical value is calculated as the value δ that satisfies

$$1-\alpha = \Pr\left(\hat{\boldsymbol{\mu}} - \delta\mathbf{v} \leq \boldsymbol{\mu} \leq \hat{\boldsymbol{\mu}} + \delta\mathbf{v}\right) = 1 - 2\Pr\left(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}} > \delta\mathbf{v}\right),$$

using the symmetry and continuity properties of the multivariate normal distribution. Therefore, δ is calculated to satisfy

$$\Pr\left(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}} > \delta\mathbf{v}\right) = \alpha/2.$$

Assuming that the true spectral structure is known (in practice we have an asymptotically unbiased, consistent estimate), then

$$\boldsymbol{\mu} - \hat{\boldsymbol{\mu}} \sim \mathcal{N}(\mathbf{0}, R\Sigma R^{\mathsf{T}}).$$

Hence we calculate δ as the value such that the probability of this multivariate normal random variable exceeding $\delta\mathbf{v}$ is equal to $\alpha/2$, for which implementation is readily available (for example within the `mvtnorm` package in `R`).

4.4.4 Boundary Handling

Using the results from Section 4.3, we can obtain an unbiased, consistent spectral estimator for the non-boundary coefficients when the time series exhibits a smooth trend. However, this will not be true for the boundary coefficients as they will still contain the trend of the time series. This behaviour is not bias but simply a consequence of the way the boundary coefficients are handled when performing the non-decimated wavelet transform. The two most common methods for boundary handling are either to reflect the data at the boundaries or to periodise it – see Section 2.8 of Nason (2008) or Section 4.11 of Percival and Walden (2006) for more discussion. These two methods can be used effectively in a zero-mean time series, but are less suitable when the time series displays a trend.

We can adapt the reflective boundary handling in order to produce better estimates of the EWS in the coarser scales. We use reflection of the time series as this ensures that the spectral properties of the time series in the reflected part evolve slowly as we move from the end of the time series past into the reflected, extended part. This is more appealing than periodising the time series, as spectral properties at one boundary of the series may not be as similar to those at the other boundary, due to the nonstationarity of the time series. Starting with a time series of length T, we form a time series of length $4T$, whereby the original data is centred within the extended time series. That is, we add $3T/2$ data points

to the beginning, and end, of the original series.

We note that in most applications, the standard practice of boundary handling involves repeating the series half of the length in each direction, giving a series of length $2T$. However, we have found in our empirical analysis that using a series of length $4T$ offers stronger practical performance. One of the reasons for this is that in the coarser scales of the wavelet transform, the wavelet filter is often already longer than the length of the time series. Hence, in these coarse scales, the wavelet coefficients are formed only using boundary handling, which causes issues due to the trend of the time series. Therefore, extending the time series to length $4T$ ensures that a larger number of wavelet coefficients are not affected by the trend. This enables more accurate estimation in the coarser scales for the evolutionary wavelet spectrum.

We illustrate the boundary handling procedure for the right-hand boundary of the data with Figure 4.4.1. An analogous procedure is used to add data to the beginning of the series, and so here we focus on an illustration of adding data to the end of the series. In (a) we see a polynomial trend of length $T = 1024$ (in the example we use a deterministic series for illustrative purposes). We wish to extend the data at the right boundary in such a way that we continuously extend the trend in order to minimise the boundary effect caused when calculating the boundary non-decimated wavelet coefficients. To achieve this, we first (b) reflect the entire time series at the boundary, given by the dashed vertical line, motivated by the reasoning above, obtaining a series of length $2T = 2048$. Next, we multiply the boundary series (data points 1025 to 2048) by -1, to ensure that the gradient of the reflected trend part is similar to that at the edge of the time series (c). Finally in (d), we must add some value, c, to this reflected part, in order to align the edge of the time series with the boundary part. The value c used to align the boundary part with the original series is calculated using a simple pre-estimate of the trend function, obtained via a polynomial regression. From this we obtain a pre-estimate for the value of the trend at the boundary. The value c is given by

2 times this value, as this reverses the effect caused by multiplying the boundary series by -1, realigning the series.

Having performed this step, we can repeat this again to add another $T/2 = 512$ data points on the right side, giving a series of length $5T/2 = 2560$. We then repeat the whole procedure for the left side boundary, adding $3T/2 = 1536$ data points to the start of the series, giving the final series of length $4T = 4096$. In practice, we can apply the procedure when performing trend estimation, in order to produce a more reliable estimate at the boundaries of the time series. In this case, we no longer have strictly valid confidence intervals at the boundary locations. However, we have found that performing the boundary handling step significantly improves estimation at the boundaries by reducing the variability of the estimate.

Finally, we observe that this procedure can be easily adapted to time series whose length is non-dyadic. A time series of general length n is extended to one of dyadic length d, where $d = 4 \times 2^{\lfloor \log_2 n \rfloor}$. Indeed, the data example used in Section 4.6 is non-dyadic.

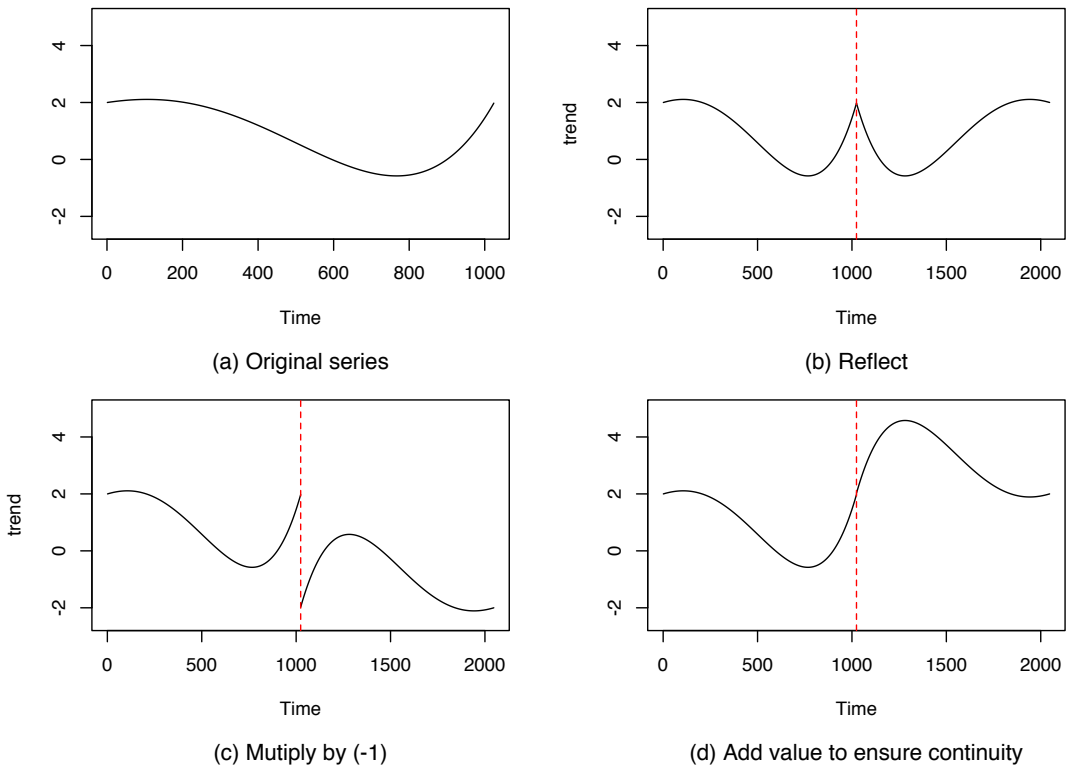

Figure 4.4.1: Boundary handling process for an example trend.

4.5 Simulation Study

To illustrate the ability to incorporate a smooth trend within the LSW framework, we perform a simulation study. In the simulation we define three different evolutionary wavelet spectra, shown in Figure 4.5.1. Spectrum S^1, studied in Nason (2008), displays coarse-scale, slowly-evolving sinusoidal behaviour with a fine-scale burst in power at time point 800; spectrum S^2 is a concatenation of moving average processes and contains power moving from fine to coarser scales, and was examined in Nason et al. (2000); spectrum S^3 contains slowly-evolving power at fine scales. The three spectra used in the simulations are explicitly defined as:

$$S_j^1(z) = \begin{cases} \sin^2(4\pi z) & \text{for } j = -5, z \in (0,1), \\ 1 & \text{for } j = -1, z \in (800/1024, 900/1024), \\ 0 & \text{otherwise,} \end{cases} \tag{4.5.1}$$

$$S_j^2(z) = \begin{cases} 1 & \text{for } j = -1, z \in (0/1024, 256/1024), \\ 1 & \text{for } j = -2, z \in (256/1024, 512/1024), \\ 1 & \text{for } j = -3, z \in (512/1024, 768/1024), \\ 1 & \text{for } j = -4, z \in (768/1024, 1), \\ 0 & \text{otherwise,} \end{cases} \tag{4.5.2}$$

$$S_j^3(z) = \begin{cases} \frac{1}{2} + \frac{1}{4}\sin(\pi z) - \frac{1}{2}\cos(3\pi z/2) & \text{for } j = -1, z \in (0,1), \\ \frac{1}{2} - \frac{1}{8}\sin(2\pi z) - \frac{1}{4}\cos(\pi z/2) & \text{for } j = -3, z \in (0,1), \\ 0 & \text{otherwise.} \end{cases} \tag{4.5.3}$$

We simulate time series $\{X_t\}_{t=0}^{T-1}$ of length $T = 2^{10} = 1024$ from LSW processes with those spectra, using Gaussian innovations. Depending on the simulation, different wavelets were used, in order to highlight the observations from Section 4.3. The simulations were performed in `R` (R Core Team (2019)) using the `wavethresh` package of Nason (2016b). For each spectrum, 100 LSW processes with different

trends added on were simulated.

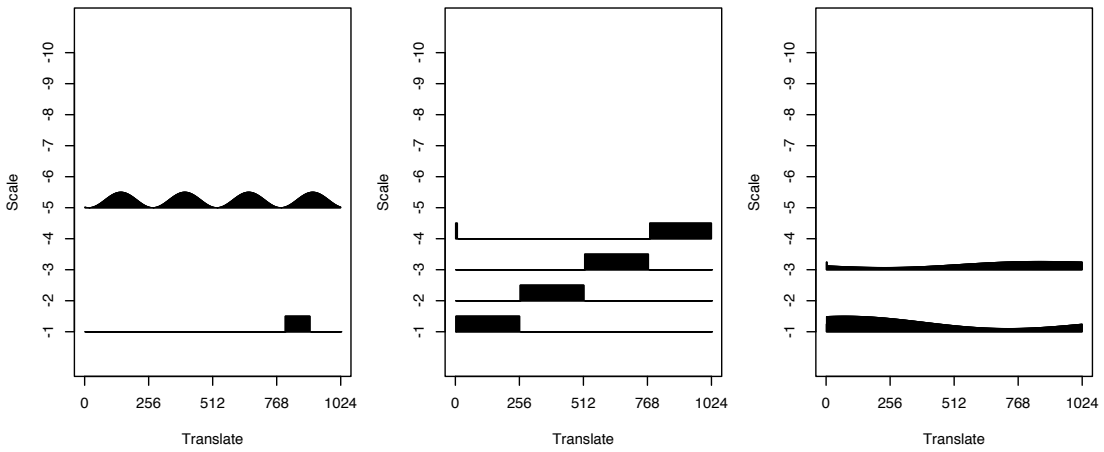

Figure 4.5.1: Spectra used in the simulation study. Left: S^1, sinusoid with "burst"; centre: S^2, concatenated moving average process; right: S^3, slowly-evolving fine-scale power.

We examine the effects of including a trend across three scenarios, corresponding to Sections 4.3.1, 4.3.2 and 4.3.3 of this chapter. In the first scenario, we add low-order trends to LSW processes simulated using a Daubechies EP wavelet with 4 vanishing moments, which corresponds to the setting of Section 4.3.1. In the second, we add low-order trends to LSW processes simulated using Haar wavelets, which corresponds to Section 4.3.2. In the final set of simulations, we add non-polynomial trends to LSW processes simulated using the Daubechies EP4 wavelet, corresponding to Section 4.3.3.

To assess the spectral estimation performance, for each realisation, the unsmoothed estimate of the EWS was calculated, which was then used to obtain an averaged estimate for the EWS across the 100 realisations. In alignment with the discussion in Section 4.3.3, we correct the EWS across the finest 7 scales and use the boundary handling procedure, which ensures that the boundary effects of the trend are minimised.

In each of the three scenarios, we also assess the trend estimation performance by reporting the averaged mean squared error of the trend estimate across the 100 realisations. The trend estimate is calculated using the Daubechies LA4 wavelet, with boundary handling applied. Our linear wavelet thresholding method is re-

ferred to as LWT in the comparison tables. We compare our method to three other trend estimation methods, two of which were used in the numerical comparisons in Chapter 3. We first compare to the wavelet-based method of von Sachs and MacGibbon (2000), which utilises a local MAD estimator in a wavelet thresholding procedure. As mentioned in Chapter 3, no code is available for the method, therefore we have implemented the method utilising `wavethresh` and following the description of the computation of the threshold in Section 2.5 of von Sachs and MacGibbon (2000). This method is referred to as MVSWT in the comparison tables. Secondly, we also compare to the spline-based method using the `R` function `smooth.spline`, where the smoothing parameter is chosen via cross validation. Lastly, we compare to a local polynomial regression estimator, calculated using the `loess` function in `R`. We use the default settings and a local quadratic polynomial for fitting the trend.

4.5.1 Low-Order Polynomial Trend, High-Order Generating Wavelet

In this section we add various low-order polynomial trends to LSW processes generated by the three different spectra. In particular, we use three different trends μ_t^1, μ_t^2, and μ_t^3 (where $\mu_t = \mu(t/T)$); polynomials of degree 1, 2, and 3 given by

$$\mu_t^1 = 4t, \qquad \mu_t^2 = 12t^2 - 12t, \qquad \mu_t^3 = 32t^3 - 48t^2 + 22t - 3. \tag{4.5.4}$$

A plot of example realisations from each trend and spectrum scenario is shown in Figure 4.5.2. We use the EP4 wavelet to analyse the series, the same wavelet that generates the LSW processes, which is a high enough order to remove the polynomial trends.

In each of the simulations, we compare the estimates of the EWS obtained when trends are added, to the estimate of the EWS obtained in the absence of any trend ("None" in Table 4.5.1), obtained in the standard way using the `ewspec3`

command within the `R` package `locits`. Table 4.5.1 contains the mean squared error across the four trends and three spectra, which shows that the trend has negligible impact. In fact, for the third spectrum, the sum of squared error is actually lower when a trend is added. From the simulation results, we see that there is no discernible difference in the estimation quality between the no trend case, and the case when a trend is added.

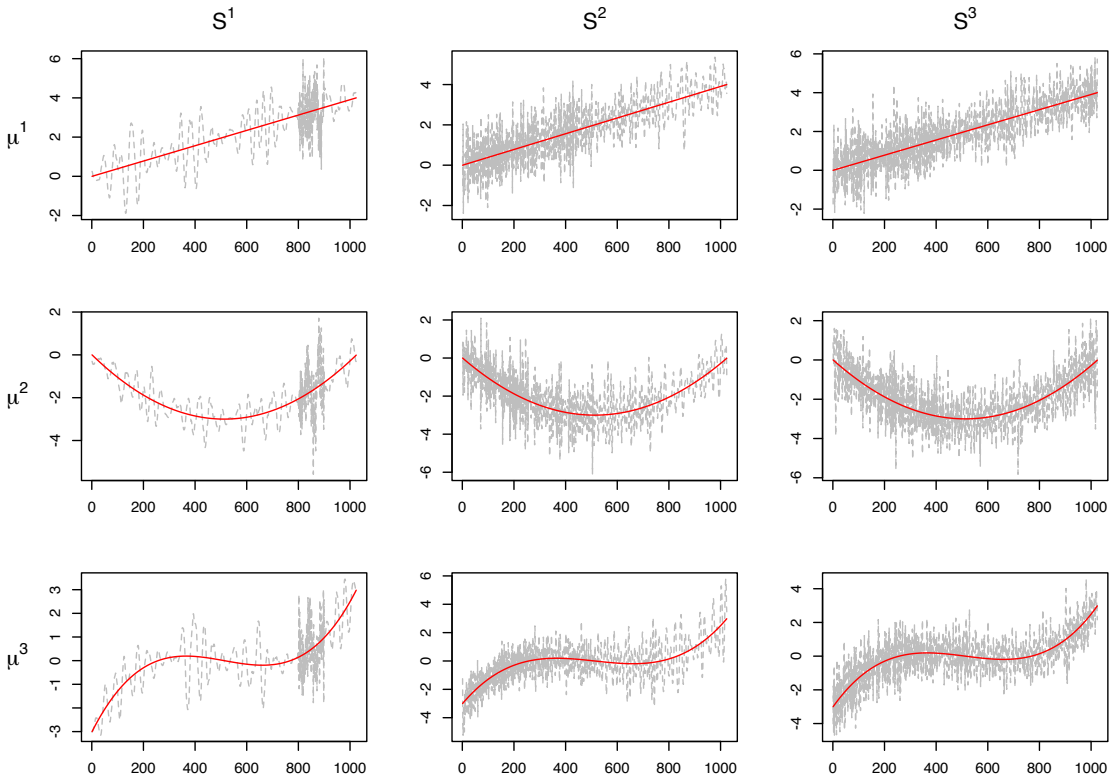

Figure 4.5.2: Example realisations from each trend and spectrum scenario. Dashed line shows time series with true underlying trend shown in solid line. Left: spectrum S^1, sinusoid with "burst" . Centre: S^2, concatenated moving average process. Right: S^3, slowly-evolving fine-scale power.

Trend	Spectrum 1	Spectrum 2	Spectrum 3
None	3.13	4.88	1.87
Linear	3.13	5.50	1.79
Quadratic	3.13	5.50	1.79
Cubic	3.13	5.50	1.79

Table 4.5.1: Mean squared error (in units $\times 10^{-3}$) of the averaged spectral estimate over 100 realisations.

In Figure 4.5.3 we see a comparison across the four trend scenarios for the averaged spectral estimate of spectrum S^1. From the figure we can see that the plots are nearly identical by eye, and no bias is incurred when estimating the spectrum in the presence of a polynomial trend.

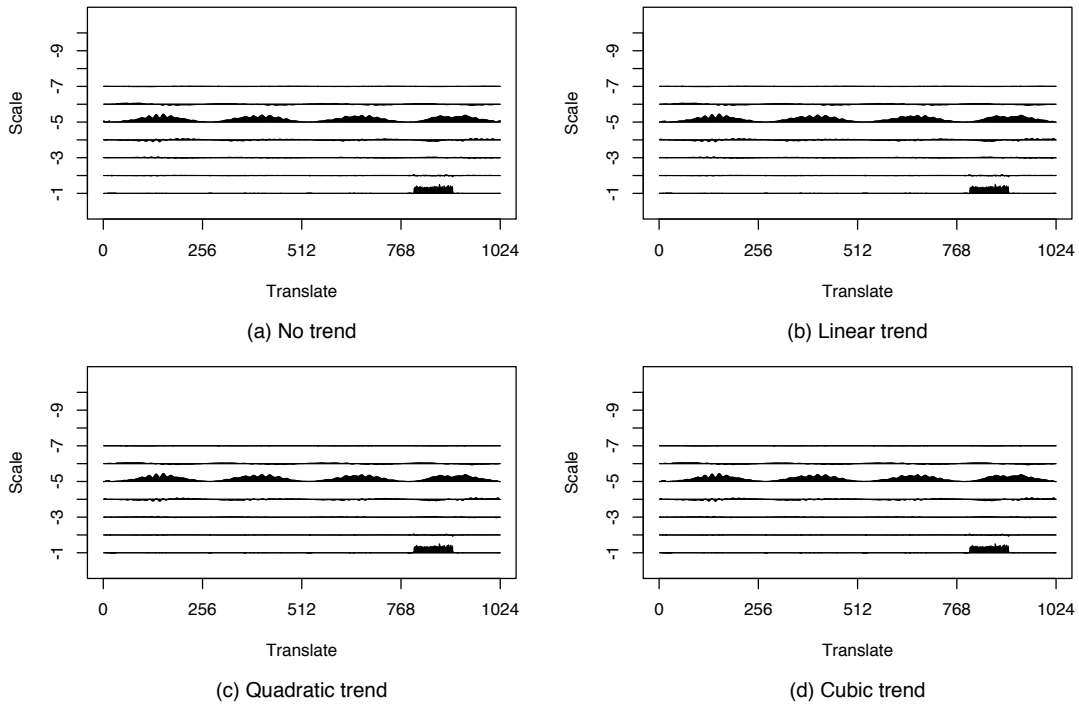

Figure 4.5.3: Averaged EWS estimate comparisons using realisations from the spectrum S^1.

In Table 4.5.2, we report the average mean squared error for the trend estimates, calculated over 100 realisations, with standard deviation given in brackets. Bold values in the table indicate the lowest mean squared error for each trend and spectrum combination. From the table, we see that our method performs well and gives a low mean squared error across the various scenarios. The method consistently outperforms the competing methods, with only one scenario in which it does not give the lowest mean squared error. In particular, we vastly outperform the wavelet-based method of von Sachs and MacGibbon (2000), which is in part due to the lack of proper boundary handling. As we might expect, the local polynomial method also performs well. The spline-based method performs well for spectrum 3, which is likely due to the fact this process contains only weak autocorrelation.

Trend	Spectrum	Method			
		LSWT	VSMWT	Spline	Loc. Poly.
Linear	1	$\mathbf{7.10 \times 10^{-4}}$ (4.66×10^{-4})	0.273 (0.094)	0.438 (0.115)	1.59×10^{-3} (1.75×10^{-3})
	2	$\mathbf{5.32 \times 10^{-4}}$ (7.71×10^{-4})	0.226 (0.064)	0.154 (0.087)	7.32×10^{-4} (6.45×10^{-4})
	3	1.01×10^{-4} (1.31×10^{-4})	0.161 (0.041)	$\mathbf{2.26 \times 10^{-5}}$ (2.99×10^{-5})	4.63×10^{-4} (3.37×10^{-4})
Quadratic	1	$\mathbf{7.05 \times 10^{-4}}$ (4.50×10^{-4})	0.284 (0.101)	0.453 (0.118)	1.51×10^{-3} (1.48×10^{-3})
	2	$\mathbf{4.61 \times 10^{-4}}$ (6.91×10^{-4})	0.208 (0.061)	0.154 (0.086)	6.79×10^{-4} (7.77×10^{-4})
	3	$\mathbf{9.69 \times 10^{-5}}$ (1.13×10^{-4})	0.157 (0.039)	1.55×10^{-3} (8.81×10^{-5})	4.87×10^{-4} (4.15×10^{-4})
Cubic	1	$\mathbf{7.18 \times 10^{-4}}$ (4.49×10^{-4})	0.285 (0.098)	0.466 (0.122)	9.07×10^{-3} (3.14×10^{-3})
	2	$\mathbf{6.77 \times 10^{-4}}$ (8.41×10^{-4})	0.228 (0.070)	0.161 (0.092)	6.79×10^{-3} (1.52×10^{-3})
	3	$\mathbf{1.07 \times 10^{-4}}$ (1.41×10^{-4})	0.177 (0.044)	2.57×10^{-3} (8.17×10^{-4})	6.85×10^{-3} (1.25×10^{-3})

Table 4.5.2: Average mean squared error and standard deviation of trend estimate over 100 realisations.

We next plot example trend estimates for each of the three trend scenarios for processes generated with spectrum S^1, shown in Figure 4.5.4. The estimate is given by the solid line, with the truth shown in dotted line. A 95% pointwise confidence interval is shown in dotted lines. The boundary handling method from Section 4 was used for these estimates, which can result in a slightly biased estimate at the endpoints. As previously discussed, confidence intervals near the boundary are not strictly valid due to the boundary handling, and will be larger. However, we believe this to be an acceptable trade-off as the boundary handling greatly improves the trend estimate itself near the boundaries. This is reflected in the consistently low mean squared error observed in Table 4.5.2.

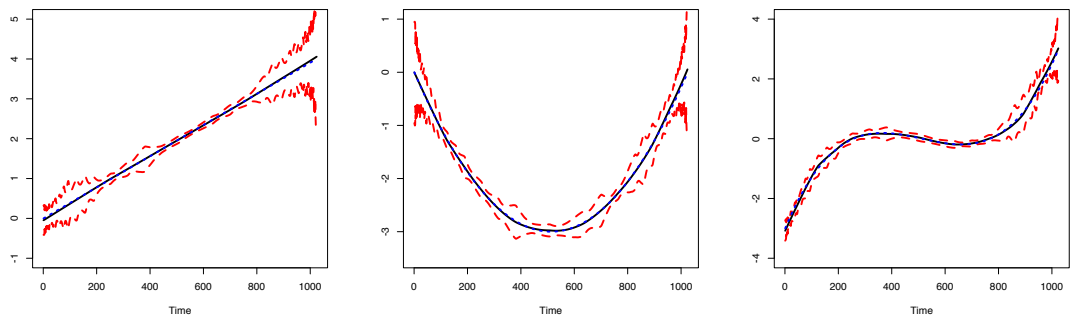

Figure 4.5.4: Example trend estimates (solid line) with 95% pointwise confidence intervals (dashed lines) for the three polynomial trend scenarios (dotted line), using realisations from spectrum S^1.

4.5.2 Low-Order Polynomial Trend, Low-Order Generating Wavelet

Here we consider a similar scenario to that in Section 4.5.1, however we use Haar wavelets to simulate the LSW processes rather than EP4 wavelets. In this case, using a Haar wavelet to analyse the series will cause bias as the Haar wavelet only has 1 vanishing moment and cannot zero out non-constant polynomial trends. Instead, we use the EP4 wavelet to analyse the series, and correct the estimates using the C-inverse matrix, motivated by the discussion in Section 4.3.2. In the no trend case, the estimate is calculated in the standard way with the Haar wavelet using the `ewspec3` function in `locits`.

We see from the results that we can still obtain an unbiased estimate of the spectrum in this scenario. Figure 4.5.5 shows the estimation comparisons for spectrum S^3 across the trend scenarios, while Table 4.5.3 reports the mean squared error for the averaged spectral estimates across the polynomial trend and spectrum scenarios. For the first spectrum S^1, the no trend estimate does have a lower mean squared error. However, for the other two spectra, the mean squared error is significantly lower when a trend is included. This is due to the ability of a smoother wavelet to provide a better estimate, which was discussed in Section 4.3.2. Further simulation results demonstrating this are provided in Section 4.5.4.

Trend	Spectrum 1	Spectrum 2	Spectrum 3
None	9.45	17.05	6.40
Linear	10.41	13.11	4.11
Quadratic	10.41	13.10	4.11
Cubic	10.41	13.11	4.11

Table 4.5.3: Mean squared error (in units $\times 10^{-3}$) of the averaged spectral estimate over 100 realisations.

In Table 4.5.4, we report the average mean squared error for the trend estimates, calculated over 100 trend estimates, with standard deviation given in brackets. Bold values indicate the lowest recorder mean squared error for a particular trend and spectrum combination. We see that our proposed trend estimator

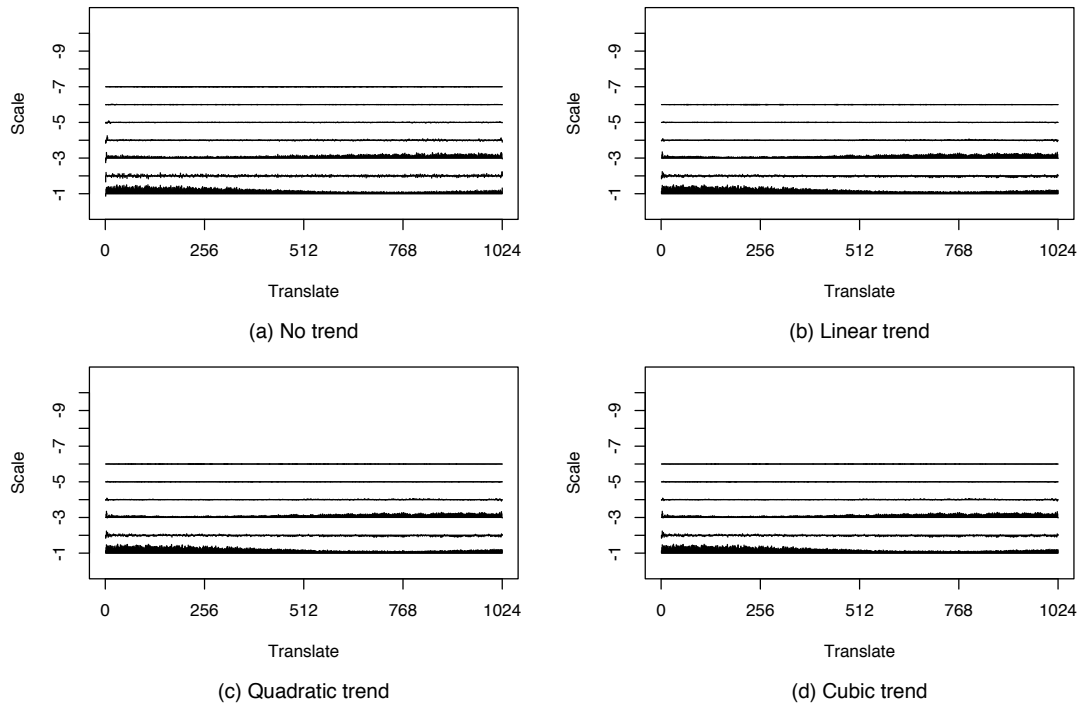

Figure 4.5.5: Averaged EWS estimate comparisons using realisations from the spectrum S^3.

performance is not significantly affected by the wavelet used to generate the LSW process, with similar results to those in Table 4.5.2. Our method is generally the best or second-best performing, with the local polynomial method also performing well. Lastly, we plot example trend estimates for each of the three polynomial trend scenarios for LSW processes generated using spectrum S^3, shown in Figure 4.5.6. The estimate is given by the solid line, with the truth shown in dotted line. A 95% pointwise confidence interval is shown in dotted lines.

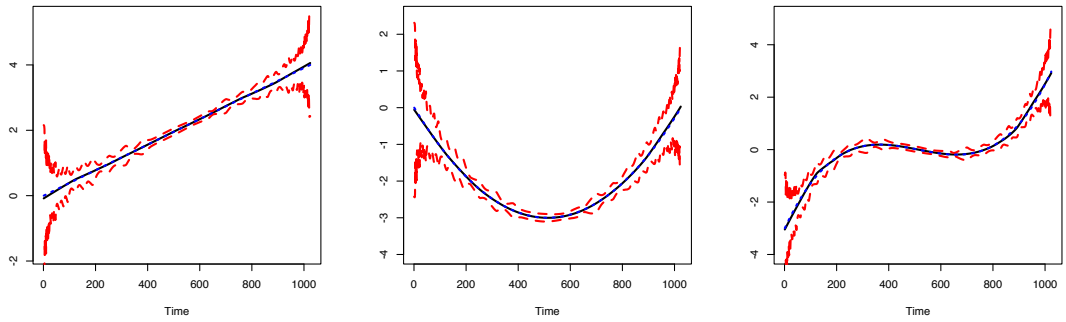

Figure 4.5.6: Example trend estimates (solid line) with 95% pointwise confidence intervals (dashed lines) for the three polynomial trend scenarios (dotted line), using realisations from spectrum S^3.

Trend	Spectrum	Method			
		LSWT	VSMWT	Spline	Loc. Poly.
Linear	1	2.56×10^{-3} (1.77×10^{-3})	0.286 (0.079)	0.383 (0.083)	$\mathbf{2.48 \times 10^{-3}}$ (1.83×10^{-3})
	2	1.67×10^{-3} (1.39×10^{-3})	0.220 (0.063)	0.116 (0.089)	$\mathbf{9.02 \times 10^{-4}}$ (9.34×10^{-4})
	3	2.96×10^{-4} (2.38×10^{-4})	0.166 (0.031)	$\mathbf{3.94 \times 10^{-5}}$ (5.22×10^{-5})	6.07×10^{-4} (4.86×10^{-4})
Quadratic	1	$\mathbf{2.26 \times 10^{-3}}$ (1.41×10^{-3})	0.287 (0.073)	0.384 (0.088)	2.63×10^{-3} (2.11×10^{-3})
	2	1.98×10^{-3} (1.48×10^{-3})	0.211 (0.056)	0.127 (0.095)	$\mathbf{1.05 \times 10^{-3}}$ (9.11×10^{-4})
	3	$\mathbf{3.13 \times 10^{-4}}$ (1.97×10^{-4})	0.158 (0.037)	1.58×10^{-3} (1.23×10^{-4})	5.36×10^{-4} (4.05×10^{-4})
Cubic	1	$\mathbf{3.20 \times 10^{-3}}$ (2.10×10^{-3})	0.269 (0.072)	0.383 (0.091)	9.43×10^{-3} (3.09×10^{-3})
	2	$\mathbf{1.87 \times 10^{-3}}$ (1.64×10^{-3})	0.228 (0.057)	0.118 (0.094)	7.22×10^{-3} (1.82×10^{-3})
	3	$\mathbf{3.32 \times 10^{-4}}$ (3.16×10^{-4})	0.177 (0.033)	2.62×10^{-3} (1.09×10^{-3})	6.83×10^{-3} (1.05×10^{-3})

Table 4.5.4: Average mean squared error and standard deviation of trend estimate over 100 realisations.

4.5.3 Non-Polynomial Trend, High-Order Analysing Wavelet

Lastly, we show simulation results when the trend is less smooth. In this simulation, the spectra are simulated using the Daubechies EP4 wavelet, and we add three non-polynomial trends to the series. The three trends are a sinusoidal, logistic, and exponential trend, defined as follows:

$$\mu_t^s = \sin(2\pi t) + \cos(\pi t), \qquad \mu_t^l = \frac{4}{1 + \exp(4 - 7\log 4t)}, \qquad \mu_t^e = 2\exp(3t). \tag{4.5.5}$$

The sinusoidal and logistic trends are Lipschitz continuous; however, the exponential trend is not, and further it is not Hölder continuous. We use a EP10 wavelet to compute the spectral estimate, which is the smoothest available wavelet in this family within `wavethresh` and will therefore be the best at removing the effect of the trend. In Figure 4.5.7, we see example realisations from each of the non-polynomial trend and spectrum scenarios.

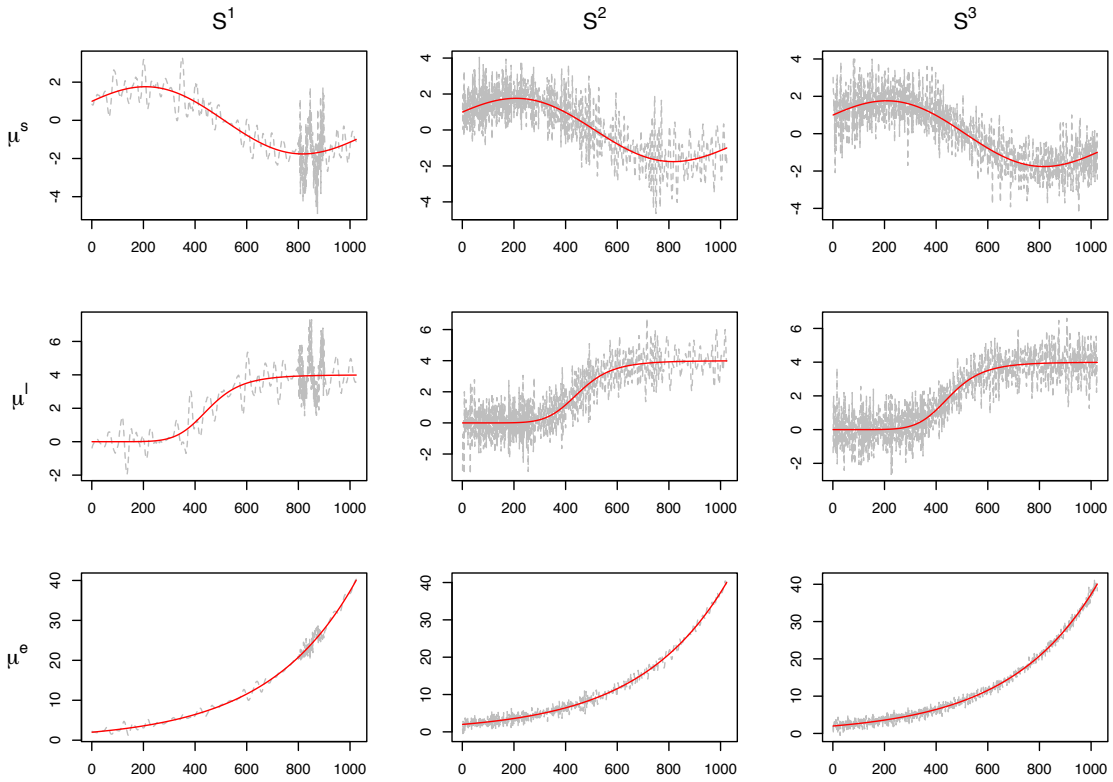

Figure 4.5.7: Example realisations of time series with the three non-polynomial trends defined in Equation (4.5.5) and three spectra shown in Figure 4.5.1.

Trend	Spectrum 1	Spectrum 2	Spectrum 3
None	2.60	5.00	1.80
Sine	2.93	4.34	1.53
Logistic	2.94	4.35	1.54
Exponential	2.94	4.35	1.54

Table 4.5.5: Mean squared error (in units $\times 10^{-3}$) of the averaged spectral estimate over 100 realisations, with trend scenarios in Equation (4.5.5).

In Table 4.5.5 we compare the mean squared error for the averaged spectral estimates, and find similar results to that in Section 4.5.1. The error when including a trend is higher for spectrum S^1, while lower for the other two spectra. These results confirm the finding in Section 4.3.3, that including a smooth non-polynomial trend will cause little bias to the estimate of the EWS. They also serve to highlight that using a smoother wavelet can achieve a more accurate spectral estimate. In Figure 4.5.8 we see a comparison across the four trend scenarios for the averaged

spectrum estimate of spectrum S^2. Again we see that the plots are similar, and the incorporation of a non-polynomial trend has negligible impact on the estimate of the spectrum.

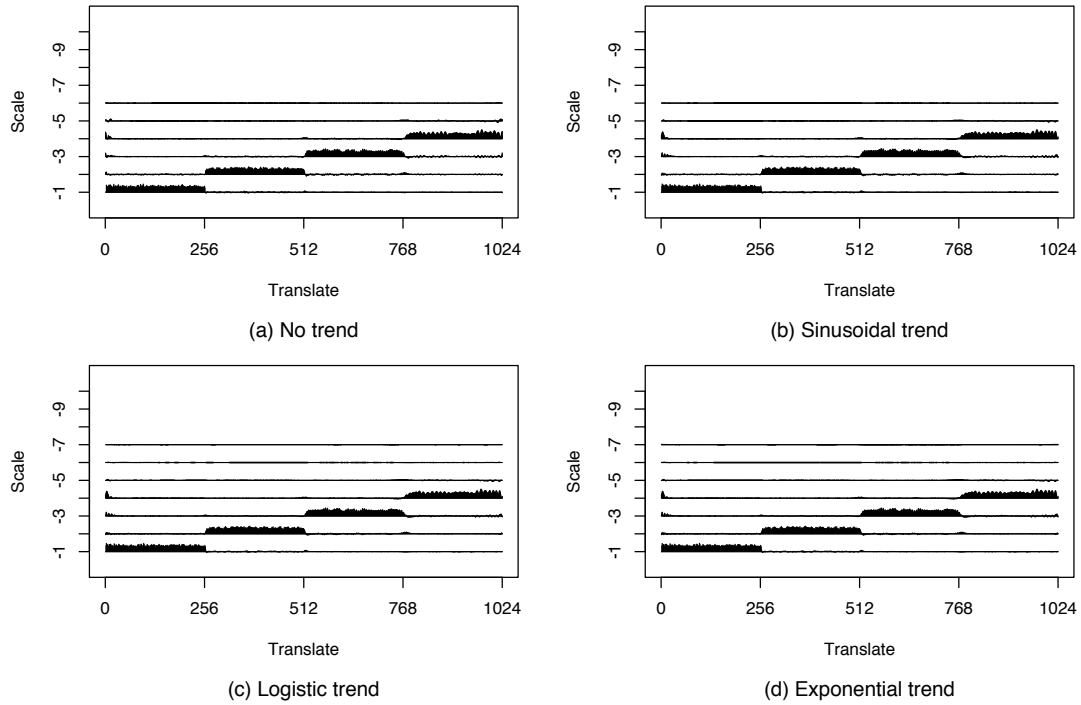

Figure 4.5.8: Averaged EWS estimate comparisons using realisations from the spectrum S^2.

Finally, we assess the trend estimation performance. In Table 4.5.6, we report the average mean squared error for the methods calculated over 100 realisations, with standard deviation given in brackets. Despite the fact that the trends are non-polynomial, the wavelet-based estimator still performs well. Our method is consistently the best performing across all of the trend and spectrum scenarios. Overall, the simulation results show that the linear wavelet thresholding estimator can perform well in a variety of trend settings in the presence of locally stationary errors. In particular, we observe that the linear wavelet thresholding estimator outperforms the non-linear thresholding method of von Sachs and MacGibbon (2000).

Trend	Spectrum	Method			
		LSWT	VSMWT	Spline	Loc. Poly.
Sine	1	$\mathbf{7.17 \times 10^{-4}}$ (4.93×10^{-4})	0.289 (0.108)	0.467 (0.127)	3.26×10^{-3} (2.49×10^{-3})
	2	$\mathbf{6.35 \times 10^{-4}}$ (8.61×10^{-4})	0.201 (0.064)	0.152 (0.097)	1.88×10^{-3} (9.37×10^{-4})
	3	$\mathbf{1.19 \times 10^{-4}}$ (1.57×10^{-4})	0.157 (0.0387)	1.49×10^{-3} (5.76×10^{-4})	1.65×10^{-3} (5.47×10^{-4})
Logistic	1	$\mathbf{2.63 \times 10^{-3}}$ (1.08×10^{-3})	0.284 (0.110)	0.451 (0.124)	1.58×10^{-2} (3.65×10^{-3})
	2	$\mathbf{2.52 \times 10^{-3}}$ (1.73×10^{-3})	0.214 (0.061)	0.148 (0.085)	1.41×10^{-2} (1.37×10^{-3})
	3	$\mathbf{2.15 \times 10^{-3}}$ (7.56×10^{-4})	0.158 (0.039)	2.32×10^{-3} (1.41×10^{-4})	1.38×10^{-2} (7.94×10^{-4})
Exponential	1	$\mathbf{9.45 \times 10^{-4}}$ (4.91×10^{-4})	0.386 (0.088)	0.446 (0.098)	2.99×10^{-2} (6.36×10^{-3})
	2	$\mathbf{8.61 \times 10^{-4}}$ (9.94×10^{-4})	0.347 (0.072)	0.154 (0.084)	2.27×10^{-2} (3.62×10^{-3})
	3	$\mathbf{2.54 \times 10^{-4}}$ (1.54×10^{-4})	0.294 (0.040)	3.61×10^{-3} (8.57×10^{-4})	2.29×10^{-2} (1.85×10^{-3})

Table 4.5.6: Average mean squared error and standard deviation of trend estimate over 100 realisations.

In Figure 4.5.9, we see example trend estimates for each of the three non-polynomial trend scenarios for processes generated with spectrum S^2. The estimate is given by the solid line, with the truth shown in dotted line. A 95% pointwise confidence interval is shown in dotted lines.

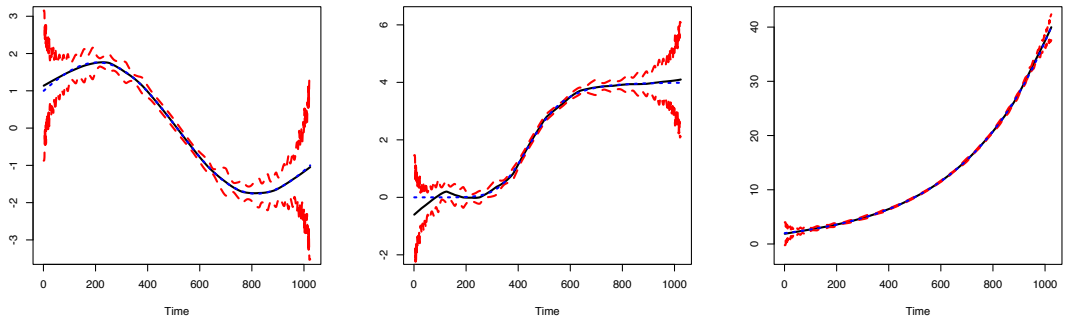

Figure 4.5.9: Example trend estimates (solid line) with 95% pointwise confidence intervals (dashed lines) for the three non-polynomial trend scenarios (dotted line), using realisations from spectrum S^2.

4.5.4 Using Smoother Wavelets for Improved EWS Estimation

Numerical results show the improved performance of using a smoother wavelet in the estimation procedure. In particular, they highlight the ability of the smoother wavelet to reduce the mean squared error of the spectral estimate. We simulated 100 realisations of Haar wavelet LSW processes of length $2^{10} = 1024$, using the

Spectrum	Haar	LA10	LP250
S^1, burst	0.85	0.32	**0.19**
S^2, moving average	1.18	0.54	**0.53**
S^3, slowly-evolving	0.55	**0.26**	0.27

Table 4.5.7: Average mean squared error comparisons between Haar, LA10, and LP250 wavelets, across the three defined spectra.

three spectra defined previously, without adding any trend term. We compared the average mean squared error of the un-smoothed spectral estimates in the case when a Haar wavelet, a Daubechies Least Asymmetric wavelet with ten vanishing moments (LA10), and the Littlewood-Paley wavelet with filter number 250 (LP250) were used to analyse the series. The raw wavelet periodograms were corrected using the standard A-inverse matrix for the Haar wavelet, and the appropriate C-inverse matrices for the LA10 and LP250 wavelets. The Littlewood-Paley wavelet can be thought of as an approximation to the Shannon wavelet; a larger filter number gives a closer approximation. The Haar wavelet is not smooth, while the Littlewood-Paley wavelet is the smoothest of the three.

The results of the simulation are shown in Table 4.5.7. The Littlewood-Paley estimate is the best for the first two spectra, while the LA10 wavelet is the best for the third spectra. The Haar wavelet struggles due to the leakage of power to nearby scales, which is characterised by the slow decay in the off-diagonal entries of the Haar A matrix. Performance is particularly poor for spectrum S^2, since this spectrum contains power transferring from fine to coarser scales so one can expect more leakage to occur during estimation. Although the Littlewood-Paley wavelet performs best overall, we recommend using a smooth wavelet within the Daubechies extremal phase or least asymmetric family. This is because the length of the Littlewood-Paley (or Shannon) wavelet filter is large and hence causes more pronounced boundary effects than that of wavelet filters with shorter length. The LA10 wavelet, for example, provides a good compromise between smoothness, which lowers leakage and removes trend, and filter length, which causes boundary effects.

However, we note that there will still be some spectral structures which are better estimated using a less smooth analysing wavelet. Intuitively, the Haar wavelet is piecewise constant and so should be better at estimating spectra whose behaviour is piecewise constant, whereas a smoother wavelet may struggle. We may still have higher leakage into nearby scales, but the estimate should perform better when the spectral structure has many jumps only at one scale. In general, when the wavelet used to generate the process is not smooth, the simulation results show that it can be better to use a smoother wavelet in the spectral estimation in order to reduce leakage and obtain a more accurate estimate.

4.6 Global Mean Sea Temperature Data Application

We now highlight the ability to model time series with time-varying first and second-order structure on the monthly global mean sea temperature dataset (GMST) version 4.6 from the Goddard Institute for Space Studies (GISS) Surface Temperature Analysis (GISTEMP). The GISTEMP dataset combines land and sea surface temperatures from the Global Historical Climatology Network, version 3.3.0 (GHCNv3.3.0) and the Extended Reconstructed Sea Surface Temperature, version 4 (ERSST.v4; Huang et al. (2015); Liu et al. (2015)). It also includes the Scientific Committee on Antarctic Research (SCAR) stations over Antarctica (Hansen et al. (2006). The data are available at `http://data.giss.nasa.gov/gistemp`, and are shown in Figure 4.6.1. The GMST time series exhibits periods of warming separated by a long pause from approximately the mid-1940s to the mid-1970s (Kellogg, 1993) and possibly a second, shorter one, since the late 1990s/early 2000s, although this is highly debated (Karl et al., 2015; Trenberth, 2015; Beaulieu and Killick, 2018). The series should be well-suited to our methodology, as it appears to contain both long-term trends and short-term variability that may be nonstationary.

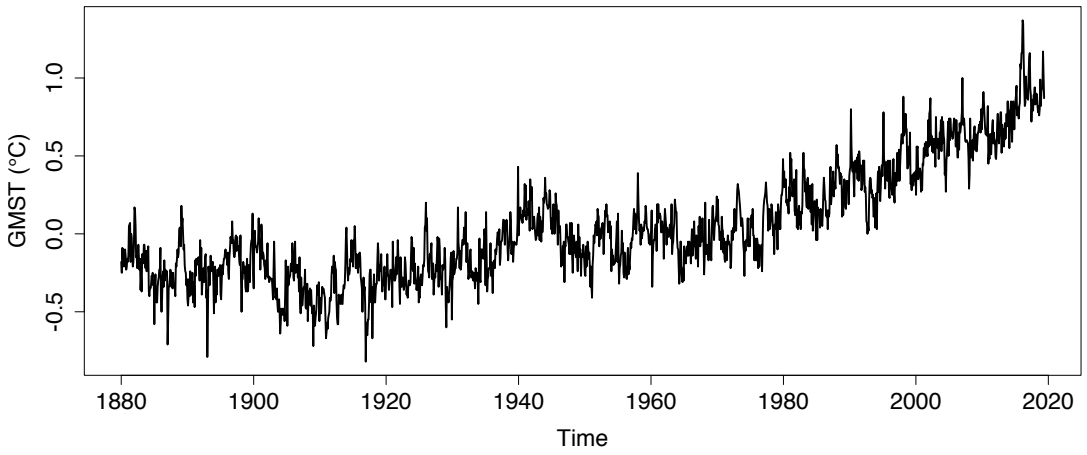

Figure 4.6.1: Global average monthly sea temperature data.

In general the series exhibits an upward trend therefore using a high-order wavelet should remove the trend from the wavelet coefficients, and allow us to obtain an estimate for the spectrum. In addition, we might reasonably expect there to be strong dependency within the data: as with many environmental time series this may be a result of certain physical phenomena.

In Figure 4.6.2, we see the spectral estimate for the GMST time series. This is estimated using the Daubechies LA10 wavelet, while the raw wavelet periodogram is smoothed using a running mean smoother with bin width 120, corresponding to 10 years. Each level is scaled individually for clarity. Additionally, we have used the boundary handling procedure outlined in Section 4.4.4 to properly handle the endpoints of the series. Motivated by the discussion in Section 4.3.3, we only calculate the EWS estimate across the finest 7 scales, which corresponds to 70% of the scales. We see that the spectral behaviour is generally slowly-evolving with peaks at various locations: for example note the burst of power that occurs in scale -4 around 1910. We also see an appearance of power in scale -7 at approximately 1940 and scale -6 slightly later, corresponding to the proposed period of pause in warming.

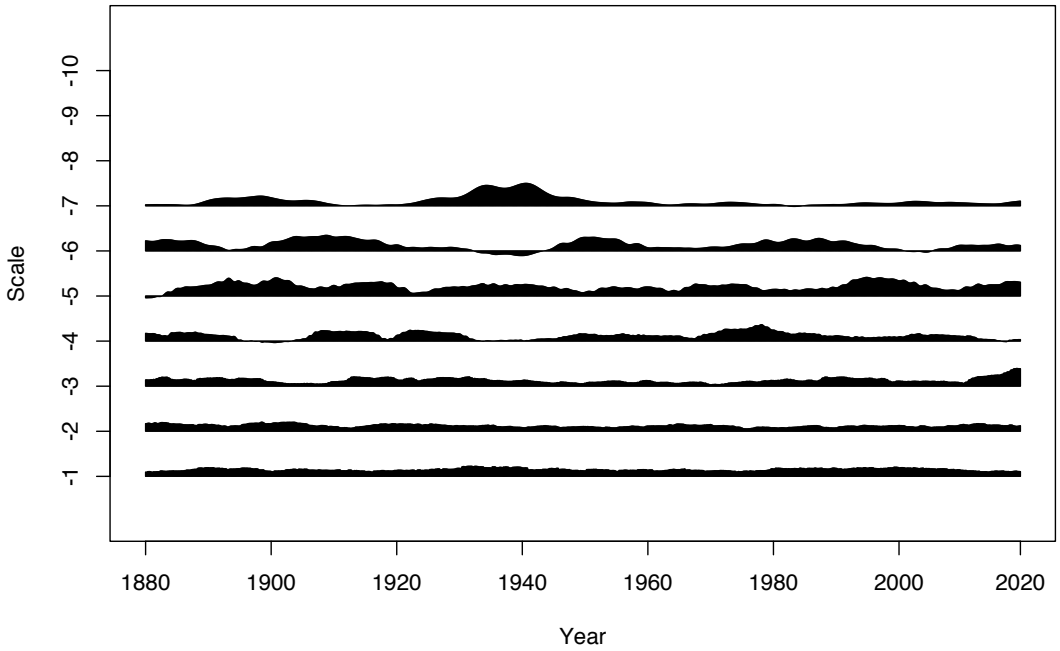

Figure 4.6.2: EWS estimate of the GMST series. Each level is scaled individually for clarity.

In Figure 4.6.3, we see the estimate for the trend shown in the solid line, with the 95% pointwise confidence interval shown in the dashed lines. This was calculated with the Daubechies EP10 wavelet, using the methodology described in Section 4.4, using the boundary extension from Section 4.4.4. Note that the estimate near the endpoints is therefore dependent on this boundary handling and inference should be carefully considered. Overall, our estimate agrees with many previous analyses of the series: we see a long pause between approximately the 1940s and 1970s, as well as a sustained period of warming post-1970 to the early 2000s.

Lastly, we investigate the autocorrelation structure of the GMST time series. A common assertion in the Climate literature is that the autocorrelation within the Global Mean Sea Temperature Data is constant across time. Here we use the estimate of the spectrum to look at the autocorrelation function at different periods in the series.

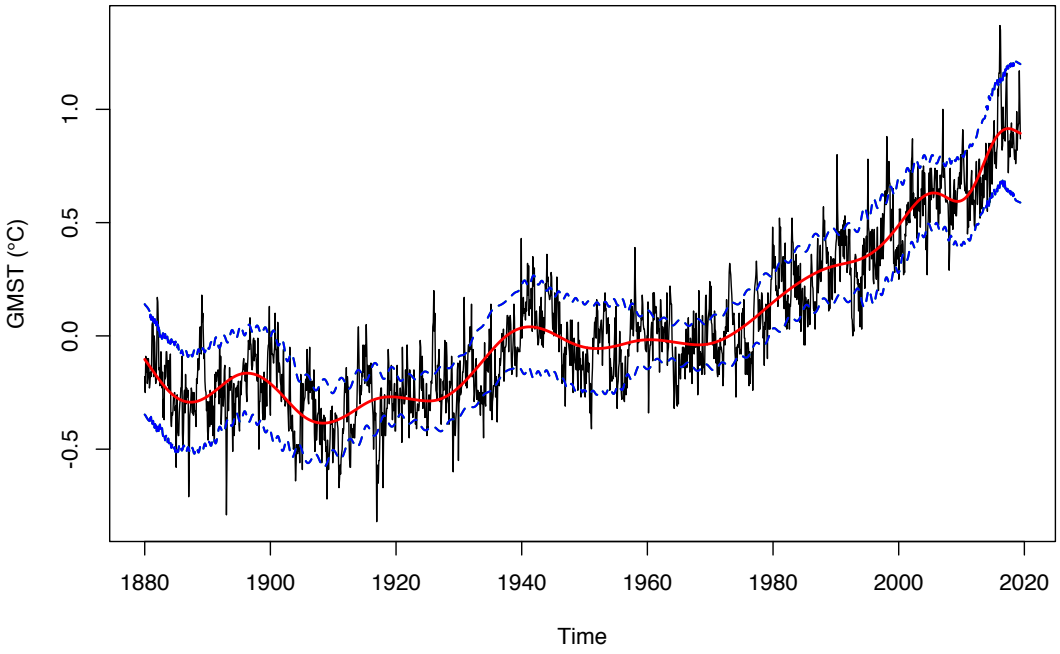

Figure 4.6.3: Global average monthly sea temperature with trend estimate (solid line) and 95% pointwise confidence intervals (dashed lines).

In Figure 4.6.4, we see a plot of the autocorrelation function of the time series at 6 different time-points across 25-year intervals, starting at 1890 and ending at 2015. Note that the estimate of the autocorrelation is dependent on the choice of wavelet. However, we have found empirically that using other wavelets within the Daubechies LA and EP family that possess several vanishing moments yield almost identical results. Commonly, the error structure of climate change data is modelled using a low-order AR process, for example Beaulieu and Killick (2018), Karl et al. (2000) and Rodionov (2004). In similar analyses on gridded datasets, Lenton et al. (2017) find an increase in variance and autocorrelation in global surface sea temperature data, while Boulton and Lenton (2015) find a "reddening" - a gradual increase in autocorrelation - in North Pacific surface sea temperature data.

Our analysis suggests that overall, the low-order AR assumption may be appropriate at certain time points, however the degree of autocorrelation changes over time, and the assumption of a stationary AR model may not be appropriate. From the plot, the nonstationary nature of the second-order structure can be seen.

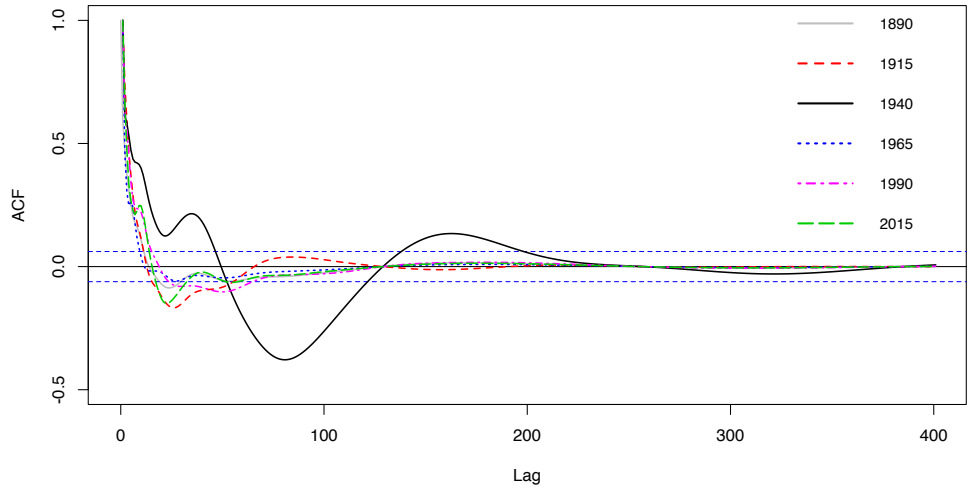

Figure 4.6.4: Autocorrelation of the GMST series across 6 time-points, up to lag 400.

In 1890, 1915, 1965, 1990 and 2015, the autocorrelation exhibits a similar decay behaviour, falling within the upper and lower confidence lines after approximately lag 60, corresponding to a time period of 5 years. However at time 1940, given by the dark solid line, we see much stronger autocorrelation, exhibiting a shape that is not well represented by a low-order AR model. This increase in autocorrelation can be attributed to the appearance of power at the coarsest scale of the spectrum, as seen in Figure 4.6.2. Moreover, our analysis shows that in the period between 2005 – 2020 – which is debated to be another warming pause – the variance of the series increases by roughly 63% from 0.0153 to 0.0240. This may explain the appearance of a pause due to increased variability, as many methods assume the series to be second-order stationary. Note that the trend estimate does suggest the appearance of a pause in this time period, however since this occurs at the boundary the estimate should be used with caution.

4.7 Concluding Remarks

In this chapter we have addressed the commonly encountered problem of modelling time series that display first and second-order nonstationarity. Our model extends the locally stationary wavelet model of Nason et al. (2000), used for modelling

second-order nonstationarity, to include a nonstationary, smooth mean structure. Within our framework we can model both smooth deterministic trends and slowly-evolving second-order behaviour, allowing for a flexible model that can be applicable to time series in many applications.

We have established results concerning the theoretical properties of the estimators of the first and second-order quantities of interest, which are shown to have good consistency properties and strong practical performance. Furthermore, we have found using a smoother wavelet to be useful even in the case of a zero-mean time series. In order to ensure the applicability of our methodology, we propose a new method for handling the boundary of a time series when computing the discrete non-decimated wavelet transform, which incorporates data of any length and is suitable for data that display trend behaviour. In our framework, the results and methodology are suitable for polynomial and smooth trends; however, a polynomial or smooth function is a good approximation to many trends of practical interest. In practice, the techniques described here will work well when we depart from the assumptions of the model, as evidenced within the simulation study.

We demonstrated our methodology on a data example that shows that our model can be successfully applied to a variety of data sets that exhibit time-varying first and second-order structure. In our analysis of the global mean sea temperature dataset, we find that the second-order structure of the series varies over time, while observing similar findings as previously described of a general increasing trend of global mean sea temperature.

A number of interesting research questions arise from our work, such as how to incorporate other LSW-based work within this framework, for example forecasting as in Fryzlewicz et al. (2003) or changepoint analysis as in Killick et al. (2013). Another potential avenue is to incorporate the more recent work of Cardinali and Nason (2017) on locally stationary packet processes into the methodology. We leave these as areas for future work.

Chapter 5

Detecting Changes in Mean in the Presence of Time-Varying Autocovariance

5.1 Introduction

In this chapter, we consider the problem of detecting changes in the mean of a univariate time series that exhibits a non-trivial, unknown, time-varying autocovariance structure. The study of changepoint problems dates back to Page (1954, 1955), and continues to be a highly active area of research, with applications including genetics (Hocking et al., 2013), cyber security (Adams and Heard, 2016), and climatology (Carr et al., 2017). A survey of changepoint methods can be found in Tartakovsky et al. (2014). State-of-the-art methods for identifying multiple changepoints in univariate data include Pruned Exact Linear Time method (PELT) (Killick et al. (2012)), Wild Binary Segmentation (WBS) (Fryzlewicz (2014)), and Simultaneous Multiscale Changepoint Estimator (SMUCE) (Frick et al. (2014)). In the case of detecting changes in mean, most state-of-the-art methods focus their attention on the case where the error structure is IID with known variance.

There is comparatively little attention given to the problem of detecting changes in mean while the second-order structure is unknown, non-trivial and must be estimated. Further still, there is even less attention given to the case where this second-order structure is allowed to vary over time. One approach for dealing with autocorrelation when testing for changes in mean is to increase the threshold above which changes are detected, see Lavielle (1999) for example. However, it is difficult to systematically choose the threshold without performing some sort of pre-estimation of the autocovariance, which is highly challenging in the presence of mean changes (Lund and Shi, 2020). If one simply raises the threshold, then there will be a trade-off between a decreased false positive rate and a decreased true positive rate. Further, using standard autocovariance estimation techniques will result in bias, and subsequent use of these biased estimates can significantly reduce the efficiency of changepoint estimators, as described in Section 3 of Jandhyala et al. (2013).

One can estimate the autocovariance structure and incorporate it within the changepoint detection algorithm: for example, Dette et al. (2018) consider modifications to the SMUCE algorithm for dependent, not necessarily Gaussian data. Tecuapetla-Gómez and Munk (2017) estimate the autocovariance structure using a difference-based approach under the assumption of m-dependent errors, which is then used in order to estimate a piecewise constant signal. Chakar et al. (2017) studies the problem of estimating mean changes with an underlying AR(1) process. Lavielle (1999) considers dependent processes and shows that the least-squares estimators of the changepoint locations and of the parameters of each segment are consistent under mild conditions, however the number of changepoints is assumed to be known. The narrowest-over-threshold (NOT) approach of Baranowski et al. (2019) is shown to be consistent in the presence of short memory dependence, however the autocorrelation is assumed known. Kokoszka and Leipus (1998) use a cumulative sum approach to detect changes in mean with dependent errors, while Bücher and Kojadinovic (2016) use a nonparametric bootstrap approach to detect

changepoints in structures including the mean. All of the above approaches make the assumption of second-order stationarity.

In practice, the autocovariance structure of a time series may be time-varying, and is not known a priori. Failure to take into account this autocovariance can lead to serious errors in the changepoint detection procedure. In particular, the presence of autocorrelation may lead to overfitting of mean changes. Our approach enables the detection of changes in mean while allowing for unknown, possibly time-varying second-order structure, by modelling the error process of the time series as a locally stationary wavelet (LSW) process (Nason et al., 2000). The LSW model enables the modelling of a zero-mean time series with slowly evolving, time-varying second-order structure: an overview of the necessary concepts can be found in Section 5.2.2. The LSW model is a flexible framework that can incorporate a wide variety of error processes – both stationary and nonstationary – for example ARMA and GARCH models.

In this chapter, we describe a likelihood-based test, in which the test statistic incorporates the unknown and potentially time-varying autocovariance of the time series. In order to utilise the theory of LSW processes, which is designed for zero-mean processes, we adapt the estimation procedure of Nason et al. (2000) to our piecewise constant mean structure. This is used in a standard likelihood ratio framework to test for changes in mean in the observed time series. By including the autocovariance information, we can both lower the false discovery rate by accurately distinguishing between mean changes and patterns due to autocorrelation, and increase the accuracy of true change detection. Furthermore, we demonstrate that allowing for the possibility of non-trivial autocovariance does not significantly lower the power of the likelihood ratio test when the underlying noise in the time series is in fact independent. This approach provides practitioners with a simple yet flexible alternative to assuming an IID error structure without incurring significant loss of power.

The remainder of the chapter is organised as follows. Section 5.2 recalls the like-

lihood ratio test for detecting single changes in mean and introduces the paradigm for our methodology. Section 5.3 describes the method for estimating the autocovariance of the time series, and discusses our final likelihood-based test for mean changepoint detection. The performance of the method is tested and compared via a simulation study in Section 5.4, and applied to two data examples in Section 5.5. Concluding remarks are given in Section 5.6.

5.2 Background

In this section we give the necessary background on changepoint detection and locally stationary wavelet process modelling, which will form the basis of our approach.

5.2.1 Detecting a Single Change in Mean

We aim to detect a finite number of changes in mean within a time series of length T. We begin by considering the problem of detection a single change in mean, using a likelihood-based framework. The likelihood-based approach is one of the most common and well studied methods used within the changepoint literature: for a thorough review see for example Chen and Gupta (2011). Assume that the time series is given by

$$X_t = \mu_t + \epsilon_t \tag{5.2.1}$$

for $t = 0, \ldots, T-1$. The mean function is given by μ_t and the error terms ϵ_t are Gaussian, zero-mean random variables with variance-autocovariance matrix Σ. The hypothesis to test for a single change in mean at time p is

$$H_0 : \mu_0 = \mu_1 = \cdots = \mu_{T-1}$$

$$H_1 : \mu_0 = \cdots = \mu_p \neq \mu_{p+1} = \cdots = \mu_{T-1}.$$

Using this formulation we can derive the likelihood ratio test for a single change in mean. Let $\boldsymbol{X} = (X_0, \ldots, X_{T-1})^\mathsf{T}$ be the observed time series. Under the assumption that the ϵ_t are Gaussian, the log-likelihood of the time series under the null hypothesis is given by

$$l(\boldsymbol{\mu}, \Sigma | \boldsymbol{X}) = \frac{T}{2} \log 2\pi - \frac{1}{2} \log |\Sigma| - \frac{1}{2} (\boldsymbol{X} - \boldsymbol{\mu}_0)^\mathsf{T} \Sigma^{-1} (\boldsymbol{X} - \boldsymbol{\mu}_0),$$

where the superscript T denotes the transpose, $\boldsymbol{\mu}_0 = (\mu_0, \ldots, \mu_0)^\mathsf{T}$ is the (constant) mean vector of the series and Σ is the variance-autocovariance matrix, with entries given by $\Sigma_{ts} = \mathrm{Cov}(X_t, X_s)$. The log-likelihood under the alternative hypothesis is given by

$$l(\boldsymbol{\mu}, \Sigma, p | \boldsymbol{X}) = \frac{T}{2} \log 2\pi - \frac{1}{2} \log |\Sigma| - \frac{1}{2} (\mathbf{X} - \boldsymbol{\mu}_1)^\mathsf{T} \Sigma^{-1} (\mathbf{X} - \boldsymbol{\mu}_1),$$

where $\boldsymbol{\mu}_1 = (\mu_0, \ldots, \mu_0, \mu_{T-1}, \ldots, \mu_{T-1})^\mathsf{T}$ is the mean vector with change at time p. Note that Σ is the same under both hypotheses. Hence the likelihood ratio test statistic for a single change in mean, after plugging in estimators for the unknown $\boldsymbol{\mu}_0$, $\boldsymbol{\mu}_1$ and Σ, is given by

$$\lambda \propto \max_{1<p<T-2} \left\{ (\boldsymbol{X} - \hat{\boldsymbol{\mu}}_0)^\mathsf{T} \hat{\Sigma}^{-1} (\boldsymbol{X} - \hat{\boldsymbol{\mu}}_0) - (\boldsymbol{X} - \hat{\boldsymbol{\mu}}_1)^\mathsf{T} \hat{\Sigma}^{-1} (\boldsymbol{X} - \hat{\boldsymbol{\mu}}_1) \right\}. \quad (5.2.2)$$

Using (5.2.2), a changepoint is detected when $\lambda > c$, where c is a constant. If a change is detected, then the location of the changepoint, $\hat{p}$, is estimated as the value of p that maximises λ. Note that a similar test statistic was applied to the problem of detecting changes in a piecewise constant autocovariance function, in Killick et al. (2013). Under the simplifying assumption of IID Gaussian errors with known variance, this reduces to the standard likelihood test for a change in mean, equivalent to the cumulative sum (CUSUM) test statistic. The appropriate threshold c can be chosen using the asymptotic distribution of (5.2.2).

In our more general scenario we must estimate the unknown autocovariance

matrix Σ. Under the null hypothesis, this is straightforward; however under the alternative hypothesis this becomes highly challenging as any standard estimation procedure will be biased due to the presence of the mean change. This is why several authors have assumed a known (autoco)variance, as once we have a sensible estimate we can then calculate λ and test for a change in mean in the time series in the standard way.

5.2.2 Modelling Series with Time-Varying Autocovariance

As discussed in Section 5.2.1, in order to use the likelihood-based framework we must estimate the unknown autocovariance of the time series. To this end, we represent the ϵ_t in Equation (5.2.1), the error component of the time series, as a locally stationary wavelet (LSW) process. The LSW model was introduced in Nason et al. (2000) and provides methodology for modelling time series whose second-order structure evolves slowly over time. Wavelets are useful in estimating time varying quantities as they are compactly-supported oscillatory functions that can be translated and dilated efficiently to provide location-scale decompositions. For an overview of wavelet techniques, see for example Nason (2008) or Vidakovic (2009).

In the remainder of this section, we give a brief overview of the LSW framework. Following Nason et al. (2000), a triangular stochastic array $\{\epsilon_{t,T}\}_{t=0}^{T-1}$ for $T = 1, 2, \ldots,$ is in a class of locally stationary wavelet (LSW) processes if there exists a mean-square representation

$$\epsilon_{t,T} = \sum_{j=-\infty}^{-1} \sum_{k} w_{j,k;T} \psi_{j,k-t} \xi_{j,k}, \tag{5.2.3}$$

where j and $k \in \mathbb{Z}$ are scale and location parameters respectively, and the $\xi_{j,k}$ are zero-mean, orthonormal identically distributed random variables. The $\{\psi_{j,k-t}\}_{j,k}$ are a set of discrete non-decimated wavelets, and the $\{w_{j,k;T}\}$ are a set of time-varying amplitudes. There exists, for each $j \leq -1$, a Lipschitz continuous function

$W_j(z)$ for $z \in (0,1)$ which satisfies $\sum_j |W_j(z)|^2 < \infty$ uniformly in $z \in (0,1)$. The Lipschitz constants L_j are uniformly bounded in j and $\sum_j 2^{-j} L_j < \infty$. There exists a sequence of constants C_j such that for each T

$$\sup_k |w_{j,k;T} - W_j(k/T)| \leq C_j T^{-1}, \tag{5.2.4}$$

where for each $j \leq -1$ the supremum is over $k = 0, \ldots, T-1$, and where the sequence $\{C_j\}$ satisfies $\sum_j C_j < \infty$.

We further assume that the increment sequence $\{\xi_{j,k}\}$ is Gaussian, which ensures that $\epsilon_{t,T}$ is Gaussian. As with classical time series theory, the second-order structure of an LSW process is encoded in the spectrum. The evolutionary wavelet spectrum (EWS) of an LSW process is defined as $S_j(z) = |W_j(z)|^2$ for rescaled time $z = k/T$ and measures the contribution to variance at a particular rescaled time z and scale j. Since the W_j are Lipschitz continuous, the spectrum at level j, S_j, is also Lipschitz continuous. However, Fryzlewicz and Nason (2006) and Van Bellegem and von Sachs (2008) extend the LSW model to consider piecewise constant spectra and those of bounded total variation respectively. No further assumptions are required on the dependence structure of the time series, beyond the requirements of the LSW model.

Note that LSW processes include the class of stationary processes with absolutely summable autocovariance $\sum_\tau |c_X(\tau)| < \infty$. Hence, this modelling framework can account for a large variety of classes of short memory stationary process, such as ARMA and GARCH models. Conversely, any LSW process with a spectrum independent of time is stationary, provided that the additional assumption $\sum_j 2^{-j} S_j(z) < \infty$ is satisfied (Nason et al., 2000).

In Nason et al. (2000), rigorous estimation theory for the EWS and time-varying autocovariance of a zero-mean series are outlined, which we will now briefly describe here. The EWS is estimated via the empirical wavelet coefficients of the time series, given by $d_{j,k;T} := \langle \epsilon_{t,T}, \psi_{j,k-t} \rangle = \sum_t \epsilon_{t,T} \psi_{j,k-t}$. As with Fourier approaches, the raw wavelet periodogram $I^j_{k;T} := |d_{j,k;T}|^2$ is a biased, inconsistent

estimator of the EWS (Nason et al. (2000), Proposition 4):

$$\mathbb{E}\left(I^j_{k;T}\right) = \sum_l A_{jl} S_l(k/T) + \mathcal{O}(T^{-1}), \tag{5.2.5}$$

$$\mathrm{Var}(I^j_{k;T}) = 2\left(\sum_l A_{jl} S_l(k/T)\right)^2 + \mathcal{O}(2^{-j}T^{-1}), \tag{5.2.6}$$

where the operator $A = (A_{jl})_{j,l<0}$ is given by $A_{jl} := \langle \Psi_j, \Psi_l \rangle = \sum_\tau \Psi_j(\tau)\Psi_l(\tau)$, and the autocorrelation wavelets are defined by $\Psi_j(\tau) := \sum_{k\in\mathbb{Z}} \psi_{j,k}\psi_{j,k-\tau},\ j < 0, \tau \in \mathbb{Z}$.

The vector of periodograms $\mathbf{I}(z) := \{I^l_{\lfloor zT \rfloor,T}\}_{l=-1,\ldots,-J},\ J = \lfloor \log_2(T) \rfloor$ is bias-corrected using the curtailed J-dimensional matrix $A_J := (A_{jl})_{j,l=-1,\ldots,-J}$ which is shown to be invertible in Nason et al. (2000). The raw wavelet periodogram is first smoothed and then corrected by A_J^{-1} to produce an asymptotically unbiased, consistent estimator. Smoothing can be carried out using a number of techniques, for example via a running mean as in Nason (2013) or using wavelet thresholding as in Nason et al. (2000).

The local autocovariance (LACV) function, $c(z, \tau)$, of an LSW process with EWS $\{S_j(z)\}$ is defined as $c(z, \tau) = \sum_{j=-\infty}^{-1} S_j(z)\Psi_j(\tau)$, for $\tau \in \mathbb{Z}, z \in (0, 1)$. The LACV can be thought of as a decomposition of the autocovariance of a process over scales and rescaled time locations. In practice, the local autocovariance is estimated by plugging in the smoothed, corrected estimate for the EWS into the LACV definition; this results in a consistent estimator of the LACV (Nason et al., 2000).

5.3 Method for Detecting Mean Changes in the Presence of Time-Varying Autocovariance

In this section, we discuss our proposed methodology for estimating the time-varying autocovariance in the presence of mean changes, and the likelihood-based test for mean changepoint detection.

5.3.1 Autocovariance Estimation in the Presence of Mean Changes

Having assumed that the error process of the time series is LSW, we now discuss the methodology for estimating the autocovariance of this process. In Section 5.2.2 we recalled the standard estimation procedure for estimating the EWS, and hence the autocovariance, however this method is only valid in the case where the time series has a constant mean. A key ingredient in our approach is providing methodology to estimate the nonstationary second-order structure of a time series under the relaxed assumption of a piecewise constant mean signal.

Under the alternative hypothesis, the overall model for X_t is not LSW due to changes in mean. Wavelet coefficients informally describe weighted localised changes in average, therefore changes in mean will cause the raw wavelet periodogram to become contaminated with information about the mean, rather than purely information about the spectrum. This will cause additional bias in the raw wavelet periodogram. If we were to use, for example, a running mean to smooth the wavelet periodogram – as is standard in LSW spectral estimation to achieve consistency – this would result in a biased estimator.

To illustrate this phenomenon, in Figure 5.3.1 left, we see the empirical Haar wavelet periodogram of a piecewise constant function of length 512 with a single jump in value at time point 300, without noise. Coefficients localised around the jump contain information about the change, which propagates through to coarser scales in a cone of influence. In time locations where the coefficients are localised entirely within one of the two piecewise constants, the coefficients are unaffected by the jump and are exactly zero. Hence, for a fixed scale, the empirical wavelet periodogram will be contaminated with the jump in a fixed number of coefficients, localised around the jump's location. All other coefficients in that scale will be unaffected. In particular, ignoring boundary effects, for the Haar wavelet, at scale j, 2^{-j} coefficients will be affected by a single jump.

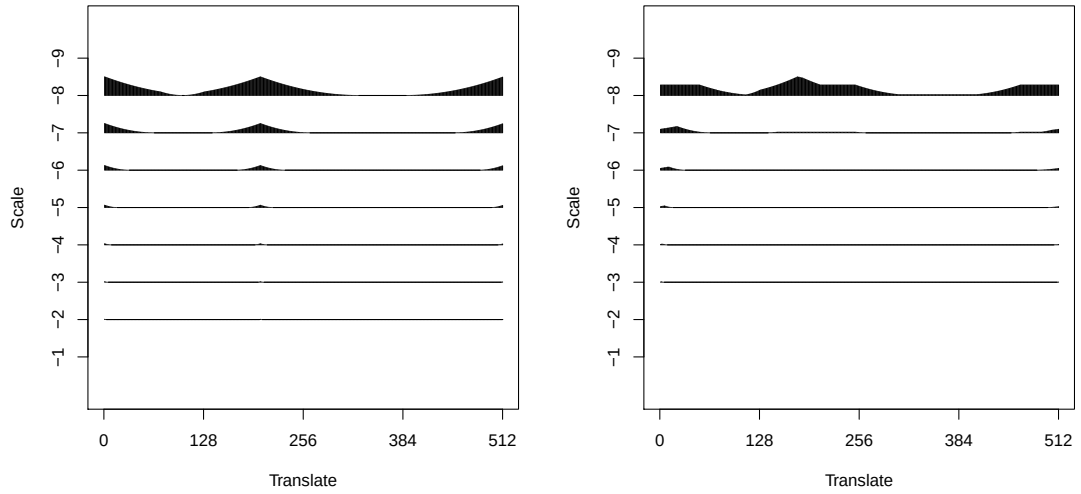

Figure 5.3.1: Left: empirical wavelet periodogram of a piecewise constant signal with one jump. Right: median smoothed version.

As described in Nason et al. (2000), asymptotically, as $T \to \infty$, we observe more data at an increasingly finer level on the interval $(0, 1)$. Therefore, as $T \to \infty$, we do not consider an increasing number of jumps in the mean, but rather we are observing a (finite) fixed number of jumps and observe more information about their structure as T increases. Suppose there are m changes in mean within the series. If we consider a fixed scale j^*, then a maximum number of wavelet coefficients, $Cm2^{-j^*}$, for some constant C, will be affected by the finite number of jumps in the mean of the series, due to the compact support of the wavelet. (For the Haar wavelet, $C = 1$). As T increases, an increasing number of the coefficients will contain information about the wavelet spectrum that is uncontaminated by the jumps.

Therefore, we wish to use these coefficients as an estimator for the wavelet spectrum. Provided the number of contaminated coefficients is small relative to the number of time points, a natural alternative to the running mean smoother – which would be affected by the jumps of the time series – is to use a running median instead. This is analogous to robust estimation of the variance using a median absolute deviation (MAD) estimator – by taking a median we can reduce the influence that the jumps have on the spectrum estimator. The running median

smoothed wavelet periodogram, with bin width $2n + 1$, is given by

$$\tilde{I}_k^j = \text{Median}\left(I_{k-n}^j, I_{k-n+1}^j, \ldots, I_{k+n-1}^j, I_{k+n}^j\right). \tag{5.3.1}$$

In Figure 5.3.1 right, we see the running median smoothed version of Figure 5.3.1 left, with a bin width of size 201. We see that the median smoothing is able to negate the effect of the change in mean of the signal on the raw wavelet periodogram, at least in the finer scales.

The above discussion motivates the following autocovariance estimation procedure. First, compute the raw wavelet periodogram I_k^j of the time series. Then, for each scale of interest j, obtain the smoothed wavelet periodogram by computing a running median of the I_k^j. Next, we must correct this median estimate to ensure unbiasedness. The raw wavelet periodogram is distributed asymptotically as a scaled χ_1^2 random variable, so we divide the smoothed wavelet periodogram by a suitable scale factor. This is analogous to the scale factor multiplication that occurs when using an MAD estimator. In our case, we divide the median smoothed estimate by $(7/9)^3 \approx 0.471$, as the mean of a χ_1^2 random variable is equal to 1, while the median is equal to 0.471. Denote the scaled, median smoothed estimate of I_k^j by $\hat{I}_k^j$. Next, we correct this estimate using the inverse of the A matrix to give the final estimate of the EWS:

$$\hat{S}_j(k/T) = \sum_{l=-J_0}^{-1} A_{jl}^{-1} \hat{I}_k^l, \tag{5.3.2}$$

for some $J_0 = \alpha \log_2 T < J$, where $\alpha \in (0, 1)$. Lastly, we estimate the autocovariance by plugging in the EWS estimate into the equation for the local autocovariance of an LSW process, i.e.

$$\hat{c}(k/T, \tau) = \sum_{j=-J_0}^{-1} \hat{S}_j(k/T)\Psi_j(\tau). \tag{5.3.3}$$

Choice of contributing scales and bin width. We must decide on a value

of α, the proportion of scales over which we correct. Higher values of α ensure a decomposition over a larger number of scales. In practice, not all of the scales will be informative. We propose using the proportion $\alpha = 3/5$. This is slightly less than the proportion 2/3 discussed in Sanderson et al. (2010), reflecting the fact that the time series may display mean changes.

The choice of the bin width parameter n in Equation (5.3.1) will also affect the quality of the estimate. We can choose to use level-dependent smoothing of the periodogram, and so a larger bin width is used in coarser scales. Choosing a bin width in this way is natural for two reasons: coarser scales will experience stronger autocorrelation, and coarser scales have larger number of coefficients affected by potential mean changes. If, instead, we assume second-order stationarity, we can drop the dependence on k in the estimation equations (5.3.2) and (5.3.3). In this case we can use a global scale-wise median when smoothing the wavelet periodogram, obtaining a time-independent estimate of the EWS and autocovariance.

5.3.2 Mean Change Detection

Having discussed our estimator of the autocovariance of the time series in the presence of mean changes, we now return to the problem of detecting the mean changes within a likelihood-based framework.

5.3.2.1 Likelihood-Based Test Statistic

Recall that the likelihood ratio test statistic for a single change in mean is given by

$$\lambda = \max_{1<p<T-2} \left\{ (\boldsymbol{X} - \hat{\boldsymbol{\mu}}_0)^{\mathsf{T}} \hat{\Sigma}^{-1} (\boldsymbol{X} - \hat{\boldsymbol{\mu}}_0) - (\mathbf{X} - \hat{\boldsymbol{\mu}}_1)^{\mathsf{T}} \hat{\Sigma}^{-1} (\mathbf{X} - \hat{\boldsymbol{\mu}}_1) \right\}.$$

In our approach, we also use sample means for the mean estimates $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_1$. The sample mean is asymptotically consistent, provided that the LSW process satisfies $\sup_{z\in(0,1)} \sum_\tau |c(z,\tau)| < \infty$, which can be thought of as a short memory assumption. The entries of the variance-autocovariance matrix Σ are estimated

using Equation (5.3.3).

The consistency of likelihood methods for detecting changepoints has been shown in Csörgö and Horváth (1997). It is well known that plug-in estimates that are consistent possess the same properties as maximum likelihood estimates, therefore the proof of consistency for identifying changepoints using plug-in estimates follows directly provided the estimates used are consistent. In our case, the sample means are consistent if we assume that the process has absolutely summable autocovariance. The mean smoothed estimate of the autocovariance is consistent, so intuitively we would expect the median to be also, although we give no formal proof of this here.

In practice, the variance-autocovariance matrices can be ill-conditioned and cause the test statistic to become numerically unstable. To combat this issue, we regularise the covariance matrix estimate using the approach of Rothman (2012), one of many possible regularisation methods. Denote by R the correlation matrix calculated using the initial covariance matrix estimate $\hat{\Sigma}$, and let $\theta \succ 0$ indicate that θ is symmetric and positive definite. Let θ^+ be the diagonal matrix with the same diagonal entries as θ. The method utilises a lasso-type procedure and computes an estimate of the correlation matrix θ given by

$$\hat{\theta} = \underset{\theta \succ 0}{\arg\min} \left\{ ||\theta - R||_F/2 - \tau \log|\theta| + \gamma|\theta^+ - \theta|_1 \right\}, \tag{5.3.4}$$

where the subscript F denotes the Frobenius norm, $\tau = 10^{-4}$ is a fixed constant, and $\gamma \geq 0$ is a tuning parameter. We recommend choosing γ as low as possible while still ensuring positive-definiteness, in order to capture the largest amount of autocovariance information within the series.

5.3.2.2 Detecting Multiple Changes in Mean

In the previous sections we have focussed on the single change in mean case, however often one is interested in detecting multiple changes. The likelihood ratio test statistic is consistent when estimating a single changepoint in the presence of

multiple changepoints, as shown in Vostrikova (1981). There are many proposed methods in the changepoint literature which tackle the multiple changepoint problem, for example Scott and Knott (1974); Bai and Perron (2003); Killick et al. (2012); and Fryzlewicz (2014).

In our work, we implement the Binary Segmentation algorithm of Scott and Knott (1974). Binary Segmentation is a technique for multiple changepoint detection where initially the entire data set is searched for a changepoint. If a changepoint is detected, the data are split into two subsegments defined by the detected changepoint. This procedure is then recursively repeated on subsequent subsegments until no further changepoints are detected. The algorithm is outlined in detail in Algorithm 1 below.

Algorithm 1: Binary Segmentation algorithm.

Input: A set of data of the form $\boldsymbol{X} = \{X_0, X_1, \ldots, X_{T-1}\}$.
A test statistic $\lambda(\cdot)$ dependent on the data.
An estimator of the changepoint location $\hat{p}(\cdot)$.
A rejection threshold (penalty) c.
Initialise: Let $\mathcal{C} = \emptyset$, and $\mathcal{S} = \{[0, T-1]\}$
while $\mathcal{S} \neq \emptyset$ **do**

1. Choose an element of $\mathcal{S}$, denoted $[s, t]$.
2. If $\lambda(X_{s:t}) < c$, remove $[s, t]$ from $\mathcal{S}$
3. If $\lambda(X_{s:t}) \geq c$, then:
    1. remove $[s, t]$ from $\mathcal{S}$;
    2. calculate $r = \hat{p}(X_{s:t}) + s - 1$, and add r to $\mathcal{C}$;
    3. if $r \neq s$ add $[s, r]$ to $\mathcal{S}$;
    4. if $r \neq t - 1$ add $[r + 1, t]$ to $\mathcal{S}$.

Output: the set of changepoints recorded $\mathcal{C}$.

5.3.2.3 Choice of Threshold/Penalty

In changepoint problems, it is common practice to use a penalty term to protect against overfitting, i.e. fitting too many changepoints. In the multiple changepoint setting, the penalty is usually a function of the number of changepoints; for ex-

ample see Harchaoui and Lévy-Leduc (2010). We outline two possible approaches for the selection of the appropriate penalty/threshold c.

First, one option is to use the asymptotic distribution of the test statistic to inform an appropriate value for c. If the autocovariance matrix was known, then an appropriate choice of penalty would be $2 \log T$. However, in our scenario the autocovariance is unknown and must be estimated. We therefore propose that a practical choice of threshold is given by $3 \log T$, which is motivated by the modified Bayesian information criterion as used in Zhang and Siegmund (2007). As remarked in Fisch et al. (2018), we utilise an inflated penalty in order to reflect the uncertainty of the unknown parameters. We have found in our empirical analysis that this threshold works well in practice, allowing for a low false discovery rate and a high detection rate.

In the second approach, we choose the threshold c via a Monte Carlo procedure. Using the estimate of the evolutionary wavelet spectrum of the time series, Monte Carlo simulations allow us to compute an appropriate $(100 - \alpha)\%$-quantile, which we choose to be c. We then detect a change if the test statistic of the observed series is greater than c.

The Monte Carlo approach works well in practice, and provides a fully automated method for detecting changes. However, it may become prohibitively slow, due to the need to calculate an inverse covariance matrix for each simulated series. To address this problem, we can instead use a modified test statistic, in which no inverse calculation is required. We use the alternative test statistic

$$\lambda = \max_{1<p<T-2} \left\{ \frac{1}{\hat{\sigma}_L^2(p)} \left((\boldsymbol{X} - \hat{\boldsymbol{\mu}}_0)^\mathsf{T} (\boldsymbol{X} - \hat{\boldsymbol{\mu}}_0) - (\mathbf{X} - \hat{\boldsymbol{\mu}}_1)^\mathsf{T} (\mathbf{X} - \hat{\boldsymbol{\mu}}_1) \right) \right\},$$

where $\hat{\sigma}_L^2(p)$ is an estimate of $\sigma_L^2(p) = \sum_\tau c(p/T, \tau)$, the possibly time-varying long run variance of the time series. We estimate this via the estimate of the local autocovariance and weighting via a kernel approach, in a similar fashion to

Tecuapetla-Gómez and Munk (2017). The estimator is given by

$$\hat{\sigma}_L^2(p) = \hat{c}(p/T, 0) + 2\sum_{\tau=1}^{2^{J_0}} \left(1 - \frac{\tau}{2^{J_0}+1}\right) \hat{c}(p/T, \tau).$$

The kernel approach is a common technique for estimation of the long run variance (Newey and West, 1986) and helps to ensure positive-definiteness of the resulting estimate. This approach enables faster implementation of the Monte Carlo threshold choice, while still taking the autocorrelation of the time series into account in the procedure. The approach of using the long run variance estimate in a changepoint detection procedure is also used, for example, in Dette et al. (2018) and Dette and Wu (2019).

Collecting the above ideas together, Algorithm 2 describes our LSW likelihood-based method which we henceforth refer to as LSWL. The output of Algorithm 2 is then used as the input for Binary Segmentation (Algorithm 1) in order to detect multiple changes in mean.

Algorithm 2: LSWL algorithm

Input: A set of data of the form $\boldsymbol{X} = \{X_0, X_1, \ldots, X_{T-1}\}$.
An estimate of the LACV, $\hat{c}$
Number of simulations, $nsim$, if using Monte Carlo method
Significance level, α, if using Monte Carlo method.
if *Method = Monte Carlo* **then**
  Calculate $\hat{\sigma}_L^2$
  Calculate λ, $\hat{p}$
  for $n \in \{1, ..., nsims\}$ **do**
    Simulate $\boldsymbol{X}_n$
    Calculate $\hat{\sigma}_{L,n}^2$
    Calculate λ_n,
  $\boldsymbol{\lambda} = \{\lambda_1, \ldots, \lambda_{nsims}\}$
  $c = q_{1-\alpha}(\boldsymbol{\lambda})$
else
  $c = 3 \log T$
  Calculate $\hat{\Sigma}$
  Calculate λ, $\hat{p}$
Output: test statistic λ, threshold c, estimated change location $\hat{p}$.

5.4 Simulation Study

In this section we assess the performance of LSWL via a simulation study. We simulate time series, with and without changes in mean, under a variety of error structures. In all simulations the results were obtained using Haar wavelets, and where applicable Haar wavelets were used to simulate from wavelet spectra. Other wavelet families yield similar results. In all simulations, 100 realisations were simulated, all of length 512. In each of the scenarios listed below the innovations ϵ_t are standard normal random variables.

If the estimated autocovariance matrix is numerically stable and can be regularised using the penalty value $\gamma = 0.4$, we utilise the standard version of LSWL, otherwise we use the Monte Carlo version with 100 replications, at a significance level of 5%. In alignment with the discussion in Section 5.3.1, we estimate the EWS at the finest $J_0 = 5$ scales, and hence the autocovariance up to lag $\tau = 2^5 = 32$. For the second-order stationary scenarios in the simulation study, we utilise global scale-wise median smoothing for spectral estimation. For the nonstationary scenarios, we use a running median of bin width 151 at each scale for simplicity. Our implementation of LSWL utilises the `R` package `wavethresh` (Nason, 2016b) in order to fit the LSW model.

Where applicable, we compare our method to the AR1Seg method of Chakar et al. (2017) and the NPCP method of Bücher and Kojadinovic (2016), whose implementations are readily available in `R` (Chakar et al., 2014; Kojadinovic, 2015). AR1Seg can detect multiple mean changes in the presence of (stationary) AR(1) errors. The method robustly estimates the autocorrelation parameter of the AR(1) process, which is then plugged in to a dynamic programming algorithm in order to detect multiple changes. NPCP can detect a single change in the presence of stationary autocorrelation subject to mild technical assumptions. NPCP utilises a multiplier bootstrap in order to derive critical values for hypothesis testing in the single change case. In our simulations the significance level for NPCP is set at 5%. For illustrative purposes, we also compare to the performance of NOT,

implemented in the `R` package `not` (Baranowski et al., 2016), to highlight the effect of not incorporating the autocorrelation within the changepoint method. NOT detects multiple changepoints by applying a generalised likelihood ratio test to randomly selected subsamples of the data. NOT is shown to be consistent in the presence of autocorrelated errors, however there is no option for including autocorrelation information within the implementation.

5.4.1 Null Hypothesis Performance

We first assess the performance of the method under the null hypothesis of a constant mean (without loss of generality equal to zero) of the time series. We simulate a variety of stationary and nonstationary error structures, that highlight the ability of LSWL to successfully account for autocorrelation in order to maintain a low false positive rate. In each simulation, a false positive is recorded if the method detected any changes, and we record the proportion of the 100 simulations resulting in a false positive.

5.4.1.1 No Changepoint, Various AR(1) Parameters

We investigate the effect of varying AR(1) parameter values on the false positive rate of the methods. Data sets are simulated from

$$X_t = \phi X_{t-1} + \epsilon_t,$$

where ϕ is given by the values in Table 5.4.1. Included is the case of white noise where $\phi = 0$. The values underlined indicate performance where the type 1 error is clearly not controlled. We see from Table 5.4.1 that negative values for the AR(1) parameter do not hamper the performance of the methods, while large positive values do. Furthermore, the results show that LSWL compares similarly or favourably with AR1Seg and NPCP. AR1Seg has a high false positive rate when the AR parameter is equal to 0.9. As expected, NOT confuses changes in mean

with strongly autocorrelated errors.

False Discovery Rate	AR(1) Parameter						
Method	-0.9	-0.6	-0.3	0	0.3	0.6	0.9
LSWL	0.00	0.00	0.00	0.00	0.08	0.09	0.05
AR1Seg	0.01	0.02	0.01	0.02	0.05	0.03	0.45
NPCP	0.01	0.02	0.02	0.05	0.07	0.05	0.03
NOT	0.00	0.00	0.00	0.01	0.34	0.99	1.00

Table 5.4.1: False discovery rates across different AR(1) scenarios with constant mean equal to zero.

5.4.1.2 No Changepoint, Various Autocorrelation Scenarios

Next, we investigate a variety of autocorrelation scenarios, representing some of the most common time series models. Data sets are simulated from models A – H, as described below:

(A) AR(2) model with parameters $\phi_1 = 0.5$ and $\phi_2 = 0.3$.

(B) ARCH(3) model with parameters $\alpha_1 = 1$, $\alpha_2 = 0.5$, $\alpha_3 = 0.4$.

(C) Stationary LSW model with $S_{-4} = S_{-5} = 1$, $S_j = 0$ otherwise.

(D) ARMA(1,6) model with AR parameter $\phi = 0.5$, MA parameters $\theta_1 = 1$, $\theta_2 = -1$, $\theta_3 = 0.5$, $\theta_4 = 0.5$, $\theta_5 = 1$, $\theta_6 = 0.5$.

(E) Time-varying AR(1) model where the AR(1) parameter moves from 0.7 to 0.3 linearly, i.e.

$$X_t = \phi_t X_{t-1} + \epsilon_t, \qquad \phi_t = \frac{0.4(1-t)}{511} + 0.7.$$

(F) Nonstationary LSW model with spectrum given by

$$S_j(z) = \begin{cases} 1/5 + \sin^2(2\pi z) & \text{for } j = -5, z \in (0,1), \\ 1 & \text{for } j = -1, z \in (100/512, 200/512), \\ 0 & \text{otherwise.} \end{cases}$$

(G) Nonstationary LSW model with spectrum given by

$$S_j(z) = \begin{cases} 1 & \text{for } j = -1, z \in (0, 100/512), \\ 1 & \text{for } j = -2, z \in (100/512, 300/512), \\ 1 & \text{for } j = -3, z \in (300/512, 1), \\ 0 & \text{otherwise.} \end{cases}$$

(H) $X_t = \sigma_t \epsilon_t$, where

$$\sigma_t = \begin{cases} 24(t/512)^2 + 4(t/512) + 4 & \text{for } t \in [0, 299], \\ -32(t/512)^2 + 8(t/512) + 7.62 & \text{for } t \in [300, 511]. \end{cases}$$

The scenarios (A) – (D) represent second-order stationary time series, while scenarios (E) – (H) are second-order nonstationary. The results are reported in Table 5.4.2, again with underlined values indicating uncontrolled type 1 error. We see that LSWL is able to maintain a low false positive rate in all scenarios, comparing similarly or favourably to NPCP. Whilst we expect our method under models C, F and G to perform well, it is reassuring to find good performance in the cases where the error structure is not simulated using an LSW process. AR1Seg performs poorly in the case of an AR(2) model and ARCH(3) model, and model F. As expected, NOT generally performs poorly due to autocorrelation, while performing well in the presence of nonstationary IID errors (model H).

False Discovery Rate	Scenario							
Method	A	B	C	D	E	F	G	H
LSWL	0.02	0.07	0.00	0.01	0.09	0.00	0.00	0.00
AR1Seg	0.97	0.96	0.11	0.08	0.14	0.94	0.05	0.15
NPCP	0.07	0.03	0.00	0.06	0.05	0.00	0.00	0.07
NOT	1.00	0.60	1.00	**1.00**	0.92	1.00	0.14	0.03

Table 5.4.2: False discovery rates across different autocorrelated error scenarios, with constant mean equal to zero.

5.4.2 Alternative Hypothesis Assessment

Next, we assess the method's performance in detecting changes in mean. We use the same error structure scenarios defined in the previous section, and examine the performance of the method in both the single and multiple changepoint settings.

5.4.2.1 Single Changepoint, Various Stationary Autocorrelation Scenarios

In these scenarios there is a single change in mean at time 300, with the size of the change, δ, chosen so that that $\delta/\sigma = 1$, where σ is the standard deviation of the time series. We investigate the performance of the methods for the IID model, AR(1) model with parameter $\phi = 0.6$, and models (A) – (D), by reporting the number of detected changepoints of each method. We also report the average mean squared error (MSE) of the estimated signal $\hat{\mu}$ across the 100 simulations, where the MSE is given by

$$\text{MSE} = \frac{1}{T}\sum_{t=0}^{T-1}\left(\mu_t - \hat{\mu}_t\right)^2.$$

The MSE is an alternative comparison to examine how well each method estimates the mean function in the model. The results are given in Table 5.4.3, with the correct number of changepoints in the top row in bold and the modal value for

Model	Method	No. of changepoints				MSE
		0	**1**	2	≥ 3	
IID	LSWL	0	**98**	2	0	**0.010**
	AR1Seg	0	**94**	4	2	0.012
	NPCP	0	**100**	-	-	**0.010**
	NOT	0	**96**	4	0	0.011
AR(1) 0.6	LSWL	0	**95**	5	0	0.070
	AR1Seg	27	**64**	7	2	0.165
	NPCP	4	**96**	-	-	**0.056**
	NOT	0	2	4	**94**	0.441
A	LSWL	12	**76**	12	0	0.291
	AR1Seg	0	2	5	**93**	1.652
	NPCP	35	**65**	-	-	**0.174**
	NOT	0	0	0	**100**	1.040
B	LSWL	0	**71**	27	2	0.180
	AR1Seg	0	5	1	**94**	2.044
	NPCP	0	**100**	-	-	**0.061**
	NOT	0	40	2	**58**	1.013
C	LSWL	2	**85**	12	1	0.060
	AR1Seg	**61**	22	8	9	0.374
	NPCP	0	**100**	-	-	**0.034**
	NOT	0	0	0	**100**	1.26
D	LSWL	6	**80**	14	0	0.616
	AR1Seg	**49**	40	5	6	1.226
	NPCP	9	**91**	-	-	**0.363**
	NOT	0	0	1	**99**	3.415

Table 5.4.3: Method performance comparisons for the single changepoint and stationary second-order scenarios, reporting number of changepoints detected and average mean squared error of the estimated mean.

each method also in bold. The table indicates that LSWL performs well across all considered scenarios. Surprisingly, LSWL outperforms AR1Seg in the case of AR(1) errors. Departure from an AR(1) error structure results in poor performance for AR1Seg. LSWL also performs well in the IID case, in part due to the ability to obtain an accurate estimate of the trivial autocovariance. Note that NPCP is capable of detecting only a single change: running the simulations for at-most-one-change LSWL gives similar results to NPCP.

Model	Method	No. of changepoints				MSE
		0	**1**	2	≥ 3	
E	LSWL	0	**78**	15	7	0.059
	AR1Seg	17	**59**	7	17	0.160
	NPCP	0	**100**	-	-	**0.037**
	NOT	0	8	2	**90**	0.292
F	LSWL	0	**89**	7	4	0.026
	AR1Seg	3	3	3	**91**	0.302
	NPCP	0	**100**	-	-	**0.010**
	NOT	0	0	0	**100**	0.519
G	LSWL	0	**100**	0	0	**0.004**
	AR1Seg	**62**	34	3	1	0.162
	NPCP	0	**100**	-	-	**0.004**
	NOT	0	**88**	3	9	0.012
H	LSWL	0	**100**	0	0	**0.375**
	AR1Seg	0	**68**	14	18	0.508
	NPCP	0	**100**	-	-	0.388
	NOT	0	**94**	1	5	0.721

Table 5.4.4: Method performance comparisons for the single changepoint and non-stationary second-order scenarios, reporting number of changepoints detected and average mean squared error of the estimated mean.

5.4.2.2 Single Changepoint, Time-Varying Autocorrelation Scenarios

In these scenarios there is again a single change in mean at time 300, with the size of the change, δ, chosen so that that $\delta/\sigma = 1$, where σ is the maximum standard deviation of the time series. We investigate the performance of the methods for models (E) – (H), with the results reported in Table 5.4.4. We see that LSWL performs well in general, with a high proportion of simulations identifying the correct number of changepoints. NPCP also performs well, despite the fact that it is not designed to handle nonstationary second-order structure.

5.4.2.3 Multiple Changepoints, Various Stationary Autocorrelation Scenarios

Next, we analyse the performance of the methods in the case where there are multiple changepoints, with the error structures the same as in the single changepoint case. In these simulations, the signal contains 3 changepoints, at times $p = 100$, 180 and 380. For the IID and AR(1) case, the mean within each segment alternates

Model	Method	No. of changepoints					MSE
		≤ 1	2	**3**	4	≥ 5	
IID	LSWL	8	0	**92**	0	0	0.054
	AR1Seg	4	0	**98**	1	1	0.026
	NOT	0	0	**99**	1	0	**0.025**
AR(1) 0.6	LSWL	22	7	**55**	14	2	**0.271**
	AR1Seg	**48**	3	36	5	8	0.346
	NOT	0	0	4	1	**95**	0.471
A	LSWL	3	0	**71**	21	5	**0.661**
	AR1Seg	2	0	4	3	**91**	1.122
	NOT	0	0	0	0	**100**	1.393
B	LSWL	0	0	**66**	20	14	**1.018**
	AR1Seg	0	0	1	3	**96**	3.388
	NOT	0	0	28	6	**66**	2.037
C	LSWL	40	0	**59**	1	0	**0.554**
	AR1Seg	**75**	1	5	2	17	1.037
	NOT	0	0	0	0	**100**	1.303
D	LSWL	2	0	**70**	25	3	**1.688**
	AR1Seg	**45**	2	39	4	10	3.670
	NOT	0	0	0	1	**99**	3.381

Table 5.4.5: Method performance comparisons for the multiple changepoint and stationary second-order scenarios, reporting number of changepoints detected and average mean squared error of the estimated mean.

between 0 and δ, where δ is calculated so that $\delta/\sigma = 1.25$. For cases (A) – (D), δ is calculated so that $\delta/\sigma = 2$, to reflect the difficulty of the scenarios. Note that this up, down, up pattern is not favourable to the Binary Segmentation algorithm, and represents a “worst case" scenario. The results of the simulation are reported in Table 5.4.5. We see that, for all but the IID scenario, LSWL offers the strongest performance, both in terms of detecting changepoints and mean squared error. In the IID case, LSWL is still able to achieve good performance. As in the single changepoint case, LSWL outperforms AR1Seg in the case of AR(1) errors.

5.4.2.4 Multiple Changepoints, Time-Varying Autocorrelation Scenarios

Finally we examine the multiple changepoint setting with nonstationary second-order structure. We use the same signal from the previous section, where the mean within each segment alternates between 0 and δ, where δ is calculated so

Model	Method	No. of changepoints					MSE
		≤ 1	2	**3**	4	≥ 5	
E	LSWL	13	4	**68**	14	1	**0.219**
	AR1Seg	28	3	**59**	8	2	0.282
	NOT	0	0	10	18	**82**	0.300
F	LSWL	35	0	**56**	8	1	**0.269**
	AR1Seg	1	0	6	5	**88**	0.484
	NOT	0	0	0	0	**100**	0.543
G	LSWL	0	0	**99**	1	0	**0.017**
	AR1Seg	**79**	3	15	2	1	0.281
	NOT	0	0	**83**	1	16	0.026
H	LSWL	0	0	**99**	1	0	1.232
	AR1Seg	0	0	**73**	13	14	1.057
	NOT	0	0	**89**	9	2	**0.832**

Table 5.4.6: Method performance comparisons for the multiple changepoint non-stationary second-order scenarios, reporting number of changepoints detected and average mean squared error of the estimated mean.

that $\delta/\sigma = 1.25$ for maximum standard deviation σ. Error structures are again simulated from models (E) – (H) as in the single changepoint case. The results of the simulation are shown in Table 5.4.6. We see that LSWL has the strongest performance in terms of detected changepoints, and has the lowest MSE for all but one scenario. Therefore, we find that LSWL performs well in both single and multiple changepoint scenarios, as well as in the presence of stationary and nonstationary second-order structure.

5.5 Data Applications

In this section we demonstrate the potential uses of LSWL with analysis of two data sets. Firstly, we perform changepoint analysis on historical Ebay stock price data, and secondly we examine UK house price index data that was previously analysed in, amongst others, Baranowski et al. (2019). Accurate segmentation of these data sets is crucial to ensure an informed interpretation and analysis – falsely detecting too many changes in mean may lead to false assertions about the underlying series. False detection of changes in stock price data may lead to incorrect investing decisions, while inaccurate assessment of the house price index

may detrimentally inform housing policy.

5.5.1 Ebay Stock Price Data

Shown in Figure 5.5.1 is the historic daily closing stock price in dollars for Ebay over a roughly 4 year time period from approximately April 2013 to April 2017, with corresponding fitted changepoints. One obvious changepoint is apparent in mid-July of 2015: this is due to Ebay and Paypal splitting into separate public companies at this time, roughly 12 years after Ebay had acquired Paypal in 2003. It is clear that we will detect this change using our methodology, as will any multiple changepoint detection method. However, it is not immediately obvious how many subsequent changes there are: one might expect that a financial time series of this sort may exhibit strong autocorrelation, with stock prices close in time being highly correlated. Therefore, we may take the view that this time series has a small number of changes, with a strong level of autocorrelation, as opposed to possessing many mean changes with relatively low autocorrelation.

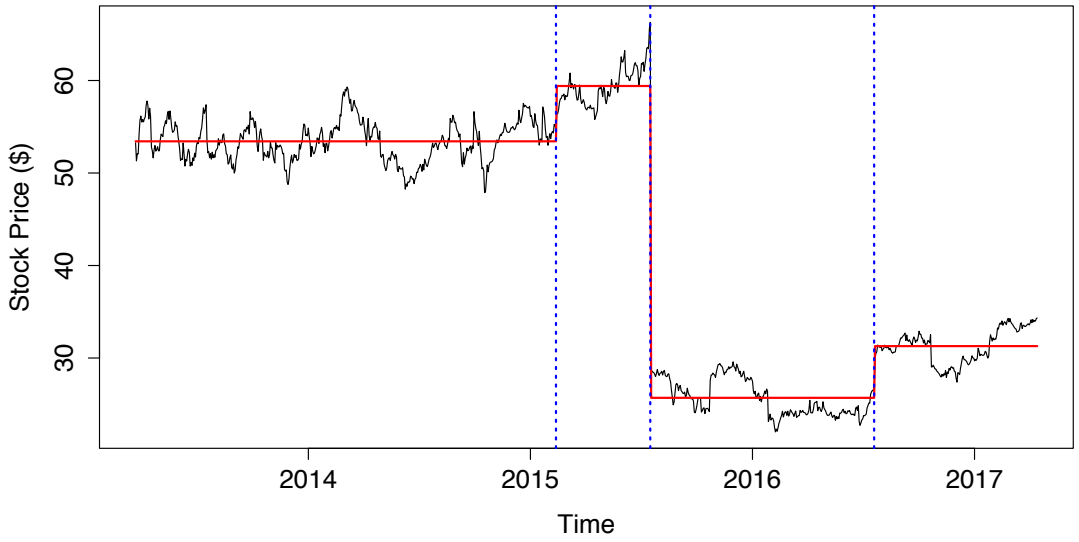

Figure 5.5.1: Ebay closing stock price over roughly 4 year period. Fitted mean function in solid line.

In this example we have used a global scale-wise median smoothing and anal-

ysed the finest $J_0 = 6$ scales in the spectrum (and hence autocovariance) estimate. We detect three total changes in mean, with locations shown by the vertical dotted lines in Figure 5.5.1. The fitted estimate for the mean is shown in the red solid line. We detect the change associated to the Ebay-Paypal split, and detect two subsequent changes that occur at times close to the release of Ebay quarterly results. Therefore, we can interpret the LSWL segmentation as saying that any other large fluctuations in the data that are noticeable by eye can be attributed to serial autocorrelation.

In our analysis of the Ebay data, we used a global scale-wise estimate of the wavelet spectrum when estimating the autocovariance, i.e. we assumed the second-order structure is stationary. Having fitted our changepoint model to the series, we now validate this assumption. Running a test for second-order stationarity on the mean removed series (for example using the work of Nason (2013) or Bücher and Kojadinovic (2016)), we find that we fail to reject the null hypothesis of second-order stationarity. This is reassuring on two counts: firstly that the series can be considered second-order stationary, and secondly that our segmentation of the series was able to identify all mean changes. Furthermore, estimating the autocorrelation of the series highlights the large levels of autocorrelation present in the series. In Figure 5.5.2 we see the estimated autocorrelation of the mean removed series. We can see that autocorrelation persists into very high lags in the series, with significant positive autocorrelation at lag $\tau = 1$ to approximately at least lag 25.

Running standard multiple changepoint detection methods on this series will result in detecting a large number of changes. Under default settings, PELT obtains a total of 35 detected changepoints, however setting a penalty value of approximately $610 \approx 88 \log(1024)$ yields the exact 3 changepoints that LSWL detected. (Choosing this penalty value in a principled manner is in practice extremely challenging). This result however is reassuring, as PELT uses the standard Gaussian likelihood for a cost function when fitting changepoints. Similarly, WBS fits 48

changepoints, NOT fits 25, and AR1Seg fits 12 changepoints.

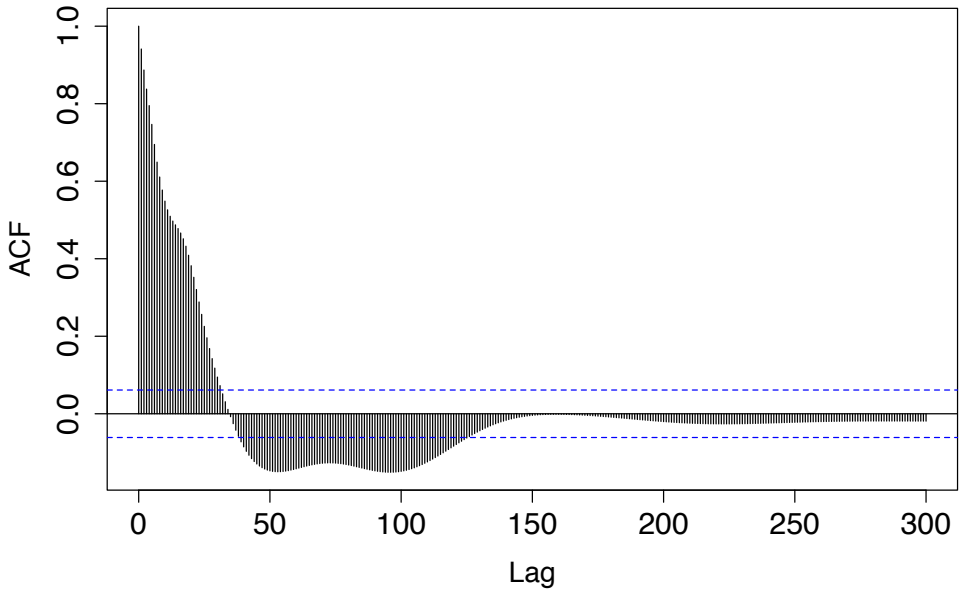

Figure 5.5.2: Estimate of the autocorrelation of the mean subtracted Ebay stock price data.

5.5.2 UK House Price Index

In this example we analyse the monthly percentage changes in UK house price index (HPI) data set, that provides insight into the estimated overall changes in house prices across the UK. The data are available to download online (`https://www.gov.uk/government/statistical-data-sets/uk-house-price-index-data-downloads-march-2020`), while a detailed description for the calculation of the HPI is also given online from the UK Land Registry (2020).

Fryzlewicz (2018) and Baranowski et al. (2019) analyse the percentage changes in the HPI for three London boroughs; Hackney, Newham and Tower Hamlets, all of which are in East London. In both of the analyses, the authors do not allow for the possibility of autocorrelated errors, which could potentially cause spurious changepoints to be detected. Baranowski et al. (2019) allow for changepoints in the variance of the time series, while Fryzlewicz (2018) does not. We analyse the Newham HPI data as an illustrative example, since it is this series for which Fryzlewicz (2018) and Baranowski et al. (2019) detect the most changepoints out

of the three boroughs.

Figure 5.5.3 shows monthly percentage changes in the HPI for the Newham borough, and the corresponding fitted changepoints obtained using LSWL. In this example we have used a bin width of size 71 for the running median smoothing and analysed the finest $J_0 = 5$ scales in the spectrum estimate. The autocovariance matrix is regularised using $\gamma = 0.45$. We see that there are few changepoints fitted to the series. This is in contrast to Fryzlewicz (2018), whose methodology estimates at least 10 changepoints, and Baranowski et al. (2019) who find 5 changepoints in the series. We believe that the most likely explanation for this is the presence of significant autocorrelation within the series, which may cause the methods to overfit the number of changepoints, as noted in Lund and Shi (2020). To illustrate this point, in Figure 5.5.4, we plot the estimated autocorrelation for the mean removed Newham series, at lags 1, 2 and 3. We see that there is significant (time-varying) autocorrelation within the series at these lags, which can inhibit the performance of a changepoint detection algorithm if the autocorrelation is not taken into account.

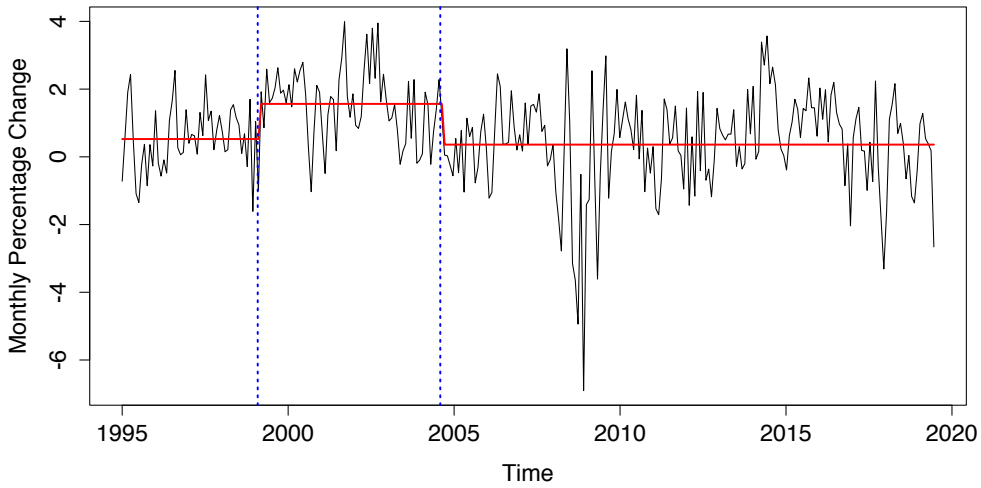

Figure 5.5.3: Newham HPI series with fitted changepoints (solid lines).

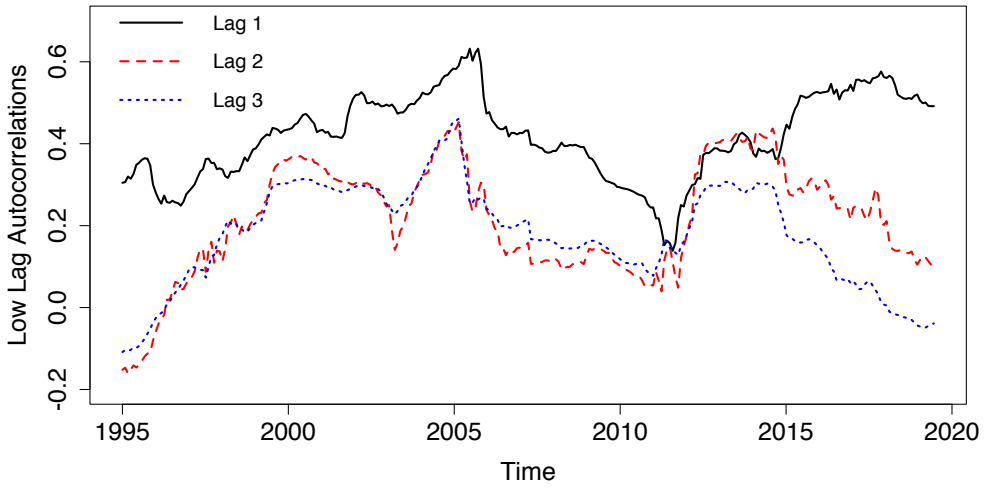

Figure 5.5.4: Estimated lag 1 (solid line), lag 2 (dashed line), and lag 3 (dotted line) autocorrelation of the Newham HPI series.

5.6 Concluding Comments

Time series can be commonly modelled using changepoint methods whereby the mean function is assumed to be piecewise constant. It is often the case that mean changepoint detection methods make the assumption that the observations are independent, when in reality this may not be true. The work presented in this chapter addresses this problem by introducing a likelihood-based test under the LSW framework to detect changes in the mean of time series whose unknown, non-trivial autocovariance can vary over time. The proposed approach is shown to work well on simulated data, and compares similarly or favourably with other methods in the literature. The potential uses of the method are shown on two data examples, which highlight the ability to account for highly autocorrelated time series.

LSWL performs well in both single and multiple changepoint scenarios, as well as in the presence of stationary and nonstationary second-order structure. We note, however, that LSWL is likely to be less effective in the very frequent changepoint setting. Firstly, this is due to the limitations of Binary Segmentation

itself, which is effectively a greedy method for locating changepoints, as described in Fryzlewicz (2014). Secondly, in the case of frequent changepoints, the estimate of the EWS is likely to become increasingly contaminated by the changepoints. This will cause inaccurate estimation of the autocovariance, which will impact the power of LSWL. Lastly, we remark that the procedure may become prohibitively slow for large sample sizes. This is due to the high computational cost of the autocovariance matrix inversion and regularisation procedure, which are both $\mathcal{O}(T^3)$. In situations where the sample size is very large, it may be advantageous to employ the Monte Carlo version of LSWL, which has computational cost $\mathcal{O}(mT \log T)$, where m is the number of Monte Carlo replications performed.

One interesting avenue for future investigation is the incorporation of the 'robust' autocovariance estimate within more sophisticated changepoint algorithms. For example, the NOT method of Baranowski et al. (2019) is shown to be consistent in the presence of a stationary short memory process, however this is achieved assuming a known variance and given pre-estimate of the autocorrelation. We could use the wavelet-based autocovariance estimate described in this chapter within NOT or similar changepoint detection algorithms, which could enable the estimation of changepoints under unknown autocorrelation.

Chapter 6

Conclusions and Further Work

In this thesis, we have considered wavelet-based methodology for the modelling of time series that exhibit both first and second-order nonstationary structure. In this final chapter, we summarise the main contributions made in Chapters 3 – 5. Furthermore, we outline some potential avenues for future research that have arisen as a result of the work undertaken for this thesis.

6.1 Contributions ad Discussion

In Chapter 3, we considered the problem of modelling first and second-order nonstationary time series, which are often observed in practice. We employed the method of differencing in order to remove the trend; a commonly used technique in classical time series analysis. We derived a wavelet-based method for estimating the nonstationary second-order structure of the original series using differencing, and provided theoretical results related to the asymptotic behaviour of the estimator. We developed a wavelet thresholding-based estimator of the trend of the time series, in which the threshold is set by incorporating the difference-based estimate of the second-order structure.

We performed a simulation study that highlighted the strong performance of the methodology. We showed that, in the presence of a trend, the difference-based wavelet estimator of the spectrum is able to perform similarly to the standard

spectrum estimator in the absence of a trend. Furthermore, we showed that our proposed trend estimator compared favourably to both a stationary wavelet thresholding approach, and the local MAD-based wavelet thresholding approach of von Sachs and MacGibbon (2000). The trend estimator was shown to work well for a variety of trends and both stationary and nonstationary error structures.

Lastly, we demonstrated the utility of the method by analysing two data examples. In the first example we analysed North Atlantic wave height data, showing wave height and wave height variability to be larger in the winter than the summer. In the second example, we re-analysed the baby electrocardiogram data set that was examined in the original LSW work of Nason et al. (2000). In the authors' work, analysis of the first-differenced series was performed, however in our work we performed analysis on the original series using our proposed methodology. This enabled us to gain a greater insight into the second-order structure of the data, and revealed a more apparent link between the electrocardiogram data and the accompanying sleep state of the baby.

In Chapter 4, we also considered the problem of modelling first and second-order nonstationary time series, but approached the problem from a different angle. In this chapter, we considered different assumptions on the trend function of the time series to those of Chapter 3. To tackle this problem, we developed wavelet-based techniques that expand upon the work on locally stationary wavelet (LSW) process modelling of Nason et al. (2000). We removed the restrictive zero-mean assumption of locally stationary wavelet processes, and extended the applicability of the LSW model to include a trend component. We developed an associated estimation theory for both first and second-order time series quantities and showed that our estimators achieve good properties in isolation of each other by making appropriate assumptions on the series trend.

We illustrated the strong performance of the proposed methodology via a simulation study. We demonstrated that the spectral estimate in the presence of a smooth trend displays similar performance to the standard spectral estimate in

the absence of a trend. We showed that using a smoother wavelet for spectral estimation can result in a more accurate estimate. Moreover, we showed that the linear wavelet thresholding trend estimator offers strong performance in a variety of scenarios and compared favourably to competing methods. We demonstrated the potential uses of the method by analysing the global mean sea temperature time series. In doing so, we provided new insights into the data set, whilst aligning with other aspects of analyses in the literature.

Furthermore, we proposed a new way to perform boundary handling of the discrete wavelet transform in the case where a trend is present. This method enables one to perform a discrete wavelet transform of a time series whose length is not dyadic, ensuring wider applicability for our methodology. Finally, we provided an important theoretical contribution to the LSW literature. We showed that, for all Daubechies compactly supported wavelets, the wavelet-based operator B, with entries given by $B_{jl} = 2^{j/2} A_{jl} 2^{l/2}$, possesses a bounded inverse. This result is of general interest to the LSW literature as it extends many results, that were only shown to hold for Haar (and Shannon) wavelets, to all wavelets within the Daubechies compactly supported wavelet family.

Due to the similarities between Chapters 3 and 4, it is helpful to now comment on the relations and differences between the chapters. Ultimately, both chapters tackle the problem of modelling and estimation of time series that are both first and second-order nonstationary. The work in the chapters are in the same vein, although the two chapters attack the problem from different angles. The work in Chapter 3 can be viewed as a rigorous investigation into the uses of differencing in the case of locally stationary time series. Differencing is an ubiquitous practice in time series analysis: the work here shows that one should be wary of the potential effects differencing has on the spectral properties of a nonstationary time series. The work in Chapter 3 may be useful for analysing the effect of differencing on the second-order properties of the time series, and as an exploratory data analysis tool to inform whether differencing is appropriate in a given context.

The work in Chapter 4 can be viewed as more of a complete modelling framework within the locally stationary time series setting. We have shown that the well-known wavelet-based methods that are used for modelling first *or* second-order nonstationary time series can be adapted to account for first *and* second-order nonstationarity. In particular, trend estimation can be carried out through linear wavelet thresholding, avoiding the difficult problem of incorporating the second-order properties in the procedure. Spectral estimation can make use of the LSW process construction and vanishing moments property of wavelets to incur minimal additional bias in the presence of a trend. The methodology allows for simultaneous modelling of first and second-order quantities, enabling the estimation of these quantities in isolation of one another, through the use of a simple, intuitive procedure.

For very smooth trends, it is likely that the linear wavelet thresholding technique in Chapter 4 will produce stronger practical performance. For polynomial trends, and trends that can be well approximated by polynomials, the methods described in Chapter 4 are recommended. In the case of more complex trends, it is recommended to use the trend estimator proposed in Chapter 3, as it is adaptive to the trend. For example, the piecewise continuous quadratic trend which is non-differentiable at two points, used in the simulation study in Chapter 3, is one setting where the use of the methods in Chapter 3 would be preferred.

We also stress that, with respect to the modelling of the mean function, the emphasis in these chapters has been on modelling smooth trends. Therefore, it should be expected that the performance of the methods will suffer when the trend present in the time series is not smooth, or is discontinuous. For example, we would expect the methods to perform worse in the presence of trends like the "Doppler" and "Bumps" signals that were introduced by Donoho and Johnstone (1994). This fact is one of the motivations behind the work in Chapter 5, in which we consider a piecewise constant mean function instead. The effect of a very non-smooth trend would be detrimental to both the spectral and trend estimation. For

example, in the work of Chapter 3, if the trend is not sufficiently smooth, then the differencing will not remove the trend, which causes the spectral estimate to be positively biased. This in turn causes the threshold for the trend estimator to be artificially inflated, resulting in too many coefficients being thresholded.

It may therefore be beneficial to perform some empirical checks when using the techniques in practice, such as those carried out in the first data example in Chapter 3. One can compare the spectral estimate obtained using the difference-based method, to the one obtained after detrending the series using the trend estimate. One could also compare to the spectral estimate obtained by analysing the original series without detrending or differencing. Indeed, in the case of a more abruptly changing smooth trend, one possible workaround is an iterative approach to spectrum and trend estimation, by iterating between the two estimates until some sort of convergence is reached. Another possible approach is to vary the coarsest scale analysed in the trend estimation procedure. This could perhaps be chosen in a more systematic fashion using a cross validation approach as in Nason (1996). Finally, one could use a median smoother for the spectral estimate, instead of a running mean, in a similar fashion to the work in Chapter 5, to further improve practical performance.

In Chapter 5, we considered the scenario where the mean function of the time series is given by a piecewise constant function. There has been much attention in the literature to the problem of detecting changes in mean in a piecewise constant time series. However, many such methods only consider the case where the error structure of the time series is independent, identically distributed Gaussian noise. We developed methodology for detecting the location of mean changepoints in a time series that exhibits nonstationary second-order behaviour. We proposed a likelihood-based method (LSWL) using wavelets to detect changes in mean in time series that can display a much more general error structure than simply IID errors. Our proposed technique was shown to work well for time series with a variety of error structures through the use of a simulation study. The method was applied

to two financial time series data sets, which highlighted the proposed method's ability to work well in the presence of significant autocorrelation.

In practice, LSWL will work well in distinguishing between strong autocorrelation and changepoints, as shown by the simulation results under the null hypothesis. The method is also able to detect multiple changepoints in the presence of both stationary and time-varying autocorrelation, while also performing well in the simple case of IID errors. Practical performance of the method is likely to suffer in the presence of frequent changepoints. This is due to the inherent limitations of binary segmentation as a 'greedy' algorithm, and the increasing effect of the changepoints on the EWS estimate. Finally, we note that the computational cost of the autocovariance matrix-based version of LSWL scales poorly with the sample size T. Therefore, we recommend using the faster Monte Carlo version in the case where the computational speed is of concern.

6.2 Further Work

We conclude this thesis by considering a few possible areas for future research, arising from the work undertaken in Chapters 3 – 5. A natural avenue for future research, stemming from the work of Chapters 3 – 4, is to incorporate other work on LSW process modelling within our framework. In the literature, there are a number of applications of the LSW model to classical time series problems, however these works all rely on the zero-mean assumption of the original LSW model. Applying our framework, which can model time-varying first and second-order behaviour, could allow for new approaches to these classical problems.

For example, we could research the problem of forecasting as in Fryzlewicz et al. (2003). This would allow for the forecasting of time series that display a trend behaviour. Another possibility is the application to changepoint analysis as in Killick et al. (2013). This approach would involve the detection of changes in the second-order structure of the time series in the presence of a smooth trend. Another potential direction is to incorporate the more recent work of Cardinali

and Nason (2017) on locally stationary packet processes into our methodology. This would allow for a wider class of time series models to be considered within our proposed work.

Furthermore, developing a statistical test for trend is another possible extension of the work in Chapters 3 – 4 concerning wavelet-based trend estimation. One possible approach in this regard is to adopt a similar strategy to that of Nason (2013). In this work, the author uses the Haar wavelet transform of the raw wavelet periodogram to test for nonstationarity of the second-order structure. If the time series is second-order stationary, then the raw wavelet periodogram will be a constant function, and so the Haar wavelet coefficients should be zero. In our scenario, the null hypothesis would be that the trend is constant across time, which would mean the Haar wavelet coefficients would be zero also. We should be able to derive appropriate critical values for testing for a nonstationary trend using the Haar wavelet coefficients. This would be achieved using asymptotic normality of the wavelet coefficients, along with the estimators we developed for the nonstationary second-order structure of the time series.

In Chapter 3 we proved that the operator P possessed a bounded inverse for the Haar and Shannon wavelets; the two wavelets that can be viewed as the extreme ends of the Daubechies family. It would be of interest to extend this result to all wavelets within the compactly supported Daubechies wavelet family. This could perhaps be achieved by modifying the proof strategy of the result on bounded invertibility of A used in Chapter 4. In this case, the difficulty lies in showing linear independence for the family of vectors given by $\{2^{-j/2}[\Psi_j(\tau) - \Psi_j(\tau - 1)]\}_{j \leq -1}$, for all compactly supported Daubechies wavelets. This problem could be approached using similar techniques from Nason et al. (2000), who show that the family $\{\Psi_j(\tau)\}_{j \leq -1}$ is linearly independent for all Daubechies compactly supported wavelets. Such a rigorous investigation into the properties of P was beyond the scope of the work in Chapter 3.

Further interesting research questions emanate from the work of Chapter 3.

We provided theoretical results describing the effect of differencing has on the asymptotic behaviour of the raw wavelet periodogram. This work complements the work of Roueff and von Sachs (2011), who propose a class of locally stationary long memory processes. An interesting avenue for further research is to develop a similar modelling framework with LSW long memory processes, which one could think of as *integrated* LSW processes. An integrated LSW process of order n could be defined as a process whose n-th order difference is given by an LSW process. This would be akin to an ARIMA process whereby the differenced series is represented as an ARMA process. Within this framework, another exciting idea is to investigate the possibility of developing formal model selection procedures, which is an often neglected aspect of LSW modelling. Cardinali and Nason (2017) propose methodology for basis selection for LSW packet processes. This approach could be adapted into a model selection procedure that yields the order of the integrated LSW process, as well as the correct wavelet family and selection of basis functions for modelling the time series.

Finally, the work of Chapter 5 also lends itself to a number of future research questions. For example, it may be of interest to implement our methodology, for estimating the nonstationary second-order structure in the presence of mean changepoints, within the work of other authors. There are many algorithms and approaches for detecting mean changes in the presence of IID Gaussian noise – it would be interesting to investigate a way to incorporate our estimate of the nonstationary second-order structure within these more sophisticated methods. This may help overcome some of the shortcomings of our binary segmentation-based approach for detecting multiple changes. For example, the Narrowest-Over-Threshold method of Baranowski et al. (2019) is shown to be consistent in the presence of autocorrelated errors. However, this approach requires a pre-estimate of the autocorrelation of the time series, which could be achieved using our method employing the median-smoothed wavelet spectrum estimate. Similarly, a challenging research question would be how to incorporate our second-order estimator

within dynamic programming-based changepoint algorithms, such as the Pruned Exact Linear Time (PELT) method of Killick et al. (2012).

Appendix A

Appendix for 'Modelling Time-Varying First and Second-Order Structure of Time Series via Wavelets and Differencing'

A.1 Proofs of Results

In this section we provide proofs of all results stated in Chapter 3.

A.1.1 Proof of Proposition 3.3.1

The matrix D_J^1 is symmetric and positive semi-definite, since it can be expressed as a Gram matrix of vector inner products. The matrix D_J is invertible, since the family of vectors $\{\Psi_j(\tau)-\Psi_j(\tau-1)\}_{j=-J}^{-1}$ is linearly independent. This follows from observing that this family of vectors can be expressed as $\{\nabla\Psi_j(\tau)\}_{j=-J}^{-1}$, where ∇ is the invertible first-differencing matrix with diagonal entries 1, above the main diagonal entries equal to -1, and all other entries equal to zero.

A.1.2 Proof of Proposition 3.3.2

The expectation is given by

$$\begin{aligned}\mathbb{E}(\tilde{I}_k^j) &= \mathbb{E}\left[\left(\sum_t \left[\epsilon_t - \epsilon_{t-1} + \mathcal{O}(T^{-1})\right]\psi_{j,k-t}\right)^2\right] \\ &= \mathbb{E}\left[\left(\sum_t \epsilon_t \psi_{j,k-t}\right)^2\right] + \mathbb{E}\left[\left(\sum_t \epsilon_{t-1}\psi_{j,k-t}\right)^2\right] \\ &\quad - 2\mathbb{E}\left(\sum_s \sum_t \epsilon_t \epsilon_{s-1}\psi_{j,k-t}\psi_{j,k-s}\right) + \mathcal{O}(T^{-1}) \\ &:= I + II + III + \mathcal{O}(T^{-1}),\end{aligned}$$

where the remainder term can come out of the inner bracket since, for fixed j, the sum is finite as the wavelet is compactly supported. Now we evaluate each expectation individually. The term I is simply the expectation of the raw wavelet periodogram of the original LSW model. Hence,

$$I = \sum_l A_{jl} S_l\left(\frac{k}{T}\right) + \mathcal{O}(T^{-1}).$$

Next, observe that the term II is equal to $\mathbb{E}(I_{k-1}^j)$, since $d_{j,k-1} = \sum_s \epsilon_s \psi_{j,k-1-s}$, and by substituting $s = t-1$, we obtain $\sum_t \epsilon_{t-1}\psi_{j,k-t} = \sum_s \epsilon_s \psi_{j,k-1-s}$. Next, setting $u = k-1$, we obtain

$$\mathbb{E}(I_u^j) = \sum_l \sum_m w_{lm}^2 \left(\sum_t \psi_{j,k-t}\psi_{j,u-t}\right)^2.$$

Now, following a similar argument to Nason et al. (2000), we obtain

$$\begin{aligned}II &= \sum_l S_l\left(\frac{u+1}{T}\right)\sum_n \sum_t \sum_s \psi_{j,-s}\psi_{j,-t}\psi_{l,n-s}\psi_{l,n-t} + \mathcal{O}(T^{-1}) \\ &= \sum_l A_{jl} S_l\left(\frac{k}{T}\right) + \mathcal{O}(T^{-1}).\end{aligned}$$

Finally,

$$
\begin{aligned}
III &= -2\mathbb{E}\left(\sum_s \sum_t \epsilon_t \epsilon_{s-1} \psi_{j,k-t} \psi_{j,k-s}\right) \\
&= -2\sum_s \sum_t \psi_{j,k-t}\psi_{j,k-s}\mathbb{E}(\epsilon_t \epsilon_{s-1}) \\
&= -2\sum_s \sum_t \psi_{j,k-t}\psi_{j,k-s} \sum_l S_l\left(\frac{t+s-1}{2T}\right)\Psi_l(s-t-1) + \mathcal{O}(T^{-1}),
\end{aligned}
$$

since

$$
\mathbb{E}(\epsilon_t \epsilon_{s-1}) = c\left(\frac{t+s-1}{2T}, s-t-1\right) + \mathcal{O}(T^{-1}).
$$

The remainder of the proof is similar to that of the proof of Theorem 1 of Nason (2013). Let $u = k - t$ and $v = k - s$, substituting into the above to obtain

$$
\begin{aligned}
III &= -2\sum_l \sum_u \sum_v \psi_{j,u}\psi_{j,v} S_l\left(\frac{k}{T} - \frac{u+v-1}{2T}\right)\Psi_l(u-v-1) + \mathcal{O}(T^{-1}) \\
&= -2\sum_l \sum_u \sum_v \psi_{j,u}\psi_{j,v}\left[S_l\left(\frac{k}{T}\right) + \mathcal{O}\left(\frac{|u+v-1|}{2T}\right)\right]\Psi_l(u-v-1) + \mathcal{O}(T^{-1}).
\end{aligned}
$$

The remainder term in the above expression can be shown to be $\mathcal{O}(T^{-1})$ (A.2.1 in Nason (2013)). Hence, we are left with

$$
III = -2\sum_l S_l\left(\frac{k}{T}\right)\sum_u \sum_v \psi_{j,u}\psi_{j,v}\Psi_l(u-v-1) + \mathcal{O}(T^{-1}).
$$

Finally, substituting $r = u - v$, we obtain

$$
\begin{aligned}
C &= -2\sum_l S_l\left(\frac{k}{T}\right)\sum_u \sum_r \psi_{j,u}\psi_{j,u-r}\Psi_l(r-1) + \mathcal{O}(T^{-1}) \\
&= -2\sum_l S_l\left(\frac{k}{T}\right)\sum_r \Psi_j(r)\Psi_l(r-1) + \mathcal{O}(T^{-1}) \\
&= -2\sum_l A^1_{jl} S_l\left(\frac{k}{T}\right) + \mathcal{O}(T^{-1}).
\end{aligned}
$$

Hence, we obtain

$$\begin{aligned}\mathbb{E}(\tilde{I}_k^j) &= I + II + III \\ &= \sum_l A_{jl} S_l\left(\frac{k}{T}\right) + \sum_l A_{jl} S_l\left(\frac{k}{T}\right) - 2\sum_l A_{jl}^1 S_l\left(\frac{k}{T}\right) + \mathcal{O}(T^{-1}) \\ &= 2\sum_l \left(A_{jl} - A_{jl}^1\right) S_l\left(\frac{k}{T}\right) + \mathcal{O}(T^{-1}).\end{aligned}$$

For the variance part, we follow the same argument as in the proof of Proposition 4 of Nason et al. (2000). The wavelet coefficients of the differenced series are asymptotically Gaussian. Hence, the wavelet periodograms are asymptotically scaled χ^2-distributed, and hence the variance is asymptotically proportional to the expectation squared.

A.1.3 Proof of Theorem 3.3.3

(Proof for the Haar wavelet). The proof follows the same strategy to that of the proof of Theorem 2 in Nason et al. (2000). We show that there exists $\delta > 0$ such that $\lambda_{\min}(P) \geq \delta$, by using the following property from Toeplitz matrix theory. Let T be Toeplitz (and Hermitian) with elements $\{t_0, t_1, \ldots\}$. Let $f(z) = \sum_{n=-\infty}^{\infty} t_n z^n$ for $z \in \mathbb{C}$ be the symbol of the operator associated with T. If $\sum_n |t_n| < \infty$, then $f(z)$ is analytic in the open unit disc D in the complex plane and continuous in the closed unit disc $\Delta = D \cup S$, where S denotes the unit circle. The spectrum Λ of the (Laurent) operator T is $\Lambda(T) = f(S)$. If T is symmetric then an estimate of the smallest eigenvalue of T is

$$\min_{|z|=1}\{f(z)\} = \min_{|z|=1}\left\{t_0 + 2\mathrm{Re}\left(\sum_{n=1}^{\infty} t_n z^n\right)\right\}.$$

(Reichel and Trefethen (1992), Theorem 3.1 (i)). For ease of notation, indices will now run from 1 to ∞ instead of -1 to $-\infty$. Using straightforward algebra, we can derive explicit formulae for the entries of P, which are given by the following lemma.

Lemma A.1.1. In the case of the Haar wavelet, the elements of the matrix P are given by

$$P_{jj} = 10,$$

$$P_{j,j+m} = 6 \times 2^{-m/2}, \text{ for } l = j + m, m > 0.$$

Proof. In general, we can derive the discrete autocorrelation wavelets $\Psi_j(\tau)$ by discretising the continuous wavelet autocorrelation function $\Psi(\tau)$, using the relationship

$$\Psi_j(\tau) = \Psi\left(\frac{|\tau|}{2^j}\right).$$

For Haar wavelets, the continuous wavelet autocorrelation function is given by

$$\Psi(\tau) = \int_{-\infty}^{\infty} \psi_H(x)\psi_H(x-\tau)dx = \begin{cases} 1 - 3|\tau| & \text{for } |\tau| \in [0, 1/2], \\ |\tau| - 1 & \text{for } |\tau| \in (1/2, 1], \end{cases}$$

where $\psi_H(x)$ is the Haar mother wavelet. The discretisation formula holds for $\tau = -(2^j - 1), \ldots, 0, \ldots, (2^j - 1)$, and is equal to zero for all other values of τ. By Nason et al. (2000), we have that the elements of A are given by

$$A_{jj} = \frac{1}{3}2^j + \frac{5}{3}2^{-j}, \qquad A_{jl} = 2^{2j-l-1} + 2^{-l}, \quad l > j > 0.$$

We must therefore calculate the elements of A^1, from which we can obtain the elements of D^1, and hence P. Note that replacing the $(\tau - 1)$ terms with $(\tau + 1)$ results in an equivalent definition of the operators A^1 and D^1. When $l = j$, we have that

$$\begin{aligned} A^1_{jj} &= \sum_{\tau=-(2^j-1)}^{2^j-1} \Psi_j(\tau)\Psi_j(\tau+1) \\ &= 2\left[\sum_{\tau=1}^{2^{j-1}-1} \Psi_j(\tau)\Psi_j(\tau+1) + \sum_{\tau=2^{j-1}+1}^{2^j-1} \Psi_j(\tau)\Psi_j(\tau+1)\right] - 2^{-j-1} + \frac{1}{4} - 2^{-j-1} + \frac{1}{4} \end{aligned}$$

$$= 2\left[\sum_{\tau=1}^{2^{j-1}-1}\left(1-\frac{3\tau}{2^j}\right)\left(1-\frac{3(\tau+1)}{2^j}\right)+\sum_{\tau=2^{j-1}+1}^{2^j-1}\left(\frac{\tau}{2^j}-1\right)\left(\frac{\tau+1}{2^j}-1\right)\right]-2^{-j}+\frac{1}{2}$$

$$= \frac{1}{3}2^j-\frac{10}{3}2^{-j}.$$

Hence $D^1_{jj}=2A_{jj}-2A^1_{jj}=10\times 2^{-j}$, and therefore $P_{jj}=10$. When $j\neq l$, without loss of generality assume $l>j$. We have that

$$A^1_{jl}=\sum_{\tau=-(2^j-1)}^{2^j-1}\Psi_j(\tau)\Psi_l(\tau+1)=\sum_{\tau=-(2^j-1)}^{2^j-1}\Psi_j(\tau)\left(1-\frac{3|\tau+1|}{2^l}\right)$$

$$= 1-3\times 2^{-l}+\sum_{\tau=-(2^j-1)}^{-1}\Psi_j(\tau)\left(1-\frac{3|\tau+1|}{2^l}\right)+\sum_{\tau=1}^{2^j-1}\Psi_j(\tau)\left(1-\frac{3(\tau+1)}{2^l}\right),$$

since the value of $|\tau+1|$ is always less than or equal to $1/2$, and so we use the $1-3|\tau|$ part of the autocorrelation wavelet formula. The first sum is equal to

$$\sum_{\tau=-(2^j-1)}^{-2^{j-1}-1}\Psi_j(\tau)\left(1+\frac{3(\tau+1)}{2^l}\right)+\sum_{\tau=-2^{j-1}}^{-1}\Psi_j(\tau)\left(1+\frac{3(\tau+1)}{2^l}\right)$$

$$=\sum_{\tau=-(2^j-1)}^{-2^{j-1}-1}\left(-1-\frac{\tau}{2^j}\right)\left(1+\frac{3(\tau+1)}{2^l}\right)+\sum_{\tau=-2^{j-1}}^{-1}\left(1+\frac{3\tau}{2^j}\right)\left(1+\frac{3(\tau+1)}{2^l}\right)$$

$$=\left(-2^{j-l-2}-2^{j-3}+2^{-l-1}+\frac{1}{4}-2^{j-l-1}+2^{2j-l-2}\right)+\left(3\times 2^{j-l-2}+2^{j-3}-3\times 2^{-l-1}-\frac{3}{4}\right)$$

$$=2^{2j-l-2}-2^{-l}-\frac{1}{2}.$$

The second sum is given by

$$\sum_{\tau=1}^{2^{j-1}}\Psi_j(\tau)\left(1-\frac{3(\tau+1)}{2^l}\right)+\sum_{\tau=2^{j-1}+1}^{2^j-1}\Psi_j(\tau)\left(1-\frac{3(\tau+1)}{2^l}\right)$$

$$=\sum_{\tau=1}^{2^{j-1}}\left(1-\frac{3\tau}{2^j}\right)\left(1-\frac{3(\tau+1)}{2^l}\right)+\sum_{\tau=2^{j-1}+1}^{2^j-1}\left(\frac{\tau}{2^j}-1\right)\left(1-\frac{3(\tau+1)}{2^l}\right)$$

$$=\left(2^{j-3}+3\times 2^{-l}-\frac{3}{4}\right)+\left(2^{2j-l-2}-2^{j-3}-2^{-l}+\frac{1}{4}\right)$$

$$=2^{2j-l-2}+2^{1-l}-\frac{1}{2}.$$

From this, we finally obtain that

$$\begin{aligned} A^1_{jl} &= \left(2^{2j-l-2} - 2^{-l} - \frac{1}{2}\right) + \left(2^{2j-l-2} + 2^{1-l} - \frac{1}{2}\right) + \left(1 - 3 \times 2^{-l}\right) \\ &= 2^{2j-l-1} - 2^{1-l}. \end{aligned}$$

Therefore, we have that $D^1_{jl} = 2A_{jl} - 2A^1_{jl} = 6 \times 2^{-l}$, from which the required form for P follows. □

Returning to the main proof, we have that P is symmetric, however the formula above only refers to the upper triangular portion of the matrix P. Now, P is a Toeplitz matrix with $p_0 = 10$ and $p_m = 6 \times 2^{-m/2}$. We can now show that $\lambda_{\min}(P) \geq \delta > 0$. Substituting in the formula for the symbol of the symmetric Toeplitz P, we obtain

$$\begin{aligned} \min_{|z|=1}\{f(z)\} &= \min_{|z|=1}\left\{10 + 2\mathrm{Re}\left(\sum_{n=1}^{\infty} 6 \times 2^{-n/2} z^n\right)\right\} \\ &= 10 + 12 \min_{|z|=1}\left\{\mathrm{Re}\left(\frac{-\sqrt{2}z}{\sqrt{2}z - 2}\right)\right\} \\ &= 10 + 12 \min_{|z|=1}\left\{\frac{2\sqrt{2}\mathrm{Re}(z) - 2}{6 - 4\sqrt{2}\mathrm{Re}(z)}\right\}. \end{aligned}$$

This function is strictly monotonically increasing on $-1 \leq \mathrm{Re}(z) \leq 1$, therefore it follows that $\min_{|z|=1}\{f(z)\} = f(-1)$. Hence,

$$\lambda_{\min}(P) \geq f(-1) = 10 + \frac{12(-2\sqrt{2} - 2)}{6 + 4\sqrt{2}} = \frac{18 + 8\sqrt{2}}{3 + 2\sqrt{2}} > 0.$$

(Proof for the Shannon wavelet). Note that the indices now run over the negative integers. We can compute the entries of P using a variant of the Fourier domain formula shown in Equation (A.1.1) below, which is a consequence of Parseval's relation.

$$A_{jl} = \sum_{\tau} \Psi_j(\tau)\Psi_l(\tau) = \frac{1}{2\pi} \int \hat{\Psi}_j(\omega)\hat{\Psi}_l(\omega) d\omega, \tag{A.1.1}$$

where $\hat{\Psi}_j(\omega)$ denotes the Fourier transform of $\Psi_j(\tau)$, which is equal to the squared modulus of the Fourier transform of the non-decimated wavelet coefficients, $\hat{\psi}_j(\omega)$. Explicitly,

$$\hat{\Psi}_j(\omega) = \left|\hat{\psi}_j(\omega)\right|^2 = 2^{-j} \left|m_1\left(2^{-(j+1)}\omega\right)\right|^2 \prod_{l=0}^{-(j+2)} \left|m_0\left(2^l\omega\right)\right|^2, \tag{A.1.2}$$

where $m_0(\omega) = 2^{-1/2}\sum_k h_k \exp(-i\omega k)$ is the transfer function; $\{h_k\}$ is is the high-pass quadrature mirror filter with $\sum_k h_k^2 = 1$ and $\sum_k h_k = 2^{1/2}$; and $|m_1(\omega)|^2 = 1 - |m_0(\omega)|^2$.

The formula for the Fourier transform of the non-decimated wavelets given in Equation (A.1.2) and the corresponding formulae for $m_0(\omega)$ and $m_1(\omega)$ for the Shannon wavelet can be found using the Fourier transform of the continuous time mother and father wavelets, which can be found in Chui (1997), pages 46 and 64. Define the set C_j, for $j < 0$, to be

$$C_j = \left[-\frac{\pi}{2^{-j-1}}, -\frac{\pi}{2^{-j}}\right] \cup \left[\frac{\pi}{2^{-j}}, \frac{\pi}{2^{-j-1}}\right].$$

As in Nason et al. (2000), the Fourier transform of the non-decimated Shannon wavelets is given by

$$\hat{\psi}_j(\omega) = -2^{-j/2}\exp(-2^{-j-1}i\omega)\mathbb{I}_{C_j}(\omega),$$

where $\mathbb{I}_{C_j}$ is the indicator function on the set C_j. From this the Fourier transform of the autocorrelation wavelets can be obtained as

$$\hat{\Psi}_j(\omega) = \left|\hat{\psi}_j(\omega)\right|^2 = 2^{-j}\mathbb{I}_{C_j}(\omega).$$

Using Parseval's relation and the shifting property of the Fourier transform, we

have that

$$D^1_{jl} = 2\sum_{\tau} \Psi_j(\tau)(\Psi_l(\tau) - \Psi_l(\tau+1)) = \frac{1}{2\pi}\int \hat{\Psi}_j(\omega)\hat{\Psi}_l(\omega)d\omega - \frac{1}{2\pi}\int e^{-i\omega}\hat{\Psi}_j(\omega)\hat{\Psi}_l(\omega)d\omega.$$

The first term in the sum is exactly the entries of the original A-matrix which are given by $A_{jj} = 2^{-j}$ for $j < 0$, $A_{jl} = 0$ for $j \neq l$. When $j \neq l$, the second term is equal to zero since the supports of different $\hat{\Psi}_j(\omega)$ do not overlap. Thus, the matrix is diagonal, and when $j = l$, the second term is given by

$$\begin{aligned}
-\sum_{\tau} \Psi_j(\tau)\Psi_j(\tau+1) &= -\frac{1}{2\pi}\int e^{-i\omega}\hat{\Psi}_j(\omega)^2 d\omega \\
&= -\frac{2^{-2j-1}}{\pi}\int e^{-i\omega}\mathbb{I}_{C_j}(\omega)d\omega \\
&= -\frac{2^{-2j-1}}{\pi}\left(\int_{-2^{j+1}\pi}^{-2^j\pi} e^{-i\omega}d\omega + \int_{2^j\pi}^{2^{j+1}\pi} e^{-i\omega}d\omega\right) \\
&= -\frac{2^{-2j-1}}{\pi}\left(\left[ie^{-i\omega}\right]_{-2^{j+1}\pi}^{-2^j\pi} + \left[ie^{-i\omega}\right]_{2^j1\pi}^{2^{j+1}\pi}\right) \\
&= -\frac{2^{-2j-1}i}{\pi}\left(\exp(2^j\pi i) - \exp(2^{j+1}\pi i) + \exp(-2^{j+1}\pi i) - \exp(-2^j\pi i)\right).
\end{aligned}$$

Now, expanding the complex exponential terms into trigonometric functions, we obtain

$$\begin{aligned}
-\sum_{\tau} \Psi_j(\tau)\Psi_j(\tau+1) = -\frac{2^{-2j-1}i}{\pi}\big\{&\cos(2^j\pi) + i\sin(2^j\pi) - \cos(2^{j+1}\pi) - i\sin(2^{j+1}\pi) \\
&+ \cos(-2^{j+1}\pi) + i\sin(-2^{j+1}\pi) - \cos(-2^j\pi) - i\sin(-2^j\pi)\big\}.
\end{aligned}$$

After much simplification in which the cosine terms will vanish, we obtain

$$-\sum_{\tau} \Psi_j(\tau)\Psi_j(\tau+1) = \frac{2^{-2j}}{\pi}\left(\sin(2^j\pi) - \sin(2^{j+1}\pi)\right).$$

Hence, the diagonal entries of the matrix D^1 are given by

$$D^1_{jj} = 2^{-j+1} + \frac{2^{-2j+1}}{\pi}\left(\sin(2^j\pi) - \sin(2^{j+1}\pi)\right).$$

Hence, $P_{jj} = 2^{-2j+1} + 2^{-3j+1}\left[\sin(2^j\pi) - \sin(2^{j+1}\pi)\right]/\pi$, while the off-diagonal terms are zero. Next, approximate the diagonal terms using a Maclaurin series:

$$\begin{aligned} P_{jj} &\approx 2^{-2j+1} + \frac{2^{-3j+1}}{\pi}\left(2^j\pi - \frac{(2^j\pi)^3}{6} - 2^{j+1}\pi + \frac{(2^{j+1}\pi)^3}{6}\right) \\ &= 2^{-2j+1} + 2^{-2j+1} - 2^{-2j+2} - \frac{\pi^2}{3} + \frac{8\pi^2}{3} \\ &= \frac{7\pi^2}{3}. \end{aligned}$$

Therefore, the matrix is diagonal, with all diagonal entries being uniformly bounded away from 0, with $P_{jj} \geq P_{-1,-1} \approx 13.09$ for all j. Hence, we have shown that there exists some $\delta > 0$ such that $\lambda_{\min}(P) \geq \delta$.

A.1.4 Proof of Theorem 3.3.4

As $T \to \infty$, the eigenvalues of D^1 tend to zero and hence, when viewed as an operator acting on the sequence space $\ell^2(\mathbb{N})$, its inverse is unbounded. In the locally stationary Fourier time series setting, a similar relationship is found, as given in Equation (5) of Roueff and von Sachs (2011). Loosely speaking, the original spectrum at frequency ω is related to the differenced one through multiplication of the term $|1 - e^{-i\omega}|^{-2}$. As $\omega \to 0$ (corresponding to low frequencies) the equation blows up. This mirrors our scenario, where, as the correction matrix grows in size – and we consider coarse-scale (low frequency) behaviour – the inverse matrix norm becomes larger.

We can account for the unboundedness of the inverse of D^1 by using a rescaling of the LSW process itself. Consider the auxiliary process $\epsilon_t = \sum_{j,k} \tilde{w}_{jk}\tilde{\psi}_{j,k-t}\xi_{lm}$, where $\tilde{w}_{jk} = 2^{j/4}w_{jk}$ and $\tilde{\psi}_{j,k-t} = 2^{-j/4}\psi_{j,k-t}$. Then, the expectation of the raw wavelet periodogram (with respect to the rescaled wavelet) is given by

$$\mathbb{E}(\tilde{I}_k^j) = \sum_l P_{jl}\tilde{S}_l(k/T) + \mathcal{O}(T^{-1}), \tag{A.1.3}$$

where $\tilde{S}_j(k/T) = 2^{j/2}S_j(k/T)$ and $P_{jl} = 2^{-j/2}D^1_{jl}2^{-l/2}$. We can therefore use

the (bounded) P-inverse matrix to correct the smoothed, rescaled periodogram, and then multiply by $2^{-j/2}$, since $S_j(k/T) = 2^{-j/2}\tilde{S}_j(k/T)$. To determine the appropriate threshold for the wavelet-based estimator, we use the following lemma:

Lemma A.1.2. For a Gaussian trend LSW process and using a wavelet ψ' of bounded variation, the wavelet coefficients $\hat{v}^j_{rs}$ of the rescaled wavelet periodogram $\tilde{I}^j_k$, with $2^r = o(T)$, obey uniformly in s,

$$\mathbb{E}(\hat{v}^j_{rs}) - \int_0^1 \sum_n P_{jn}\tilde{S}_n(z)\psi'_{rs}(z)dz = \mathcal{O}\left(2^{r/2}T^{-1}\right),$$

and

$$\mathrm{Var}(\hat{v}^j_{rs}) = 2T^{-1}\int_0^1 \left(\sum_n P_{jn}\tilde{S}_n(z)\right)^2 \psi'^2_{rs}(z)dz + \mathcal{O}\left(2^r T^{-2}\right).$$

Lemma A.1.2 is analogous to Theorem 3 of Nason et al. (2000). The result of mean square consistency follows due to a combination of Equation (A.1.3) and Theorem 4 of Nason et al. (2000). The mean squared error of the smoothed, corrected wavelet periodogram is given by

$$\mathbb{E}\left[\int_0^1 \left(\widehat{S}_j(z) - S_j(z)\right)^2 dz\right] = 2^{-j}\mathbb{E}\left[\int_0^1 \left(\hat{\tilde{S}}_j(z) - \tilde{S}_j(z)\right)^2 dz\right]$$
$$\leq 2^{-j+1}\mathbb{E}\left[\int_0^1 \left(\sum_{l=-J}^{-1}\left(\hat{I}^l_{\lfloor zT\rfloor} - \Lambda_l(z)\right)P^{-1}_{jl}\right)^2 dz\right] + 2^{-j+1}\int_0^1 \left(\sum_{l<-J}\Lambda_l(z)P^{-1}_{jl}\right)^2 dz,$$

where $\hat{I}^l_{\lfloor zT\rfloor}$ is the smoothed estimate of the rescaled raw wavelet periodogram and $\Lambda_l(z) = \sum_n P_{nl}\tilde{S}_n(z)$. The first term can be bounded as

$$I \leq 2^{-j+1}\left(\sum_{l=-J}^{-1} P^{-1}_{jl}\left(\mathbb{E}\left[\int_0^1 \left(\hat{I}^l_{\lfloor zT\rfloor} - \Lambda_l(z)\right)^2 dz\right]\right)^{1/2}\right)^2$$
$$\leq 2^{-j+1}\left(\sum_{l=-J}^{-1} P^{-1}_{jl}\mathcal{O}\left(T^{-2/3}\log^2(T)\right)^{1/2}\right)^2 = \mathcal{O}\left(2^{-j}T^{-2/3}\log^2(T)\right),$$

which follows since P possesses a bounded inverse with exponentially decaying entries, and using the rate of convergence of the mean squared error of the wavelet

thresholding estimator derived in Neumann and von Sachs (1995), Theorem 3.1 A. The second term is asymptotically dominated by the first, since it can be bounded as

$$II \leq 2^{-j+1} \left(\sum_{l<-J} \Lambda_l(z) P_{jl}^{-1} \right)^2 \leq 2^{-j+1} \left(\sum_{l<-J} P_{jl}^{-1} \sum_{n=-\infty}^{-1} 2^{-l/2} D_{ln}^1 S_n(z) \right)^2$$
$$\leq 2^{-j+1} \left(\sum_{l<-J} \mathcal{O}(2^{l/2}) \sum_{n=-\infty}^{-1} S_n(z) \right)^2 = \mathcal{O}(2^{-j} \times 2^{-J}) = \mathcal{O}(2^{-j} T^{-1}),$$

which follows since P_{jl}^{-1} is bounded, $D_{ln}^1 = \mathcal{O}(2^l)$, $\sum_n S_n(z) < \infty$, and $T = 2^J$. Hence, the mean squared error is given by $\mathcal{O}\left(2^{-j} T^{-2/3} \log^2(T)\right)$.

A.1.5 Proof of Proposition 3.3.5

We can write the mean squared error as

$$\mathbb{E}\left[\int_0^1 \left(\hat{c}(z,\tau) - c(z,\tau) \right)^2 dz \right] \leq 2\mathbb{E}\left[\int_0^1 \left(\sum_{j=-J_0}^{-1} \left(\widehat{S}_j(z) - S_j(z) \right) \Psi_j(\tau) \right)^2 dz \right] + 2R_{J_0},$$

where R_{J_0} can be bounded as

$$R_{J_0} = \left(\sum_{j<-J_0} S_j(z) \Psi_j(\tau) \right)^2 \leq \left(\sum_{j<-J_0} S_j(z) \right)^2$$
$$\leq \left(\sum_{j<-J_0} \mathcal{O}(2^{\gamma j}) \right)^2 = \mathcal{O}(2^{-2\gamma J_0}) = \mathcal{O}(T^{-2\gamma\alpha}) = \mathcal{O}(T^{\alpha - 2/3}).$$

For the first term, we obtain

$$\mathbb{E}\left[\int_0^1 \left(\sum_{j=-J_0}^{-1} \left(\widehat{S}_j(z) - S_j(z) \right) \Psi_j(\tau) \right)^2 dz \right]$$
$$\leq \left(\sum_{j=-J_0}^{-1} \Psi_j(\tau) \left(\mathbb{E}\left[\int_0^1 \left(\widehat{S}_j(z) - S_j(z) \right)^2 dz \right] \right)^{1/2} \right)^2$$
$$\leq \left(\sum_{j=-J_0}^{-1} \mathcal{O}\left(2^{-j} T^{-2/3} \log^2(T) \right)^{1/2} \right)^2$$

$$= \mathcal{O}\left(\left(T^{\alpha-2/3}\log^2(T)\right)^{1/2}\right)^2 = \mathcal{O}\left(T^{\alpha-2/3}\log^2(T)\right),$$

by using the mean squared error of the EWS estimator, and using that $\Psi_j(\tau) \leq 1$ for all j and τ. Hence, provided that $T^{\alpha-2/3}\log^2(T) \to 0$ as $T \to \infty$, the estimator is mean square consistent.

A.1.6 Proof of Proposition 3.3.6

In order to derive the formula for the expectation of the squared wavelet coefficients of a general n-th difference, we require the formula the n-th difference itself, which is a well-known result. Denote by $\nabla^n X_t$ the n-th difference of the time series $\{X_t\}$ at time t. Then,

$$\nabla^n X_t = \sum_{k=0}^{n}(-1)^k\binom{n}{k}X_{t-k} = \sum_{k=0}^{n}(-1)^k\binom{n}{k}\epsilon_{t-k} + \mathcal{O}(T^{-1}), \tag{A.1.4}$$

which follows from the differentiability assumption of the trend μ. Now, observe that, for an n-th difference, the expectation will involve the sum of the spectrum over all scales, multiplied by a linear combination of lagged inner product A-matrix entries, denoted A^τ, from lag 0 to lag n. To calculate the coefficient in front of each of the A^τ, we simply calculate the sum of the coefficients of the $\{\epsilon_t\epsilon_s\}$ for each particular lag in the expansion obtained by squaring the n-th difference of the time series, for which we can use Equation (A.1.4).

For example, to calculate the coefficient of A, we add together the coefficients of the squared terms in the square of the differenced series, i.e. add the coefficients of $\epsilon_t^2, \epsilon_{t-1}^2, \ldots \epsilon_{t-n}^2$. For the coefficient of A^1, we add together the coefficients of the terms in the square of the difference that differ in index by 1, i.e. we add the coefficients of $\epsilon_t\epsilon_{t-1}, \epsilon_{t-1}\epsilon_{t-2}, \epsilon_{t-1}\epsilon_t, \ldots, \epsilon_{n-1}\epsilon_n$. In particular, if we were interested in the third difference, then the coefficient in front of A would be $1+9+9+1$, and the coefficient in front of A^1 would be $-3-3-9-9-3-3 = -30$, and so

on. Hence the formula for the coefficient in front of A is given by

$$\sum_{r=0}^{n} \binom{n}{r}^2 = \binom{2n}{n},$$

and similarly, the formula for the coefficient in front of A^τ, for $\tau \geq 1$, is given by

$$2(-1)^\tau \sum_{r=0}^{n-\tau} \binom{n}{r}\binom{n}{r+\tau} = 2(-1)^\tau \binom{2n}{n+\tau},$$

where the multiplication by 2 arises due to symmetry (for example we must add both the coefficients of $\epsilon_{t-1}\epsilon_{t-2}$ and $\epsilon_{t-2}\epsilon_{t-1}$). The equality on the right follows from a counting argument. The number of ways to choose $n+\tau$ objects from $2n$ choices is the same as the number of ways of choosing r objects from the first n and choosing $r+\tau$ from the remaining n objects, for $0 \leq r \leq n-\tau$. The expectation is accurate up to order $\mathcal{O}(T^{-1})$ by the same argument as in the proof of Proposition 3.3.2. Hence, the squared expectation of the wavelet coefficients of the n-th differenced series are given by

$$\mathbb{E}(\tilde{I}_k^j) = \sum_l S_l\left(\frac{k}{T}\right)\left[\binom{2n}{n}A_{jl} + 2\sum_{\tau=1}^{n}(-1)^\tau \binom{2n}{n+\tau}A_{jl}^\tau\right] + \mathcal{O}(T^{-1}).$$

A.1.7 Proof of Theorem 3.3.8

By the Daubechies characterisation of Hölder spaces, rescaling of the LSW process, and Proposition 3.3.2, the expectation of the rescaled raw wavelet periodogram of the differenced time series is given by

$$\mathbb{E}(\tilde{I}_k^j) = \sum_l P_{jl}\tilde{S}_l(k/T) + \mathcal{O}(2^{-7j/2}T^{-2}) + \mathcal{O}(T^{-1}). \tag{A.1.5}$$

Hence, the mean squared error of the smoothed wavelet periodogram is given by

$$\mathbb{E}\left[\int_0^1 \left(\widehat{S}_j(z) - S_j(z)\right)^2 dz\right] = 2^{-j}\mathbb{E}\left[\int_0^1 \left(\hat{\tilde{S}}_j(z) - \tilde{S}_j(z)\right)^2 dz\right]$$

$$\leq 2^{-j+1}\mathbb{E}\left[\int_0^1\left(\sum_{l=-J_1}^{-1}\left(\hat{I}^l_{\lfloor zT\rfloor}-\Lambda_l(z)\right)P^{-1}_{jl}\right)^2dz\right]+2^{-j+1}\int_0^1\left(\sum_{l<-J_1}\Lambda_l(z)P^{-1}_{jl}\right)^2dz$$

where $\hat{I}^l_{\lfloor zT\rfloor}$ is the smoothed estimate of the raw wavelet periodogram and $\Lambda_l(z)=\sum_n P_{nl}\tilde{S}_n(z)$. The first term can be bounded as

$$\begin{aligned}
I &\leq 2^{-j+1}\left(\sum_{l=-J_1}^{-1}P^{-1}_{jl}\left(\mathbb{E}\left[\int_0^1\left(\hat{I}^l_{\lfloor zT\rfloor}-\Lambda_l(z)\right)^2dz\right]\right)^{1/2}\right)^2\\
&\leq 2^{-j+1}\left(\sum_{l=-J_1}^{-1}P^{-1}_{jl}\left(\mathcal{O}\left(2^{-7l}T^{-4}\right)+\mathcal{O}\left(T^{-2/3}\log^2(T)\right)\right)^{1/2}\right)^2\\
&= \mathcal{O}\left(2^{-j}T^{7\beta-4}\right)+\mathcal{O}\left(2^{-j}T^{-2/3}\log^2(T)\right),
\end{aligned}$$

which follows since P possesses a bounded inverse with exponentially decaying entries, and using Equation (A.1.5). The second term can be bounded in the same fashion as in the proof of Theorem 3.3.4, and is of order $\mathcal{O}(2^{-j}T^{-\beta})$, which gives the stated consistency result. Similarly, the mean squared error of the LACV estimator is calculated as

$$\mathbb{E}\left[\int_0^1\left(\hat{c}(z,\tau)-c(z,\tau)\right)^2dz\right]\leq 2\mathbb{E}\left[\int_0^1\left(\sum_{j=-J_0}^{-1}\left(\widehat{S}_j(z)-S_j(z)\right)\Psi_j(\tau)\right)^2dz\right]+2R_{J_0},$$

where R_{J_0} is asymptotically negligible by the argument in Proposition 3.3.5. For the first term, we obtain

$$\begin{aligned}
&\mathbb{E}\left[\int_0^1\left(\sum_{j=-J_0}^{-1}\left(\hat{S}_j(z)-S_j(z)\right)\Psi_j(\tau)\right)^2dz\right]\\
&\leq\left(\sum_{j=-J_0}^{-1}2^{-j/2}\left(\mathcal{O}\left(T^{7\beta-4}\right)+\mathcal{O}\left(T^{-2/3}\log^2(T)\right)+\mathcal{O}(T^{-\beta})\right)^{1/2}\right)^2\\
&=\left(\mathcal{O}\left(T^{7\beta-4}\right)+\mathcal{O}\left(T^{-2/3}\log^2(T)\right)+\mathcal{O}(T^{-\beta})\right)\left(\sum_{j=-J_0}^{-1}2^{-j/2}\right)^2\\
&=\mathcal{O}\left(T^{\alpha+7\beta-4}\right)+\mathcal{O}\left(T^{\alpha-2/3}\log^2(T)\right)+\mathcal{O}(T^{\alpha-\beta}).
\end{aligned}$$

A.1.8 Proof of Proposition 3.4.1

The appropriate threshold is derived by using an analogous result to Lemma A.1.2, from which we obtain

$$\mathbb{E}(\hat{v}_{rs}) - \int_0^1 \mu(z)\psi'_{rs}(z)dz = \mathcal{O}\left(2^{r/2}T^{-1}\right),$$

and

$$\mathrm{Var}(\hat{v}_{rs}) = T^{-1}\int_0^1 \sum_n S_n(z)\psi'^2_{rs}(z)dz + \mathcal{O}\left(2^r T^{-2}\right).$$

The mean squared error rate is obtained by using Theorem 1 of von Sachs and MacGibbon (2000), with the specific case of a Lipschitz continuous trend.

Appendix B

Appendix for 'Trend Locally Stationary Wavelet Processes with Applications to Environmental Data'

B.1 Proofs of Results

In this appendix, we provide proofs of the results described in Chapter 4.

B.1.1 Proof of Lemma 4.3.1

For simplicity, let

$$\tilde{d}_{j,k} = \sum_t \psi_{j,k-t} \sum_l \sum_m w_{l,m;T} \psi_{l,m-t} \xi_{lm},$$

i.e. $\tilde{d}_{j,k}$ is the non-decimated wavelet transform of the LSW process part of the model. Then,

$$d_{j,k} = \sum_t \mu\left(\frac{t}{T}\right) \psi_{j,k-t} + \tilde{d}_{j,k} = \tilde{d}_{j,k}.$$

The last equality holds due to the vanishing moments property of the wavelet. Hence,

$$\mathbb{E}\left(I^j_{k,T}\right) = \mathbb{E}(d^2_{j,k}) = \mathbb{E}(\tilde{d}^2_{j,k}) = \sum_l A_{jl} S_l(k/T) + \mathcal{O}(T^{-1}),$$

where the last part follows from Proposition 4 of Nason et al. (2000). The proof for the form of the variance is also obtained using Proposition 4.

B.1.2 Proof of Proposition 4.3.2

The proof of the proposition takes a similar approach to Cardinali and Nason (2017), but our approach crucially enables us to bound the entries of the A-inverse operator, which can be used when bounding the error of LSW process-based estimators. We follow the strategy of the proof of Theorem 1 of Cardinali and Nason (2017), who show that the autocorrelation wavelet packet inner product has bounded inverse. Firstly, we show that B is positive definite. Recall that the indices j run from -1 onwards. We can write

$$B = DAD^\mathsf{T} = D\Psi\Psi^\mathsf{T}D^\mathsf{T} = (D\Psi)(D\Psi)^\mathsf{T},$$

where Ψ is constructed from the autocorrelation wavelets, and D is a diagonal matrix with entries given by $D_{jj} = 2^{j/2}$. Hence, for any $x \in \ell^2(\mathbb{N})$,

$$x^\mathsf{T}Bx = x^\mathsf{T}(D\Psi)(D\Psi)^\mathsf{T}x = ((D\Psi)^\mathsf{T}x)^\mathsf{T}((D\Psi)^\mathsf{T}x) := y^\mathsf{T}y \geq 0.$$

Finally, the family $\{2^{j/2}\Psi_j(\tau)\}_{j=-\infty}^{-1}$, which induces the Gram matrix B, is linearly independent since $\{\Psi_j(\tau)\}_{j=-\infty}^{-1}$ is. If $x^TBx = 0$ for some x, then

$$\sum_\tau \left[\sum_j x_j 2^{j/2}\Psi_j(\tau)\right]\left[\sum_l x_l 2^{l/2}\Psi_l(\tau)\right] = 0.$$

Therefore,

$$\sum_\tau \left[\sum_j x_j 2^{j/2}\Psi_j(\tau)\right]^2 = 0,$$

and hence

$$\sum_j x_j 2^{j/2} \Psi_j(\tau) = \sum_j \tilde{x}_j \Psi_j(\tau) = 0,$$

for all τ, where $\tilde{x}_j = 2^{j/2} x_j$. Hence, $\tilde{x}_j = 0$ for all j since $\{\Psi_j(\tau)\}_{j=-\infty}^{-1}$ is linearly independent. Therefore, if $x^\mathsf{T} B x = 0$, then $x_j = 0$ for all j. Thus B is positive definite. Next, we require the following definition.

Definition B.1.1. A matrix $A = (A_{ij})_{i,j \in \mathcal{J}}$ *decays exponentially* if there exist constants $c > 0$ and $0 < \lambda < 1$ such that $|A_{ij}| \leq c\lambda^{|i-j|}$ for all $i, j \in \mathcal{J}$.

In our situation the set $\mathcal{J}$ is equal to $\mathbb{N}$. Observe that B has exponential decay on the off-diagonal entries, because

$$|B_{jl}| = |2^{j/2} A_{jl} 2^{l/2}| \leq |A_{jl}|,$$

and since the off-diagonal entries of A are shown to be exponentially decaying in Cardinali and Nason (2017). Now, for the diagonal entries $j = l$, we have that

$$|B_{jj}| = 2^j |A_{jj}| = 2^j \left| \sum_\tau \Psi_j(\tau)^2 \right| \leq 2^j \sum_\tau 1,$$

since $|\Psi_j(\tau)| \leq 1$ for all τ. The autocorrelation wavelets are compactly supported, with the length of the support given by $2L_j - 1$, where $L_j = (2^{-j} - 1)(N_h - 1) + 1$, where N_h is the length of the high-pass quadrature mirror filter associated to the given wavelet. Hence, $L_j \leq K 2^{-j}$ for some positive constant K which will depend on the choice of wavelet. Therefore,

$$|B_{jj}| \leq 2^j \sum_\tau 1 \leq 2^j K 2^{-j} = K.$$

Hence, B is exponentially decaying. By Lemma 2.1 of Goodman et al. (1995), which states that exponentially decaying matrices are bounded, B is a bounded operator. We next utilise Theorem A from Goodman et al. (1995), which states that any bounded, positive definite matrix B possesses a unique Cholesky fac-

torisation $B = LL^{\mathsf{T}}$, where L is lower triangular with positive diagonal entries, is invertible, and has inverse L^{-1} which is also lower triangular. We can use Theorem A for the operator B, since we have shown above that the operator B is bounded and positive definite. This is a key difference to the proof in Cardinali and Nason (2017), and why we consider B, since the operator A is not bounded as the diagonal entries diverge.

Furthermore, from Corollary 1 of Krishtal et al. (2015), we have that the Cholesky factor L of the exponentially decaying B is itself exponentially decaying. Next, we note that L^{-1} is also exponentially decaying, by well-known results on matrix inverse localisation, see for example Theorem 1 of Baskakov (1990) or Proposition 2 of Jaffard (1990). Note that the rate of exponential decay of the inverse is not necessarily the same as for the original matrix.

Lemma 2.2 of Goodman et al. (1995) states that the product of exponentially decaying matrices is also exponentially decaying, and hence we have that $B^{-1} = (L^{-1})^{\mathsf{T}}L^{-1}$ decays exponentially. Finally, we again use Lemma 2.1 of Goodman et al. (1995), and hence we have that B^{-1} is bounded in $\ell^2(\mathbb{N})$.

B.1.3 Proof of Corollary 4.3.3

This follows from Proposition 4.3.2 which shows that B has bounded inverse, and the fact that $A_{jl}^{-1} = 2^{j/2}B_{jl}^{-1}2^{l/2}$.

B.1.4 Proof of Proposition 4.3.4

To determine the appropriate threshold in the estimator we use the following corollary:

Corollary B.1.1. For a polynomial trend Gaussian LSW process and using a wavelet $\tilde{\psi}$ of bounded variation, the wavelet coefficients $\hat{v}_{rs}^{j}$ of $I_{\lfloor z \rfloor T \rfloor}^{j}$, with $2^r =$

$o(T)$, obey uniformly in s,

$$\mathbb{E}(\hat{v}_{rs}^j) - \int_0^1 \sum_n A_{jn} S_n(z) \tilde{\psi}_{rs}(z) dz = \mathcal{O}\left(2^{r/2} T^{-1}\right), \tag{B.1.1}$$

and

$$\mathrm{Var}(\hat{v}_{rs}^j) = 2T^{-1} \int_0^1 \left(\sum_n A_{jn} S_n(z) \right)^2 \tilde{\psi}_{rs}^2(z) dz + \mathcal{O}\left(2^r T^{-2}\right). \tag{B.1.2}$$

Now, the mean squared error of the smoothed, corrected wavelet periodogram is given by

$$\mathbb{E}\left[\int_0^1 \left(\hat{S}_j(z) - S_j(z) \right)^2 dz \right] \leq 2\mathbb{E}\left[\int_0^1 \left(\sum_{l=-J}^{-1} \left(\hat{I}^l_{\lfloor zT \rfloor} - \beta_l(z) \right) A_{jl}^{-1} \right)^2 dz \right] + 2R_J,$$

where $\hat{I}^l_{\lfloor zT \rfloor}$ is the smoothed estimate of the raw wavelet periodogram and $\beta_l(z) = \sum_n A_{nl} S_n(z)$. The remainder R_J can be bounded as

$$\begin{aligned} R_J &= \left(\sum_{l<-J} \beta_l(z) A_{jl}^{-1} \right)^2 \leq \left(\sum_{l<-J} A_{jl}^{-1} \sum_{n=-\infty}^{-1} A_{nl} S_n(z) \right)^2 \\ &\leq D^2 \left(\sum_{l<-J} A_{jl}^{-1} \sum_{n=-\infty}^{-1} A_{nl} 2^{5n/6} \right)^2, \end{aligned}$$

since $S_j(z) \leq D 2^{5j/6}$ by assumption. This can be further bounded as

$$\begin{aligned} R_J &\leq D^2 \left(\sum_{l<-J} A_{jl}^{-1} \left[\sum_{n<l} A_{nl} 2^{5n/6} + A_{ll} 2^{5l/6} \sum_{n=l+1}^{-1} A_{nl} 2^{5n/6} \right] \right)^2 \\ &\leq D^2 \left(\sum_{l<-J} A_{jl}^{-1} \left[\sum_{n<l} \mathcal{O}(2^{n-l}) 2^{5n/6} + \mathcal{O}(2^{-l}) 2^{5l/6} \sum_{n=l+1}^{-1} \mathcal{O}(2^{l-n}) 2^{5n/6} \right] \right)^2 \\ &\leq D^2 \left(\sum_{l<-J} A_{jl}^{-1} \left[\sum_{n<l} \mathcal{O}(2^{11n/6-l}) + \mathcal{O}(2^{-l/6}) + \sum_{n=l+1}^{-1} \mathcal{O}(2^{l-n/6}) \right] \right)^2 \\ &\leq \left(\sum_{l<-J} A_{jl}^{-1} \mathcal{O}(2^{-l/6}) \right)^2 = 2^j \left(\sum_{l<-J} \mathcal{O}(2^{l/3}) \right)^2 = 2^j \mathcal{O}(2^{-2J/3}) = 2^j \mathcal{O}(T^{-2/3}). \end{aligned}$$

We have used that the off-diagonal entries of A_{jl} are of order $\mathcal{O}(2^{-|j-l|})$, while $A_{jj} = \mathcal{O}(2^{-j})$. Further, we have used that $A_{jl}^{-1} = 2^{j/2}B_{jl}^{-1}2^{l/2}$, where B^{-1} is bounded by Proposition 4.3.2. The first term can be bounded as

$$\begin{aligned} I &\leq 2\left(\sum_{l=-J}^{-1} A_{jl}^{-1}\left(\mathbb{E}\left[\int_0^1 \left(\hat{I}^l_{\lfloor zT\rfloor} - \beta_l(z)\right)^2 dz\right]\right)^{1/2}\right)^2 \\ &\leq 2\left(\sum_{l=-J}^{-1} 2^{j/2}B_{jl}^{-1}2^{l/2}\mathcal{O}\left(T^{-2/3}\log^2(T)\right)^{1/2}\right)^2 = 2^j\mathcal{O}\left(T^{-2/3}\log^2(T)\right), \end{aligned}$$

by a similar argument, and using the rate of convergence of the mean squared error of the wavelet thresholding estimator in Nason et al. (2000). Hence, the stated mean square consistency result follows.

B.1.5 Proof of Proposition 4.3.5

We can write the mean squared error as

$$\mathbb{E}\left[\int_0^1 \left(\hat{c}(z,\tau) - c(z,\tau)\right)^2 dz\right] \leq 2\mathbb{E}\left[\int_0^1 \left(\sum_{j=-J}^{-1}\left(\hat{S}_j(z) - S_j(z)\right)\Psi_j(\tau)\right)^2 dz\right] + 2R_J,$$

where R_J can be bounded as

$$\begin{aligned} R_J &= \left(\sum_{j<-J} S_j(z)\Psi_j(\tau)\right)^2 \leq \left(\sum_{j<-J} S_j(z)\right)^2 \\ &\leq \left(\sum_{j<-J} \mathcal{O}(2^{5j/6})\right)^2 = \mathcal{O}(2^{-5J_0/3}) = \mathcal{O}(T^{-5/3}). \end{aligned}$$

The first term can be bounded as

$$\begin{aligned} &\mathbb{E}\left[\int_0^1 \left(\sum_{j=-J}^{-1}\left(\hat{S}_j(z) - S_j(z)\right)\Psi_j(\tau)\right)^2 dz\right] \\ &\leq \left(\sum_{j=-J}^{-1}\Psi_j(\tau)\left(\mathbb{E}\left[\int_0^1 \left(\hat{S}_j(z) - S_j(z)\right)^2 dz\right]\right)^{1/2}\right)^2 \end{aligned}$$

$$\leq \left(\sum_{j=-J}^{-1} 2^j \mathcal{O}\left(T^{-2/3}\log^2(T)\right)^{1/2} \right)^2 = \mathcal{O}\left(T^{-2/3}\log^2(T)\right),$$

which proves the result.

B.1.6 Proof of Lemma 4.3.6

The proof is the analogous result of Proposition 4 of Nason et al. (2000), where in that proof $\psi^0 = \psi^1$. The wavelet coefficients are given by

$$\begin{aligned} d_{j,k} &= \sum_t \mu\left(\frac{t}{T}\right)\psi^1_{j,k-t} + \sum_t \sum_l \sum_m w_{lm}\psi^0_{l,m-t}\xi_{lm}\psi^1_{j,k-t} \\ &= \sum_t \sum_l \sum_m w_{lm}\psi^0_{l,m-t}\xi_{lm}\psi^1_{j,k-t}, \end{aligned}$$

where the last equality holds due to the vanishing moments property of the wavelet. Then, it follows from Proposition 3.1 of Gott and Eckley (2013) that

$$\mathbb{E}(d^2_{j,k}) = \sum_l S_l\left(\frac{k}{T}\right) C^{1,0}_{jl} + \mathcal{O}(T^{-1}).$$

B.1.7 Proof of Lemma 4.3.7

For simplicity let $\mu_t = \mu(t/T)$. We have that

$$\mathbb{E}(I^j_k) = \sum_l A_{jl}S_j(k/T) + \left(\sum_t \mu_t\psi_{j,k-t}\right)^2 + \mathcal{O}(T^{-1}).$$

Using the characterisation of Hölder spaces in terms of wavelet coefficients (Daubechies, 1992, page 299), we have that for any Hölder continuous function μ with exponent β defined on rescaled time $z \in (0,1)$, and any Daubechies compactly supported wavelet ψ,

$$\left|\sum_t \mu_t\psi_{j,k-t}\right| = \mathcal{O}\left(2^{-j(\beta+1/2)}T^{-\beta}\right),$$

and hence

$$\left|\sum_t \mu_t\psi_{j,k-t}\right|^2 = \mathcal{O}\left(2^{-j(2\beta+1)}T^{-2\beta}\right).$$

The result follows by plugging this term into the expression for the expectation.

B.1.8 Proof of Theorem 4.3.8

First, the expectation and variance of the smoothed wavelet periodogram are described by the following lemma.

Lemma B.1.2. Suppose that $S_j(z) \leq D2^j$ for all $z \in (0,1)$ and some constant D and μ is Hölder continuous with exponent β. The expectation of the smoothed wavelet periodogram is given by

$$\mathbb{E}\left(\tilde{I}_k^j\right) = \sum_l A_{jl} S_l(k/T) + \mathcal{O}(nT^{-1}) + \mathcal{O}(2^{-j(2\beta+1)}T^{-2\beta}).$$

The asymptotic variance is given by

$$\mathrm{Var}\left(\tilde{I}_k^j\right) = \mathcal{O}(2^{-2j}n^{-1}) + \mathcal{O}(2^{-j}nT^{-1}),$$

and hence

$$\mathbb{E}\left(\tilde{I}_k^j - I_k^j\right)^2 = \mathcal{O}(2^{-2j}n^{-1}) + \mathcal{O}(2^{-j}nT^{-1}) + \mathcal{O}(2^{-2j(2\beta+1)}T^{-4\beta}).$$

Proof.

$$\begin{aligned}
\mathbb{E}\left(\tilde{I}_k^j\right) &= \mathbb{E}\left(\frac{1}{2n+1}\sum_{t=k-n}^{k+n} I_t^j\right) = \frac{1}{2n+1}\sum_{t=k-n}^{k+n}\mathbb{E}\left(I_t^j\right) \\
&= \frac{1}{2n+1}\sum_{t=k-n}^{k+n}\left(\sum_l A_{jl}S_l(t/T) + \mathcal{O}(2^{-j(2\beta+1)}T^{-2\beta}) + \mathcal{O}(T^{-1})\right) \\
&= \frac{1}{2n+1}\left(\sum_{t=k-n}^{k+n}\sum_l A_{jl}S_l(t/T)\right) + \mathcal{O}(2^{-j(2\beta+1)}T^{-2\beta}) + \mathcal{O}(T^{-1}) \\
&= \sum_l A_{jl}S_l(k/T) + \mathcal{O}(nT^{-1}) + \mathcal{O}(2^{-j(2\beta+1)}T^{-2\beta}),
\end{aligned}$$

which follows as an analogous result to Lemma A1 of Sanderson et al. (2010), as well as using the fact that S_j is Lipschitz with constant $K_j = \mathcal{O}(L_j)$, where L_j is the Lipschitz constant of W_j, and where the L_j satisfies $\sum_j 2^{-j} L_j < \infty$, and that $\sum_j 2^j A_{jl} = 1$. The expression for the variance can be derived via Lemma A1 of Sanderson et al. (2010). The variance can be written as

$$\mathrm{Var}\left(\tilde{I}_k^j\right) = \frac{1}{(2n+1)^2} \sum_{m=-n}^{n} \sum_{\tau} \mathrm{Cov}\left(I_{k+m}^j, I_{k+m+\tau}^j\right), \tag{B.1.3}$$

where $\tau = m' - m$. By expanding the expectation, using Isserlis theorem and Lemma A1, the covariance term is given by

$$\begin{aligned}\mathrm{Cov}\left(I_{k+m}^j, I_{k+m+\tau}^j\right) &= 2\mathbb{E}\left(\tilde{d}_{j,k+m}\tilde{d}_{j,k+m+\tau}\right)^2 + \\ &= 2\left(\sum_l S_l(k/T) A_{lj}^{\tau}\right)^2 + \mathcal{O}(2^{-j}|m|T^{-1}),\end{aligned}$$

where $A_{jl}^{\tau} = \sum_s \Psi_{l,j}(s)\Psi_{l,j}(s+\tau)$. Substituting this into Equation (B.1.3), we obtain

$$\mathrm{Var}\left(\tilde{I}_k^j\right) = \frac{2}{(2n+1)^2} \sum_{m=-n}^{n} \sum_{\tau} \left[\left(\sum_l S_l(k/T) A_{lj}^{\tau}\right)^2 + \mathcal{O}(2^{-j}|m|T^{-1})\right].$$

Using Park et al. (2014), examining the term

$$\begin{aligned}\sum_{\tau}\left(\sum_l S_l(k/T) A_{lj}^{\tau}\right)^2 &\leq \left(\sum_{\tau}\left|\sum_l S_l(k/T) A_{lj}^{\tau}\right|\right)\left(\sum_{\tau}\left|\sum_m S_m(k/T) A_{mj}^{\tau}\right|\right) \\ &\leq \left(\sum_{\tau}|c(k/T,\tau)|\sum_n |\Psi_j(n+\tau)|\right)^2 \\ &= \mathcal{O}(2^{-2j}),\end{aligned}$$

where we have assumed that $\sup_{z\in(0,1)}\sum_{\tau}|c(z,\tau)| < \infty$. Hence, we have that

$$\mathrm{Var}\left(\tilde{I}_k^j\right) = \frac{2}{(2n+1)^2} \sum_{m=-n}^{n}\left(\mathcal{O}(2^{-2j}) + \sum_{\tau}\mathcal{O}(2^{-j}|m|T^{-1})\right)$$

$$= \mathcal{O}(2^{-2j}n^{-1}) + \mathcal{O}(2^{-j}nT^{-1}).$$

Therefore, the mean squared error is given by

$$\text{MSE}\left(\tilde{I}_k^j\right) = \mathcal{O}(2^{-2j}n^{-1}) + \mathcal{O}(2^{-j}nT^{-1}) + \mathcal{O}(2^{-2j(2\beta+1)}T^{-4\beta}).$$

□

Finally, the expectation of the estimator is given by

$$\begin{aligned}
\mathbb{E}\left(\hat{S}_j(k/T)\right) &= \sum_{l=-J_0}^{-1} \mathbb{E}\left(\tilde{I}_k^l\right) A_{lj}^{-1} \\
&= \sum_{l=-J_0}^{-1} \left[\sum_p A_{pl} S_p(k/T) + \mathcal{O}(2^{j/2}nT^{-1}) + \mathcal{O}(2^{-l(2\beta+1)}T^{-2\beta})\right] A_{lj}^{-1} \\
&= S_j(k/T) + \mathcal{O}(nT^{-1}) + \mathcal{O}(2^{j/2}T^{\alpha(2\beta+1/2)-2\beta}),
\end{aligned}$$

which follows since $A_{jl}^{-1} \leq C2^{l/2}2^{j/2}$, $J_0 = \alpha \log_2 T$, and Lemma B.1.2. The mean squared error is given by

$$\text{MSE}\left(\hat{S}_j(k/T)\right) = \mathbb{E}\left(\sum_{l=-J_0}^{-1} \tilde{I}_k^l A_{lj}^{-1} - \sum_{l=-\infty}^{-1} \beta_l(k/T) A_{lj}^{-1}\right)^2,$$

where $\beta_l(k/T) = \sum_n S_n(k/T)A_{nl}$. As in Sanderson et al. (2010), we split this into two terms to obtain

$$\begin{aligned}
\text{MSE}\left(\hat{S}_j(k/T)\right) &\leq 2\mathbb{E}\left(\sum_{l=-J_0}^{-1} \left(\tilde{I}_k^l - \beta_l(k/T)\right) A_{lj}^{-1}\right)^2 + 2\left(\sum_{l<-J_0} \beta_l(k/T) A_{lj}^{-1}\right)^2 \\
&:= I + II.
\end{aligned}$$

The first term is

$$I \leq 2 \sum_{l=-J_0}^{-1} \mathbb{E}\left(\tilde{I}_k^l - \beta_l(k/T)\right)^2 \left(A_{lj}^{-1}\right)^2$$

$$\leq \mathcal{O}(2^j n^{-1} T^{\alpha}) + \mathcal{O}(2^j n T^{-1}) + \mathcal{O}(2^j T^{4\beta(\alpha-1)+\alpha}).$$

The result follows since $J_0 = \alpha \log_2 T$, $A_{lj}^{-1} \leq C 2^{l/2} 2^{j/2}$, and using Lemma B.1.2. For the second term, we use the assumption that $S_j \leq D 2^j$, as in Sanderson et al. (2010) and Fryzlewicz and Nason (2006). From this we obtain that $\beta_l(k/T) \leq D$, and hence $II = \mathcal{O}(2^{-J_0}) = \mathcal{O}(T^{-\alpha})$. Therefore, the estimator is consistent for each fixed scale j, provided that $n^{-1}T^{\alpha} \to 0$, $nT^{-1} \to 0$ and $T^{4\beta(\alpha-1)+\alpha} \to 0$ as $T \to \infty$ and $n \to \infty$.

B.1.9 Proof of Theorem 4.3.9

The mean squared error is given by:

$$\text{MSE}\left(\hat{c}(k/T, \tau)\right) = \mathbb{E}\left(\sum_{j=-J_0}^{-1} \hat{S}_j(k/T)\Psi_j(\tau) - \sum_{l=-\infty}^{-1} S_j(k/T)\Psi_j(\tau) \right)^2,$$

which again we split into two terms to obtain

$$\text{MSE}\left(\hat{c}(k/T, \tau)\right) \leq 2\mathbb{E}\left(\sum_{j=-J_0}^{-1} \left(\hat{S}_j(k/T) - S_j(k/T) \right) \Psi_j(\tau) \right)^2 + 2\left(\sum_{j<-J_0} S_j(k/T)\Psi_j(\tau) \right)^2$$

$$:= I + II.$$

The first term is

$$\begin{aligned} I &\leq 2 \sum_{j=-J_0}^{-1} \mathbb{E}\left(\hat{S}_j(k/T) - S_j(k/T) \right)^2 \Psi_j(\tau)^2 \leq 2 \sum_{j=-J_0}^{-1} \mathbb{E}\left(\hat{S}_j(k/T) - S_j(k/T) \right)^2 \\ &\leq 2 \sum_{j=-J_0}^{-1} \left(\mathcal{O}(2^j n^{-1} T^{\alpha}) + \mathcal{O}(2^j n T^{-1}) + \mathcal{O}(2^j T^{4\beta(\alpha-1)+\alpha}) \right) \\ &= \mathcal{O}(n^{-1} T^{\alpha}) + \mathcal{O}(n T^{-1}) + \mathcal{O}(T^{4\beta(\alpha-1)+\alpha}), \end{aligned}$$

which follows using the mean squared error of the EWS estimator, and that $\Psi_j(\tau)^2 \leq 1$ for all j, τ. The second term is $II = \mathcal{O}(T^{-2\alpha})$, since we have that $\sum_{j<-J_0} S_j(k/T)\Psi_j(\tau) = \mathcal{O}(2^{-J_0}) = \mathcal{O}(T^{-\alpha})$. Hence, the estimator is consistent,

provided that $n^{-1}T^{\alpha} \to 0$, $nT^{-1} \to 0$ and $T^{4\beta(\alpha-1)+\alpha} \to 0$ as $T \to \infty$ and $n \to \infty$.

B.1.10 Proof of Proposition 4.4.1

The non-boundary wavelet coefficients have expectation zero, since the LSW process is zero mean and the wavelet zeroes out the polynomial trend. Therefore, a linear combination of the non-boundary wavelet coefficients such as in the estimate $\hat{\mu}$ is also zero mean. Therefore,

$$\mathbb{E}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) = \mathbb{E}((\mathbf{x} - \hat{\boldsymbol{\epsilon}}) - (\mathbf{x} - \boldsymbol{\epsilon})) = \mathbb{E}(\boldsymbol{\epsilon} - \hat{\boldsymbol{\epsilon}}) = 0.$$

B.1.11 Proof of Proposition 4.4.2

Using Parseval's relation, the mean squared error can be written as

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\left[\left(\mu\left(\frac{t}{T}\right) - \hat{\mu}\left(\frac{t}{T}\right)\right)^2\right] = \frac{1}{T}\sum_{j,k}\mathbb{E}\left[\left(\hat{\theta}_{j,k} - \theta_{j,k}\right)^2\right],$$

where $\hat{\theta}_{j,k} = d_{j,k}I(j < -j_0)$, $\theta_{j,k}$ are the wavelet coefficients of $\mu(t/T)$, $d_{j,k}$ are the wavelet coefficients of the observed time series. Then,

$$\mathbb{E}\left[\left(\hat{\theta}_{j,k} - \theta_{j,k}\right)^2\right] = \mathbb{E}(\hat{\theta}_{j,k}^2) = \mathbb{E}(d_{j,k}^2)I(j < -j_0).$$

Assume that the process is generated with wavelet ψ_0, and we use wavelet ψ^1 for trend estimation. From Lemma 4.3.6, we have that

$$\mathbb{E}(d_{j,k}^2) = \sum_{l} C_{jl}^{1,0} S_l\left(k/T\right) + \mathcal{O}(T^{-1}),$$

and hence

$$\begin{aligned}\frac{1}{T}\sum_{j,k}\mathbb{E}\left[\left(\hat{\theta}_{j,k} - \theta_{j,k}\right)^2\right] &= \frac{1}{T}\sum_{j<-j_0}\sum_{k}\left(\sum_{l} C_{j,l}^{1,0} S_l(k/T) + \mathcal{O}(T^{-1})\right) \\ &= \frac{1}{T}\sum_{j<-j_0}\sum_{k}\sum_{l} C_{j,l}^{1,0} S_l(k/T) + \mathcal{O}(T^{-2})\sum_{j<-j_0}\sum_{k}\mathcal{O}(1)\end{aligned}$$

$$= \frac{1}{T} \sum_{j<-j_0} \sum_k \sum_l C_{j,l}^{1,0} S_l(k/T) + \mathcal{O}(T^{-2}) \sum_{j<-j_0} \mathcal{O}(2^j T),$$

where the second term is at least $\mathcal{O}(T^{-1})$, since $\sum_{j<-j_0} 2^j = \mathcal{O}(2^{-j_0})$. Next, we require the following lemma:

Lemma B.1.3.

$$\sum_j 2^j C_{j,l}^{1,0} = 1.$$

Proof. The result is analogous to Lemma 8 of Fryzlewicz et al. (2003). Using the fact that $\sum_j 2^j \Psi_j(\tau) = \delta_0(\tau)$, we have that

$$\sum_j 2^j C_{j,l}^{1,0} = \sum_j 2^j \sum_\tau \Psi_j^1(\tau)\Psi_l^0(\tau) = \sum_\tau \Psi_l^0(\tau) \sum_j 2^j \Psi_j^1(\tau) = \sum_\tau \Psi_l^0(\tau)\delta_0(\tau) = 1.$$

□

By assumption, $S_j(k/T) \leq D2^j$, from which we obtain that $\sum_l C_{j,l}^{1,0} S_l(k/T) \leq D$ using the above lemma. Hence, we obtain:

$$\begin{aligned}
\frac{1}{T} \sum_{j,k} \mathbb{E}\left[\left(\hat{\theta}_{j,k} - \theta_{j,k}\right)^2\right] &= \frac{1}{T} \sum_{j<-j_0} \sum_k \sum_l C_{j,l}^{1,0} S_l(k/T) + \mathcal{O}(T^{-1}) \\
&\leq \frac{1}{T} \sum_{j<-j_0} \sum_k \sum_l C_{j,l}^{1,0} D2^l + \mathcal{O}(T^{-1}) \\
&\leq \frac{1}{T} \sum_{j<-j_0} \sum_k D = \mathcal{O}(2^{j_0}) = \mathcal{O}(T^{-\alpha}).
\end{aligned}$$

B.1.12 Proof of Proposition 4.4.3

Using Parseval's relation,

$$\begin{aligned}
&\frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}\left[\left(\mu\left(\frac{t}{T}\right) - \hat{\mu}\left(\frac{t}{T}\right)\right)^2\right] = \frac{1}{T} \sum_{j,k} \mathbb{E}\left[\left(\hat{\theta}_{j,k} - \theta_{j,k}\right)^2\right] \\
&= \frac{1}{T} \sum_{j,k} \mathbb{E}\left[\left(\hat{\theta}_{j,k} - \theta_{j,k} I(j < -j_0) + \theta_{j,k} I(j < -j_0) - \theta_{j,k}\right)^2\right] \\
&\leq \frac{1}{T} \sum_{j,k} 2\mathbb{E}\left(\hat{\theta}_{j,k} - \theta_{j,k} I(j < -j_0)\right)^2 + 2\mathbb{E}\left(\theta_{j,k} I(j < -j_0) - \theta_{j,k}\right)^2
\end{aligned}$$

$$= I + II.$$

The first term can be bounded as:

$$\begin{aligned} I &\leq \frac{2}{T}\sum_{j,k}\mathbb{E}\left(d_{j,k}I(j<-j_0)-\theta_{j,k}I(j<-j_0)\right)^2 \\ &= \frac{2}{T}\sum_{j,k} I(j<-j_0)\mathbb{E}\left(d_{j,k}-\theta_{j,k}\right)^2 \\ &= \frac{2}{T}\sum_{j<-j_0}\sum_{k}\mathrm{Var}(d_{j,k}) = \frac{2}{T}\sum_{j<-j_0}\sum_{k}\left(\sum_{l} C_{jl}^{1,0}S_l\left(\frac{k}{T}\right)+\mathcal{O}(T^{-1})\right) \\ &= \mathcal{O}(T^{-\alpha}), \end{aligned}$$

where the last line follows from the proof of Proposition 4.4.2. The second term can be bounded as

$$\begin{aligned} I &\leq \frac{2}{T}\sum_{j,k}\mathbb{E}\left(\theta_{j,k}I(j<-j_0)-\theta_{j,k}\right)^2 = \frac{2}{T}\sum_{j,k} I(j\geq -j_0)\theta_{j,k}^2 \\ &= \mathcal{O}(T^{-1})\sum_{j\geq -j_0}\sum_{k}\theta_{j,k}^2 = \mathcal{O}(T^{-1})\sum_{j\geq -j_0}\sum_{k}\mathcal{O}(T^{-2\beta}2^{-j(2\beta+1)}) \\ &= \mathcal{O}(T^{-1})\sum_{j\geq -j_0}\mathcal{O}(T^{1-2\beta}2^{-2j\beta}) = \mathcal{O}(T^{2\beta(\alpha-1)}), \end{aligned}$$

since $\theta_{j,k} = \mathcal{O}(T^{-\beta}2^{-j(\beta+1/2)})$ by the wavelet characterisation of Hölder spaces on rescaled time.

Bibliography

Adams, N. M. and Heard, N. (2016). *Dynamic Networks and Cyber-Security*, volume 1. World Scientific.

Andersen, T. G., Davis, R. A., Kreiß, J.-P., and Mikosch, T. V. (2009). *Handbook of Financial Time Series*. Springer Science & Business Media.

Bai, J. and Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, 18(1):1–22.

Baranowski, R., Chen, Y., and Fryzlewicz, P. (2016). *not: Narrowest-over-threshold change-point detection*. R Package Version 1.0.

Baranowski, R., Chen, Y., and Fryzlewicz, P. (2019). Narrowest-over-threshold detection of multiple change points and change-point-like features. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 81(3):649–672.

Baskakov, A. G. (1990). Wiener's theorem and the asymptotic estimates of the elements of inverse matrices. *Functional Analysis and Its Applications*, 24(3):222–224.

Beaulieu, C. and Killick, R. (2018). Distinguishing trends and shifts from memory in climate data. *Journal of Climate*, 31(23):9519–9543.

Benedetti, J. K. (1977). On the nonparametric estimation of regression functions. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(2):248–253.

Beran, J. and Feng, Y. (2002). Semifar models – a semiparametric approach to modelling trends, long-range dependence and nonstationarity. *Computational Statistics & Data Analysis*, 40(2):393–419.

Boulton, C. A. and Lenton, T. M. (2015). Slowing down of north pacific climate variability and its implications for abrupt ecosystem change. *Proceedings of the National Academy of Sciences*, 112(37):11496–11501.

Brillinger, D. R. (1994). Some river wavelets. *Environmetrics*, 5(3):211–220.

Brockwell, P. J., Davis, R. A., and Fienberg, S. E. (1991). *Time Series: Theory and Methods*. Springer Science & Business Media.

Bücher, A. and Kojadinovic, I. (2016). Dependent multiplier bootstraps for non-degenerate U-statistics under mixing conditions with applications. *Journal of Statistical Planning and Inference*, 170:83–105.

Cai, T. T. and Brown, L. D. (1998). Wavelet shrinkage for nonequispaced samples. *The Annals of Statistics*, 26(5):1783–1799.

Cai, T. T. and Silverman, B. W. (2001). Incorporating information on neighbouring coefficients into wavelet estimation. *Sankhyā: The Indian Journal of Statistics, Series B*, pages 127–148.

Cardinali, A. and Nason, G. P. (2017). Locally stationary wavelet packet processes: Basis selection and model fitting. *Journal of Time Series Analysis*, 38(2):151–174.

Carr, J. R., Bell, H., Killick, R., and Holt, T. (2017). Exceptional retreat of Novaya Zemlya's marine-terminating outlet glaciers between 2000 and 2013. *The Cryosphere.*, 11(5):2149–2174.

Chakar, S., Lebarbier, E., Lévy-Leduc, C., and Robin, S. (2014). *AR1Seg: Segmentation of an AR(1) Gaussian process*. R package version 1.0.

Chakar, S., Lebarbier, E., Lévy-Leduc, C., and Robin, S. (2017). A robust approach for estimating change-points in the mean of an AR(1) process. *Bernoulli*, 23(2):1408–1447.

Chan, K. H., Hayya, J. C., and Ord, J. K. (1977). A note on trend removal methods: the case of polynomial regression versus variate differencing. *Econometrica*, 45(3):737.

Chan, T. F. and Shen, J. (2005). *Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods*. SIAM.

Chen, J. and Gupta, A. K. (2011). *Parametric Statistical Change Point Analysis: with Applications to Genetics, Medicine, and Finance*. Springer Science & Business Media.

Chui, C. K. (1997). *Wavelets: A Mathematical Tool for Signal Analysis*, volume 1. SIAM.

Cohen, A., Daubechies, I., and Feauveau, J.-C. (1992). Biorthogonal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, 45(5):485–560.

Coifman, R. R. and Donoho, D. L. (1995). Translation-invariant de-noising. In *Wavelets and Statistics*, pages 125–150. Springer.

Coifman, R. R. and Wickerhauser, M. V. (1992). Entropy-based algorithms for best basis selection. *IEEE Transactions on Information Theory*, 38(2):713–718.

Craigmile, P. F., Guttorp, P., and Percival, D. B. (2004). Trend assessment in a long memory dependence model using the discrete wavelet transform. *Environmetrics*, 15(4):313–335.

Craigmile, P. F., Guttorp, P., and Percival, D. B. (2005). Wavelet-based parameter estimation for polynomial contaminated fractionally differenced processes. *IEEE Transactions on Signal Processing*, 53(8):3151–3161.

Csörgö, M. and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis*, volume 18. John Wiley & Sons Inc.

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *The Annals of Statistics*, 25(1):1–37.

Dahlhaus, R. (2000). A likelihood approximation for locally stationary processes. *The Annals of Statistics*, 28(6):1762–1794.

Dahlhaus, R. (2012). Locally Stationary Processes. In *Handbook of Statistics*, volume 30, pages 351–413. Elsevier.

Dahlhaus, R. and Neumann, M. H. (2001). Locally adaptive fitting of semiparametric models to nonstationary time series. *Stochastic Processes and their Applications*, 91(2):277–308.

Dai, W., Ma, Y., Tong, T., and Zhu, L. (2015). Difference-based variance estimation in nonparametric regression with repeated measurement data. *Journal of Statistical Planning and Inference*, 163:1–20.

Das, S. and Politis, D. N. (2020). Predictive inference for locally stationary time series with an application to climate data. *Journal of the American Statistical Association*, pages 1–16.

Daubechies, I. (1988). Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, 41(7):909–996.

Daubechies, I. (1992). *Ten Lectures on Wavelets*. SIAM.

Dette, H., Eckle, T., and Vetter, M. (2018). Multiscale change point detection for dependent data. *Scandinavian Journal of Statistics*.

Dette, H. and Wu, W. (2019). Detecting relevant changes in the mean of nonstationary processes – a mass excess approach. *The Annals of Statistics*, 47(6):3578–3608.

Dette, H. and Wu, W. (2020). Prediction in locally stationary time series. *arXiv preprint arXiv:2001.00419*.

Donoho, D. L. (1993). Unconditional bases are optimal bases for data compression and for statistical estimation. *Applied and Computational Harmonic Analysis*, 1(1):100–115.

Donoho, D. L. (1995). De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, 41(3):613–627.

Donoho, D. L. and Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 90(432):1200–1224.

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G., and Picard, D. (1995). Wavelet shrinkage: asymptopia? *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(2):301–337.

Donoho, D. L. and Johnstone, J. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications: Monographs on Statistics and Applied Probability*, volume 66. CRC Press.

Ferreira, G., Rodríguez, A., and Lagos, B. (2013). Kalman filter estimation for a regression model with locally stationary errors. *Computational Statistics & Data Analysis*, 62:52–69.

Fisch, A., Eckley, I. A., and Fearnhead, P. (2018). A linear time method for the detection of point and collective anomalies. *arXiv preprint arXiv:1806.01947*.

Frick, K., Munk, A., and Sieling, H. (2014). Multiscale change point inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(3):495–580.

Fryzlewicz, P. (2005). Modelling and forecasting financial log-returns as locally stationary wavelet processes. *Journal of Applied Statistics*, 32(5):503.

Fryzlewicz, P. (2014). Wild binary segmentation for multiple change-point detection. *The Annals of Statistics*, 42(6):2243–2281.

Fryzlewicz, P. (2018). Tail-greedy bottom-up data decompositions and fast multiple change-point detection. *The Annals of Statistics*, 46(6B):3390–3421.

Fryzlewicz, P. and Nason, G. P. (2006). Haar–Fisz estimation of evolutionary wavelet spectra. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(4):611–634.

Fryzlewicz, P., Van Bellegem, S., and von Sachs, R. (2003). Forecasting non-stationary time series by wavelet process modelling. *Annals of the Institute of Statistical Mathematics*, 55(4):737–764.

Fryzlewicz, P. Z. (2003). *Wavelet Techniques for Time Series and Poisson Data.* PhD thesis, University of Bristol.

Gao, H.-Y. and Bruce, A. G. (1997). Waveshrink with firm shrinkage. *Statistica Sinica*, pages 855–874.

Goodman, T. N. T., Micchelli, C. A., Rodriguez, G., and Seatzu, S. (1995). On the Cholesky factorization of the Gram matrix of locally supported functions. *BIT Numerical Mathematics*, 35(2):233–257.

Gott, A. N. and Eckley, I. A. (2013). A note on the effect of wavelet choice on the estimation of the evolutionary wavelet spectrum. *Communications in Statistics – Simulation and Computation*, 42(2):393–406.

Green, P. J. and Silverman, B. W. (1993). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach.* CRC Press.

Haar, A. (1910). Zur Theorie der orthogonalen Funktionensysteme. *Mathematische Annalen*, 69(3):331–371.

Hansen, J., Sato, M., Ruedy, R., Lo, K., Lea, D. W., and Medina-Elizade, M. (2006). Global temperature change. *Proceedings of the National Academy of Sciences*, 103(39):14288–14293.

Harchaoui, Z. and Lévy-Leduc, C. (2010). Multiple change-point estimation with a total variation penalty. *Journal of the American Statistical Association*, 105(492):1480–1493.

Härdle, W. (1990). *Applied Nonparametric Regression*. Number 19. Cambridge university press.

Härdle, W. K. (1991). *Smoothing Techniques: with Implementation in S*. Springer Science & Business Media.

Hargreaves, J. K., Knight, M. I., Pitchford, J. W., Oakenfull, R. J., and Davis, S. J. (2018). Clustering nonstationary circadian rhythms using locally stationary wavelet representations. *Multiscale Modeling & Simulation*, 16(1):184–214.

Hart, J. D. (1989). Differencing as an approximate de-trending device. *Stochastic Processes and their Applications*, 31(2):251–259.

Hocking, T. D., Schleiermacher, G., Janoueix-Lerosey, I., Boeva, V., Cappo, J., Delattre, O., Bach, F., and Vert, J.-P. (2013). Learning smoothing models of copy number profiles using breakpoint annotations. *BMC Bioinformatics*, 14(1):164.

Hu, L., Huang, T., and You, J. (2019). Two-step estimation of time-varying additive model for locally stationary time series. *Computational Statistics & Data Analysis*, 130:94–110.

Huang, B., Banzon, V. F., Freeman, E., Lawrimore, J., Liu, W., Peterson, T. C., Smith, T. M., Thorne, P. W., Woodruff, S. D., and Zhang, H.-M. (2015). Extended reconstructed sea surface temperature version 4 (ERSST. v4). Part I: upgrades and intercomparisons. *Journal of Climate*, 28(3):911–930.

Jaffard, S. (1990). Propriétés des matrices «bien localisées» près de leur diagonale et quelques applications. In *Annales de l'Institut Henri Poincare (C) Non Linear Analysis*, volume 7, pages 461–476. Elsevier.

Jandhyala, V., Fotopoulos, S., MacNeill, I., and Liu, P. (2013). Inference for single and multiple change-points in time series. *Journal of Time Series Analysis*, 34(4):423–446.

Johnstone, I. M. and Silverman, B. W. (1997). Wavelet threshold estimators for data with correlated noise. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(2):319–351.

Johnstone, I. M. and Silverman, B. W. (2005). Empirical Bayes selection of wavelet thresholds. *Annals of Statistics*, pages 1700–1752.

Kallache, M., Rust, H. W., and Kropp, J. (2005). Trend assessment: applications for hydrology and climate research. *Nonlinear Processes in Geophysics*, 12(2):201–210.

Karl, T. R., Arguez, A., Huang, B., Lawrimore, J. H., McMahon, J. R., Menne, M. J., Peterson, T. C., Vose, R. S., and Zhang, H.-M. (2015). Possible artifacts of data biases in the recent global surface warming hiatus. *Science*, 348(6242):1469–1472.

Karl, T. R., Knight, R. W., and Baker, B. (2000). The record breaking global temperatures of 1997 and 1998: Evidence for an increase in the rate of global warming? *Geophysical Research Letters*, 27(5):719–722.

Kellogg, W. W. (1993). An apparent moratorium on the greenhouse warming due to the deep ocean. *Climatic Change*, 25(1):85–88.

Khismatullina, M. and Vogt, M. (2020). Multiscale inference and long-run variance estimation in non-parametric regression with time series errors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(1):5–37.

Killick, R. and Eckley, I. (2014). changepoint: An R package for changepoint analysis. *Journal of Statistical Software*, 58(3):1–19.

Killick, R., Eckley, I., and Jonathan, P. (2013). A wavelet-based approach for detecting changes in second order structure within nonstationary time series. *Electronic Journal of Statistics*, 7:1167–1183.

Killick, R., Fearnhead, P., and Eckley, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, 107(500):1590–1598.

Knight, M. I., Nunes, M. A., and Nason, G. P. (2012). Spectral estimation for locally stationary time series with missing observations. *Statistics and Computing*, 22(4):877–895.

Kojadinovic, I. (2015). *npcp: Some nonparametric tests for change-point detection in possibly multivariate observations*. R package version 0.1-6.

Kokoszka, P. and Leipus, R. (1998). Change-point in the mean of dependent observations. *Statistics & probability letters*, 40(4):385–393.

Krampe, J., Kreiss, J.-P., and Paparoditis, E. (2015). Hybrid wild bootstrap for nonparametric trend estimation in locally stationary time series. *Statistics & Probability Letters*, 101:54–63.

Krishtal, I., Strohmer, T., and Wertz, T. (2015). Localization of matrix factorizations. *Foundations of Computational Mathematics*, 15(4):931–951.

Krzemieniewska, K., Eckley, I. A., and Fearnhead, P. (2014). Classification of non-stationary time series. *Stat*, 3(1):144–157.

Lavielle, M. (1999). Detection of multiple changes in a sequence of dependent variables. *Stochastic Processes and their Applications*, 83(1):79–102.

Lawton, W. (1993). Applications of complex valued wavelet transforms to subband decomposition. *IEEE Transactions on Signal Processing*, 41(12):3566–3568.

Lenton, T. M., Dakos, V., Bathiany, S., and Scheffer, M. (2017). Observed trends in the magnitude and persistence of monthly temperature variability. *Scientific Reports*, 7(1):5940.

Lina, J.-M. and Mayrand, M. (1995). Complex Daubechies wavelets. *Applied and Computational Harmonic Analysis*, 2(3):219–229.

Liu, W., Huang, B., Thorne, P. W., Banzon, V. F., Zhang, H.-M., Freeman, E., Lawrimore, J., Peterson, T. C., Smith, T. M., and Woodruff, S. D. (2015). Extended reconstructed sea surface temperature version 4 (ERSST. v4): Part II. parametric and structural uncertainty estimations. *Journal of Climate*, 28(3):931–951.

Lund, R. and Shi, X. (2020). Short communication: Detecting possibly frequent change-points: Wild binary segmentation 2 and steepest-drop model selection. *arXiv preprint arXiv:2006.10845v1*.

Mallat, S. G. (1989a). Multiresolution approximations and wavelet orthonormal bases of $L^2(\mathbb{R})$. *Transactions of the American Mathematical Society*, 315(1):69–87.

Mallat, S. G. (1989b). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693.

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability & Its Applications*, 9(1):141–142.

Nason, G. (2008). *Wavelet Methods in Statistics with R*. Springer Science & Business Media.

Nason, G. (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 (Statistical Methodology)*, 75(5):879–904.

Nason, G. (2016a). *locits: Tests of stationarity and localized autocovariance*. R package version 1.7.3.

Nason, G. (2016b). *wavethresh: Wavelet transforms and associated statistical methodology*. R package version 4.6.8.

Nason, G. and Stevens, K. (2015). Bayesian wavelet shrinkage of the Haar-Fisz transformed wavelet periodogram. *PloS one*, 10(9):e0137662.

Nason, G. P. (1996). Wavelet shrinkage using cross-validation. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(2):463–479.

Nason, G. P. and Silverman, B. W. (1995). The Stationary Wavelet Transform and some Statistical Applications. *Lecture Notes in Statistics*, pages 281–281.

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 (Statistical Methodology)*, 62(2):271–292.

Neumann, M. H. and von Sachs, R. (1995). Wavelet thresholding: beyond the Gaussian IID situation. In *Wavelets and Statistics*, pages 301–329. Springer.

Neumann, M. H. and von Sachs, R. (1997). Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra. *The Annals of Statistics*, 25(1):38–76.

Newey, W. K. and West, K. D. (1986). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Technical report, National Bureau of Economic Research.

Ombao, H., Raz, J., von Sachs, R., and Guo, W. (2002). The SLEX model of a non-stationary random process. *Annals of the Institute of Statistical Mathematics*, 54(1):171–200.

Ombao, H. C., Raz, J. A., von Sachs, R., and Malow, B. A. (2001). Automatic statistical analysis of bivariate nonstationary time series. *Journal of the American Statistical Association*, 96(454):543–560.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, 41(1/2):100–115.

Page, E. S. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika*, 42(3/4):523–527.

Park, T., Eckley, I. A., and Ombao, H. C. (2014). Estimating time-evolving partial coherence between signals via multivariate locally stationary wavelet processes. *IEEE Transactions on Signal Processing*, 62(20):5240–5250.

Percival, D. B. and Walden, A. T. (2006). *Wavelet Methods for Time Series Analysis*, volume 4. Cambridge university press.

Pesquet, J.-C., Krim, H., and Carfantan, H. (1996). Time-invariant orthonormal wavelet representations. *IEEE Transactions on Signal Processing*, 44(8):1964–1970.

Priestley, M. (1983). *Spectral Analysis and Time Series*. Academic Press, London.

Priestley, M. B. (1965). Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 27(2):204–229.

Priestley, M. B. and Chao, M. (1972). Non-parametric function fitting. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(3):385–392.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Reichel, L. and Trefethen, L. N. (1992). Eigenvalues and pseudo-eigenvalues of Toeplitz matrices. *Linear Algebra and its Applications*, 162:153–185.

Rodionov, S. N. (2004). A sequential algorithm for testing climate regime shifts. *Geophysical Research Letters*, 31(9).

Rothman, A. J. (2012). Positive definite estimators of large covariance matrices. *Biometrika*, 99(3):733–740.

Roueff, F. and von Sachs, R. (2011). Locally stationary long memory estimation. *Stochastic Processes and their Applications*, 121(4):813–844.

Roueff, F. and von Sachs, R. (2019). Time-frequency analysis of locally stationary Hawkes processes. *Bernoulli*, 25(2):1355–1385.

Sanderson, J., Fryzlewicz, P., and Jones, M. (2010). Estimating linear dependence between nonstationary time series using the locally stationary wavelet model. *Biometrika*, 97(2):435–446.

Sardy, S., Percival, D. B., Bruce, A. G., Gao, H.-Y., and Stuetzle, W. (1999). Wavelet shrinkage for unequally spaced data. *Statistics and Computing*, 9(1):65–75.

Scott, A. J. and Knott, M. (1974). A cluster analysis method for grouping means in the analysis of variance. *Biometrics*, pages 507–512.

Shen, C. (2015). Analysis of detrended time-lagged cross-correlation between two nonstationary time series. *Physics Letters A*, 379(7):680–687.

Shumway, R. H. and Stoffer, D. S. (2010). *Time Series Analysis and its Applications: with R Examples*. Springer Science & Business Media.

Simonoff, J. S. (2012). *Smoothing Methods in Statistics*. Springer Science & Business Media.

Stevens, N. K. (2013). *New Methods for Spectral Estimation with Confidence Intervals for Locally Stationary Wavelet Processes*. PhD thesis, University of Bristol.

Sweldens, W. (1995). Lifting scheme: a new philosophy in biorthogonal wavelet constructions. In *Wavelet Applications in Signal and Image Processing III*, volume 2569, pages 68–79. International Society for Optics and Photonics.

Tartakovsky, A., Nikiforov, I., and Basseville, M. (2014). *Sequential Analysis: Hypothesis Testing and Changepoint Detection*. Chapman and Hall/CRC.

Tecuapetla-Gómez, I. and Munk, A. (2017). Autocovariance estimation in regression with a discontinuous signal and m-dependent errors: A difference-based approach. *Scandinavian Journal of Statistics*, 44(2):346–368.

Trenberth, K. E. (2015). Has there been a hiatus? *Science*, 349(6249):691–692.

Tunyavetchakit, S. (2010). *On the Optimal Segment Length for Tapered Yule-Walker Estimates for Time-Varying Autoregressive Processes*. PhD thesis, Heidelberg University.

UK Land Registry (2020). UK house price index. `http://landregistry.data.gov.uk/app/ukhpi`. Accessed 01/06/2020.

Van Bellegem, S. and Dahlhaus, R. (2006). Semiparametric estimation by model selection for locally stationary processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(5):721–746.

Van Bellegem, S. and von Sachs, R. (2008). Locally adaptive estimation of evolutionary wavelet spectra. *The Annals of Statistics*, 36(4):1879–1924.

Vidakovic, B. (2009). *Statistical Modeling by Wavelets*, volume 503. John Wiley & Sons.

Vogt, M. (2012). Nonparametric regression for locally stationary time series. *The Annals of Statistics*, 40(5):2601–2633.

Vogt, M. and Dette, H. (2015). Detecting gradual changes in locally stationary processes. *The Annals of Statistics*, 43(2):713–740.

von Sachs, R. and MacGibbon, B. (2000). Non-parametric curve estimation by wavelet thresholding with locally stationary errors. *Scandinavian Journal of Statistics*, 27(3):475–499.

von Sachs, R., Nason, G. P., and Kroisandt, G. (1997). Adaptive estimation of the evolutionary wavelet spectrum. Technical report, Department of Mathematics, University of Kaiserslautern and Department of Mathematics, University of Bristol.

von Sachs, R. and Schneider, K. (1996). Wavelet smoothing of evolutionary spectra by nonlinear thresholding. *Applied and Computational Harmonic Analysis*, 3(3):268–282.

Vostrikova, L. Y. (1981). Detecting "disorder" in multidimensional random processes. In *Doklady Akademii Nauk*, volume 259, pages 270–274. Russian Academy of Sciences.

Wang, Y. (1998). Smoothing spline models with correlated random errors. *Journal of the American Statistical Association*, 93(441):341–348.

Watson, G. S. (1964). Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 359–372.

Williams, J. R. and Amaratunga, K. (1994). Introduction to wavelets in engineering. *International Journal for Numerical Methods in Engineering*, 37(14):2365–2388.

Wu, W. and Zhou, Z. (2020). Multiscale jump testing and estimation under complex temporal dynamics. *arXiv preprint arXiv:1909.06307*.

Xiao, Z., Linton, O. B., Carroll, R. J., and Mammen, E. (2003). More efficient local polynomial estimation in nonparametric regression with autocorrelated errors. *Journal of the American Statistical Association*, 98(464):980–992.

Zhang, N. R. and Siegmund, D. O. (2007). A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data. *Biometrics*, 63(1):22–32.