Detecting patterns in Time Series Data with applications in Official Statistics

Ben Norwood, MSci (Hons.)

Department of Mathematics and Statistics

Lancaster University

This thesis is submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

September 2019

Abstract

This thesis examines the issue of detecting components or features within time series data in automatic procedures. We begin by introducing the concept of Wavelets and briefly show their usage as a tool for detection. This leads to our first contribution which is a novel method using wavelets for identifying correlation structures in time series data which are often ambiguous with very different contexts. Using the properties of the wavelet transform we show the ability to distinguish between short memory models with changepoints and long memory models. The next two Chapters consider seasonality within data, which is often present in time series used in Offical Statistics. We first describe the historical evolution of identification of seasonality, comparing and contrasting methodology as it has expanded throughout time. Following this, motivated by the increased use of high-frequency time series in Official Statistics and a lack of methods for identifying low-frequency seasonal components within high-frequency data, we present a method for identifying periodicity in a series with the use of a simple wavelet decomposition. Presented with theoretical results and simulations, we show how the seasonality of a series is uniquely represented within a wavelet transform and use this to identify low frequency components which are often overlooked in favour of a trend, with very different interpretations. Finally, beginning with the motivation of forecasting European Area GDP at the current time point, we show the effectiveness of an algorithm which detects the most useful data and structures for a Dynamic Factor Model. We show its effectiveness in reducing forecasting errors but show that under large scale simulation that the recovery of the true structure

over two dimensions is a difficult task. All the chapters of this thesis are motivated by, and give applications to, time series from different areas of Official Statistics.

Acknowledgements

I'd like to take this oppurtunity to mention and thank everyone who has made this happen. Firstly I would like to thank my colleagues and friends Laura Barlow, Callum Vyner and Jessica Welding, for always providing an interesting workplace and keeping me grounded. You have all taught me so much. Thank you also to my friends who have continually waited for me to finish this: Jon, Ben and Jake, you have stuck by me whilst my head has been firmly in this thesis and I can't acknowledge that enough. To my parents who have continually pushed me to higher levels, I would not have made it without you or your love. A very special thank you to my supervisor, and now dear friend, Dr. Rebecca Killick whom without her expertise, patience and kindness this would not have been possible.

Financial support for this work was granted by the Economic and Social Research Council, and for that I gratefully acknowledge them. My industrial supervisor, Duncan Elliott, and his team at The Office of National Statistics have made my time working with them invaluable.

Finally, I'd like to thank my partner Anna France. Your love and support has kept me going in ways I cannot find the words to explain. You challenge me to become better and you've always kept me level-headed. I consider myself very lucky that we could complete our work together.

Declaration

I declare that this thesis is my own work, and has not been submitted by myself in substantially the same form for the award of a higher degree elsewhere.

Ben Norwood

Published Work

This thesis includes the following published work

- Norwood, B., Killick, R. (2018) *Long Memory and Changepoint Models: A Spectral Classification Procedure*. Statistics and Computing 28(2) 291–302
- Norwood, B., Killick, R. (2016) *Evolution of Seasonal Adjustment Methods*. EuroStat Report, ESTAT/11111.2013.001-2013.254
- Norwood, B., Killick, R. (2019) *Nowcasting: Practical Results with Theoretical Flaws*. EuroStat Report, ESTAT/11111.2013.001-2013.254

Each piece of this work is given as Chapters 3, 4 and 6. Chapters 4 and 6 were created for Eurostat under contract number ESTAT/11111.2013.001-2013.254 through a subcontract between Lancaster University and Sogeti.

Motivations and appropriate literature reviews are provided in each of the chapters. Notation, where possible, has been kept consistent throughout and is introduced separately in each chapter. Therefore they may be read individually or as a whole. Relevant appendices are given at the end of each chapter. The bibliography for all chapters is given at the end.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

As technology has evolved, so has the data which we collect. Whilst accuracy and length of data is often the issue for most pracitioners, a new emerging area is the breadth, or width, of the data we collect. As the ability to store data efficiently has increased and is no longer a limiting factor to the decision maker, data is now collected from a wide variety of sources. The issue of automatically analysing and utilising data becomes ever more important when it is predicted that the total amount of new data generated in 2025 will be over five times more than that generated in 2018 (Reinsel et al., 2019) and thus we must develop methodology to meet this sudden rise in supply.

Wavelets have been used as a tool to efficiently store important features of a data set. They have grown in their popularity as a tool to succinctly decompose a series, and this thesis develops methods that continue in this vein. We begin by reviewing the key concepts and techniques used such as Wavelet Decomposition within Chapter 2. Here we present the key advantages of a time varying decomposition, and show how they have been used in other areas of statistics such as denoising and compression. Following this, in Chapter 3 we apply these particular advantages to the case of classification between ambiguous models with very different interpretations. Distinguishing between Long Memory or Short Memory with Changepoints can significantly change interpretation, and we present a solution to this ambiguity alongside applications to Price Inflation and Stock Correlation

Data.

To meet this high demand on knowledge we develop methodology for the use of data which is collected more often, or at a higher frequency. However with this comes additional concerns over interpretation, particularly in the area of seasonal adjustment. Chapter 4 reviews the methodology currently available for seasonal adjustment, providing one page overviews of the methods as they have developed through time. We then look to use the advantages of Wavelets in Chapter 5 to aid in the discovery of low frequency seasonal components which may often be overlooked as trends. We provide theoretical results for a particular type of seasonality, or periodicity and present methodology to detect and identify this component, before applying it French National Birth data.

With this increasing data availability, expectations are also rising as to the lead time of publication of information. In the domain of Official Statistics, this raises an interesting issue where data collection can only occur within a certain timeframe, but the output still needs to be estimated to meet the growing demand for timely information. Therefore methodology, colloquially referred to as 'Now-casting', was created for the purpose of estimating large complex systems at the current timepoint using their relationships with other more readily available data. With the motivation of European Area Gross Domestic Product, Chapter 6 looks at how these relationships change over time and how this may aid estimation of the system. With encouraging results we then extend our methodology further and test its capability to recover the truth under simulation.

Chapter 2

Wavelet Literature Review

The following chapter introduces the concept of wavelets and explains their common use cases. We begin by introducing them in terms of a Multi-Resolution Analysis, before showing their relationships in the Fourier Domain. Following this we give the example of the Haar Wavelet, then show how a wavelet decomposition is constructed. We then explore this further by reviewing the decimated and non-decimated approaches. Continuing, we then explain the concept of Locally Stationary Wavelet Processes before giving a brief review of the literature of wavelet usage. Finally we conclude by mentioning a number of other wavelet methods which we do not cover here.

2.1 Introduction

Defined seminally in Daubechies (1992) wavelets, or 'little waves', are a specially constructed localised function which are used to capture information and features on a local and global scale of a function or data. We begin first by formalising this concept of multi-resolution analysis, before showing the relevance of wavelet functions in this context.

Definition 2.1. *A multi-resolution analysis is a collection of closed subspaces of functions* $V_j \in \mathbb{L}_2(\mathbb{R})$ *for* $j \in \mathbb{Z}$ *such that the following conditions are satisfied:*

1. *A containment hierarchy exists as such:*

$$\cdots \subset V_{j+2} \subset V_{j+1} \subset V_j \subset V_{j-1} \subset V_{j-2} \subset \ldots \tag{2.1}$$

2. *There is a trivial intersection and dense union of these spaces,*

$$\bigcap_j V_j = \{\mathbf{0}\}, \qquad \bigcup_j V_j = \mathbb{L}_2(\mathbb{R}).$$

3. *There is a self-similarity in the spaces by scale,*

$$f(2^{-j}x) \in V_j \iff f(x) \in V_0.$$

4. *Each space is invariant to translation,*

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

5. *There exists a function $\phi(x) \in \mathbb{L}_2(\mathbb{R})$, termed the scaling function, such that for all $j \in \mathbb{Z}$ and $k \in \mathbb{Z}$; the functions $2^{-\frac{j}{2}}\phi(2^{-j}x + k)$ form an orthonormal basis for V_j.*

Here and going forward, we refer to the parameters j and k as scale and translate respectively. It should be noted that in other literature often the index of the subspaces is negatively reversed, however we proceed as this is more natural for the algorithm development in the works that follow, as noted in Vidakovic (2009). If we now consider the scaling function, usually referred to as the *father wavelet* $\phi(x) \in V_0$, and the hierarchy shown in Equation (2.1) we can then determine a relationship between scaling functions. This is such that

$$\phi_{j+1}(x) = \sum_{k=-\infty}^{\infty} h_k 2^{-\frac{j}{2}} \phi(2^{-j}x + k) = \sum_{k \in \mathbb{Z}} h_k \phi_{j,k}(x) \tag{2.2}$$

so we can form an orthonormal basis for V_j based upon a function from V_0 and

additional constant h_k where $k \in \mathbb{Z}$. It is the set of h_k which form a wavelet filter, of which we are particularly interested in. Each of these wavelet filters satisfy two main conditions: firstly normalization such that $\sum_{k\in\mathbb{Z}} h_k = \sqrt{2}$; and orthogonality such that

$$\sum_{k\in\mathbb{Z}} h_k h_{k-2l} = \delta_l, \qquad \text{for} \quad l \in \mathbb{Z}.$$

Consider next that following Equation (2.1) there exists a difference space between each subspace, such that

$$W_j = V_{j-1} \ominus V_j. \tag{2.3}$$

Given that we are able to determine a wavelet filter h_k on a scaling function $\phi(x)$, this implies there exists a non-unique orthonormal basis in $\mathbb{L}_2(\mathbb{R})$ for the different space W_j. We denote this basis by the set

$$\{\psi_{j,k}(x) = 2^{\frac{-j}{2}}\psi(2^{-j}x + k) : j, k \in \mathbb{Z}\}. \tag{2.4}$$

The function $\psi_{0,0}(x)$ or $\psi(x)$ is referred to as the wavelet function, or colloquially the *mother wavelet*. As $\psi(x) \in W_0 \subset V_{-1}$ we can relate the mother to the father wavelet as

$$\psi(x) = \sum_{k=-\infty}^{\infty} g_k \frac{1}{\sqrt{2}}\phi(\frac{1}{2}x + k)$$

or more generally

$$\psi_{j,k}(x) = \sum_{k=-\infty}^{\infty} g_k 2^{-\frac{j}{2}}\phi(2^{-j-1}x + k).$$

2.2 Fourier Analysis

It is of interest to look at the properties of the mother and father wavelet in the time-frequency dimensions. Such an analysis can be done by reviewing the functions under a Fourier transformation.

Definition 2.2. *For a given function $f \in \mathbb{L}_1(\mathbb{R})$, the Fourier transformation is*

given by

$$\hat{f}(\omega) = \mathcal{F}[f(x)] = \int_{\mathbb{R}} f(x)e^{-i\omega x}\,dx. \tag{2.5}$$

This transformation can be reversed also:

$$f(x) = \mathcal{F}^{-1}[\hat{f}(\omega)] = \frac{1}{2\pi}\int \hat{f}(\omega)e^{i\omega x}\,d\omega.$$

The Fourier transformation of a function allows analysis over frequency, but sacrifices the information from the time dimension of the original function. There are many properties attached to this transformation, which we will make use of in the definitions and explanations to follow, however we do not detail them here but refer the reader to Vidakovic (2009). It can be shown that by using the properties associated with the Fourier transform, that there is a likewise recurrence relation on the scaling function. This is such that

$$\Phi(\omega) = \mathcal{F}[\phi(x)] = 2^{-\frac{1}{2}}\sum_{k\in\mathbb{Z}} h_k \exp^{-ik\omega}\,\Phi\left(\frac{\omega}{2}\right) = m_0\left(\frac{\omega}{2}\right)\Phi\left(\frac{\omega}{2}\right) = \prod_{j=1}^{\infty}\Phi\left(\frac{\omega}{2^j}\right).$$

Similarly we can show how the set of translates $\{\phi(x+k)\}_{k\in\mathbb{Z}}$ using the Fourier transform as

$$\sum_{k=-\infty}^{\infty} |\Phi(\omega + 2\pi k)|^2,$$

or similarly (but not equivalently)

$$|m_0(\omega)|^2 + |m_o(\omega+\pi)|^2 = 1.$$

For more information on the difference between these two statements, please consult Vidakovic (2009).

Moving next to the wavelet filter, we have already shown a relationship between both functions, but we further investigate it within the Fourier domain. Following

similar calculations to those performed on the scaling function, we find that

$$\Psi(\omega) = \mathcal{F}[\psi(x)] = 2^{-\frac{1}{2}} \sum_{k\in\mathbb{Z}} g_k \exp^{-ik\omega} \Psi\left(\frac{\omega}{2}\right) = m_1\left(\frac{\omega}{2}\right) \Psi\left(\frac{\omega}{2}\right) = \prod_{j=1}^{\infty} \Psi\left(\frac{\omega}{2^j}\right).$$

Most importantly, it can be shown that there exists a direct relationship between m_0 and m_1:

$$|m_0(\omega)|^2 + |m_1(\omega)|^2 = 1.$$

With these relationships we are now able to directly compute the wavelet filter coefficients g_k from h_k:

$$g_k = (-1)^k h_{1-k}. \tag{2.6}$$

Known as the quadrature mirror relation, this allows us to calculate a high-pass filter (containing detail) from a low-pass (averaging) filter.

2.3 Haar Wavelet

An often used example of a wavelet is the Haar Wavelet. First discovered by Alfred Haar (Haar, 1910), its simple nature and intuitive construction aid often in the explanation of wavelets.

The scaling function for the Haar Wavelet is a simple step function of the form

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

It can be verified quickly that the set of functions $\{\phi(x+k)\}_{k\in\mathbb{Z}}$ forms an orthonormal basis for V_0, as the function is non-overlapping due to the non-inclusive upper bound on the domain, and also $\int_{\mathbb{R}} \phi(x)\mathrm{d}x = 1$. Following Equation (2.2) with $j = -1$ we determine that the only relevant values of k are $k = -1, 0$ giving,

$$\phi(x) = h_0 2^{\frac{1}{2}} \phi(2x) + h_1 2^{\frac{1}{2}} \phi(2x+1) = h_0 \phi_{-1,0}(x) + h_1 \phi_{-1,1}(x).$$

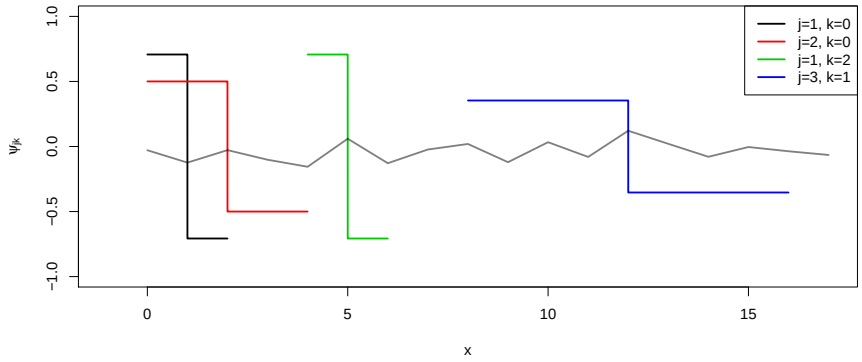


Figure 2.1: A number of Haar Wavelets over different scales j and position k superposed upon a small white noise series ($\mathcal{N}(0, 0.1^2)$) for clarity. Note that values of zero for the wavelets are not shown.

Given that we have already determined that integer translates of the haar scaling functions are non-overlapping, we can determine that $h_0 = h_1 = \frac{1}{\sqrt{2}}$. Following on now, using the quadrature mirror filter relationship given in Equation (2.6) we can now determine that the mother wavelet ψ has the following form

$$\psi(x) = \begin{cases} 1 & \text{if } 0 \leq t < \frac{1}{2}, \\ -1 & \text{if } \frac{1}{2} \leq t < 1, \\ 0 & \text{else,} \end{cases}$$

such that there is a set of wavelet coefficients $g_k = \{\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\}$. We can now translate and scale this function as per the set definition given in Equation (2.4). An example of a number of wavelets drawn up is given in Figure 2.1.

2.4 Wavelet Decomposition

As we are now able to construct wavelet filters sufficiently, given an appropriate scaling function, we now look at a transformation to decompose a function using the orthonormal wavelet bases. First we define the coefficients generated by the

inner product of the wavelet filter and scaling function as

$$c_{j,k} = < f(x), \phi_{j,k}(x) > = \int_{-\infty}^{\infty} f(x)\phi_{j,k}(x)\mathrm{d}x,$$
$$d_{j,k} = < f(x), \psi_{j,k}(x) > = \int_{-\infty}^{\infty} f(x)\psi_{j,k}(x)\mathrm{d}x.$$

Then, working from the scaling equation defined earlier in Equation (2.2), we are able to create an approximation to a function $f(x)$ at scale j:

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

which can be combined with the properties of the MRA given in Equation (2.1) and (2.3) to extend the approximation. This leads to

$$\begin{aligned}\hat{f}_j(x) &= \hat{f}_{j-1}(x) + \sum_{k\in\mathbb{Z}} d_{j,k}\psi_{j,k}(x) \\ &= \sum_{k\in\mathbb{Z}} c_{j,k}\phi_{j,k}(x) + \sum_{k\in\mathbb{Z}} d_{j,k}\psi_{j,k}(x) \\ &= \sum_{k\in\mathbb{Z}} c_{j_0,k}\phi_{j_0,k}(x) + \sum_{i=j_0}^{j}\sum_{k\in\mathbb{Z}} d_{i,k}\psi_{i,k}(x), \qquad j_0 \leq j.\end{aligned}$$

This is such that we can approximate a function at scale j by the information provided at a coarser level j_0 and the detail in between. If we then take an increasing amount of scales, such that $j \to \infty$ we converge upon

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

This has a direct discrete counterpart, such that we can decompose a series $y = \{y_0, y_1, \ldots, y_{n-1}\}$ of dyadic length ($\exists J \in \mathbb{N} : n = 2^J$) into 'smooth' components and a number of 'detail' components:

$$\{c_{J,0}, d_{1,0}, d_{1,1}, \ldots, d_{1,2^{J-1}-1}, d_{2,0}, \ldots, d_{2,2^{J-2}-1}, \ldots d_{J,1}\}.$$

2.5 Decimated Wavelet Decomposition

A direct calculation can be made for each of these components, however this can be very costly computationally. If however the scaling function satisfies the properties of an MRA, then we can use the pyramidal algorithm detailed in Mallat (1989). This algorithm uses the relations between each scale to calculate the discrete wavelet transform of a series with order $\mathcal{O}(n)$ computations. To see this, consider again the scaling relationships given in Equation (2.2) and see that we can rewrite the calculation of the 'smooth' coefficient

$$c_{j,k} = \int_{-\infty}^{\infty} f(x)\phi_{j,k}(x)\mathrm{d}x = \sum_{i\in\mathbb{Z}} h_i c_{j-1,i+2k} = \sum_{i\in\mathbb{Z}} h_{i-2k} c_{j-1,i},$$

such that it relies upon the previous scale and a translation of the original h_k filter. This applies similarly to the wavelet function,

$$d_{j,k} = \sum_{i\in\mathbb{Z}} g_{i-2k} c_{j-1,i}.$$

Note that there are half as many coefficients that can be computed at scale j than scale $j-1$. This is a consequence of the decimation that occurs between scales which is described further within Mallat (1989). Here we briefly overview it following similar notation.

Definition 2.3. *Given a series $y = \{\dots, y_{-1}, y_0, y_1, y_2, \dots\}$, define the application of filters $\mathcal{H}$ and $\mathcal{G}$ as*

$$(\mathcal{H}y)_k = \sum_{i\in\mathbb{Z}} h_{i-k} y_i, \qquad (\mathcal{G}y)_k = \sum_{i\in\mathbb{Z}} g_{i-k} y_i.$$

Definition 2.4. *Define the selection of each even element of a series as the binary decimation operator $\mathcal{D}_0$ such that $(\mathcal{D}_0 y)_j = y_{2j}$.*

If we now let the coarest information of the transform be the original series, such that $c_0 = y$, then we can describe the successive operations required to compute

the full transform by

$$c_j = \mathcal{D}_0 \mathcal{H} c_{j-1}, \qquad d_j = \mathcal{D}_0 \mathcal{G} c_{j-1}.$$

An example of this, using the Haar Wavelet given in Section 2.3 is shown in Figure 2.2, applied to the sawtooth function

$$f(x) = \begin{cases} 4x^2 + \frac{4}{3}x & \text{if } 0 \leq x \leq \frac{1}{3} \\ \frac{12}{8}x^2 + \frac{4}{8}x & \text{if } \frac{1}{3} < x \leq \frac{2}{3} \\ x^2 + \frac{4}{8}x - \frac{1}{2} & \text{if } \frac{2}{3} < x \leq 1 \end{cases}. \tag{2.7}$$

2.6 Non-Decimated Wavelet Decomposition

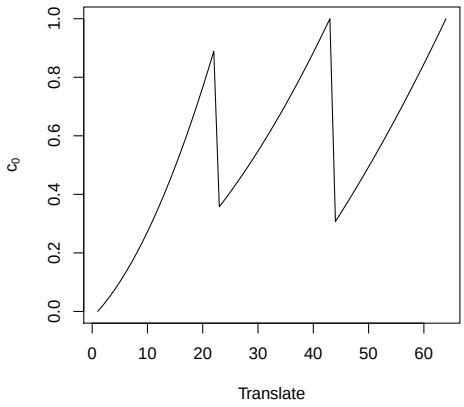
Whilst computationally efficient, the decimated wavelet decomposition has a main disadvantage that can be seen by merely shifting the data by a single point, as shown in Figure 2.2. In particular note the shift in magnitude at the lowest scales at half the length of the series. Due to the decimation process $\mathcal{D}_0$, shifting the series by a single point creates a complete new set of coefficients. This is known as the translation invariance property of the decimated wavelet transform.

To define the non-decimated transform, we must describe how to transform the filtering operations and decimation process.

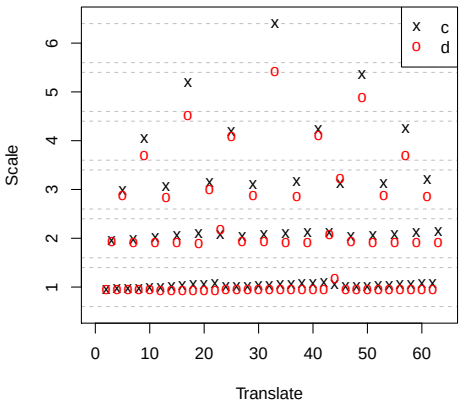
Definition 2.5. *Let the inverse of the binary decimation operator be such that upon application to a series $y = \{\dots, y_{-1}, y_0, y_1, \dots\}$, it adds zeros between each element. This is such that*

$$\mathcal{D}_0^{-1} y = \tilde{y}, \quad \textit{where} \quad \tilde{y} = \{\dots, 0, y_{-1}, 0, y_0, 0, y_1, 0, \dots\}.$$

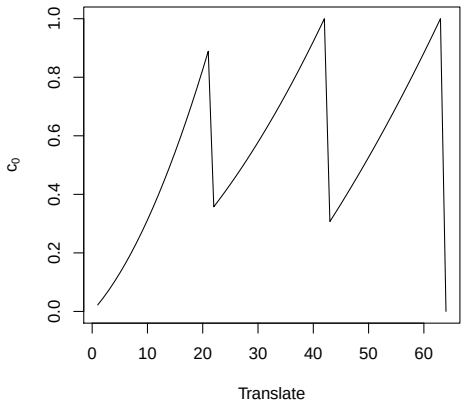
Definition 2.6. *The non-decimated wavelet transform can be computed by using*



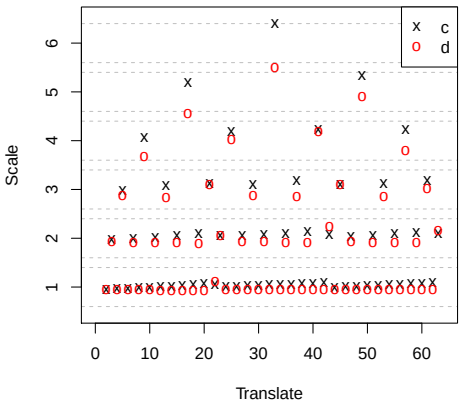
(a) Original series.



(b) c and d coefficients for each scale.



(c) Original series shifted by one point.



(d) c and d coefficients for each scale.

Figure 2.2: Wavelet transform of a sawtooth function. Note in Figures 2.2b and 2.2d that each coefficient has been scaled by the maximum value and is centered on a zero line, bounded above and below by 1 and -1 respectively shown by the dotted lines.

filters which are recursively related such that

$$\mathcal{H}^{[0]} = \mathcal{H}, \qquad \mathcal{H}^{[j]} = \mathcal{D}_0^{-1}\mathcal{H}^{[j-1]},$$
$$\mathcal{G}^{[0]} = \mathcal{G}, \qquad \mathcal{G}^{[j]} = \mathcal{D}_0^{-1}\mathcal{G}^{[j-1]}.$$

Then the non-decimated smooth and detail coefficients sets, $\bar{c}_j$ and $\bar{d}_j$ are computed by

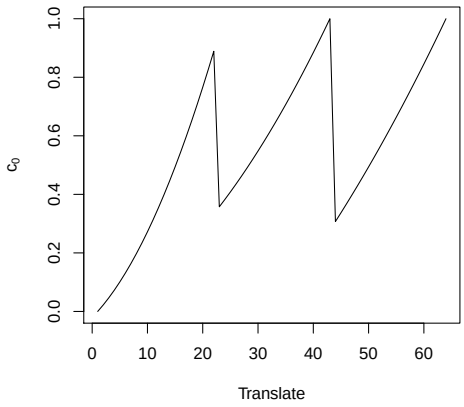
$$\bar{c}_j = \mathcal{H}^{[j-1]}\bar{c}_{j-1}, \qquad \bar{d}_j = \mathcal{G}^{[j-1]}\bar{c}_{j-1}.$$

This then creates an overcomplete transform which considers both even and odd decimations. As such, each coefficient set is of length 2^J. In contrary to the decimated wavelet transform, this is translation invariant, and often referred to as the stationary wavelet transform. Our original example given by Equation (2.7), is presented under a non-decimated wavelet transform, with a shifted version for comparison, in Figure 2.3.

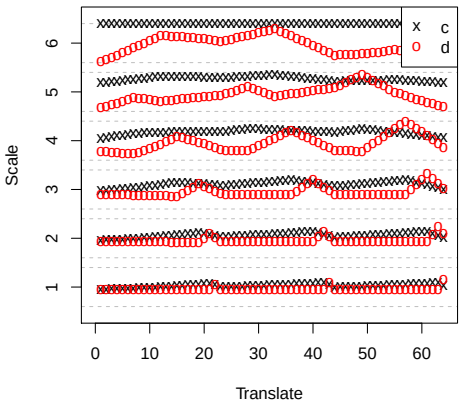
2.7 Locally Stationary Wavelet Processes

We must first define a time series, before defining Locally Stationary Wavelet Processes. A time series is formally defined as a collection of data over time, such that we have a series y_t for $t = 1, 2, \ldots, n$. A common assumption in the modelling of a time series is that it is stationary, such that its properties do not vary over time. This is often not valid, but a similar assumption of local stationarity, such that a segment of data is stationary, may often prove more useful. For a review of such processes and an overview of the literature surrounding them, Dahlhaus (2012) is recommended.

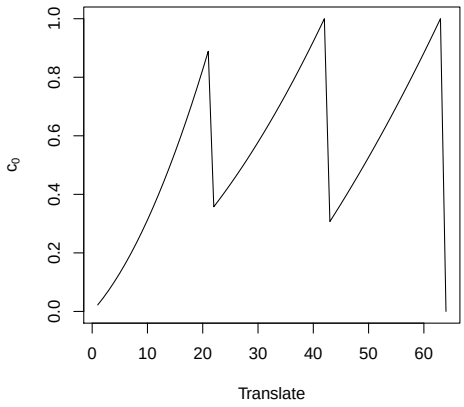
Wavelets, with their ability to localise particular characteristics across different scales and translations, provide a framework to model locally stationary wavelet processes, introduced in Nason et al. (2000).



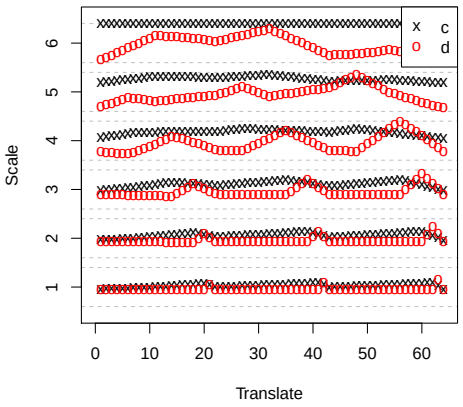
(a) Original series.



(b) c and d coefficients for each scale.



(c) Original series shifted by one point.



(d) c and d coefficients for each scale.

Figure 2.3: Non-decimated wavelet transform of a sawtooth function. Note in Figures 2.3b and 2.3d that each coefficient has been scaled by the maximum value and is centered on a zero line, bounded above and below by 1 and -1 respectively shown by the dotted lines.

Definition 2.7. *A time series is of the locally stationary wavelet processes class if it can be represented, in the mean-square sense, by a doubly indexed stochastic process* $\{X_{t,N}\}_{t=0}^{N-1}$ *(where* $N = 2^J \geq 1$*) such that*

$$X_{t,N} = \sum_{j=1}^{N} \sum_{k \in \mathbb{Z}} w^0_{j,k;N} \psi_{j,k}(t) \xi_{j,k}$$

where j*,* k*, are the scale and translation parameters seen previously and* ψ *is a family of discrete non-decimated wavelets. Further:*

1. *The components* ξ *are an orthonormal random increment sequence, such that* $\mathbb{E}(\xi_{j,k}) = 0$ *for all* j, k *and* $Cov(\xi_{j,k}, \xi_{l,m}) = \delta_{j,l}\delta_{k,m}$*.*

2. *There exists* $W_j(z) : [0,1] \to \mathbb{R}$ *which are Lipschitz continuous functions on* $z \in (0,1)$ *for each* j*. Each function* $W_j(z)$ *satisfies*

   (a) *The Lipschitz constants* L_j *are uniformly bounded in* j *and*

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

   (b) *A sequence of constants* C_j *exists such that for each* N

   $$\sup_k \left| w^0_{j,k;N} - W_j\left(\frac{k}{N}\right) \right| \leq \frac{C_j}{N}$$

   where for each j *the supremum is over* $k = 0, \ldots, N-1$ *and the constants* C_j *are bounded such that* $\sum_{j=1}^{\infty} C_j < \infty$*.*

   (c) $\sum_{j=1}^{\infty} |W_j(z)|^2 < \infty$ *uniformly in* z*.*

Note that there is a further variation to this definition where jumps are allowed within the amplitudes, given in Fryzlewicz and Nason (2006). From Definition 2.7 we are then able to construct a spectrum of the data, similarly to the Fourier spectrum that can be created by taking the modulus squared of the Fourier transform given in Equation (2.5).

Definition 2.8. *The locally stationary wavelet process spectrum is defined as*

$$S_j\left(\frac{k}{N}\right) = \left|W_j\left(\frac{k}{N}\right)\right|^2$$

for a locally stationary wavelet process $X_{t,N}$, where $j = 1, \ldots, J = \log_2(N)$ and $k = 1, \ldots, N$.

There exists a biased approximation to this spectrum through the non-decimated wavelet coefficients. This is defined as the local wavelet periodogram:

Definition 2.9. *The LSW process spectrum can be approximated by*

$$\hat{S}_j\left(\frac{k}{N}\right) = \mathbf{A}_j \left|\sum_{l=0}^{N-1} X_{t,N}\psi_{j,k}(l)\right|^2 = \mathbf{A}_j \left|d_{j,k}(l)\right|^2 \quad j = 1, \ldots, J.$$

The bias correction matrix $\mathbf{A}$ is defined within Nason et al. (2000).

2.8 Statistical Applications of Wavelets

Wavelets provide a strong basis for decomposing data into a succinct set of information, and providing the methodology to reconstruct it. As such there are several properties of interest to many statisticians that have led to research in a number of areas.

Most prominent is the process of denoising, such that we improve the signal-to-noise ratio of a piece of data by successfuly removing all or some of the noise. The use of wavelets for this problem was first introduced within Donoho (1993), where they presented methodology which thresholded wavelet coefficients which could then be reconstructed to return denoised data. This work was further built upon by extending the generation of wavelet coefficients through the use of a non-decimated transform in Coifman and Donoho (1995) and similarly whilst using bivariate thresholds in Bui and Guangyi Chen (1998). Indeed determining a suitable threshold to apply to the coefficients is studied closely, with procedures for an adaptive threshold given in Donoho and Johnstone (1995) and selecting using the

information from a non-decimated transform given within Gao and Bruce (1997). Research has also been conducted into the correlation between wavelet coefficients and the effect of this upon denoising in Cai and Silverman (2001) and He et al. (2008) where the interscale correlation is used to determine a suitable threshold. Further, these techniques can be used as a dimensional reduction tool as in Bruce et al. (2002) and Kaewpijit et al. (2003). These methods have been used in many areas such as ECG data (Singh and Tiwari, 2006), geospatial data (To et al., 2009) and medical imaging (Dansena and Dewangan, 2015). The literature of denoising is vast and thus we recommend Chen et al. (2013a) for further reading on this area.

Given the scope of such methodology, usage has been extended into many Machine Learning areas. In the area of pattern matching where the aim is to determine if similar behaviour is exhibited within or between series, wavelet coefficients are used in cardiac data in Senhadji et al. (1995). Similarly work from Du et al. (2006) thresholds coefficients to determine peaks within mass spectrums. Interest has also been drawn to using wavelets in a classification method, where contexts such as image annotation (Blume and Ballard, 1997) have used a Haar Wavelet Transform to identify similar images and identification of seizure patterns within EEG data (Panda et al., 2010). Work has also been conducted in clustering, such that we group data together by their properties, where as examples: Vlachos et al. (2003) uses multi-resolution properties of a wavelet transform to optimise a k-means clustering approach; and Misiti et al. (2007) uses a subset of wavelet coefficients to cluster directly onto. The scope of these applications and context is large and thus Li et al. (2002) and Abbas and Raina (2018) are recommended as good reviews of the area.

2.9 Other Wavelet Methods

Beyond the wavelet transformations detailed already, we must make note of a number of further wavelet transforms that have gained popularity. In particular

we draw attention to four areas: Bi-Orthogonal Wavelet Transforms; Multiple Wavelet Bases; Wavelet Lifting; and Wavelet Packet Transforms. Each of these is related in part to the descriptions already given and we explain them further here.

So far we have used the same wavelet to decompose a series, and then translated/scaled it into the next wavelet to continue the decomposition. Moving beyond this principle we encounter what are referred to as 'Second Generation Wavelets'. A key example of this are Bi-Orthogonal wavelets, which separate these two processes such that one wavelet is used in decomposition, and another informs the wavelet to be used in the following transform. These wavelets are specifically designed to be orthogonal only to each other, thus for two wavelets $\psi_{a,b}$ and $\tilde{\psi}_{c,d}$ we have

$$< \psi_{a,b}, \tilde{\psi}_{c,d} >= \delta_{a,c}\delta_{b,d}.$$

For more information on the construction and usage of Bi-Orthogonal wavelets an interesting article is given by Karoui and Vaillancourt (1994).

The use of second generation wavelets is further extended beyond Bi-Orthogonality into Multiple Wavelet Bases. Here rather than decomposing a series by the application of a vector of coefficients, this is extended into a matrix. This is such that we apply more than one scaling and smooth functions to the data. Originally introduced in Alpert (1993), appropriate pre-processing and post-processing must be used to prepare and combine the results of the decompositions to adequately describe the data. An example of such usage can be found in Geronimo et al. (1994) where symmetric scaling functions are used simultaneously.

Further use of the relationship between multiple wavelet bases can be found through Wavelet Lifting, or Wavelets on an interval. First shown in Sweldens (1998) the process extends the idea of non-decimated transforms, sparse representation of data and recalculating wavelets. Using both the even and odd decimations of the data, at each stage of the decomposition a preditive step occurs to determine the most suitable wavelet to continue decomposing, minimising over the sparsity of the return. This comes with many advantages, such as computational efficiency

gains and in-place memory requirements (such that a fixed size memory space is needed throughout). A strong review of the work conducted on this methodology can be found in Acharya and Chakrabarti (2006).

Looking beyond efficiency gains and more towards sparsity requirements, the Wavelet Packet Transform can be seen as a further generalisation of a wavelet transform. Rather than continually decomposing the smooth coefficients, wavelet packets look to exploit additional information that could be contained by decomposing the detail coefficients further. Each of these decompositions is then referred to as a 'packet' which approximate a certain part of the function in question. With such a selection of information, algorithms such as Best Bases (Coifman and Wickerhauser, 1992) can be used to select a subset of packets to give a sparse representation. For further information on the wavelet packet transform we direct the reader to Nason (2010).

Chapter 3

Long Memory and Changepoint Models: A Spectral Classification Procedure

Abstract Time series within fields such as Finance and Economics are often modelled using long memory processes. Alternative studies on the same data can suggest that series may actually contain a 'changepoint' (a point within the time series where the data generating process has changed). These models have been shown to have elements of similarity, such as within their spectrum. Without prior knowledge this leads to an ambiguity between these two models, meaning it is difficult to assess which model is most appropriate. We demonstrate that considering this problem in a time varying environment using the time varying spectrum removes this ambiguity. Using the wavelet spectrum we then use a classification approach to determine the most appropriate model (long memory or changepoint). Simulation results are presented across a number of models followed by an application to stock cross correlations and US inflation. The results indicate that the proposed classification outperforms an existing hypothesis testing approach on a number of models and performs comparatively across others.

3.1 Introduction

It is not often the case that a given data set has a known explicit model from which it is generated. Analysts will look to fit an appropriate model to such a series in the hopes of understanding the underlying mechanisms or to make predictions into the future. The models proposed are expected to be distinct in their properties such that there is a clear prevalence of a suitable model for the data. However, models with certain structural features have been known to have similar properties to other models (Granger and Hyung, 2004). This overlap will be here referred to as an 'ambiguity' between the models. This is such that either model may appear similar to one another in some metrics, but provide very different interpretations on the data generating process, and lead to different predictions into the future.

In this paper we consider the ambiguity between long memory and changepoint models. This ambiguity has been documented in fields such as Finance and Economics which are modelled using long memory models (Granger and Ding, 1996; Pivetta and Reis, 2007) and changepoint models (Levin and Piger, 2004; Starica and Granger, 2005). Thus it is reasonable to assert that there is an element of ambiguity between these two models. Following the discussion and in-depth analysis within Diebold and Inoue (2001), it has been shown that both models share some similar properties, especially within the spectrum. Often a decision on a model can not be made with the 'luxury' of prior knowledge, and as such assuming the data derives from either of these models comes at a risk of mis-specification.

Existing work in Yau and Davis (2012) conducts a hypothesis test to determine between the changepoint and long memory model. The authors choose to use the changepoint model as a null model with the justification that this is the more plausible model. However in some circumstances this may not be the case so it leads to the question as to which model should be the null model. It would be entirely feasible to choose the changepoint model as the null model, not reject H_0 and then flip to have the long memory model as the null model and also not reject H_0. This does not give a clear answer to the question of an appropriate model.

As an alternative this paper introduces a classifier, which places no such assumptions on which model is preferred. Instead the purpose of a classifier is only to give a measure of which category provides the best fit. In the context here, it can measure which model best describes a time series, without assuming that this model is where the data was originally generated from. Classification of time series has been previously used in Grabocka et al. (2012) and Krzemieniewska et al. (2014). It was shown in Yau and Davis (2012) that the autocorrelation function and periodogram of data generated from a changepoint model and a long memory model exhibit similar structures (i.e. slow decay in the autocorrelation and spectral pole at zero). However, if we consider a time-varying periodogram, then the stationarity of a long memory model can be seen (constant structure over time), whilst a changepoint model exhibits the piecewise stationarity expected (see for example Killick et al. (2013)). As the time varying spectrum shows evidence of a difference between these models, we use it as the basis for our classification procedure.

The structure of this article is as follows. The background and methods to our approach are given in detail in Section 3.2. A simulation study of the proposed classification method, with a comparison to the Likelihood Ratio Test from Yau and Davis (2012), can be found in Section 3.3. Applications of the classifier are then given using US price inflation and stock cross-correlations in Section 3.4. Finally, concluding remarks and a discussion is given in Section 3.5.

3.2 Methods

3.2.1 Changepoint and Long Memory Models

The aim of our method is to distinguish between data which arises from either a changepoint or a long memory model. To define these, we first define the general Autoregressive Integrated Moving Average (ARIMA) model, characterised by its Autoregressive (AR) parameters $\boldsymbol{\phi} \in \mathbb{R}^p$, Moving Average (MA) parameters $\boldsymbol{\theta} \in \mathbb{R}^q$

and the Integration (I) parameter $d \in \mathbb{N}$. For random variables $X_1, X_2, \ldots, X_n$ this is formally defined as,

$$\left(1 - \sum_{k=1}^{p} \phi_k B^k\right)(1-B)^d X_t = \left(1 + \sum_{k=1}^{q} \theta_k B^k\right)\epsilon_t$$

where ϵ_t are independent identically distributed as $N(0, \sigma^2)$ variables, and B is the backward shift operator such that $BX_t = X_{t-1}$ and $B\epsilon_t = \epsilon_{t-1}$. A variation of this, Autoregressive Fractional Integrated Moving Average (ARFIMA) is such that $d \in \mathbb{R}$, allowing it to be fractional. This modification allows long memory behaviour to be captured through dependence over a large number of previous observations.

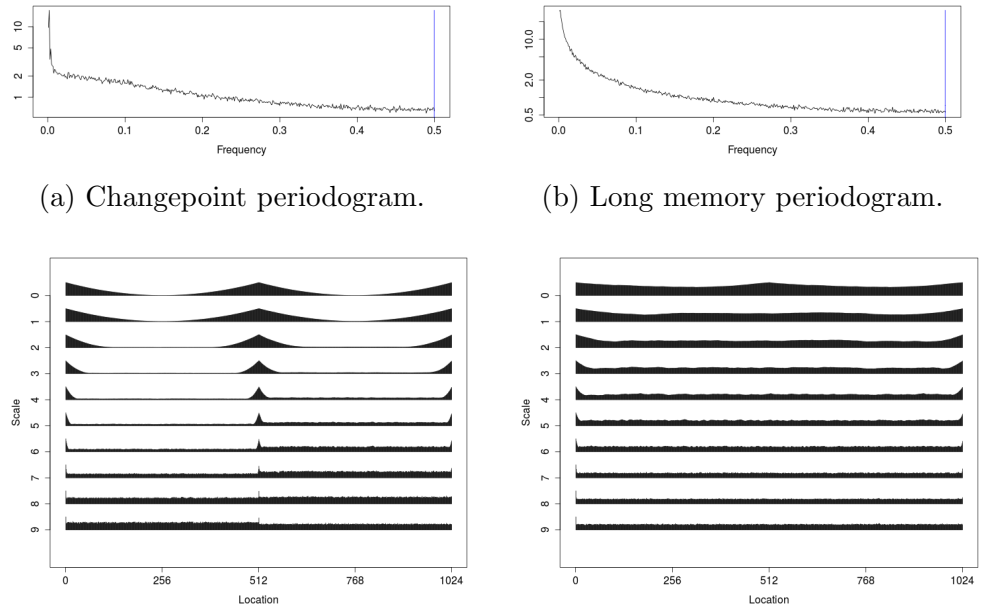
For the purpose of this paper, we define the changepoint and long memory models as:

$$X_t \sim \begin{cases} \mu_1 + \text{ARMA}(\boldsymbol{\phi_1}, \boldsymbol{\theta_1}) & \text{if} \quad t = 1, 2, \ldots \tau \\ \mu_2 + \text{ARMA}(\boldsymbol{\phi_2}, \boldsymbol{\theta_2}) & \text{if} \quad t = \tau+1, \tau+2, \ldots n. \end{cases} \tag{3.1}$$

$$X_t \sim \mu + \text{ARFIMA}(\boldsymbol{\phi}, d, \boldsymbol{\theta}) \quad t = 1, 2, \ldots, n \tag{3.2}$$

Note that we depict a single changepoint $\tau = \lfloor n\lambda \rfloor$ for notational ease, but the software we provide (see Section 3.5) contains the generalisation to multiple changes through use of the PELT algorithm Killick et al. (2012) and extending Equation (3.1) to include multiple τ. Other models such as ARCH models and Fractional Gaussian Noise (Molz et al., 1997) could also be used but we restrict our consideration to ARFIMA here. In the general case we allow $p, q \in \mathbb{N}$, but in the simulations and applications given in Section 3.3 and 3.4 we restrict their range for computational reasons. Further, we restrict the range of the fractional differencing parameter to $d \in (0, 0.5)$ and this resolves to a stationary model as will be necessary for the work that follows.

3.2.2 Wavelet Spectrum



(a) Changepoint periodogram. (b) Long memory periodogram.

(c) Changepoint wavelet spectrum. (d) Long memory wavelet spectrum.

Figure 3.1: Empirical periodogram and wavelet spectrum averaged over 500 realizations.

The ambiguity present between diagnostics of the competing models given in Equation (3.1) and (3.2) can cause issues in identifying the correct model. Figure 3.1 shows the average empirical periodograms from realisations of long memory (ARFIMA(0,0.4,0)) and changepoint (AR(1), $\lambda = 0.5$, $\phi_1 = 0.1$, $\phi_2 = 0.4$, $\mu_1 = 0$, $\mu_2 = 1$) models. It can be seen that the periodogram for the changepoint model has a pole at zero and shows similar behaviour to that of long memory.

Before discussing the wavelet spectrum, we provide a brief background to wavelets and the specific spectrum we propose to use.

Wavelets capture properties of the data through a location-scale decomposition using compactly supported oscillating functions. Through dilation and translation, a wavelet is applied across a number of a scales and locations to capture behaviour occurring over different parts of a series. Further information on them and their application can be found in Daubechies (1992) and Nason (2010). In this work we use the model framework of the Locally Stationary Wavelet process which provides a stochastic model for second order structure using wavelets as building blocks.

We follow the definition in Fryzlewicz and Nason (2006) for a Locally Stationary Wavelet (LSW) process.

Definition 3.1. *Define the triangular stochastic array* $\{X_{t,N}\}_{t=0}^{N-1}$ *which is in the class of LSW processes given it has the mean-square representation*

$$X_{t,N} = \sum_{j=1}^{\infty} \sum_{k} W_j \left(\frac{k}{n}\right) \psi_{j,k-t} \xi_{j,k},$$

where $j \in 1, 2, \ldots$ *and* $k \in \mathbb{Z}$ *are scale and location parameters respectively,* $\psi_j = (\psi_{j,0}, \ldots, \psi_{j,L_j-1})$ *are discrete, compactly supported, real-valued non-decimated wavelet vectors of support length* L_j*. If the* ψ_j *are Daubechies wavelets Daubechies (1992) then* $L_j = (2^j - 1)(N_h - 1) + 1$ *where* N_h *is the length of the Daubechies wavelet filter, finally the* $\xi_{j,k}$ *are orthonormal, zero-mean, identically distributed random variables. The amplitudes* $W_j(z) : [0, 1] \to \mathbb{R}$ *at each* $j \geq 1$ *are time varying, real-valued, piecewise constant functions which have an unknown (but finite) amount of jumps. The constraints on* $W_j(z)$ *are such that if* $\mathcal{P}_j$ *are constants representing the total magnitude of jumps in* $W_j^2(z)$*, then the variability of* $W_j(z)$ *is controlled by*

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

Note that we apply this framework to our models given in Equations 3.1 and 3.3 as both are stationary models under the assumption that $d \in (0, 0.5)$. Thus by Wold's Decomposition Theorem (Geary, 1956), which states that any stationary model can be represented as a Moving Average (MA) model with infinite parameters, we can represent the ARFIMA model as an MA(∞). Further, by Nason et al. (2000) it is shown that any MA model can be represented in as an LSW and thus are applicable here.

As in the traditional Fourier setting, the spectrum is the square of the ampli-

tudes and as such the Evolutionary Wavelet Spectrum can be defined as

$$S_j\left(\frac{k}{N}\right) = \left|W_j\left(\frac{k}{N}\right)\right|^2$$

which changes over both scale (frequency band) j and location (time) k.

Considering both scale and location, the two dimensions allow the differences between the proposed models to be seen. Examples of the differences in these spectra are given in Figure 3.1 for both the changepoint and long memory models. To interpret the wavelet spectrum: scale corresponds to frequency bands with high frequency at the bottom to low frequency at the top. Further details on the spectrum and its applicability can be found in Fryzlewicz and Nason (2006), Nason (2010) and Killick et al. (2013). Note that there is a clear difference between the wavelet spectra of the two models with the changepoint model being piecewise stationary (pre and post change), with the change occurring in the spectrum where the change occurs in the data. In contrast the long memory model remains flat across each scale and time reflecting the stationarity of the original series.

Due to the fact that the wavelet spectrum gives a distinction between the two models we propose to use this as the basis for our inference regarding the most appropriate model. Whilst the Fourier spectrum could be used here as in Janacek et al. (2005), we choose to use the Evolutionary Wavelet Spectrum. As shown in Figure 3.1 this is advantageous for characterising the non-stationarity changepoint data due to the Scale-Location transformation used. This is since the $W_j(z)$ are constant for stationary models, but for non-stationary models the break in the second order structure of the original data causes breaks in the wavelet spectra, as described in Cho and Fryzlewicz (2012).

In the next section we detail how to use the wavelet spectrum of the two models in a classification procedure.

3.2.3 Classification

Testing whether a long memory or changepoint model is more appropriate whilst under model uncertainty comes with the hazard of mis-specification. A formal hypothesis test places assumptions on the underlying model in both the null and alternative, but the allocation of the null is hazardous - should the changepoint model be the null or alternative? It would be entirely feasible to choose the changepoint model as the null model, not reject H_0 and then flip to have the long memory model as the null model and also not reject H_0. Given the absence of a clear null model, which result to proceed with is unclear. Instead it may be preferable to quantify the evidence for each model separately. A classification method such as the one proposed here gives a candidate series a measure of distance from a number of groups, which can then be used to select the most appropriate group.

In the previous subsection it was demonstrated that the wavelet spectrum can be used to distinguish the changepoint model from the long memory model, and the classifier proposed here builds on this. However, to begin a classification method must first 'teach' itself on the structure of the classes through sets of training data. These are data sets already determined to be in each category and are the basis for calculating the distances from each group. This previous knowledge allows for determination of patterns and features of each category (that are unique from other categories) for comparison to the candidate data set. A common example is the spam filter on mailboxes, which is trained on previous spam emails so that it can classify if a new email that arrives is spam or not. The decision is made by comparing it to a number of patterns already determined to be features in spam email for example, short messages or hidden sender identities. Further information on classification methods and training them can be found within Michie et al. (1994).

In our example we only have a single data set of length n, the classifier has no previous information to train on. To remedy this we create training data through

simulation. Given a candidate series we first fit the competing models in Equations (3.1) and (3.2) choosing the best fit for each model. For the changepoint model the best fit uses the ARMA likelihood within the PELT multiple changepoint framework to identify multiple changes in ARMA structure (Hyndman and Khandakar, 2008; Killick et al., 2012). When considering fitted long memory models, a number of ARFIMA models are fitted (Veenstra, 2012) and selection occurs according to Bayesian Information Criterion (following Beran et al. (1998)).

Following the identification of the best changepoint and long memory models, the training data is then simulated as (Monte Carlo) realisations from these, denoted by

$$\boldsymbol{X}_m^g = \left\{X_{i,m}^g\right\}_{i=1,2,\ldots,n} \quad m = 1, 2, \ldots, M. \qquad g = 1, 2.$$

where the group, $g = 1$ for changepoint simulations and $g = 2$ for long memory simulations, M is the number of simulated series and n is the length of the original series. Note that we are not sampling from the original series, we are generating realizations from the fitted models.

Now we have the training data and the observed data, denoted $\boldsymbol{X}^o$, a measure of distance of the observed data from each group is calculated. As discussed previously we will use a comparison of their evolutionary wavelet spectra as the distance metric. Before detailing the metric, we first define the estimated wavelet spectrum of the original series as

$$\hat{\boldsymbol{S}}^o = \left\{\hat{S}_k^o\right\}_{k=1,2,\ldots n*J}$$

where we remove the index over scale j by concatenating scales, hence $k = 1, 2, \ldots n * J$, where $J = \lfloor \log_2(n) \rfloor$. Similarly we define the spectra for each simulated series:

$$\hat{\boldsymbol{S}}_m^g = \left\{\hat{S}_{k,m}^g\right\}_{k=1,2,\ldots n*J}.$$

To obtain a group spectra, an average is then taken over the M simulated series at each position of each scale for each group,

$$\bar{\boldsymbol{S}}^g = \left\{ \frac{1}{M} \sum_{m=1}^{M} \hat{S}_{k,m}^g \right\}_{k=1,2,\ldots n*J}.$$

Based on these spectra the distance metric proposed is a variance corrected squared distance, across all spectral coefficients as proposed in Krzemieniewska et al. (2014),

$$D^g = \frac{M}{(M+1)} \sum_{k=1}^{n*J} \frac{(\hat{S}_k^o - \bar{S}_k^g)^2}{\sum_{m=1}^{M} (\hat{S}_{k,m}^g - \bar{S}_k^g)^2} \tag{3.3}$$

Note that the variance correction occurs within the denominator to account for potentially different variability seen across simulations for each group. This is modified from Krzemieniewska et al. (2014) to allow different variances within each group. The theoretical consistency of the classification was shown in Theorem 3.1 from Fryzlewicz and Ombao (2009) where the error for misclassifying two spectra $\{S_k^{(1)}\}_k$ and $\{S_k^{(2)}\}_k$ (whose difference summed over k is larger than CN) is bounded by $\mathcal{O}\left(N^{-1}\log_2^3 N + N^{1/\{2\log_2(a)-1\}-1}\log_2^2 N\right)$. However this result requires a short memory assumption that is clearly not satisfied for our long memory processes. Thus we prove a similar bound under the assumption that the spectra are created from ARFIMA processes. We first replicate the required assumptions from Fryzlewicz and Ombao (2009) for completeness:

Assumption 3.2. *(Assumption 2.1 from Fryzlewicz and Ombao (2009))*
The set of those locations z where (possibly infinitely many) functions $S_j(z)$ contain a jump is finite. In other words, let $\mathcal{B} := \{z : \exists j \lim_{u\to z^-} S_j(u) \neq \exists j \lim_{u\to z^+}\}$. We assume $B := \#\mathcal{B} < \infty$.

Assumption 3.3. *(Assumption 2.2 from Fryzlewicz and Ombao (2009))*
There exists a positive constant C_1 such that for all $j, S_j(z) \leq C_1 2^j$.

Theorem 3.4. *Suppose that Assumptions 3.2 and 3.3 hold, and that the constants $\mathcal{P}_j$ from Definition 3.1 decay as $\mathcal{O}(a^j)$ for $a > 2$. Let $S_j^{(1)}(z)$ and $S_j^{(2)}(z)$ be two*

non-identical wavelet spectra from Changepoint ARFIMA processes. Let $I_{k,N}^{(J)}$ *be the wavelet periodogram constructed from a process with spectrum* $S^{(1)}(z)$*, and let* $L_{k,N}^{(j)}$ *be the corresponding bias-corrected periodogram, with* $J^* = \log_2 N$*. Let*

$$\sum_{j,k} \left\{ S_j^{(1)}(k/N) - S_j^{(2)}(k/N) \right\}^2 = \mathcal{O}(N).$$

The probability of misclassifying $L_{k,N}^{(j)}$ *as coming from a process with spectrum* $S_j^{(2)}(z)$ *can be bounded as follows:*

$$P(D_1 > D_2) = \mathcal{O}\left(\log_2^2 N \left[N^{-1} + N^{\frac{1}{(2\log_2 a - 1)} - 1} \right] \right)$$

Proof. The proof is given in Appendix 3.6.1. □

A summary of the proposed procedure is given in Algorithm 1.

3.3 Simulation Study

To test the empirical accuracy of our proposed approach, simulations were conducted over a number of models. Here, these models are chosen over a number of parameter magnitudes and combinations to show the effectiveness of the approach outlined in Section 3.2. A number of these models also appear in Yau and Davis (2012) which uses a likelihood-ratio method to test the null hypothesis of a changepoint model. As part of their notation they introduce λ which represents the location in the series where the changepoint occurs, such that $\lambda = \frac{\tau}{n}$. Their results for these models are correspondingly given as a comparison.

For each model given in the tables below, 500 realisations of each model were generated and classified, using $M = 1000$ training simulations for each fit. For computational efficiency, the maximum order of the fitted models are constrained to $p, q \leq 1$. Three different time series lengths were computed for each model; $n = 512, 1024, 2048$. It is expected that as a series grows larger, more evidence of long memory features will become prevalent, and as such the effect of length of

Initialization:

$\mathbf{X} : \{X_i\}_{i=1}^{n}$ observed series.

n : Length of series

M : Number of bootstrap simulations

$\bar{\boldsymbol{S}}^1$, $\bar{\boldsymbol{S}}^2$: Empty Spectra 1, 2.

Algorithm:

1. Fit: $\mathcal{M}_1$ - best changepoint model (Equation (3.1)) to $\mathbf{X}$.
2. Fit: $\mathcal{M}_2$ - best long memory model (Equation (3.2)) to $\mathbf{X}$.
3. Calculate training spectra

   for $m = 1, 2, \ldots, M$ **do**
   Simulate n observations from $\mathcal{M}_1$, denote as $\mathbf{Y_1}$
   Calculate Evolutionary Wavelet Spectra $\hat{\boldsymbol{S}}_m^1$ of $\mathbf{Y_1}$
   Let $\bar{\boldsymbol{S}}^1 = \bar{\boldsymbol{S}}^1 + \hat{\boldsymbol{S}}_m^1$
   Simulate n observations from $\mathcal{M}_2$, $\mathbf{Y_2}$
   Calculate Evolutionary Wavelet Spectra $\hat{\boldsymbol{S}}_m^2$ of $\mathbf{Y_2}$
   Let $\bar{\boldsymbol{S}}^2 = \bar{\boldsymbol{S}}^2 + \hat{\boldsymbol{S}}_m^2$
   end
4. Calculate the average Evolutionary Wavelet Spectra for each group $\bar{\boldsymbol{S}}^1 = \frac{\bar{\boldsymbol{S}}^1}{M}$, $\bar{\boldsymbol{S}}^2 = \frac{\bar{\boldsymbol{S}}^2}{M}$.
5. Calculate Evolutionary Wavelet Spectrum of $\mathbf{X}$, $\hat{\boldsymbol{S}}^o$.
6. Compute the distance D_1, D_2, between $\hat{\boldsymbol{S}}^o$ and $\bar{\boldsymbol{S}}^1$, $\bar{\boldsymbol{S}}^2$ respectively (Equation (3.3)).

Output: Distances D_1, D_2.

Algorithm 1: Wavelet Classifier Algorithm

series on accuracy is investigated.

We have used $n = 2^J$ as the length of the series as the wavelet decomposition software in Nason (2016b) requires that the series transformed is of dyadic length. This is not a desirable trait as data sets come in many different sizes. Thus we overcome this using a standard padding technique described in Nason (2010) that adds 0's to the left of each series until the data is of length 2^J. The extended wavelet coefficients are then removed before calculating the distance metric. Finally, we use the Haar Wavelet across all simulations.

3.3.1 Changepoint Observations

	Model Parameters						Classification Rate ($n =$)			Y&D LR ($n =$)	
Ref	λ	μ	ϕ_1	θ_1	ϕ_2	θ_2	512	1024	2048	500	1000
1	0.5	1	0.1	0.3	0.4	0.2	1.00	1.00	1.00	0.99	0.97
2	0.5	2	0.1	0.3	0.4	0.2	1.00	1.00	1.00	0.95	0.93
3	0.5	1	0.1	0.3	0.8	0.2	1.00	1.00	1.00	0.97	0.99
4	0.5	2	0.1	0.3	0.8	0.2	1.00	1.00	1.00	0.94	0.95
5	0.7	1	0.1	0.3	0.8	0.2	1.00	1.00	1.00	0.94	0.94
6	0.7	2	0.1	0.3	0.8	0.2	1.00	1.00	1.00	0.91	0.93

Table 3.1: Changepoint observations results with Likelihood Ratio results comparison taken from Yau and Davis (2012).

For the changepoint models we used the simulations given in Yau and Davis (2012). Table 3.1 gives the parameters used in Equation (3.1) along with the correct classification rate. The results show that if the data follows a changepoint model then we have a 100% classification rate. A movement of the changepoint to a later part of the series, as in models 5 and 6, does not appear to have an effect upon classification rates unlike for the Yau and Davis method. It is not really a surprise that we are receiving 100% classification rates as if a changepoint occurs then it is a clear feature within the spectrum.

It should be noted that as the Yau and Davis method is a hypothesis test we would expect results around 0.95 for a 5% type I error.

3.3.2 Long Memory Observations

	Model Parameters				Classification Rate			Y&D LR Power
Ref	ϕ	d	θ_1	θ_2	$n = 512$	$n = 1024$	$n = 2048$	$n = 500$
7	-0.8	0.1	0.6		0.42	0.61	0.79	0.63
8	-0.8	0.2	0.6		0.56	0.83	0.94	0.97
9	-0.8	0.3	0.6		0.66	0.90	0.96	0.98
10	-0.8	0.4	0.6		0.75	0.88	0.96	0.96
11	0.1	0.1	-0.8		0.74	0.87	0.95	0.08
12	0.1	0.2	-0.8		0.84	0.96	0.99	0.09
13	0.1	0.3	-0.8		0.89	0.98	1.00	0.15
14	0.1	0.4	-0.8		0.88	0.99	1.00	0.32
15	0.1	0.1	0.8		0.54	0.78	0.90	
16	0.1	0.2	0.8		0.61	0.85	0.91	
17	0.1	0.3	0.8		0.62	0.87	0.95	
18	0.1	0.4	0.8		0.63	0.87	0.98	
19	0.6	0.1	-0.8		0.33	0.45	0.65	
20	0.6	0.2	-0.8		0.38	0.62	0.83	
21	0.6	0.3	-0.8		0.44	0.63	0.87	
22	0.6	0.4	-0.8		0.39	0.59	0.86	
23	0.0	0.1	0.7	-0.7	0.94	0.97	0.99	
24	0.0	0.2	0.7	-0.7	1.00	0.99	1.00	
25	0.0	0.3	0.7	-0.7	1.00	1.00	1.00	
26	0.0	0.4	0.7	-0.7	1.00	0.99	1.00	

Table 3.2: Long memory observations results with Likelihood Ratio results comparison taken from Yau and Davis (2012).

In contrast to the changepoint models, the classification of a long memory model is expected to be less clear. This is due to the variation within the wavelet spectrum of long memory series that could be interpreted as different levels and hence a changepoint model would be more appropriate. To demonstrate the effect of the classifier on long memory observations, a larger number of models were considered. We simulated long memory models with differing levels of long memory as measured by the d parameter, values close to 0 are closer to short memory models and values close to 0.5 are stronger long memory models (values > 0.5 are not stationary and thus not considered).

The results in Table 3.2 give an indication of the accuracy of the classifier in a number of different situations. Overall, as the length of the time series increases we see an increase in classification accuracy. This is to be expected as evidence of

long memory will be more prevalent in longer series. Similarly as we increase the long memory parameter d from 0.1 to 0.4 we improve the classification rate.

Some interesting things to note include, when there are strong AR parameters (ϕ) such as models 7-10 and 19-22 we require longer time series to achieve good classification rates. However, in contrast if there are strong MA components as in the remaining models the classifier performs better. A larger effect is found when the MA parameter is negative, seen through models 11-14 where the classifier performs strongly even at $n = 512$. This effect is further exemplified by models 23-26 which include a further MA parameter and achieve near 100% classification at $n = 512$. Here the maximum used p, q was 2.

Comparing our results to that of Yau and Davis we note that the opposite performance is seen. For the likelihood ratio method there is high power for models with strong AR components and poor performance for strong MA components. Notably the strong MA performance is much worse than our method on the strong AR components.

3.4 Application

To further demonstrate the usage of our approach, two applications to real data are given in this section. The first is an economics example based on US price inflation and this is followed by financial data on stock cross-correlations. A sensitivity analysis was conducted over the possible maximum values of p, q. It was found that no additional parameters were required beyond maximum $p, q = 4$, thus these results are presented here.

3.4.1 Price Inflation

US price inflation can be determined using the GDP index. The dataset used here is available from the Bureau of Economic Analysis, based on quarterly GDP indexes, denoted P_t, from the first quarter of 1947 to the third quarter of 2006 (227

data points). Price inflation is calculated as $\pi_t = 400\ln(P_t/P_{t-1})$ (thus $n = 226$). A plot of the inflation is given below in Figure 3.2a. Studies of the persistence of this data have been conducted to determine the level of dependence within the series. A high amount of persistence, indicating long memory, was found in Pivetta and Reis (2007). However Levin and Piger (2004) found a structural break, which when accounted for showed the series to have low persistence, indicating the presence of changepoints with short memory segments. Applying our classification approach to this series will give an additional indication as to which model is statistically more appropriate.

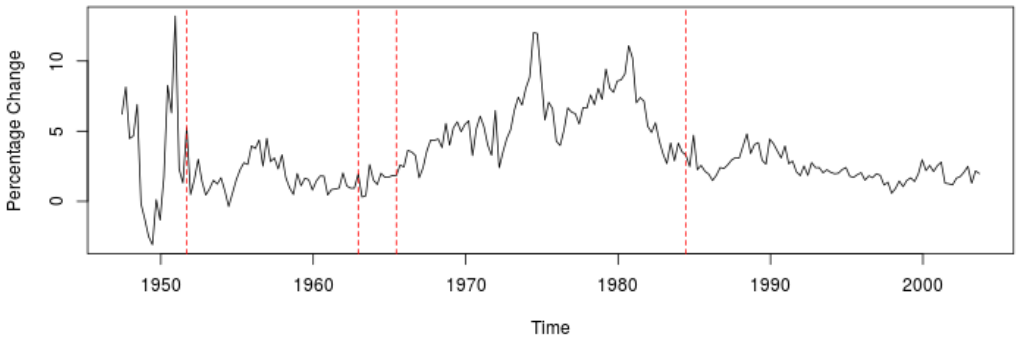
The parameters of the fitted changepoint and long memory models are given in Table 3.3. Diagnostic autocorrelation and partial autocorrelation function plots are given in Figure 3.3. The level shifts are given in respect to their position in the series, but correspond to 1951 Q3, 1962 Q4, 1965 Q2, 1984 Q2. The classifier returns a changepoint classification for this series.

3.4.2 Stock Cross Correlations

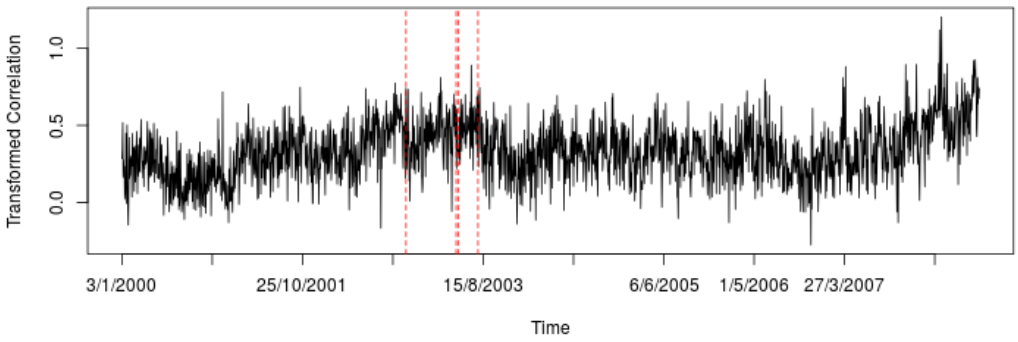
Stock Cross Correlation data has been obtained from the supplementary material of Chiriac and Voev (2011). The data consists of Open to Close stock returns for 6 companies from January 1^{st} 2001 to 30^{th} July 2008 ($n = 2156$). The data is first transformed using a Fisher Transformation, then correlations are calculated between each stock. Here analysis will look at the correlation between American Express and Home Depot.

This data has been analysed previously by Bertram et al. (2013) to determine between fractional integration (long memory behaviour) and level shifts and is given in Figure 3.2b. Parameters for the models fitted by the algorithm are also in Table 3.3. It can be seen that one of the AR coefficients is close to 1 indicating an element of non-stationarity, however we conducted a test of stationarity on this segment using the `locits` R package (Nason, 2016a) which implements the test of stationarity from Nason (2013) (no rejections) and also the `fractal` R

package (Constantine and Percival, 2016) which implements the Priestley-Subba Rao (PSR) test (Priestley and Rao, 1969) (time varying p-value 0.061). This coupled with autocorrelation and partial autocorrelation function plots given in Figure 3.4 means we conclude that the segment is stationary. Here the estimated changepoints at times 715, 841, 847 and 896 correspond 15/12/2002, 20/04/2003, 26/04/2003 and 14/06/2003. The distance scores given by the classifier indicate a strong preference for long memory over changepoints. This result stands against that found in Bertram et al. (2013) which indicated a preference for a model with similarly 4 changepoints. The difference is likely due to the fact that in Bertram et al. (2013) the changepoint model does not contain any short memory dependence and we have shown here that if that short memory structure is correctly taken into account within the sub-series then the series shows greater evidence of long memory properties.



(a) Time series of US Price Inflation.



(b) Time series of the Cross Correlations of American Express and Home Depot.

Figure 3.2: Real Data Examples

	Long Memory Model	Score	Changepoint Model	Score
Data				
Inflation	ARFIMA$(4,0,0)$ $\boldsymbol{\phi} = (0.30, 0.20, 0.20, -0.16)$ $d = 0.31$ $\mu = 3.29$		AR(1) $\phi = 0.68, \quad \tau = 18, \quad \mu = 0$	Segment 1
			AR(1) $\phi = 0.63, \quad \tau = 63, \quad \mu = 1.70$	Segment 2
			ARMA$(2,1)$ $\boldsymbol{\phi} = (-0.43, -0.33)$ $\theta = 0.47, \quad \tau = 73, \quad \mu = 1.69$	Segment3
			AR(1) $\phi = 0.82, \quad \tau = 149, \quad \mu = 5.43$	Segment 4
			AR(3) $\boldsymbol{\phi} = (0.60, 0.19, 0.18)$ $\tau = 226, \quad \mu = 0$	Segment 5
		40147		**38750**
Stock	ARFIMA$(1,0,1)$ $\boldsymbol{\phi} = 0.30$ $\boldsymbol{\theta} = 0.58$ $d = 0.47, \quad \mu = 0.35$		ARMA$(4,4)$ $\boldsymbol{\phi} = (-0.05, -0.002, 0.07, 0.94)$ $\boldsymbol{\theta} = (0.17, 0.14, 0.04, -0.83)$ $\tau = 715, \quad \mu = 0.31$	Segment 1
			AR(1) $\phi = 0.29, \quad \tau = 841, \quad \mu = 0.42$	Segment 2
			ARMA$(2,1)$ $\boldsymbol{\phi} = (0.09, 0.09)$ $\theta = -0.58, \quad \tau = 847, \quad \mu = 0.30$	Segemnt 3
			MA(1) $\theta = -0.72, \quad \tau = 896, \quad \mu = 0.50$	Segment 4
			ARMA$(3,1)$ $\boldsymbol{\phi} = (1.12, -0.06, -0.07)$	Segment 5
		12128	$\theta = -0.90, \quad \tau = 2156, \quad \mu = 0.35$	5348047

Table 3.3: Model fits and scores for US Inflation (Inflation) and Stock Cross-Correlations (Stock). Bold scores are the minimum. Each segment ending at τ is separated by a dotted line.

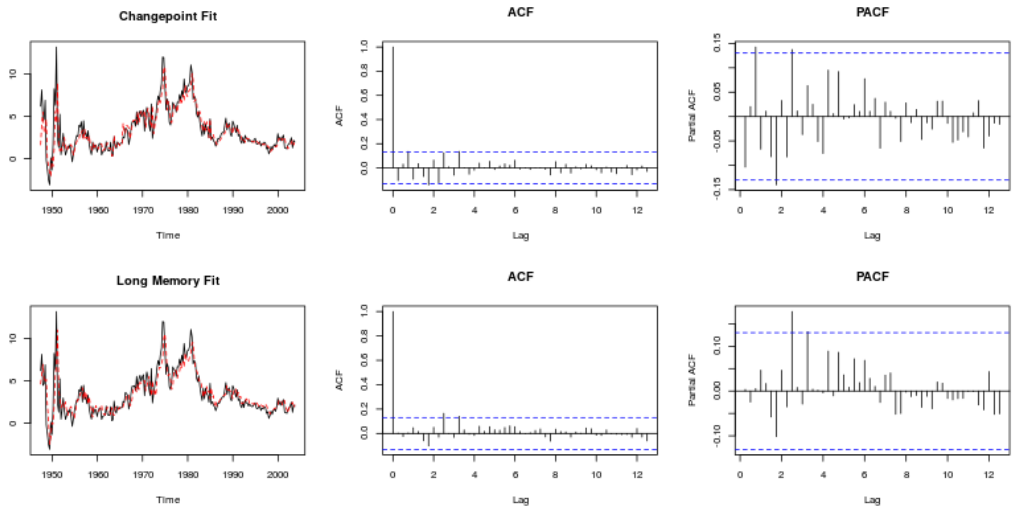


Figure 3.3: Inflation diagnostics. (Top) Left: Original data with fitted changepoint model; Middle: Autocorrelation function of changepoint model residuals; Right: Partial autocorrelations of changepoint model residuals. (Bottom) Left: Original data with fitted long memory model; Middle: Autocorrelation function of long memory model residuals; Right: Partial autocorrelations of long memory model residuals.

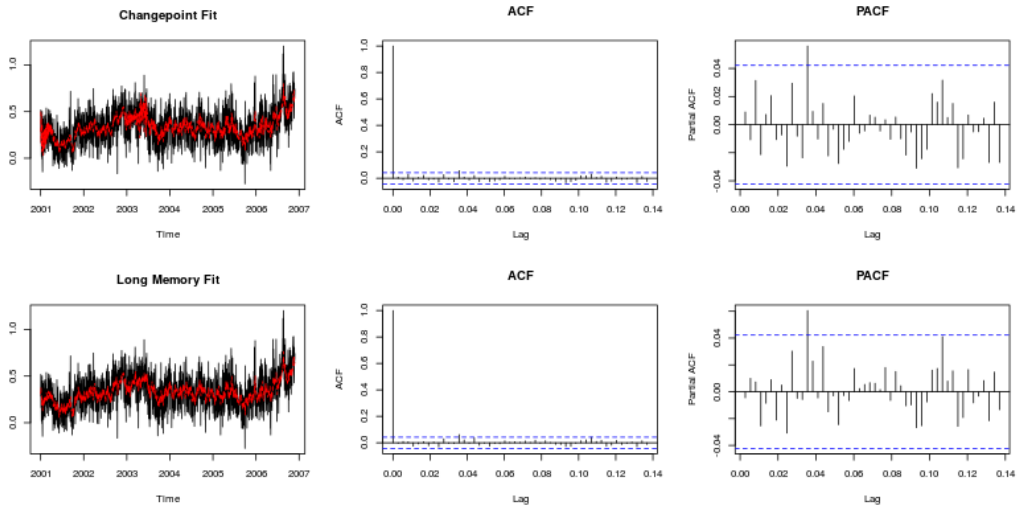


Figure 3.4: Stock diagnostics. (Top) Left: Original data with fitted changepoint model; Middle: Autocorrelation unction of changepoint model residuals; Right: Partial autocorrelations of changepoint model residuals. (Bottom) Left: Original data with fitted long memory model; Middle: Autocorrelation function of long memory model residuals; Right: Partial autocorrelations of long memory model residuals.

3.5 Conclusion

The wavelet classification process presented within this paper provides the user a distinct choice over a number of proposed models, and when explicitly applied to an ambiguity such as long memory or a changepoint as in Section 3.3, it provides an additional piece of information to aid decision making. The accuracy of the classifier over a number of simulated models has been presented within Section 3.3 and applied to data from the Financial and Economic fields in Section 3.4.

The Evolutionary Wavelet Spectrum provides a representation of non-stationarity which is lacking in the commonly used (averaged over time) spectrum. This gives an advantage when drawing comparisons between non-stationary and stationary series, since the wavelet spectrum may appear substantially different. Quantifying this visual difference allows for a direct comparison between the series and each proposed model.

The variance-corrected squared distance metric used in the proposed classifier has been demonstrated to be quite accurate under the ambiguity of long memory and changepoint models. It is particularly effective at identifying changepoint models correctly, as the results in Table 3.1 demonstrate. It was noted that there is relatively lower variation between the simulations generated for the changepoint than the long memory model, which reduces the distance metric significantly even though it is variance corrected.

As mentioned in Section 3.1 there are many series that can be found in fields such as Economics and Finance which show evidence of the ambiguity investigated here. This classification is not intended to propose a final model for these series, but instead give additional information, treated perhaps as a diagnostic. This could be to begin investigation of a series, or to confirm a previously found model fit. As this is not a formal test, the lack of assumptions allows for more flexibility in how the classification can be used. This work however is not restricted only to the ambiguity mentioned here, further work could extend it to determine between other features, such as local trends and seasonal behaviour or combining the behaviour

of both models i.e., a long memory model with a changepoint.

An aspect not covered in this paper is the precise form of ARMA and long memory models in the LSW paradigm, i.e. how the model coefficients relate to the $W_{j,k}$'s. This is an interesting area for future research which would cement the LSW model as an encompassing model but is beyond the scope of this paper.

An R package (`LSWclassify`) is available from the authors that implements the method from the paper.

Acknowledgements The authors would like to acknowledge the helpful comments of an Associate Editor, 2 reviewers and Dr. Matthew Nunes. This research was conducted whilst Ben Norwood was a graduate student funded by the Economic and Social Research Council in collaboration with the Office for National Statistics.

3.6 Appendix

3.6.1 Proof of Theorem 3.4

Proof. We replicate the steps of the proof within the Appendix of Fryzlewicz and Ombao (2009) up until (A.6), where following this step the short memory condition is used. To briefly summarise previous steps,

$$\begin{aligned} P(D_1 - D_2 > 0) &= P(X - t > 0) \leq E(\tilde{X}^2)/t^2, \\ &\qquad \text{(by Chebyshev's Inequality)} \\ E(\tilde{X}^2) &=: I + II, \\ I &\leq CJ_0J^* \sum_{j=-1}^{-J0} \sum_{i=-1}^{-J^*} 2^{i+j} E\left\{b_{i,j}^2\right\}, \\ E\left\{b_{i,j}^2\right\} &=: 2A + 2B. \end{aligned}$$

Here D_1, D_2 are the squared distance metrics from each respective group, t is the time index across the original series and C is a generic constant. Definitions for

components $\tilde{X}, t, b_{i,j}, I, II, A$ and B can be found in the Appendix of Fryzlewicz and Ombao (2009). Note that there is a negative indexing across the scales j occurring in the proof which is analogous to the positive indexing used elsewhere in the work, however this more closely aligns with the original proof.

Component A is where we alter the proof. Recall $I^{(i)}_{k,N}$ is the wavelet periodogram at a fixed scale i, at position k with total length N, with $d^{(i)}_{k,N}$ the wavelet coefficient corresponding to it through the relationship $I^{(i)}_{k,N} = \left(d^{(i)}_{k,N}\right)^2$. We continue the proof from (A.6) using the ARFIMA assumption instead. Following from above (A.6):

$$\begin{aligned} A &= E\left\{\sum_{k=1}^{N}\left\{I^{(i)}_{k,N} - E\left(I^{(i)}_{k,N}\right)\right\}c_{j,k}\right\}^2 \\ &\leq 2^{2j}\sum_{k,k'=1}^{N}\left|\mathrm{cov}\left(I^{(i)}_{k,N}, I^{(i)}_{k',N}\right)\right| \\ &= 2^{2j}\sum_{k,k'=1}^{N} 2\mathrm{cov}^2\left(d^{(i)}_{k,N}, d^{(i)}_{k',N}\right) \\ &\quad \text{(by Isserli's Theorem)} \end{aligned} \tag{3.4}$$

Jensen (2000) gives bounds for the covariance of wavelet coefficients;

$$\begin{aligned} \mathrm{cov}\left(d^{(m)}_{k,N}, d^{(j)}_{n,N}\right) &= \mathrm{C}_1|\alpha|^{2d-1-2M} + R_{2M+1} \\ \alpha &= 2^{m-j}k - n, \quad m \geq j \\ |R_{2M+1}| &\leq \mathrm{C}_2|\alpha|^{2d-2-2M}, \end{aligned}$$

where $M \geq 1$ is the number of vanishing moments in the wavelet used. Using

$|\alpha| = |2^{i-i}k - k'| = |k - k'| \geq 1$ and substituting into Equation (3.4):

$$\begin{aligned}
A &= 2^{2j+1} \sum_{k,k'=1}^{N} (\mathrm{C}_3|\alpha|^{2d-1-2M} + R_{2M+1})^2 \\
&= 2^{2j+1} \sum_{k,k'=1}^{N} |\mathrm{C}_3|\alpha|^{2d-1-2M} + R_{2M+1}|^2 \\
&\leq 2^{2j+1} \sum_{k,k'=1}^{N} (\left|\mathrm{C}_3|\alpha|^{2d-1-2M}\right| + |R_{2M+1}|)^2 \\
&\leq 2^{2j+1} \sum_{k,k'=1}^{N} (\mathrm{C}_4|\alpha|^{2d-1-2M} \\
&\quad + \mathrm{C}_5|\alpha|^{2d-2-2M})^2
\end{aligned}$$

As $|\alpha|^{2d-2-2M} \leq |\alpha|^{2d-1-2M}$ we have:

$$\begin{aligned}
A &\leq 2^{2j+1} \sum_{k,k'=1}^{N} \left(\mathrm{C}_6|k - k'|^{2d-1-2M}\right)^2 \\
&= 2^{2j+1} \sum_{k,k'=1}^{N} \mathrm{C}_7 \frac{1}{|k - k'|^{-2(2d-1-2M)}} \\
&= 2^{2j+1} \sum_{s=1}^{N-1} (N - s)\mathrm{C}_7 \frac{1}{s^{-2(2d-1-2M)}} \\
&= 2^{2j+1}\mathrm{C}_7 \left[N \sum_{s=1}^{N-1} \frac{1}{s^{-2(2d-1-2M)}} \right. \\
&\quad \left. - \sum_{s=1}^{N-1} \frac{1}{s^{-4(d-M)+1}} \right]
\end{aligned}$$

Given that $|d| < 0.5$ and $M \geq 1$ then $4 < -2(2d - 1 - 2M) = \delta_1$ and $3 < -4(d - M) + 1 = \delta_2$. We can then replace the sums using the definition of Generalised Harmonic Numbers and their convergence:

$$\begin{aligned}
H_{n,m} &= \sum_{k=1}^{n} \frac{1}{k^m} \\
H_{n,m} &= \mathcal{O}(1) \quad \text{as } n \to \infty \qquad (m > 1).
\end{aligned}$$

Thus

$$A \leq 2^{2j+1}\mathrm{C}_7 \left(NH_{N-1,\delta_1} - H_{N-1,\delta_2}\right) \qquad = 2^{2j+1}C_7\mathcal{H}_N,$$

where $\mathcal{H}_N = NH_{N-1,\delta_1} - H_{N-1,\delta_2}$. Returning to consider (A.4) from Fryzlewicz and Ombao (2009), we find a bound for component I, where $J_0, J^* = \log_2 N$ and $\Delta = \frac{1}{(2\log_2 a - 1)}$:

$$\begin{aligned}
I &= \mathrm{C}_8 J_0 J^* \sum_{j=-1}^{-J_0} \sum_{i=-1}^{-J^*} 2^{i+j} \left[2^{2j+1}\mathrm{C}_7 \left(NH_{N-1,\delta_1}\right.\right. \\
&\quad \left.\left. - H_{N-1,\delta_2}\right) + N^{1+\Delta}2^j\right] \\
&= \mathrm{C}_8 \log_2^2 N \sum_{j=-1}^{-J_0} 2^j \left[2^{2j+1}\mathrm{C}_7\mathcal{H}_N\right. \\
&\quad \left. + N^{1+\Delta}2^j\right]\left(1 - 2^{J*}\right) \\
&= \mathrm{C}_8 \log_2^2 N \left[\sum_{j=-1}^{-J_0} \mathrm{C}_7 2^{3j+1}\left(1 - 2^{J*}\right)\mathcal{H}_N\right. \\
&\quad \left. + \sum_{j=-1}^{-J_0} 2^{3j}\left(1 - 2^{J*}\right)N^{1+\Delta}\right] \\
&= \mathrm{C}_9 \log_2^2 N \left(1 - 2^{-J*}\right)\mathcal{H}_N \sum_{j=-1}^{-J_0} 2^{3j+1} \\
&\quad + \mathrm{C}_8 \log_2^2 N \left(1 - 2^{-J*}\right)N^{1+\Delta} \sum_{j=-1}^{-J_0} 2^{3j} \\
&= \mathrm{C}_9 \log_2^2 N \left(1 - 2^{-J*}\right)\mathcal{H}_N \frac{2}{7}\left(1 - 2^{-3J_0}\right) \\
&\quad + \mathrm{C}_8 \log_2^2 N \left(1 - 2^{-J*}\right)N^{1+\Delta}\frac{1}{7}\left(1 - 2^{-3J_0}\right) \\
&= \log_2^2 N \left(1 - N^{-1}\right)\left(1 - N^{-3}\right) \\
&\quad \left[\mathrm{C}_{10}\mathcal{H}_N + \mathrm{C}_{11}N^{1+\Delta}\right] \\
&= \log_2^2 N \left(1 - N^{-3} - N^{-1} + N^{-4}\right) \\
&\quad \left[\mathrm{C}_{10}\mathcal{H}_N + \mathrm{C}_{11}N^{1+\Delta}\right] \\
&\leq \mathrm{C}_{12} \log_2^2 N \left(\mathcal{H}_N + N^{1+\Delta}\right)
\end{aligned}$$

Following this, using results in Fryzlewicz and Ombao (2009) the probability of misclassification is:

$$P(X > t) = \mathcal{O}\left(\log_2^2 N \left[N^{-1} + N^{\Delta - 1}\right]\right).$$

□

Chapter 4

Evolution of Seasonal Adjustment Methods

4.1 Introduction

It is well known that many economic variables are affected by the time of year. For example, it is often the case that during the winter, unemployment will fall as businesses look to take temporary staff for the Christmas period. This similarly affects sales across businesses, and ultimately may contribute to a rise in Gross Domestic Product (GDP). It would be unfair to present such a rise in a key economic indicator with the knowledge of such seasonal movements without first explaining what effect should be expected throughout the year. Such an effect is referred to as a seasonal component, and its estimation and consequent removal from a series is known as Seasonal Adjustment.

The aim of the following report is to provide a timeline of the development of seasonal adjustment methods and to review their methodology. As part of this report, an overview of each method discussed is given using a standard template for ease of comparison, followed by a bibliography. The act of seasonally adjusting a series from empirical to adjusted can involve a number of steps beforehand known as pre-treatments. These pre-treatment methods look to identify for example,

structural breaks, outliers, calendar effects (such as trading day or Easter for example). For the purpose of this report, only the method of estimating and removing a seasonal component is given, however if any such pre-treatment is given within the work this is mentioned on the individual method reports within Section 4.6.

Before reviewing seasonal adjustment methods, it is useful to review the decomposition of a seasonal series. This was significantly formalised in Persons (1919) into a number of main components and is given as:

$$\text{Additive: } y_t = T_t + C_t + S_t + I_t \quad \text{or Multiplicative: } y_t = T_t * C_t * S_t * I_t. \quad (4.1)$$

Here T_t is referred to as the trend level, and represents the general movement of the series, considered the long-term growth. C_t is the cyclical component, representing cycles which are longer than a year and often represent the 'business cycle'. S_t is the seasonal component, which has a similar pattern each year and represents changes in frequencies less than a year (monthly for example). Finally I_t is the irregular component, the unobservable events and variation that occur throughout. Often the trend and cycle components are grouped together giving the standard form:

$$\text{Additive: } y_t = T_t + S_t + I_t \quad \text{or Multiplicative: } y_t = T_t * S_t * I_t.$$

Methods of estimating and extracting S_t form a wide and varied field of study. Within the early 20th Century, parametric methods included using multiple regression (Mendershausen, 1939) to model different periods, or harmonic analysis (Beveridge, 1921) to define cyclical components. A method of forecasting with exponential weights (Holt, 2004) (a reprint from 1957) could also be used to seasonally adjust. Non-parametric methods such as the link relatives method (Persons, 1919), moving averages (Federal Reserve Board, 1922) (accredited to F. R. Macaulay), successive arithmetic means (Kuznets, 1933), seasonal indices (Shiskin,

1942) and graphical estimates (Spurr, 1940) became popular techniques, providing effective and quick methods in a time when calculations were completed by clerical assistants. With the introduction of computing, more complex calculations could be completed efficiently. This lead first to an earlier moving average method soon becoming automated (Shiskin and Eisenpress, 1957) allowing a large amount of seasonal adjustment to be undertaken. Computation also aided a multivariate approach to moving averages on monetary series (Young, 1992). With calculations becoming more automated, more complex parametric methods were developed including varying harmonic calculation used to determine a slowly changing seasonal component (Hannan, 1964), the general linear model (Henshaw, 1966) a combination of harmonic regression and moving averages (Burman, 1965), mixed multiplicative-additive regression (Durbin and Murphy, 1975), moving medians and polynomials (Cleveland et al., 1979) and a composite of a number of applied filters (Hylleberg, 1986). As development grew further, it was found that modelling of stochastic variables (ARIMA) gave a close approximation to the non-parametric approach (Cleveland and Tiao, 1976), showing the common ground between the two approaches, where they would both later be combined (Dagum, 1980). Frequency domain extraction methods (Burman, 1980) were also developed, and a method allowing each component to vary stochastically in a regression model (Havenner and Swamy, 1981). Following that, local regression techniques (Cleveland et al., 1990) and the development of unobserved components being modelled using Markov properties (Harvey, 1990) were created. Now, methods such as variations to state space models (Tripodis and Penzer, 2004), modifications to the popular X-11 algorithm (McElroy, 2010), analysis of eigenvectors and values (Chen et al., 2013b) and wavelets (Stachura, 2014) are being investigated.

The methods briefly mentioned above will be covered in more detail in Sections 4.2, 4.3 and 4.4. Within those sections a number of parametric, non-parametric and semi-parametric (crossover) methods will be outlined and compared respectively. In Section 4.5 the report concludes on methods of seasonal adjustment and the

difficulties within the field. More details on the methods can be found in the overview sheets in Section 4.6 where further reading is given.

4.2 Non-Parametric Methods

The approach of non-parametric methods is to place little to no assumptions on the underlying form of the data. This is such that there is no requirement for the data to follow any particular statistical model. In a field where there may only be a limited amount of prior information on an economic series, reducing the assumptions on the data can be seen as a more unbiased approach. However, with few assumptions on the data, a longer series is necessary for accurate smoothing to occur.

The following section describes a number of non-parametric methods. Firstly the early use of seasonal indexes, medians and moving averages is described, along with how they are modified and their issues. This leads into the use of exponentially weighted averages as a seasonal adjustment tool. Following this, a more efficient graphical approximation is detailed, with accuracy issues mentioned. Next, a multivariate approach to a composite series is given. Lastly two developing methods are given to conclude the section, that of singular spectrum analysis and a multi-resolution approach using wavelets.

One of the earliest full methodological approaches to seasonal adjustment came from the 'link-relative' method (Persons, 1919). The decomposition within the method aimed to remove influence from cycle and seasonal variation, beyond that of the usual seasonal pattern, to estimate trend and seasonal components robustly. This is done by relating seasonal indexes to their previous values, known as 'link-relatives'. These link relatives can be calculated, for monthly seasonal indexes $z_{i,j}$

(over months and years i, j respectively) as:

$$z_{i,j}^{LR} = \begin{cases} 100 & \text{if } i,j = 1 \\ \frac{z_{i,j}}{z_{1,j-1}} & \text{if } i = 12 \\ \frac{z_{i,j}}{z_{i-1,j}} & \text{else.} \end{cases} \qquad \forall i = 1, 2, \ldots, 12. \quad \forall j = 1, 2, \ldots, n/12.$$

A median of these values would then be taken, allowing for greater protection from the influence of extreme values. However, the method assumed that a large amount of data was available (15 years) and was a complex procedure to implement in a time of clerical aid for calculations. An overview sheet for this method can be found on page 79. Seasonal indexes where further used in a model which did not distinguish between trend and cycle. Rather than compute medians of link-relatives, the method instead looked at successive larger averages of the indexes to find a stable seasonal component (Kuznets, 1933). This also aimed to reduce the effect of extreme observations, but assumed that there was a stable seasonal component. Further details of this method can be found in on page 80.

Moving averages, known for their smoothing properties, were a much more readily available method. The use of different lengths and weights leads to distinctive features being extracted from a series. A moving average filter is defined as the following:

$$M_{\boldsymbol{\theta},s}^{f}(y_t) = \sum_{i=s}^{f} \theta_{(i-s+1)} y_{t+i}$$

for a given series y_t, weighting coefficients $\boldsymbol{\theta}$ and start and finishing points s, f respectively. These can be centred (such that $f - s + 1$ is odd) or uncentered ($f - s + 1$ even); centred moving averages can also be symmetric in that the weights are reflected around the centre point. Uncentred averages can be centred to a time t by applying another moving average to the results of the filter.

Longer length filters are used to estimate a long running trend in a series, whereas shorter filters can be used to estimate short fluctuations such as those of a seasonal pattern. Work on optimal filter coefficients began within the actuarial

field where Henderson (1916) detailed a 13-point moving average ideal for trend estimation, giving both symmetric and asymmetric versions. Due to the endpoint issue, where there are not enough data points at the terminal ends of a series, asymmetric weights must be used. This can lead to large revisions when new data is available, as the estimation of the components will not be as accurate at the endpoints. However, this issue can be reduced by using forecasting methods, but this was implemented much later (Dagum, 1980). Further work on a number of a different moving averages, including the optimal 43-term filter (Macaulay, 1931a), can be found in Macaulay (1931b), Bongard (1960) and Kenny and Durbin (1982).

The first automated moving average process was that of Census Method I (Shiskin and Eisenpress, 1957) which approximated the procedure set out in Barton (1941) in use at the time by the Federal Reserve. This method used a combination of equally weighted 12 and 5 month moving averages for estimation of the trend, and 5 month moving averages for seasonal indices. An overview of this method is given on page 85.

A combination of moving averages and medians have also been applied to a series using an iterative algorithm (Cleveland et al., 1979). Using repeated medians the method looks to improve the robustness of moving medians, whilst employing moving averages to aid selection of the relevant components. These are used in collaboration with methods such as 'splitting' and 'midmeans'. The algorithm itself involves a large amount of steps and there are necessary extension rules for when a moving average cannot smooth the endpoints. An overview sheet for this method can be found on page 92.

However, as previously mentioned, smoothed endpoints of a series cannot be as accurately estimated through moving averages due to the limited information to pass to the filter. As such asymmetric weights or forecasts have to be used in their place, of which forecasts have been shown to reduce revisions in collaboration with parametric methods (Huyot et al., 1986), this is given later in Section 4.4. This issue has to be counter-balanced against the need for certain lengths in a

filter to reduce the impact of a trend estimation capturing elements of seasonality, and vice versa. A study into the gain functions, showing which components are captured, of filters was undertaken in O'Gorman (1982). Whilst the filters provide an efficient and robust estimation of a general movement, these issues must be kept in mind when choosing which is most appropriate for a series.

Another form of weighted average which was originally used for forecasting purposes is the exponentially weighted moving average (Winters, 1960). This is a continuation of a method originally proposed by Holt (2004)(reprint from 1957) in which forecasts were drawn from a seasonal series. Through use of an recursive average, a value at any time is a combination of: all previous values with deteriorating weights; previous trend estimates; and previous seasonal factors. This is in contrast to the moving average approach which requires both future and past values, removing the need for asymmetric filtering. A seasonally adjusted and detrended series y_t^* is expressed as

$$y_t^* = A\frac{y_t}{F_{t-L}} + (1-A)(y_{t-1}^* + R_{t-1}) \tag{4.2}$$

where F_t, R_t represent the seasonal and trend respectively, which each have their own recursive equation. Further details on this method and its implementation as a form of seasonal adjustment can be found on page 86.

Whilst this method appears intuitive in that it uses all the data previously available at any point, it can also be hampered by the same strength. Without prior robust estimation and consequent removal of extreme values the fitted recursive scheme can lead to large and long lasting variations from the true estimate of the seasonal and trend. However, this can make the system very adaptable to change within the series, if it is not sudden, and as such it is effective in capturing evolving seasonality. Although, consideration must be given to the length of previous values to be used in estimation, and if a static weighting variable (A in Equation (4.2)) is suitable throughout the series. Further details on exponential smoothing and recent developments can be found in Gardner (1985), Gardner

(2006) and Svetunkov and Kourentzes (2016). Variations of different smoothing techniques have also been developed, such as the weighted ratio-to-moving-average procedure outlined in Shiskin (1942), described further on page 84.

Before the time of computers, repeated use of smoothing techniques such as moving averages and exponential weights could prove time consuming, where for example exponential weights would have to be capped at a certain length. Following this, a graphical approximation to the iterative use of smoothing techniques was proposed. Such approximations are outlined in Spurr (1937), who advocate the use of only small strips of paper to compute a trend and seasonal component with 'neglible' error. The method, which would prove more inefficient today, showed only minor deviations from a calculated moving average filter, but showed large gains in efficiency. It is outlined on page 81. However, the procedure was criticised for the amount of subjectivity involved. Later a variation was detailed which replaced the most subjective steps with calculation of moving averages, at the cost of time (Spurr, 1940). This however, is a non-issue given the computational power we have access to today.

Computational aid also allows for a greater approach to dealing with composite series, those made up of smaller series but forming a greater seasonal series. A common filter is applied to all of these series with the aim of balance and fairness, using symmetric moving averages on page 97. Endpoints are accounted for by use of sequentially smaller filters where necessary. This approach was typically in use for monetary series, with previous knowledge on a relationship between the macro and micro series. This information however leads to the assumption that the true seasonal pattern of the macro series has been accounted for through the large amount of data, as such it is only typically useful where such relationships are known. Further details regarding this multivariate approach can be found in the overview sheet on page 97.

Looking ahead, a developing non-parametric method used within primarily the geophysical sciences is one of Singular Spectrum Analysis (Chen et al., 2013b).

The method looks at the eigenvectors and eigenvalues of the covariances between the series and aims to group them according to their periodic behaviour. This is performed according to three criteria, which can appear subjective, but are more formal than say the graphical approximation. There are a number of variants of this method in other work for example, Golyandina and Shlemov (2015). A description of the Singular Spectrum Analysis method is given on page 102.

A second developing non-parametric method is a multi-scale decomposition of a series using wavelets (Stachura, 2014). Through use of a pre-defined wavelet function, a series can be de-constructed over different scales and analysed separately. This allows periodicities to be seen more clearly and the coefficients from the transform can be used to decompose it the series into multiple components. These components are then grouped into seasonal and trend, similarly to singular spectrum analysis, however the determination of the decomposition leads to the grouping of components. The choice of wavelet however, is left to the users discretion and this can lead to poor specification of the components if not adequately selected. This method is described further on page 103.

Each non-parametric approach looks to allocate certain features in a series to components without an explicit model to explain them. Use of moving medians allows for greater robustness against extreme items rather than the moving-average approach, which is hampered by terminal points in a series, and the user must be careful on the specification of the filter coefficients else they may also remove periodic features. In contrary however, the exponential smoothing approach requires much less specification, and has no issue with endpoints. This unfortunately makes the method more static when dealing with changes in the series. More subjective methods such as that of the graphical approximation are able to account for changes within the seasonal patterns. However, the level of subjectivity within this method is not suitable for the reproducible world of today. Developing methods look to be using multi-level approaches to decomposition, increasing the level of complexity but allowing for greater analysis over multiple scales.

The following section contrasts the non-parametric methods described above with parametric methods which require a model assumption to be placed on the data generating process.

4.3 Parametric Methods

In contrast to non-parametric methods in Section 4.2, parametric methods allow for a direct interpretation of the data. Assuming that the models fitted to the data are correct (or at least adequate), the parameters and relationships within the model can be used to explain the movements within the different components and allow for forecasting methods under the assumption of a specific model. Prior knowledge of a series' behaviour can aid in the discovery of an appropriate model significantly. However, should the model be incorrectly specified, this can drastically affect estimation, interpretation and forecasts. As such, any fitted model needs to be rigorously tested as part of any parametric methodology.

In this section, a number of parametric methods are detailed and compared. To begin, a regression approach is outlined which looks at how the seasonal component can be modelled by sine and cosine waves, known as harmonic regression or trigonometric seasonality. An adaptation of this model to account for both additive and multiplicative seasonality is also given. Regression is then extended, attempting to explain the seasonal component in terms of potential covariates. A more general approach follows, using fixed degree models to explain movements within the same period (month on month for example), and a method which is used in collaboration with moving averages. Following this, a description of the use of ARIMA models is outlined in the context of seasonal adjustment. Structural time series models are then given, and how they can be set up in state space form. Lastly a developing method is given, a variation of the structural approach which considers sub-periods to test seasonal heteroscedasticity.

Early regression techniques looked to model seasonality by visualising the seasonal component as a combination of sine and cosine waves. This was captured

in Beveridge (1921) who used harmonic analysis from a clerically calculated periodogram. Significant periodicities could then be identified by prominent peaks p_i^* which could be then be implemented into seasonal factors:

$$s_t = \left[\sum_{i=1}^{k} a_{p_i^*} \cos \frac{2\pi t}{p_i^*} + b_{p_i^*} \sin \frac{2\pi t}{p_i^*} \right] \tag{4.3}$$

where $a_{p_i^*}, b_{p_i^*}$ can be calculated through a Fourier Transformation. More details on the method of harmonic regression can be found on page 83. Peak detection methods have developed since publication of the method, see for example Davis (1941), Granger et al. (1964) and McElroy and Holan (2009). A variation of this model has been proposed such that the components can be modelled stochastically as an autoregressive procedure. This is such that the models of the parameters for seasonal factors defined in Equation (4.3) are assumed to deviate from their overall mean with set variance. This method is detailed on page 94.

Extraction from the spectral domain for harmonic regressors has also been adapted for detailing a slowly changing seasonal (Hannan, 1964). By specifying a dependence variable, the user can set the length of time that they expect seasonality to evolve into a new structure. Time-varying harmonic regressors are then iteratively calculated, varying slowly with the period. However, the model can only capture variation which is slow, quick changes can not be accurately captured. Further details on the method can be found in an overview sheet on page 87.

Further development of harmonic regression looked to explain relationships between the variables in the basic structural model in Equation (4.1) (Durbin and Murphy, 1975). Rather than assuming an additive or multiplicative decomposition, both can be simultaneously modelled. A testing procedure is outlined to determine which type of relationship most suited to the data. Harmonic factors are then fitted with local scaling factors to aid moving seasonality. However this is only up to a degree of two. An overview of this method, including the testing procedure, can be found on page 90.

The regression model was later specified to include covariates to explain seasonal behaviour. This raised the issue of which and how many covariates should be included in a model. For example, patterns in weather will drive crop yields, and as mentioned in Section 4.1 social conventions such as Christmas will affect employment and output. These relationships have been used in an attempt to explain seasonality in a regression model (Mendershausen, 1939). The model proposed for the seasonal components of a series is:

$$s_k = \bar{s}_k + \sum_{i=1}^{m} \alpha_i x_i + \sum_{j=1}^{n} \beta_j y_j, \quad k = 1, 2, \ldots, p. \tag{4.4}$$

where $\bar{s}_k$ is the average over that period across the series, with meteorological and social covariates x_i, y_j respectively. Possible meteorological variables may include average temperature or daily rainfall, whilst social covariates may include the number of trading days or moving holidays (such as Easter and Ramadan etc.). The effect of such variables would appear intuitive, but their relationship may not be as simple as linear. Whilst modifications to Equation (4.4) would accommodate more complex relationships a consideration in the efficiency of finding these relationships should be made. If however there is prior knowledge about the series and possible covariates then this should, where possible, be efficiently introduced into the model. It must also be noted that in the model proposed above, the seasonal factors were not constrained to sum to unity over a yearly period, but over the course of the series. This is contrary to a popular assumption that the seasonal should not add or take away from the series over a yearly period. An overview sheet for this model and how to fit it is given on page 82.

Block-wise regression techniques are also employed with aid from moving averages (detailed in Section 4.2) for endpoints (Hylleberg, 1986). Seasonal estimation is computed using least squares over each block, before compositing the series together, using a number of moving averages to estimate the seasonal factors close to the end of a series. However, due to the nature of algorithm in use, only the seasonally adjusted series is returned. An overview sheet of this method can be

found on page 100.

Within economic series, it is not often that a relationship between covariates and the seasonal component will explain near all variation, due to the large number of possible events that may occur. As such the regression framework was further extended to the general linear model which attempts to model each period difference (each month across all years for example) through fixed degree models (Henshaw, 1966). The model is such that

$$y_t = \sum_{j=1}^{12} \sum_{\nu=0}^{s} \lambda_{\nu j} \Upsilon_{jt}^{\nu} + \sum_{\nu=s+1}^{c} \lambda_{\nu} t^{\nu} + u_t \tag{4.5}$$

where s, c are the degree of the seasonal and non-seasonal components respectively. More details of the method, including the rearrangement of Equation (4.5) into component form is given on page 89. Although the degree expected for these components is low (the paper implies it is often the seasonal which is lowest) it is an issue of trial and error to find the correct degree such that the errors u_t are uncorrelated white noise.

The method of modelling such correlations is the approach of the popularised ARIMA (Auto Regressive Integrated Moving Average) models popularised by Box and Jenkins (1976). These models describe series in relation to their past values only, unlike the covariate approach which uses external factors. The generalised seasonal ARIMA$(p, d, q)(P, D, Q)_m$ model for a demeaned series y_t is:

$$(y_t - \mu) = \frac{(1 + \theta_1 B + \cdots + \theta_q B^q)(1 + \Theta_1 B^m + \cdots + 1 + \Theta_Q B^{Qm})\epsilon_t}{(1 - \phi_1 B + \cdots + \phi_p B^p)(1 - \Phi_1 B^m + \cdots + \Phi_P B^{mP})(1 - B)^d (1 - B^m)^D} \tag{4.6}$$

as mentioned in Section 4.1. It was found that using such models to estimate the components of an economic series closely approximates one of the procedures used within the X-11 algorithm (Cleveland and Tiao, 1976). The method used within this work is detailed on page 91. A model derived in this work, often referred to as the 'Airline Model', is a common seasonal ARIMA model used for

seasonal adjustment of a series. However, these models require a level of pre-treatment to be able to accurately estimate their coefficients, such as accounting for structural breaks or other non-stationary behaviours that cannot be resolved through differencing. Note also that the model assumes that once the data has been sufficiently modelled by ARIMA the remaining component is stationary white noise, which with evolving data may not necessarily be true.

A more descriptive variation of ARIMA modelling is that of the recently introduced structural time series models (Harvey, 1990). These models allow direct description of each component stochastically. Structural models use the Kalman filter (Kalman, 1960) initially set up in State Space Form which directly describes the expected behaviour of given components. This allows the user to describe any features they expect to find in their series. Whereas non-stationary behaviour has to be pre-treated when using ARIMA models solely, these can be included within the state space form and estimated through the Kalman Filter. An overview of a general approach to setting up the components for dummy seasonal variables and harmonic seasonality can be found on page 96.

Extensive work has been developed within structural models, the monograph Harvey (1990) is a key reference here. Further work is contained within Hamilton (1994), Penzer (2006) and Durbin and Koopman (2012). Although the Kalman filter was originally designed for use within Engineering, it is finding increasing use within economic time series. However, whilst the Kalman Filter is efficient in identifying the coefficients of proposed models, this is only the case if the model is correctly specified and convergence is not guaranteed otherwise. Therefore this relies on a low level understanding of the process to adequately describe the possible models.

Structural time series models have been developed recently such that they are used to model each sub-period (yearly for example) separately (Tripodis and Penzer, 2004). This allows for more significant changes within the seasonal pattern on a sub-period basis. By testing the models proposed for each sub-period after

estimation using the Kalman Filter, seasonal heteroscedasticity can be determined. However, this leads to more testing procedures before the final estimate of the seasonally adjusted series can be obtained. An overview of this method can be found on page 99. Unlike regression models, this periodic structural model can not only capture an evolving seasonality, but also a structural change.

The level of description of the components varies throughout these models. Regression models can be used to give a high amount of information if the correct covariates are used. However, it is not often the case that these relationships can be found, and a more cyclical interpretation may be more representative. This however is still more information than given by ARIMA models, which are only modelled on previous values of the series. Although, the extension to structural models does lead to a greater understanding of the behaviour of the whole of the series, not just the components, if specified sufficiently.

4.4 Semi-Parametric Methods

The methods given within this section detail procedures which look to use both a modelling and smoothing procedure as part of seasonal adjustment. This is such that a model has to be estimated to explain the data, but the signal extraction that occurs is based on a non-parametric approach.

Four of the methods listed in this section use the capabilities of ARIMA models to enhance previous methods. Firstly, they are used to aid harmonic regression through extrapolation and use of a frequency response function of a filter. Next they are used to extrapolate, but instead to aid use of moving average filters in X-11. Following that, a method which uses the spectrum of a fitted ARIMA model is used to create filters to extract components. The next method uses a local form of regression to smooth a series in an iterative and robust algorithm. Lastly a developing method is given which is similar in using ARIMA models to create filters, but they are applied in a fashion similar to the X-11 algorithm.

ARIMA modifications on the harmonic regression approach given in Equation

(4.3) can be found within Burman (1965). Here trigonometric seasonal factors are determined using an extrapolated series from ARIMA models and smoothed using a chosen filter from a list of candidates. This technique uses the frequency response function of a moving average filter as part of determining the seasonal factors, where this function is often used to show how a moving average impacts upon certain periodicities. However it is pointed out in the work that the method of extrapolation used here can be hazardous when dealing with highly varying series. An overview of this model is on page 88. The use of ARIMA models here allows for more accurate use of a moving average filter, which can produce inaccurate results when smoothing terminal points.

The issue of terminal points is directly addressed however in X-11-ARIMA, proposed by Dagum (1980) and officially implemented and extended by the US Census Bureau as X-12-ARIMA (Findley et al., 1998), amongst other changes such as the introduction of regression variables for calendar effects. Within X-11-ARIMA, the issue of smoothing terminal points within a series is tackled using ARIMA models to back/forecast values, rather than the use of asymmetric filters. X-12-ARIMA implements ARIMA models similarly but also allows for regression variables to be introduced such that $(y_t - \mu)$ is replaced by $(y_t - x_t\boldsymbol{\beta})$ within Equation (4.6). These regression variables can be such that they account for trading days and moving holidays for example, a pre-treatment before the series is modelled. An overview sheet is given for X-11-ARIMA on page 95, however further detail on their evolution can be found in Ladiray and Quenneville (2012), Findley et al. (1998) and Findley (2005).

Following from the ARIMA modelling approach, a signal extraction procedure was outlined using the proposed models spectrum in Burman (1980). This fits a seasonal ARIMA model as in Equation (4.6) and calculates the spectrum, decomposes it according to constraints on the level of variation within each components. This decomposition can then be used to calculate filters for extraction. Details on how the decomposition is created can be found on page 93. Note that this method

constrains by minimising the amount of variation attributed to the seasonal and trend component, providing less flexible descriptions of these components. This is in contrast to ARIMA modelling alone however, which aims to minimise the variance of the irregular where possible.

Regression techniques have also been modified to collaborate with moving averages to smooth a series (Cleveland et al., 1990). By fitting a polynomial to a local segment of a series (such as all January values for example) an approximate value of seasonal components are estimated and then smoothed appropriately. However, the degree of these polynomials is constant across all monthly series, assuming that any evolution of the seasonal pattern will be of consistent degree across the board. Although this translates to a smoothing procedure, a parametric explanation is first given to the movements in the periodic values. Robustness is also attempted through use of weighting more extreme values to remove any harmful effects they may have on estimation, which is calculated iteratively. More details on this method are given on page 98. The weighting functions used in determination of the local regression model has the same issues with terminal endpoints as in the moving average case, but a different extrapolation approach.

A developing semi-parametric method is to follow a similar procedure to the smoothing process of X-11 but using a model based approach instead (McElroy, 2010). This process estimates the series as an ARIMA model and uses the coefficients of the model to determine filters which are iteratively applied in a similar fashion to X-11. Motivation behind such a procedure is a bias that is often not explored, created during the exponential and logarithmic transformations taken on multiplicative series (Burman, 1980). This method has been summarised on page 101. However it is noted in the work that this method could suffer from a lack of convergence, perhaps if the model is misspecified, a weakness shared amongst all ARIMA models.

4.5 Conclusion

This report has aimed to give a review of a number of methods of seasonally adjusting a series that developed throughout the 20th century. Whilst it is not completely comprehensive in that not every variation of the methods are given here, it is hoped that it will give the reader an understanding of the general approaches that are undertaken when dealing with seasonal adjustment and their evolution so far. Given a series which needs to be seasonally adjusted, there are many issues that need to be considered, not least is which method is most appropriate.

A note should be made to the consideration of the basic structural model in Equation (4.1) as to the ambiguity of the seasonal component. It is noted in Bell and Hillmer (1984) that different authors give different definitions of the seasonal component, without formally defining it. In their work they place a number of assumptions on the components of the basic model which are seen in the ARIMA model (Cleveland and Tiao, 1976) process. However, different works approach the definition of seasonality in a different manner, such as that in the harmonic regression (Beveridge, 1921) approach, where the seasonal components only sum to unity over the whole range of the series, not a year. This is contrary to the common definition that a seasonal component should not take or add anything away from a series over a year. Before approaching any seasonal adjustment, an analyst should be clear on the definition of seasonality that are they are estimating.

The report shows there is a wide and varied amount of methods available for seasonal adjustment, dependent on the level of information to be input and the expected output. For the most amount of information regarding movements in the series, a user first needs to provide a method with assumptions on what they expect. Parametric methods allow this information to become part of the model through their specification. However, if the aim is to only to determine the general movement of a series which is confounded with seasonal effects, a non-parametric approach may prove more useful. Within the overview sheets, where at all possible, a link to software which incorporates the methods has been given, such that they

can be tested. Many of the older methods can be calculated through standard functions within statistical packages such as R and SAS.

As outlined in Sections 4.2, 4.3, and 4.4 each method has its own weaknesses and strengths. Empirical studies into the efficiency of some of methods mentioned here can be found in Bianchi (1996), Hood (2002) and Jain (2012). However, considering only the methodological issues involved, and the development of the field throughout, it would appear that methods are becoming more parametric. The widely popular X-11 algorithm, currently now X-13-ARIMA-SEATS has incorporated parametric methods to aid their estimation. The increased efficiency of computing is allowing for more complex methods to be used when analysing a series, leading to more complex methods to decompose them.

4.6 Overview Sheets

Overview sheets can be found overleaf.

Median Link Relatives

1919

Persons, W. M., (1919), Outline of the Method, The Review of Economics and Statistics, 1(1), 3737.

Non-parametric **Keywords:** Least Squares, Median, Parabola, Moving Average.

Strengths

- Estimates both trend and cycle components separately.
- Medians used are more resistant to extreme observations.

Weaknesses

- Monthly data of length 15 years minimum is necessary.
- Separation of the secular trend and seasonal variation is not possible.

Description

A combination of locally fitted linear functions, link relatives, medians and a 12 month moving average are used to seasonally adjust a series. This was one of the first main methodologies for seasonal adjustment.

Pre-Adjustments

The data is assumed to be homogenous monthly data, so any calendar effects have been accounted for and that the seasonal is stable. A multiplicative decomposition is detailed but additive can be achieved using a logarithmic trnasformation.

Method

1. Given initial series y_t for $t = 1, 2, \ldots, n$ assume that it follows the model $y_t = \tau_t * c_t * s_t * \epsilon_t$ where the trend, secular trend (cycle), seasonal and irregular components are represented by $\tau_t, c_t, s_t, \epsilon_t$ respectively.

2. Estimate the trend component by fitting an appropriate function to sub-periods of the data. Often it is acceptable to fit a linear function, that which minimises the residual sum of squares. It may be necessary to vary the lengths of the sub-periods and weight appropriately. Denote the estimated trend $\hat{\tau}_t$

3. Once the estimated trend component has been removed (by subtraction $(y_t - \hat{\tau}_t)$ or division $y_t/\hat{\tau}_t)$) decompose the series into monthly (i) series across all years (j), denoted $z_{i,j}$. Then compute link relatives of the monthly changes by:

$$z_{i,j}^{LR} = \begin{cases} 100 & \text{if } i, j = 1 \\ \frac{z_{i,j}}{z_{1,j-1}} & \text{if } i = 12 \\ \frac{z_{i,j}}{z_{i-1,j}} & \text{else.} \end{cases} \qquad \forall i = 1, 2, \ldots, 12. \quad \forall j = 1, 2, \ldots, n/12.$$

4. Next take the median value of each month within all years, denoted z_i^m, and base them on the January value (set to 100) with adjustment by:

$$\frac{1}{1.011^{i-1}} 100 \prod_{i=1}^{a} z_i^m \qquad i = 2, 3, \ldots, 12.$$

   then divide by their total to center them. This is the seasonal index repeated for each year, denoted as $\hat{s}_t$.

5. To account for secular trend and seasonal variation, compute a 12 month moving average on the original series y_t to generate a curve that goes through all the data (by extension) and denote the results as λ_t. Following this, the correction (and thus seasonally adjusted series) can be calculated as:

$$y_t^* = \frac{y_t - \hat{s}_t \lambda_t}{\lambda_t}$$

Derivation

Percentage changes of monthly observations in relation to their mean are used.

Copeland, M. T., (1915), Statistical Indices of Business Conditions, The Quarterly Journal of Economics, 29(3), 522-562.

Periodicity of the cycle component is argued.

Moore, H. L., (1914), Economic cycles: Their law and cause, Macmillan Company.

Development

Disadvantages of the link relative method are further detailed.

Bauman, A. O., (1928), Thirteen-Months-Ratio-First-Difference Method of Measuring Seasonal Variation, Journal of the American Statistical Association, 23(163), 282-290.

Software available within standard packages.

Successive Mean Seasonal Index 1933

Kuznets, S., (1933), The Statistical Measurement, In Seasonal Variations in Industry and Trade, (pp. 23-40), NBER.

Non-parametric **Keywords:** Moving Average, Successive, Graphical.

Strengths

- Positional successive arithmetic means are used to guard against extreme values.

Weaknesses

- Assumes a stable seasonal component.
- Possible correlation amongst components not accounted for.

Description

A moving average of 13 points is estimated for the trend, followed by successive positional arithmetic means used as a method of estimating the seasonal index.

Pre-Adjustments

No pre-adjustments to the data are given in the work, the method can be used for additive series by using subtraction rather than division where necessary.

Method

1. Given pre-adjusted data y_t for $t = 1, 2, \ldots, n$ assume it can be decomposed as $y_t = \tau_t s_t \epsilon_t$ such that the trend, seasonal and irregular components are τ_t, s_t, ϵ_t respectively. Estimate the trend as $\hat{\tau}_t$ using a 13 point moving average with weights

$$\frac{1}{24}(1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 1).$$

2. Divide this estimate into the original data $y_t/\hat{\tau}_t = y_t^{s\epsilon}$ and multiply the ratio by 100.

3. Plot the results of each month over all years separately, and calculate successively larger averages of middle items within each month. For example, given $y_{i,j}^{s\epsilon}$ for month i and year j calculate means (assuming all years have 12 observations) for all months i:

$$\text{if } j \in 2\mathbb{N} : \frac{1}{L+1} \sum_{k=-L/2}^{L/2} y_{(j/2)+k}, \quad L = 2, 4, \ldots, j.$$

$$\text{if } j \in 2\mathbb{N}+1 : \frac{1}{L} \sum_{k=-L/2}^{L/2} y_{(j+k)/2}, \quad L = 3, 5, \ldots, j.$$

4. Center each successive mean to 100, by dividing by the sum of the components, and multiplying by 100.

5. Inspection of the indexes against the successive means should be undertaken to assess fit.

6. Select the mean which most accurately has other means clustered about it for each month. Compose these together to get the estimated seasonal index $\hat{s}_t$.

7. Comparison of this index against the relatives plotted in step 3 should be used for inspection of validity.

8. This final seasonal index is then divided into the original series to obtain the seasonally adjusted series. $y_t^{\tau\epsilon} = y_t/\hat{s}_t$

Derivation

Weekly variations are analysed.

Crum, W. L., (1927), Weekly Fluctuations in Outside Bank Debits, The Review of Economics and Statistics, 9(1), 3036.

The algorithm is very similar to the method then in use at the Federal Reserve Board.

Joy, A., & Thomas, W., (1928), The use of moving averages in the measurement of seasonal variations, Journal of the American Statistical Association, 23(163), 241-252.

Development

Seasonal indexes are used to estimate moving seasonality.

Cowden, D. J., (1942), Moving seasonal indexes, Journal of the American Statistical Association, 37(220), 523-524.

Software available within standard statistical packages.

Graphical Method

1937

Spurr, W. A. (1937), A Graphic Method of Measuring Seasonal Variation, Journal of American Statistical Association, 32.198 281-289.

Non-Parametric **Keywords:** Graphical Measure, Trend, Trend-Cycle, Seasonal Index, Subjective.

Strengths

- Less laborious than calculating Moving Averages with 'negligible' error.
- Adaptive to progressive seasonal indices.

Weaknesses

- Computer calculations now mean the save in efficiency is now neglible/defunct.
- Two subjective stages to the process could lead to greater variation if separate people perform the analysis.

Description

This method contrasts from analytical methods in that it uses free hand curve and best estimated graphical procedures. These are based off the same analytical methods but reduce computational time for the sake of efficiency. Computing times (in comparison to methods at that time) are efficient with a small level of error.

Pre-Adjustments

The original data is either plotted on semi-logarithmic paper for multiplicative series or arithmetic paper for additive series. Extreme values within the seasonal indices are accounted for within the method.

Method

1. Draw a non-seasonal baseline through the original data, this may be a trend or trend-cycle. This can be done freehand or using the method of least squares. A graphic link relative method can also be used.
2. For each period (month, quarter etc.) measure the distance in each year from the non-seasonal to the data. This is done using a vertical strip of paper which has been bisected with a dotted line. This line is placed adjacent to the non-seasonal line for that period and the position of the original value from this line is marked on the strip with the year noted. This is done for each month.
3. When turned horizontal, these strips show the behaviour of the seasonal ratios for each month and can indicate progressive seasonality. This is dealt with further below.
4. Calculate a geometric mean of the central group and mark it with a caret on the strip. Extreme items can be discarded by inspection. The distance of the caret from the dotted line is the preliminary seasonal index.
5. Next normalise the seasonal indices by plotting them vertically over a horizontal origin and dividing each distance by twelve (bisect and trisect), then summing these differences. Multiply this average difference with each monthly index caret, marking this as the corrected index.
6. To plot the seasonally adjusted series, line each month's corrected caret up on the corresponding point, then mark where on the chart the dotted line lies. A freehand curve can then be drawn through the points.

Progressive Seasonality:

a) Rather than taking a geometric mean over the monthly differences from trend, instead draw a curve through the points on the strip, and transferring that trend for each year to the left hand side of the strip.

b) The differences now between the dotted line and the seasonal trend values for each year are the seasonal indices. These are then normalised as per Steps 4 and 5.

c) When plotting the adjusted series, the point must correspond to that year's index.

Derivation

Freehand curves are used in the revision of Factory Employment by the Federal Reserve Bureau over moving averages.

Federal Reserve Bureau Bulletin, December 1936.

The adaptive mean used in the calculation of the indices is similar to the "positional arithmetic mean" by Kuznets.

Kuznets, S. (1933), The Statistical Measurement, Seasonal Variation in Trade and Industry, NBER, 23-40.

Development

Spurr alters his method, changing the subjective trend/trend-cycle estimation making it computational.

Spurr, W. A. (1940), A Graphic Short Cut to the Moving Average Method of Measuring Seasonality, Journal of the American Statistical Association, 35.212 667670.

Examples of how to use the paper strips are given in the work.

Multiple Regression Analysis 1939

Mendershausen, H. (1939), Eliminating Changing Seasonals by Multiple Regression Analysis, The Review of Economics and Statistics, 21.4 171-177.

Parametric **Keywords:** Multiple Regression, Evolving Seasonality, Variate Analysis.

Strengths

- Allows for evolving seasonality throughout the period analysed.
- No extrapolation/data loss at the terminal ends of a series are necessary.

Weaknesses

- Not all seasonal movements can be explained using the linear equations, curvilinear equations are then suggested, but their determination is not given.
- Constrains the seasonal component over the period of the data, but not over the period of a year.

Description

Mendershausen shows how multiple regression analysis can be used to develop evolving seasonal patterns dependent on variables such as weather and variables for fall of employment levels. This is done through two case studies, but an attempt at generalisation is made here, however methods for determining the regression equations have been assumed.

Pre-Adjustments

No pre-adjustments are given in the method, attempts to explain unusual movements are made with knowledge of events and adjustments to the seasonal regression equations. A logarithmic transformation should be used for suspected multiplicative series.

Method

1. Given a series $y_t, \quad t = 1, 2, \ldots, n = sN + e$ where s is the amount of periods (i.e. 12 months), N is the length of the data (5 years) and e is the additional (non-full years) data, let the seasonal normal for each period be:

$$S_k = \bar{S}_k + \sum_{i=1}^{l} \alpha_i x_i + \sum_{j=1}^{m} \beta_j z_j, \quad k = 1, 2, \ldots, s$$

   where x_i are significant meteorological variables and z_j are social variates, with lengths l, m respectively and $\bar{s}_k$ represents the average seasonal component with average social and meteorological conditions.

2. A method is not given in the work to fit the equations, however multiple linear regression using least squares techniques would enable computation of the coefficients and constants.

3. Again, the method of selection of the optimal subset of coefficients is not given, but is referred to earlier work given below.

4. Select significant seasonal normal equations by considering the percentage of variance explained by each. The statistics used and their testing procedures are given in work below.

5. If a particular months seasonal normal tested insignificant, the stable seasonal component is used instead.

6. The seasonal component can be removed using subtraction from the original series at each month.

Additionally, the irregular can be removed from the remaining series using a 3 term moving average. Curvilinear equations may also be fitted to the seasonal normals rather than a simple linear relationship, however these are drawn free-hand and must be read off a scatterplot.

Derivation

Description of the variates and the process of their selection is given in an earlier study.

Mendershausen, H. (1937), Les variations du mouvement saisonnier dans l'industrie de la construction: etude methodologique et analyse des faits, Geneve: Georg & Cie. s. a.

Statistics used to explain the percentage of variation explained are detailed.

Ezekiel, M. (1930), Methods of Correlation Analysis, Wiley, Oxford, England.

Development

Linear regression techniques for seasonal adjustment are extended to use dummy variables to account for extreme points, and corrections are made for autocorrelated residuals.

Lovell, M. C. (1963), Seasonal Adjustment of Economic Time Series and Multiple Regression Analysis, Journal of the American Statistical Association, 58.304 993-1010.

Software employing standard regression techniques can be used.

Early Harmonic Regression 1941

Davis, H. T. (1941), The Technique of Harmonic Analysis, The Analysis of Economic Time Series, Bloomington, Indiana: The Principia Press Inc, 77-83.

Non-Parametric **Keywords:** Periodogram, Harmonic Analysis, Period, Cycle Extraction.

Strengths

- Cycle extraction is a straight forward process.
- Allows period extraction over longer cycles (2 years for example).

Weaknesses

- No formal significance test is given within the method, only by eye.
- Does not account for changing seasonality.
- Construction of different periods for analysis leads to a loss of data.

Description

Harmonic analysis of economic time series allows key peaks of periodicity to be noticed through use of a constructed periodogram. Following this, prevalent cycles can be analysed and where necessary removed through use of sine and cosine regression terms.

Pre-Adjustments

No pre-adjustments are given in the work. Logarithms should be taken on multiplicative series.

Method

1. Given data $y_1, y_2, \ldots, y_N$, arrange into m rows of p observations ($N' = mp, p \in \mathbb{Z}, N' \leq N$), and calculate column totals:
$$Y_i = \sum_{j=1}^{m} y_{(j-1)p+i} \quad \text{for} \quad i = 1, 2, \ldots, p,$$
selecting a range of periods p relevant to the data.

2. Following this, for each period p calculate:
$$a_p = \frac{2}{N'} \sum_{i=1}^{p} Y_i \cos\left(\frac{2\pi i}{p}\right), \qquad b_p = \frac{2}{N'} \sum_{i=1}^{p} Y_i \sin\left(\frac{2\pi i}{p}\right), \qquad r_p^2 = a_p^2 + b_p^2.$$

3. The values of r_p can be plotted against their period to give the periodogram. Peaks on this graph indicate the possible existence of a periodicity.

4. Given significant periods $p_1^*, p_2^*, \ldots, p_k^*$ determined by eye, the percentage of energy contained in these periods can be calculated as:
$$P = \frac{\sum_{i=1}^{k} R^2(p_i^*)}{2\sigma^2}$$

5. An adjusted series can then be obtained by removing these significant peaks through:
$$Y_t = y_t - \left[\sum_{i=1}^{k} a_{p_i^*} \cos\left(\frac{2\pi t}{p_i^*}\right) + b_{p_i^*} \sin\left(\frac{2\pi t}{p_i^*}\right)\right]$$

Derivation

The Ohio Annual rainfall is investigated using a periodogram. (Chapter produced in 1914)
Moore, H. L. (1967), Cycles in Rainfall, Economic Cycles: Their Law and Cause, The Macmillan Company, New York, 6-34.

Analysis of a periodogram was conducted by Beveridge in his attempt to justify weather cycles.
W. H. Beveridge (1921), Weather and Harvest Cycles, The Economic Journal, 31.124 429-452.

Development

Decomposition of uncorrelated spectral components is described with programs.
Granger, C. W. J. & M. Hatanaka (1964), Spectral Analysis of Economic Time Series, Princeton University Press.

Harmonic regression and polynomial trend curves are used to seasonally adjust a series.
Speth, H. T., (2006), The BV4.1 procedure for decomposing and seasonally adjusting economic time series.

Software available within standard statistical software.

Weighted Ratio Indices 1942

Shiskin, J. (1942), A New Multiplicative Seasonal Index, Journal of the American Statistical Association, 37.220 507-516.

Non-Parametric **Keywords:** Moving Average, Multiplicative, Indices, Additive, Centred.

Strengths

- Shown to be quicker to compute than the monthly means indices and the ratio-to-moving average index in small empirical study.
- A multiplicative seasonal index can be converted to additive (or vice versa).

Weaknesses

- Assumed that the random factor has no disturbing effect upon any relationships between the monthly trend-cycle and seasonal factors.
- Trend-cycle curve estimated through a twelve month moving average is known to potentially capture elements of seasonality and also needs to be extended using a freehand curve.

Description

A new 'weighted ratio-to-moving-average index' or 'weighted ratio index' is proposed to calculate the seasonal indices. This new index has a close relationship and is very similar with the ratio-to-moving average index and monthly means index which is shown in the work.

Pre-Adjustments

No pre-adjustment methods are mentioned in the work, but it is mentioned that it will be difficult to determine extreme values within ratio-index charts. The method accounts for additive or multiplicative indices.

Method

1. Given original observations y_i for $i = 1, 2, \ldots, n$, calculate the trend-cycle component by a 12 month moving average and extend the curve freehand. This can be calculated by (without freehand extension):

$$T_i = \frac{1}{12} \sum_{j=i-6}^{i+5} y_j \qquad i = 7, 8, \ldots, n-5.$$

2. For each month calculate the sum of the observations and the sum of the trend cycle component as:

$$y_\tau = \sum_{i=1}^{n/12} y_{12(i-1)+\tau}, \quad T_\tau = \sum_{i=1}^{n/12} x_{12(i-1)+\tau}, \qquad \tau = 1, 2, \ldots, 12.$$

3. Each month's multiplicative uncentered weighted-ratio seasonal index is then $S^*_\tau = y_\tau / T_\tau$.
4. Centre the indices through:

$$\tilde{S}^*_\tau = \frac{S^*_{\tau=1}}{\sum_{\tau=1}^{12} S^*_\tau}.$$

To convert this multiplicative index S^*_τ to an additive seasonal index S^+_τ,

$$S^+_\tau = \frac{n(S^*_\tau - 1)}{x_\tau}.$$

Derivation

The ratio-to-moving averages method this is related to is used within Census Method I.

Shiskin, J., & Eisenpress, H. (1957), Seasonal Adjustments by Electronic Computer Methods, JASA, 52.280 415-449.

Development

Durbin describes an additive seasonal index very similar in construction, with endpoint correction.

Durbin, J. (1963), Trend elimination for the purpose of estimating seasonal and periodic components of time series, Time Series Analysis, 3-16.

Software can be created using the standard techniques quickly.

Census Method I 1954

Shiskin, J. (1954), Seasonal computations on UNIVAC, The American Statistician, 9.1 19-23.

Non-parametric **Keywords:** Moving Average, Filter, Extrapolation, Repeated.

Strengths

- Simple yet effective procedure to decompose a series into Trend-Cycle, Seasonal and Irregular signals.
- Accounts for non-constant seasonality by allowing the seasonal factors to change over time.

Weaknesses

- An ill-fitting curve can occasionally be obtained through the five month moving average used to estimate the Trend-Cycle signal.
- A trend at the beginning or end of the series can create problems when extrapolating during the 5 month moving average.
- Freehand curve estimation using moving averages as guides.

Description

Method I was the first mechanised Seasonal Adjustment Method, developed by the US National Census Bureau. Programmed on the Univac Computer, it allowed automatic adjustment of many series. Using different moving average filters, signals are extracted from the series and presented within 19 tables. A test of Seasonality is also made on the original series and the adjusted series to conclude on the effect of adjustment.

Pre-Adjustments

No pre adjustments given within the work. If the series is believed to be multiplicative, then a logarithmic transformation should be taken.

Method

1. Given series y_t for $t = 1, 2, \ldots, n = Ns$ (N is the number of years and s the periodicity, i.e. 12 for monthly) assume it can be decomposed into $y_t = T_t + S_t + I_t$, where T_t is the trend, S_t the seasonal comopent and I_t the irregular. Test for evidence of seasonality by dividing the original figures by an average of the preceding and following months value, $y_t \div \frac{1}{2}(y_{t-1} + y_{t+1})$. If the variations appear random about 1 then this suggests no seasonality.
2. Trend-Cycle signals are estimated by a freehand curve aided by plotting the series y_t with an equally weighted and uncentred 12-month moving average, denote this as T_t^*.
3. This is divided into the original series to leave the Seasonal-Irregular signal, $(S_t^* + I_t^*) = y_t / T_t^*$.
4. Calculate a five term moving average across the series representing the Seasonal-Irregular signal (seasonal factors) for each month across all years. Factors are extrapolated for this calculation by extending the previously calculated factor across the two missing initial and terminal years, denote the factors as $S_{N,s}^*$.
5. Center seasonal factors such that their sum equals 1200, such that $S_t^* = S_{N^*,e} / \sum_{i=1}^{N^*} S_{i,e}^*$ where $t = sN^* + e$.
6. Preliminary seasonally adjusted series is considered to be the original series divided by the seasonal factors calculated above, $y_t^p = y_t / S_t^*$
7. Smooth this preliminary seasonally adjusted series with a five month moving average extending it as in step 4, giving a more adaptable Trend-Cycle signal than the 12 month moving average.
8. Repeat steps 2, 3, 4 for the 5-month moving average, returning the final adjusted series.
9. Check for seasonality, by dividing an uncentered 12 month moving average of the original series by a uncentered 12 month moving average of the seasonally adjusted series.

Derivation

This is a mechanical approximation to the method given by the Federal Reserve Bureau.

Barton, H. C. Jr. (1941), Adjustment for Seasonal Variation, Federal Reserve Bulletin, 518-528.

Adaptation of the ratio-to-moving average method.

Mills, F. C. (1955), Statistical Methods, New York, 360-375.

Development

The method was further developed into Census Method II.

Shiskin, J. & Eisenpress, H. (1957), Seasonal Adjustments by Electronic Computer Methods, Journal of the American Statistical Association, 415-449.

Software available as part of *Statistica* : `http://www.statsoft.com/Products/STATISTICA/`

Exponentially Weighted Moving Average 1960

Winters, P. R. (1960), Forecasting Sales by Exponentially Weighted Moving Averages, Management Science, 6 324-342.

Non-Parametric **Keywords:** Moving Average, Exponential, Forecasting, Seasonal, Decomposition.

Strengths

- Can be computed only reliant on previous values, making it quick and efficient.
- Method can be used for effective forecasting of the series (widely used today).

Weaknesses

- Initial weights on previous values need to be chosen through trial and error.
- Initial values for the seasonal factors, trend and beginning adjustment based on averages.

Description

The work builds on previous work by Holt and is primarily for the function of forecasting. However, the method can be used as a form of seasonal adjustment. One can use the exponential weights and simple iterative relationships between the components to supply an adjusted series.

Pre-Adjustments

No pre-adjustments are given within the work. Procedures for multiplicative series are given, if it is believed that the series is additive, take exponentials.

Method

Begin with observations y_t where $t = 1, 2, \ldots, sN$ such that s is the amount of periods (12 months for example) and N the amount of years. Model the adjusted series through the equations and relationships:

$$\text{Adjusted Series:} \quad \tilde{y}_t = A\frac{y_t}{S_{t-s}} + (1-A)(\tilde{y}_{t-1} + T_{t-1}) \tag{4.7}$$

$$\text{Seasonal Factors:} \quad S_t = B\frac{y_t}{\tilde{y}_t} + (1-B)S_{t-L} \tag{4.8}$$

$$\text{Trend Estimate:} \quad T_t = C(\tilde{y}_t - \tilde{y}_{t-1}) + (1-C)T_{t-1} \tag{4.9}$$

Values of the constants A, B, and C can be found through exploratory methods, the work suggests the Gradient Method. The initialisation of $\tilde{y}_1, S_1, S_2, \ldots, S_s, T_1$ is suggested as:

1. Consider $N*$ years of data to begin initialisation and calculate the average observation for each year $i = 1, 2, \ldots, N*$ as V_i and the average trend across the period as T_1 by,

$$V_i = \sum_{t=12(N*-1)+1}^{sN*} y_t, \quad i = 1, 2, \ldots, p; \quad T_1 = \frac{V_{N*} - V_1}{N* - 1}.$$

2. Initialise $\tilde{y}_1 = V_1$ and calculate seasonal factors for each sub-period t, where j represents the position with the year (i.e. $j = 1$ for January in a monthly series), then average and normalise them by

$$\hat{S}_t = \frac{y_t}{V_i - \left(\frac{s+1}{2} - j\right)T_1} \quad t = 1, 2, \ldots, sN; \quad \bar{S}_j = \sum_{i=1}^{s} \hat{S}_{s(j-1)+i}, \quad S_j = \frac{\bar{S}_j L}{\sum_{j=1}^{s} \bar{S}_j}.$$

Equations (4.7), (4.8) and (4.9) can then be iteratively calculated and a seasonally adjusted series given by y_t/S_t.

Derivation

The exponential models were first developed in a paper by Charles C. Holt.

Holt, C. C. (2004), Forecasting seasonals and trends by exponentially weighted moving averages, International Journal of Forecasting, 20.1 5-10.

The gradient method is described by Lance in

Lance, G. N. (1959), Solution of Algebraic and Transcendental Equations on an Automatic Digital Computer, Journal of Association for Computing Machinery, 6 97-101.

Development

Gardner reviews the models and attempts to improve them using Fourier functions for the seasonal factors.

Gardner, E. S. (1985), Exponential smoothing: The state of the art, Journal of Forecasting, 4 1-28.

Software available as the **R** function **HoltWinters**.

Slowly Changing Seasonal

1964

Hannan, E. J., (1964), The Estimation of a Changing Seasonal Pattern, Journal of the American Statistical Association, 59(308), 1063-1077.

Parametric

Keywords: Harmonic Regression, Evolving Seasonality, Moving Average.

Strengths

- The method attempts to account for seasonality that isn't assumed to be constant.
- A level of expected dependence amongst seasonal components can be set.

Weaknesses

- Fast fluctuations within the seasonal component can hinder the method.
- There is no way to incorporate information regarding the behaviour of the seasonal.

Description

By decomposing the spectrum of a series, the method aims to capture both stable and varying seasonality through an iterative relationship.

Pre-Adjustments

The work assumes any pre-adjustments necessary, such as calendar effects or transformations of the data has already been done. An additive model is assumed, logarithms could be taken for a multiplicative series.

Method

1. Assume that a series y_t for $t = 1, 2, \ldots, n$ can be modelled as $y_t = \tau_t + s_t + \epsilon_t$ for trend, seasonal and irregular components τ_t, s_t, ϵ_t respectively.
2. The trend is estimated using a moving average, in an example within the work a 12 point moving average is employed, denote the results as $\hat{\tau}_t$ and subtract them from the original series to get $y_t^* = y_t - \hat{\tau}_t$.
3. Calculate:

$$\hat{\alpha}_{k,t} = \frac{1}{12}\sum_{s=-6}^{6}(2-\delta_k^6)y_{t+s}^*\cos\lambda_k(t+s) \quad \text{for } k = 1, 2, \ldots, 6. \qquad \text{Where:} \quad \delta_k^6 = \begin{cases} 0 & \text{if } k \neq 6 \\ 1 & \text{if } k = 6 \end{cases}$$

$$\hat{\beta}_{k,t} = \frac{1}{12}\sum_{s=-6}^{6}(2-\delta_k^6)y_{t+s}^*\sin\lambda_k(t+s) \quad \text{for } k = 1, 2, \ldots, 5. \qquad \lambda_k = \frac{2\pi k}{6}$$

   and smooth each series using a moving average across t to obtain $\hat{\alpha}_k, \hat{\beta}_k$.
4. Given a value for $\rho_k \geq 0.95$ and sufficiently large m such that $\hat{\beta}_k^j$ is negligible for $j > m$, calculate for an initial time point t_0:

$$\xi_{k,t_0} = (1-\rho_k^{-1}\hat{\beta}_k)\sum_{j=0}^{m}\hat{\beta}_k^j y_{t_0-j}^*\cos\lambda_k j \quad k = 1, 2, \ldots, 6.$$

$$\psi_{k,t_0} = (1-\rho_k^{-1}\hat{\beta}_k)\sum_{j=0}^{m}\hat{\beta}_k^j y_{t_0-j}^*\sin\lambda_k j \quad k = 1, 2, \ldots, 5.$$

5. These items can be iteratively calculated by the following relationships:

$$\xi_{k,t+1} = \hat{\beta}_k\{\hat{\xi}_{k,t}\cos\lambda_k - \hat{\psi}_{k,t}\sin\lambda_k\} + (1-\rho_k^{-1}\hat{\beta}_k)y_{t+1}$$

$$\psi_{k,t+1} = \hat{\beta}_k\{\hat{\psi}_{k,t}\cos\lambda_k + \hat{\xi}_{k,t}\sin\lambda_k\}.$$

6. Finally the seasonal component can be calculated as $\hat{s}_t = \sum_{k=1}^{6}\xi_{k,t}(2-\delta_k^6)$. This can be subtracted to give the seasonally adjusted series $\hat{y}_t = y_t - \hat{s}_t$.

Derivation

Spectral analysis of seasonal adjustment is undertaken.

Nerlove, M., (1964), Spectral Analysis of Seasonal Adjustment Procedures, Econometrica, 32(3), 241-286.

An early attempt to account for slowly varying seasonality is presented.

Kuznets, S., (1932), Seasonal Pattern and Seasonal Amplitude: Measurement of Their Short-Time Variations. Journal of the American Statistical Association, 27(177), 9-20.

Development

Methods of seasonal adjustment, including this method are compared.

Godfrey, M. D., & Karreman, H. F., (1964), A spectrum analysis of seasonal adjustment, Princeton University.

Software not available.

Moving Average/Harmonic Analysis 1965

Burman, J. P. (1965), Moving Seasonal Adjustment of Economic Time Series, Journal of the Royal Statistical Society: Series A, 128.4 534-558.

Parametric **Keywords:** Moving Average, Harmonic, Regression, Adaptive, ARIMA model.

Strengths

- Adaptive based algorithm which chooses the most flexible moving average within context of the series.

Weaknesses

- The extrapolation methods described can be 'hazardous' when used on highly irregular series.

Description

Burman describes an extensive method of seasonal adjustment using moving averages for trends and harmonic regression for the seasonal factors with exponential extrapolation. It is described for monthly series.

Pre-Adjustments

Extreme values are found using root mean square deviations in the method. An additive series method is described, for multiplicative series steps 1, 5 and 7 are used on the logarithms.

Method

1. Apply the optimal 13-term moving average detailed in the work to the series to estimate the trend.
2. Model the deviations of the original series from the trend estimate $y_t - T_t$ using harmonic analysis,
$$a_{kj} = \frac{2}{12g(\omega_k)} \sum_{i=1}^{12} x_{ij} \cos i\omega_k, \quad b_{kj} = \frac{2}{12g(\omega_k)} \sum_{i=1}^{12} x_{ij} \sin i\omega_k, \quad k = 1, \ldots, 6,$$
where $x_{ij} = y_{ij} - T_{ij}$ is the deviation from the trend of the ith month in the jth year and $g(\omega_k) = 1 - F(\omega_k)$ where F is the frequency response function, values of which can be found in Table 2 of the work.
3. Form a time series of each harmonic $z_{j,k}$ across all years j, then calculate the von Neumann Ratio,
$$V = \sum_{k=1}^{11} \frac{\sum_{t=2}^{n} (z_{t,k} - z_{t-1,k})^2}{\sum_{t=1}^{n} (z_t - \bar{z}_k)^2}, \quad \bar{z}_k = \frac{1}{n} \sum_{t=1}^{n} z_{t,k}$$
and test at the 25% significance level. If insignificant, choose $\bar{z}$ as the smoothed amplitudes. However, if significant, replace $z_{t,k}$ now with $z_{t,k} - \bar{z}_{t,k}$ where $\bar{z}_{t,k}$ is the smoothed amplitude calculated from the next flexible average according to page 547 of the work. If this list is exhausted, use the 5 term moving average. Denote the smoothed amplitudes $\bar{a}_{k,j}$ and $\bar{b}_{k,j}$, giving seasonal variations in the ith month and jth year as $u_{i,j}$ below. Then extrapolate the amplitudes at each end by a year by following the formula for $\hat{z}_{t+1}$:
$$u_{i,j} = \sum_{k=1}^{6} \bar{a}_{k,j} \cos i\omega_k + \sum_{k=1}^{5} \bar{b}_{k,j} \sin i\omega_k \qquad \hat{z}_{t+1} = \frac{z_t + \lambda z_{t-1} + \lambda^z_{t-2} + \cdots + \lambda^{t-1} z_1}{1 + \lambda + \lambda^2 + \cdots + \lambda^{t-1}}$$
taking λ as either the value used during smoothing or $\lambda = 0.6$ for the 3x5 moving average. If however the 5-term moving average was used, then find by least squares methods the optimal value of λ.
4. Apply the extrapolation and smoothing to the last 12 terms and combine this with the 11 smoothed amplitudes. Deduct these from the original series, giving the preliminary seasonally adjusted series (PSA).
5. Deduct the trend estimate from Step 1 and calculate the root mean square of these deviations σ_1.
6. Any term whose residuals are more than $2.5\sigma_1$ away in distance should be replaced by the trend value.
7. Extrapolate the PSA by 6 terms at the end (and if the series spans whole years, the beginning) using the Box-Jenkins 3 term adaptive predictor based on the previous (first) 60 terms. Add the seasonal deviations and their extrapolations to obtain an extended original series. Replace any extreme items as per step 6.
8. Finally repeat steps 1-4 on the extended series to provide a final seasonally adjusted series. A full measure of deviation can also be obtained by applying step 5 to the seasonally adjusted extended series.

Derivation

The method is a variation on the Method II census program of seasonal adjustment.

Shiskin, J. & Eisenpress, H. (1957), Seasonal Adjustments by Electronic Computer Methods, Journal of the American Statistical Association, 415-449.

Development

Further optimal design on filters using Frequency Response Functions is detailed.

O'Gorman, T. W. (1982), On the Design of Seasonal Adjustment Methods Using Linear Programming Techniques, Journal of the American Statistical Association, 77.380 739-742.

Software mentioned in work, but were not able to be found.

General Linear Model 1966

Henshaw, R. C. (1966), Application of the General Linear Model to Seasonal Adjustment of Economic Time Series, Econometrica, 34.2 381-395.

Parametric **Keywords:** Linear Model, Least Squares Estimation, Moving Seasonality, Autocorrelated Disturbances.

Strengths

- Adaptable process where degree of components can be changed to fit different evolutionary patterns.

Weaknesses

- Trial and Error process requiring initial knowledge of expected features of a series.

Description

This method decomposes the series into trend-cycle (referred to as non-periodicities) and seasonal (periodicities) components using linear models of varying degrees. The degree of the seasonal polynomials allows for moving seasonal components.

Pre-Adjustments

It is assumed that any systematic error variables or configurations, such as trading day effect and holidays have been removed. The method of employing polynomials in time (t) is suggested. Logarithms should be taken if it is suspected that the series is in fact multiplicative.

Method

Given the standard model of (Log) $Y_t = T_t + S_t + I_t$, model the observations (assumed monthly) by:

$$y_t = \sum_{j=1}^{12}\sum_{\nu=0}^{s}\lambda_{\nu j}\Upsilon_{jt}^{\nu} + \sum_{\nu=s+1}^{c}\lambda_\nu t^\nu + u_t, \quad t = 1, 2, \ldots T. \quad \text{where} \quad \Upsilon_{jt}^{\nu} = \begin{cases} t^\nu & \text{if } t - j \in 12\mathbb{Z} \\ 0 & \text{otherwise} \end{cases}.$$

This is such that the seasonal and non seasonal components are represented by:

$$T_t = \sum_{\nu=0}^{c}\lambda_\nu t^\nu, \quad S_t = \sum_{j=1}^{12}\sum_{\nu=0}^{s}(\lambda_{\nu j} - \lambda_\nu)\Upsilon_{jt}^{\nu}, \quad \text{where} \quad \lambda_\nu = \frac{\sum_{tj}\lambda_{\nu j}\Upsilon_{jt}^{\nu}}{\sum_{tj}\Upsilon_{jt}^{\nu}}, \quad \text{and} \quad \sum_{t=1}^{T} S_t = 0.$$

A dependence structure is defined on disturbances such that $u_t = \delta u_{t-1} + e_t, \quad 0 \leq \delta < 1$ where e_t are IID Normal variables with mean zero and constant variance σ^2. This is fitted using least square regression techniques given the degree of the (non)seasonal components (c) s. Formulae are given in the Appendices of the work. An iterative procedure is then employed:

1. Begin by choosing large values of $\hat{s}$ and $\hat{c}$ such that $\hat{\delta}$ is ***almost surely*** zero.
2. Calculate von Neumann ratio of least-squares regression disturbances:
$$d = \frac{\sum_{t=2}^{T}(\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{T}\hat{u}_t^2}$$
3. Test $H_0 : (\delta = 0)$ vs. $H_1 : (\delta > 0)$ using the Durbin-Watson test at the 5% level with d as the test statistic.
4. If H_0 is not rejected, through trial and error, reduce $(\hat{s}, \hat{c})$ such that they are minimized whilst H_0 is not rejected. ANOVA F tests and t tests on the regression coefficients are suggested to aid decision making.
5. Obtain Seasonally Adjusted series $\tilde{y}_t = y_t - \hat{S}_t$.

Derivation

Cowden suggests fitting polynomials in t by least squares to each of the monthly seasonals.

D. J. Cowden (1942), Moving Seasonal Indexes, Journal of American Statistical Association, 37 523-524.

Hald simultaneously fits 12 constant polynomials to fixed seasonal components and a general polynomial to the trend.

A. Hald (1948), The decomposition of a series of observations composed of a trend, a periodic movement and a stochastic variable, Thesis.

Development

Stephenson develops the model further to include both constant and evolving seasonal variables.

J. A. Stephenson & H. T. Farr, Seasonal Adjustment of Economic Data by Application of the General Linear Statistical Model, *JASA, 67.337*, 37-45 (1972)

Software available through standard regression methods.

Multiplicative/Additive Regression 1975

Durbin, J. & Murphy, M. J. (1975), Seasonal Adjustment Based on a Mixed Additive–Multiplicative Model, Journal of the Royal Statistical Society, Series A, 138(3), 385-410.

Parametric **Keywords:** Mixed Model, Regression, Test Procedure, Moving Average, Filters.

Strengths

- Allows for both a multiplicative and additive form of the seasonal.
- A testing procedure is outlined for detection of mixed models.

Weaknesses

- Example series show that different trend filters can influence the multiplicative/additive test.
- Moving seasonality allowed only up to a degree of two.

Description

Step-wise regression under a number of constraints is used to estimate both multiplicative and additive seasonal components, potentially including both in a model. Trend estimation is done iteratively using 21-point and Burman filters.

Pre-Adjustments

Adjustments for calendar effects are not given in the work, however a procedure that accounts for extreme values is used between steps 2-3, 5-6 and 6-7 can be found in Section 5 of the work.

Method

1. Assume that a given pre-adjusted series $y_{i,j}$ can be expressed as $y_{i,j} = \tau_{i,j} + \lambda_{i,j}(s_j^+ + s_j^*)\tau_{i,j} + \epsilon_{i,j}$ where $\tau_{i,j}$ is the trend, s_j^+, s_j^* are the harmonic regressors for additive and multiplicative seasonal regressors with local scaling factor $\lambda_{i,j}$, and $\epsilon_{i,j}$ is the irregular component. The indexes i, j represent month and year.
2. Estimate preliminary seasonal factors $\tau_{i,j}^P$ by using a 21-point moving average with weights: $(-0.04769, -0.02535, -0.00301, 0.01933, 0.04167, 0.06401, 0.08634, 0.10868, 0.13102, 0.08333, \underline{0.08333})$. Asymmetric weights are used for the endpoints. The Burman Mixed trend filter is also suggested.
3. Fit the 11 additive harmonic seasonal components $s_{i,j}^+$ and, using an F-test at either 1% or 5%, fit the multiplicative components $s_{i,j}^*$. If the multiplicative components are deemed significant then denote the event by S_M, if insignificant then N_M.
4. Repeat step 3 but instead beginning with the multiplicative components with events S_A and N_A. The choice of model is then 'Either' for $N_M N_A$, 'Additive' for $N_M S_A$, 'Multiplicative' for $S_M N_A$ and 'Mixed' for $S_M S_A$. Denote the preliminary seasonal factors by s_A^P, s_M^P.
5. The local scaling factors $\lambda_{i,j}$ can be estimated using regression on $y_{i,j} - \tau_{i,j}^P = \lambda_{i,j} s_{i,j}^P + \epsilon_{i,j}$ using 15 month windows centred on a time point. Giving preliminary seasonally adjusted series $y_{i,j}^{PS} = y_{i,j} - \lambda_{i,j}^P s_{i,j}^P$.
6. Apply a 13 point moving average to $y_{i,j}^{PS}$ with weights: $(-0.0331, -0.0208, 0.0152, 0.0755, 0.1462, 0.2039, \underline{0.2262})$ (A Burman Filter) and fit seasonal harmonics as per steps 3 and 4, giving new estimates of $s_{i,j}^P$ and $\tau_{i,j}^P$.
7. Re-calculate amplitude factors $\lambda_{i,j}^P$ as per step 6, giving a new preliminary seasonally adjusted series $y_{i,j}^{PS}$.
8. For the last time, recalculate the final trend $(\tau_{i,j}^F)$ and seasonal $(s_{i,j}^F)$, but including possible time varying parameters such that there are extra additive seasonal components based on time $t = 12j + i$:
$$y_{i,j} = \tau_{i,j}^F + \lambda_{i,j}^P(s_{0,j}^+ + s_{1,j}^+ t + s_{2,j}^+ t^2 + s_j^*)\tau_{i,j}^F + \epsilon_{i,j}$$
9. Calculate the final local amplitudes $\lambda_{i,j}^F$, giving final seasonally adjusted series $y_{i,j}^F = y_{i,j} - s_{i,j}^F$.

Derivation

A mixed model is used but without much flexibility.

Deutsche Bundesbank (1960), The Practice of Seasonal Adjustment with Regression Equations, Frankfurt (Main): Deutsche Bundesbank.

Burman filters are derived in an earlier work.

Burman, J. P. (1965), Moving Seasonal Adjustment of Economic Time Series. Journal of the Royal Statistical Society. Series A, 128(4), 534-558.

Development

A mixed seasonal model estimated in X-12-ARIMA is detailed.

Arz, S. (2006), A new mixed multiplicative-additive model for seasonal adjusment (No. 2006, 47,. Discussion paper Series 1/Volkswirtschaftliches Forschungszentrum der Deutschen Bundesbank.

Software not currently available publicly.

Seasonal ARIMA Models 1976

Cleveland, W. P. & Tiao, G. C. (1976), Decomposition of Seasonal Time Series: A Model for the Census X-11 Program, Journal of the American Statistical Association, 71.355 581-587.

Parametric

Keywords: ARIMA models, Decomposition, Autoregressive, Moving Average, Differencing.

Strengths

- Provides a parametric explanation to the behaviour of a series.
- No issues when dealing with the terminal ends of series.
- Allows nonstationary trend and seasonal components.

Weaknesses

- Larger residual sum of squares is found after modelling than the Census I procedure.
- The model given closely approximates the procedure for a 13 month moving-average, but not for a 9 term or 23 term.

Description

The work presents ARIMA models and shows how they can closely approximate the moving average procedures used in the X-11 Seasonal Adjustment procedure. These models approximate non-stationary seasonality and trend components primarily for forecasting, but can be used to decompose a series.

Pre-Adjustments

No pre-adjustments are given in the work. Logarithms should be taken on multiplicative series.

Method

1. Given a series y_t, $t = 1, 2, \ldots, n$ propose the additive decomposition into component form $y_t = T_t + S_t + I_t$ where T_t is the trend, S_t the seasonal component and I_t the noise.
2. ARIMA models are proposed for the trend and seasonal components:

$$\phi_{r_1}(B)(T_t - \mu) = \psi_{q_1}(B)b_{1,t}, \qquad \phi_{r_2}(B)S_t = \psi_{q_2}(B)b_{2,t}.$$

   Here $By_t = y_{t-1}$ such that B is that Backward Shift Operator, ϕ_j, ψ_j are real polynomials of degrees q_j. It is required that the roots of $\phi_j(B)$ are outside the unit circle, and the roots of $\psi_j(B)$ are on or outside the unit circle. Note $I_t, b_{1,t}, b_{2,t}$ are independent white noise processes each with zero mean and individual variance.
3. The full procedure for fitting such a model to the series y_t can be found in the book by Box and Jenkins below, this involves monitoring the auto-correlation, partial auto-correlation and the periodogram to aid model selection. An automatic ARIMA fitting procedure exists in many software packages such as the **auto.arima** function within **R**. The aim is to accurately measure the whole of the dependence structure of the series and their components, leaving uncorrelated residuals.
4. Given estimated seasonal component $\hat{S}_t = \frac{\hat{\psi}_{q_2}}{\hat{\phi}_{r_2}(B)}(B)b_{2,t}$ a seasonally adjusted series can obtained through subtraction: $\hat{y}_t = y_t - S_t$.

Note that in this work the commonly used 'airline model' is detailed, where

$$(1 - B)(1 - B^{12})y_t = (1 - \theta_1 B)(1 - \theta_{12}B^{12})c_t$$

and c_t is a white noise process with variance σ_{c_t}.

Derivation

Model based on a previous multiplicative model detailed in

Box, G. E. P. & Jenkins, G.M. (1970), Time series analysis: forecasting and control, Holden-Day.

When comparing the weights given by the ARIMA models to the filters, compparisons where made to the X-11 Procedure.

Shiskin, J. & Young, A. H. & Musgrave, J. C. (1965), The X-11 variant of the census method II seasonal adjustment program, US Department of Commerce, Bureau of the Census.

Development

ARIMA models are incorporated into the X-11 procedure, providing more robust extrapolation at the terminal ends of a series.

Dagum, E. B. (1978), Modelling, Forecasting and Seasonally Adjusting Economic Time Series with the X-11 ARIMA Method, Journal of the Royal Statistical Society: Series D, 27.3/4 203-216.

Software for fitting ARIMA models is available in most statistical software.

SABL

1979

Cleveland, W. S. & Dunn, D. M., & Terpenning, I. J. (1979), SABL: A resistant seasonal adjustment procedure with graphical methods for interpretation and diagnosis, In Seasonal Analysis of Economic Time Series (pp. 201-241), NBER.

Non-Parametric

Keywords: Power Transformation, Moving Median, Moving Average, Polynomial Estimation.

Strengths

- Use of moving medians makes the smoothing procedures more resistant to extreme values.
- Multiple methods used collaboratively aims to reduce 'leakage' of components.

Weaknesses

- Testing the stability of the seasonal component involved assumes it is perfectly stable.
- A large amount of smoothing steps are involved.

Description

An iterative procedure using locally fitted polynomials, moving medians and moving averages is used to estimate components. The methods are used in co-ordination with each other and are 'spliced' together.

Pre-Adjustments

An adjustment of the original series is made dependent on minimal residuals of a number of fitted models.

Method

1. Given pre-adjusted series y_t divide it into 9 sections. The midmean (the mean value between the 25th and 75th quartile) of time t and series y_t is calculated for each section. A line of best fit is then computed and a moving median of length 12 is then applied to the residuals, giving results by a_t. Then a weighted 24 point moving average is applied to the moving median values, giving b_t.
2. Apply a 12 point moving average to a_t and denote this c_l, and create preliminary trend estimate:
$$\tau_t^P = \begin{cases} c_t & \text{for } 12 \leq t \leq 17 \text{ and } n-16 \leq t \leq n-11 \\ (1-\cos((t-17.5)\pi/12) + \cos((t-17.5)\pi/12)c_t & \text{for } 18 \leq t \leq 29 \text{ and } n-28 \leq t \leq n-17 \\ b_t & \text{for } 30 \leq t \leq n-29 \end{cases}$$
3. Given a raw seasonal $s_t^R = y_T - \tau_t^P$ create monthly subseries m_j across all years.
4. Smooth each series with a moving median of length 4 (applying an extension rule in the work), then repeatedly apply a moving median of three till the values are consistent (extended). Smooth these values by a method called splitting (in work), then repeatedly smooth using a 3 point moving median. Next a 2 point moving average is used (and extended by the work), to give series m_j^*.
5. Apply step 4 to the differences $m_j - m_j^*$ (and extended). Compose these to give seasonal component s_t^p.
6. Extend s_t^P by six points either end, by extending the necessary monthly series by taking the median of: m_{j-1}; $2m_{j-1} - m_{j-2}$; $(4m_{j-1} + m_{j-2} - 2m_{j-3})/3$. This is reflected analogously for backcasting.
7. Apply step 2 then a 2 point moving average and subtract from the original seasonal estimate, giving s_t^P.
8. The trend is re-estimated then from $y_t - s_t^P$ by applying step 1, a 12 month moving median (extended) and a 2 point moving average, which is all extended. Next apply step 4 and 5 to the trend estimate. Then for moving blocks of 15 points, fit local quadratic polynomials, giving output $a_{i,k}$ for $i = 1, \ldots, 15$ and $k = 1, \ldots, n/15$. Then create the trend estimate:
$$\tau_t = \begin{cases} a_{t,1} & \text{for } 1 \leq t \leq 7 \text{ and } n-6 \leq t \leq n \\ (1-\cos((t-7.5)\pi/8) + \cos((t-7.5)\pi/8)c_t & \text{for } 8 \leq t \leq 15 \text{ and } n-14 \leq t \leq n-7 \\ b_t & \text{for } n-15 \leq t \leq n-8 \end{cases}$$
9. Re-apply steps 4,5,6,7,8 and 9 with the new trend, obtaining another trend.
10. Re-apply steps 4,5,6,7 and 8, obtaining the final seasonal component.

Derivation

The algorithm holds many similarities to the X-11 seasonal adjustment software.

Ladiray, D., & Quenneville, B. (2012), Seasonal adjustment with the X-11 method (Vol. 158), Springer Science & Business Media.

Development

Non-linear methods of smoothing are reviewed, SABL included.

Velleman, P. F. , (1982), Applied Nonlinear Smoothing. Sociological Methodology, 13, 141-177.

Software no longer available.

Minimum Signal Extraction 1980

Burman, J. P. (1980), Seasonal Adjustment by Signal Extraction, Journal of the Royal Statistical Society, Series A, 143.3 321-337.

Parametric

Keywords: ARIMA Model, Spectrum Analysis, Decomposition.

Strengths

- ARIMA models can be used to forecast and backcast to improve fit of the models.
- Parametric model gives an explanation of the series.

Weaknesses

- The method is found to be less flexible than adjustments using the X-11 method.
- The model is evaluated only for low orders.

Description

The method proposes a signal extraction technique that can extract seasonal and trend components after fitting an ARIMA model to the series, coupled with forecasts and backcasts provided by the fitted model.

Pre-Adjustments

Extreme items are identified within the method between the preliminary and final adjustments. For multiplicative series, take logarithms.

Method

1. Given a seasonal decomposition of a series $y_t = T_t + S_t + I_t$, assume that it follows a ARIMA$(p,d,q)(P,D,Q)_m$ model where the denominator can be factorised into trend components ψ_T, seasonal components ψ_S and white noise a_t. Here B is the Backward Shift Operator (such that $By_t = y_{t-1}$), and the spectrum of the series $g_y(w)$ is given by:

$$y_t = \frac{\theta(B)}{\psi_T(B)\psi_S(B)}a_t, \qquad g_y(\omega) = f(e^{i\omega})f(e^{-i\omega})\sigma_a^2 = \frac{\theta(e^{i\omega})\theta(e^{-i\omega})}{\psi_T(e^{i\omega})\psi_T(e^{-i\omega})\psi_S(e^{i\omega})\psi_S(e^{-i\omega})}\sigma_a^2$$

$$= \frac{U(x)}{V_T(x)V_S(x)}\sigma_a^2 = h_y(x)\sigma_a^2. \quad (\text{where } x = \cos\omega)$$

$$\text{Note} \quad h_y(x) = Q(x) + \frac{R(x)}{V_T(x)V_S(x)} = Q(x) + \frac{R_T(x)}{V_T(x)} + \frac{R_S(x)}{V_S(x)} = h_I(x) + h_T(x) + h_S(x)$$

2. Note that this is not a unique decomposition of the spectrum, and as such the method suggests removing the minimum of the trend and seasonal spectral components and adding them to the irregular to give a unique decomposition. This is such that $\min h_S(x) = \epsilon_s$ and $h_S^*(x) = h_S(x) - \epsilon_S$ and similarly for trend T, with irregular now $h_I^*(x) = h_I(x) + \epsilon_S + \epsilon_M$.

3. As $x = \frac{1}{2}(e^{i\omega} + e^{-i\omega}) = \frac{1}{2}(B + F)$ where F is the Forward Shift Operator ($B = F^{-1}$) then $h_S^*(x) \equiv H_S(B,F)$ giving a minimum signal extraction filters:

$$\frac{h_S^*(x)}{h_y(x)} = \frac{H_s(B,F)\psi_T(B)\psi_T(F)}{\theta(B)\theta(F)} = \frac{C_S(B,F)}{\theta(B)\theta(F)} \quad \text{similarly} \quad \frac{h_T^*(x)}{h_y(x)} = \frac{C_T(B,F)}{\theta(B)\theta(F)}.$$

4. The model and signal extraction filters are fitted in a two stage process:

   (a) Fit ARIMA model to the series y_t using maximum likelihood techniques, forecasting and backcasting.

   (b) Generate filters using parameters estimated with fore/backcasts.

Derivation

Signal extraction for unobserved components from time series is briefly explained.

Whittle, P. (1963), Prediction and regulation by linear least-square methods, English Universities Press.

The maximum likelihood technique of fitting ARIMA models given in the method.

Osborn, D. R. (1977), Exact and Approximate Maximum Likelihood Estimators for Vector Moving Average Processes, Journal of the Royal Statistical Society, Series B, 39.1 114-118.

Development

Selection of the seasonal model and a 'canonical' model is developed.

Maravall, A. & Pierce, D. A. (1987), A Prototypical Seasonal Adjustment Model, Journal of Time Series Analysis, 8.2 177-193.

TRAMO/SEATS is developed using signal extraction, with changes to the decomposition algorithm.

Gomez, V. & A. Maravall (1997a), Program TRAMO and SEATS: Instructions for the User, Beta Version, Banco de Espana.

Software available through TRAMO/SEATS:

`www.bde.es/bde/en/secciones/servicios/Profesionales/Programas_estadi/Programas.html`

Stochastic Regression Analysis 1981

Havenner, A. & Swamy, P. A. V. B. (1981), A Random Coefficient Approach to Seasonal Adjustment of Economic Time Series, Journal of Econometrics, 15.2 177-209.

Parametric **Keywords:** Regression, Stochastic Components, Harmonic.

Strengths

- Simultaneous calculation of components is believed to be superior to step-wise procedures.
- Allows for analysis of both possible deterministic and stochastic components for the trend and seasonal.

Weaknesses

- The backward eliminating method of removing coefficients can be inefficent.

Description

The work extends the typical decomposition of a seasonal time series into trend, seasonal and irregular to include stochastic counterparts using harmonic regression.

Pre-Adjustments

No pre-adjustments are given in the work, however logarithms were taken of a multiplicative series used as an example and this should be extended across all multiplicative series.

Method

1. Given a time series y_t where $t = 1, 2, \ldots, 12T$ (for a monthly series) assume the series follows the model:

$$\begin{aligned} y_t =& \beta_{0,t} + \sum_{j=1}^{5}\left[\beta_{j,t}\cos\frac{2\pi j}{12}t + \beta_{j+5,t}\sin\frac{2\pi j}{12}t\right] + \beta_{11,t}(-1)^t + \beta_{12,t}t \\ =& \left[1, \cos\frac{\pi t}{6}, \ldots, \cos\frac{5\pi t}{6}, \sin\frac{\pi t}{6}, \ldots, \sin\frac{5\pi t}{6}, (-1)^t, t\right]\left[\beta_{0,t}, \beta_{1,t}, , \ldots, \beta_{12,t},\right]' = \mathbf{x}_t'\boldsymbol{\beta}_t \end{aligned}$$

   Assume $\mathbb{E}\boldsymbol{\beta}_t = \bar{\boldsymbol{\beta}}_t = (\bar{\beta}_0, \bar{\beta}_1, \ldots, \bar{\beta}_{12})$ and $(\boldsymbol{\beta}_t - \bar{\boldsymbol{\beta}}) = \boldsymbol{\Phi}(\boldsymbol{\beta}_{t-1} - \bar{\boldsymbol{\beta}}) + \mathbf{a}_t$ where the roots of $\boldsymbol{\Phi}$ are absolutely less than 1, and $\mathbf{a}_t$ is a white noise process with fixed variance. The coefficients represent

   - $(\bar{\beta}_0)$ the level of the series;
   - $(\beta_{0,t} - \bar{\beta}_0)$ is the remainder;
   - $(\bar{\beta}_{12}t)$ the deterministic trend;
   - $(\beta_{12,t} - \bar{\beta}_{12})t$ is the stochastic trend;
   - $\sum_{j=1}^{5}\left[\bar{\beta}_j\cos\frac{2\pi j}{12}t + \bar{\beta}_{j+5}\sin\frac{2\pi j}{12}t\right] + \bar{\beta}_{11}(-1)^t$ is the deterministic seasonal;
   - $\sum_{j=1}^{5}\left[(\beta_{j,t} - \bar{\beta}_j)\cos\frac{2\pi j}{12}t + (\beta_{j+5,t} - \bar{\beta}_{j+5})\sin\frac{2\pi j}{12}t\right] + (\beta_{11,t} - \bar{\beta}_{11})(-1)^t$ is the stochastic seasonal.

2. Estimation of the parameters, their corresponding standard errors, the residuals and their structure can be found within the work.
3. Hypothesis tests can be performed on the coefficients $\bar{\beta}_j$ to determine if they are significant. Adequacy of the model can be tested using a normalised sum of squares of a variable described within the work.
4. A seasonally adjusted series can then be obtained by subtracting the stochastic and deterministic seasonal factors from the original series (and exponentiating if the series is multiplicative).

Derivation

Further details of the model are given in an earlier work by the author.

Swamy, P.A.V.B. & Tinsley, P.A. (1980), Linear prediction and estimation methods for regression models with stationary stochastic coefficients, Journal of Econometrics, 12.2 103-142.

An overview of methods determining deterministic and stochastic seasonality is given, and a model for both described.

Pierce, D. A. (1979), Seasonal Adjustment When Both Deterministic and Stochastic Seasonality are Present, In A. Zellner, Seasonal Analysis of Economic Time Series, NBER 242-280.

Development

An efficient algorithm for computation of the model is determined and given.

Chang, I. & Hallahan, C. & Swamy, P. A. V. B. (1992), Efficient Computation of Stochastic Coefficients Models, In H. M. Amman, D. A. Besley, L. F. Pau, Computational Economics and Econometrics, Springer Netherlands, 43-53.

No software given in the work.

X-11-ARIMA 1988

Dagum, E. B. (1988), The X-II-ARIMA Seasonal Adjustment Method, Statistics Canada, Seasonal Adjustment and Time Series Staff.

Semi-Parametric **Keywords:** X-11 Algorithm, ARIMA models, Forecast, Backcast, Semi-Parametric.

Strengths

- Use of ARIMA models to forecast/backcast remove the need for asymmetrical filters.
- An explanation to the behaviour of the series can be given through a model.

Weaknesses

- Only a selection of models are available to use, which may not meet all requirements when fitted.
- ARIMA models have been shown not to be able to adequately model all series.

Description

X-11-ARIMA is a modified X-11 algorithm such that the series is fore/backcasted using ARIMA models to aid the use of asymmetrical filters.

Pre-Adjustments

Logs should be taken on a multiplicative series. The algorithm is given for monthly series but can be extended to other frequencies. Upon development to X-12-ARIMA (amongst other changes) the regARIMA modelling approach was incorporated to allow for fore/backcasts and regression variables such as calendar effects to be incorporated before trend and seasonal component estimation. X-13-ARIMA-SEATS goes further by adding the option of using signal extraction in place of the X-11 algorithm.

Method

1. Given a pre-adjusted series $y_t = y_{i,j}$ for $t = 1, 2, \ldots, n$; $i = 1, 2, \ldots, P$ and $j = 1, 2, \ldots, n/P$ where n is the length of the series and P is the periodicity (i.e. $P = 12$ for monthly data) assume it can be modelled by $y_t = \tau_t * s_t * \epsilon_t$ and fit the following ARIMA models, minimising the conditions, given below.

   Possible ARIMA models:
   - ARIMA$(0,1,1)(0,1,1)_P$
   - ARIMA$(2,1,2)(0,1,1)_P$
   - ARIMA$(2,1,2)(0,1,1)_P$

   Conditions to be met:
   - Absolute average forecasting error in the last three years is $<$ 12%. This condition is minimised to choose between sufficient models.
   - Chi-squared probability value is less than 10%.
   - There is no evidence of over-differencing.

2. If a model is fitted, create fore/backcasts, giving $y^*_{t^*}$ for $t^* = 1, \ldots, n + 2P$. Else, let $y^*_{t^*} = y_t$.
3. Compute a 2x12 Moving average on the series $y^*_{t^*}$ to given an initial estimate of the trend τ_{t^*}. This gives the initial seasonal-irregular as $y^*_{t^*}/\tau_{t^*}$.
4. Calculate a 3x3 moving average on each sub-period i across all years j of the seasonal irregular $y^*_{t^*}/\tau_{t^*}$ individually to obtain initial seasonal factors s_t.
5. Apply a 2x12 Moving average to s_t, extending the series by six observations either side by repeating the beginning/last seasonal factor. Divide by the moving average values and make a correction for outliers (detail given in work). With adjusted values, repeat step 4. These are preliminary seasonal factors.
6. Repeat step 5 on the factors, then divide into y_t to obtain a preliminary seasonally adjusted series.
7. Apply a 9/13/23 term Henderson average to this series, then divide into y_t for a seasonal-irregular series.
8. Calculate a 3x5 Moving average to the seasonal-irregular series, obtaining the new seasonal factors. Repeat step 5 on this series, and divide into the original series to give the final seasonally adjusted series.

Derivation

The X-11 algorithm is outlined by the US Census Bureau.

Ladiray, D. & Quenneville, B. (2012), Seasonal adjustment with the X-11 method, Springer Science & Business Media, Vol 158.

Derivation of Henderson moving averages is detailed.

Henderson, R. (1924), A new method of graduation, Transactions of the Actuarial Society of America, 25, 29-40.

Development

ARIMA models are officially incorporated by the US Census Bureau amongst others into X-12-ARIMA.

Findley, D. F. & Monsell, B. C. & Bell, W. R. & Otto, M. C. & Chen, B. C. (1998), New capabilities and methods of the X-12-ARIMA seasonal-adjustment program, Journal of Business & Economic Statistics, 16.2, 127-152.

Software not currently publicly available.

Structural Time Series Models 1989

Harvey, A. C. (1989), Forecasting, Structural Time Series Models and the Kalman Filter, Cambridge University Press.

Parametric

Keywords: State Space Models, Kalman Filter, Trigonometric, Dummy Variable.

Strengths

- Allows use of previous statistical knowledge of the components.
- Separable components allow for new elements to be added given knowledge of events.

Weaknesses

- A low level of previous knowledge of the behaviour of the series is required to begin.
- Complex models can have issues with stability and convergence if model is incorrectly specified.

Description

Structural time series models allow for the model of a time series to be defined in terms of unobserved components with certain behaviour attached to each component which then has a direct interpretation.

Pre-Adjustments

Pre-adjustments can be made to the data, however there are methods to add components into the model to account for outliers and calendar effects. Taking logarithms of multiplicative series is advised.

Method

1. Given a series represented in basic structural form with seasonal dummy variables $y_t = T_t + S_t + I_t$, for $t = 1, \dots, T$ with seasonal period s, the trend T_t and seasonal S_t are represented by

$$T_t = T_{t-1} + \beta_{t-1} + \eta_t, \quad \beta_t = \beta_{t-1} + \zeta_t, \quad \sum_{j=0}^{s-1} S_{t-j} = \omega_t,$$

where $\eta_t, \zeta_t, \omega_t, \epsilon_t$ are uncorrelated white noise. This leads to a state space form of $y_t = [1\ 0\ 1\ 0\ 0]\,\boldsymbol{\alpha}_t + I_t$:

$$\boldsymbol{\alpha}_t = \begin{bmatrix} T_t \\ \beta_t \\ \vdots \\ S_t \\ S_{t-1} \\ \vdots \\ S_{t-s+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 & \vdots & & \mathbf{0} & & \\ 0 & 1 & \vdots & & & & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ & \mathbf{0} & \vdots & -1 & -1 & \cdots & -1 \\ & & \vdots & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ & & \vdots & 0 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} T_{t-1} \\ \beta_{t-1} \\ \vdots \\ S_{t-1} \\ S_{t-2} \\ \vdots \\ S_{t-s} \end{bmatrix} + \begin{bmatrix} \eta_t \\ \zeta_t \\ \vdots \\ \omega_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

2. If a trigonometric seasonality is to be defined, then the series $y_t = T_t + S_t + I_t$ is defined by:

$$S_t = \sum_{j=1}^{[s/2]} S_{j,t}, \quad S_{j,t} = S_{j,t-1}\cos\lambda_j + S^*_{j,t-1}\sin\lambda_j + \omega_{j,t}, \quad S^*_{j,t} = -S_{j,t-1}\sin\lambda_j + S^*_{j,t-1}\cos\lambda_j + \omega^*_{j,t},$$

This leads to a state space form of $y_t = [1\ 0\ 1\ 0\ 1]\,\boldsymbol{\alpha}_t + \epsilon_t$ where

$$\boldsymbol{\alpha}_t = \begin{bmatrix} T_t \\ \beta_t \\ \vdots \\ S_{1,t} \\ S^*_{1,t} \\ \vdots \\ S_{[s/2],t} \end{bmatrix} = \begin{bmatrix} 1 & 1 & \vdots & & \mathbf{0} & \vdots & 0 \\ 0 & 1 & \vdots & & & \vdots & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ & \mathbf{0} & \vdots & & \mathbf{C}^1_{[s/2]} & \vdots & -1 \\ & & \vdots & & & \vdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ & \mathbf{0}' & \vdots & & \mathbf{0}' & \vdots & -1 \end{bmatrix} \begin{bmatrix} T_{t-1} \\ \beta_{t-1} \\ \vdots \\ S_{t-1} \\ S_{t-2} \\ \vdots \\ S_{t-s} \end{bmatrix} + \begin{bmatrix} \eta_t \\ \zeta_t \\ \vdots \\ \omega_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \qquad \mathbf{C}^1_{[s/2]} = \begin{bmatrix} \mathbf{C}_1 \\ \mathbf{C}_2 \\ \vdots \\ \mathbf{C}_{[s/2]} \end{bmatrix},$$

$$\mathbf{C}_j = \begin{bmatrix} \cos\frac{2\pi j}{s} & \sin\frac{2\pi j}{s} \\ -\sin\frac{2\pi j}{s} & \cos\frac{2\pi j}{s} \end{bmatrix}.$$

Derivation of the components can then be calculated using the Kalman Filter, given an initial estimate of $\boldsymbol{\alpha}_0$.

Derivation

The Kalman Filter:

Kalman, R. E. (1960), A New Approach to Linear Filtering and Prediction Problems, Transactions of the ASME - Journal of Basic Engineering, 82.1 35-45.

State space models are used in systems theory:

Hutchinson, C. E. (1984), The Kalman Filter Applied to Aerospace and Electronic Systems, IEEE Transactions on Aerospace and Electronic Systems, 20 500-504.

Development

State Space models are used to look for shifts in the seasonal pattern.

Penzer, J. (2006), Diagnosing seasonal shifts in time series using state space models, Statistical Methodology, 3.3 193-210.

Software available as part of STAMP `http://stamp-software.com/`

General Linear Abstraction of Seasonality 1990

Young, T. & Young, J. (1990), The Bank's seasonal adjustment method: GLAS.

Non-parametric

Keywords: Moving Averages, Multivariate, Component Series, Cumulative.

Strengths

- The procedure is believed to be flexible to all expected variations in a seasonal series.
- Linearity of the procedure allows for explanation and reconstruction.

Weaknesses

- Assuming that the series used within adjustment explain all seasonal variations.
- Only a additive model can be used.
- Asymmetrical weights are used when filtering endpoints.

Description

A multivariate seasonal adjustment procedure in the context of monetary statistics is outlined and described, using a common linear filter across a number of series with the aim of balance.

Pre-Adjustments

Series which are deemed multiplicative are modelled in an additive approach, logarithmic transformation is not suggested. Manual modifications are made for detected outliers.

Method

1. Given a system of m time series all of length n with month and yearly indexes i, j, denote their matrix by $\mathbf{Y}$. Assume that each row (denoting a component series now as $y_t, t = 1, 2, \ldots, n$) can be modelled by the additive seasonal decomposition $y_t = \tau_t + s_t + \epsilon_t$, where τ_t is the trend, s_t is the seasonal component and ϵ_t is the irregular or residual component. Now for each row complete steps 2-5.
2. Calculate a simple 12x12 (23-point) triangular moving average to estimate trend. This has weights:

$$(0.0069, 0.0139, 0.0208, 0.0278, 0.0347, 0.0417, 0.0486, 0.0556, 0.625, 0.0694, 0.0764, \underline{0.0833}).$$

   Asymmetric weights are used on a filter of the same length for endpoints.
3. Subtract this trend estimate from the original series, $y_t - \hat{\tau}_t$. Decompose this series into yearly observations for each month, $y^*_{i,j}$
4. Now calculate a 3x3 (5-point) triangular moving average applied to the $y^*_{i,j}$. This is weighted as:

$$(0.1111, 0.2222, 0.3333, 0.2222, 0.1111).$$

   At endpoints, the size of the filter is decreased sequentially towards the end but maintains a simple moving average.
5. This can be then composed back into a seasonal estimate $\hat{s}_t$, which can be subtracted from the original series to give the seasonally adjusted series $y^S_t = y_t - \hat{s}_t$ with irregular $\hat{\epsilon}_t = y^S_t - \hat{\tau}_t$.
6. Denoting each series component with a subscript $k = 1, 2, \ldots, m$, the cumulative seasonal, trend, irregular and seasonally adjusted series are estimated simply as:

$$\bar{s}_t = \frac{1}{m}\sum_{k=1}^{m} \hat{s}_{t,k}, \quad \bar{\tau}_t = \frac{1}{m}\sum_{k=1}^{m} \hat{\tau}_{t,k}, \quad \bar{\epsilon}_t = \frac{1}{m}\sum_{k=1}^{m} \hat{\epsilon}_{t,k}, \quad \bar{y}^S_t = \frac{1}{m}\sum_{k=1}^{m} y_{t,k} - \hat{s}_t$$

Derivation

Weights of the asymmetrical filters are derived using the 'minimum revision' technique.

Lane, R. O. D. (1972), Minimal revision trend estimates, Technical Report Research Exercise Note 8/72, Central Statistical Office, London.

Development

The Bank compares X-12-ARIMA to GLAS.

Thorp, J. (2003), Change of seasonal adjustment method to X-12-ARIMA, Monetary and Financial Statistics, 4-8.

Software not publicly available.

Local Regression Smoothing 1990

Cleveland, R. B. & Cleveland, W. S. & McRae, J. E. & Terpenning, I. (1990), STL: A seasonal-trend decomposition procedure based on loess, Journal of Official Statistics, 6.1 3-73.

Parametric **Keywords:** Local Regression, Moving Average, Decomposition, Iterative.

Strengths

- Level of variation within each component can be set by user.
- Adaptable algorithm which can handle different types of evolution within each component.

Weaknesses

- Large amount of parameters to choose, poor choice can lead to poor estimation.
- It has been shown that there is heavier than expected smoothing at the ends of a series.

Description

STL is a seasonal trend decomposition using Local Regression with moving averages. Choice of the parameters allows allocation of variance amongst each component, methods for selecting them are given in the work.

Pre-Adjustments

No pre-adjustments for calendar effects are given (though a method is suggested), outliers are dealt with using weights to remove their impact on estimation. Logarithms should be taken for multiplicative series.

Method

Given a series y_t, assume the basic structural model $y_t = T_t + S_t + I_t, t = 1, 2, \ldots, N$ with periodicity s ($s = 12$ for monthly data). Choose smoothing parameters q for the seasonal, n_l for the low-pass filter and n_t for the trend. Also choose the degree of the fitted seasonal polynomials d, dependent on the level of variation. Lastly choose p the moving average parameter when smoothing for the low-pass filter. Initialise $T_t = 0, \rho_t = 1 \ \ \forall t$.

1. Remove the trend from the series, giving $y_t^{-T} = y_t - T_t$.
2. For each cycles subseries (each month for example) smooth these by Local Regression. Given a subseries of the de-trended values for each period ($j = 1, 2, \ldots s$); $x_{i,j}, (i = 1, 2, \ldots, N_j \leq N/s)$ for each data point $\{x^*_{0,j}, x_{1,j}, \ldots x_{N_j,j}, x^*_{N_j+1,j}\}$ (where * denotes an unobserved but still estimated observation):
   (a) Take q values around $x_{i,j}$ as the subset $\{x_{k,j}\}$, $k = i - q/2, (i + 1) - q/2, \ldots, i, \ldots i + q/2$. Let $\lambda_q(x_{i,j})$ be the distance of the q^{th} farthest $x_{k,j}$ from $x_{i,j}$, and calculate weights $v_{k,j}$:
   $$v_{k,j}(x_{k,j}) = W\left(\frac{|x_{k,j} - x_{i,j}|}{\lambda_q(x_{i,j})}\right)\rho_{(j-1)s+k}, \quad W(u) = \begin{cases}(1-u^3)^3 & 0 \leq u < 1\\ 0 & u \geq 1\end{cases}$$
   Note if $k \notin \{i\}$ then $\lambda_q(x_{i,j}) = \lambda_{N_j}(x_{i,j})\frac{q}{N_j}$.
   (b) Fit a polynomial g of degree d to the set of points $\{v_{k,j}(x_{k,j}) \cdot x_{k,j} : i = 1, 2, \ldots, N_j.\}$. Evaluate $g(x_{i,j})$ and record as S_a, where $a = (j-1)s + i = -s, -s+1, \ldots, 0, 1, \ldots, N, N+1, \ldots N+s$.
3. Given initial seasonal estimates $\{S_a\}$, apply a p x p x 3 moving average to the series, and apply the loess smoothing described in (a) and (b) on the new series, with $d = 1$ and $q = n_l$. Define the output as L_t.
4. Re-estimate the seasonal components now as $S^*_t = S_t - L_t$ and obtain a deseasonalised series, $y^*_t = y_t - S^*_t$. Use the loess smoothing above on y^*_t with $d = 1$ and $q = n_t$ to obtain a new estimate of the trend τ^*_t.
5. Calculate robustness weights ρ_t:
   $$\rho_t = B\left(\frac{|y^*_t - T^*_t|}{6\text{median}(|y^*_t - T^*_t|)}\right), \quad B(u) = \begin{cases}(1-u^2)^2 & \text{for } 0 \leq u < 1\\ 0 & \text{for } u \geq 1.\end{cases}$$

Step 5 is the *outer loop* and is optional. Steps 1-4 are the *inner loop*. Both can be reiterated.

Derivation

Local Regression is introduced looking at scatterplots.

Cleveland, W. S. (1979), Robust Locally Weighted Regression and Smoothing Scatterplots, Journal of the American Statistical Association, 74.368 829-836.

The robustness procedure in Step 5 is dervied.

Andrews, D. F. (1974), A Robust Method for Multiple Linear Regression, Technometrics, 16.4 523-531.

Development

Linear combinations after using STL for decomposition are used for prediction.

Theodosiou, M. (2010), Forecasting Issues: Ideas of Decomposition and Combination, Working Papers 2010-4, Central Bank of Cyprus.

Software available through the STL function in R

`https://stat.ethz.ch/R-manual/R-devel/library/stats/html/stl.html`

Periodic Structural Model 2004

Tripodis, Y. & Penzer, J. (2004), Periodic time series models: a structural approach, Technical report, London School of Economics.

Parametric **Keywords:** Structural Model, Periodic, Multivariate, Kalman Filter.

Strengths

- Allows for changing seasonality over each period.
- Gives direct interpretation of the components through the structural model.

Weaknesses

- Multiple tests must be used to arrive at the most appropriate model.

Description

The periodic structural model breaks a series into sub-periods (such as years) and models each year using the basic structural decomposition represented in state space form. This is then passed into the Kalman Filter for estimation.

Pre-Adjustments

No pre-adjustments are given in the work. For multiplicative series, logarithms should be taken.

Method

1. Given a series $y_{p,t}$ where $t = 1, 2, \ldots, N$ represents the year and N the amount of years in the series, $p = 1, 2, \ldots, s$ represents the current season (or month for example) and s is the amount of periods in a year, we represent each years observations by $y_n = (y_{1,n}, y_{2,n}, \ldots, y_{p,n})$.

2. Assume that

$$\begin{aligned} \text{(Basic Structural Decomposition):} \quad & y_{p,t} = \mu_{p,t} + \gamma_{p,t} + \epsilon_{p,t}, \quad \epsilon_{p,t} \sim NID(0, \sigma^2_{\epsilon,p}) \\ \text{(Trend Component):} \quad & \mu_{p,t} = \mu_{-1,t} + \eta_{p,t}, \quad \eta_{p,t} \sim NID(0, \sigma^2_{\eta,p}) \end{aligned}$$

   For the seasonal component $\gamma_{p,t}$ there are two choices.

   (a) The first is $\gamma_{p,t} = \gamma_{p,t-1} + \omega_{p,t}$, such that if $D = \text{diag}\{\sigma^2_{\omega,1}, \ldots, \sigma^2_{\omega,s}\}$ and $\mathbb{1}_s = [1, \ldots, 1]$ then vectorising as in y_n earlier, the following condition constrains the seasonal such that it sums to zero over each year:

$$\text{Var}(\omega_t) = D - \frac{D\mathbb{1}_p\mathbb{1}_p^{'}D}{\mathbb{1}_p^{'}D\mathbb{1}_p}$$

   (b) The second is a dummy variable approach, which can be expressed as:

$$\begin{pmatrix} 1 & 0 & \ldots & 0 \\ 1 & 1 & \ddots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \ldots & 1 \end{pmatrix} \begin{pmatrix} \gamma_{1,t} \\ \gamma_{2,t} \\ \vdots \\ \gamma_{s,t} \end{pmatrix} = \begin{pmatrix} 0 & -1 & \ldots & -1 \\ 0 & 0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & -1 \\ 0 & 0 & \ldots & 0 \end{pmatrix} \begin{pmatrix} \gamma_{1,t-1} \\ \gamma_{2,t-1} \\ \vdots \\ \gamma_{s,t-1} \end{pmatrix} + \begin{pmatrix} \omega_{1,t} \\ \omega_{2,t} \\ \vdots \\ \omega_{s,t} \end{pmatrix}$$

3. The Kalman filter can be applied to the appropriate decomposition using likelihood inference to conduct likelihood ratio tests on the seasonal heteroscedasticity

4. Aikaike's Information Criterion is then used to determine the significant periodic variances.

5. Tests for normality and independence of residuals are conducted using methods such as the Ljung-Box.

Derivation

State space models and described in detail.

Harvey, A. (1990), Forecasting, Structural Time Series Models and the Kalman Filter, Cambridge University Press.

Periodic models are used with economic series and forecasts are compared.

Novales, A. & de Fruto, R. F. (1997), Forecasting with periodic models. A comparison with time invariant coefficient models, International Journal of Forecasting, 13.3 393-405.

Development

Periodic structural models are used as part of wavelet benchamrking.

Sayal, H. & Aston, J. A. D. & Elliott, D. & Ombao, H. (2016). An introduction to applications of wavelet benchmarking with seasonal adjustment. Journal of the Royal Statistical Society: Series A (Statistics in Society), 180(3), 863-889. doi: 10.1111/rssa.12241

Software referenced in work: SSFPack in Ox `http://www.ssfpack.com/`

DAINTIES 2007

Ladiray, D., (2007), The DAINTIES Method, Eurostat.

Parametric

Keywords: Linear Regression, Moving Average, Asymmetric, Weighted Filters.

Strengths

- The range of filters when weighted can capture both additive and multiplicative seasonality simultaneously.

Weaknesses

- Only returns the seasonally adjusted series, not returning trend.

Description

A weighted composite of numerous filters applied to the data result in a seasonally adjusted series. The filters are applied according to conditions on length and the range of the series.

Pre-Adjustments

Filters are applied assuming that the series is multiplicative/additive and are weighted afterwards. Modification of extreme values is given with the work.

Method

1. Assume a series y_t (for $t = 1, 2, \ldots, n$ and periodicity p) can be decomposed such that $y_t = \tau_t + s_t + \epsilon_t$ where τ_t is the trend, s_t is the seasonal and ϵ_t is the irregular. Model this series, with constraint as:

$$y_t = a_0 + a_1 t + a_2 t^2 + a_3 t^3 + \sum_{i=1}^{p} \alpha_i \mathbb{I}_{[t \equiv i-1 \mod p]} + \epsilon_t, \quad \sum_{i=1}^{p} \alpha_i = 0.$$

   Here the cubic polynomial represents trend τ_t and α_t are the seasonal estimated factors for s_t.

2. Using least squares regression, an estimate of the seasonal part of the model above is calculated on blocks of m observations. For the first m observations, those calculated by regression are kept as the component estimated s_t for $t = 1, 2, \ldots, m$, then for the next block from 2 to m only the components estimated in position $m+1$ are kept, and similarly until the components are estimated till $n-m$, denoted as y_t^{SA} for $t = 1, 2, \ldots, n-m$.

3. For the endpoints, a number of asymmetric moving averages are applied to the series to estimate each point. These filters are given in the work and not detailed here, but they are used based on the following conditions:

Length of series	$y_t > 0 \ \forall t$	$\exists t : y_t \leq 0$
$n < 3p+1$	No filters applied	No filters applied
$3p+1 \leq n < 4p+1$	A_S	A_S, M_S
$4p+1 \leq n < 5p+1$	A_S, A_M	A_S, A_M, M_S, M_M
$5p+1 \leq n$	A_S, A_M, A_L	$A_S, A_M, A_L, M_S, M_M, M_L$

   Where A_S, A_M, A_L are small, medium and long filters with coefficients given within the work, and M_S, M_M, M_L are the same filters applied to the logarithm of the original series y_t.

4. Let the results of result of each filter applied to the series as $y_{t,k}^*$ for k estimations. At time t we consider the observations $y_{t-p+1,k}, y_{t-p+2,k}, \ldots, y_{t,k}$ and fit a straight line, computing the residual sum of squares at each point giving $\delta_{t,k}$ (where $\delta_{t,k} = \delta_{p,k}$ for $t < p$). Then calculate for all t: $\bar{\delta}_t^{add} = \frac{1}{3} \sum_{l=1}^{k/2} \delta_{t,l}^{add}$ such that only additive estimations are used.

5. Calculate weights, and calculate endpoint seasonally adjusted values:

$$\lambda_{t,k} = \begin{cases} 0 & \text{if } k \text{ is multiplicative and } \delta_{t,k} > 0.8\bar{\delta}_t^{add} \\ 1 & \text{if } \delta_{p,k} = 0 \text{ (Note: All other weights = 0)} \\ 1/\delta_{p,k} & else \end{cases} \qquad y_t^{SA} = \frac{\sum_{l=1}^{k} \lambda_{t,k} y_{t,k}^*}{\sum_{l=1}^{k} \lambda_{t,k}} \quad (t = n-m+1, \ldots, n)$$

Derivation

DAINTIES was preceeded by an algorithm named SEABIRD

Mesnage, M. (1968), Elimination des variations saisonnires: La nouvelle mthode de l'OSCE, Statistische Studien und Erhebungen / Statistisches Amt der Europischen Gemeinschaften, 1.

Development

A number of seasonal adjustment procedures are compared including DAINTIES.

Franses, P. H. & Paap, R. & Fok, D. (2005), Performance of seasonal adjustment procedures: Simulation and empirical results. Erasmus University Rotterdam, Econometric Institute.

Software not publicly available.

Model Based X-11 2010

McElroy, T. (2010), A Nonlinear Algorithm for Seasonal Adjustment in Multiplicative Component Decompositions, Studies in Nonlinear Dynamics & Econometrics, 14.4 1-23.

Parametric

Keywords: State Space, X-11, Filter, Model Based, Developing Method.

Strengths

- The additive algorithm has been shown to converge to true expectations.
- Removes trend bias shown to be in multiplicative series.

Weaknesses

- The multiplicative algorithm does not always converge.
- Relies on a previously fitted ARIMA model which could be misspecified.

Description

The method is a model based approach to an adaptation of the X-11 algorithm. The aim is to reduce the bias caused by taking logarithms of multiplicative series and using ARIMA models on the new series. The procedure is outlined for both additive and multiplicative versions of the X-11 algorithm.

Pre-Adjustments

No pre-adjustments are given in the work.

Method

1. Given a series y_t, assume it can be decomposed into the basic structural model such that $y_t = T_t + S_t + I_t$, where s_t is the seasonal, T_t the trend, and I_t the irregular. If the series is multiplicative, replace addition (subtraction) with multiplication (division) where necessary throughout. Begin by fitting an ARIMA model.

2. Fit the stationary process $x_t = \delta(B)y_t$, where B is the Backward Shift Operator. Here $\delta(B)$ has roots only on the unit circle and $\delta(0) = 1$, representing the seasonal and non-seasonal differencing. Factorise $\delta(B) = \delta^S(B)\delta^T(B)$ for the seasonal and trend, with no common zeros and $\deg(\delta^S) + \deg(\delta^T) = \deg(\delta)$.

3. Construct matrices Δ^S, where $d^S = \deg(\delta^S)$ (with a similar construction for Δ^T):

$$\Delta^S = \begin{bmatrix} \delta^S_{d^S} & \delta^S_{d^S-1} & \dots & \delta^S_1 & 1 & 0 & \dots & 0 \\ 0 & \delta^S_{d^S} & \delta^S_{d^S-1} & \dots & \delta^s_1 & 1 & 0 & \dots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \delta^S_{d^S} & \delta^S_{d^S-1} & \dots & \delta^S_1 & 1 \end{bmatrix}, \Delta^S_{i,j} = \begin{cases} \delta_{i-j+d^S} & 0 \le i-j+d^S \le d^S \\ 0 & \text{otherwise} \end{cases}$$

where each δ_i are the coefficients of the appropriate differencing operator.

4. Let $y^S_t = S_t + I_t$ such that the trend has been removed, and similarly for y^T_t and construct row vectors from them. These can be calculated using the model fitted in Step (1). Calculate covariance matrices $\Sigma_I, \Sigma_{\Delta^S(y^S)'}, \Sigma_{\Delta^T(y_T)'}$ of each row vector I, y^S, t^T respectively.

5. Given that $\mathbb{I}$ is the (n x n) identity matrix, construct filters:

$$F^S = \mathbb{I} - \Sigma_I \Delta_S^{'} \Sigma^{-1}_{\Delta^S(y^S)'} \Delta^S, \quad F^T = \mathbb{I} - \Sigma_I \Delta_T^{'} \Sigma^{-1}_{\Delta^T(y^T)'} \Delta^T$$

6. Iteratively apply the filters to the data following the appropriate algorithm:

Additive	*Multiplicative*
Initialise $\hat{S}^{(0)}$ as any given vector	Initialise $\hat{S}^{(0)} = \mathbb{1}_n = (1, 1, \dots, 1)$
For $i = 1$ to sufficient convergence	For $i = 1$ to sufficient convergence
$\hat{T}^{(i)} = F^T(y - \hat{S}^{(i-1)})$,	$\hat{T}^{(i)} = F^T(y \div \hat{S}^{(i-1)})$,
$\hat{S}^{(i)} = F^S(y - \hat{T}^{(i)})$	$\hat{S}^{(i)} = \mathbb{1}_n + F^S(y \div \hat{T}^{(i)} - \mathbb{1}_n)$
end	end

7. A seasonally adjusted series can then be achieved by subtracting (dividing) $\hat{S}^{conv}$ from y_t.

Derivation

Proof of the algorithm converging to true expectations.

Mc Elroy, T. & Sutcliffe, A. (2004), An Iterated Parametric Approach to Nonstationary Signal Extraction, Statistical Research Division, U.S. Bureau of the Census.

Original X-11 Algorithm detailed.

Shiskin, J. & Young, A. H. & Musgrave, J. C. (1965), The X-11 Variant of the Census Method II Seasonal Adjustment Program, U.S. Department of Commerce, Bureau of the Census.

Software available in supplementary materials: `http://www.degruyter.com/view/j/snde.2010.14.4/snde.2010.14.4.1756/snde1756_supplementary_5.zip`

Singular Spectrum Analysis 2013

Chen Q. & van Dam, T. & Sneeuw, N. & Collilieux X. & Weigelt M. & Rebischung, P. (2013), Singular spectrum analysis for modeling seasonal signals from GPS time series, Journal of Geodynamics, 72 25-35.

Non-Parametric **Keywords:** Developing Method, Spectrum Analysis, Eigenvalue Decomposition.

Strengths

- Cyclic behaviour can be analysed over a number of scales simultaneously.
- Can account for changing patterns in seasonality.

Weaknesses

- Analysis of which components are seasonal or trend can be subjective.
- No analysis of error is made through the method.

Description

SSA decomposes a series into components by considering their cyclical behaviour through eigenvector analysis over a moving window. Selection of the length of window M is not given with the work, but it should be greater than all lengths of expected periodicities.

Pre-Adjustments

No details are given of any pre-adjustments. For a multiplicative series, logarithms should be taken.

Method

1. Given a time series $y_t, t = 1, 2, \dots N$ embed it into a trajectory matrix D through a sliding window of length M, such that $N' = N - M + 1$, and calculate the covariance matrix C:

$$D = \begin{pmatrix} y_1 & y_2 & \cdots & y_M \\ y_2 & y_3 & \cdots & y_{M+1} \\ \vdots & \vdots & \vdots & \vdots \\ y_{N'} & y_{N'+1} & \cdots & y_N \end{pmatrix}, \quad C = \begin{pmatrix} c_0 & c_1 & \cdots & c_{M-1} \\ c_1 & c_0 & \cdots & c_{M-2} \\ \vdots & \vdots & \vdots & \vdots \\ c_{M-1} & c_{M-2} & \cdots & c_0 \end{pmatrix}, \quad c_j = \frac{\sum_{i=1}^{N-j} x_i x_{i+j}}{N-j}, \quad 0 \le j \le M-1$$

2. Apply an eigenvalue decomposition to the matrix C to obtain eigenvalues λ_k and eigenvectors E^k, sort the vectors into descending order of λ_k for $k = 1, 2, \dots, M$.

3. Define the principal components as a_k^i and reconstructed components y_i^k as:

$$a_i^k = \sum_{j=1}^{M} y_{i+j} E_j^k, \quad 0 \le i \le N - M \qquad y_i^k = \begin{cases} \sum_{j=1}^{i} \frac{a_{i-j}^k E_j^k}{i} & 1 \le i \le M-1 \\ \sum_{j=1}^{M} \frac{a_{i-j}^k E_j^k}{M} & M \le i \le N-M+1 \\ \sum_{j=i-N+M}^{M} \frac{a_{i-j}^k E_j^k}{N-i+1} & N-M+2 \le i \le N \end{cases}$$

4. Seasonal behaviour can then be identified by comparing neighbouring eigenvalues and eigenvectors, following three criteria:
   - consecutive eigenvalues λ_k and λ_{k+1} are approximately equal;
   - the sequences given by eigenvectors E^k and E^{k+1} are periodic and approximately phase shifted by $\pi/2$;
   - their principal components a_i^k and a_i^{k+1} are approximately phase shifted by $\pi/2$. For all k^s considered part of the seasonal component, group together: $s_t = \sum_{k^s} y_t^{k^s}$.

5. Those with long running cycles are the trend estimates, and should be grouped similarly: $\tau_t = \sum_{k^t} y_t^{k^t}$.

6. A seasonally adjusted series can then be obtained by $y_t^* = y_t - s_t$.

Derivation

The criteria for selection of each component is discussed.

Plaut, G. & Vautard, R. (1994), Spells of Low-Frequency Oscillations and Weather Regimes in the Northern Hemisphere, Journal of the Atmospheric Sciences, 51.2 210-236.

A more thorough description of SSA is given.

Broomhead, D. S. & King, G. P. (1986), Extracting qualitative dynamics from experimental data, Physica D: Nonlinear Phenomena, 20.2 217-236.

Development

Different variations of SSA are presented and implemented on seasonal series.

Golyandina, N., & Shlemov, A. (2015). Variations of singular spectrum analysis for separability improvement: non-orthogonal decompositions of time series. Statistics and Its Interface, 8(3), 277-294. doi: 10.4310/sii.2015.v8.n3.a3

Software available within the RSSA package in R
`https://cran.r-project.org/web/packages/Rssa/index.html`

Wavelet Analysis 2014

Stachura, M. (2014), Detecting Seasonality via Wavelet Methods, Studia Ekonomiczne, 207 223-232.

Non-Parametric **Keywords:** Wavelet, Multi-Resolution Analysis, Filter, Devleoping Method.

Strengths

- Analysis at multiple scales allows different length cycles to be analysed.
- Seasonal and trend are decomposed simultaneously, not iteratively.

Weaknesses

- Choice of wavelet used based on experimentation.
- An irregular is not described in the work.
- Discrete Wavelet Transform is not rotation invariant.

Description

Using a wavelet decomposition, a series can be reduced into the sum of the wavelet transformation across different scales. Each scale represents a different length of cycle and thus can be used to extract trend (very short cycles) and seasonal components (longer cycles) simultaneously.

Pre-Adjustments

No pre-adjustments are given in the work, logarithms should be taken for multiplicative series.

Method

1. Given a series $Y = (y_1, y_2, \dots, y_n)^T$, scales $j = 1, 2, \dots J$ each at locations $k = 1, 2, \dots 2^j$, define the wavelet coefficients as

$$W_{j,k} = \sum_{t=1}^{T} y_t \psi_{j,k}\left(\frac{t}{n}\right), \quad \psi_{j,k}(s) = 2^{-\frac{j}{2}}\psi(2^{-j}s - k),$$

   where ψ is the chosen mother wavelet, and $\psi_{j,k}$ form an orthonormal basis.

2. Collect the coefficients such that $\mathbf{W}_j = (W_{j,1}, W_{j,2}, \dots, W_{j,2^J-j})$ and $\mathbf{W} = (\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_J, \mathbf{V}_J)^T$ where $\mathbf{W}$ is the discrete wavelet transform of Y and $\mathbf{V}$ is the residual vector. Note that the calculation of the coefficients can be done recursively:

   (a) The chosen wavelet has associated scaling filter $\{h_l\}_{l=0,1,\dots,L-1}$ and conjugate filter $\{g_l\}_{l=0,1,\dots,L-1}$ such that $g_l = (-1)^{l+1}h_{L-l-1}$.

   (b) Apply $\{h_l\}, \{g_l\}$ to Y to get $\mathbf{W}_1, V_1$ respectively.

   (c) Repeat previous step, but apply $\{h_l\}, \{g_l\}$ to V_1 to get $\mathbf{W}_2, V_2$ respectively.

   (d) Repeat until V_{n-2^J+1} is obtained.

   Then $\mathbf{V}_J = (V_1, \dots, V_{n-2^J+1})$.

3. With residual vector and wavelet coefficients, it can be seen that $\mathbf{W} = \mathbf{\Psi}Y$ and thus, $Y = \mathbf{\Psi}^T\mathbf{W}$ where the orthonormal matrix $\mathbf{\Psi}$ is determined by the wavelet used ψ. As such Y can now be expressed:

$$Y = \mathbf{\Psi}_1^T\mathbf{W}_1 + \mathbf{\Psi}_2^T\mathbf{W}_2 + \cdots + \mathbf{\Psi}_J^T\mathbf{W}_J + \mathbf{\Phi}_J^T\mathbf{V}_J = \mathbf{D}_1 + \mathbf{D}_2 + \cdots + \mathbf{D}_J + \mathbf{T}_J.$$

   Where $\mathbf{\Psi}_1, \dots, \mathbf{\Psi}_J, \mathbf{\Phi}_J$ are submatrices of $\mathbf{\Psi}$ that partition $\mathbf{W}$.

4. The $\mathbf{D}_j$ represent the cycles over scales j and the $\mathbf{T}_J$ represents the trend throughout the series. Therefore the series can then be represented by seasonal $\mathbf{S}$ and trend $\mathbf{T}$ and can thus be removed:

$$Y = \mathbf{D}_1 + \mathbf{D}_2 + \cdots + \mathbf{D}_J + \mathbf{T}_J = \mathbf{S} + \mathbf{T}$$

Derivation

Multi resolution analysis using a bi-orthogonal wavelet filter is used on GDP.

Yogo, M. (2008), Measuring business cycles: A wavelet analysis of economic time series, Economics Letters, 100.2 208-212.

Calculation of the wavelet coefficients using the recursive scheme and filters is outlined.

Percival, D. B. & Walden, A. T. (2006), Wavelet methods for time series analysis, C. U. P.

Development

Wavelet filtering is used for forecasting after decomposing a series into trend variations.

Joo, T. W. & Kim, S. B. (2015), Time series forecasting based on wavelet filtering, Expert Systems with Applications, 42.8 3868-3874.

Software available within R using packages such as Wavethresh:
`https://cran.r-project.org/web/packages/wavethresh/index.html`

Chapter 5

Wavelet Frequency Detection

Abstract Detecting the presence of a periodic component in a series can have significant impact upon both modelling and prediction. Common methods for detection and prediction of these components often use Fourier analysis and spectrum decomposition, or require information *a priori* about the periodicity expected. Here we present an alternative based upon Wavelet Decompositions, showing how a 'Variance Profile' of a periodic series can be seen from the wavelet spectrum. We present results on theoretical distributions and an empirical analysis of the effectiveness of a proposed algorithm. This is followed by an application upon daily data which contains a low frequency component within it.

5.1 Introduction

Detection of any repeating components within a series, such as seasonality, is an important step in understanding and predicting data. Often a series of interest is studied carefully to determine the most likely component relative to its context. This allows for careful consideration to be made between a number of components including (but not limited to) seasonality/periodicity. This is such that a part of the series of interest will reoccur within p points of the series. As a motivating example consider the series of French births given in Figure 5.1. There are different levels of seasonality occuring within the series, some on a smaller scale perhaps of

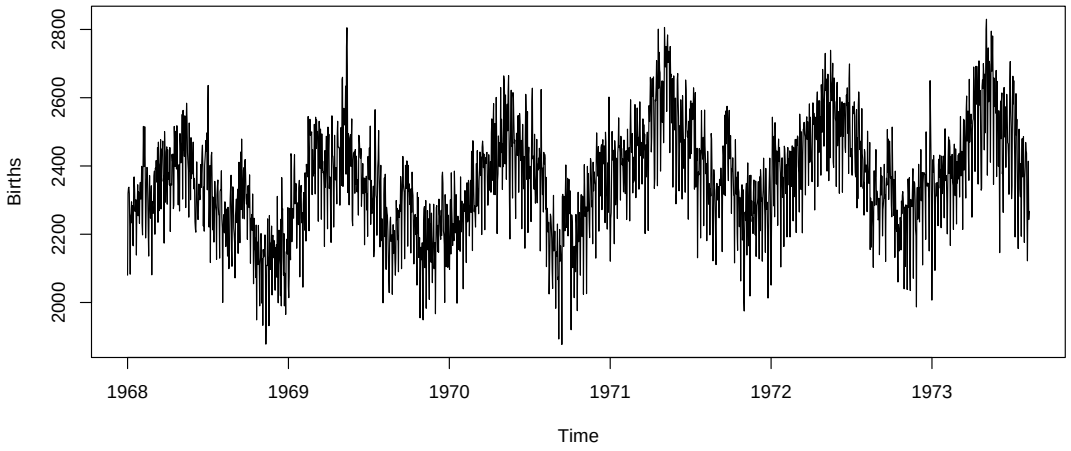


Figure 5.1: Daily published figures on French Births from January 1968 to August 1973.

weekly/monthly, but also on a larger scale of yearly. However to determine and test for these components, particularly those who do not show as often such as weekly, is no straightforward task task.

Typical seasonal adjustment software such as X-12-ARIMA-SEATS and JDemetra+ have a suite of tests available for testing such expected components. For example, aggregating the data across its expected period leads to the Kruskal-Wallis Test (Kruskal and Wallis, 1952) which tests the period specific mean ranks of a series for significant differences. Similarly an F-test can be computed on monthly means to determine seasonality (Ladiray and Quenneville, 2012). Monitoring the autocorrelation function and looking for significantly positive values at seasonal lags forms the QS test, transforming the same diagnostic into a periodogram and taking a weighted sum at seasonal periods can also be used to test for seasonality (Grudkowska, 2016). Further, one can visually inspect the periodogram at known seasonal frequencies also (Soukup and Findley, 1999). Model based approaches can also be taken, such as testing the significance of monthly variables fitted in a RegARIMA model. A number of these tests are further detailed and compared in Lytras et al. (2007) in a review of statistical software X12-ARIMA-SEATS, and similarly a discussion and combination of tests used in the software JDemetra+ is

given in Webel and Ollech (2018) where a random forest test is constructed from those described.

However, these tests require *a priori* knowledge of the periodicity p to be expected, but with the cost of data collection decreasing, the amount of series collected is growing exponentially. This is such that giving careful attention to any series becomes relatively expensive in comparison to an automatic procedure which comprises of two main steps: firstly the detection of a periodicity, and secondly estimating the value of p.

Detection of periodic components is often based on the use of diagnostic measures such as the periodogram, which measures the amplitude of a set of frequencies within a series, and the autocorrelation function, which measures the dependence of a time point to another in the future/past. As such often they test against the alternative that there is no periodicity present within a series. Such a test was developed in Fisher (1929) to determine if there were any significant peaks within the periodogram by comparing values different from the average under normal distributed data. Further tests for white noise can be conducted on the periodogram using the Bartlett B test (Bartlett, 1967) which compares against an expected Exponential distribution for points within the periodogram under the null. Similarly a test of independence can be used such as the Ljung-Box test (Ljung and Box, 1978), which tests the departure of autocorrelations from zero. Further, a test based on the haar wavelet transform (termed HWWN for Haar Wavelet White Noise), of which we discuss further in this work, looks to identify white noise by comparing against an expected flat spectrum (Nason and Savchev, 2014).

Combining these two aspects, detection and prediction of periodicity, has been a key research area known as 'periodicity mining' particularly prevalent for symbolic time series data, such that $y_t = aab, aba, baa, \ldots$. Key works include the use of 'sketches' which transformed the data into a different form to compare distances more efficiently was used in Indyk et al. (2000). Further, convolution over a number of proposed periodicities to determine rate and patterns was applied in Elfeky et al.

(2005). The literature is rich for periodicity mining of symbolic data and provides many motivating methodologies, but it is not our focus here and we refer the reader to Fournier-Viger et al. (2017) for a comprehensive review of their applications.

The use of decomposition was also applied in a semi-automatic procedure within Kanjilal et al. (1999) which applied a singular value decomposition to a numeric time series, such that each period becomes a row, to determine periodicity. However, at least two cycles were required and it becomes computationally expensive when used as an exhaustive search. Collection over a period is also featured in Schaidnagel and Laux (2013) where each period is summed appropriately and the pearson correlations are calculated to determine the likely period, given a pre-defined range however. However, not all methodology requires prior information, such as the fully automatic 'Autoperiod' algorithm (Vlachos et al., 2005) which combines the aforementioned periodogram and autocorrelation information to confirm 'hints' of periodicities within a series, however as noted within Puech and Boussard (2019) the algorithm can suffer under noise and they combine clustering and filtering techniques to aid this in a trade off with computational efficiency. Additional methodology was given within the forecast package (Hyndman, 2017) under the function `find.frequency` which detects local peaks within a spectrum and returns a proposed frequency, though there is no formal test. Yang and Su (2010) determine periodicities from peaks which are picked from a smoothed periodogram, and models an increasing amount of hamronics against the series of interest to model possible periodic effects (termed ARSER).

Our work introduces a new automatic procedure for detection and prediction of a periodic component by combining many of the features described in these works to aid in the case of low frequency and high noise environments. Through the use of Wavelets, which can be seen as a filtering or convolution process, we similarly decompose a series onto an efficient scale and compare distances between proposed periodicities expressed in the same domain. However, through the use of the information gained during this decomposition, we are able to determine a viable

search region without prior knowledge and reduce our computation significantly.

Wavelets have been used often within periodicity detection due to their ability to decompose a series across both time and frequency. There have been many previous applications of wavelets for periodicity detection with particular usage in Astronomy. In Szatmary et al. (1994) and Szatmary et al. (1996) wavelets were applied to a number of different occurances of periodicity and the resulting coefficients were studied widely. Periodic aggregation of generalised haar wavelet transforms to detect periodicities over a known range was used in Benedetto and Pfander (2002). Wavelets were used alongside autocorrelation and peak analysis for the determination of heart rate through non-contact in Sekine and Maeno (2011). Reconstruction from a wavelet transform was also used to remove low frequency components in head accelerometry measures in Sani et al. (2013). Further investigation into the behaviour of different wavelet coefficients was studied for critical time interval identification related to posture control in Zhang et al. (2015). Similarly, sign sequences of wavelet coefficients were used by Park and Kim (2018) within a search region to analyze repetitive tasks. Each of these studies looks to employ the unique properties of a wavelet decomposition, but none apply them in a fully automatic way, always requiring precursor knowledge of the behaviour expected, which is a particular strength of our work which we describe fully in the sections to follow.

Our work is structured as follows. Section 5.2 describes the discrete wavelet transform that we use, the findings on periodicity detection and proofs therein, discusses the creation of 'variance profiles' and a formal test before presenting the final algorithm. Section 5.3 then tests the algorithm over a number of cases and provides commentary on the results found. We then provide our motivating application of Daily French Birth data in Section 5.4. Finally we give concluding remarks overviewing the paper and proposing next steps in Section 5.5. Within the Appendix proofs can be found for results stated, and additional graphs not given in Section 5.2.

5.2 Methodology

5.2.1 Wavelet Tranforms

We review the methodology used in our work by introducting wavelets. Seminally described in Daubechies (1992), they allow the decomposition of a time series in two dimensions, scale and position. Scale is the resolution of the data that you view at a point; going from 'fine' where you view only a small amount of detail, to 'coarse' where you review a large portion of the data. Position then represents the point within the series where you view this. These two dimensions allow for identification of components within a series on both large and small scales, and over time. For the work conducted here we use the discrete version of wavelets of a time series y_t for $t = 1, 2, \ldots, n$ where each point is regularly spaced. This is such that wavelet coefficients (in respect to a wavelet basis ψ) are calculated by

$$d_{j,k} = \sum_{t=0}^{n} y_t \psi_{j,k}(t), \tag{5.1}$$

where $\psi_{j,k}(t) = 2^{-\frac{j}{2}}\psi(2^j t - k)$ is the relationship of each wavelet to the mother wavelet $\psi(x)$. The simplest wavelet is the Haar (Haar, 1910) given as:

$$\psi(x) = \begin{cases} 1 & \text{when } 0 \leq x < \frac{1}{2} \\ -1 & \text{when } \frac{1}{2} \leq x < 1 \\ 0 & \text{else} \end{cases}.$$

Within our work we proceed with the 'non-decimated' wavelet transform. This is such that we allow each individually calculated wavelet coefficient $d_{j,k}$ to possibly have overlapping supports. This can be seen as an overcomplete transformation, but we use the additional information gained from this to bolster our methodology. For further information on the different transformations and the benefits to each, we direct the reader to Nason (2010).

Continuing our example, the non-decimated Haar Wavelet Coefficients for scale

j and position k can be calculated by

$$d_{j,k} = \frac{1}{\sqrt{2}^j} \left(\sum_{i=0}^{2^{j-1}-1} y_{k+i} - \sum_{i=2^{j-1}}^{2^j-1} y_{k+i} \right).$$

This can be seen as a centered moving average filter with weights $\frac{1}{\sqrt{2}^j}$ for the first half of point considered and $-\frac{1}{\sqrt{2}^j}$ for the second half.

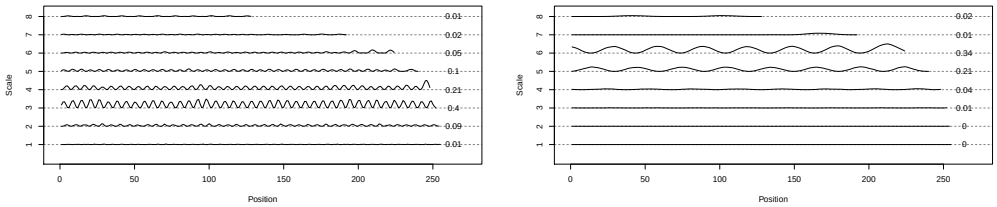
5.2.2 Transforms of periodicity

For the purpose of identifying periodicity, we focus on a periodic signal generated by a sinusoid with noise. This is a common component used in harmonic analysis where periodicities are explained by a combination of sine and cosine waves, and we begin with this component as we explore the effectiveness of our work. It is this model which we assume in our work now going forward, formally defined as:

$$y_t = \sin\left(\frac{2\pi t}{p}\right) + \epsilon_t \quad \text{where} \quad \epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2), \qquad t = 1, 2, \ldots, n \tag{5.2}$$

where p is the periodicity of the sinusoid and t is the time index. This model represents a simple repeating pattern over p points plus a level of noise which obscures it. We can decompose this series under a non-decimated wavelet transform and square the coefficients to get a wavelet spectrum, which shows the magnitude of each coefficient on each scale. This spectrum is useful for the identification of features of series. Such spectra of series which follow our model are given in Figure 5.2.

It is clear to see from the spectra that the periodicity of the sinusoid has a significant impact upon coefficients. For example, it can be seen that in Figure 5.2a that scales 3 and 4 show more variation than that of any other scale. Whilst scales 5 and 6 show the most in Figure 5.2b. Note that the periodic behaviour can be seen within the prominent scales as the periodicity is smaller than the wavelet length, such that even shifting the series would not remove this effect.



(a) Periodicity of 12. (b) Periodicity of 60.

Figure 5.2: Averaged Wavelet spectra of sinusoid with noise where $\sigma = 0.1$. We do not calculate the coefficients that would require points beyond the length of the series. To the right of each scale is the average magnitude. Within the transformation a Haar Wavelet is used.

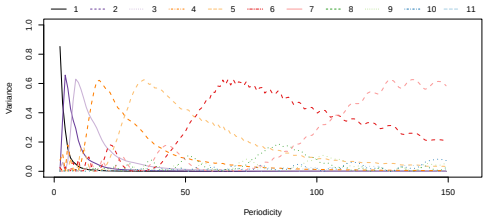
5.2.3 Variance Profiles

The level of periodic behaviour presented within each scale of a wavelet transform can be measured by monitoring the variance of the scale. We can calculate this for each scale by

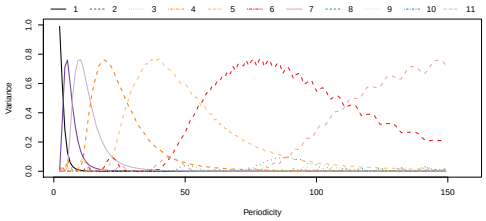
$$\hat{\sigma}^2_{d_j} = \frac{1}{2^J - 2^{j-1} - 1} \sum_{k=0}^{2^J - 2^{j-1} - 1} d^2_{j,k},$$

which can be compared across all scales. This is calculated and shown on the right hand side of Figure 5.2 for the average spectra. This calculation allows the quantification of the difference between the scales and periodicities, which we study further. By varying the periodicity we can study the changes in this variance through simulation to explore this relationship between periodicity and wavelet transform. To do this we simulate from a sinusoid with no noise, vary the periodicity over a range and use a number of wavelets chosen from the Daubechies Extremal Phase family. We do not go into detail regarding the differences between these wavelets as they are not our focus, but more detail can be found in Daubechies (1992). The results of this exploration are shown in Figure 5.3.

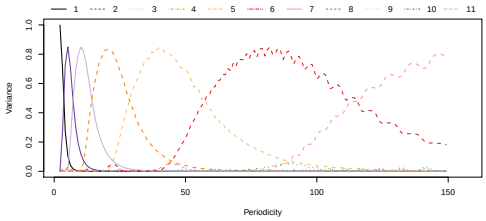
Notice that as before, regardless of the wavelet chosen, each scale becomes the most varying at a point before declining to zero as the periodicity increases. This behaviour is of particular interest as it can be used to distinguish between other periodicities. Going forward we analyse the effect of using the Haar Wavelet exclusively, as although the differences are less distinct in comparison to other



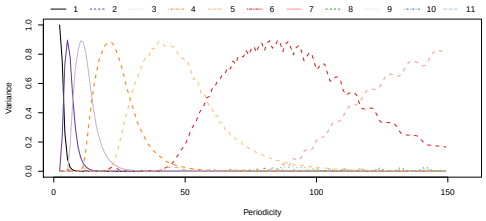
(a) Daubechies Extremal Phase wavelet 1 (Haar)



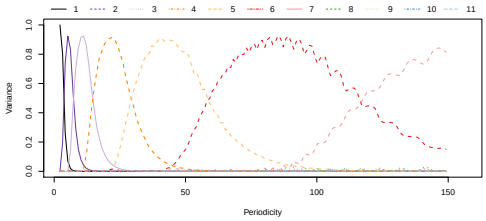
(b) Daubechies Extremal Phase wavelet 2



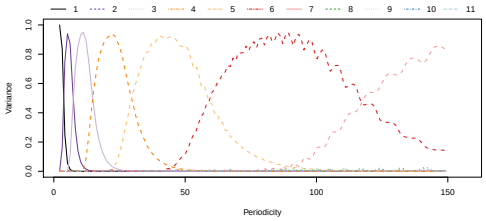
(c) Daubechies Extremal Phase wavelet 3



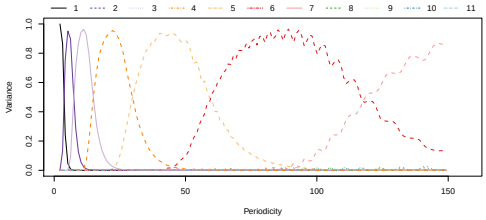
(d) Daubechies Extremal Phase wavelet 4



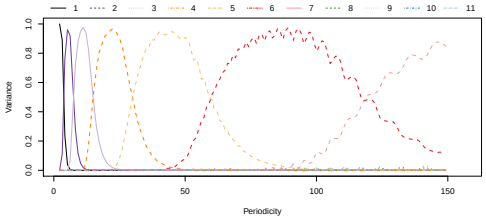
(e) Daubechies Extremal Phase wavelet 5



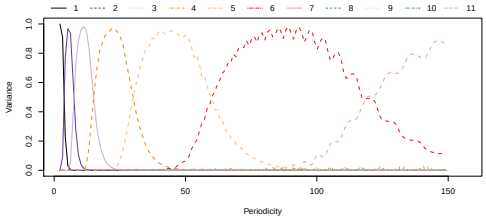
(f) Daubechies Extremal Phase wavelet 6



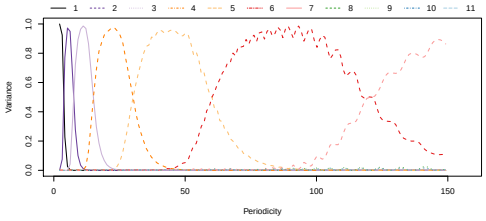
(g) Daubechies Extremal Phase wavelet 7



(h) Daubechies Extremal Phase wavelet 8



(i) Daubechies Extremal Phase wavelet 9



(j) Daubechies Extremal Phase wavelet 10

Figure 5.3: Proportional Variance of each scale of a wavelet transform for a sinusoid as periodicity varies and the wavelet chosen varies. Note that we have standardised the variances such that they are scaled by their sum for each periodicity to show where the majority of variance lies. Simulated series are of length $n = 2048$.

wavelets, there is still clearly distinguishable behaviour and the formulation shown in Equation (5.2.1) is favourable when determining the theoretical properties.

Focusing now on Figure 5.3a each of these curves can be quantified using the following theorem:

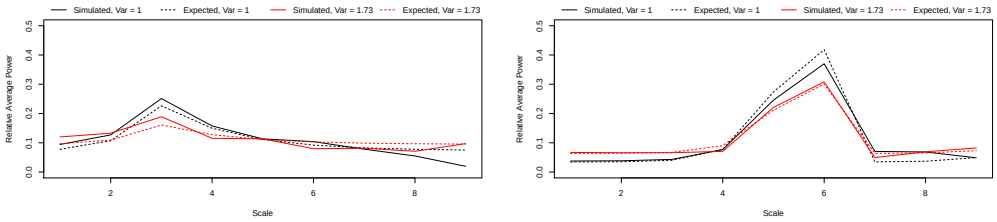
Theorem 5.1. *Given a time series follows the model described in Equation* (5.2)*, the variance of the discrete Non-Decimated Haar Wavelet Transform coefficients* $d_{j,k}$ *(as calculated in Equation* (5.1)*) have variance* $\sigma^2_{d_j}$ *where*

$$\mathbb{E}\left[\frac{1}{2^J - 2^j + 1}\sum_{k=1}^{2^J-2^j+1} d^2_{j,k}\right] = \sigma^2_{d_j} = \frac{\sin^4\left(\frac{2^{j-1}\pi}{p}\right)}{2^{j-1}\sin^2\left(\frac{\pi}{p}\right)} + \sigma^2_\epsilon + o(1)$$

such that $o(1) \to 0$ *as* $n \to \infty$*, where* $n = 2^J$*.*

Proof: See Appendix A, where the term $o(1)$ is also detailed.

It is this particularly interesting relationship which forms the basis for our work. Using the evolution of the wavelet variances across scales we can calculate the variance of each scale to create a 'variance profile'. Examples of such are given in Figure 5.4 for our sinusoidal model in Equation (5.2). Using the formula from Theorem 1 we can deduce the predicted variance profile and we plot these alongside those simulated. Note that despite an increase in noise, the structure of the variance profile is still prevalant, particularly at coarser scales.

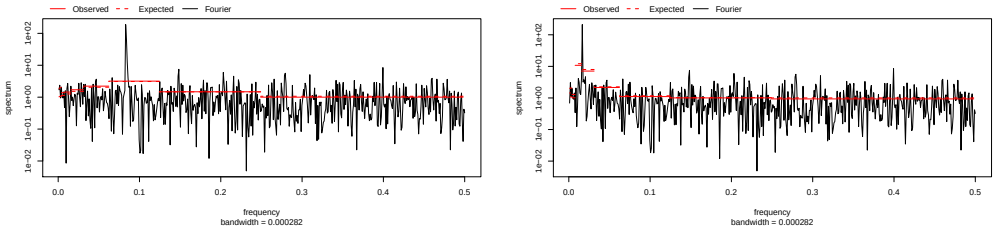


(a) $p = 12$. (b) $p = 60$.

Figure 5.4: Variance profiles over different periodicities and noise levels. Note each has been normalised by their sum.

As should be expected, these profiles align nicely with the usual Fourier spectrum. Using the MATLAB function `centfreq` (Misiti et al., 2018) we can calculate

the frequency each scale corresponds to for the Haar wavelet. In Figure 5.5 we overlay the variance profiles in this form above a Fourier spectrum. There it can be clearly seen that the peaks in the profile also overlay the peaks in the Fourier spectrum, however they are not as pronounced.



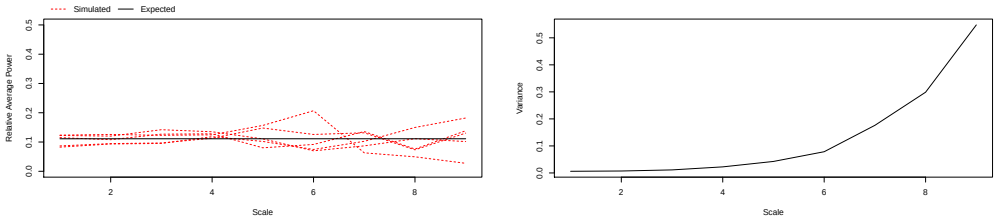
(a) $p = 12$. (b) $p = 60$.

Figure 5.5: Fourier spectra overlaid with a corresponding variance profile. Noise level of $\sigma^2 = 1$. Scale indices for the wavelet transform are such that 1 is on the right, J is on the left. Note both the red and black lines are on equal y scales.

5.2.4 Test statistic

These variance profiles allow us to make an estimation on the periodicity present within a series, but we must also consider the possibility that there is no periodicity. Note that in Figure 5.3 we did not show the behaviour of no periodicity, where in our model form (Equation (5.2)) $p = 1$, such that we are only dealing with a gaussian white noise series. Following Theorem 1 this means that the wavelet variance across scales would be a constant σ_ϵ^2. However, this is in the case where $n \to \infty$ and as we can see in Figure 5.6 that although the expected values are constant, the variance of the simulated profiles grows significantly towards the higher scales. As such we must characterise this behaviour to determine a test for the presence of periodicity in our data.

Theorem 5.2. *Given a gaussian white noise series $y_t = \epsilon_t$ for $t = 1, \ldots, n = 2^J$ the sum of the square of Non-Decimated Haar Wavelet Coefficients calculated by*



(a) Variance profiles of a white noise series (where p = 1), by scalem, with $\sigma^2 = 1$.

(b) Variance of white noise variance profiles by scale. Calculated empirically.

Figure 5.6: Exploration of variance profiles under gaussian white noise.

Equation (5.2.1) *is distributed as:*

$$\sum_{k=1}^{2^J-2^j+1} d_{j,k}^2 \sim \begin{cases} Gamma\left(\frac{1}{2}, 2\sigma_\epsilon^2\right) & if\ J = j \\ Gamma\left(\frac{(2^J-2^j+1)^2}{r_j}, \frac{r_j\sigma_\epsilon^2}{2^J-2^j+1}\right) & if\ J > j \end{cases}$$

where $r_j = 3^{-1} \cdot 2^{-j-1}\left(20 - 11(2^{j+1}) + 5(2^{2+J}) + 2^{2+2j+J} + 4^{j-1} - 5(8^j)\right)$.

Proof: See Appendix B.

With this information, we can then construct a multiple hypothesis test using the Bonferroni approach (Bonferroni, 1936). This is such that we create a composite hypothesis test for which we do not reject the null of no periodicity if the following is satisfied:

$$\frac{\alpha_j}{2} < \mathbb{P}\left(\sum_{k=1}^{2^J-2^j+1} d_{j,k}^2 < \sum_{k=1}^{2^J-2^j+1} \hat{d}_{j,k}^2\right) < 1 - \frac{\alpha_j}{2}, \qquad \forall j.$$

Determination of the α_j is open so long as they satisfy $\sum_{j=1}^{J-1} \alpha_j = \alpha$ where α is the pre-determined power level. Note that we only test up to scale $J-1$, which has $2^{J-1}+1$ coefficients, as the scale J has only 1 coefficient.

5.2.5 Algorithm

Given that we are now have the methodology to test for periodicity and generate variance profiles for series and their expected values, we now look to match them

to the correct profile. If we have an observed profile $\hat{\sigma}^2_{d_j}$ we have to measure its distance from a true profile at periodicity p, σ^2_{p,d_j}. As such there are a number of distance metrics we could employ, however here we use an absolute difference which weights according to the amount of coefficients in the scale. This is such that our observed difference is calculated as

$$\hat{d}_p = \delta\left(\hat{\sigma}^2_{d_j}, \sigma^2_{p,d_j}\right) = \sum_{j=1}^{J-1} \frac{1}{2^J - 2^j + 1} \left|\hat{\sigma}^2_{d_j} - \sigma^2_{p,d_j}\right|,$$

thus we estimate the periodicity of a series as

$$\hat{p} = \arg \min_{p=2,\ldots,2^J} \hat{d}_p.$$

However, dependent on the value of J we may have a large and computationally inefficient set of p to explore. To resolve this problem we return to our original exploration into the evolution of power within the scales seen in Figure 5.3a. There it can be seen that for each periodicity there is at most two dominant scales for each value of p. We can use this information to reduce the search region significantly by determining the region where the absolute power of a scale crosses with another.

Lemma 5.3. *Given that a series is generated by the model in Equation* (5.2) *as* $n \to \infty$,

$$\sum_{k=1}^{2^J - 2^j + 1} d^2_{j,k} = \sum_{k=1}^{2^J - 2^{j+1} + 1} d^2_{j+1,k} \iff p = \frac{2^{j-1}\pi}{\cos^{-1}\left(2^{-\frac{3}{4}}\right)}.$$

Proof: See Appendix C.

A representation of this is given in Figure 5.7 where we have calculated the cross over points and overlayed them onto Figure 5.3a. Note that due to the finite sample size, we begin to see an error in the approximation in the latter scales.

We now construct an algorithm to detect and predict the periodicity within a series based upon the methodology discovered. This is Algorithm 2 and we name

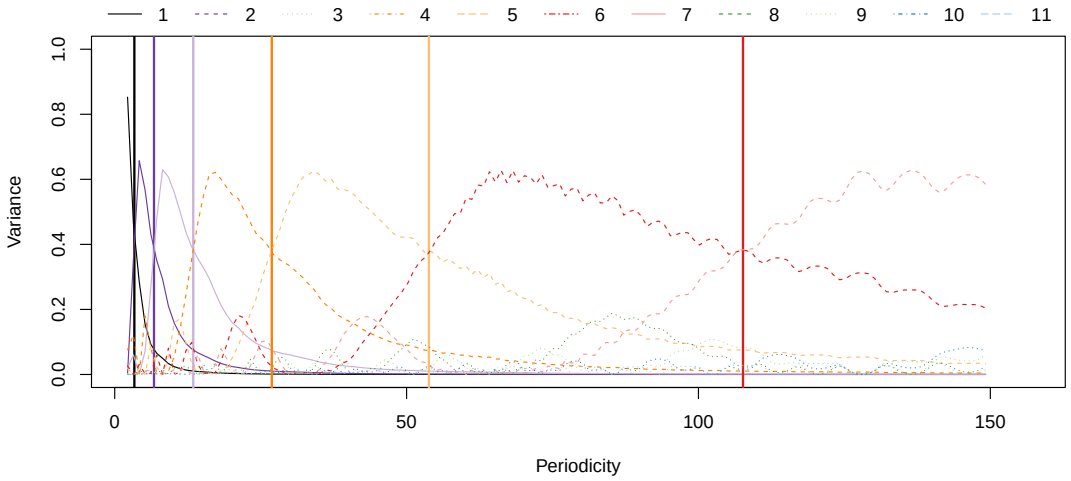


Figure 5.7: The calculated cross over points between the scales as per Lemma 1. Note we have emphasised the cross over lines rather than the Variance.

it coloquially the 'WavePer Algorithm'.

Data: y_t for $t = 1, 2, \ldots n = 2^J$: Series of interest;

Data: $\hat{\sigma}_{d_j}$ for $j = 1, \ldots, J$: Variance Profile;

input : $\hat{d}_p = \infty \quad \forall p = 2, \ldots 2^J$;

input : $\hat{p} = NA$;

Test the presence of periodicity ;

if *Periodicity Found* **then**

  Determine Search Region $(\hat{p}_{low}, \hat{p}_{high})$;

  $i = \hat{p}_{high}$;

  while $i \leq \hat{p}_{high}$ **do**

    Calculate profile distance $\hat{d}_i = \delta\left(\hat{\sigma}^2_{d_j}, \sigma^2_{i,d_j}\right)$;

  end

  Choose $\hat{p} = \arg\min_{p=2,\ldots,2^J} \hat{d}_p$.

end

Result: $\hat{p}$ estimated periodicity.

Algorithm 2: WavePer Algorithm.

5.3 Empirical Analysis

We now look to apply the methodology discussed in Section 5.2 in an empirical analysis to detail its effectiveness. Focus lies on two areas on interest: power of detection, and accuracy in prediction. As such we construct a simulation study which also compares to other methods in the field. These include many of those introduced in Section 5.1, where specicifally we compare against those presented in Table 5.1 in two areas. Firstly the detection of a periodic component in a series, and secondly given the component is present a prediction on p. Note in the case of FindFrequency and ARSER, their source code was modified to not identify any linear trends as we do not expect them to be present.

Method	Detect if $p = 1$ or $p > 1$	Estimate p
WavePer	✓	✓
HWWN	✓	
BartlettB	✓	
Ljung-Box	✓	
Fisher	✓	
FindFrequency	✓	✓
ARSER	✓	✓

Table 5.1: Methodology compared against within the simulation study.

Regarding implementation, all simulations where run within the R statistical software (R Core Team, 2018) where the WavePer algorithms have been developed. The HWWN and BartlettB tests are taken from the `hwwntest` package (Savchev and Nason, 2018). The Ljung-Box is from the core R `stats` package. Fisher's test comes from the `GeneCycle` package (Ahdesmaki et al., 2019). FindFrequency is given in the `forecast` package (Hyndman, 2017). Finally, the R implementation of ARSER can be found through Github (Yang, 2019).

5.3.1 Test Power

As previously mentioned in Section 5.2 we use a composite hypothesis test to determine if there is evidence of periodicity in a series. To do this we use the

Bonferroni approach which allows flexibility in the individual bounds of each test performed under the composite. Here we present two proposals on these bounds and show the results. The first is simply that at each bound we have $\alpha_j = \frac{\alpha}{J-1}$ so that each bound is exactly the same. The second we propose uses the variance calculated under Theorem 2. This is such that

$$\alpha_j = \frac{\alpha r_j^{-2}}{\sum_{l=1}^{J-1} (r_l^2)^{-1}}.$$

Given that each scale has a decreasing amount of coefficients and therefore information than the previous scale, the high increase in variability causes the α_j to become very small for the higher scales. This is such that we are less likely to reject on these scales, but more on those which have more information and coefficients. We will refer to the second set of α_j as 'Decay' for clarity in the results.

We test at the standard $\alpha = 0.05$ level, performing 1000 iterations under a number of different lengths and noise levels. We vary the length from $n = 128, 256, 512, 1024$ and vary the noise levels as $\sigma_\epsilon^2 = 1, 3, 5$. We compare against the methods introduced in Section 5.1 and outlined in Table 5.1. We present the results of no periodicity (such that $p = 1$ in Equation (5.2)) in Table 5.2. The table outlines the power of these tests, where 'WavePer (Decay)' uses the decaying α_j and 'WavePer' is using the unmodified version.

All the formal tests (thus not including FindFrequency or ARSER) performed are below or near to the expected 5%, with our algorithm under decaying weights performing the best. This is then followed by the same procedure with non-decaying weights. Both the HWWN and BatlettB are consistently below or equal to 5% whereas the Ljung-Box and Fisher tests show a greater reliance on sample size before converging to 0.05, but it is a neglible difference. Of note is that it would appear that smaller sample sizes perform better than larger for increasing variability in the non-decaying WavePer algorithm, but this effect cannot be seen in the decaying case. In the case of the prediction methods FindFrequency and ARSER, FindFrequency has a much greater power in the lower variance cases, but

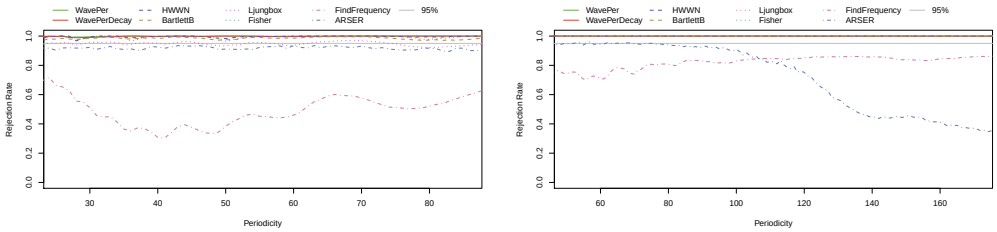
is troubled when there is a small sample with large variance. ARSER however attempts to fit a harmonic, and thus estimate a periodicity, from almost any peak it can find in the periodogram, making it especially weak in the case of Gaussian white noise.

5.3.2 Detection Rates

Further exploring our algorithm's effectiveness, we now look to its ability to detect periodicities where they are present. As we have previously mentioned in Section 5.1 we are not aiming for detection of high frequency components as this is not our aim, instead looking at components at lower frequencies which are often overlooked. As such we simulate for $p = \frac{n}{5}, \frac{n}{5} + 1 \ldots, \frac{2n}{3}$ so that we are studying a periodicity that goes from appearing 5 times in a series to only once and a half. The results for $n = 128, 256$ can be seen in Figure 5.8.

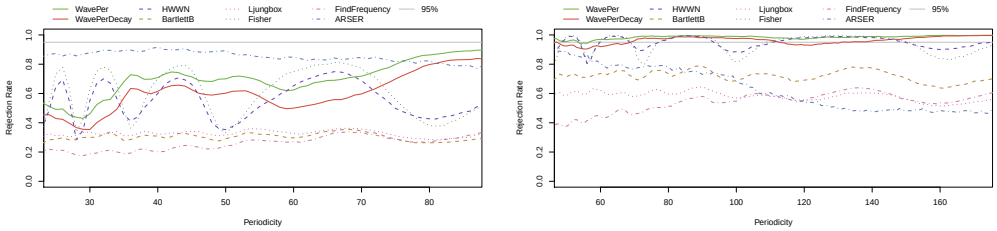
Though there is varying sample size n and signal-to-noise ratio through σ, a similar story can be seen across these plots. Firstly we can see that FindFrequency is consistently below the 95% level across all scenarios given. In both cases where $\sigma = 1$ we can see that most tests detect periodicity 95% of the time, with ARSER often being the lowest. However as we increase the variance of the noise, ARSER begins to become the strongest to detect any periodicity, a result that may be expected given the level of false positives in Table 5.2. It can be seen that most often the FindFrequency, Ljung-Box and BartlettB often share the lowest rates. Most interestingly is that the HWWN and Fisher tests show a increasingly periodic pattern in their detection rates. This leads to them alternating as the next strongest result (after ARSER). Where they are at a trough WavePer shows increasing strength as periodicity increases to a lower frequency component, with no obvious signs of periodic behaviour affecting their detection rates. Indeed as the noise increases WavePer shows greater strength and resilience against its competitors.

Note that we do not include the results for $n = 512, 1024$ here as most results



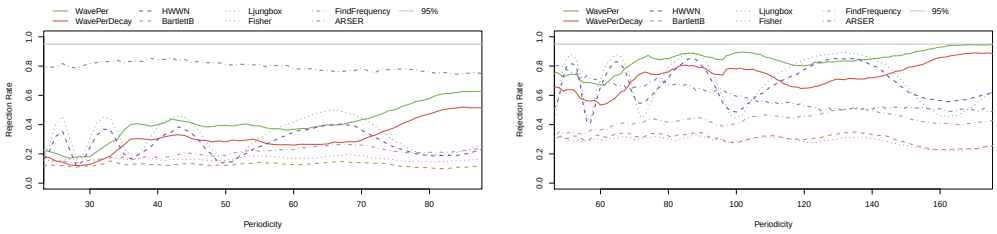
(a) $n = 128, \quad \sigma^2 = 1.$

(b) $n = 256, \quad \sigma^2 = 1.$



(c) $n = 128, \quad \sigma^2 = 3.$

(d) $n = 256, \quad \sigma^2 = 3.$



(e) $n = 128, \quad \sigma^2 = 5.$

(f) $n = 256, \quad \sigma^2 = 5.$

Figure 5.8: Probability of detecting periodicity over a number of sizes and noise levels.

	$\sigma^2 = 1$				$\sigma^2 = 3$				$\sigma^2 = 5$			
Method	128	256	512	1024	128	256	512	1024	128	256	512	1024
WavePer	1.9	1.8	1.4	1.8	1.6	1.7	2.1	2.6	1.2	1	0.8	2.5
WavePer (Decay)	0.7	0.8	0.5	0.7	1	0.4	0.7	0.7	0.8	0.2	0.6	0.6
HWWN	3.8	3.7	3.8	**4.1**	2.9	3.9	3.8	2.9	**4**	3.3	3.2	**4.8**
BartlettB	**4.3**	**4.1**	3.7	3.9	2.6	**4.3**	**4.3**	**4.1**	2.8	3.1	3.6	**5**
Ljung-Box	**5.5**	**5.4**	**4.1**	**5**	**4.2**	**5.2**	**4.7**	**5.1**	**4**	**5.1**	**4.6**	**4.9**
Fisher	**4.7**	**5.1**	**4.8**	**5.1**	**5.3**	**4.6**	**4.3**	**4**	**4.7**	**5.1**	3.7	**5.1**
FindFrequency	0	0	0	0	0.9	0.6	0.1	0	6.9	3.8	1.1	0.4
ARSER	71.3	56.9	42.4	33.9	69.9	50.4	44.3	36.4	68.7	54.2	43.4	34.8

Table 5.2: Percentage of simulations which rejected the null hypothesis that $p = 1$. Note we have emboldened those within 1% of the expected 5% value, and changed the colour of those less than 4% to green, similarly those above 6% are red.

converged to 1 rapidly and exhibit similar behaviour to that of $n = 128,256$ as shown in Figure 5.8, but they can be found in Appendix 5.6.4. Two points are noteworthy from those plots, the decline of ARSER for detection as the sample size increases, and that as noise increases, Ljung-Box is the first test to drop below the expected level.

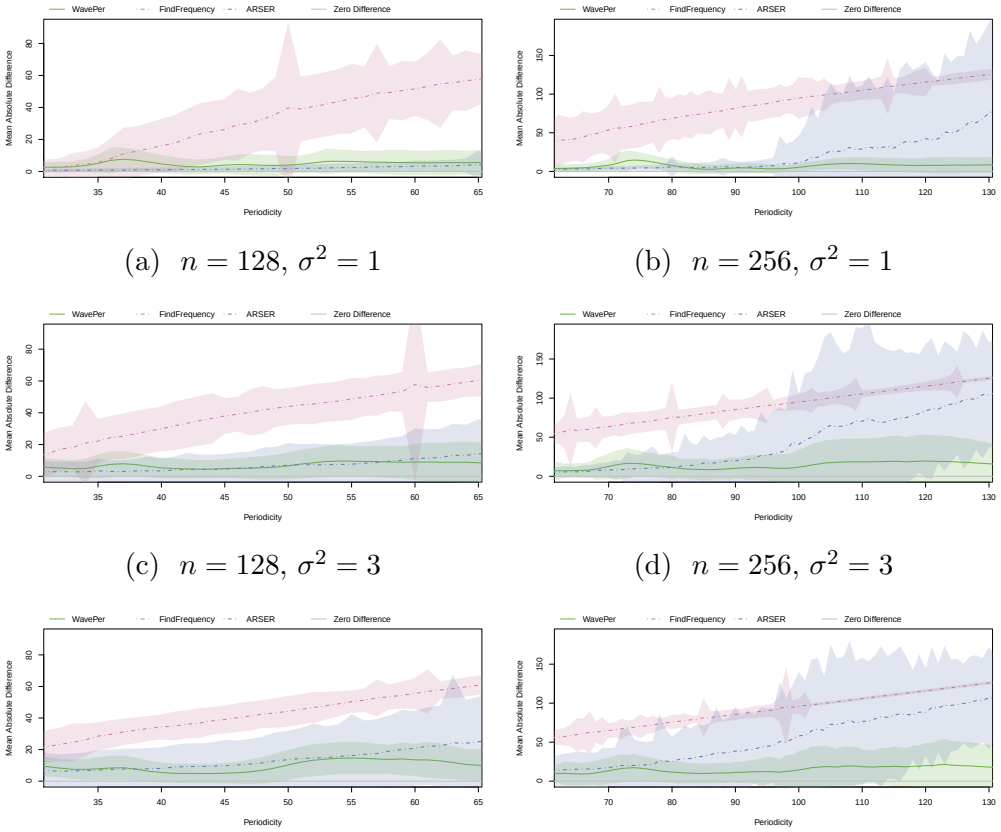
5.3.3 Estimation of p

We now turn our attention to the final part of our algorithm, which predicts the periodicity present. Looking upon the same range of p as in Section 5.3.2, the results are shown in terms of their Mean Absolute Difference, calculated by:

$$MAD_j = \frac{1}{|\tilde{n}_j|} \sum_{i \in \tilde{n}_j} |\hat{p}_i - p|$$

where j is the index for the method used (as given in Table 5.1), n_j is the set of indexes where the method found evidence of periodicity, $\hat{p}_i$ is the periodicity estimated in simulation $i = 1, 2, \ldots, 1000$ and p is the true periodicity. Thus we only consider those results which indicated periodicity initially. Additionally we study the standard deviation of our results $\hat{\sigma}_{\text{MAD }_j}$ by banding them with $\pm\hat{\sigma}_{\text{MAD }_j}$.

The results of this part of the simulation study are presented in Figure 5.9. Looking closely at the plots there are a number of features to note. Firstly in the lower n case, the level of noise causes the FindFrequency to diverge more initially from the results of WavePer and ARSER, before continuing to diverge at a greater magnitude. Noise similarly effects ARSER such that it does not follow WavePer's results as closely as the signal becomes more obscure. This occurs to the point that WavePer has a clear advantage in high noise cases at lower frequencies. Increasing the sample size causes these issues to become greater as both FindFrequency and ARSER attempt to fit a higher frequency component than WavePer. Noise continues to affect both and cause them to diverae much more substantially as the periodicity increases whereas WavePer shows no clear



(a) $n = 128$, $\sigma^2 = 1$ (b) $n = 256$, $\sigma^2 = 1$

(c) $n = 128$, $\sigma^2 = 3$ (d) $n = 256$, $\sigma^2 = 3$

(e) $n = 128$, $\sigma^2 = 5$ (f) $n = 256$, $\sigma^2 = 5$

Figure 5.9: Mean Absolute Difference from Truth. Shaded areas are the Mean Absolute Difference plus or minus the standard deviation of the estimate.

sign of this.

Studying the bands around the estimate, we can see that the highest widths occur from the ARSER procedure, with the narrowest occurring from FindFrequency as it continues to diverge from the expected value. However in comparison to WavePer both methods show inconsistent widths as periodicity increases, whereas WavePer stays approximately consistent. Note that as the sample size increases that the bands for WavePer increase unlike other methods, but this is due to the other methods consistently approximating very distant results.

As before we show the results for $n = 512, 1024$ in Appendix 5.6.5, where the divergence between FindFrequency and WavePer continues.

5.4 Application: Live Births in Metropolitan France

It can be anticipated that for daily birth we would expect to find a yearly periodicity and as such we provide this as an application for our methodology. Here we study a dataset produced by National Institute of Statistics and Economic Studies (INSEE), the national statistics bureau of France. In particular we focus our attention on a daily series of Live Births for Metropolitan France (excluding Mayotte) which spans from 1st January 1968 to 10th August 1973 (INSEE, 2017). A large and insightful analysis of this data can be found in Regnier-Loilier and Divinagracia (2010). The series which we study is shown in Figure 5.10.

We focus on different parts of the series in our application here, in the absence of a confidence interval we take an increasing amount of data to show how our procedure encloses upon the yearly periodicity, given an increasingly limited amount of full cycles completed. Currently our algorithm can only support series which are of dyadic length, such that they are of length 2^J for some J. However there are many processes which could be appended to our algorithm to account for a non-dyadic sample size, such as maximal overlap and periodic bounding, for more

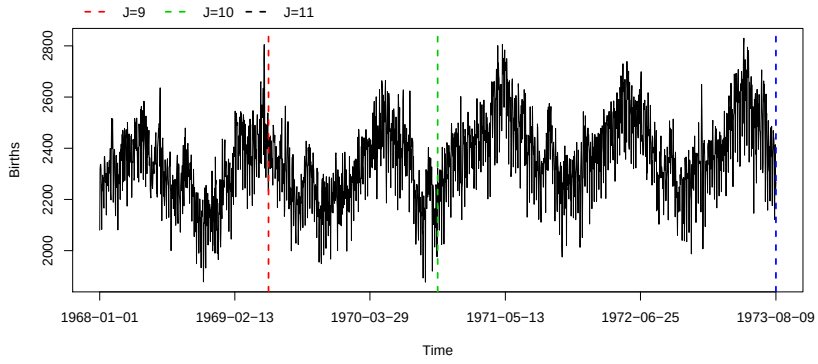


Figure 5.10: Live Births in France, from 1st January 1968 to 9th August 1973.

details of which we refer the reader to Nason (2010).

Continuing, we apply our methodology to the series beginning at 1st January 1968 and taking 2^J points for $J = 9, 10, 11$ and does not include a potential non-stationarity that appears shortly after. The cut off point for each series is shown in Figure 5.10. In the each segment we are analysing a year and 5 months, 2 years and 10 monts in the second, 5 years and 7 months in the third and 11 years and 3 months respectively. To best replicate the model we assume in Equation (5.2), we standardise each segment by subtracting the sample mean and dividing by the sample standard deviation, as we assume a amplitude of 1 on the sinusoid. Further, from Figure 5.10 it is possible that there may be an increasing trend present in the data, as such we attempt to find and remove a linear trend through simple linear modelling, as trend estimation is not our focus.

Results of our algorithm can be found in Table 5.3 for each segment. We also show if a linear trend was removed for clarity. Note that for each of the samples both forms of our test reject the null hypothesis of white noise as expected. Further the estimated periodicities indicate a yearly pattern in the data. For visual results from $n = 512, 1024, 2048$ we refer the reader to Figure 5.11 where we review the distance metric's over the search regions and the variance profiles observed and those expected. For $n = 512$ it can be seen in Figure 5.11a that although 233 was chosen as the minimum, there is a noticeable dip in the function around 360, but there was not enough information at that point to weight it appropriately. This

is due to lack of information in the eighth scale which is calculated and presented in the next larger subset in Figure 5.11d. This allows us to get much closer to the expected periodicity. Further information is then gathered in our final subset which moves our estimate just beyond our expected value, seemingly due to a large amount of noise in the 10th scale which reduces the peak structure. This coincides with the statistical evidence of a trend appearing in the data which may obfuscate the low frequency components we are searching for.

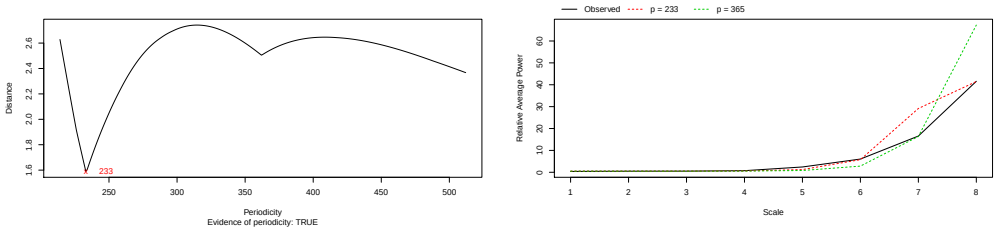
Segment Length	Test result	Test result (Decay)	Estimated Periodicity	Linear Trend
512	True	True	233	No
1024	True	True	359	No
2048	True	True	377	Yes

Table 5.3: Estimation of periodicity in Daily Births

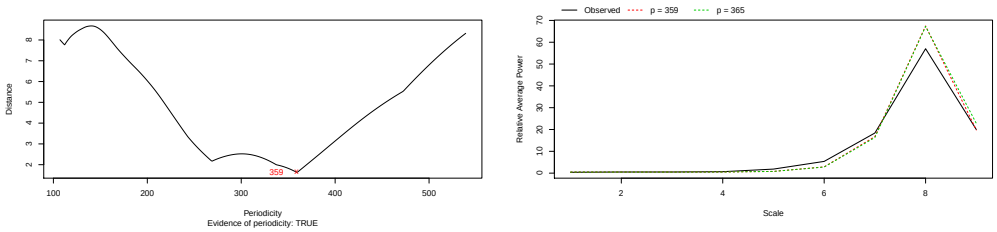
5.5 Conclusion

Within this paper we have shown the validity of a wavelet based periodicity detection algorithm. Using the unique representation of each periodicity in a 'Variance Profile' we note a number of observations. Firstly we can determine if there is periodicity present within the series, with the required power. Secondly we are able to calculate that if such a periodicity exists, the likely value it holds. We have shown this through an extensive empirical study and presented an example of Daily Birth data from France, showing the efficacy of our work.

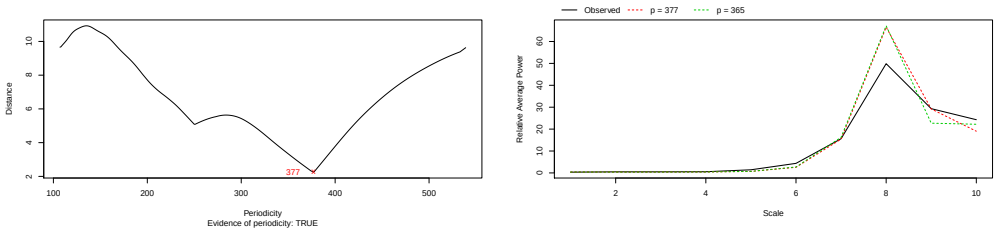
We find that our work has particular strength in the detection of lower frequency components with only a small amount of full cycles within a series, at a more competitive rate than those of other methods. The unique representation of each periodicity allows lower frequency components to be more prominent than those of more traditional methods such as the Fourier spectrum. It has been shown that in contradiction to other methods our work does not prefer to determine a higher frequency component where it is not present, but infact detects an often over looked lower frequency component instead. With strong detection rates and



(a) $n = 512$ (b) $n = 512$



(c) $n = 1024$ (d) $n = 1024$



(e) $n = 2048$ (f) $n = 2048$

Figure 5.11: Distance metric over the search region (Left) and variance profile comparisons (Right) for each sample.

a non-diverging distance from the true value of the periodicity present, our work is a promising step to automatic modelling of such components.

However, we must draw attention to further work that is required to further validate this method for extensive use. We have restricted our potential model to only that of a sinusoid with noise as a proof of concept, however there are few series which conform so nicely. Furthermore, more work needs to be conducted to explore the application of this work to higher frequency components. As the study of the unique profiles show in Figure 5.3, as we progress to lower values of p (high frequency components), there is much more movement in the variance profiles and thus we expected it would be difficult to distinguish between them in large noise cases. However, the slower movement in the low frequency components is where we draw our strength.

This leaves many avenues to explore in the future. As in Benedetto and Pfander (2002) where they use a generalisation of the Haar Wavelet to suitable fit their problem, we could explore the effect of alternative wavelets on this process. Indeed, given the plots given in Figure 5.3, of interest would be similar wavelets performance and theory in the 'Daubechies Extremal Phase Family', of which the Haar wavelet belongs, and more wavelets beyond this. Furthermore the resilience of this methodology under different model structures would require further study to determine the convergent behaviour and improve the algorithm. What we have found however is a method of detection for a component which would often be overlooked, which can not only have a significant impact upon prediction, but could be used unsupervised on a larger scale across the bigger data we encounter today.

5.6 Appendix

5.6.1 Proof of Theorem 1

Before directly determining the functional form of the Haar Non-Decimated Wavelet Variance for scales j we first calculate the first two moments and covariance structure of the assumed model for calculations further ahead.

Consider the model described in Equation (5.2) which is periodic in $p \in \mathbb{R}$ can be described also by

$$Y_T = \sin \frac{2\pi T}{p} + \epsilon_T, \qquad \epsilon_T \sim N(0, \sigma_\epsilon^2)$$

where $T \in \{1, 2, \ldots, n\}$. We can model this without knowledge of the current time T by:

$$Y_t | T = t \sim N\left(\sin \frac{2\pi t}{p}, \sigma_\epsilon^2\right), \qquad T \sim_{Discrete} U(1, n).$$

We can then calculate expectation of the series:

$$\mathbb{E}(Y_t) = \sum_{t=1}^{n} \mathbb{E}(Y_T | T = t)\mathbb{P}(T = t) = \frac{1}{n}\sum_{t=1}^{n} \sin \frac{2\pi t}{p} = \frac{1}{n}\frac{1}{\sin \frac{\pi}{p}} \sin \frac{n\pi}{p} \sin \frac{(1+n)\pi}{p}.$$

Now consider

$$\begin{aligned}
\mathbb{E}(Y_t^2) &= \mathbb{E}(\mathbb{E}(Y_t^2 | T = t)) \\
&= \sum_{t=1}^{n} \mathbb{E}(Y_t^2 | T = t)\mathbb{P}(T = t) \\
&= \frac{1}{n}\sum_{t=1}^{n} \mathbb{E}(\sin^2 \frac{2\pi t}{p} + 2\epsilon_t \sin \frac{2\pi t}{p} + \epsilon_t^2) \\
&= \frac{1}{n}\sum_{t=1}^{n} (\sin^2 \frac{2\pi t}{p} + 0 + \sigma_\epsilon^2) \\
&= \sigma_\epsilon^2 + \frac{1}{n}\sum_{t=1}^{n} \sin^2 \frac{2\pi t}{p}\sigma_\epsilon^2 + \frac{1 + 2n - \frac{\sin \frac{2(\pi + 2n\pi)}{p}}{\sin \frac{2\pi}{p}}}{4n}.
\end{aligned}$$

Now we consider the covariance at lag s:

$$\mathrm{Cov}(Y_t, Y_{t+s}) = \mathbb{E}(Y_t Y_{t+s}) - \mathbb{E}(Y_t)\mathbb{E}(Y_{t+s})$$

Firstly consider:

$$\begin{aligned}
\mathbb{E}(Y_t Y_{t+s}) &= \mathbb{E}(\mathbb{E}(Y_t Y_{t+s}|T=t)) = \sum_{t=1}^{n} \mathbb{E}(Y_t Y_{t+s}|T=t)\mathbb{P}(T=t) \\
&= \frac{1}{n}\sum_{t=1}^{n-s} \mathbb{E}\left[(\sin\frac{2\pi t}{p} + \epsilon_t)(\sin\frac{2\pi(t+s)}{p} + \epsilon_{t+s})\right] \\
&= \frac{1}{n}\sum_{t=1}^{n-s} \mathbb{E}(\sin\frac{2\pi t}{p}\sin\frac{2\pi(t+s)}{p} + \epsilon_{t+s}\sin\frac{2\pi t}{p} + \epsilon_t \sin\frac{2\pi(t+s)}{p} + \epsilon_t\epsilon_{t+s}) \\
&= \frac{1}{n}\left[\sum_{t=1}^{n-s} \sin\frac{2\pi t}{p}\sin\frac{2\pi(t+s)}{p}\right] + \delta_{0,s}\sigma_\epsilon^2\frac{n-s}{n} \\
&= \frac{2(n-s)\cos\frac{2\pi s}{p} + \csc\frac{2\pi}{p}(-\sin\frac{2\pi(1+2n-s)}{p} + \sin\frac{2\pi(1+s)}{p})}{4n} + \delta_{0,s}\sigma_\epsilon^2\frac{n-s}{n} \\
&= \frac{1}{2}\cos\frac{2\pi s}{p} - \frac{s}{2n}\cos\frac{2\pi s}{p} + \frac{\sin\frac{2\pi(1+s)}{p} - \sin\frac{2\pi(1+2n-s)}{p}}{4n\sin\frac{2\pi}{p}} + \delta_{0,s}\sigma_\epsilon^2\frac{n-s}{n}.
\end{aligned}$$

Where $\delta_{i,j}$ is the Kronecker Delta. We can now solve:

$$\begin{aligned}
\mathrm{Cov}(Y_t, Y_{t+s}) &= \mathbb{E}(Y_t Y_{t+s}) - \mathbb{E}(Y_t)\mathbb{E}(Y_{t+s}) \\
&= \frac{1}{2}\cos\frac{2\pi s}{p} - \frac{s}{2n}\cos\frac{2\pi s}{p} + \frac{\sin\frac{2\pi(1+s)}{p} - \sin\frac{2\pi(1+2n-s)}{p}}{4n\sin\frac{2\pi}{p}} + \delta_{0,s}\sigma_\epsilon^2\frac{n-s}{n} \\
&\quad - \left(\frac{1}{n}\frac{1}{\sin\frac{\pi}{p}}\sin\frac{n\pi}{p}\sin\frac{(1+n)\pi}{p}\right)^2 \\
&= \sigma_\epsilon^2\delta_{0,s} + \frac{2n(n-s)\cos\left(\frac{2\pi s}{p}\right)}{4n^2} \\
&\quad + \frac{n\csc\left(\frac{2\pi}{p}\right)\left(\sin\left(\frac{2\pi(s+1)}{p}\right) - \sin\left(\frac{2\pi(2n-s+1)}{p}\right)\right)}{4n^2} \\
&\quad - \frac{4\csc^2\left(\frac{\pi}{p}\right)\sin^2\left(\frac{\pi n}{p}\right)\sin^2\left(\frac{\pi(n+1)}{p}\right)}{4n^2} \\
&= \frac{1}{2}\cos\left(\frac{2\pi s}{p}\right)\left(1 - \frac{s}{n}\right) + \sigma_\epsilon^2\delta_{0,s} \\
&\quad + \frac{1}{n}\left[\frac{\sin\left(\frac{2\pi(s+1)}{p}\right) - \sin\left(\frac{2\pi(2n-s+1)}{p}\right)}{\sin\left(\frac{2\pi}{p}\right)}\right] \\
&\quad - \frac{1}{n^2}\left[\frac{\sin\left(\frac{\pi n}{p}\right)\sin\left(\frac{\pi(n+1)}{p}\right)}{\sin\left(\frac{\pi}{p}\right)}\right]^2
\end{aligned}$$

Now we take the Haar Non-Decimated Wavelet Transformation of this series across scale j and position T shown in Equation (5.2.1). We now consider the first two moments, then the covariance which leads ultimately to the variance formula.

$$\begin{aligned}
\mathbb{E}(d_{j,t}) &= \mathbb{E}(\frac{1}{\sqrt{2^j}}\left[\left\{\sum_{i=0}^{2^{j-1}-1} Y_{t+i}\right\} - \left\{\sum_{i=2^{j-1}}^{2^j-1} Y_{t+i}\right\}\right]) \\
&= \frac{1}{\sqrt{2^j}}\left[\left\{\sum_{i=0}^{2^{j-1}-1} \mathbb{E}(Y_{t+i})\right\} - \left\{\sum_{i=2^{j-1}}^{2^j-1} \mathbb{E}(Y_{t+i})\right\}\right] = 0
\end{aligned}$$

Consider next the covariance, we need only to solve:

$$\begin{aligned}
\mathbb{E}(d_{j,t}d_{j,t+s}) &= \mathbb{E}(\mathbb{E}(d_{j,T}d_{j,T+s}|T=t)) = \sum_{t=1}^{n}\mathbb{E}(d_{j,T}d_{j,T+s}|T=t)\mathbb{P}(T=t) \\
&= \frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left(\frac{1}{\sqrt{2}^{j}}\left[\sum_{i=0}^{2^{j-1}-1}Y_{T+i}-\sum_{i=2^{j-1}}^{2^{j}-1}Y_{T+i}\right]\right. \\
&\qquad \left.\cdot\frac{1}{\sqrt{2}^{j}}\left[\sum_{i=0}^{2^{j-1}-1}Y_{T+s+i}-\sum_{i=2^{j-1}}^{2^{j}-1}Y_{T+s+i}\right]\middle| T=t\right) \\
&= \frac{1}{2^{j}n}\sum_{t=1}^{n}\mathbb{E}\left[\sum_{i=0}^{2^{j-1}-1}Y_{T+i}\sum_{i=0}^{2^{j-1}-1}Y_{T+s+i}-\sum_{i=2^{j-1}}^{2^{j}-1}Y_{T+i}\sum_{i=0}^{2^{j-1}-1}Y_{T+s+i}\right. \\
&\qquad \left.-\sum_{i=0}^{2^{j-1}-1}Y_{T+i}\sum_{i=2^{j-1}}^{2^{j}-1}Y_{T+s+i}+\sum_{i=2^{j-1}}^{2^{j}-1}Y_{T+i}\sum_{i=2^{j-1}}^{2^{j}-1}Y_{T+s+i}\middle| T=t\right] \\
&= \frac{1}{2^{j}n}\sum_{t=1}^{n}\mathbb{E}\left[A_{j,T}-B_{j,T}-C_{j,T}+D_{j,T}|\,T=t\right]
\end{aligned}$$

We break the sum into four parts and solve each individually. The first part is:

$$
\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}[A_{j,T}|T=t] &= \frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left(\sum_{i=0}^{2^{j-1}-1}Y_{T+i}\sum_{i=0}^{2^{j-1}-1}Y_{T+s+i}\middle|T=t\right)\\
&= \frac{1}{n}\sum_{t=1}^{n}\sum_{i=0}^{2^{j-1}-1}\sum_{l=0}^{2^{j-1}-1}\mathbb{E}\left(Y_{T+i}Y_{T+s+l}|\,T=t\right)\\
&= \sum_{i=0}^{2^{j-1}-1}\sum_{l=0}^{2^{j-1}-1}\mathbb{E}\left(Y_{t+i}Y_{t+s+l}\right)\\
&= \sum_{i=0}^{2^{j-1}-1}\sum_{l=0}^{2^{j-1}-1}\frac{1}{2}\cos\frac{2\pi(s+l-i)}{p}-\frac{s+l-i}{2n}\cos\frac{2\pi(s+l-i)}{p}\\
&\quad+\frac{\sin\frac{2\pi(1+s+l-i)}{p}-\sin\frac{2\pi(1+2n-(s+l-i))}{p}}{4n\sin\frac{2\pi}{p}}\\
&\quad+\delta_{0,s+l-i}\sigma_\epsilon^2\frac{n-(s+l-i)}{n}\\
&= \frac{-2(n-s)\left(\cos\left(\frac{\pi\left(2^j+2s\right)}{p}\right)+\cos\left(\frac{\pi\left(2^j-2s\right)}{p}\right)-2\cos\left(\frac{2\pi s}{p}\right)\right)}{16n\sin^2\frac{\pi}{p}}\\
&\quad+\frac{-\sin\left(\frac{\pi\left(2^j-4n+2s-2\right)}{p}\right)+\sin\left(\frac{\pi\left(2^j+4n-2s+2\right)}{p}\right)}{16n\sin^2\frac{\pi}{p}}\\
&\quad+\frac{-\sin\left(\frac{\pi\left(2^j+2s+2\right)}{p}\right)+\sin\left(\frac{\pi\left(2^j-2s-2\right)}{p}\right)}{16n\sin^2\frac{\pi}{p}\sin\frac{2\pi}{p}}\\
&\quad+\frac{-2\sin\left(\frac{2\pi(2n-s+1)}{p}\right)+2\sin\left(\frac{2\pi(s+1)}{p}\right)}{16n\sin^2\frac{\pi}{p}\sin\frac{2\pi}{p}}\\
&\quad+\frac{2\sin\left(\frac{2\pi s}{p}\right)\left(4\cot\left(\frac{\pi}{p}\right)\sin^2\left(\frac{\pi 2^{j-1}}{p}\right)-2^j\sin\left(\frac{\pi 2^j}{p}\right)\right)}{16n\sin^2\frac{\pi}{p}}\\
&\quad+\sigma_\epsilon^2(2^{j-1}-s)
\end{aligned}
$$

Rearranging into powers of n:

$$
\begin{aligned}
&= -\frac{\cos\left(\frac{\pi\left(2^j+2s\right)}{p}\right)+\cos\left(\frac{\pi\left(2^j-2s\right)}{p}\right)-2\cos\left(\frac{2\pi s}{p}\right)}{8\sin^2\frac{\pi}{p}}+\sigma_\epsilon^2(2^{j-1}-s)\\
&+\frac{1}{n}\left[\frac{s\left(\cos\left(\frac{\pi\left(2^j+2s\right)}{p}\right)+\cos\left(\frac{\pi\left(2^j-2s\right)}{p}\right)-2\cos\left(\frac{2\pi s}{p}\right)\right)}{8\sin^2\frac{\pi}{p}}\right.\\
&\quad-\frac{\sin\left(\frac{\pi\left(2^j-4n+2s-2\right)}{p}\right)+\sin\left(\frac{\pi\left(2^j+4n-2s+2\right)}{p}\right)-\sin\left(\frac{\pi\left(2^j+2s+2\right)}{p}\right)}{16\sin^2\frac{\pi}{p}\sin\frac{2\pi}{p}}\\
&\quad+\frac{\sin\left(\frac{\pi\left(2^j-2s-2\right)}{p}\right)-2\sin\left(\frac{2\pi(2n-s+1)}{p}\right)+2\sin\left(\frac{2\pi(s+1)}{p}\right)}{16\sin^2\frac{\pi}{p}\sin\frac{2\pi}{p}}\\
&\quad\left.+\frac{2\sin\left(\frac{2\pi s}{p}\right)\left(4\cot\left(\frac{\pi}{p}\right)\sin^2\left(\frac{\pi 2^{j-1}}{p}\right)-2^j\sin\left(\frac{\pi 2^j}{p}\right)\right)}{16\sin^2\frac{\pi}{p}}\right]
\end{aligned}
$$

Note that:

$$\lim_{n\to\infty}\frac{\sin f(n)}{n}\to 0 \iff \deg f(n)<2$$

by L'Hopitals Rule. The second part is therefore:

$$
\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}[B_{j,T}|T=t] &= \frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left(\sum_{i=2^{j-1}}^{2^j-1}Y_{T+i}\sum_{i=0}^{2^{j-1}-1}Y_{T+s+i}\middle|T=t\right)\\
&= \frac{1}{n}\sum_{t=1}^{n}\sum_{i=2^{j-1}}^{2^j-1}\sum_{l=0}^{2^{j-1}-1}\mathbb{E}\left(Y_{T+i}Y_{T+s+l}|\,T=t\right)\\
&= \sum_{i=2^{j-1}}^{2^j-1}\sum_{l=0}^{2^{j-1}-1}\mathbb{E}\left(Y_{t+i}Y_{t+s+l}\right)\\
&= \sum_{i=2^{j-1}}^{2^j-1}\sum_{l=0}^{2^{j-1}-1}\frac{1}{2}\cos\frac{2\pi(s+l-i)}{p}-\frac{s+l-i}{2n}\cos\frac{2\pi(s+l-i)}{p}\\
&\quad+\frac{\sin\frac{2\pi(1+s+l-i)}{p}-\sin\frac{2\pi(1+2n-(s+l-i))}{p}}{4n\sin\frac{2\pi}{p}}\\
&\quad+\delta_{0,s+l-i}\sigma_\epsilon^2\frac{n-(s+l-i)}{n}_{s=s+l-i}\\
&= \frac{-2\left(2^j+n-s\right)\cos\left(\frac{2\pi\left(2^j-s\right)}{p}\right)}{16n\sin^2\frac{\pi}{p}}\\
&\quad+\frac{2\left(2^j+2n-2s\right)\cos\left(\frac{\pi\left(2^j-2s\right)}{p}\right)}{16n\sin^2\frac{\pi}{p}}\\
&\quad+\frac{\sin\left(\frac{2\pi\left(2^j+2n-s+1\right)}{p}\right)-2\sin\left(\frac{\pi\left(2^j+4n-2s+2\right)}{p}\right)}{16n\sin^2\frac{\pi}{p}\sin\frac{2\pi}{p}}\\
&\quad+\frac{\sin\left(\frac{2\pi\left(2^j-s-1\right)}{p}\right)-2\sin\left(\frac{\pi\left(2^j-2s-2\right)}{p}\right)}{16n\sin^2\frac{\pi}{p}\sin\frac{2\pi}{p}}\\
&\quad+\frac{\sin\left(\frac{2\pi(2n-s+1)}{p}\right)-\sin\left(\frac{2\pi(s+1)}{p}\right)}{16n\sin^2\frac{\pi}{p}\sin\frac{2\pi}{p}}\\
&\quad+\frac{-2\sin\left(\frac{\pi\left(2^j-2s-1\right)}{p}\right)-2\sin\left(\frac{\pi\left(2^j-2s+1\right)}{p}\right)}{16n\sin^3\frac{\pi}{p}}\\
&\quad+\frac{\sin\left(\frac{\pi\left(2^{j+1}-2s-1\right)}{p}\right)+\sin\left(\frac{\pi\left(2^{j+1}-2s+1\right)}{p}\right)}{16n\sin^3\frac{\pi}{p}}\\
&\quad+\frac{-2\cos\left(\frac{\pi}{p}\right)\sin\left(\frac{2\pi s}{p}\right)}{16n\sin^3\frac{\pi}{p}}+\frac{2(s-n)\cos\left(\frac{2\pi s}{p}\right)}{16n\sin^2\frac{\pi}{p}}+s\sigma_\epsilon^2
\end{aligned}
$$

Collecting terms in powers of n:

$$
\begin{aligned}
=&\frac{1}{16\sin^2\frac{\pi}{p}}\left(-2\cos\frac{2\pi(2^j-s)}{p}+4\cos\frac{\pi(2^j-2s)}{p}-2\cos\frac{2\pi s}{p}\right)+\sigma_\epsilon^2(2^{j-1}-s)\\
&-\frac{1}{n}\left(\frac{2\left(2^j-s\right)\cos\left(\frac{2\pi\left(2^j-s\right)}{p}\right)-2\left(2^j-2s\right)\cos\left(\frac{\pi\left(2^j-2s\right)}{p}\right)-2s\cos\left(\frac{2\pi s}{p}\right)}{16\sin^2\frac{\pi}{p}}\right.\\
&\quad-\frac{\sin\left(\frac{2\pi\left(2^j+2n-s+1\right)}{p}\right)-2\sin\left(\frac{\pi\left(2^j+4n-2s+2\right)}{p}\right)+\sin\left(\frac{2\pi\left(2^j-s-1\right)}{p}\right)}{16\sin^2\frac{\pi}{p}\sin\frac{2\pi}{p}}\\
&\quad+\frac{2\sin\left(\frac{\pi\left(2^j-2s-2\right)}{p}\right)+\sin\left(\frac{2\pi(2n-s+1)}{p}\right)-\sin\left(\frac{2\pi(s+1)}{p}\right)}{16\sin^2\frac{\pi}{p}\sin\frac{2\pi}{p}}\\
&\quad\left.-\frac{\sin\left(\frac{\pi\left(2^{j+1}-2s+1\right)}{p}\right)-2\cos\left(\frac{\pi}{p}\right)\sin\left(\frac{2\pi s}{p}\right)}{16\sin^3\frac{\pi}{p}}\right)+s\sigma_\epsilon^2
\end{aligned}
$$

Considering next the third part

$$
\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}[C_{j,T}|T=t] &= \frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left(\sum_{i=0}^{2^{j-1}-1}Y_{T+i}\sum_{i=2^{j-1}}^{2^j-1}Y_{T+s+i}\middle|T=t\right)\\
&= \frac{1}{n}\sum_{t=1}^{n}\sum_{i=0}^{2^{j-1}-1}\sum_{l=2^{j-1}}^{2^j-1}\mathbb{E}\left(Y_{T+i}Y_{T+s+l}|\,T=t\right)\\
&= \sum_{t=1}^{n}\sum_{i=0}^{2^{j-1}-1}\sum_{l=2^{j-1}}^{2^j-1}\mathbb{E}\left(Y_{t+i}Y_{t+s+l}\right)\\
&= \sum_{t=1}^{n}\sum_{i=0}^{2^{j-1}-1}\sum_{l=2^{j-1}}^{2^j-1}\frac{1}{2}\cos\frac{2\pi(s+l-i)}{p}\\
&\quad -\frac{s+l-i}{2n}\cos\frac{2\pi(s+l-i)}{p}\\
&\quad +\frac{\sin\frac{2\pi(1+s+l-i)}{p}-\sin\frac{2\pi(1+2n-(s+l-i))}{p}}{4n\sin\frac{2\pi}{p}}\\
&\quad +\delta_{0,s+l-i}\sigma_\epsilon^2\frac{n-(s+l-i)}{n}{}_{s=s+l-i}\\
&= \frac{2\left(2^j-n+s\right)\cos\left(\frac{2\pi\left(2^j+s\right)}{p}\right)}{16n\sin^2\frac{\pi}{p}}\\
&\quad -\frac{2\left(2^j-2n+2s\right)\cos\left(\frac{\pi\left(2^j+2s\right)}{p}\right)}{16n\sin^2\frac{\pi}{p}}-\frac{\sin\left(\frac{2\pi\left(2^j-2n+s-1\right)}{p}\right)}{16n\sin^2\frac{\pi}{p}\sin\frac{2\pi}{p}}\\
&\quad +\frac{2\left(\sin\left(\frac{\pi\left(2^j-4n+2s-2\right)}{p}\right)+\sin\left(\frac{\pi\left(2^j+2s+2\right)}{p}\right)\right)}{16n\sin^2\frac{\pi}{p}\sin\frac{2\pi}{p}}\\
&\quad -\frac{\sin\left(\frac{2\pi\left(2^j+s+1\right)}{p}\right)-\sin\left(\frac{2\pi(2n-s+1)}{p}\right)+\sin\left(\frac{2\pi(s+1)}{p}\right)}{16n\sin^2\frac{\pi}{p}\sin\frac{2\pi}{p}}\\
&\quad +\frac{2\sin\left(\frac{\pi\left(2^j+2s-1\right)}{p}\right)+2\sin\left(\frac{\pi\left(2^j+2s+1\right)}{p}\right)}{16n\sin^3\frac{\pi}{p}}\\
&\quad -\frac{\sin\left(\frac{\pi\left(2^{j+1}+2s-1\right)}{p}\right)+\sin\left(\frac{\pi\left(2^{j+1}+2s+1\right)}{p}\right)+\sin\left(\frac{\pi(2s-1)}{p}\right)}{16n\sin^3\frac{\pi}{p}}\\
&\quad -\frac{\sin\left(\frac{2\pi s+\pi}{p}\right)}{16n\sin^3\frac{\pi}{p}}-\frac{2(n-s)\cos\left(\frac{2\pi s}{p}\right)}{16n\sin^2\frac{\pi}{p}}
\end{aligned}
$$

Note that there is no overlap between the ranges of i and l such that σ_ϵ^2 exists.

Collecting these terms in powers of n:

$$
\begin{aligned}
=&\frac{1}{16\sin^2\frac{\pi}{p}}\left(2\cos\frac{2\pi(2^j+s)}{p}-4\cos\frac{\pi(2^j+2s)}{p}+2\cos\frac{2\pi s}{p}\right)\\
&+\frac{1}{n}\left(\frac{2\left(2^j+2s\right)\cos\left(\frac{\pi\left(2^j+2s\right)}{p}\right)-2\left(2^j+s\right)\cos\left(\frac{2\pi\left(2^j+s\right)}{p}\right)-2s\cos\left(\frac{2\pi s}{p}\right)}{16\sin^2\frac{\pi}{p}}\right.\\
&\quad+\frac{\sin\left(\frac{2\pi\left(2^j-2n+s-1\right)}{p}\right)-2\sin\left(\frac{\pi\left(2^j-4n+2s-2\right)}{p}\right)+\sin\left(\frac{2\pi\left(2^j+s+1\right)}{p}\right)}{16\sin^2\frac{\pi}{p}\sin\frac{2\pi}{p}}\\
&\quad-\frac{2\sin\left(\frac{\pi\left(2^j+2s+2\right)}{p}\right)-\sin\left(\frac{2\pi(2n-s+1)}{p}\right)+\sin\left(\frac{2\pi(s+1)}{p}\right)}{16\sin^2\frac{\pi}{p}\sin\frac{2\pi}{p}}\\
&\quad\left.+\frac{\sin\left(\frac{\pi\left(2^{j+1}+2s+1\right)}{p}\right)+\sin\left(\frac{\pi(2s-1)}{p}\right)+\sin\left(\frac{2\pi s+\pi}{p}\right)}{16\sin^3\frac{\pi}{p}}\right)
\end{aligned}
$$

Finally the fourth part:

$$
\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}[D_{j,T}|T=t] &= \frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left(\sum_{i=2^{j-1}}^{2^j-1}Y_{T+i}\sum_{i=2^{j-1}}^{2^j-1}Y_{T+s+i}\middle|\,T=t\right)\\
&= \frac{1}{n}\sum_{t=1}^{n}\sum_{i=2^{j-1}}^{2^j-1}\sum_{l=2^{j-1}}^{2^j-1}\mathbb{E}\left(Y_{T+i}Y_{T+s+l}|\,T=t\right)\\
&= \sum_{i=2^{j-1}}^{2^j-1}\sum_{l=2^{j-1}}^{2^j-1}\mathbb{E}\left(Y_{t+i}Y_{t+s+l}\right)\\
&= \sum_{i=2^{j-1}}^{2^j-1}\sum_{l=2^{j-1}}^{2^j-1}\frac{1}{2}\cos\frac{2\pi(s+l-i)}{p}-\frac{s+l-i}{2n}\cos\frac{2\pi(s+l-i)}{p}\\
&\quad+\frac{\sin\frac{2\pi(1+s+l-i)}{p}-\sin\frac{2\pi(1+2n-(s+l-i))}{p}}{4n\sin\frac{2\pi}{p}}\\
&\quad+\delta_{0,s+l-i}\sigma_\epsilon^2\frac{n-(s+l-i)}{n}_{s=s+l-i}\\
&= \frac{2(n-s)\left(\cos\left(\frac{\pi\left(2^j+2s\right)}{p}\right)-\cos\left(\frac{\pi\left(2^j-2s\right)}{p}\right)+2\cos\left(\frac{2\pi s}{p}\right)\right)}{16n\sin^2\frac{\pi}{p}}\\
&\quad+\frac{-\sin\left(\frac{\pi\left(2^j-4n+2s-2\right)}{p}\right)+\sin\left(\frac{\pi\left(2^j+4n-2s+2\right)}{p}\right)}{16n\sin^2\frac{\pi}{p}\sin\frac{2\pi}{p}}\\
&\quad-\frac{\sin\left(\frac{\pi\left(2^j+2s+2\right)}{p}\right)-\sin\left(\frac{\pi\left(2^j-2s-2\right)}{p}\right)}{16n\sin^2\frac{\pi}{p}\sin\frac{2\pi}{p}}\\
&\quad+\frac{-2\sin\left(\frac{2\pi(2n-s+1)}{p}\right)+2\sin\left(\frac{2\pi(s+1)}{p}\right)}{16n\sin^2\frac{\pi}{p}\sin\frac{2\pi}{p}}\\
&\quad+\frac{2\sin\left(\frac{2\pi s}{p}\right)\left(4\cot\left(\frac{\pi}{p}\right)\sin^2\left(\frac{\pi 2^{j-1}}{p}\right)-2^j\sin\left(\frac{\pi 2^j}{p}\right)\right)}{16n\sin^2\frac{\pi}{p}}\\
&\quad+\sigma_\epsilon^2(2^{j-1}-s)
\end{aligned}
$$

Rearranging into powers of n:

$$
\begin{aligned}
&= -\frac{\cos\left(\frac{\pi\left(2^j+2s\right)}{p}\right)+\cos\left(\frac{\pi\left(2^j-2s\right)}{p}\right)-2\cos\left(\frac{2\pi s}{p}\right)}{8\sin^2\frac{\pi}{p}}+\sigma_\epsilon^2(2^{j-1}-s)\\
&+\frac{1}{n}\left[\frac{s\left(\cos\left(\frac{\pi\left(2^j+2s\right)}{p}\right)+\cos\left(\frac{\pi\left(2^j-2s\right)}{p}\right)-2\cos\left(\frac{2\pi s}{p}\right)\right)}{8\sin^2\frac{\pi}{p}}\right.\\
&\quad-\frac{\sin\left(\frac{\pi\left(2^j-4n+2s-2\right)}{p}\right)+\sin\left(\frac{\pi\left(2^j+4n-2s+2\right)}{p}\right)-\sin\left(\frac{\pi\left(2^j+2s+2\right)}{p}\right)}{16\sin^2\frac{\pi}{p}\sin\frac{2\pi}{p}}\\
&\quad+\frac{\sin\left(\frac{\pi\left(2^j-2s-2\right)}{p}\right)-2\sin\left(\frac{2\pi(2n-s+1)}{p}\right)+2\sin\left(\frac{2\pi(s+1)}{p}\right)}{16\sin^2\frac{\pi}{p}\sin\frac{2\pi}{p}}\\
&\quad\left.+\frac{2\sin\left(\frac{2\pi s}{p}\right)\left(4\cot\left(\frac{\pi}{p}\right)\sin^2\left(\frac{\pi 2^{j-1}}{p}\right)-2^j\sin\left(\frac{\pi 2^j}{p}\right)\right)}{16\sin^2\frac{\pi}{p}}\right]
\end{aligned}
$$

Note that the fourth part is equivalent to the first part.

Combining all this together we get:

$$
\begin{aligned}
\mathbb{E}(d_{j,t}d_{j,t+s}) &= \frac{1}{2^j n}\sum_{t=1}^{n}\mathbb{E}[A_{j,T}|T=t]-\mathbb{E}[B_{j,T}|T=t] \\
&\quad -\mathbb{E}[C_{j,T}|T=t]+\mathbb{E}[D_{j,T}|T=t] \\
&= \frac{2\left(2^j+n-s\right)\cos\left(\frac{2\pi\left(2^j-s\right)}{p}\right)+2n\cos\left(\frac{2\pi\left(2^j+s\right)}{p}\right)}{2^{j+4}n\sin\frac{\pi}{p}} \\
&\quad + \frac{-8n\cos\left(\frac{\pi\left(2^j+2s\right)}{p}\right)-2\left(2^j+4n-4s\right)\cos\left(\frac{\pi\left(2^j-2s\right)}{p}\right)}{2^{j+4}n\sin\frac{\pi}{p}} \\
&\quad + \frac{-\csc\left(\frac{2\pi}{p}\right)\sin\left(\frac{2\pi\left(2^j+2n-s+1\right)}{p}\right)+\csc\left(\frac{2\pi}{p}\right)\sin\left(\frac{2\pi\left(2^j-2n+s-1\right)}{p}\right)}{2^{j+4}n\sin\frac{\pi}{p}} \\
&\quad - \frac{4\csc\left(\frac{2\pi}{p}\right)\sin\left(\frac{\pi\left(2^j-4n+2s-2\right)}{p}\right)-4\csc\left(\frac{2\pi}{p}\right)\sin\left(\frac{\pi\left(2^j+4n-2s+2\right)}{p}\right)}{2^{j+4}n\sin\frac{\pi}{p}} \\
&\quad - \frac{2^{j+2}\sin\left(\frac{\pi 2^j}{p}\right)\sin\left(\frac{2\pi s}{p}\right)+2^{j+1}\cos\left(\frac{2\pi\left(2^j+s\right)}{p}\right)}{2^{j+4}n\sin\frac{\pi}{p}} \\
&\quad + \frac{-2s\cos\left(\frac{2\pi\left(2^j+s\right)}{p}\right)+2^{j+1}\cos\left(\frac{\pi\left(2^j+2s\right)}{p}\right)+8s\cos\left(\frac{\pi\left(2^j+2s\right)}{p}\right)}{2^{j+4}n\sin\frac{\pi}{p}} \\
&\quad + \frac{16\cot\left(\frac{\pi}{p}\right)\sin^2\left(\frac{\pi 2^{j-1}}{p}\right)\sin\left(\frac{2\pi s}{p}\right)-\csc\left(\frac{2\pi}{p}\right)\sin\left(\frac{2\pi\left(2^j-s-1\right)}{p}\right)}{2^{j+4}n\sin\frac{\pi}{p}} \\
&\quad + \frac{\csc\left(\frac{2\pi}{p}\right)\sin\left(\frac{2\pi\left(2^j+s+1\right)}{p}\right)-2\csc\left(\frac{\pi}{p}\right)\sin\left(\frac{\pi\left(2^j+2s-1\right)}{p}\right)}{2^{j+4}n\sin\frac{\pi}{p}} \\
&\quad + \frac{-2\csc\left(\frac{\pi}{p}\right)\sin\left(\frac{\pi\left(2^j+2s+1\right)}{p}\right)-4\csc\left(\frac{2\pi}{p}\right)\sin\left(\frac{\pi\left(2^j+2s+2\right)}{p}\right)}{2^{j+4}n\sin\frac{\pi}{p}} \\
&\quad + \frac{\csc\left(\frac{\pi}{p}\right)\sin\left(\frac{\pi\left(2^{j+1}+2s-1\right)}{p}\right)+\csc\left(\frac{\pi}{p}\right)\sin\left(\frac{\pi\left(2^{j+1}+2s+1\right)}{p}\right)}{2^{j+4}n\sin\frac{\pi}{p}} \\
&\quad + \frac{4\csc\left(\frac{2\pi}{p}\right)\sin\left(\frac{\pi\left(2^j-2s-2\right)}{p}\right)+2\csc\left(\frac{\pi}{p}\right)\sin\left(\frac{\pi\left(2^j-2s-1\right)}{p}\right)}{2^{j+4}n\sin\frac{\pi}{p}} \\
&\quad + \frac{2\csc\left(\frac{\pi}{p}\right)\sin\left(\frac{\pi\left(2^j-2s+1\right)}{p}\right)-\csc\left(\frac{\pi}{p}\right)\sin\left(\frac{\pi\left(2^{j+1}-2s-1\right)}{p}\right)}{2^{j+4}n\sin\frac{\pi}{p}}
\end{aligned}
$$

+ . . . (continued overleaf)

$$
\begin{aligned}
\mathbb{E}(d_{j,t}d_{j,t+s}) = \cdots + \\
&+ \frac{-\csc\left(\frac{\pi}{p}\right)\sin\left(\frac{\pi\left(2^{j+1}-2s+1\right)}{p}\right) + 12n\cos\left(\frac{2\pi s}{p}\right)}{2^{j+4}n\sin^{\frac{\pi}{p}}} \\
&- \frac{6\csc\left(\frac{2\pi}{p}\right)\sin\left(\frac{2\pi(2n-s+1)}{p}\right) + 12s\cos\left(\frac{2\pi s}{p}\right)}{2^{j+4}n\sin^{\frac{\pi}{p}}} \\
&+ \frac{2\cot\left(\frac{\pi}{p}\right)\sin\left(\frac{2\pi s}{p}\right) + 6\csc\left(\frac{2\pi}{p}\right)\sin\left(\frac{2\pi(s+1)}{p}\right)}{2^{j+4}n\sin^{\frac{\pi}{p}}} \\
&+ \frac{\csc\left(\frac{\pi}{p}\right)\sin\left(\frac{\pi(2s-1)}{p}\right) + \csc\left(\frac{\pi}{p}\right)\sin\left(\frac{2\pi s+\pi}{p}\right)}{2^{j+4}n\sin^{\frac{\pi}{p}}} \\
&+ \sigma_\epsilon^2 - \frac{3s}{2^j}\sigma_\epsilon^2 \\
&= \mathrm{Cov}(d_{j,t}, d_{j,t+s})
\end{aligned}
$$

Collecting into terms of n:

$$\begin{aligned}
\mathrm{Cov}(d_{j,t}, d_{j,t+s}) = &\frac{2\cos\left(\frac{2\pi\left(2^j-s\right)}{p}\right) + 12\cos\left(\frac{2\pi s}{p}\right) + 2\cos\left(\frac{2\pi\left(s+2^j\right)}{p}\right)}{2^{j+4}\sin^2\frac{\pi}{p}} \\
&- \frac{8\cos\left(\frac{\pi\left(2^j-2s\right)}{p}\right) + 8\cos\left(\frac{\pi\left(2s+2^j\right)}{p}\right)}{2^{j+4}\sin^2\frac{\pi}{p}} + \sigma_\epsilon^2 - \frac{3s}{2^j}\sigma_\epsilon^2 \\
&+ \frac{1}{n}\left[\frac{1}{2^{j+4}\sin\frac{\pi}{p}}\left(16\cot\left(\frac{\pi}{p}\right)\sin\left(\frac{2\pi s}{p}\right)\sin^2\left(\frac{2^{j-1}\pi}{p}\right)\right.\right. \\
&\quad + 2^{j+1}\cos\left(\frac{2\pi\left(2^j-s\right)}{p}\right) - 2s\cos\left(\frac{2\pi\left(2^j-s\right)}{p}\right) \\
&\quad - 12s\cos\left(\frac{2\pi s}{p}\right) - 2^{j+1}\cos\left(\frac{2\pi\left(s+2^j\right)}{p}\right) \\
&\quad - 2s\cos\left(\frac{2\pi\left(s+2^j\right)}{p}\right) + 2^{j+1}\cos\left(\frac{\pi\left(2s+2^j\right)}{p}\right) \\
&\quad + 8s\cos\left(\frac{\pi\left(2s+2^j\right)}{p}\right) - 2^{j+1}\cos\left(\frac{\pi\left(2^j-2s\right)}{p}\right) \\
&\quad + 8s\cos\left(\frac{\pi\left(2^j-2s\right)}{p}\right) - \csc\left(\frac{2\pi}{p}\right)\sin\left(\frac{2\pi\left(-s+2^j-1\right)}{p}\right) \\
&\quad - 6\csc\left(\frac{2\pi}{p}\right)\sin\left(\frac{2\pi(2n-s+1)}{p}\right) \\
&\quad - \csc\left(\frac{2\pi}{p}\right)\sin\left(\frac{2\pi\left(2n+2^j-s+1\right)}{p}\right) \\
&\quad + 2\cot\left(\frac{\pi}{p}\right)\sin\left(\frac{2\pi s}{p}\right) - 2^{j+2}\sin\left(\frac{2^j\pi}{p}\right)\sin\left(\frac{2\pi s}{p}\right) \\
&\quad + 6\csc\left(\frac{2\pi}{p}\right)\sin\left(\frac{2\pi(s+1)}{p}\right) \\
&\quad + \csc\left(\frac{2\pi}{p}\right)\sin\left(\frac{2\pi\left(s+2^j+1\right)}{p}\right) \\
&\quad + \csc\left(\frac{2\pi}{p}\right)\sin\left(\frac{2\pi\left(-2n+2^j+s-1\right)}{p}\right) \\
&\quad + \csc\left(\frac{\pi}{p}\right)\sin\left(\frac{\pi(2s-1)}{p}\right) \\
&\quad - 2\csc\left(\frac{\pi}{p}\right)\sin\left(\frac{\pi\left(2s+2^j-1\right)}{p}\right) \\
&\quad - 2\csc\left(\frac{\pi}{p}\right)\sin\left(\frac{\pi\left(2s+2^j+1\right)}{p}\right) \\
&\quad - 4\csc\left(\frac{2\pi}{p}\right)\sin\left(\frac{\pi\left(2s+2^j+2\right)}{p}\right) \\
&\quad + \ldots \text{(continued overleaf)}
\end{aligned}$$

$$
\begin{aligned}
\operatorname{Cov}(d_{j,t}, d_{j,t+s}) = \ldots \\
&+ \csc\left(\frac{\pi}{p}\right) \sin\left(\frac{\pi\left(2s + 2^{j+1} - 1\right)}{p}\right) \\
&+ \csc\left(\frac{\pi}{p}\right) \sin\left(\frac{\pi\left(2s + 2^{j+1} + 1\right)}{p}\right) \\
&- 4\csc\left(\frac{2\pi}{p}\right) \sin\left(\frac{\pi\left(-4n + 2^{j} + 2s - 2\right)}{p}\right) \\
&+ \csc\left(\frac{\pi}{p}\right) \sin\left(\frac{2\pi s + \pi}{p}\right) \\
&+ 4\csc\left(\frac{2\pi}{p}\right) \sin\left(\frac{\pi\left(-2s + 2^{j} - 2\right)}{p}\right) \\
&+ 2\csc\left(\frac{\pi}{p}\right) \sin\left(\frac{\pi\left(-2s + 2^{j} - 1\right)}{p}\right) \\
&+ 2\csc\left(\frac{\pi}{p}\right) \sin\left(\frac{\pi\left(-2s + 2^{j} + 1\right)}{p}\right) \\
&- \csc\left(\frac{\pi}{p}\right) \sin\left(\frac{\pi\left(-2s + 2^{j+1} - 1\right)}{p}\right) \\
&- \csc\left(\frac{\pi}{p}\right) \sin\left(\frac{\pi\left(-2s + 2^{j+1} + 1\right)}{p}\right) \\
&+ 4\csc\left(\frac{2\pi}{p}\right) \sin\left(\frac{\pi\left(4n + 2^{j} - 2s + 2\right)}{p}\right)\Bigg)\Bigg]
\end{aligned}
$$

Thus we finally have:

$$\begin{aligned}
\mathrm{Cov}(d_{j,t}, d_{j,t+s}) &= \frac{2\cos\left(\frac{2\pi\left(2^j-s\right)}{p}\right) + 12\cos\left(\frac{2\pi s}{p}\right) + 2\cos\left(\frac{2\pi\left(s+2^j\right)}{p}\right)}{2^{j+4}\sin^2\frac{\pi}{p}} \\
&\quad + \frac{-8\cos\left(\frac{\pi\left(2s+2^j\right)}{p}\right) - 8\cos\left(\frac{\pi\left(2^j-2s\right)}{p}\right)}{2^{j+4}\sin^2\frac{\pi}{p}} \\
&\quad + \sigma_\epsilon^2 - \frac{3s}{2^j}\sigma_\epsilon^2 + o(1)_{s,j,p} \\
\mathrm{Var}(d_{j,t}) &= \frac{2\cos\left(\frac{2\pi 2^j}{p}\right) + 12 + 2\cos\left(\frac{2\pi 2^j}{p}\right) - 8\cos\left(\frac{\pi 2^j}{p}\right) - 8\cos\left(\frac{\pi 2^j}{p}\right)}{2^{j+4}\sin^2\frac{\pi}{p}} \\
&\quad + \sigma_\epsilon^2 + o(1)_{0,j,p} \\
&= \frac{4\cos\left(\frac{2^{j+1}\pi}{p}\right) - 16\cos\left(\frac{2^j\pi}{p}\right) + 12}{2^{j+4}\sin^2\frac{\pi}{p}} + \sigma_\epsilon^2 + o(1)_{0,j,p} \\
&= \frac{32\sin^4\left(\frac{2^{j-1}\pi}{p}\right)}{2^{j+4}\sin^2\left(\frac{\pi}{p}\right)} + \sigma_\epsilon^2 + o(1)_{0,j,p}
\end{aligned}$$

Thus finally:

$$\mathrm{Var}(d_{j,t}) = \frac{\sin^4\left(\frac{2^{j-1}\pi}{p}\right)}{2^{j-1}\sin^2\frac{\pi}{p}} + \sigma_\epsilon^2 + o(1)_{s=0}$$

as expected.

5.6.2 Proof of Theorem 2

Here we prove the the distribution of the square of Non-Decimated Haar Wavelet cofficients summed across each scale as described in Theorem 5.2. First we determine the distribution of the squared coefficients, then the correlation structure between each coefficient within a scale, before determining the distribution of the sum.

First we begin by considering the distribution of the squared coefficient. We begin by detailing the distribution of our white noise series Y_t as normal and

considering the Non-Decimated Haar Wavelet transform as a linear combination:

$$\begin{aligned}
&Y_t \sim N(0, \sigma_\epsilon^2) \\
\Rightarrow &d_{j,t} = \frac{1}{\sqrt{2}^j}\left[\sum_{i=0}^{2^{j-1}-1} Y_{t+i} - \sum_{i=2^{j-1}}^{2^j-1} Y_{t+i}\right] \sim N\left(0, \sigma_\epsilon^2\right) \\
\Rightarrow &\frac{d_{j,t}^2}{\sigma_\epsilon^2} \sim \mathcal{X}_1^2 \Rightarrow d_{j,t}^2 \sim \Gamma\left(\frac{1}{2}, 2\sigma_\epsilon^2\right)
\end{aligned}$$

Now we look to study the correlation structure of the squared coefficients $d_{j,k}^2$,

$$\text{Corr}(d_{j,s}^2, d_{j,v}^2) = \frac{\text{Cov}(d_{j,s}^2, d_{j,v}^2)}{\sqrt{\text{Var}(d_{j,s}^2)\text{Var}(d_{j,v}^2)}}.$$

We begin by studying the covariance:

$$\text{Cov}(d_{j,s}^2, d_{j,v}^2) = \mathbb{E}(d_{j,s}^2 d_{j,v}^2) - \mathbb{E}(d_{j,s}^2)\mathbb{E}(d_{j,v}^2)$$

Note that if $|v - s| \geq 2^j$ then $d_{j,s}^2$ and $d_{j,v}^2$ are independent, making the covariance zero. Thus we study the range $|v - s| < 2^j$, beginning with:

$$\begin{aligned}
\mathbb{E}(d_{j,s}^2) &= \mathbb{E}\left[\frac{1}{2^j}\left[\sum_{i=0}^{2^{j-1}-1} \epsilon_{s+i} - \sum_{i=2^{j-1}}^{2^j-1} \epsilon_{s+i}\right]^2\right] \\
&= \mathbb{E}\left[\frac{1}{2^j}\left[\sum_{i=0}^{2^{j-1}-1}\sum_{k=0}^{2^{j-1}-1} \epsilon_{s+i}\epsilon_{s+k} - 2\sum_{i=0}^{2^{j-1}-1}\sum_{k=2^{j-1}}^{2^j-1} \epsilon_{s+i}\epsilon_{s+k}\right.\right. \\
&\qquad \left.\left. + \sum_{i=2^{j-1}}^{2^j-1}\sum_{k=2^{j-1}}^{2^j-1} \epsilon_{s+i}\epsilon_{s+k}\right]\right] \\
&= \frac{1}{2^j}\left[2^{j-1}\sigma_\epsilon^2 + 0 + 2^{j-1}\sigma_\epsilon^2\right] \\
&= \sigma_\epsilon^2.
\end{aligned}$$

Next we consider the joint expectation,

$$
\begin{aligned}
\mathbb{E}(d_{j,s}^2 d_{j,v}^2) &= \mathbb{E}\left[\frac{1}{2^{2j}}\left[\sum_{i=0}^{2^{j-1}-1}\epsilon_{s+i} - \sum_{i=2^{j-1}}^{2^j-1}\epsilon_{s+i}\right]^2\left[\sum_{i=0}^{2^{j-1}-1}\epsilon_{v+i} - \sum_{i=2^{j-1}}^{2^j-1}\epsilon_{v+i}\right]^2\right] \\
&= \frac{1}{2^{2j}}\mathbb{E}\left[\left[\sum_{i=0}^{2^{j-1}-2}\sum_{l=0}^{2^{j-1}-1}\epsilon_{s+i}\epsilon_{s+l} - 2\sum_{i=0}^{2^{j-1}-2}\sum_{l=2^{j-1}}^{2^j-1}\epsilon_{s+i}\epsilon_{s+l}\right.\right. \\
&\quad \left.+ \sum_{i=2^{j-1}}^{2^j-1}\sum_{l=2^{j-1}}^{2^j-1}\epsilon_{s+i}\epsilon_{s+l}\right]\left[\sum_{i=0}^{2^{j-1}-2}\sum_{l=0}^{2^{j-1}-1}\epsilon_{v+i}\epsilon_{v+l}\right. \\
&\quad \left.\left.-2\sum_{i=0}^{2^{j-1}-2}\sum_{l=2^{j-1}}^{2^j-1}\epsilon_{v+i}\epsilon_{v+l} + \sum_{i=2^{j-1}}^{2^j-1}\sum_{l=2^{j-1}}^{2^j-1}\epsilon_{v+i}\epsilon_{v+l}\right]\right] \\
&= \frac{1}{2^{2j}}\left[\sum_{i=0}^{2^{j-1}-1}\sum_{l=0}^{2^{j-1}-1}\sum_{m=0}^{2^{j-1}-1}\sum_{n=0}^{2^{j-1}-1}\mathbb{E}(\epsilon_{s+i}\epsilon_{s+l}\epsilon_{v+n}\epsilon_{v+m})\right. \\
&\quad - 2\sum_{i=0}^{2^{j-1}-1}\sum_{l=0}^{2^{j-1}-1}\sum_{m=0}^{2^{j-1}-1}\sum_{n=2^{j-1}}^{2^j-1}\mathbb{E}(\epsilon_{s+i}\epsilon_{s+l}\epsilon_{v+n}\epsilon_{v+m}) \\
&\quad + \sum_{i=0}^{2^{j-1}-1}\sum_{l=0}^{2^{j-1}-1}\sum_{m=2^{j-1}}^{2^j-1}\sum_{n=2^{j-1}}^{2^j-1}\mathbb{E}(\epsilon_{s+i}\epsilon_{s+l}\epsilon_{v+n}\epsilon_{v+m}) \\
&\quad - 2\sum_{i=0}^{2^{j-1}-1}\sum_{l=2^{j-1}}^{2^j-1}\sum_{m=0}^{2^{j-1}-1}\sum_{n=0}^{2^{j-1}-1}\mathbb{E}(\epsilon_{s+i}\epsilon_{s+l}\epsilon_{v+n}\epsilon_{v+m}) \\
&\quad + 4\sum_{i=0}^{2^{j-1}-1}\sum_{l=2^{j-1}}^{2^j-1}\sum_{m=0}^{2^{j-1}-1}\sum_{n=2^{j-1}}^{2^j-1}\mathbb{E}(\epsilon_{s+i}\epsilon_{s+l}\epsilon_{v+n}\epsilon_{v+m}) \\
&\quad - 2\sum_{i=0}^{2^{j-1}-1}\sum_{l=2^{j-1}}^{2^j-1}\sum_{m=2^{j-1}}^{2^j-1}\sum_{n=2^{j-1}}^{2^j-1}\mathbb{E}(\epsilon_{s+i}\epsilon_{s+l}\epsilon_{v+n}\epsilon_{v+m}) \\
&\quad + \sum_{i=2^{j-1}}^{2^j-1}\sum_{l=2^{j-1}}^{2^j-1}\sum_{m=0}^{2^{j-1}-1}\sum_{n=0}^{2^{j-1}-1}\mathbb{E}(\epsilon_{s+i}\epsilon_{s+l}\epsilon_{v+n}\epsilon_{v+m}) \\
&\quad - 2\sum_{i=2^{j-1}}^{2^j-1}\sum_{l=2^{j-1}}^{2^j-1}\sum_{m=0}^{2^{j-1}-1}\sum_{n=2^{j-1}}^{2^j-1}\mathbb{E}(\epsilon_{s+i}\epsilon_{s+l}\epsilon_{v+n}\epsilon_{v+m}) \\
&\quad \left.+ \sum_{i=2^{j-1}}^{2^j-1}\sum_{l=2^{j-1}}^{2^j-1}\sum_{m=2^{j-1}}^{2^j-1}\sum_{n=2^{j-1}}^{2^j-1}\mathbb{E}(\epsilon_{s+i}\epsilon_{s+l}\epsilon_{v+n}\epsilon_{v+m})\right] \\
&= \lambda_{s,v}
\end{aligned}
$$

Note that by Isserli's Theorem:

$$\begin{aligned}\mathbb{E}(\epsilon_{s+i}\epsilon_{s+l}\epsilon_{v+n}\epsilon_{v+m}) &= \mathbb{E}(\epsilon_{s+i}\epsilon_{s+l})\mathbb{E}(\epsilon_{v+n}\epsilon_{v+m}) + \mathbb{E}(\epsilon_{s+i}\epsilon_{v+n})\mathbb{E}(\epsilon_{s+l}\epsilon_{v+m}) + \\ &\qquad \mathbb{E}(\epsilon_{s+i}\epsilon_{v+m})\mathbb{E}(\epsilon_{v+n}\epsilon_{s+l}) \\ &= \left[\delta_{i,l}\delta_{n,m} + \delta_{s+i,v+n}\delta_{s+l,v+m} + \delta_{s+i,v+m}\delta_{v+n,s+l}\right]\sigma_\epsilon^4\end{aligned}$$

Thus we continue:

$$\begin{aligned}\lambda_{s,v} = \frac{\sigma_\epsilon^4}{2^{2j}} \Bigg[&\sum_{i=0}^{2^{j-1}-1}\sum_{l=0}^{2^{j-1}-1}\sum_{m=0}^{2^{j-1}-1}\sum_{n=0}^{2^{j-1}-1} \delta_{i,l}\delta_{n,m} + \delta_{s+i,v+n}\delta_{s+l,v+m} + \delta_{s+i,v+m}\delta_{v+n,s+l} \\ &- 2\sum_{i=0}^{2^{j-1}-1}\sum_{l=0}^{2^{j-1}-1}\sum_{m=0}^{2^{j-1}-1}\sum_{n=2^{j-1}}^{2^{j}-1} \delta_{i,l}\delta_{n,m} + \delta_{s+i,v+n}\delta_{s+l,v+m} + \delta_{s+i,v+m}\delta_{v+n,s+l} \\ &+ \sum_{i=0}^{2^{j-1}-1}\sum_{l=0}^{2^{j-1}-1}\sum_{m=2^{j-1}}^{2^{j}-1}\sum_{n=2^{j-1}}^{2^{j}-1} \delta_{i,l}\delta_{n,m} + \delta_{s+i,v+n}\delta_{s+l,v+m} + \delta_{s+i,v+m}\delta_{v+n,s+l} \\ &- 2\sum_{i=0}^{2^{j-1}-1}\sum_{l=2^{j-1}}^{2^{j}-1}\sum_{m=0}^{2^{j-1}-1}\sum_{n=0}^{2^{j-1}-1} \delta_{i,l}\delta_{n,m} + \delta_{s+i,v+n}\delta_{s+l,v+m} + \delta_{s+i,v+m}\delta_{v+n,s+l} \\ &+ 4\sum_{i=0}^{2^{j-1}-1}\sum_{l=2^{j-1}}^{2^{j}-1}\sum_{m=0}^{2^{j-1}-1}\sum_{n=2^{j-1}}^{2^{j}-1} \delta_{i,l}\delta_{n,m} + \delta_{s+i,v+n}\delta_{s+l,v+m} + \delta_{s+i,v+m}\delta_{v+n,s+l} \\ &- 2\sum_{i=0}^{2^{j-1}-1}\sum_{l=2^{j-1}}^{2^{j}-1}\sum_{m=2^{j-1}}^{2^{j}-1}\sum_{n=2^{j-1}}^{2^{j}-1} \delta_{i,l}\delta_{n,m} + \delta_{s+i,v+n}\delta_{s+l,v+m} + \delta_{s+i,v+m}\delta_{v+n,s+l} \\ &+ \sum_{i=2^{j-1}}^{2^{j}-1}\sum_{l=2^{j-1}}^{2^{j}-1}\sum_{m=0}^{2^{j-1}-1}\sum_{n=0}^{2^{j-1}-1} \delta_{i,l}\delta_{n,m} + \delta_{s+i,v+n}\delta_{s+l,v+m} + \delta_{s+i,v+m}\delta_{v+n,s+l} \\ &- 2\sum_{i=2^{j-1}}^{2^{j}-1}\sum_{l=2^{j-1}}^{2^{j}-1}\sum_{m=0}^{2^{j-1}-1}\sum_{n=2^{j-1}}^{2^{j}-1} \delta_{i,l}\delta_{n,m} + \delta_{s+i,v+n}\delta_{s+l,v+m} + \delta_{s+i,v+m}\delta_{v+n,s+l} \\ &+ \sum_{i=2^{j-1}}^{2^{j}-1}\sum_{l=2^{j-1}}^{2^{j}-1}\sum_{m=2^{j-1}}^{2^{j}-1}\sum_{n=2^{j-1}}^{2^{j}-1} \delta_{i,l}\delta_{n,m} + \delta_{s+i,v+n}\delta_{s+l,v+m} + \delta_{s+i,v+m}\delta_{v+n,s+l} \Bigg]\end{aligned}$$

The first column of sums can be easily calculated to give,

$$\begin{aligned}
\lambda_{s,v} = \frac{\sigma_\epsilon^4}{2^{2j}} \Bigg[& \left(4^{j-1} + \sum_{i=0}^{2^{j-1}-1} \sum_{l=0}^{2^{j-1}-1} \sum_{m=0}^{2^{j-1}-1} \sum_{n=0}^{2^{j-1}-1} \delta_{s+i,v+n}\delta_{s+l,v+m} + \delta_{s+i,v+m}\delta_{v+n,s+l} \right) \\
& - 2 \left(0 + \sum_{i=0}^{2^{j-1}-1} \sum_{l=0}^{2^{j-1}-1} \sum_{m=0}^{2^{j-1}-1} \sum_{n=2^{j-1}}^{2^j-1} \delta_{s+i,v+n}\delta_{s+l,v+m} + \delta_{s+i,v+m}\delta_{v+n,s+l} \right) \\
& + \left(4^{j-1} + \sum_{i=0}^{2^{j-1}-1} \sum_{l=0}^{2^{j-1}-1} \sum_{m=2^{j-1}}^{2^j-1} \sum_{n=2^{j-1}}^{2^j-1} \delta_{s+i,v+n}\delta_{s+l,v+m} + \delta_{s+i,v+m}\delta_{v+n,s+l} \right) \\
& - 2 \left(0 + \sum_{i=0}^{2^{j-1}-1} \sum_{l=2^{j-1}}^{2^j-1} \sum_{m=0}^{2^{j-1}-1} \sum_{n=0}^{2^{j-1}-1} \delta_{s+i,v+n}\delta_{s+l,v+m} + \delta_{s+i,v+m}\delta_{v+n,s+l} \right) \\
& + 4 \left(0 + \sum_{i=0}^{2^{j-1}-1} \sum_{l=2^{j-1}}^{2^j-1} \sum_{m=0}^{2^{j-1}-1} \sum_{n=2^{j-1}}^{2^j-1} \delta_{s+i,v+n}\delta_{s+l,v+m} + \delta_{s+i,v+m}\delta_{v+n,s+l} \right) \\
& - 2 \left(0 + \sum_{i=0}^{2^{j-1}-1} \sum_{l=2^{j-1}}^{2^j-1} \sum_{m=2^{j-1}}^{2^j-1} \sum_{n=2^{j-1}}^{2^j-1} \delta_{s+i,v+n}\delta_{s+l,v+m} + \delta_{s+i,v+m}\delta_{v+n,s+l} \right) \\
& + \left(4^{j-1} + \sum_{i=2^{j-1}}^{2^j-1} \sum_{l=2^{j-1}}^{2^j-1} \sum_{m=0}^{2^{j-1}-1} \sum_{n=0}^{2^{j-1}-1} \delta_{s+i,v+n}\delta_{s+l,v+m} + \delta_{s+i,v+m}\delta_{v+n,s+l} \right) \\
& - 2 \left(0 + \sum_{i=2^{j-1}}^{2^j-1} \sum_{l=2^{j-1}}^{2^j-1} \sum_{m=0}^{2^{j-1}-1} \sum_{n=2^{j-1}}^{2^j-1} \delta_{s+i,v+n}\delta_{s+l,v+m} + \delta_{s+i,v+m}\delta_{v+n,s+l} \right) \\
& + \left(4^{j-1} + \sum_{i=2^{j-1}}^{2^j-1} \sum_{l=2^{j-1}}^{2^j-1} \sum_{m=2^{j-1}}^{2^j-1} \sum_{n=2^{j-1}}^{2^j-1} \delta_{s+i,v+n}\delta_{s+l,v+m} + \delta_{s+i,v+m}\delta_{v+n,s+l} \right) \Bigg]
\end{aligned}$$

Reducing the second column we get,

$$\lambda_{s,v} = \sigma_\epsilon^4 + \frac{\sigma_\epsilon^4}{2^{2j}} \Bigg[\Bigg(\mathbb{I}\{1 \le |s-v| \le 2^{j-1}\}(2^{j-1} - |s-v|)^2 + \sum_{i=0}^{2^{j-1}-1} \sum_{l=0}^{2^{j-1}-1} \sum_{m=0}^{2^{j-1}-1} \sum_{n=0}^{2^{j-1}-1} \delta_{s+i,v+m}\delta_{v+n,s+l} \Bigg)$$

$$- 2 \Bigg(\mathbb{I}\{1 \le s-v \le 2^{j-1}\}(s-v)(2^{j-1} - (s-v)) + \sum_{i=0}^{2^{j-1}-1} \sum_{i=0}^{2^{j-1}-1} \sum_{l=0}^{2^{j-1}-1} \sum_{m=0}^{2^{j-1}-1} \sum_{n=2^{j-1}}^{2^j-1} \delta_{s+i,v+m}\delta_{v+n,s+l} \Bigg)$$

$$+ \Bigg(\mathbb{I}\{1 \le s-v \le 2^{j}\}(2^{j-1} - |s-v-2^{j-1}|)^2 + \sum_{i=0}^{2^{j-1}-1} \sum_{l=0}^{2^{j-1}-1} \sum_{m=2^{j-1}}^{2^j-1} \sum_{n=2^{j-1}}^{2^j-1} \delta_{s+i,v+m}\delta_{v+n,s+l} \Bigg)$$

$$- 2 \Bigg(\mathbb{I}\{1 \le v-s \le 2^{j-1}\}(v-s)(2^{j-1} - (v-s)) + \sum_{i=0}^{2^{j-1}-1} \sum_{l=2^{j-1}}^{2^j-1} \sum_{m=0}^{2^{j-1}-1} \sum_{n=0}^{2^{j-1}-1} \delta_{s+i,v+m}\delta_{v+n,s+l} \Bigg)$$

$$+ 4 \Bigg(0 + \sum_{i=0}^{2^{j-1}-1} \sum_{l=2^{j-1}}^{2^j-1} \sum_{m=0}^{2^{j-1}-1} \sum_{n=2^{j-1}}^{2^j-1} \delta_{s+i,v+m}\delta_{v+n,s+l} \Bigg)$$

$$- 2 \Bigg(\mathbb{I}\{1 \le s-v \le 2^{j-1}\}(s-v)(2^{j-1} - (s-v)) + \sum_{i=0}^{2^{j-1}-1} \sum_{l=2^{j-1}}^{2^j-1} \sum_{m=2^{j-1}}^{2^j-1} \sum_{n=2^{j-1}}^{2^j-1} \delta_{s+i,v+m}\delta_{v+n,s+l} \Bigg)$$

$$+ \Bigg(\mathbb{I}\{1 \le v-s \le 2^{j}\}(2^{j-1} - |v-s-2^{j-1}|)^2 + \sum_{i=2^{j-1}}^{2^j-1} \sum_{l=2^{j-1}}^{2^j-1} \sum_{m=0}^{2^{j-1}-1} \sum_{n=0}^{2^{j-1}-1} \delta_{s+i,v+m}\delta_{v+n,s+l} \Bigg)$$

$$+ \ldots \text{(continued overleaf)}$$

$$\lambda_{s,v} = \ldots$$

$$- 2 \Bigg(\mathbb{I}\{1 \leq v - s \leq 2^{j-1}\}(v-s)(2^{j-1} - (v-s))$$

$$+ \sum_{i=2^{j-1}}^{2^j-1} \sum_{l=2^{j-1}}^{2^j-1} \sum_{m=0}^{2^{j-1}-1} \sum_{n=2^{j-1}}^{2^j-1} \delta_{s+i,v+m}\delta_{v+n,s+l} \Bigg)$$

$$+ \Bigg(\mathbb{I}\{1 \leq |s - v| \leq 2^{j-1}\}(2^{j-1} - |s-v|)^2$$

$$+ \sum_{i=2^{j-1}}^{2^j-1} \sum_{l=2^{j-1}}^{2^j-1} \sum_{m=2^{j-1}}^{2^j-1} \sum_{n=2^{j-1}}^{2^j-1} \delta_{s+i,v+m}\delta_{v+n,s+l} \Bigg) \Bigg]$$

Noting the reflections we combine to get:

$$\lambda_{s,v} = \frac{\sigma_\epsilon^4}{2^{2j}} \Bigg[\left(\sum_{i=0}^{2^{j-1}-1} \sum_{l=0}^{2^{j-1}-1} \sum_{m=0}^{2^{j-1}-1} \sum_{n=0}^{2^{j-1}-1} \delta_{s+i,v+m}\delta_{v+n,s+l} \right)$$
$$- 2 \left(\sum_{i=0}^{2^{j-1}-1} \sum_{l=0}^{2^{j-1}-1} \sum_{m=0}^{2^{j-1}-1} \sum_{n=2^{j-1}}^{2^j-1} \delta_{s+i,v+m}\delta_{v+n,s+l} \right)$$
$$+ \left(\sum_{i=0}^{2^{j-1}-1} \sum_{l=0}^{2^{j-1}-1} \sum_{m=2^{j-1}}^{2^j-1} \sum_{n=2^{j-1}}^{2^j-1} \delta_{s+i,v+m}\delta_{v+n,s+l} \right)$$
$$- 2 \left(\sum_{i=0}^{2^{j-1}-1} \sum_{l=2^{j-1}}^{2^j-1} \sum_{m=0}^{2^{j-1}-1} \sum_{n=0}^{2^{j-1}-1} \delta_{s+i,v+m}\delta_{v+n,s+l} \right)$$
$$+ 4 \left(\sum_{i=0}^{2^{j-1}-1} \sum_{l=2^{j-1}}^{2^j-1} \sum_{m=0}^{2^{j-1}-1} \sum_{n=2^{j-1}}^{2^j-1} \delta_{s+i,v+m}\delta_{v+n,s+l} \right)$$
$$- 2 \left(\sum_{i=0}^{2^{j-1}-1} \sum_{l=2^{j-1}}^{2^j-1} \sum_{m=2^{j-1}}^{2^j-1} \sum_{n=2^{j-1}}^{2^j-1} \delta_{s+i,v+m}\delta_{v+n,s+l} \right)$$
$$+ \left(\sum_{i=2^{j-1}}^{2^j-1} \sum_{l=2^{j-1}}^{2^j-1} \sum_{m=0}^{2^{j-1}-1} \sum_{n=0}^{2^{j-1}-1} \delta_{s+i,v+m}\delta_{v+n,s+l} \right)$$
$$- 2 \left(\sum_{i=2^{j-1}}^{2^j-1} \sum_{l=2^{j-1}}^{2^j-1} \sum_{m=0}^{2^{j-1}-1} \sum_{n=2^{j-1}}^{2^j-1} \delta_{s+i,v+m}\delta_{v+n,s+l} \right)$$
$$+ \left(\sum_{i=2^{j-1}}^{2^j-1} \sum_{l=2^{j-1}}^{2^j-1} \sum_{m=2^{j-1}}^{2^j-1} \sum_{n=2^{j-1}}^{2^j-1} \delta_{s+i,v+m}\delta_{v+n,s+l} \right) \Bigg]$$
$$+ \sigma_\epsilon^4 \begin{cases} 1 - \frac{|s-v|(2^{j-1}-|s-v|)}{2^{2j-2}} + \frac{(2^{j-1}-|s-v|)^2}{2^{2j-1}} & \\ \qquad + \frac{(2^{j-1}-||s-v|-2^{j-1}|)^2}{2^{2j}} & \text{if } 1 \le |s-v| \le 2^{j-1} \\ 1 + \frac{(2^{j-1}-||s-v|-2^{j-1}|)^2}{2^{2j}} & \text{if } 2^{j-1}+1 \le |s-v| \le 2^j \end{cases}$$

Finally we solve the sums to get:

$$
\begin{aligned}
\lambda_{s,v} =& \frac{\sigma_\epsilon^4}{2^{2j}} \big[\mathbb{I}\{1 \le |s-v| \le 2^{j-1}\}(2^{j-1} - |s-v|)^2 \\
&\quad - 2\left(\mathbb{I}\{1 \le s-v \le 2^{j-1}\}(s-v)(2^{j-1} - (s-v))\right) \\
&\quad + \mathbb{I}\{1 \le s-v \le 2^{j}\}(2^{j-1} - |s-v-2^{j-1}|)^2 \\
&\quad - 2\left(\mathbb{I}\{1 \le v-s \le 2^{j-1}\}(v-s)(2^{j-1} - (v-s))\right) \\
&\quad + 4\left(\mathbb{I}\{1 \le |s-v| \le 2^{j-1}\}(2^{j-1} - |s-v|)^2\right) \\
&\quad - 2\left(\mathbb{I}\{1 \le s-v \le 2^{j-1}\}(s-v)(2^{j-1} - (s-v))\right) \\
&\quad + \left(\mathbb{I}\{1 \le v-s \le 2^{j}\}(2^{j-1} - |v-s-2^{j-1}|)^2\right) \\
&\quad - 2\left(\mathbb{I}\{1 \le v-s \le 2^{j-1}\}(v-s)(2^{j-1} - (v-s))\right) \\
&\quad + \left(\mathbb{I}\{1 \le |s-v| \le 2^{j-1}\}(2^{j-1} - |s-v|)^2\right) \big] \\
&+ \sigma_\epsilon^4 \begin{cases} 1 - \frac{|s-v|(2^{j-1}-|s-v|)}{2^{2j-2}} + \frac{(2^{j-1}-|s-v|)^2}{2^{2j-1}} & \\ \qquad + \frac{(2^{j-1}-||s-v|-2^{j-1}|)^2}{2^{2j}} & \text{if } 1 \le |s-v| \le 2^{j-1} \\ 1 + \frac{(2^{j-1}-||s-v|-2^{j-1}|)^2}{2^{2j}} & \text{if } 2^{j-1}+1 \le |s-v| \le 2^{j} \end{cases}
\end{aligned}
$$

Again noting the symmetry we reduce to:

$$\lambda_{s,v} = \sigma_\epsilon^4 \begin{cases} 1 - \frac{|s-v|(2^{j-1}-|s-v|)}{2^{2j-2}} + \frac{(2^{j-1}-|s-v|)^2}{2^{2j-1}} & \\ \quad + \frac{(2^{j-1}-||s-v|-2^{j-1}|)^2}{2^{2j}} & \text{if } 1 \le |s-v| \le 2^{j-1} \\ 1 + \frac{(2^{j-1}-||s-v|-2^{j-1}|)^2}{2^{2j}} & \text{if } 2^{j-1}+1 \le |s-v| \le 2^j \end{cases}$$

$$+ \sigma_\epsilon^4 \begin{cases} -\frac{|s-v|(2^{j-1}-|s-v|)}{2^{2j-2}} + \frac{3(2^{j-1}-|s-v|)^2}{2^{2j-1}} & \\ \quad + \frac{(2^{j-1}-||s-v|-2^{j-1}|)^2}{2^{2j}} & \text{if } 1 \le |s-v| \le 2^{j-1} \\ \frac{(2^{j-1}-||s-v|-2^{j-1}|)^2}{2^{2j}} & \text{if } 2^{j-1}+1 \le |s-v| \le 2^j \end{cases}$$

$$= \sigma_\epsilon^4 \begin{cases} 1 - \frac{|s-v|(2^{j-1}-|s-v|)}{2^{2j-3}} + \frac{(2^{j-1}-|s-v|)^2}{2^{2j-3}} & \\ \quad + \frac{(2^{j-1}-||s-v|-2^{j-1}|)^2}{2^{2j-1}} & \text{if } 1 \le |s-v| \le 2^{j-1} \\ 1 + \frac{(2^{j-1}-||s-v|-2^{j-1}|)^2}{2^{2j-1}} & \text{if } 2^{j-1}+1 \le |s-v| \le 2^j \end{cases}$$

$$= \sigma_\epsilon^4 \begin{cases} 4 + \frac{9(s-v)^2}{2^{2j-1}} - \frac{14|s-v|+|2^j-2|s-v||}{2^j} & \text{if } 1 \le |s-v| \le 2^{j-1} \\ 1 + \frac{(2^{j-1}-||s-v|-2^{j-1}|)^2}{2^{2j-1}} & \text{if } 2^{j-1}+1 \le |s-v| \le 2^j \end{cases}$$

Recall from earlier that $d_{j,k}^2 \sim \text{Gamma}(\frac{1}{2}, 2\sigma_\epsilon^2)$. We can now combine this information to get:

$$\text{Corr}(d_{j,s}^2, d_{j,v}^2) = \frac{\text{Cov}(d_{j,s}^2, d_{j,v}^2)}{\sqrt{\text{Var}(d_{j,s}^2)\text{Var}(d_{j,v}^2)}}$$

$$= \begin{cases} \frac{\left[4+\frac{9(s-v)^2}{2^{2j-1}} - \frac{14|s-v|+|2^j-2|s-v||}{2^j}\right]\sigma_\epsilon^4 - \sigma_\epsilon^4}{2\sigma_\epsilon^4} & \text{if } 1 \le |s-v| \le 2^{j-1} \\ \frac{\left[1+\frac{(2^{j-1}-||s-v|-2^{j-1}|)^2}{2^{2j-1}}\right]\sigma_\epsilon^4 - \sigma_\epsilon^4}{2\sigma_\epsilon^4} & \text{if } 2^{j-1}+1 \le |s-v| \le 2^j \end{cases}$$

$$= \begin{cases} \frac{3}{2} + \left(\frac{3(s-v)}{2^j}\right)^2 - \frac{7|s-v|+|2^{j-1}-|s-v||}{2^j} & \text{if } 1 \le |s-v| \le 2^{j-1} \\ \left(\frac{(2^{j-1}-||s-v|-2^{j-1}|)}{2^j}\right)^2 & \text{if } 2^{j-1}+1 \le |s-v| \le 2^j \end{cases}$$

Now that we have the covariance and correlation of the $d^2_{j,k}$ we follow the approach of (Feng et al., 2016) which says that if we have correlated gamma random variables

$$\gamma_n \sim \Gamma\left(m_n, \frac{\Omega_n}{m_n}\right) \qquad n = 1, 2, \ldots, N.$$

With covariance $\text{Cov}(\gamma_i, \gamma_j) = \mathbf{R}_\gamma(i,j)$ then the sum of these variables $\sum_{n=1}^N \gamma_n = \gamma$ can be approximately distributed as

$$\gamma \sim \text{Gamma}\left(\frac{\left(\sum_{n=1}^N \Omega_n\right)^2}{\sum_{i=1}^N \sum_{j=1}^N \mathbf{R}_\gamma(i,j)}, \frac{\sum_{i=1}^N \sum_{j=1}^N \mathbf{R}_\gamma(i,j)}{\sum_{n=1}^N \Omega_n}\right).$$

Given we know that $m_n = \frac{1}{2}$ and thus $\Omega_n = \sigma^2_\epsilon$, we need to determine:

$$\sum_{s=1}^N \sum_{v=1}^N \mathbf{R}_{d^2_j}(s,v) = \sum_{s=1}^N \sum_{v=1}^N \text{Cov}(d^2_{j,s}, d^2_{j,v})$$

Note that we let $N = 2^J - 2^j + 1$ as this captures the non-decimated wavelet coefficients which do not overlap the end of the series on scale j. Thus we continue

$$\begin{aligned}
&\sum_{s=1}^{2^J-2^j+1} \sum_{v=1}^{2^J-2^j+1} \mathbf{R}_{d^2_j}(s,v) \\
&= \sum_{s=1}^{2^J-2^j+1} \sum_{v=1}^{2^J-2^j+1} \sigma^4_\epsilon \begin{cases} 3 + \frac{9(s-v)^2}{2^{2j-1}} - \frac{14|s-v|+|2^j-2|s-v||}{2^j} & \text{if } 0 \leq |s-v| \leq 2^{j-1} \\ \frac{(2^{j-1}-||s-v|-2^{j-1}|)^2}{2^{2j-1}} & \text{if } 2^{j-1}+1 \leq |s-v| \leq 2^j \end{cases} \\
&= \sigma^4_\epsilon \sum_{s=1}^{2^J-2^j+1} \sum_{v=1}^{2^J-2^j+1} \begin{cases} f_1(s-v) & \text{if } 0 \leq |s-v| \leq 2^{j-1} \\ f_2(s-v) & \text{if } 2^{j-1}+1 \leq |s-v| \leq 2^j \end{cases}
\end{aligned}$$

This can split into two cases: $J = j$ and $J > j$. In the first case we are left with:

$$\mathbf{R}_{d^2_j}(1,1) = \sigma^4_\epsilon f_1(0) = 2\sigma^4_\epsilon$$

Considering the second case we continue with:

$$
\begin{aligned}
&\sum_{s=1}^{2^J-2^j+1}\sum_{v=1}^{2^J-2^j+1}\mathbf{R}_{d_j^2}(s,v)\\
&\quad=\sigma_\epsilon^4\sum_{s=1}^{2^J-2^j+1}\sum_{v=1}^{2^J-2^j+1}\begin{cases}f_1(s-v) & \text{if } 0\le|s-v|\le 2^{j-1}\\ f_2(s-v) & \text{if } 2^{j-1}+1\le|s-v|\le 2^j\end{cases}\\
&\quad=(2^J-2^j+1)f_1(0)+2(2^J-2^j+1-1)f_1(1)+\dots\\
&\qquad\qquad\cdots+2(2^J-2^j+1-2^{j-1})f_1(2^{j-1})\\
&\qquad+2(2^J-2^j+1-(2^{j-1}+1))f_2(2^{j-1}+1)+\dots\\
&\qquad\qquad\cdots+f_2(2^j)(2^J-2^j+1-2^j)\\
&\quad=(2^J-2^j+1)2\sigma_\epsilon^4+2\sum_{i=1}^{2^{j-1}}(2^J-2^j+1-i)f_1(i)\\
&\qquad+2\sum_{i=2^{j-1}+1}^{2^j-1}(2^J-2^j+1-i)f_2(i)\\
&\quad=(2^J-2^j+1)2\sigma_\epsilon^4+2(2^J-2^j+1)\sum_{i=1}^{2^{j-1}}f_1(i)-2\sum_{i=1}^{2^{j-1}}if_1(i)\\
&\qquad+2(2^J-2^j+1)\sum_{i=2^{j-1}+1}^{2^j-1}f_2(i)-2\sum_{i=2^{j-1}+1}^{2^j-1}if_2(i)\\
&\quad=2(2^J-2^j+1)\left[\sigma_\epsilon^4+\sum_{i=1}^{2^{j-1}}f_1(i)+\sum_{i=2^{j-1}+1}^{2^j-1}f_2(i)\right]-\left[\sum_{i=1}^{2^{j-1}}if_1(i)+\sum_{i=2^{j-1}+1}^{2^j-1}if_2(i)\right]\\
&\quad=2(2^J-2^j+1)\left[\sigma_\epsilon^4+\frac{\sigma_\epsilon^4}{2^{j+2}}\left(6-3\cdot 2^j+4^j\right)+\frac{\sigma_\epsilon^4}{3\cdot 2^{j+2}}\left(2-3\cdot 2^j+4^j\right)\right]\\
&\qquad-\left[\frac{\sigma_\epsilon^4}{32}\left(2+2^j\right)^2+\frac{\sigma_\epsilon^4}{96}\left(-2+2^j\right)\left(-2+5\cdot 2^j\right)\right]
\end{aligned}
$$

The final results resolves to:

$$
\sum_{s=1}^{2^J-2^j+1}\sum_{v=1}^{2^J-2^j+1}\mathbf{R}_{d_j^2}(s,v)=\begin{cases}2\sigma_\epsilon^4 & \text{if } J=j\\ r_j\sigma_\epsilon^4 & \text{if } J>j\end{cases}
$$

where $r_j = 3^{-1} \cdot 2^{-j-1}\left(20 - 11(2^{j+1}) + 5(2^{2+J}) + 2^{2+2j+J} + 4^{j-1} - 5(8^j)\right)$.

With this we know that:

$$\sum_{k=1}^{2^J-2^j+1} d_{j,k}^2 \sim \begin{cases} \text{Gamma}\left(\frac{(\sigma_\epsilon^2)^2}{2\sigma_\epsilon^4}, \frac{2\sigma_\epsilon^4}{\sigma_\epsilon^2}\right) & \text{if } J = j \\ \text{Gamma}\left(\frac{((2^J-2^j+1)\sigma_\epsilon^2)^2}{r_j\sigma_\epsilon^4}, \frac{r_j\sigma_\epsilon^4}{(2^J-2^j+1)\sigma_\epsilon^2}\right) & \text{if } J > j \end{cases}$$

$$\sim \begin{cases} \text{Gamma}\left(\frac{1}{2}, 2\sigma_\epsilon^2\right) & \text{if } J = j \\ \text{Gamma}\left(\frac{(2^J-2^j+1)^2}{r_j}, \frac{r_j\sigma_\epsilon^2}{2^J-2^j+1}\right) & \text{if } J > j \end{cases}$$

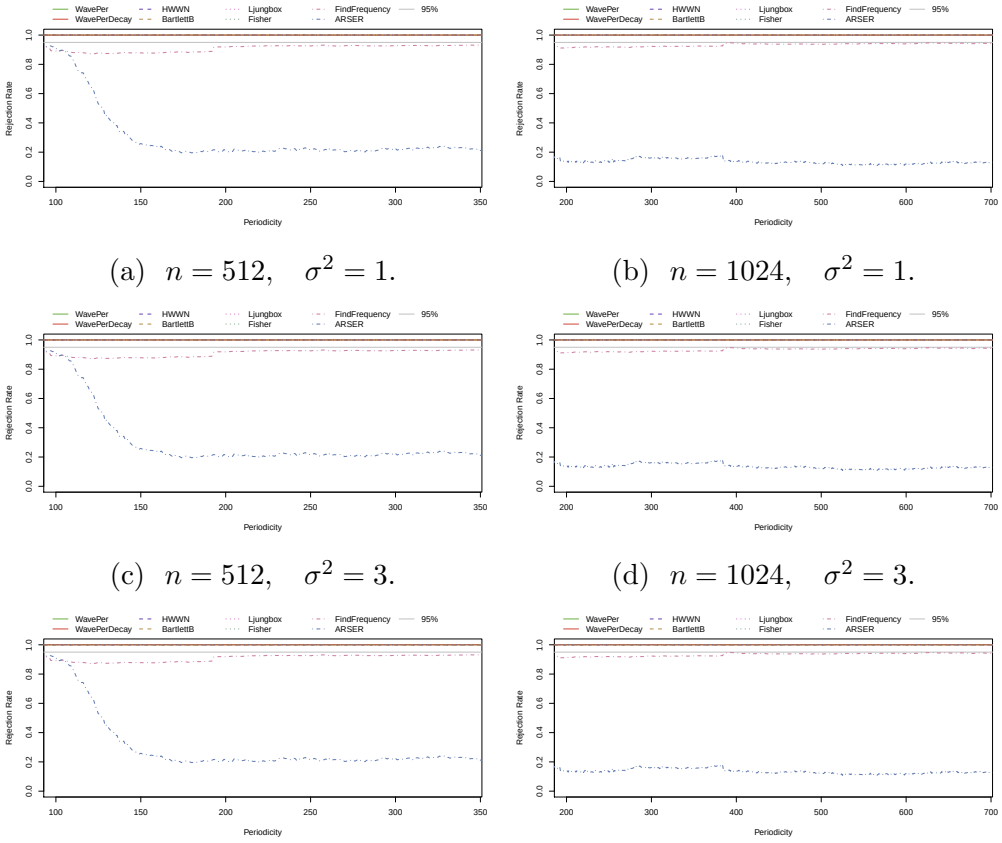
5.6.3 Proof of Theorem 3

This proof shows the quick calculation necessary to determine the periodicity where two scales cross by having the same absolute power.

We consider this in the case where $n \to \infty$ such that following Theorem 2 if

$$\begin{aligned} &\sum_{k=1}^{2^J-2^j+1} d_{j,k}^2 = \sum_{k=1}^{2^J-2^{j+1}+1} d_{j+1,k}^2 \\ \iff &\frac{\sin^4\left(\frac{2^{j-1}\pi}{p}\right)}{2^{j-1}\sin^2\left(\frac{\pi}{p}\right)} = \frac{\sin^4\left(\frac{2^j\pi}{p}\right)}{2^j\sin^2\left(\frac{\pi}{p}\right)} \\ \iff &2\sin^4\left(\frac{2^{j-1}\pi}{p}\right) = \sin^4\left(\frac{2^j\pi}{p}\right) \\ \iff &2\sin^4\left(\frac{2^{j-1}\pi}{p}\right) = 2^4\sin^4\left(\frac{2^{j-1}\pi}{p}\right)\cos^4\left(\frac{2^{j-1}\pi}{p}\right) \\ \iff &1 = 2^3\cos^4\left(\frac{2^{j-1}\pi}{p}\right) \\ \iff &2^{-\frac{3}{4}} = \cos\left(\frac{2^{j-1}\pi}{p}\right) \\ \iff &p = \frac{2^{j-1}\pi}{\cos^{-1}\left(2^{-\frac{3}{4}}\right)} \end{aligned}$$

5.6.4 Detection Graphs $n = 512, 1024$



(a) $n = 512, \quad \sigma^2 = 1.$

(b) $n = 1024, \quad \sigma^2 = 1.$

(c) $n = 512, \quad \sigma^2 = 3.$

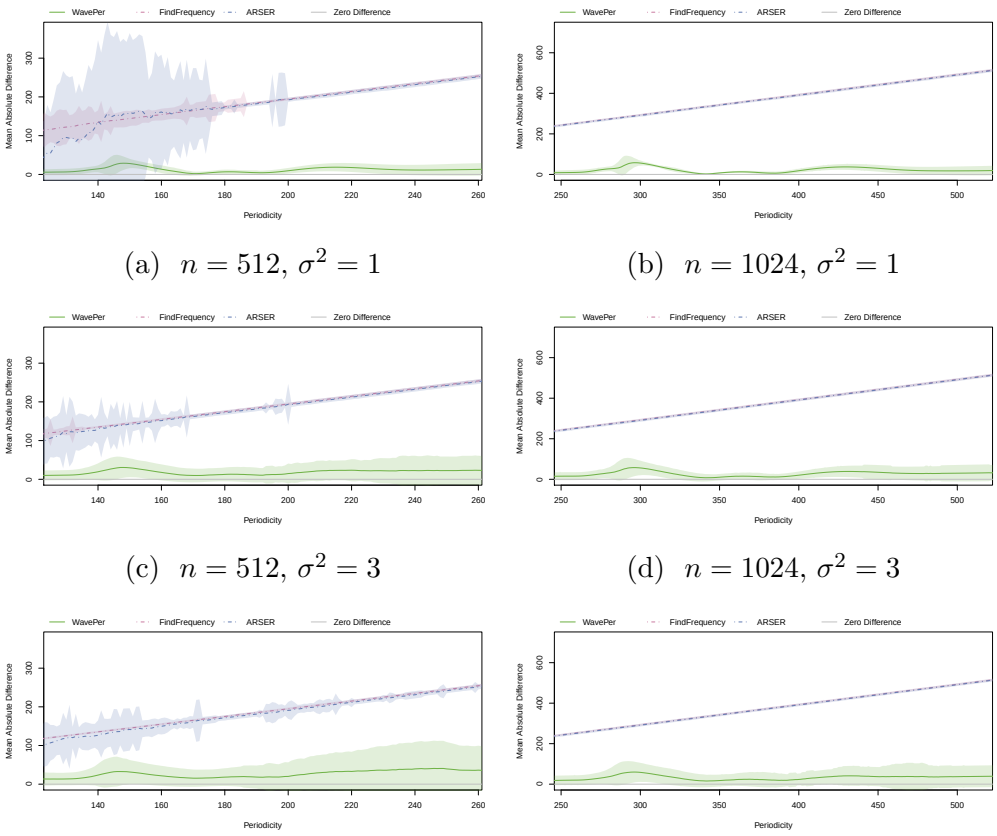
(d) $n = 1024, \quad \sigma^2 = 3.$

(e) $n = 512, \quad \sigma^2 = 5.$

(f) $n = 1024, \quad \sigma^2 = 5.$

Figure 5.12: Probability of Detecting Periodicity over a number of sizes and noise levels.

5.6.5 Estimation Graphs $n = 512, 1024$



(a) $n = 512$, $\sigma^2 = 1$ (b) $n = 1024$, $\sigma^2 = 1$

(c) $n = 512$, $\sigma^2 = 3$ (d) $n = 1024$, $\sigma^2 = 3$

(e) $n = 512$, $\sigma^2 = 5$ (f) $n = 1024$, $\sigma^2 = 5$

Figure 5.13: Mean Absolute Difference from Truth (Left) and amount of series considered (Right). Shaded areas are the Mean Absolute Difference plus or minus the standard deviation of the estimate.

Chapter 6

Automatic Dynamic Factor Model

6.1 Introduction

In comparison to the usual forecasting, which looks into the future, Nowcasting is the prediction of a series right now. This is particularly important for European Area Gross Domestic Product (EAGDP), a key measure of the output of the European Area (a full defintion provided by OECD can be found in OECD (2018)). Whilst this may seem unusual for such a key quantity, there are a significant number of series forming this measure for which the truth is often very hard to determine. Given the time taken for data collection and processing, coupled with an asynchronous schedule of publication, the truth of EAGDP is difficult to determine quickly. Nowcasting methodology allows for a prediction of a complex system such as GDP to be made, despite a revision schedule which suggests the true figure will not be determined till much after the fact. It is this problem which is of particular interest to us, more specifically for the European Area Gross Domestic Product (EAGDP) which we focus on here.

Nowcasting EAGDP is a complex and large task to undertake, given the actual values are created from an aggregation of a number of countries (of which their GDP will be nowcasted) which make up its final figure. However there are many models in existance which attempt to provide a guidance of the value, such as the well known Eurostat Flash Estimate (Eurostat, 2016), a weighted aggregate of the

member countries estimates upon their own GDP. Many other methodology exists such as: Grasmann and Keereman (2001) using linear regression with a moving average error term; Marcellino et al. (2003) combining univariate autoregressions of each country; Rünstler and Sédillot (2003) and Baffigi et al. (2004) with bridge equations; Banerjee et al. (2005) employing leading indicators in a number of models including a dynamic factor model; Diron (2008) explores the use of vintages (non-revised data) with bridge equations; Angelini et al. (2011) combining bridge equations with factor models, to name but a few.

As can be seen by the brief highlight of literature, there are many approaches that can be taken to model such a complex system, using many different datasets and components. Given the changing nature of EAGDP, such that there are recessions and booms in the business cycle, and that certain industries become more prevalent through time, it can be expected that measuring any such value must account for changing conditions. Much methodology attempts to do this by using a large and vast dataset covering many areas from Industrial Production to consumer expectations. However, this still relies upon a wealth of prior knowledge of series which are currently relevant to the estimation being made, which may not always be the case or may change on too frequent of a basis.

Here we approach this modelling problem from a naive perspective. Rather than determine the necessary relationships at any one time point, we employ a stepwise algorithm which selects the necessary components based upon a predictive measure fitted by a Dynamic Factor Model. These models are a common approach in Nowcasting, being widely used across the literature worldwide. This is such that we do not model directly against the datasets we are given, but instead upon hidden factors shared amongst coincident series. This provides strength to our particular problem as the dataset, which we describe further in Section 6.2.1, we use is based upon quickly produced figures from surveys and exchange rate indicators, which may not best directly represent EAGDP, however their shared behaviour between them may be more indicative. This therefore aims to account

for the naive approach we take alongside the varied publication scehdule of data.

Our work is structured as follows. We begin by exploring our motivation of EAGDP in Section 6.2 reviewing the data used and the previous methodology employed. Section 6.3 introduces the Dynamic Factor Model, used to measure the different movements and review the literature of these models and their usage in different fields. Further, we detail the algorithm constructed around these models to estimate the EAGDP data naively. We experiment with the algorithm through an empirical study in Section 6.4. Following this we then generalise our approach and explore additional options in our selection process in Section 6.5. Finally we conclude with closing remarks in Section 6.6.

6.2 European Area GDP

Our original motivation behind the work conducted in this Chapter comes from Nowcasting European Area Gross Domestic Product (EAGDP), and improving upon the Nowcasting generated by competing models. It is this problem we focus upon here in this section. Firstly we outline the data to be used within Section 6.2.1. Following this we outline the previous models used in this problem in Section 6.2.2.

6.2.1 Data

Within the context of nowcasting, the availability of information is a key consideration when it comes to selection of the correct variables. In our investigation we work with data which is published quickly due to its nature. These variables are assumed to have a relevance to our problem, but no formal testing or explorative analysis has been conducted and we leave this to our work that follows. Further, due to there being no investigation into these variables in this context, we are given no information *a priori* regarding the structure of any model we may wish to fit.

Our work was motivated from a problem presented by Eurostat, and it is from them that we acquire our data, listed in Table 6.1. Their original implementation used proprietary data whereas the new implementation is based on open source data. We apply transformations to each variable in an attempt to normalise them for fitting a DFM. Each of these datasets has been classified as 'soft' or 'hard' as a representation of their qualitative nature, note that we do not use this classification within the modelling. For example, the Euro/Dollar Exchange Rate is non-subjective and is thus classified 'hard' whereas industrial production expectations is collected via a survey of companies where they predict their outputs, which can be seen as subjective and thus classified as 'soft'. However, this data is particularly useful as they can be published quickly and thus aid in nowcasting as whilst we may suspect they are not the most indicative of EAGDP, they may still provide insight. This data is given over the range of February 1985 - August 2017.

Our variable of interest for modelling is the quarterly European Area Gross Domestic Product. This quantity is revised throughout the quarter as more data becomes available, but it remains quarterly. Therefore, as our indicator variables are on a monthly frequency we must employ a form of temporal disaggregation to calculate the data on a monthly frequency. We use the Litterman (Litterman, 1983) method as is commonly employed within Eurostat, using a sum disaggregation such that the sum of three months would equal a quarter. Note that as is standard within the industry, we do not estimate the actual figure of EAGDP, but instead its growth rate.

As will be seen in Section 6.3.4 we use both the monthly and quarterly versions of EAGDP separately to see which provides the most accurate nowcasts when measuring our estimates.

ID	Data Name	Data Ref	Soft/Hard
1	Euro/Dollar real exchange rate		Hard
2	Industrial (manufacturing) production expectations	INDU.EA.TOT.5.BS.M	Soft
3	Economic sentiment indicator	EA.ESI	Soft
4	Households opinion on economic situation over next 12 months	CONS.EA.TOT.4.BS.M	Soft
5	Households opinion (consumer survey)	CONS.EA.TOT.9.BS.M	Soft
6	Construction confidence indicators	BUIL.EA.TOT.COF.BS.M	Soft
7	Industrial (manufacturing) export order book positions	INDU.EA.TOT.3.BS.M	Soft
8	The industrial (manufacturing) order book positions	INDU.EA.TOT.2.BS.M	Soft
9	Retail trade current business situation	RETA.EA.TOT.1.BS.M	Soft

Table 6.1: Data sources used within the application.

6.2.2 Previous Models

As part of our investigations we compare the estimates we create for EAGDP against previous methodologies. We do not run these models but instead compare the values we would have created historically to those generated by these previous models. An earlier model is not detailed beyond the estimates created, however estimates from a Vector Autoregressive Model is given. This is such that the full dataset is used in the model VAR(β) with β lags:

$$\{\mathbf{Y},\mathbf{X}\}_t = \boldsymbol{\mu} + \boldsymbol{\alpha}_1\{\mathbf{Y},\mathbf{X}\}_{t-1} + \cdots + \boldsymbol{\alpha}_\beta\{\mathbf{Y},\mathbf{X}\}_{t-\beta} + \mathbf{e}_t,$$

$$\begin{bmatrix} Y_t \\ X_{1,t} \\ \vdots \\ X_{9,t} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_1 \end{bmatrix} + \begin{bmatrix} \alpha_{1,1,1} & \dots & \alpha_{1,1,9} \\ \alpha_{1,2,1} & \dots & \alpha_{1,2,9} \\ \vdots & \ddots & \vdots \\ \alpha_{1,9,1} & \dots & \alpha_{1,9,9} \end{bmatrix} \begin{bmatrix} Y_{t-1} \\ X_{1,t-1} \\ \vdots \\ X_{9,t-1} \end{bmatrix} + \dots$$

$$\dots + \begin{bmatrix} \alpha_{\beta,1,1} & \dots & \alpha_{\beta,1,9} \\ \alpha_{\beta,2,1} & \dots & \alpha_{\beta,2,9} \\ \vdots & \ddots & \vdots \\ \alpha_{\beta,9,1} & \dots & \alpha_{\beta,9,9} \end{bmatrix} \begin{bmatrix} Y_{t-\beta} \\ X_{1,t-\beta} \\ \vdots \\ X_{9,t-\beta} \end{bmatrix} + \begin{bmatrix} e_{0,t} \\ e_{1,t} \\ \vdots \\ e_{9,t} \end{bmatrix},$$

where α are the estimated coefficients, $\mathbf{e}_t$ is the estimation error which satisfies $\mathbb{E}(\mathbf{e}_t) = 0$, $\mathbb{E}(\mathbf{e}_t\mathbf{e}_{t-k}^{'})(\forall k \in \mathbb{Z} : k \neq 0)$ and $\mathbb{E}(\mathbf{e}_t\mathbf{e}_t^{'}) = \Omega$ where Ω is a covariance matrix.

6.3 Dynamic Factor Models

Here we explore the model which we focus on throughout this Chapter. Firstly we introduce the model formally and review the literature in Section 6.3.1. Following this we detail the considerations that need to be made regarding the structure of a Dynamic Factor Model in Section 6.3.2. Next we propose new methodology for detection of the structure of EAGDP and detail the steps taken in the constructed algorithm in Section 6.3.3. Following this in Section 6.3.4 we present the results

of our algorithm and the structures it finds. This is followed by a discussion of these results and their interpretation in Section 6.3.5. The testing of this approach through simulation is then reserved for Section 6.4.

6.3.1 Overview

Highly prevalent within the Nowcasting literature, the general Dynamic Factor Model (DFM) looks to determine and use hidden trends between correlated datasets to make inference on a series of interest. This representation is particularly coveted in econometric literature through Real Business Cycle theory and models such as the Dynamic Stochastic General Equilibrium (Fernandez-Villaverde, 2009) which look for 'common forces' amongst a plethora of available information. This may come from a very large and populated dataset in both time and dimension. The general Dynamic Factor Model can be represented as

$$Z_t = \lambda(L)F_t + \epsilon_t \qquad \epsilon_t \sim N(0, V) \tag{6.1}$$

$$F_t = \psi(L)F_{t-1} + \upsilon_t \qquad \upsilon_t \sim N(0, W). \tag{6.2}$$

where all data is presented as the $[(m+1)$ x $1]$ column matrix Z_t, which we model using the 'hidden' factor $[p$ x $(m+1)]$ matrix F_t, where $p < (m+1)$. Both $\lambda(L)$ and $\psi(L)$ are polynomial matrices, of dimensions $[(m+1)$ x $p]$ and $[p$ x $p]$ respectively, acting upon the lag operator L. The idiosyncratic error ϵ_t and υ_t are assumed to be uncorrelated at all lags such that $\mathbb{E}(\epsilon_t \upsilon_{t+\tau}^T) = 0 \ \forall \tau \in \mathbb{Z}$ and in the exact Dynamic Factor Model it is assumed that $\mathbb{E}(\epsilon_{i,t}\upsilon_{j,t+\tau}) = 0 \ \forall \tau \in \mathbb{Z} \iff i \neq j$. Finally both Z_t and F_t are assumed to be stationary. Given the large amount of parameters which may need to be estimated at fitting, often the matrix $\lambda(L)$ is reduced to A by excluding lags, similarly for $\psi(L)$. However, as noted in Stock and Watson (2009), this is a rearrangement of the Dynamic Factor Model as the

factors are still related to their past values. This is represented by

$$Z_t = AF_t + \epsilon_t \qquad \epsilon_t \sim N(0, V) \tag{6.3}$$

$$F_t = F_{t-1} + \upsilon_t \qquad \upsilon_t \sim N(0, W). \tag{6.4}$$

such that there are no longer any lag polynomials to estimate, but instead a $[(m+1)$ x $p]$ factor loadings matrix. With the reduction in free parameters these models lend themselves particularly to those with a smaller dataset. It is this particular model which of interest to our investigations. It should be noted that the re-arrangement of F_t causes non-stationarity. However this model is noted as an extension to the Dynamic Factor Model in Stock and Watson (2009).

Dynamic Factor Models have been used in forecasting since Geweke (1977) proposed them as an extension of a cross-sectional data model. In the same year Sargent and Sims (1977) applied the methods to macroeconomic variables in the US. Stock and Watson (1988) then model the state of the economy using the dynamic factor models to create an economic indicator. More recently forecasts on GDP specifically have been done for the US (Banerjee and Marcellino, 2006) and the Eurozone (Angelini et al., 2011).

Outside of economics, dynamic factor models have been used in many forecasting applications. A variety of information at different scales and timings were used to model credit risk in Creal et al. (2014). Within psychology, dynamic factor models have been used to represent changes in behaviour from one timepoint to another and the differences between individuals in Ram et al. (2013). Air quality assessments have also been conducted using dynamic factor models to reduce the dimensionality of data within Calder (2007). Investigation has also been done into the major climate event El Nino in Li et al.(2019+).

Extensive work has been conducted on adapting the Dynamic Factor Model into a number of extensions. Given the changing conditions that can be expected within such a complex system as GDP, much work has been done to explore adaptation to the DFM framework. Diebold and Rudebusch (1994) looked to use a

Markov Regime Switching approach in collaboration with a DFM to model a system separately dependent upon the current phase of an economy. The idea of a sudden change in modelling was continued in Stock and Watson (2009) where investigations which conducted into the effect of a sudden change in the factor relationships. A slowly varying approach to these changes was taken in Negro and Otrok (2008), studying the changes which occurred in the post Bretton-Woods period. Beyond structural considerations, work has been done in the use of the estimated factors as tools in other modelling approaches such as the Vector Autoregression Model (VAR) to create a factor augmented version (FAVAR) in Bernanke et al. (2005) and a Factor Augmented Error Correction Model in Banerjee and Marcellino (2009).

Fitting and Identifiability

There are a number of methods available which can be used to determine the parameters of a Dynamic Factor Model. These can be split into a number of methodological approaches. Firstly is that of cross-sectional averaging, such that cross-section averages of a dataset are regressed against augmented variables, more concisely explained in Pesaran (2006). Secondly principal component analysis can be used to estimate the unobservable factors F_t as described in Forni et al. (2000). A further methodology is that of the Kalman Filter using a Maximum Likelihood Estimator (Watson and Engle, 1983) by rewriting Equations (6.1) and (6.2) in State Space Form, of which Equations (6.3) and (6.4) are already. There is also further work in using both Principal Components Analysis and the Kalman Filter together (Giannone et al., 2008) and Bayesian Analysis using Monte Carlo Markov Chains (Otrok and Whiteman, 1996).

Our work proceeds through usage of the Kalman Filter, given the relatively small number of series we will be investigating and the optimal estimates we can gain under our assumptions. Using available methods in the `dlm` package (Petris, 2010) within `R` and modifications to represent the DFM of Equations (6.3) and

(6.4) our work fits a DFM giving fitted values as the one-step-ahead forecasts. However, as there are numerous versions of one factor model, we must parameterise it according to the alternative parameterisation mentioned within Petris (2010). This is such that A is an upper triangular matrix with diagonal entries of 1, and coefficients $a_{i,j} \in \mathbb{R}$. Further, W is a diagonal matrix with positive coefficients $w_i \in \mathbb{R}_+$ as such

$$A = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ a_{2,1} & 1 & 0 & \cdots & 0 \\ a_{3,1} & a_{3,2} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{p,1} & a_{p,2} & a_{p,3} & \cdots & 1 \\ a_{p+1,1} & a_{p+1,2} & a_{p+1,3} & \cdots & a_{p+1,p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m+1,1} & a_{m+1,2} & a_{m+1,4} & \cdots & a_{m+1,p} \end{bmatrix}, \qquad W = \begin{bmatrix} w_1 & 0 & 0 & \cdots & 0 \\ 0 & w_2 & 0 & \cdots & 0 \\ 0 & 0 & w_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & w_p \end{bmatrix}. \tag{6.5}$$

6.3.2 Structural Considerations

Given the identifiability of our factor loadings in Equation (6.5) there are two main considerations that can be seen, the selections of the parameters p and m. These are the number of factors to estimate, and the amount of series (excluding the series of interest) to model concurrently. Further, there is also the consideration that such components may change throughout time, given the dynamic conditions of the systems we study, as mentioned previously. These considerations have a large impact upon the resultant modelling and thus forecasting, and as such have been a point of interest for research.

Choice of the amount of dynamic factors has received particular attention, here we mention a few. Often used is *a prioiri* information, provided perhaps by experts in the field or a continuation of modelling performed previously. This may be

consulted alongside a Scree Plot (Cattell, 1966) often used in Principal Component Analysis for the selection of the optimal amount of components, done through a visual inspection of the eigenvalues of the dataset against their corresponding index, and picking the 'elbow' of the plot. This procedure can be formalised into a test as in the work of Onatski (2009), which considered a range of values for p in high dimensions. Further, there are a number of information criterion given in Bai and Ng (2002) which combine information on the scale and size of the dataset with a mean squared error approach. A number of other methodologies have been studied, of which there is a strong review to be found in Hoyle and Duvall (2004) which looked at this issue regarding a general factor analysis problem.

The selection of datasets however, has received less attention. Often it is the case that large datasets are used and the effects of potential weakly correlated information is not explored due to the gains of a larger dataset. However, work completed in Boivin and Ng (2006) showed that reduction from a large dataset to a dataset of key variables can enhance estimation of the factors F_t. Further, in a meta analysis Eickmeier and Ziegler (2008) find a reduction in Root Mean Squared Error when studies have specified they performed a preselection of variables. In particular Bai and Ng (2008) use Least Angle Regression techniques such as Elastic Net and LASSO (of which the soft-thresholding rules are special cases) to reduce their dataset to targeted predictors, by measuring their predictive capabilities. Further Rünstler (2016) uses the marginal predictive gains of the addition of a dataset as a measure of the effectiveness of a variable to determine dataset size. In specific relation to EAGDP, Girardi et al. (2016) find that softer data (surverys) are often of use within a certain period of a nowcast in comparison to other data. For a review of papers which consider the variable problem, the meta analysis conducted in Eickmeier and Ziegler (2008) is recommended.

Considering these aspects throughout time, and how they may change is of particular interest for the factor loadings matrix A. Negro and Otrok (2008) highlighted potential changes in the structure of business cycles over a changing

economy. Determining and accounting for these changes was considered highly relevant to the performance of a factor model in Banerjee et al. (2008) where a simulation study investigated the effect of factor changes and how the amount of factors impacted upon forecasts. Further investigation is conducted within Su and Wang (2017) where time varying factors were fitted using local Principal Component Analysis and an information criterion to determine the number of factors. Considering further changes to these factors, Pelger and Xiong (2018) estimate the model under two switching states to represent differing periods within an economy, allowing the factor loadings and number of factors to potentially change. Beyond this Cheung (2018) consider the case of changing the number of factors alongside the factor loadings using a sieve estimator.

It is clear that there are several factors at play here in our model; the number of factors, the number of variables, parameter estimation technique, should time varying factor structure or variable inclusion be considered. The current literature selects from these problems and considers it in isolation to the rest. It is likely this affects the final model form and thus interpretation and forecasts. In contrast, we create an approach that acknowledges these interdependencies and simultaneously selects the number of factors and variables.

6.3.3 Methodology For EAGDP

As discussed, the use of the Dynamic Factor Model for nowcasting comes with a number of considerations which must be taken into account. Often the selection of the best fitting Dynamic Factor Model is made from a number of approaches, but to the best of our knowledge no one has considered the estimation of the variable and factor selection simultaneously, and it is this approach we take.

Before continuing it is important to clarify the notation which is used going forward. We now split the data Z_t shown in Equation (6.1) into two components. The data which we are modelling is termed the 'reference series' Y_t (with estimated values $\hat{Y}_t$) and those which we believe are related are termed the 'explanatory series'

$X_{t,j}$ (with estimated values $\hat{X}_{t,j}$). Note that we index the exploratory series by $j = 1, \ldots, m$, and both series time points by $t = 1, \ldots, n$. The number of factors we model this data with is p. When referring to n^* recent points, we are referring the data points at time points $n - n^*, n - n^* + 1, \ldots, n$.

Exploration of all potential models would require (for a given amount of factors p) that $2^m - \sum_{i=1}^{p} \binom{m}{i}$ models to be computed. This becomes a very computationally expensive process, particularly when fitting with a large amount of variables or factors. To mitigate this, we apply a forward selection algorithm which reviews each nearby model (in terms of variables or factors) using a Mean Squared Error (MSE) estimated on the most recent n^* points,

$$m\left(\mathbf{Y}, \hat{\mathbf{Y}}, n, n^*\right) = \sum_{t=n-n*}^{n} \left(Y_t - \hat{Y}_t\right)^2. \tag{6.6}$$

similar to that of Rünstler (2016). By taking a forward rather than backward approach, we look to be prudent in our inclusion of variables/factors. Further, we use n^* to restrict our fitting to the most recent points in our estimation as for nowcasting we are not necessarily looking to explain behaviour which has happened in the past, which may have been revised a number of times, but we are more concerned with recent data which is still due revision or has been revised only recently.

We fit these models through the use of the state space representation given in Equations (6.3) and (6.4) where we optimise over the unknown components A, V and W to fully specify the model. Following this we apply the Kalman Filter (Kalman, 1960) to estimate the unobserved factors, implemented through the `dlm` R package. This then means that we can easily generate one-step ahead forecasts for all time points due to the nature of the Kalman Filter, which we use in our MSE calculations. Further details of this process can be found in Beck (1989). As part of this fitting process, we pre-process the series through a mean and variance changepoint detection algorithm from the `changepoint` package in R (Killick and Eckley, 2014) to standardise each series to mean 0 and standard deviation 1, and

removing sudden structural changes which may have occurred.

Combining these elements of methodology together leads to our proposed work, given as a pseudo-algorithm in Algorithm 3. Here we initialise with a reference series Y_t and potential indicator series $X_{t,j}$, and begin by fitting each separately with the reference series and a single dynamic factor. Following this, we then attempt to continue adding additional variables until we find no decrease in MSE. At this point we shift our attention to the number of dynamic factors and attempt to increase them alongside adding an additional variable. If this is successful in reducing the MSE, we continue only searching for more variables. If it is unsuccessful, we consider the algorithm to have converged upon a structure.

6.3.4 Results of Application

We explore the applicability of our algorithm to the EAGDP task by varying two dimensions. Firstly, we consider the calculation of our model fit (by recent MSE) on the monthly or quarterly series. This is such that we can predict the growth rate of the EAGDP after our model is fit, then choose to compare it directly to the disaggregated series or instead transform it back to a quarterly series and compare it against the actual values (though they are still under revision). Comparing on a monthly level allows for a direct comparison without a loss in information during transformation, however only the quarterly values are actual observations. Our second dimension is time, such that we restrict the data to end a certain number of months earlier. This allows us to monitor the change in the structure selected by our algorithm through time. We vary this by up to 3 and a half years, or 42 months.

Results are presented on two main areas. Firstly we review the models chosen by the algorithm in their primary endpoint, nowcasting efficiency. This is done by a comparison of estimated growth rates against competing models, the actual observed and the Eurostat $t+30$ flash estimate produced 30 days after the quarter ends. Similarly a comparison of the squared errors is presented for clarity. Our

Data: Y_t, Euro Area Gross Domestic Product Monthly for $t = 1, \dots n$;
Data: $X_{t,j}$ For $j = 1, \dots, m$ indicator series over timepoints $t = 1, \dots n$;
input : n^*, The amount of recent points to consider in MSE calculation ;
$\hat{\delta} = \infty - \epsilon, \ \epsilon > 0$ The minimum MSE of the current stage ;
$\delta = \infty$ The minimum MSE of fitted models in all previous stages ;
$\delta_i^* = \infty \ \forall i \in \mathbb{N}$ The current MSE of fitted models ;
$\hat{p} = 1$ The current amount of dynamic factors ;
while $\hat{\delta} \leq \delta$ *and* $m > 0$ **do**
  for j *in* $1 : m$ **do**
    Fit DFM using data $\{\mathbf{Y}, \mathbf{X}_j\}$ with $\hat{p}$ dynamic factors. ;
    Calculate MSE δ_i^* over n^* recent points (Equation (6.6)) ;
  end
  Let $\hat{\delta} = \min(\{\delta_j^* : j = 1, \dots, m\})$;
  if $\hat{\delta} < \delta$ **then**
    Reset $\delta_j^* = \infty \ \forall j \in \mathbb{N}$;
    Re-assign $\delta = \hat{\delta}$;
    Expand the forecasting dataset: $\{\mathbf{Y}\} = \{\mathbf{Y}, \{\mathbf{X}_{j:\delta_j^* = \min(\boldsymbol{\delta}^*)}\}\}$;
    Remove that variable from the indicators dataset: $\{\mathbf{X}\} = \{\mathbf{X} \backslash \{\mathbf{X}_{j:\delta_j^* = \min(\boldsymbol{\delta}^*)}\}\}$;
  else
    if $\hat{p} \leq m$ **then**
      Let $\hat{p} = \hat{p} + 1$;
      for j *in* $1 : m$ **do**
        Fit DFM using data $\{\mathbf{Y}, \mathbf{X}_j\}$ with $\hat{p}$ dynamic factors ;
        Calculate MSE δ_j^* over n^* recent points (Equation (6.6)) ;
      end
      Let $\hat{\delta} = \min(\{\delta_j^* : j = 1, \dots, m\})$;
      if $\hat{\delta} < \delta$ **then**
        Re-assign $\delta = \hat{\delta}$;
        Reset $\delta_j^* = \infty \ \forall j \in \mathbb{N}$;
        Expand the forecasting dataset: $\{\mathbf{Y}\} = \{\mathbf{Y}, \mathbf{X}_{j:\delta_j^* = \min(\boldsymbol{\delta}^*)}\}$;
        Remove that variable from the indicators dataset: $\{\mathbf{X}\} = \{\mathbf{X} \backslash \{\mathbf{X}_{j:\delta_j^* = \min(\boldsymbol{\delta}^*)}\}\}$;
      end
    end
    Recalculate the amount of explanatory series still unselected $m = |\mathbf{X}|$. ;
  end
end
output: Best dataset $\{\mathbf{Y}\}$ and best number of dynamic factors p. ;

Algorithm 3: Proposed Model Selection: Variable Selection and Number of Dynamic Factors.

second area of interest is the chosen structure of the models from our algorithm. We explore the datasets chosen at each of the given time points and the amount of factors chosen.

Forecast Efficiency

To begin our comparison of results, we look at Quarterly GDP growth against our estimated growth rates and those given by competing models. As we estimate a monthly growth rate we have to consider how we transform this into a comparable quarterly rate given which month we are in. For comparison we present the Eurostat $t + 30$ estimate, which we extend two months previous to their generation for comparison. Thus when we are in the first month of a new quarter, we use the fitted monthly growth rates from the prior three months to create a quarterly growth rate of the previous quarter. However, if we are in the second month of a quarter, we then look to nowcast the current quarter we are in, thus we require a one step ahead forecast alongside the fitted values. For the third month, we use the current and previous two values of the fitted series to compute a nowcast for the current quarter's GDP. Table 6.3 shows how the months and the predicted quarter line up more succinctly.

A comparison of these calculated quarterly growth rates is given in Figure 6.1a alongside competing models. Note that the graph separates out the estimations made by a previous model and those under consideration. These are the estimates from a newly proposed VAR model implemented within Matlab with the full data set (unknown for the previous model). Additionally we include the Eurostat Flash Estimate of GDP for that quarter. Note the discontinuous nature of the actual growth rate and the flash estimate is such that we are comparing the actual quarter at month T with that which would be predicted at that time.

It can be seen that the DFM with quarterly variable selection is not as smooth as the monthly variable selection to begin, but smooths out throughout the period. Overall it appears to follow closely to the truth and the flash estimate. Most

notably in comparison to the currently implemented model the proposed model predicts a higher level, following more closely with the flash estimates and truth. The fit generally follows consistently the growth rate over the period 2014-2017 except for the jump and correction in early 2015. Numerically, in comparison to the VAR model estimates given towards the end, the DFM reduces the MSE over the period by 67.45%.

When using a monthly MSE the predictions can be more erratic on a month by month basis. However this allows for quick adaptation to the movements within the true growth rate such as in 2016 and the placement between the truth and the flash estimate in late 2014. However, as in the quarterly case, the level of the series follows that of the truth without deviating too much and in the very recent past draws closest to the truth. Here also there is a reduction in the MSE in comparison to VAR by 82.23%.

Figure 6.1b depicts the point wise squared error (SE) of the predictions from Figure 6.1a. The large departures prior to 2015 can be seen to be an overestimate of the jump to come in the truth, before recalculating a much lower estimate in the following monthly as the Eurostat Flash Estimate had. Beyond 2015 the jumps are much less erratic for both comparisons and often below that of the previous contractor until an underprediction of a large jump, which the previous model consistently overestimated. However, in comparison to the newly proposed VAR model, we see large reductions in error for both the monthly and quarterly variable selection procedures.

Choice of Structure

As detailed in Algorithm 3, a selection process was put in place such that only the most favourable data is used and most predictive factor structure is selected in the prediction for time period which we restrict to. Here we present which variables were being used throughout the previous 42 month period and the amount of factors that were selected. The results in Figure 6.2a show variable selection

Variable	1	2	3	4	5	6	7	8	9
Quarterly	0	27.91	0	65.12	100	32.56	6.98	53.49	16.28
Monthly	46.51	34.88	11.63	25.58	41.86	48.84	37.21	58.14	20.93

Table 6.2: Inclusion rate (%) of variable within each comparison methodology.

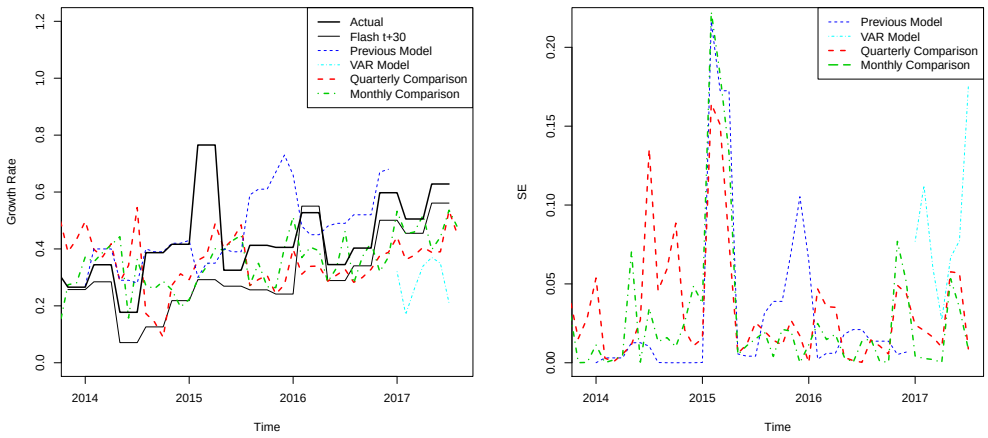
Month	1	2	3	4	5	6	7	8	9	10	11	12
Actual Quarter	1			2			3			4		
Predicted Quarter	4	1			2			3			4	

Table 6.3: How each calendar month corresponds to the actual quarter and the quarter predicted at that time.

throughout the period, where as Figure 6.2c shows the amount of factors selected. The index on the y axis corresponds to the ID within Table 6.1, but also to the amount of variables selected in total. Further we show the rate of inclusion as a percentage of each variable in Table 6.2.

Reviewing the variable selection it can be seen that this is a highly variable process which selects a diferent number and selection at each time point. Most prevalent is the complete inclusion of Variable 5 for the quarterly predictions. Further, there are some continuous streaks of variable usage, such as variable 8 during 2014 and partly 2015. An interesting trend is the inclusion of variable 1 towards the end of 2015 within the monthly comparison, and variable 4 similarly for the quarterly. However the other variables are only chosen for short periods of time before being removed. There is evidence of indicative behaviour here, but it is not necessarily consistent between the two comparisons, potentially reflecting the quick changes that can be expected in a monthly comparison rather than a quarterly. Certainly there appears to be a change in behaviour around the end of 2015.

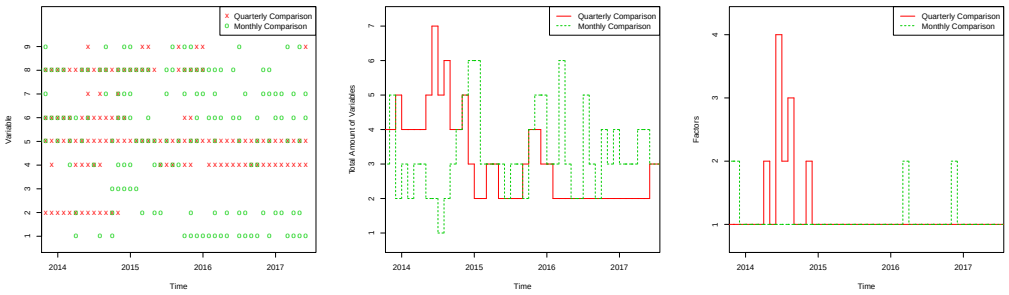
Figure 6.2c shows how the behaviour of the comparisons differ greatly. Prior to 2015 it would appear that the quarterly comparison was changing the amount of factors it was using to predict very frequently, before returning to 1 and holding consistent as 2015 began. However, the monthly comparison shows little change in the amount of factors, choosing only to depart from 1 sporadically in 2016.



(a) Predictions for each quarter with aligned prediction period.

(b) Squared Error (SE) in prediction of quarterly growth rate.

Figure 6.1: Rolling forecasts and errors for quarterly Euro-area GDP. $n^* = 24$.



(a) Variables selected over the 42 months.

(b) Total variables selected throughout period.

(c) Amount of dynamic factors estimated.

Figure 6.2: Structure of dataset selected throughout the 42 months.

6.3.5 Discussion of Application

The results shown in Section 6.3.4 are both promising and puzzling. We can see that gains in predictions of the EAGDP can be successful through this methodology, with reduced errors in particularly tumultuous times, but structure of the models chosen is not consistent. There are a number of variables which appear to be consistent in usage, but only through a particular time period or varying by comparison metric. The differing variability of the factors chosen indicates that aggregation of a series can potentially hide changing behaviour through time.

Regardless of the changing structure, the gains in forecasting are particularly strong. Both Monthly and Quarterly comparisons follow closely to the truth and

flash estimate, particularly so throughout 2016 and 2017 when growth becomes less erratic. Our comparison to a previous model shows that whilst we are not as stable as we would like in 2014/2015, we are capable of predicting closer to the truth than the flash estimate. Changes in behaviour of variables and, in the case of a quarterly comparison, appear to aid this.

However, whilst we can compare to a truth on our forecasting efficiency, there is no such truth for structure. Given the nature of the Dynamic Factor Model, the states we wish to know will always be unobservable and we cannot know for certain if we are close to matching the true structure at any time within our application. Through empirical observations however, we can measure the effectiveness of such methodology in Section 6.4.

6.4 Structural Recovery Simulations

Following on from Section 6.3.5 we look to determine the capability of our algorithm not in forecasting efficiency, but in structural selection. Whilst we cannot know the truth in the case of our application to European Area Gross Domestic Product, we can instead simulate from a known truth and compare the results of our algorithm to our expected values. It is this approach that we take to measure the effectiveness of our algorithm in three areas: the selection of correlated variables (termed dependent variables here); the removal of variables uncorrelated to the reference series (independent series); and the selection of the true amount of factors.

We choose to simulate from a number of dependent structures using a small amount of series with a single independent series chosen from Table 6.4. To do this we define a number of rules we wish to satisfy such that we are equally monitoring the structural recovery in each case. These rules are such that

1. We have twice as many variables as we do factors,

$$m - 1 = 2p; \tag{6.7}$$

2. Each series contributes to the overall system equally,

$$\sum_{i=1}^{p} |a_{i,j}| = \frac{p}{m}; \tag{6.8}$$

3. Each factor is contributed to equally,

$$\sum_{j=1}^{m} |a_{i,j}| = p. \tag{6.9}$$

We define these dependency structures following Equations (6.3) and (6.4). We keep the variances of the simulated data as unit diagonal matrices such that $V = \mathbb{I}_m$ and $W = \mathbb{I}_p$ and define further A, the factor loadings matrix as

$$\text{Model 1:} \quad p = 1, \quad m = 3, \quad A = \begin{bmatrix} 1 & 0.5 & -0.5 \end{bmatrix}^T,$$

$$\text{Model 2:} \quad p = 2, \quad m = 5, \quad A = \begin{bmatrix} 1 & 0.8 & -0.6 & 0.4 & -0.2 \\ 1 & 0.2 & 0.4 & 0.6 & 0.8 \end{bmatrix}^T, \tag{6.10}$$

$$\text{Model 3:} \quad p = 2, \quad m = 5, \quad A = \begin{bmatrix} 1 & 0.7 & -0.1 & 0.9 & -0.3 \\ 1 & 0.3 & 0.9 & 0.1 & 0.7 \end{bmatrix}^T, \tag{6.11}$$

$$\text{Model 4:} \quad p = 2, \quad m = 5, \quad A = \begin{bmatrix} 1 & 0.66 & 0.1 & 0.2 & -0.285 \\ 1 & 0.33 & -0.4 & 0.3 & 0.215 \end{bmatrix}^T. \tag{6.12}$$

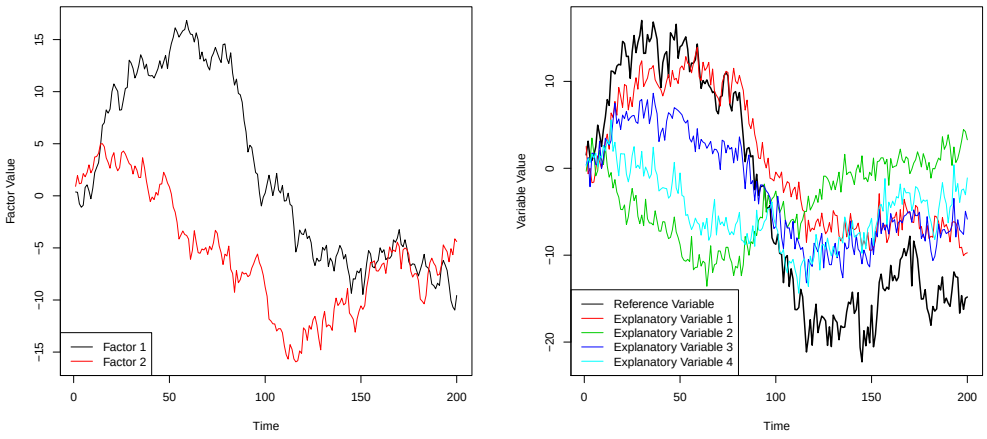
Series	AR	MA
1	0.7, 0.2	
2		0.7, 0.2
3	0.8	0.7
4		

Table 6.4: Independent models considered in the simulation study.

Note that Model 4 has been constructed such that it applies to all rules but Equation (6.8). This is such that the second series has twice as much contribution

to the overall system than other series. We can then monitor if this affects the amount of times the series is selected by the algorithm.

An example plot of a series generated from Model 3 and 4 (Equations (6.11) and (6.12)) are given in Figures 6.3 and 6.4. The dependence of each factor within each series can be seen by how each series follows similar patterns but take different turns, such as how Explanatory Variables 1 and 3 more closely followed factor 1 than 2 in Figure 6.3b due to their higher depedendence. More so the relationship between the variables can be seen more clearly in Figure 6.4b as Explanatory Variable 1 mimics the similar pattern of movement (to a lesser magnitude) than the other explanatory variables. As such we hope to detect this type of variable more frequently in our investigations.

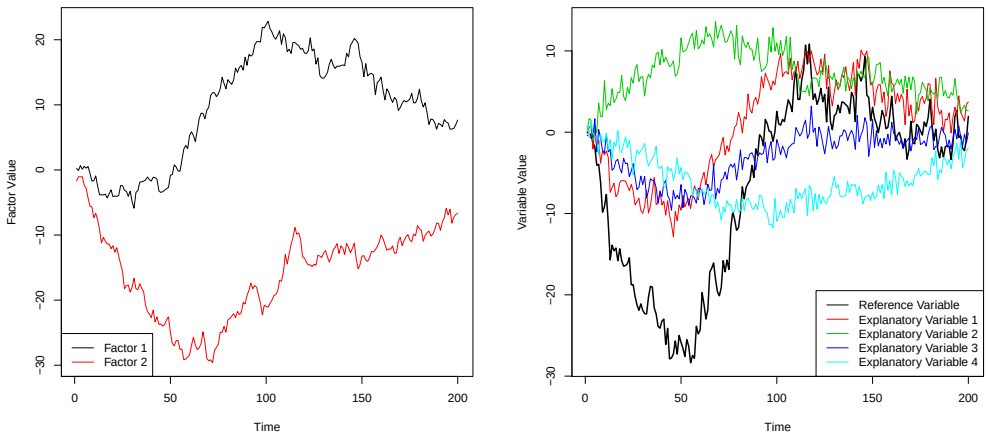


(a) Simulated factors. (b) Simulated variables.

Figure 6.3: Example simulated factors and variables from Model 3 (Equation (6.11)).

6.4.1 Results of Simulations

The results of these empirical simulations are displayed within Table 6.5 with further details about variable and factor selection in Tables 6.6a, 6.6b and 6.6c. Table 6.5 shows the selection of the correct structure in three different areas; removal of the independent series, inclusion of all dependent series, and inclusion of the correct amount of factors. Tables 6.6a, 6.6b and 6.6c show the number of dependent



(a) Simulated factors. (b) Simulated variables.

Figure 6.4: Example simulated factors and variables from Model 4 (Equation (6.12)).

series selected per dependent model, the number of factors selected per dependent model and the dependent series selected per dependent model respectively.

Reviewing the correct structure selection, it can be seen that certain structural considerations are more difficult to converge upon than others. It would appear that the algorithm is consistently able to remove the independent series approximately 80% of the time on average regardless of the dependent or independent models. There appears to be particular strength under dependent model 4 with a 93% removal rate of independent model 1. Whilst some results appear to increase with sample size, this is not consistent across all model combinations. However it does not appear to be linked to any particular selection of models and may be due to simulation size constraints.

With respect to selection of the dependent series, it would appear to be very difficult to truly select the correct amount of variables. Model 1, the simplest model, holds the highest results at an average of 25%, but all other models have very low selection rates. As in the case of independent removal, whilst some inclusion rates increase with sample size, not all do. Of particular interest is that no iteration selected the correct amount of variables when independent model 4 was present. We can break this down further by reviewing Table 6.6a, noticing

Dep. Model	Indep. Model	Independent (n=) 100	200	500	Dependent (n=) 100	200	500	Factors (n=) 100	200	500
1	1	0.78	0.81	0.83	0.24	0.20	0.26	0.91	0.90	0.95
1	2	0.73	0.75	0.83	0.25	0.25	0.25	0.90	0.92	0.95
1	3	0.75	0.86	0.78	0.27	0.23	0.24	0.86	0.95	0.94
1	4	0.76	0.72	0.81	0.00	0.00	0.00	0.91	0.90	0.93
2	1	0.80	0.84	0.75	0.02	0.04	0.04	0.07	0.09	0.10
2	2	0.75	0.82	0.74	0.02	0.02	0.01	0.10	0.04	0.07
2	3	0.80	0.78	0.74	0.02	0.04	0.02	0.10	0.08	0.07
2	4	0.77	0.82	0.75	0.00	0.00	0.00	0.08	0.05	0.09
3	1	0.83	0.88	0.86	0.01	0.03	0.03	0.08	0.10	0.03
3	2	0.76	0.83	0.80	0.02	0.03	0.02	0.12	0.08	0.08
3	3	0.85	0.81	0.80	0.02	0.03	0.05	0.06	0.09	0.09
3	4	0.80	0.84	0.79	0.00	0.00	0.00	0.09	0.10	0.06
4	1	0.81	0.80	0.93	0.04	0.05	0.02	0.11	0.16	0.04
4	2	0.84	0.71	0.85	0.04	0.07	0.01	0.11	0.13	0.06
4	3	0.84	0.75	0.81	0.04	0.07	0.01	0.11	0.17	0.08
4	4	0.83	0.71	0.83	0.00	0.00	0.00	0.12	0.14	0.05

Table 6.5: Proportion of structural components successfully selected in the simulation performed. *Independent* refers to the proportion of simulations which removed the independent series from the final model. *Dependent* refers to the number of simulations that successfully selected the correct amount of dependent series in the final model. *Factors* refers to the number of simulations that successfully select the correct amount of factors in the final model. **Results closest to 1 are favourable in all cases.**

that often we only select 1 dependent variable in total. It would appear that the next option is to include only the independent variable beyond this. It is unusual to select more than 1 variable. This is further shown in Table 6.6b which shows that there appears to be a particular series of interest. This is contrary to the way we have set up our dependent models following rules in Equation (6.7), (6.8) and (6.9), particularly in the case of dependent model 4 which has been set up to prefer series 1.

Looking next to factor selection, the results are very similar to that of dependent variable inclusion. Model 1 holds the highest rates, such that it appears rare for the algorithm to select any more factors than 1. Beyond this, it is particularly difficult to select the correct amount of factors for any of the more complex models with rates as low as 3% up to 17%. As in the previous structural considerations, there does not appear to be a clear effect of increasing sample size upon this.

Dep. Model	Variables				
	0	1	2	3	4
1	0.30	0.52	0.18		
2	0.34	0.45	0.13	0.07	0.02
3	0.30	0.49	0.12	0.07	0.02
4	0.28	0.43	0.15	0.10	0.03

(a) Number of dependent variables chosen across all simulations generated from each dependent model.

Dep. Model	Dependent Series			
	1	2	3	4
1	54	66		
2	55	25	26	24
3	32	12	59	22
4	38	18	55	23

(b) Selection of each variable (%) by dependent model.

Dep. Model	Factors			
	1	2	3	4
1	0.92	0.08		
2	0.91	0.08	0.01	0.00
3	0.91	0.08	0.01	0.00
4	0.88	0.11	0.01	0.00

(c) Number of factors chosen across all simulations generated from each dependent model.

Table 6.6: Specific components chosen within simulations for each dependence structure. Blank values appear where result is not possible.

Reviewing the amount of factors in more detail in Table 6.6c it can be seen that regardless of model, a single factor is most preferred overall.

6.4.2 Discussion of Simulations

It is clear from these simulations that it is very difficult to converge upon a true structure of a Dynamic Factor Model. As there are a number of considerations to be made at any one time, it is not suprising. The relationships between each of the variables and the reference series, of which we look to model and use to our benefit, would appear to be the very difficulty.

Overall in the results it would appear that one truth holds, a simpler model is almost always chosen. Unfortunately this is often in the case of the amount of dynamic factors. It can be suspected that this will then affect variable selection as we are not seeing the true correlation of each variable to the reference series. There is particular evidence of this through the preferred dependent series in Table 6.6b. It can be seen that the series which is preferred for dependent model 2 and 3 is that which has the highest coefficient in first column (which represents the reference series relationship to the first dynamic factor) in their corresponding factor loadings matrix A in Equation (6.10) and (6.11). However, this does not hold true for dependent model 4, which should prefer Series 1, however there was small increase in preference for this series.

It is therefore appropriate to study these effects further, by varying more the parameters and structure of Algorithm 3 to attempt to remove the tendency to choose a simpler model and missing key relationships between variables.

6.5 Generalised Algorithm Approach

Given the results shown in Section 6.4 we now look to generalise our approach to further aid in the recovery of the correct structure of a fitted Dynamic Factor Model. To do this we make changes in two significant areas of Algorithm 3.

Firstly, we approach the exploration of the potential model structures differently, such that we are reviewing changes to variables selected and the amount of factors concurrently. Secondly, we extend our comparison metric into a number of different functions found within the literature. As we go into the details of these changes, the reader should note that a new pseudo-algorithm has been given for reference in Algorithm 4.

6.5.1 Exploration of Models

Before evaluating the feasibility of a fit, we must explore the space of possible fits. Such exploration must be approached in an efficient manner, as we now move through two dimensions of exploration. These are the amount of dynamic factors p, and the subset of data required for modelling $\mathbf{X}$. As a consideration towards the computing time required for either of these approaches, we look to explore the space of both structural considerations simultaneously and outline that here.

We begin with a notational setup. As before, let $\mathbf{X}_j$ be a time series for an indicator variable j such that $1 \leq j \leq m$, where m is the total amount of indicator variables provided to the algorithm for selection. Define the set $S = \{s_1, \ldots, s_m\}$ such that $s_i \in \{1, 0\}$ where each s_i is an indicator, such that $s_j = 1$ represents whether the variable is in the subset. From this we can then define a subset of indicator variables as $\hat{\mathbf{X}} = \{\mathbf{X}_j : s_j = 1\}$. Further we can define the amount of indicators in this subset as $\hat{m} = \sum_{j=1}^{m} s_j$. We can also split this set into two sets, $S^+ = \{s_j : s_j = 1\}$ and $S^- = \{s_j : s_j = 0\}$, corresponding to the sets of variable being kept and removed respectively. Now we must restrict our exploration of $\hat{p}$ such that $\hat{p} \leq \hat{m}$ at all times. Note that due to the presence of a reference series which we are fitting against, whilst we choose $\hat{m}$ indicator series, the system is fitted with $\hat{m} + 1$ total time series.

As with model exploration in linear modelling, there are also directional considerations. Firstly we could start from a full model, such that $\hat{m} = m$ and $\hat{p} = m$, where we remove variables and reduce the value of $\hat{p}$. Secondly, we could begin

$\hat{p}$	s_1	s_2	s_3	...	s_m	$\sum s_i$
1	1	0	0	...	0	1
1	0	1	0	...	0	1
1	0	0	1	...	0	1
⋮	⋮	⋮	⋮	⋱	⋮	⋮
1	0	0	0	...	1	1

(a) Beginning from a null model.

$\hat{p}$	s_1	s_2	s_3	...	s_m	$\sum s_i$
m	1	1	1	...	1	m
$m-1$	1	1	1	...	1	m
$m-2$	1	1	1	...	1	m
⋮	⋮	⋮	⋮	⋱	⋮	⋮
1	1	1	1	...	1	m
$m-1$	0	1	1	...	1	$m-1$
$m-1$	1	0	1	...	1	$m-1$
$m-1$	1	1	0	...	1	$m-1$
⋮	⋮	⋮	⋮	⋱	⋮	⋮
$m-1$	1	1	1	...	0	$m-1$

(b) Beginning from a full model.

Table 6.7: Initial set of models chosen within the algorithm.

from a null model, such that $\hat{p} = 1$ and $\hat{m} = 0$. These two directions lead to a different first step due to the constraint $\hat{p} \leq \hat{m}$ such that we must first explore reducing $\hat{p}$ in the full model, but increasing $\hat{m}$ in the null. A representation of these steps is given in Tables 6.7a and 6.7b. To denote this difference we set a variable, $\eta = 1$, for forward selection; and similarly $\eta = 0$ for backward selection.

Following on from this, we explore three possible options for models concurrently. Firstly, we explore the current subset of variables across all possible values of $\hat{p}$. Secondly, we consider the current subset with a new variable and (where possible) explore the values of $\hat{p}$ as the current value, one below and one above. Similarly we explore the same values of p whilst removing a variable also, shown in Table 6.8.

6.5.2 Scoring Functions

Once models have been proposed within the algorithm, a determination of the most optimal (at that point) must be made, before moving onto the next iteration of the algorithm. As mentioned previously in Section 6.3, there have been investigations into measures of a dynamic factor models performance in terms of variable selection and selection of the amount of dynamic factors. The first measures used mean squared error (MSE), whilst the second used information criterion. It is these

$\hat{p}$	s_1^+	s_2^+	$\ldots$	$s_{\hat{m}}^+$	s_1^-	s_2^-	$\ldots$	$s_{m-\hat{m}}^-$	$\sum s_i$
$\hat{p}-1$	1	1	$\ldots$	1	0	0	$\ldots$	0	$\hat{m}$
$\hat{p}-2$	1	1	$\ldots$	1	0	0	$\ldots$	0	$\hat{m}$
$\hat{p}-3$	1	1	$\ldots$	1	0	0	$\ldots$	0	$\hat{m}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
1	1	1	$\ldots$	1	0	0	$\ldots$	0	$\hat{m}$
$\hat{p}+1$	1	1	$\ldots$	1	1	0	$\ldots$	0	$\hat{m}+1$
$\hat{p}+1$	1	1	$\ldots$	1	0	1	$\ldots$	0	$\hat{m}+1$
$\hat{p}+1$	1	1	$\ldots$	1	0	0	$\ldots$	0	$\hat{m}+1$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$\hat{p}+1$	1	1	$\ldots$	1	0	0	$\ldots$	1	$\hat{m}+1$
$\hat{p}$	1	1	$\ldots$	1	1	0	$\ldots$	0	$\hat{m}+1$
$\hat{p}$	1	1	$\ldots$	1	0	1	$\ldots$	0	$\hat{m}+1$
$\hat{p}$	1	1	$\ldots$	1	0	0	$\ldots$	0	$\hat{m}+1$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$\hat{p}$	1	1	$\ldots$	1	0	0	$\ldots$	1	$\hat{m}+1$
$\hat{p}-1$	1	1	$\ldots$	1	1	0	$\ldots$	0	$\hat{m}+1$
$\hat{p}-1$	1	1	$\ldots$	1	0	1	$\ldots$	0	$\hat{m}+1$
$\hat{p}-1$	1	1	$\ldots$	1	0	0	$\ldots$	0	$\hat{m}+1$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$\hat{p}-1$	1	1	$\ldots$	1	0	0	$\ldots$	1	$\hat{m}+1$
$\hat{p}+1$	0	1	$\ldots$	1	0	0	$\ldots$	0	$\hat{m}-1$
$\hat{p}+1$	1	0	$\ldots$	1	0	0	$\ldots$	0	$\hat{m}-1$
$\hat{p}+1$	1	1	$\ldots$	1	0	0	$\ldots$	0	$\hat{m}-1$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$\hat{p}+1$	1	1	$\ldots$	0	0	0	$\ldots$	0	$\hat{m}-1$
$\hat{p}$	0	1	$\ldots$	1	0	0	$\ldots$	0	$\hat{m}-1$
$\hat{p}$	1	0	$\ldots$	1	0	0	$\ldots$	0	$\hat{m}-1$
$\hat{p}$	1	1	$\ldots$	1	0	0	$\ldots$	0	$\hat{m}-1$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$\hat{p}$	1	1	$\ldots$	0	0	0	$\ldots$	0	$\hat{m}-1$
$\hat{p}-1$	0	1	$\ldots$	1	0	0	$\ldots$	0	$\hat{m}-1$
$\hat{p}-1$	1	0	$\ldots$	1	0	0	$\ldots$	0	$\hat{m}-1$
$\hat{p}-1$	1	1	$\ldots$	1	0	0	$\ldots$	0	$\hat{m}-1$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$\hat{p}-1$	1	1	$\ldots$	0	0	0	$\ldots$	0	$\hat{m}-1$

Table 6.8: Set of models considered at a non inital step.

measures which we adapt to our algorithm.

Each of these methods will be used to apply a score to a model fit, thus we refer to them as scoring functions, which we denote generically as the function Λ but specify them specifically here. Firstly we use the existing Mean Squared Error function defined by Equation (6.6). Following on from this we extend this application of MSE to the m indicator series fitted giving

$$\bar{m}(\mathbf{Y}, \hat{\mathbf{Y}}, \mathbf{X}, \hat{\mathbf{X}}, n, n^*, m) = m\left(\mathbf{Y}, \hat{\mathbf{Y}}, n, n^*\right) + \sum_{j=1}^{m} m\left(\mathbf{X}_j, \hat{\mathbf{X}}_j, n, n^*\right). \tag{6.13}$$

Moving beyond the MSE functions, we also apply information criterion proposed in Bai and Ng (2002), given by:

$$\begin{aligned} IC_1\left(\mathbf{Y}, \hat{\mathbf{Y}}, \hat{p}, n, n^*, m\right) &= \log\left(m\left(\mathbf{Y}, \hat{\mathbf{Y}}, n, n^*\right)\right) \\ &\quad + \hat{p}\frac{n+m+1}{n(m+1)}\log\left(\frac{n(m+1)}{n+m+1}\right), \end{aligned} \tag{6.14}$$

$$\begin{aligned} IC_2\left(\mathbf{Y}, \hat{\mathbf{Y}}, \hat{p}, n, n^*, m\right) &= \log\left(m\left(\mathbf{Y}, \hat{\mathbf{Y}}, n, n^*\right)\right) \\ &\quad + \hat{p}\frac{n+m+1}{n(m+1)}\log\left(\min\left(\sqrt{n}, \sqrt{m+1}\right)^2\right), \end{aligned} \tag{6.15}$$

$$\begin{aligned} IC_3\left(\mathbf{Y}, \hat{\mathbf{Y}}, \hat{p}, n, n^*, m\right) &= \log\left(m\left(\mathbf{Y}, \hat{\mathbf{Y}}, n, n^*\right)\right) \\ &\quad + \hat{p}\frac{\log\left(\min\left(\sqrt{n}, \sqrt{m+1}\right)^2\right)}{\min\left(\sqrt{n}, \sqrt{m+1}\right)^2}. \end{aligned} \tag{6.16}$$

6.5.3 Results of Simulations

We focus our analysis by measuring the algorithm in its success in 8 outcomes:

1. Removal of an independent series;
2. Selection of the correct amount of dependent series;
3. Selection of the correct amount of dynamic factors;
4. Determination of the correct model (combination of outcomes 1, 2 and 3);

Data: Y_t, Reference Series for $t = 1, \dots n$;
Data: $X_{t,j}$ For $j = 1, \dots, m$: Explanatory series over time $t = 1, \dots n$;
input : n^*, Amount of recent points to consider in a MSE calculation ;
$\eta = 1$ (or 0). Variable signifying forward/backward selection. ;
Λ: Scoring function. ;
$\hat{\delta} = \infty - \epsilon, \ \epsilon > 0$ The minimum Score of the current stage ;
$\delta = \infty$ The minimum score from all previous stages ;
$\delta_i^* = \infty \ \forall i$ The score of the fitted models in the current stage ;
$\hat{p}$: The current amount of dynamic factors ;
$\hat{\mathbf{s}} = \{\hat{s}_1, \dots, \hat{s}_m\}$. Variables to signify current inclusion of explanatory variables. ;

if $\tau = 1$ **then**
  Set $p = 1$ and $s_j = 0$ for $j = 1, \dots, m$. ;
  Propose initial k models as per Table (6.7a) with variable selection $\mathbf{s}_1, \dots, \mathbf{s}_k$ and dynamic factors $p_1, \dots, p_k$.
else
  Set $p = m$ and $s_j = 1$ for $j = 1, \dots, m$. ;
  Propose initial k models as per Table (6.7b) with variable selection $\mathbf{s}_1, \dots, \mathbf{s}_k$ and dynamic factors $p_1, \dots, p_k$.
end
while $\hat{\delta} \leq \delta$ *and* $k > 0$ **do**
  for l *in* $1 : k$ **do**
    Fit Dynamic Factor Model $\mathcal{M}_l$ using data $\{\mathbf{Y}, \{\mathbf{X}_{j:\hat{s}_{l,j}=1}\}\}$ with $\hat{p}_l$ dynamic factors. ;
    Calculate Score for fitted model, $\delta_l^* = \Lambda(\hat{\mathcal{M}}_l, n^*, \hat{p}_l)$;
  end
  Let $\hat{\delta} = \min(\{\delta_i^* : i = 1, \dots, k\})$;
  if $\hat{\delta} < \delta$ **then**
    Re-assign $\delta = \hat{\delta}$;
    Reset $\delta_i^* = \infty \ \forall i \in \mathbb{N}$;
    Set the current best variable selection, $\hat{\mathbf{s}} = \mathbf{s}_{i:\delta_i^* = \min(\boldsymbol{\delta}^*)}$;
    Set the current best amount of factors selection, $\hat{p} = p_{i:\delta_i^* = \min(\boldsymbol{\delta}^*)}$;
    Set the current best Model, $\hat{\mathcal{M}} = \mathcal{M}_{i:\delta_i^* = \min(\boldsymbol{\delta}^*)}$;
    Propose new k models to test as per Table (6.8) with variable selection $\mathbf{s}_1, \dots, \mathbf{s}_k$ and dynamic factors $p_1, \dots, p_k$.
  end
end
output: Best Model $\hat{\mathcal{M}}$, Best Variable Selection $\hat{\mathbf{s}}$ and best number of dynamic factors $\hat{p}$. ;

Algorithm 4: Generic Algorithm extension for choosing the optimal selection of variables and dynamic factors based upon a scoring function.

5. Consideration rates of the true model;

6. Mean squared error for the reference series over the most recent 24 points;

7. Mean squared error for the reference series over all points;

8. Ability to isolate correct components for dependency structure 4 (given the differing structure).

The results of each outcome is given in the tables that follow. Note that for comparison we computed the simulations over $n^* = 24,500$ but found the results to be very similar. As such we present the tables for $n^* = 24$ here, but move the tables for $n^* = 500$ to Appendix 6.7.1 and highlight any differences.

We begin by reviewing the results in Table 6.9 where we consider the amount of simulations which successfully removed the independent seies. Note that for comparison with other tables we have kept in the column 'No Series' representing no independent series was present to begin with, but thus these results automatically default to 1. It can be seen that there is more strength in this outcome when using forward selection, where we do not begin with the assumption that the series should be included. However, when models with higher complexity are involved (dependent models 2, 3 and 4) the information criterion scoring functions are more robust than that of the Mean Squared Error measures. There does not appear to be any significant variability between the information criterion results, but it can be seen that using the indicator series additionally in the MSE measures increases the removal rates when using a forward selection approach. In comparison to the results for $n^* = 500$ in Table 6.19, increasing the comparison region used in the scoring function appeared to only benefit the simpler dependent model 1 for MSE measures, and more so when the indicators series are used.

Moving next to Table 6.10 we consider here the amount of time we successfully selected the correct amount of dependent variables. It can be seen that this particular aspect of the model is very difficult to attain, with very low success rates. It most often occurs in a forward selection approach for Model 1, however this is not

Score Function	Dep. Model	Back	Indep. 1	Indep. 2	Indep. 3	Indep. 4	No Series
Mean Squared	1	T	0.46	0.41	0.43	0.40	1.00
Error using	1	F	0.94	0.89	0.90	0.88	1.00
Reference	2	T	0.36	0.32	0.21	0.34	1.00
Series	2	F	0.84	0.79	0.80	0.80	1.00
(Equation (6.6))	3	T	0.22	0.18	0.18	0.17	1.00
	3	F	0.90	0.86	0.87	0.86	1.00
	4	T	0.25	0.23	0.17	0.25	1.00
	4	F	0.94	0.92	0.91	0.90	1.00
Mean Squared	1	T	0.51	0.50	0.52	0.71	1.00
Error using	1	F	0.99	0.98	0.99	0.98	1.00
Reference	2	T	0.53	0.50	0.48	0.82	1.00
Series and	2	F	0.83	0.79	0.79	0.80	1.00
Indicator	3	T	0.51	0.52	0.47	0.76	1.00
Series	3	F	0.90	0.89	0.90	0.88	1.00
(Equation (6.13))	4	T	0.42	0.46	0.41	0.69	1.00
	4	F	0.96	0.94	0.96	0.94	1.00
Information	1	T	0.93	0.86	0.84	0.86	1.00
Criterion	1	F	0.95	0.89	0.90	0.87	1.00
1	2	T	0.80	0.71	0.59	0.70	1.00
(Equation (6.14))	2	F	0.85	0.80	0.80	0.80	1.00
	3	T	0.78	0.71	0.56	0.75	1.00
	3	F	0.91	0.88	0.88	0.87	1.00
	4	T	0.68	0.55	0.48	0.67	1.00
	4	F	0.94	0.93	0.91	0.92	1.00
Information	1	T	0.93	0.87	0.83	0.86	1.00
Criterion	1	F	0.95	0.89	0.90	0.87	1.00
2	2	T	0.78	0.72	0.64	0.69	1.00
(Equation (6.15))	2	F	0.85	0.80	0.80	0.80	1.00
	3	T	0.78	0.70	0.56	0.74	1.00
	3	F	0.90	0.88	0.88	0.87	1.00
	4	T	0.66	0.57	0.51	0.66	1.00
	4	F	0.95	0.93	0.91	0.92	1.00
Information	1	T	0.93	0.88	0.83	0.86	1.00
Criterion	1	F	0.95	0.89	0.90	0.87	1.00
3	2	T	0.80	0.71	0.57	0.70	1.00
(Equation (6.16))	2	F	0.85	0.80	0.80	0.80	1.00
	3	T	0.77	0.72	0.54	0.73	1.00
	3	F	0.91	0.87	0.88	0.87	1.00
	4	T	0.67	0.57	0.46	0.66	1.00
	4	F	0.94	0.92	0.91	0.92	1.00

Table 6.9: The proportion of models which converged upon the removal of an independent series. Separated by the information criterion used, the dependence model, usage of a backward or forward selection procedure, and the type of independent series included. **Results closest to 1 and favourable in all cases.**

Score Function	Dep. Model	Back	Indep. 1	Indep. 2	Indep. 3	Indep. 4	No Series
Mean Squared	1	T	0.42	0.35	0.32	0.46	0.05
Error using	1	F	0.01	0.01	0.01	0.01	0.02
Reference	2	T	0.04	0.05	0.09	0.06	0.00
Series	2	F	0.00	0.00	0.00	0.00	0.00
(Equation (6.6))	3	T	0.03	0.03	0.10	0.04	0.01
	3	F	0.00	0.00	0.00	0.00	0.00
	4	T	0.05	0.08	0.13	0.01	0.02
	4	F	0.00	0.00	0.00	0.00	0.00
Mean Squared	1	T	0.17	0.12	0.12	0.10	0.01
Error using	1	F	0.00	0.00	0.00	0.00	0.00
Reference	2	T	0.01	0.02	0.03	0.01	0.00
Series and	2	F	0.00	0.00	0.00	0.00	0.00
Indicator	3	T	0.00	0.01	0.02	0.00	0.00
Series	3	F	0.00	0.00	0.00	0.00	0.00
(Equation (6.13))	4	T	0.00	0.00	0.01	0.00	0.00
	4	F	0.00	0.00	0.00	0.00	0.00
Information	1	T	0.04	0.03	0.04	0.03	0.02
Criterion	1	F	0.02	0.02	0.02	0.02	0.02
1	2	T	0.05	0.04	0.05	0.03	0.06
(Equation (6.14))	2	F	0.00	0.00	0.00	0.00	0.00
	3	T	0.04	0.05	0.07	0.04	0.03
	3	F	0.00	0.00	0.00	0.00	0.00
	4	T	0.07	0.10	0.11	0.05	0.02
	4	F	0.00	0.00	0.00	0.00	0.00
Information	1	T	0.04	0.03	0.05	0.03	0.02
Criterion	1	F	0.02	0.02	0.02	0.02	0.03
2	2	T	0.06	0.04	0.07	0.04	0.07
(Equation (6.15))	2	F	0.00	0.00	0.00	0.00	0.00
	3	T	0.04	0.05	0.08	0.05	0.03
	3	F	0.00	0.00	0.00	0.00	0.00
	4	T	0.06	0.09	0.10	0.05	0.02
	4	F	0.00	0.00	0.00	0.00	0.00
Information	1	T	0.04	0.03	0.05	0.04	0.02
Criterion	1	F	0.02	0.02	0.02	0.02	0.03
3	2	T	0.06	0.04	0.07	0.04	0.07
(Equation (6.16))	2	F	0.00	0.00	0.00	0.00	0.00
	3	T	0.04	0.04	0.06	0.04	0.03
	3	F	0.00	0.00	0.00	0.00	0.00
	4	T	0.06	0.09	0.10	0.05	0.03
	4	F	0.00	0.00	0.00	0.00	0.00

Table 6.10: The proportion of models converged upon the correct selection in the amount of dependent series. Separated by the information criterion used, the dependence model, usage of a backward or forward selection procedure, and the type of independent series included. **Results closest to 1 are favourable in all cases.**

the case for the more complex models. There we find greater strength through a backward selection approach, but this is still a sparse amount of times. Whilst the MSE approaches have the highest results, particularly for model 1, the Information Criterion appear to be consistent across all models. Of note however is that not including an independent series appears to hamper the MSE approaches, but this effect is not present for the information criterion approaches. Otherwise there is no other significant difference between the independent series used. Comparing to Table 6.20 we can only see one signficant difference, such that the shorter comparison region ($n^* = 24$) is an improvement on the results for Model 1, using a backward approach with just the reference series MSE comparison.

Selection rates for the amount of dynamic factors, given in Table 6.11, show more promise. The most strength can be found in a forward based approach for model 1, however this seems likely given this would begin the algorithm with the correct amount of factors. Beyond this, a backward approach is more reasonable, showing consistent gains in strength across all scoring functions. Information Criterion hold the greatest strength across all models however, but have very similar results to one another. Looking to the independent series there appears to be evidence that using the ARMA series specified reduces the success rate, but this could be a negligible difference. In comparison to Table 6.21 we find increasing strength for the simple model 1, using backward selection for $n^* = 500$.

Table 6.12 shows the results of considering the success of all three previous components simultaneously, where we consider the amount of simulations where the model structure was exactly specified. It can be seen that as expected from the previous tables, selection of the correct model is a very rare event. The strongest results are through the use of an Information Criterion, but these are still weak. Similar to previous analysis, using a backward approach aids in strength for more complex models. The results appear to be agnostic to independent series, but it is difficult to truly determine given the low results. Comparing to the same results for $n^* = 500$ in Table 6.18 most results are consistent with some being marginally

Score Function	Dep. Model	Back	Indep. 1	Indep. 2	Indep. 3	Indep. 4	No Series
Mean Squared	1	T	0.48	0.48	0.48	0.47	0.96
Error using	1	F	0.99	0.98	0.98	0.98	0.99
Reference	2	T	0.44	0.39	0.27	0.45	0.88
Series	2	F	0.10	0.08	0.08	0.08	0.13
(Equation (6.6))	3	T	0.26	0.28	0.22	0.25	0.82
	3	F	0.11	0.10	0.08	0.10	0.11
	4	T	0.36	0.32	0.18	0.39	0.63
	4	F	0.05	0.05	0.05	0.05	0.04
Mean Squared	1	T	0.51	0.51	0.52	0.72	1.00
Error using	1	F	1.00	1.00	1.00	1.00	1.00
Reference	2	T	0.54	0.49	0.48	0.79	0.96
Series and	2	F	0.03	0.03	0.03	0.03	0.06
Indicator	3	T	0.54	0.54	0.48	0.74	0.92
Series	3	F	0.03	0.03	0.02	0.03	0.04
(Equation (6.13))	4	T	0.42	0.45	0.39	0.61	0.71
	4	F	0.01	0.01	0.01	0.01	0.01
Information	1	T	0.99	0.97	0.94	0.99	1.00
Criterion	1	F	1.00	1.00	1.00	1.00	1.00
1	2	T	0.73	0.68	0.59	0.73	0.75
(Equation (6.14))	2	F	0.05	0.05	0.04	0.04	0.09
	3	T	0.66	0.62	0.54	0.67	0.70
	3	F	0.04	0.05	0.04	0.05	0.04
	4	T	0.44	0.45	0.30	0.50	0.50
	4	F	0.02	0.01	0.02	0.01	0.01
Information	1	T	0.99	0.98	0.93	0.99	1.00
Criterion	1	F	1.00	1.00	1.00	1.00	1.00
2	2	T	0.73	0.69	0.58	0.74	0.74
(Equation (6.15))	2	F	0.05	0.05	0.04	0.04	0.08
	3	T	0.66	0.60	0.54	0.67	0.70
	3	F	0.05	0.05	0.04	0.05	0.05
	4	T	0.44	0.45	0.33	0.50	0.50
	4	F	0.01	0.01	0.02	0.01	0.01
Information	1	T	0.99	0.98	0.93	0.99	1.00
Criterion	1	F	1.00	1.00	1.00	1.00	1.00
3	2	T	0.72	0.66	0.56	0.74	0.76
(Equation (6.16))	2	F	0.05	0.05	0.04	0.04	0.08
	3	T	0.67	0.63	0.52	0.67	0.70
	3	F	0.04	0.05	0.04	0.05	0.05
	4	T	0.44	0.47	0.30	0.50	0.50
	4	F	0.02	0.01	0.02	0.01	0.01

Table 6.11: The proportion of models converged upon the correct amount of dynamic factors. Separated by the information criterion used, the dependence model, usage of a backward or forward selection procedure, and the type of independent series included. **Results closest to 1 are favourable in all cases.**

Score Function	Dep. Model	Back	Indep. 1	Indep. 2	Indep. 3	Indep. 4	No Series
Mean Squared	1	T	0.00	0.00	0.00	0.00	0.01
Error using	1	F	0.00	0.00	0.00	0.00	0.01
Reference	2	T	0.00	0.00	0.00	0.00	0.00
Series	2	F	0.00	0.00	0.00	0.00	0.00
(Equation (6.6))	3	T	0.00	0.00	0.00	0.00	0.00
	3	F	0.00	0.00	0.00	0.00	0.00
	4	T	0.00	0.00	0.00	0.00	0.01
	4	F	0.00	0.00	0.00	0.00	0.00
Mean Squared	1	T	0.00	0.00	0.00	0.00	0.00
Error using	1	F	0.00	0.00	0.00	0.00	0.00
Reference	2	T	0.00	0.00	0.00	0.00	0.00
Series and	2	F	0.00	0.00	0.00	0.00	0.00
Indicator	3	T	0.00	0.00	0.00	0.00	0.00
Series	3	F	0.00	0.00	0.00	0.00	0.00
(Equation (6.13))	4	T	0.00	0.00	0.00	0.00	0.00
	4	F	0.00	0.00	0.00	0.00	0.00
Information	1	T	0.02	0.02	0.02	0.02	0.02
Criterion	1	F	0.02	0.02	0.02	0.02	0.02
1	2	T	0.03	0.03	0.02	0.02	0.06
(Equation (6.14))	2	F	0.00	0.00	0.00	0.00	0.00
	3	T	0.01	0.01	0.01	0.02	0.03
	3	F	0.00	0.00	0.00	0.00	0.00
	4	T	0.02	0.01	0.01	0.01	0.02
	4	F	0.00	0.00	0.00	0.00	0.00
Information	1	T	0.02	0.02	0.02	0.02	0.02
Criterion	1	F	0.02	0.02	0.02	0.02	0.03
2	2	T	0.03	0.03	0.04	0.02	0.06
(Equation (6.15))	2	F	0.00	0.00	0.00	0.00	0.00
	3	T	0.02	0.01	0.02	0.02	0.03
	3	F	0.00	0.00	0.00	0.00	0.00
	4	T	0.01	0.01	0.01	0.01	0.02
	4	F	0.00	0.00	0.00	0.00	0.00
Information	1	T	0.02	0.02	0.02	0.02	0.02
Criterion	1	F	0.02	0.02	0.02	0.02	0.03
3	2	T	0.03	0.03	0.02	0.02	0.06
(Equation (6.16))	2	F	0.00	0.00	0.00	0.00	0.00
	3	T	0.02	0.01	0.01	0.01	0.03
	3	F	0.00	0.00	0.00	0.00	0.00
	4	T	0.02	0.02	0.01	0.01	0.02
	4	F	0.00	0.00	0.00	0.00	0.00

Table 6.12: The proportion of models converged upon that were exactly correct in three aspects: selection of the amount of dynamic factors, selection in the amount of dependent series and removal of an independent series. Separated by the information criterion used, the dependence model, usage of a backward or forward selection procedure, and the type of independent series included. **Results closest to 1 are favourable in all cases.**

lower, but this is a negligible amount.

Following these results it is important that we put them in the correct context, as due to the nature of the algorithm, whilst certain components may have been difficult to detect there is also the possibility that we did not consider the correct model in our process. As such we present the results in Table 6.13 to study the amount of time we considered the true model in our simulations. Note that in the case of no independent series and a backward selection process, we will always consider the true model. We find we most often consider the true structure for model 1 with a forward selection process when there is an independent series present. However it appears that the strength found in a forward selection does not transfer to more complex model structures, where backward selection appears to be the most viable option. For these more complex models we find that the information criterion hold a greater strength than the MSE scoring functions. As with other results, it does not appear that the choice of independent series has a significant impact upon consideration rates. In comparison to the results for $n^* = 500$ we find there is a small increase in strength for the MSE scoring function which also considers the indicator series, with potential minor differences in the Information Criterion scoring functions, but these appear negligible.

The previous results have focussed primarily on the structure selection of the model, but of most interest to a practitioner is the efficiency in forecasting. As such we present results on the average MSE of the fits simulated in Table 6.14, as discussed in Section 6.3. As we use one-step ahead forecasts for all time points (through use of the Kalman Filter), this can be seen as a measure of forecasting efficiency. We begin by reviewing the average mean squared error of the reference series over the past 24 points, measuring its accuracy at predicting the series at its most recent activity. It can be seen that the complex dependent structure of model 2 gives the highest MSE when using forward selection but this is countered by the low results which occur through backward selection.

This is the case for all dependent structures except model 1 which appears to

Score Function	Dep. Model	Back	Indep. 1	Indep. 2	Indep. 3	Indep. 4	No Series
Mean Squared	1	T	0.79	0.71	0.67	0.79	1.00
Error using	1	F	0.95	0.91	0.92	0.89	1.00
Reference	2	T	0.05	0.06	0.09	0.07	1.00
Series	2	F	0.00	0.00	0.00	0.00	0.01
(Equation (6.6))	3	T	0.07	0.07	0.11	0.05	1.00
	3	F	0.05	0.05	0.04	0.05	0.06
	4	T	0.14	0.16	0.18	0.11	1.00
	4	F	0.01	0.02	0.02	0.01	0.02
Mean Squared	1	T	0.60	0.59	0.60	0.80	1.00
Error using	1	F	0.99	0.98	0.99	0.98	1.00
Reference	2	T	0.04	0.03	0.05	0.30	1.00
Series and	2	F	0.00	0.00	0.00	0.00	0.00
Indicator	3	T	0.03	0.02	0.04	0.28	1.00
Series	3	F	0.00	0.00	0.00	0.00	0.00
(Equation (6.13))	4	T	0.02	0.02	0.04	0.17	1.00
	4	F	0.00	0.00	0.00	0.00	0.00
Information	1	T	0.99	0.97	0.93	0.99	1.00
Criterion	1	F	0.95	0.91	0.92	0.89	1.00
1	2	T	0.52	0.61	0.57	0.51	1.00
(Equation (6.14))	2	F	0.02	0.02	0.01	0.01	0.02
	3	T	0.58	0.63	0.58	0.56	1.00
	3	F	0.04	0.04	0.04	0.04	0.05
	4	T	0.57	0.59	0.58	0.54	1.00
	4	F	0.01	0.01	0.01	0.01	0.01
Information	1	T	0.99	0.97	0.93	0.99	1.00
Criterion	1	F	0.95	0.91	0.92	0.89	1.00
2	2	T	0.55	0.64	0.59	0.50	1.00
(Equation (6.15))	2	F	0.02	0.02	0.01	0.02	0.03
	3	T	0.59	0.63	0.54	0.57	1.00
	3	F	0.04	0.04	0.04	0.04	0.05
	4	T	0.58	0.61	0.61	0.54	1.00
	4	F	0.01	0.01	0.01	0.01	0.01
Information	1	T	0.99	0.97	0.93	0.99	1.00
Criterion	1	F	0.95	0.91	0.92	0.89	1.00
3	2	T	0.54	0.60	0.56	0.50	1.00
(Equation (6.16))	2	F	0.02	0.01	0.01	0.01	0.03
	3	T	0.58	0.63	0.60	0.56	1.00
	3	F	0.04	0.04	0.04	0.04	0.05
	4	T	0.56	0.61	0.55	0.53	1.00
	4	F	0.01	0.01	0.01	0.01	0.01

Table 6.13: The proportion of models which considered the true model as part of their exploration. Separated by the information criterion used, the dependence model, usage of a backward or forward selection procedure, and the type of independent series included. **Results closest to 1 are favourable in all cases.**

Score Function	Dep. Model	Back	Indep. 1	Indep. 2	Indep. 3	Indep. 4	No Series
Mean Squared	1	T	0.45	0.44	0.43	0.45	0.44
Error using	1	F	0.44	0.44	0.44	0.44	0.44
Reference	2	T	0.29	0.28	0.29	0.30	0.30
Series	2	F	2.70	2.96	3.07	2.81	3.37
(Equation (6.6))	3	T	0.28	0.28	0.28	0.28	0.29
	3	F	0.84	0.80	0.82	0.79	1.02
	4	T	0.26	0.25	0.24	0.26	0.27
	4	F	0.36	0.35	0.35	0.35	0.37
Mean Squared	1	T	0.50	0.50	0.49	0.50	0.48
Error using	1	F	0.50	0.49	0.49	0.50	0.48
Reference	2	T	0.41	0.42	0.40	0.43	0.44
Series and	2	F	3.46	3.64	3.90	3.45	3.82
Indicator	3	T	0.41	0.42	0.40	0.42	0.42
Series	3	F	1.33	1.20	1.30	1.27	1.35
(Equation (6.13))	4	T	0.41	0.38	0.38	0.39	0.39
	4	F	0.54	0.51	0.54	0.51	0.51
Information	1	T	0.44	0.43	0.43	0.44	0.44
Criterion	1	F	0.44	0.44	0.44	0.44	0.44
1	2	T	0.31	0.30	0.29	0.31	0.32
(Equation (6.14))	2	F	2.71	2.96	3.06	2.82	3.33
	3	T	0.30	0.30	0.30	0.30	0.31
	3	F	0.84	0.80	0.83	0.79	1.03
	4	T	0.28	0.27	0.24	0.29	0.28
	4	F	0.36	0.35	0.35	0.35	0.38
Information	1	T	0.44	0.43	0.43	0.44	0.44
Criterion	1	F	0.44	0.44	0.44	0.44	0.44
2	2	T	0.31	0.30	0.30	0.31	0.32
(Equation (6.15))	2	F	2.71	2.96	3.06	2.82	3.34
	3	T	0.30	0.30	0.29	0.30	0.31
	3	F	0.84	0.80	0.83	0.79	1.03
	4	T	0.28	0.27	0.23	0.28	0.28
	4	F	0.36	0.35	0.35	0.35	0.38
Information	1	T	0.44	0.44	0.43	0.44	0.44
Criterion	1	F	0.44	0.44	0.44	0.44	0.44
3	2	T	0.31	0.30	0.29	0.31	0.31
(Equation (6.16))	2	F	2.71	2.96	3.06	2.82	3.34
	3	T	0.30	0.30	0.29	0.30	0.31
	3	F	0.84	0.80	0.83	0.79	1.03
	4	T	0.28	0.27	0.24	0.29	0.28
	4	F	0.36	0.35	0.35	0.35	0.38

Table 6.14: Mean squared error of the model fits over the most recent 24 points in the reference series. Separated by the information criterion used, the dependence model, usage of a backward or forward selection procedure, and the type of independent series included. **Results closest to 0 are favourable in all cases.**

be agnostic to this, given its simpler structure. The lowest MSE occur consistently (unsuprisingly) for the MSE scoring function which only considers the reference series, but there are only minor differences between those of the information criterion. Similarly we find the highest results occur when also considering the indicator series as part of the MSE scoring function rather than focusing on the reference series. Beyond this there does not appear to be any significant difference between the independent series when using a backward selection, but it appears to have an adverse affect in the forward selection approach.

If we compare these results to that similarly of $n^* = 500$ in Table 6.23 most results change negligibly. However we see that the unique behaviour of model 2 grows with higher MSE results for the forward selection approaches. There also appears to be similar effects on model 3 and (to a lesser extent) model 4 when considering them using the reference series also. Similarly the effect of having no independent series with a forward selection grows.

If we now consider, as a measure of model fit rather than forecasting efficiency, the MSE of the reference series across all points, the average of this outcome is given in Table 6.15. We find that much of the results discussed similarly over the most recent 24 points hold true also across all the points. Of note however is that now the MSE scoring function which also considers the indicator series now decreases (rather than increasing in the previous case) the MSE as we proceed from $n^* = 24$ to the results of $n^* = 500$ in Table 6.24. This happens aross all results with that scoring function except those of dependent structure 2 and 3 with forward selection, where similarly we see the largest increases in MSE in the previous Tables 6.14 and 6.23.

As a particular example of component selection, we now refer back to our simulated dependency structure of model 4 given in Equation (6.12). Here we structured the matrix A such that Series 2 would have double the dependency as the other series by using higher weights within the matrix. Table 6.16 shows the selection of the series and amount of dynamic factors across the different

Score Function	Dep. Model	Back	Indep. 1	Indep. 2	Indep. 3	Indep. 4	No Series
Mean Squared	1	T	0.51	0.49	0.49	0.50	0.48
Error using	1	F	0.49	0.47	0.48	0.47	0.48
Reference	2	T	0.35	0.35	0.35	0.36	0.35
Series	2	F	4.17	3.29	4.07	3.99	4.35
(Equation (6.6))	3	T	0.35	0.34	0.33	0.35	0.35
	3	F	0.70	0.78	0.83	0.62	1.02
	4	T	0.30	0.29	0.27	0.30	0.31
	4	F	0.35	0.34	0.36	0.34	0.35
Mean Squared	1	T	0.69	0.62	0.53	0.62	0.48
Error using	1	F	0.89	0.66	0.53	0.67	0.48
Reference	2	T	0.60	0.74	0.49	0.71	0.78
Series and	2	F	9.10	8.92	8.72	8.38	6.54
Indicator	3	T	0.55	0.79	0.56	0.73	0.67
Series	3	F	3.65	3.14	4.24	4.20	2.52
(Equation (6.13))	4	T	0.78	0.58	0.80	0.74	0.62
	4	F	1.12	1.30	1.71	1.24	0.83
Information	1	T	0.49	0.47	0.47	0.47	0.48
Criterion	1	F	0.49	0.47	0.48	0.47	0.48
1	2	T	0.39	0.38	0.38	0.39	0.39
(Equation (6.14))	2	F	4.22	3.29	4.02	4.00	4.06
	3	T	0.39	0.38	0.35	0.38	0.38
	3	F	0.71	1.14	0.83	0.63	1.03
	4	T	0.32	0.31	0.29	0.33	0.33
	4	F	0.35	0.35	0.36	0.34	0.35
Information	1	T	0.49	0.47	0.47	0.47	0.48
Criterion	1	F	0.49	0.47	0.48	0.47	0.48
2	2	T	0.39	0.39	0.36	0.39	0.39
(Equation (6.15))	2	F	4.22	3.29	4.02	3.99	4.18
	3	T	0.39	0.38	0.35	0.39	0.38
	3	F	0.70	1.14	0.83	0.63	1.02
	4	T	0.32	0.32	0.29	0.33	0.33
	4	F	0.35	0.35	0.36	0.34	0.35
Information	1	T	0.49	0.47	0.47	0.47	0.48
Criterion	1	F	0.49	0.47	0.48	0.47	0.48
3	2	T	0.40	0.38	0.37	0.39	0.40
(Equation (6.16))	2	F	4.22	3.30	4.02	4.00	4.18
	3	T	0.39	0.38	0.35	0.39	0.37
	3	F	0.71	1.14	0.83	0.63	1.02
	4	T	0.32	0.32	0.29	0.34	0.33
	4	F	0.35	0.35	0.36	0.34	0.35

Table 6.15: Mean squared error of the model fits over all points in the reference series. Separated by the scoring function used, the dependence model, usage of a backward or forward selection procedure, and the type of independent series included. **Results closest to 0 are favourable in all cases.**

Score Function	Backward Selection	Series 2	Series 3	Series 4	Series 5	Independent Model	Independent Incl.	p = 1	p = 2	p = 3	p = 4
Mean Squared	T	0.21	0.63	0.62	0.69	1	0.75	0.04	0.36	0.42	0.17
Error using	T	0.25	0.64	0.61	0.73	2	0.77	0.05	0.32	0.42	0.20
Reference	T	0.33	0.71	0.64	0.76	3	0.83	0.04	0.18	0.49	0.24
Series	T	0.16	0.59	0.64	0.63	4	0.75	0.05	0.39	0.40	0.15
(Equation (6.6))	T	0.20	0.63	0.53	0.71	5	0.00	0.21	0.63	0.14	0.01
	F	0.15	0.14	0.63	0.15	1	0.06	0.94	0.05	0.01	0.00
	F	0.15	0.13	0.62	0.15	2	0.08	0.94	0.05	0.01	0.00
	F	0.16	0.14	0.61	0.15	3	0.09	0.93	0.05	0.01	0.00
	F	0.15	0.14	0.62	0.15	4	0.10	0.94	0.05	0.01	0.00
	F	0.16	0.14	0.67	0.16	5	0.00	0.95	0.04	0.01	0.00
Mean Squared	T	0.46	0.41	0.45	0.42	1	0.58	0.17	0.42	0.33	0.08
Error using	T	0.44	0.46	0.44	0.44	2	0.54	0.15	0.45	0.32	0.08
Reference	T	0.46	0.45	0.48	0.43	3	0.59	0.15	0.39	0.37	0.08
Series and	T	0.42	0.43	0.43	0.39	4	0.31	0.21	0.61	0.15	0.03
Indicator	T	0.50	0.34	0.46	0.41	5	0.00	0.29	0.71	0.00	0.00
Series	F	0.06	0.01	0.87	0.02	1	0.04	0.99	0.01	0.00	0.00
(Equation (6.13))	F	0.06	0.01	0.85	0.02	2	0.06	0.99	0.01	0.00	0.00
	F	0.06	0.02	0.87	0.02	3	0.04	0.99	0.01	0.00	0.00
	F	0.06	0.01	0.85	0.02	4	0.06	0.99	0.01	0.00	0.00
	F	0.06	0.02	0.91	0.02	5	0.00	0.99	0.01	0.00	0.00
Information	T	0.25	0.54	0.56	0.61	1	0.32	0.43	0.44	0.12	0.01
Criterion	T	0.28	0.61	0.54	0.61	2	0.45	0.35	0.45	0.17	0.03
(Equation (6.14))	T	0.29	0.64	0.55	0.67	3	0.52	0.31	0.30	0.34	0.05
	T	0.25	0.51	0.56	0.54	4	0.33	0.46	0.50	0.04	0.00
	T	0.22	0.60	0.53	0.60	5	0.00	0.50	0.50	0.00	0.00
	F	0.16	0.15	0.65	0.15	1	0.06	0.98	0.02	0.00	0.00
	F	0.16	0.14	0.63	0.15	2	0.07	0.99	0.01	0.00	0.00
	F	0.17	0.14	0.62	0.15	3	0.09	0.98	0.02	0.00	0.00
	F	0.16	0.14	0.64	0.14	4	0.08	0.99	0.01	0.00	0.00
	F	0.17	0.15	0.68	0.15	5	0.00	0.99	0.01	0.00	0.00
Information	T	0.26	0.54	0.56	0.61	1	0.34	0.42	0.44	0.13	0.01
Criterion	T	0.29	0.61	0.51	0.60	2	0.43	0.37	0.45	0.15	0.03
2	T	0.29	0.65	0.53	0.67	3	0.49	0.31	0.33	0.33	0.04
(Equation (6.15))	T	0.24	0.51	0.56	0.53	4	0.34	0.45	0.50	0.04	0.00
	T	0.22	0.59	0.53	0.60	5	0.00	0.50	0.50	0.00	0.00
	F	0.16	0.14	0.65	0.15	1	0.05	0.99	0.01	0.00	0.00
	F	0.16	0.14	0.63	0.15	2	0.07	0.99	0.01	0.00	0.00
	F	0.17	0.14	0.62	0.15	3	0.09	0.98	0.02	0.00	0.00
	F	0.16	0.14	0.63	0.15	4	0.08	0.99	0.01	0.00	0.00
	F	0.17	0.14	0.68	0.15	5	0.00	0.99	0.01	0.00	0.00
Information	T	0.26	0.55	0.56	0.61	1	0.33	0.42	0.44	0.13	0.01
Criterion	T	0.28	0.63	0.55	0.59	2	0.43	0.34	0.47	0.17	0.01
3	T	0.29	0.67	0.55	0.65	3	0.54	0.29	0.30	0.36	0.05
(Equation (6.16))	T	0.24	0.52	0.56	0.55	4	0.34	0.46	0.50	0.04	0.00
	T	0.23	0.61	0.53	0.60	5	0.00	0.50	0.50	0.00	0.00
	F	0.16	0.15	0.65	0.15	1	0.06	0.98	0.02	0.00	0.00
	F	0.16	0.15	0.64	0.14	2	0.08	0.99	0.01	0.00	0.00
	F	0.17	0.14	0.62	0.15	3	0.09	0.98	0.02	0.00	0.00
	F	0.16	0.15	0.64	0.14	4	0.08	0.99	0.01	0.00	0.00
	F	0.17	0.14	0.68	0.15	5	0.00	0.99	0.01	0.00	0.00

Table 6.16: Selection rates for each dependent and independent series and the amount of dynamic factors selected in Model 4 using $n^* = 24$. Separated by the scoring function used, usage of a backward of forward selection procedure. **Results closest to 1 for Series 2-5, closest to 0 for independent series, closest to 1 for $p = 2$ and closest 0 for all other p are favourable.**

algorithm setups for $n^* = 24$. Additionally we present the results for $n^* = 500$ in the Appendix within Table 6.25 but find no significant departures in the analysis that follows.

It can be immediately seen that the opposite effect has occurred in relation to Series 2. It is not the most included series as expected, in fact it is often the most excluded series particularly when using a MSE approach with only the reference series or an information criterion. It does however show a marginally higher than fair selection rate when using the MSE approach which incorporates the indicator series with backward selection. Additionally we can study the inclusion of an independent series here (note model 5 is always 0 as there is no series present), and find that backward selection provided the greatest strength, particularly when considering only the reference series in an MSE approach. Finally we review the selection for the amount of dynamic factors, where 2 is the true amount. It can be seen that a backward selection is the most accurate, with varying degrees across the choice of indepenent series. Particular weakness can be found from independent series 3, however the greatest strength occurs with model 5 as can be expected, where using a MSE approach with reference and indicator series provides the best results.

6.5.4 Discussion of Results

Given the extent of the results given in the previous tables, it would be prudent to summarise what results we have found. As such we present such a summary in Table 6.17 for $n^* = 24$ (in the Appendix we present Table 6.26 for $n^* = 500$). Here we review each of the desired outcomes stated in Section 6.5.3 (minus the last which was a specific investigation) and review which approach would be most suitable. We present that which has the most favourable result, the margin by which that occurs, and the standard deviation to consider this result with.

It can quickly be seen that there is no one approach to cover all of these outcomes. This has been clearly present throughout the previous tables, where

Strategy	Scoring Function	Backward or Forward	Difference from Mean	Standard Deviation
Removal of Independent Series*	Mean Squared Error using Reference Series and Indicator Series (Equation (6.13))	Forward	0.17	0.07
Selection of correct Dependent Series	Mean Squared Error using Reference Series (Equation (6.6))	Backward	0.08	0.14
Selection of correct amount of Dynamic Factors	Information Criterion 2 (Equation (6.15))	Backward	0.23	0.20
Consideration of correct model structure	Information Criterion 2 (Equation (6.15))	Backward	0.32	0.21
Selection of correct model structure	Information Criterion 2 (Equation (6.15))	Backward	0.01	0.01
Minimisation of Mean Squared Error over previous 24 points in Reference Series	Mean Squared Error using Reference Series (Equation (6.6))	Backward	-0.47	0.08
Minimisation of Mean Squared Error over all points in Reference Series	Mean Squared Error using Reference Series (Equation (6.6))	Backward	-0.75	0.08

Table 6.17: Optimal algorithmic approaches for desired outcome for $n^* = 24$. Results are averaged across all dependent and independent model choices and the most optimal is selected as that which departs from the mean of the results the most in the favourable direction. The standard deviation of these results is presented for context also. **We remove the case of no independent series here as to not skew the results.*

certain algorithm approaches hold strength in different areas. When considering the dataset to model with, MSE approaches hold best in general. To remove an independent series it is best to consider the MSE also of the indicator series being used, as we can expect the independent series to be poorly modelled. Similarly selection of the correct dependent series is best represented in the MSE of the reference series only. However, to identify the correct amount of dynamic factors, it is recommended that an Information Criterion approach is best. Note that although we recommend Information Criterion 2, we have seen little difference between the three studied. This similarly holds true for reviewing and selecting the correct structure, but these results are particularly inconclusive in their difference. As can be expected, focussing on reducing the MSE of the reference series is best approached by that of the MSE itself. Note that unlike the other results, we find that the reduction is outside of twice the standard deviation of the aggregated results we studied, giving more weight to this finding.

6.6 Conclusion

Within this work we have looked to approach to the task of Nowcasting from a position of naiviety. We began by first exploring the Dynamic Factor Model which is often used in such a task, reviewing and highlighting the issues and considerations that must be made. We found that in the design of such an approach there are two main areas of interest, firstly that of the construction of the dataset used, ensuring that we are using series which indicate correctly the movement of our reference series, and disregarding any which are independent of our interest. Secondly we must review the structure of these resulting relationships to determine how many correlated walks, or dynamic factors, they follow. Each of these considerations must be made in best respect to determining the future behaviour of our series of interest, the European Area Gross Domestic Product.

Designing an initial algorithm to search through the possible combinations of data and factors, we discovered our approach can reduce the Mean Squared Error of

the resulting fit, and with greater predictive power (seen in Figure 6.1b). Studying the results further we found that the variables and factors used at each stage where highly dynamic and unexpected. This prompted a further investigation into the structural recovery properties of the inital algorithm to discover that determining the correct data and factors was a task not easily covered by the algorithm (as seen in Table 6.5), despite appealing results from a predictive perspective. Looking to understand and explore possible solutions to exploring both of these structural issues, we extended the algorithm further with scoring functions found within the literature, and different approaches to exploration akin to linear modelling. The large scale simulation study which followed concluded in results best aggregated within Table 6.17 where we found that there is no one direct approach to resolving all of the considerations concurrently, but that some approaches are best for certain situations.

Indeed this work has made clear that approaching these issues from a position of naivety is a difficult task. We have found throughout the literature that often these are considered separately either using prior knowledge of the system or previous studies to form one part of the structure before studying the second part independently. By considering these together, we have found the construction of a dynamic factor model which may have appealing predictive powers, but this does not conclude that the correct model has been selected. This aligns particularly with recent results given in Gao and Tsay (2018) where it was found that information criterion methods would often capture white noise as a factor. Further, the results here show that it most often not the case that the true model is in use, but that some structure has been found to be useful despite not being the most optimal.

Perhaps expanding the scope of the study here may shed further light on this situation. Often Dynamic Factor Models are used with significantly more series than presented here, however we restricted ours due to the computational time required to explore the possible models that could be constructed. Further work

could be done to investigate if a larger dataset would facilitate a greater structural recovery, but considerable time and computing resources would be necessary. It would be interesting to determine more closely the magntiude of difference in predictive power given the encouraging results found within our motivation of EAGDP here. Further, the selections which led to the most encouraging results in Section 6.2 showed great variation between the variables selected and the factors through time, where further study would be necessary to determine if our algorithm can successfully recover a changing state of variables and factors. However, what we have shown is that the correct selection of structure for such a Nowcasting motivation is a perilous task, one which must be approached carefully, but that a naive approach may still approximate the system sufficiently. As in the popular quote from Box (1976), "all models are wrong, some are useful".

6.7 Appendix

6.7.1 Simulation results where $n^* = 500$

Score Function	Dep. Model	Back	Indep. 1	Indep. 2	Indep. 3	Indep. 4	No Series
Mean Squared	1	T	0.00	0.00	0.00	0.00	0.00
Error using	1	F	0.00	0.00	0.00	0.00	0.00
Reference	2	T	0.00	0.00	0.00	0.00	0.00
Series	2	F	0.00	0.00	0.00	0.00	0.00
(Equation (6.6))	3	T	0.00	0.00	0.00	0.00	0.00
	3	F	0.00	0.00	0.00	0.00	0.00
	4	T	0.00	0.00	0.00	0.00	0.00
	4	F	0.00	0.00	0.00	0.00	0.00
Mean Squared	1	T	0.00	0.00	0.00	0.00	0.00
Error using	1	F	0.00	0.00	0.00	0.00	0.00
Reference	2	T	0.00	0.00	0.00	0.00	0.00
Series and	2	F	0.00	0.00	0.00	0.00	0.00
Indicator	3	T	0.00	0.00	0.00	0.00	0.00
Series	3	F	0.00	0.00	0.00	0.00	0.00
(Equation (6.13))	4	T	0.00	0.00	0.00	0.00	0.00
	4	F	0.00	0.00	0.00	0.00	0.00
Information	1	T	0.00	0.00	0.00	0.00	0.00
Criterion	1	F	0.00	0.00	0.00	0.00	0.00
1	2	T	0.01	0.01	0.01	0.00	0.03
(Equation (6.14))	2	F	0.00	0.00	0.00	0.00	0.00
	3	T	0.01	0.01	0.00	0.01	0.01
	3	F	0.00	0.00	0.00	0.00	0.00
	4	T	0.01	0.01	0.01	0.01	0.01
	4	F	0.00	0.00	0.00	0.00	0.00
Information	1	T	0.00	0.00	0.00	0.00	0.00
Criterion	1	F	0.00	0.00	0.00	0.00	0.00
2	2	T	0.01	0.01	0.01	0.00	0.03
(Equation (6.15))	2	F	0.00	0.00	0.00	0.00	0.00
	3	T	0.01	0.00	0.00	0.01	0.01
	3	F	0.00	0.00	0.00	0.00	0.00
	4	T	0.01	0.01	0.01	0.01	0.02
	4	F	0.00	0.00	0.00	0.00	0.00
Information	1	T	0.00	0.00	0.00	0.00	0.00
Criterion	1	F	0.00	0.00	0.00	0.00	0.00
3	2	T	0.01	0.01	0.01	0.00	0.03
(Equation (6.16))	2	F	0.00	0.00	0.00	0.00	0.00
	3	T	0.01	0.01	0.00	0.01	0.01
	3	F	0.00	0.00	0.00	0.00	0.00
	4	T	0.01	0.01	0.01	0.01	0.01
	4	F	0.00	0.00	0.00	0.00	0.00

Table 6.18: The proportion of models converged upon that were exactly correct in three aspects: selection of the amount of dynamic factors, selection in the amount of dependent series and removal of an independent series. Separated by the information criterion used, the dependence model, usage of a backward or forward selection procedure, and the type of independent series included. **Results closest to 1 are favourable in all cases.**

Score Function	Dep. Model	Back	Indep. 1	Indep. 2	Indep. 3	Indep. 4	No Series
Mean Squared	1	T	0.63	0.54	0.53	0.58	1.00
Error using	1	F	0.94	0.88	0.91	0.88	1.00
Reference	2	T	0.45	0.38	0.27	0.33	1.00
Series	2	F	0.88	0.84	0.84	0.84	1.00
(Equation (6.6))	3	T	0.28	0.22	0.17	0.23	1.00
	3	F	0.90	0.87	0.86	0.87	1.00
	4	T	0.24	0.21	0.14	0.20	1.00
	4	F	0.95	0.92	0.94	0.90	1.00
Mean Squared	1	T	0.80	0.79	0.73	0.97	1.00
Error using	1	F	1.00	1.00	1.00	0.99	1.00
Reference	2	T	0.51	0.48	0.45	0.85	1.00
Series and	2	F	0.88	0.88	0.87	0.85	1.00
Indicator	3	T	0.49	0.47	0.46	0.87	1.00
Series	3	F	0.92	0.90	0.90	0.89	1.00
(Equation (6.13))	4	T	0.57	0.47	0.49	0.84	1.00
	4	F	0.96	0.95	0.96	0.93	1.00
Information	1	T	0.94	0.87	0.86	0.88	1.00
Criterion	1	F	0.94	0.88	0.91	0.88	1.00
1	2	T	0.82	0.78	0.66	0.75	1.00
(Equation (6.14))	2	F	0.88	0.84	0.84	0.85	1.00
	3	T	0.83	0.77	0.61	0.80	1.00
	3	F	0.91	0.88	0.88	0.88	1.00
	4	T	0.70	0.64	0.50	0.71	1.00
	4	F	0.96	0.93	0.94	0.91	1.00
Information	1	T	0.94	0.87	0.88	0.88	1.00
Criterion	1	F	0.94	0.88	0.91	0.88	1.00
2	2	T	0.81	0.78	0.63	0.75	1.00
(Equation (6.15))	2	F	0.88	0.84	0.84	0.84	1.00
	3	T	0.83	0.76	0.67	0.80	1.00
	3	F	0.91	0.88	0.88	0.88	1.00
	4	T	0.69	0.59	0.48	0.71	1.00
	4	F	0.96	0.92	0.94	0.91	1.00
Information	1	T	0.94	0.87	0.86	0.88	1.00
Criterion	1	F	0.94	0.88	0.91	0.88	1.00
3	2	T	0.82	0.75	0.65	0.74	1.00
(Equation (6.16))	2	F	0.88	0.84	0.84	0.84	1.00
	3	T	0.86	0.74	0.66	0.80	1.00
	3	F	0.91	0.88	0.88	0.88	1.00
	4	T	0.73	0.63	0.48	0.71	1.00
	4	F	0.96	0.92	0.94	0.91	1.00

Table 6.19: The proportion of models which converged upon the removal of an independent series. Separated by the information criterion used, the dependence model, usage of a backward or forward selection procedure, and the type of independent series included. **Results closest to 1 are favourable in all cases.**

Score Function	Dep. Model	Back	Indep. 1	Indep. 2	Indep. 3	Indep. 4	No Series
Mean Squared	1	T	0.24	0.19	0.20	0.24	0.00
Error using	1	F	0.00	0.00	0.00	0.00	0.00
Reference	2	T	0.01	0.02	0.05	0.02	0.00
Series	2	F	0.00	0.00	0.00	0.00	0.00
(Equation (6.6))	3	T	0.02	0.04	0.07	0.01	0.00
	3	F	0.00	0.00	0.00	0.00	0.00
	4	T	0.01	0.02	0.10	0.00	0.00
	4	F	0.00	0.00	0.00	0.00	0.00
Mean Squared	1	T	0.16	0.16	0.17	0.03	0.01
Error using	1	F	0.00	0.00	0.00	0.00	0.00
Reference	2	T	0.00	0.00	0.00	0.00	0.00
Series and	2	F	0.00	0.00	0.00	0.00	0.00
Indicator	3	T	0.00	0.01	0.01	0.00	0.00
Series	3	F	0.00	0.00	0.00	0.00	0.00
(Equation (6.13))	4	T	0.01	0.02	0.02	0.00	0.00
	4	F	0.00	0.00	0.00	0.00	0.00
Information	1	T	0.00	0.01	0.01	0.00	0.00
Criterion	1	F	0.00	0.00	0.00	0.00	0.00
1	2	T	0.03	0.03	0.03	0.01	0.03
(Equation (6.14))	2	F	0.00	0.00	0.00	0.00	0.00
	3	T	0.03	0.02	0.05	0.03	0.01
	3	F	0.00	0.00	0.00	0.00	0.00
	4	T	0.07	0.06	0.08	0.04	0.02
	4	F	0.00	0.00	0.00	0.00	0.00
Information	1	T	0.00	0.01	0.00	0.00	0.00
Criterion	1	F	0.00	0.00	0.00	0.00	0.00
2	2	T	0.04	0.03	0.04	0.01	0.04
(Equation (6.15))	2	F	0.00	0.00	0.00	0.00	0.00
	3	T	0.03	0.04	0.04	0.03	0.01
	3	F	0.00	0.00	0.00	0.00	0.00
	4	T	0.08	0.07	0.09	0.03	0.02
	4	F	0.00	0.00	0.00	0.00	0.00
Information	1	T	0.00	0.01	0.01	0.00	0.00
Criterion	1	F	0.00	0.00	0.00	0.00	0.00
3	2	T	0.03	0.04	0.04	0.01	0.03
(Equation (6.16))	2	F	0.00	0.00	0.00	0.00	0.00
	3	T	0.02	0.04	0.05	0.03	0.01
	3	F	0.00	0.00	0.00	0.00	0.00
	4	T	0.05	0.08	0.11	0.04	0.02
	4	F	0.00	0.00	0.00	0.00	0.00

Table 6.20: The proportion of models converged upon the correct selection in the amount of dependent series. Separated by the information criterion used, the dependence model, usage of a backward or forward selection procedure, and the type of independent series included. **Results closest to 1 are favourable in all cases.**

Score Function	Dep. Model	Back	Indep. 1	Indep. 2	Indep. 3	Indep. 4	No Series
Mean Squared	1	T	0.68	0.64	0.59	0.70	1.00
Error using	1	F	1.00	1.00	1.00	1.00	1.00
Reference	2	T	0.52	0.44	0.32	0.47	0.90
Series	2	F	0.09	0.09	0.08	0.09	0.10
(Equation (6.6))	3	T	0.35	0.31	0.24	0.34	0.90
	3	F	0.11	0.11	0.09	0.10	0.10
	4	T	0.40	0.36	0.19	0.46	0.66
	4	F	0.03	0.03	0.02	0.03	0.02
Mean Squared	1	T	0.80	0.79	0.72	0.97	0.99
Error using	1	F	1.00	1.00	1.00	1.00	1.00
Reference	2	T	0.52	0.48	0.45	0.84	0.99
Series and	2	F	0.11	0.10	0.10	0.11	0.11
Indicator	3	T	0.49	0.47	0.46	0.86	0.98
Series	3	F	0.08	0.07	0.07	0.07	0.07
(Equation (6.13))	4	T	0.48	0.44	0.45	0.70	0.87
	4	F	0.03	0.03	0.02	0.03	0.02
Information	1	T	1.00	0.99	0.95	1.00	1.00
Criterion	1	F	1.00	1.00	1.00	1.00	1.00
1	2	T	0.77	0.72	0.64	0.73	0.76
(Equation (6.14))	2	F	0.02	0.02	0.01	0.02	0.03
	3	T	0.73	0.68	0.57	0.69	0.73
	3	F	0.02	0.03	0.02	0.02	0.02
	4	T	0.42	0.43	0.35	0.43	0.51
	4	F	0.00	0.00	0.01	0.00	0.00
Information	1	T	1.00	0.98	0.96	1.00	1.00
Criterion	1	F	1.00	1.00	1.00	1.00	1.00
2	2	T	0.77	0.71	0.64	0.72	0.76
(Equation (6.15))	2	F	0.02	0.02	0.01	0.02	0.03
	3	T	0.73	0.68	0.60	0.69	0.73
	3	F	0.02	0.02	0.02	0.02	0.02
	4	T	0.41	0.44	0.31	0.43	0.50
	4	F	0.00	0.00	0.01	0.00	0.00
Information	1	T	1.00	0.98	0.95	1.00	1.00
Criterion	1	F	1.00	1.00	1.00	1.00	1.00
3	2	T	0.77	0.71	0.67	0.73	0.77
(Equation (6.16))	2	F	0.02	0.02	0.01	0.02	0.03
	3	T	0.73	0.64	0.59	0.69	0.73
	3	F	0.02	0.02	0.02	0.02	0.02
	4	T	0.42	0.39	0.36	0.43	0.50
	4	F	0.00	0.00	0.01	0.00	0.00

Table 6.21: The proportion of models converged upon the correct amount of dynamic factors. Separated by the information criterion used, the dependence model, usage of a backward or forward selection procedure, and the type of independent series included. **Results closest to 1 are favourable in all cases.**

Score Function	Dep. Model	Back	Indep. 1	Indep. 2	Indep. 3	Indep. 4	No Series
Mean Squared	1	T	0.80	0.69	0.65	0.81	1.00
Error using	1	F	0.94	0.88	0.91	0.88	1.00
Reference	2	T	0.01	0.04	0.07	0.01	1.00
Series	2	F	0.00	0.00	0.00	0.00	0.00
(Equation (6.6))	3	T	0.04	0.06	0.09	0.05	1.00
	3	F	0.03	0.03	0.02	0.03	0.03
	4	T	0.05	0.07	0.11	0.03	1.00
	4	F	0.01	0.00	0.00	0.00	0.01
Mean Squared	1	T	0.80	0.79	0.73	0.97	1.00
Error using	1	F	1.00	1.00	1.00	0.99	1.00
Reference	2	T	0.18	0.11	0.11	0.76	1.00
Series and	2	F	0.01	0.01	0.01	0.01	0.01
Indicator	3	T	0.22	0.17	0.18	0.85	1.00
Series	3	F	0.00	0.00	0.00	0.00	0.00
(Equation (6.13))	4	T	0.12	0.08	0.11	0.60	1.00
	4	F	0.00	0.00	0.00	0.00	0.00
Information	1	T	0.99	0.99	0.95	1.00	1.00
Criterion	1	F	0.94	0.88	0.91	0.88	1.00
1	2	T	0.63	0.72	0.61	0.52	1.00
(Equation (6.14))	2	F	0.02	0.02	0.02	0.02	0.03
	3	T	0.66	0.69	0.60	0.64	1.00
	3	F	0.02	0.02	0.02	0.02	0.03
	4	T	0.50	0.58	0.65	0.43	1.00
	4	F	0.00	0.00	0.00	0.00	0.00
Information	1	T	0.99	0.98	0.96	1.00	1.00
Criterion	1	F	0.94	0.88	0.91	0.88	1.00
2	2	T	0.64	0.71	0.65	0.53	1.00
(Equation (6.15))	2	F	0.02	0.02	0.02	0.02	0.03
	3	T	0.68	0.66	0.63	0.63	1.00
	3	F	0.02	0.02	0.02	0.02	0.03
	4	T	0.52	0.59	0.59	0.43	1.00
	4	F	0.00	0.00	0.00	0.00	0.00
Information	1	T	0.99	0.98	0.95	1.00	1.00
Criterion	1	F	0.94	0.88	0.91	0.88	1.00
3	2	T	0.64	0.68	0.59	0.52	1.00
(Equation (6.16))	2	F	0.02	0.02	0.02	0.02	0.03
	3	T	0.69	0.67	0.65	0.63	1.00
	3	F	0.02	0.02	0.02	0.02	0.03
	4	T	0.51	0.60	0.61	0.43	1.00
	4	F	0.00	0.00	0.00	0.00	0.00

Table 6.22: The proportion of models which considered the true model as part of their exploration. Separated by the information criterion used, the dependence model, usage of a backward or forward selection procedure, and the type of independent series included. **Results closest to 1 are favourable in all cases.**

Score Function	Dep. Model	Back	Indep. 1	Indep. 2	Indep. 3	Indep. 4	No Series
Mean Squared	1	T	0.48	0.47	0.46	0.47	0.47
Error using	1	F	0.47	0.46	0.47	0.46	0.47
Reference	2	T	0.32	0.32	0.31	0.33	0.31
Series	2	F	4.73	4.28	4.44	4.27	6.87
(Equation (6.6))	3	T	0.31	0.30	0.32	0.31	0.32
	3	F	0.97	0.86	0.89	0.85	1.15
	4	T	0.28	0.26	0.25	0.28	0.29
	4	F	0.38	0.37	0.38	0.36	0.39
Mean Squared	1	T	0.52	0.51	0.52	0.50	0.49
Error using	1	F	0.49	0.49	0.49	0.49	0.49
Reference	2	T	0.42	0.41	0.49	0.43	0.43
Series and	2	F	7.98	8.73	8.59	8.01	12.89
Indicator	3	T	0.42	0.42	0.43	0.42	0.49
Series	3	F	7.55	7.86	6.09	6.23	9.66
(Equation (6.13))	4	T	0.37	0.37	0.38	0.39	0.40
	4	F	0.84	0.83	0.80	0.79	1.23
Information	1	T	0.47	0.47	0.46	0.47	0.47
Criterion	1	F	0.47	0.46	0.47	0.46	0.47
1	2	T	0.34	0.34	0.33	0.35	0.34
(Equation (6.14))	2	F	4.74	4.29	4.45	4.27	6.84
	3	T	0.33	0.33	0.32	0.33	0.33
	3	F	0.98	0.85	0.89	0.86	1.16
	4	T	0.31	0.29	0.27	0.31	0.31
	4	F	0.38	0.37	0.38	0.36	0.39
Information	1	T	0.47	0.46	0.46	0.47	0.47
Criterion	1	F	0.47	0.46	0.47	0.46	0.47
2	2	T	0.35	0.34	0.33	0.35	0.34
(Equation (6.15))	2	F	4.74	4.29	4.45	4.27	6.84
	3	T	0.34	0.33	0.32	0.34	0.33
	3	F	0.98	0.85	0.89	0.85	1.15
	4	T	0.31	0.29	0.25	0.31	0.31
	4	F	0.38	0.37	0.38	0.36	0.39
Information	1	T	0.47	0.46	0.46	0.47	0.47
Criterion	1	F	0.47	0.46	0.47	0.46	0.47
3	2	T	0.35	0.34	0.33	0.35	0.34
(Equation (6.16))	2	F	4.74	4.29	4.45	4.27	6.84
	3	T	0.34	0.32	0.31	0.33	0.33
	3	F	0.98	0.85	0.89	0.85	1.16
	4	T	0.30	0.29	0.27	0.31	0.31
	4	F	0.38	0.37	0.38	0.36	0.39

Table 6.23: Mean squared error of the model fits over the most recent 24 points in the reference series. Separated by the information criterion used, the dependence model, usage of a backward or forward selection procedure, and the type of independent series included. **Results closest to 0 are favourable in all cases.**

Score Function	Dep. Model	Back	Indep. 1	Indep. 2	Indep. 3	Indep. 4	No Series
Mean Squared	1	T	0.47	0.46	0.46	0.47	0.46
Error using	1	F	0.46	0.46	0.46	0.46	0.47
Reference	2	T	0.32	0.31	0.31	0.33	0.31
Series	2	F	2.57	2.26	2.44	2.26	3.67
(Equation (6.6))	3	T	0.31	0.30	0.31	0.31	0.31
	3	F	0.66	0.60	0.64	0.59	0.78
	4	T	0.28	0.26	0.25	0.28	0.29
	4	F	0.33	0.33	0.33	0.33	0.34
Mean Squared	1	T	0.50	0.50	0.50	0.49	0.48
Error using	1	F	0.48	0.48	0.48	0.48	0.48
Reference	2	T	0.42	0.40	0.47	0.42	0.43
Series and	2	F	5.35	5.79	5.79	5.36	8.15
Indicator	3	T	0.43	0.43	0.41	0.43	0.45
Series	3	F	3.14	3.24	3.07	3.14	4.61
(Equation (6.13))	4	T	0.36	0.37	0.37	0.36	0.37
	4	F	0.61	0.61	0.54	0.54	0.73
Information	1	T	0.46	0.46	0.46	0.46	0.47
Criterion	1	F	0.46	0.46	0.46	0.46	0.46
1	2	T	0.33	0.33	0.32	0.34	0.33
(Equation (6.14))	2	F	2.57	2.26	2.45	2.26	3.57
	3	T	0.33	0.32	0.31	0.33	0.33
	3	F	0.67	0.60	0.64	0.59	0.78
	4	T	0.30	0.28	0.26	0.30	0.30
	4	F	0.33	0.33	0.33	0.33	0.34
Information	1	T	0.46	0.46	0.46	0.46	0.47
Criterion	1	F	0.46	0.46	0.46	0.46	0.47
2	2	T	0.33	0.33	0.32	0.34	0.33
(Equation (6.15))	2	F	2.57	2.26	2.45	2.26	3.57
	3	T	0.33	0.32	0.31	0.33	0.33
	3	F	0.67	0.60	0.64	0.59	0.78
	4	T	0.30	0.28	0.26	0.30	0.30
	4	F	0.33	0.33	0.33	0.33	0.34
Information	1	T	0.46	0.46	0.46	0.46	0.47
Criterion	1	F	0.46	0.46	0.46	0.46	0.47
3	2	T	0.33	0.33	0.32	0.34	0.33
(Equation (6.16))	2	F	2.57	2.26	2.45	2.26	3.57
	3	T	0.33	0.32	0.31	0.33	0.33
	3	F	0.67	0.60	0.64	0.59	0.78
	4	T	0.30	0.29	0.27	0.30	0.30
	4	F	0.33	0.33	0.33	0.33	0.34

Table 6.24: Mean squared error of the model fits over all points in the reference series. Separated by the information criterion used, the dependence model, usage of a backward or forward selection procedure, and the type of independent series included. **Results closest to 0 are favourable in all cases.**

Score Function	Backward Selection	Series 2	Series 3	Series 4	Series 5	Independent Model	Independent Incl.	p = 1	p = 2	p = 3	p = 4
Mean Squared	T	0.05	0.60	0.63	0.60	1	0.76	0.05	0.40	0.43	0.11
Error using	T	0.11	0.60	0.58	0.65	2	0.79	0.07	0.36	0.42	0.15
Reference	T	0.25	0.70	0.62	0.78	3	0.86	0.03	0.19	0.49	0.26
Series	T	0.02	0.56	0.62	0.53	4	0.80	0.07	0.46	0.37	0.10
(Equation (6.6))	T	0.03	0.67	0.50	0.72	5	0.00	0.23	0.66	0.11	0.00
	F	0.13	0.12	0.65	0.14	1	0.05	0.97	0.03	0.00	0.00
	F	0.12	0.12	0.62	0.15	2	0.08	0.97	0.03	0.00	0.00
	F	0.14	0.11	0.62	0.14	3	0.06	0.98	0.02	0.00	0.00
	F	0.13	0.11	0.62	0.13	4	0.10	0.97	0.03	0.00	0.00
	F	0.14	0.12	0.68	0.14	5	0.00	0.98	0.02	0.00	0.00
Mean Squared	T	0.47	0.54	0.44	0.53	1	0.43	0.13	0.48	0.25	0.13
Error using	T	0.51	0.58	0.40	0.51	2	0.53	0.10	0.44	0.33	0.12
Reference	T	0.58	0.56	0.43	0.50	3	0.51	0.09	0.45	0.30	0.15
Series and	T	0.47	0.45	0.42	0.54	4	0.16	0.15	0.70	0.11	0.04
Indicator	T	0.61	0.38	0.37	0.52	5	0.00	0.13	0.87	0.00	0.00
Series	F	0.07	0.01	0.89	0.01	1	0.04	0.97	0.03	0.00	0.00
(Equation (6.13))	F	0.06	0.01	0.89	0.01	2	0.05	0.97	0.03	0.00	0.00
	F	0.07	0.01	0.89	0.01	3	0.04	0.97	0.02	0.00	0.00
	F	0.07	0.01	0.87	0.01	4	0.07	0.97	0.03	0.00	0.00
	F	0.07	0.01	0.93	0.01	5	0.00	0.98	0.02	0.00	0.00
Information	T	0.20	0.56	0.55	0.58	1	0.30	0.47	0.42	0.11	0.01
Criterion	T	0.19	0.61	0.46	0.63	2	0.36	0.42	0.43	0.13	0.02
1	T	0.27	0.67	0.51	0.67	3	0.50	0.30	0.35	0.33	0.02
(Equation (6.14))	T	0.15	0.52	0.52	0.50	4	0.29	0.54	0.43	0.03	0.00
	T	0.14	0.69	0.48	0.69	5	0.00	0.49	0.51	0.00	0.00
	F	0.14	0.14	0.66	0.15	1	0.04	1.00	0.00	0.00	0.00
	F	0.14	0.13	0.64	0.15	2	0.07	1.00	0.00	0.00	0.00
	F	0.15	0.13	0.64	0.15	3	0.06	0.99	0.01	0.00	0.00
	F	0.14	0.13	0.63	0.14	4	0.09	1.00	0.00	0.00	0.00
	F	0.15	0.14	0.68	0.15	5	0.00	1.00	0.00	0.00	0.00
Information	T	0.20	0.56	0.56	0.58	1	0.31	0.47	0.41	0.11	0.01
Criterion	T	0.21	0.64	0.50	0.63	2	0.41	0.39	0.44	0.16	0.02
2	T	0.26	0.65	0.53	0.69	3	0.52	0.30	0.31	0.35	0.04
(Equation (6.15))	T	0.15	0.50	0.53	0.49	4	0.29	0.54	0.43	0.02	0.00
	T	0.15	0.68	0.48	0.68	5	0.00	0.50	0.50	0.00	0.00
	F	0.14	0.14	0.66	0.15	1	0.04	1.00	0.00	0.00	0.00
	F	0.14	0.13	0.64	0.15	2	0.08	1.00	0.00	0.00	0.00
	F	0.15	0.13	0.64	0.15	3	0.06	0.99	0.01	0.00	0.00
	F	0.14	0.13	0.63	0.14	4	0.09	1.00	0.00	0.00	0.00
	F	0.15	0.14	0.68	0.15	5	0.00	1.00	0.00	0.00	0.00
Information	T	0.16	0.55	0.55	0.60	1	0.27	0.47	0.42	0.11	0.01
Criterion	T	0.22	0.60	0.53	0.63	2	0.37	0.43	0.39	0.15	0.02
3	T	0.26	0.68	0.55	0.71	3	0.52	0.27	0.36	0.32	0.05
(Equation (6.16))	T	0.15	0.51	0.54	0.49	4	0.29	0.55	0.43	0.03	0.00
	T	0.14	0.68	0.48	0.69	5	0.00	0.50	0.50	0.00	0.00
	F	0.14	0.14	0.66	0.15	1	0.04	1.00	0.00	0.00	0.00
	F	0.14	0.13	0.64	0.15	2	0.08	1.00	0.00	0.00	0.00
	F	0.15	0.13	0.64	0.15	3	0.06	0.99	0.01	0.00	0.00
	F	0.14	0.13	0.63	0.14	4	0.09	1.00	0.00	0.00	0.00
	F	0.15	0.14	0.68	0.15	5	0.00	1.00	0.00	0.00	0.00

Table 6.25: Selection rates for each dependent and independent series and the amount of dynamic factors selected in model 4 using $n^* = 500$. Separated by the scoring function used, usage of a backward of forward selection procedure. **Results closest to 1 for Series 2-5, closest to 0 for independent series, closest to 1 for $p = 2$ and closest 0 for all other p are favourable.**

Strategy	Scoring Function	Backward or Forward	Difference from Mean	Standard Deviation
Removal of Independent Series*	Mean Squared Error using Reference Series and Indicator Series (Equation (6.13))	Forward	0.15	0.05
Selection of correct Dependent Series	Mean Squared Error using Reference Series (Equation (6.6))	Backward	0.04	0.08
Selection of correct amount of Dynamic Factors	Information Criterion 1 (Equation (6.14))	Backward	0.23	0.21
Consideration of correct model structure	Information Criterion 2 (Equation (6.15))	Backward	0.32	0.20
Selection of correct model structure	Information Criterion 1 (Equation (6.14))	Backward	0.01	0.01
Minimisation of Mean Squared Error over previous 24 points in Reference Series	Mean Squared Error using Reference Series (Equation (6.6))	Backward	-0.97	0.08
Minimisation of Mean Squared Error over all points in Reference Series	Mean Squared Error using Reference Series (Equation (6.6))	Backward	-0.52	0.08

Table 6.26: Optimal algorithmic approaches for each desired outcome for $n^* = 500$. Results are averaged across all dependent and independent model choices and the most optimal is selected as that which departs from the mean of the results the most in the favourable direction. The standard deviation of these results is presented for context also. **We remove the case of no independent series here as to not skew the results.*

Chapter 7

Conclusion

Within this thesis we presented work which considered issues of automatically detecting components and structure within data. Chapter 3 used the Evolutionary Wavelet Spectrum as a tool to distinguish between Long Memory and Short Memory with changepoints. The unique properties of this time varying spectrum allowed us to distinguish particular behaviour which is not present in standard spectra. Results were promising in selecting between both models, with particular strength in identifying short memory structures. This work was then applied to Price Inflation and Stock Correlation data to show how different models give varying interpretations.

It is the recovery of hidden components, such as seasonality, which drove the work given in Chapter 5. Here we explored the use of wavelets in the detection and estimation of lower frequency components within high frequency data, which is often identified as trend. Using the unique properties of a wavelet decomposition, and theoretical results found for a particular form of seasonality, we showed that in comparison to the traditional methods for detecting seasonality, we more accurately recover the periodicity within the data.

Following this we studied the use of Dynamic Factor Models in an automatic procedure to select the appropriate predictors and amount of dynamic factors in Chapter 6. With the European Area Gross Domestic Product motivating dataset, we developed an algorithm which would explore the space of factors and predictors

to determine the most suitable combination at each time point. We found encouraging results in terms of lower forecasting errors than a static combination across time. Extending our work further we explored the effectiveness of the algorithm in recovering the true predictors and factors. We found that whilst the algorithm performed well practically, it did not perform well in simulations.

7.0.1 Further Work

To conclude this thesis, we remark on areas of work which could develop what has been achieved here further. Considering the long-memory / changepoint classification, we could explore the effectiveness of the algorithm for different model types. Often there are numerous components additional to those we have studied (stationary mean) which may appear within the data, such as trends or seasonality. We have completed initial work in this direction (Beaulieu et al., 2019) where the models are extended to include trend in the context of sea surface temperatures. This could then be extended further into seasonal components such as those we have studied in other parts of this thesis.

The following chapter identified a low frequency seasonal component within high frequency data. In Official Statistics data, it would be unrealistic to expect a single low frequency seasonal component as there are often several periodicities - daily, weekly, monthly, quarterly, yearly. Thus an area of extension to our periodicity detection would be to include multiple sinusoids as in the case of Harmonic Regression seen in Chapter 4. Indeed this work could be conducted through the use of Wavelet Packets, as mentioned in Chapter 2, which would provide a more overcomplete decomposition which may aid in the discovery of multiple seasonalities. A further avenue that could be explored would be look at minimising the number of components needing to be tested. Reducing the aggregated Bonferroni test down to analysing the single greatest component could be a quicker indication of whether periodicity is present before exploring further.

As has already been seen, the work on automatic selection of components for

a Dynamic Factor Model has much scope for further exploration. Indeed the primary direction to investigate would be to find a scoring function which has been theoretically designed for the problem of simultaneous predictor and factor selection. It would then be beneficial to increase the scope of the simulations conducted. This would allow further investigation into the issues which stymie the current selection procedure, but retain the encouraging results for nowcasting.

Bibliography

Abbas, Z. and Raina, P. (2018). A study on applications of wavelets to data mining. *International Journal of Applied Engineering Research*, 13(12):10886 – 108966.

Acharya, T. and Chakrabarti, C. (2006). A survey on lifting-based discrete wavelet transform architectures. *Journal of VLSI signal processing systems for signal, image and video technology*, 42(3):321–339.

Ahdesmaki, M., Fokianos, K., and Strimmer., K. (2019). *GeneCycle: Identification of Periodically Expressed Genes*. R package version 1.1.4.

Alpert, B. K. (1993). A class of bases in L2 for the sparse representations of integral operators. *SIAM J. Math. Anal.*, 24(1):246–262.

Angelini, E., Camba-Mendez, G., Giannone, D., Reichlin, L., and Rünstler, G. (2011). Short-term forecasts of Euro Area GDP growth. *The Econometrics Journal*, 14(1).

Baffigi, A., Golinelli, R., and Parigi, G. (2004). Bridge models to forecast the Euro Area GDP. *International Journal of Forecasting*, 20(3):447 – 460.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Bai, J. and Ng, S. (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics*, 146(2):304 – 317. Honoring the research contributions of Charles R. Nelson.

Banerjee, A. and Marcellino, M. (2006). Are there any reliable leading indicators for US inflation and GDP growth? *International Journal of Forecasting*, 22(1):137 – 151.

Banerjee, A. and Marcellino, M. (2009). Factor-augmented error correction models. *In: Castle and Shephard, The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*, pages 227–254.

Banerjee, A., Marcellino, M., and Masten, I. (2005). Leading indicators for euro-area inflation and GDP growth. *Oxford Bulletin of Economics and Statistics*, 67(s1):785–813.

Banerjee, A., Marcellino, M., and Masten, I. (2008). Chapter 4: Forecasting macroeconomic variables using diffusion indexes in short samples with structural change. In *Forecasting in the Presence of Structural Breaks and Model Uncertainty*, pages 149–194. Emerald Group Publishing Limited.

Bartlett, M. S. (1967). Some remarks on the analysis of time-series. *Biometrika*, 54(1/2):25–38.

Barton, H. C. J. (1941). Adjustment for seasonal variation. *Federal Reserve Bulletin*, pages 518–528.

Beaulieu, C., Killick, R., Ireland, D., and Norwood, B. (2019). Considering long-memory when testing for changepoints in surface temperature: A classification approach based on the time-varying spectrum. *Environmetrics*.

Beck, N. (1989). Estimating dynamic models using kalman filtering. *Political Analysis*, 1:121156.

Bell, W. R. and Hillmer, S. C. (1984). Issues involved with the seasonal adjustment of economic time series. *Journal of Business & Economic Statistics*, 2(4):291–320.

Benedetto, J. J. and Pfander, G. E. (2002). Periodic wavelet transforms and periodicity detection. *SIAM Journal on Applied Mathematics*, 62(4):1329–1368.

Beran, J., J., B. R., and Ocker, D. (1998). On unified model selection for stationary and nonstationary short and long-memory autoregressive processes. *Biometrika*, 85(4):921–934.

Bernanke, B. S., Boivin, J., and Eliasz, P. (2005). Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach. *The Quarterly Journal of Economics*, 120(1):387–422.

Bertram, P., Kruse, R., and Sibbertsen, P. (2013). Fractional integration versus level shifts: the case of realized asset correlations. *Statistical Papers*, 54(4):977–991.

Beveridge, W. H. (1921). Weather and harvest cycles. *The Economic Journal*, 31(124):429–452.

Bianchi, M. (1996). A comparison of methods for seasonal adjustment of the monetary aggregates. Bank of England Working Paper 44, Bank of England.

Blume, M. and Ballard, D. R. (1997). Image annotation based on learning vector quantization and localized haar wavelet transform features. In *Applications and Science of Artificial Neural Networks III*, volume 3077, pages 181–190. International Society for Optics and Photonics.

Boivin, J. and Ng, S. (2006). Are more data always better for factor analysis? *Journal of Econometrics*, 132(1):169–194. Common Features.

Bonferroni, C. (1936). Teoria statistica delle classi e calcolo delle probabilita. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commericiali di Firenze*, 8:3–62.

Bongard, J. (1960). Some remarks on moving averages.

Box, G. E. and Jenkins, G. M. (1976). Time series analysis: Forecasting and control. In *Holden-Day series in time series analysis*. Holden-Day.

Box, G. E. P. (1976). Science and statistics. *Journal of the American Statistical Association*, 71(356):791–799.

Bruce, L. M., Koger, C. H., and Li, J. (2002). Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction. *IEEE Transactions on Geoscience and Remote Sensing*, 40(10):2331–2338.

Bui, T. D. and Guangyi Chen (1998). Translation-invariant denoising using multiwavelets. *IEEE Transactions on Signal Processing*, 46(12):3414–3420.

Burman, J. P. (1965). Moving seasonal adjustment of economic time series. *Journal of the Royal Statistical Society. Series A (General)*, 128(4):534–558.

Burman, J. P. (1980). Seasonal adjustment by signal extraction. *Journal of the Royal Statistical Society. Series A (General)*, 143(3):321–337.

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.

Calder, C. A. (2007). Dynamic factor process convolution models for multivariate space–time data with application to air quality assessment. *Environmental and Ecological Statistics*, 14(3):229–247.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1(2):245–276.

Chen, G., Xie, W., and Zhao, Y. (2013a). Wavelet-based denoising: A brief review. In *2013 Fourth International Conference on Intelligent Control and Information Processing (ICICIP)*, pages 570–574.

Chen, Q., van Dam, T., Sneeuw, N., Collilieux, X., Weigelt, M., and Rebischung, P. (2013b). Singular spectrum analysis for modeling seasonal signals from GPS time series. *Journal of Geodynamics*, 72:25 – 35.

Cheung, Y. L. (2018). Sieve estimation of time-varying factor loadings: Estimating the conditional CAPM. *Available at SSRN 3122671*.

Chiriac, R. and Voev, V. (2011). Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics*, 26(6):922–947.

Cho, H. and Fryzlewicz, P. (2012). Multiscale and multilevel technique for consistent segmentation of nonstationary time series. *Statistica Sinica*, 22(1):207–229.

Cleveland, R. B., Cleveland, W. S., McRae, J. E., and Terpenning, I. (1990). STL: A seasonal-trend decomposition procedure based on Loess. *Journal of Official Statistics*, 6(1):3–73.

Cleveland, W. P. and Tiao, G. C. (1976). Decomposition of seasonal time series: A model for the census X-11 program. *Journal of the American Statistical Association*, 71(355):581–587.

Cleveland, W. S., Dunn, D. M., and Terpenning, I. J. (1979). *SABL: A Resistant Seasonal Adjustment Procedure with Graphical Methods for Interpretation and Diagnosis*, pages 201–241. NBER.

Coifman, R. R. and Donoho, D. L. (1995). *Translation-Invariant De-Noising*, pages 125–150. Springer New York, New York, NY.

Coifman, R. R. and Wickerhauser, M. V. (1992). Entropy-based algorithms for best basis selection. *IEEE Transactions on information theory*, 38(2):713–718.

Constantine, W. and Percival, D. (2016). *fractal: Fractal Time Series Modeling and Analysis*. R package version 2.0-1.

Creal, D., Schwaab, B., Koopman, S. J., and Lucas, A. (2014). Observation-driven mixed-measurement dynamic factor models with an application to credit risk. *The Review of Economics and Statistics*, 96(5):898–915.

Dagum, E. B. (1980). *The X-II-ARIMA seasonal adjustment method*. Statistics Canada, Seasonal Adjustment and Time Series Staff.

Dahlhaus, R. (2012). Locally stationary processes. In *Handbook of statistics*, volume 30, pages 351–413. Elsevier.

Dansena, P. and Dewangan, O. (2015). Comparision and analysis of wavelet based image denoising techniques. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, 4(2).

Daubechies, I. (1992). *Ten lectures on wavelets*, volume 61. SIAM.

Davis, H. T. (1941). *The Analysis of Economic Time Series*. Principia press, inc.

Diebold, F. X. and Inoue, A. (2001). Long memory and regime switching. *Journal of Econometrics*, 105(1):131 – 159. Forecasting and empirical methods in finance and macroeconomics.

Diebold, F. X. and Rudebusch, G. D. (1994). Measuring business cycles: A modern perspective. Working Paper 4643, National Bureau of Economic Research.

Diron, M. (2008). Short-term forecasts of Euro Area real GDP growth: an assessment of real-time performance based on vintage data. *Journal of Forecasting*, 27(5):371–390.

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. and Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 90(432):1200–1224.

Du, P., Kibbe, W. A., and Lin, S. M. (2006). Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. *Bioinformatics*, 22(17):2059–2065.

Durbin, J. and Koopman, S. (2012). *Time Series Analysis by State Space Methods: Second Edition*. Oxford Statistical Science Series. OUP Oxford.

Durbin, J. and Murphy, M. J. (1975). Seasonal adjustment based on a mixed additive–multiplicative model. *Journal of the Royal Statistical Society. Series A (General)*, 138(3):385–410.

Eickmeier, S. and Ziegler, C. (2008). How successful are dynamic factor models at forecasting output and inflation? A meta-analytic approach. *Journal of Forecasting*, 27(3):237–265.

Elfeky, M. G., Aref, W. G., and Elmagarmid, A. K. (2005). Periodicity detection in time series databases. *IEEE Transactions on Knowledge and Data Engineering*, 17(7):875–887.

Eurostat (2016). *Euro Area and European Union GDP flash estimates at 30 days, Eurostat Statistical working papers*.

Federal Reserve Board (1922). Index of production in selected basic industries. *Federal Reserve Bulletin*, pages 1414–1421.

Feng, Y., Wen, M., Zhang, J., Ji, F., and Ning, G. (2016). Sum of arbitrarily correlated gamma random variables with unequal parameters and its application in wireless communications. In *Computing, Networking and Communications (ICNC), 2016 International Conference on*, pages 1–5. IEEE.

Fernandez-Villaverde, J. (2009). The econometrics of DSGE models. Working Paper 14677, National Bureau of Economic Research.

Findley, D. F. (2005). Some recent developments and directions in seasonal adjustment. *Journal of official statistics*, 21(2):343.

Findley, D. F., Monsell, B. C., Bell, W. R., Otto, M. C., and Chen, B. (1998). New capabilities and methods of the X-12-ARIMA seasonal adjustment program. *Journal of Business & Economic Statistics*, 16(2):127–152.

Fisher, R. A. (1929). Tests of significance in harmonic analysis. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 125(796):54–59.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The generalized dynamic-factor model: Identification and estimation. *The Review of Economics and Statistics*, 82(4):540–554.

Fournier-Viger, P., Lin, J. C., Kiran, R. U., Koh, Y. S., and Thomas, R. (2017). A survey of sequential pattern mining.

Fryzlewicz, P. and Nason, G. P. (2006). Haar-fisz estimation of evolutionary wavelet spectra. *JRSSB*, 68(4):611–634.

Fryzlewicz, P. and Ombao, H. (2009). Consistent classification of nonstationary time series using stochastic wavelet representations. *Journal of the American Statistical Association*, 104(485):299–312.

Gao, H. and Bruce, A. G. (1997). Waveshrink with firm shrinkage. *Statistica Sinica*, 7(4):855–874.

Gao, Z. and Tsay, R. S. (2018). Structural-factor modeling of high-dimensional time series: Another look at approximate factor models with diverging eigenvalues. *arXiv preprint arXiv:1808.07932*.

Gardner, E. S. (1985). Exponential smoothing: The state of the art. *Journal of Forecasting*, 4(1):1–28.

Gardner, E. S. (2006). Exponential smoothing: The state of the art - Part II. *International Journal of Forecasting*, 22(4):637 – 666.

Geary, R. C. (1956). A Study in the Analysis of Stationary Time Series. *The Economic Journal*, 66(262):327–330.

Geronimo, J., Hardin, D., and Massopust, P. (1994). Fractal functions and wavelet expansions based on several scaling functions. *Journal of Approximation Theory*, 78(3):373 – 401.

Geweke, J. (1977). The dynamic factor analysis of economic time series models. *Latent variables in socio-economic models*, pages 365–383.

Giannone, D., Reichlin, L., and Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, 55(4):665–676.

Girardi, A., Gayer, C., and Reuter, A. (2016). The role of survey data in nowcasting Euro Area GDP growth. *Journal of Forecasting*, 35(5):400–418.

Golyandina, N. and Shlemov, A. (2015). Variations of singular spectrum analysis for separability improvement: non-orthogonal decompositions of time series. *Statistics and its Interface*, 8(3):277–294.

Grabocka, J., Nanopoulos, A., and Schmidt-Thieme, L. (2012). Invariant time-series classification. In Flach, P. A., De Bie, T., and Cristianini, N., editors, *Machine Learning and Knowledge Discovery in Databases*, volume 7524 of *Lecture Notes in Computer Science*, pages 725–740. Springer Berlin Heidelberg.

Granger, C. W. and Ding, Z. (1996). Varieties of long memory models. *Journal of Econometrics*, 73(1):61 – 77.

Granger, C. W. and Hyung, N. (2004). Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns. *Journal of Empirical Finance*, 11(3):399–421.

Granger, C. W. J., Hatanaka, M., et al. (1964). Spectral analysis of economic time series. *Spectral analysis of economic time series.*

Grasmann, P. and Keereman, F. (2001). An indicator-based short-term forecast for quarterly GDP in the Euro Area. European Economy - Economic Papers 2008 - 2015 154, Directorate General Economic and Financial Affairs (DG ECFIN), European Commission.

Grudkowska, S. (2016). *JDemetra+ Reference Manual*. Narodowy Bank Polski, version 2.1 edition.

Haar, A. (1910). Zur theorie der orthogonalen funktionensysteme. *Mathematische Annalen*, 69(3):331–371.

Hamilton, J. D. (1994). Chapter 50: State-space models. In Engle, R. and McFadden, D., editors, *Handbook of Econometrics*, volume 4 of *Handbook of Econometrics*, pages 3039 – 3080. Elsevier.

Hannan, E. J. (1964). The estimation of a changing seasonal pattern. *Journal of the American Statistical Association*, 59(308):1063–1077.

Harvey, A. C. (1990). *Forecasting, structural time series models, and the Kalman filter*. Cambridge.

Havenner, A. and Swamy, P. (1981). A random coefficient approach to seasonal adjustment of economic time series. *Journal of Econometrics*, 15(2):177 – 209.

He, L., Huang, G., Zeng, G., and Lu, H. (2008). Wavelet-based multiresolution analysis for data cleaning and its application to water quality management systems. *Expert Systems with Applications*, 35(3):1301 – 1310.

Henderson, R. (1916). Note on graduation by adjusted averages. *American Society of Actuaries*, 17:43–48.

Henshaw, R. C. (1966). Application of the general linear model to seasonal adjustment of economic time series. *Econometrica*, 34(2):381–395.

Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, 20(1):5 – 10.

Hood, C. C. (2002). Comparison of time series characteristics for seasonal adjustments from SEATS and X-12-ARIMA. *ASA proceedings, business and economic statistics section, alexandria, VA: ASA*.

Hoyle, R. H. and Duvall, J. L. (2004). Determining the number of factors in exploratory and confirmatory factor analysis. *Handbook of quantitative methodology for the social sciences*, pages 301–315.

Huyot, G., Chiu, K., Higginson, J., and Gait, N. (1986). Analysis of revisions in the seasonal adjustment of data using X-11-ARIMA model-based filters. *International Journal of Forecasting*, 2(2):217–229.

Hylleberg, S. (1986). *Seasonality in Regression*. Elsevier, 1 edition.

Hyndman, R. J. (2017). *forecast: Forecasting functions for time series and linear models*. R package version 8.0.

Hyndman, R. J. and Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, 26(3):1–22.

Indyk, P., Koudas, N., and Muthukrishnan, S. (2000). Identifying representative trends in massive time series data sets using sketches. In *Proceedings of the 26th International Conference on Very Large Data Bases*, VLDB 2000, pages 363–372, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

INSEE (2017). Les naissances en 2016 - tableaux de sries longues. `https://insee.fr/fr/statistiques/2904770`. Accessed: 2018-09-27.

Jain, R. K. (2012). The quality of seasonal adjustment: An empirical comparison of various seasonal adjustment methods.

Janacek, G., Bagnall, A., and Powell, M. (2005). A likelihood ratio distance measure for the similarity between the Fourier transform of time series. In Ho, T., Cheung, D., and Liu, H., editors, *Advances in Knowledge Discovery and*

Data Mining, volume 3518 of *Lecture Notes in Computer Science*, pages 737–743. Springer Berlin Heidelberg.

Jensen, M. J. (2000). An alternative maximum likelihood estimator of long-memory processes using compactly supported wavelets. *Journal of Economic Dynamics and Control*, 24(3):361–387.

Kaewpijit, S., Le Moigne, J., and El-Ghazawi, T. (2003). Automatic reduction of hyperspectral imagery using wavelet spectral analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 41(4):863–871.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45.

Kanjilal, P. P., Bhattacharya, J., and Saha, G. (1999). Robust method for periodicity detection and characterization of irregular cyclical series in terms of embedded periodic components. *Phys. Rev. E*, 59:4013–4025.

Karoui, A. and Vaillancourt, R. (1994). Families of biorthogonal wavelets. *Computers & Mathematics with Applications*, 28(4):25–39.

Kenny, P. B. and Durbin, J. (1982). Local trend estimation and seasonal adjustment of economic and social time series. *Journal of the Royal Statistical Society. Series A (General)*, 145(1):1–41.

Killick, R. and Eckley, I. A. (2014). changepoint: An R package for changepoint analysis. *Journal of statistical software*, 58(3):1–19.

Killick, R., Eckley, I. A., and Jonathan, P. (2013). A wavelet-based approach for detecting changes in second order structure within nonstationary time series. *Electron. J. Statist.*, 7:1167–1183.

Killick, R., Fearnhead, P., and Eckley, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *J. Am. Stat. Assoc.*, 107(500):1590–1598.

Kruskal, W. H. and Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260):583–621.

Krzemieniewska, K., Eckley, I., and Fearnhead, P. (2014). Classification of non-stationary time series. *Stat*, 3(1):144–157.

Kuznets, S. (1933). *The Statistical Measurement*, pages 23–40. NBER.

Ladiray, D. and Quenneville, B. (2012). *Seasonal adjustment with the X-11 method*, volume 158. Springer Science & Business Media.

Levin, A. T. and Piger, J. M. (2004). Is inflation persistence intrinsic in industrial economies? Working Paper Series 0334, European Central Bank.

Li, M., Koopman, S. J., Lit, R., and Petrova, D. (forthcoming). Long term forecasting of El Nino events via dynamic factor simulations. *Journal of Econometrics*.

Li, T., Li, Q., Zhu, S., and Ogihara, M. (2002). A survey on wavelet applications in data mining. *ACM SIGKDD Explorations Newsletter*, 4(2):49–68.

Litterman, R. B. (1983). A random walk, markov model for the distribution of time series. *Journal of Business & Economic Statistics*, 1(2):169–173.

Ljung, G. M. and Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303.

Lytras, D. P., Feldpausch, R. M., and Bell, W. R. (2007). Determining seasonality: a comparison of diagnostics from X-12-ARIMA. *US Census Bureau*.

Macaulay, F. R. (1931a). *The Smoothing of Economic Time Series, Curve Fitting and Graduation*, pages 31–42. NBER.

Macaulay, F. R. (1931b). The Whittaker-Henderson method of graduation. In *The Smoothing of Time Series*, pages 89–99. NBER.

Mallat, S. G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693.

Marcellino, M., Stock, J. H., and Watson, M. W. (2003). Macroeconomic forecasting in the Euro Area: Country specific versus area-wide information. *European Economic Review*, 47(1):1 – 18.

McElroy, T. (2010). A nonlinear algorithm for seasonal adjustment in multiplicative component decompositions. *Studies in Nonlinear Dynamics & Econometrics*, 14.

McElroy, T. and Holan, S. (2009). Using spectral peaks to detect seasonality. In *2009 Proceedings of the Federal Conference on Statistical Methodology*, volume 1292.

Mendershausen, H. (1939). Eliminating changing seasonals by multiple regression analysis. *The Review of Economics and Statistics*, 21(4):171–177.

Michie, D., Spiegelhalter, D. J., Taylor, C. C., and Campbell, J. (1994). *Machine Learning, Neural and Statistical Classification*. Ellis Horwood, Upper Saddle River, NJ, USA.

Misiti, M., Misiti, Y., Oppenheim, G., and Poggi, J. (2007). Clustering signals using wavelets. In Sandoval, F., Prieto, A., Cabestany, J., and Graña, M., editors, *Computational and Ambient Intelligence*, pages 514–521, Berlin, Heidelberg. Springer Berlin Heidelberg.

Misiti, M., Misiti, Y., Oppenheim, G., and Poggi, J. (2018). *Wavelet Toolbox, Reference*. MATLAB.

Molz, F. J., Liu, H. H., and Szulga, J. (1997). Fractional brownian motion and fractional gaussian noise in subsurface hydrology: A review, presentation of fundamental properties, and extensions. *Water Resources Research*, 33(10):2273–2286.

Nason, G. P. (2010). *Wavelet methods in statistics with R*. Springer Science & Business Media.

Nason, G. P. (2013). A test for second-order stationarity and approximate confidence intervals for localized autocovariances for locally stationary time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(5):879–904.

Nason, G. P. (2016a). *locits: Tests of stationarity and localized autocovariance*. R package version 1.7.1.

Nason, G. P. (2016b). *wavethresh: Wavelets Statistics and Transforms*. R package version 4.6.8.

Nason, G. P. and Savchev, D. (2014). White noise testing using wavelets. *Stat*, 3(1):351–362.

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.

Negro, M. D. and Otrok, C. (2008). Dynamic factor models with time-varying parameters: measuring changes in international business cycles. Staff Reports 326, Federal Reserve Bank of New York.

O'Gorman, T. W. (1982). On the design of seasonal adjustment methods using linear programming techniques. *Journal of the American Statistical Association*, 77(380):739–742.

Onatski, A. (2009). Testing hypotheses about the number of factors in large factor models. *Econometrica*, 77(5):1447–1479.

Otrok, C. and Whiteman, C. (1996). Bayesian Leading Indicators: Measuring and Predicting Economic Conditions in Iowa. Working Papers 96-14, University of Iowa, Department of Economics.

Panda, R., Khobragade, P. S., Jambhule, P. D., Jengthe, S. N., Pal, P. R., and Gandhi, T. K. (2010). Classification of EEG signal using wavelet transform and support vector machine for epileptic seizure diction. In *2010 International Conference on Systems in Medicine and Biology*, pages 405–408.

Park, J. W. and Kim, D. Y. (2018). Standard time estimation of manual tasks via similarity measure of unequal scale time series. *IEEE Transactions on Human-Machine Systems*, 48(3):241–251.

Pelger, M. and Xiong, R. (2018). State-varying factor models of large dimensions. *Available at SSRN 3109314*.

Penzer, J. (2006). Diagnosing seasonal shifts in time series using state space models. *Statistical Methodology*, 3(3):193 – 210.

Persons, W. M. (1919). Indices of business conditions. *Review of Economics and Statistics*, 1:5–107.

Pesaran, M. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74(4):967–1012. cited By 946.

Petris, G. (2010). An R package for dynamic linear models. *Journal of Statistical Software*, 36(12):1–16.

Pivetta, F. and Reis, R. (2007). The persistence of inflation in the United States. *Journal of Economic dynamics and control*, 31(4):1326–1358.

Priestley, M. B. and Rao, T. S. (1969). A test for non-stationarity of time-series. *Journal of the Royal Statistical Society. Series B*, 31(1):140–149.

Puech, T. and Boussard, M. (2019). A fully automated periodicity detection in time series.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ram, N., Brose, A., and Molenaar, P. C. M. (2013). *Dynamic factor analysis: Modeling person-specific process.*, chapter 21, pages 441–457. Oxford library of psychology. Oxford University Press, New York, NY, US.

Regnier-Loilier, A. and Divinagracia, E. (2010). Changes in the seasonality of births in france from 1975 to the present. *Population (English Edition, 2002-)*, 65(1):145–185.

Reinsel, D., Venkatraman, A., J.F., G., and Rydning, J. (2019). The EMEA datasphere: Rapid growth and migration to the edge.

Rünstler, G. (2016). On the design of data sets for forecasting with dynamic factor models. In *Dynamic Factor Models*, pages 629–662. Emerald Group Publishing Limited.

Rünstler, G. and Sédillot, F. (2003). Short-term estimates of Euro Area real GDP by means of monthly data. ECB Working Paper 276, European Central Bank (ECB), Frankfurt a. M.

Sani, S. N., Roscoe, B. J., Cai, X., Alamgir, Z. M., Wang, R., and Robinson, C. J. (2013). Wavelet decomposition to detect periodic signals in head accelerometry measures. In *2013 39th Annual Northeast Bioengineering Conference*, pages 209–210.

Sargent, T. and Sims, C. (1977). Business cycle modeling without pretending to have too much a priori economic theory. Working Papers 55, Federal Reserve Bank of Minneapolis.

Savchev, D. and Nason, G. P. (2018). *hwwntest: Tests of White Noise using Wavelets*. R package version 1.3.1.

Schaidnagel, M. and Laux, F. (2013). Time series predition with automated periodicity detection. *International Journal on Advances in Systems and Measurements*, 6(3 & 4):394–404.

Sekine, M. and Maeno, K. (2011). Non-contact heart rate detection using periodic variation in doppler frequency. In *2011 IEEE Sensors Applications Symposium*, pages 318–322.

Senhadji, L., Carrault, G., Bellanger, J. J., and Passariello, G. (1995). Comparing wavelet transforms for recognizing cardiac patterns. *IEEE Engineering in Medicine and Biology Magazine*, 14(2):167–173.

Shiskin, J. (1942). A new multiplicative seasonal index. *Journal of the American Statistical Association*, 37(220):507–516.

Shiskin, J. and Eisenpress, H. (1957). Seasonal adjustments by electronic computer methods. *Journal of the American Statistical Association*, 52(280):415–449.

Singh, B. N. and Tiwari, A. K. (2006). Optimal selection of wavelet basis function applied to ECG signal denoising. *Digital Signal Processing*, 16(3):275 – 287.

Soukup, R. J. and Findley, D. F. (1999). On the spectrum diagnostics used by X-12-ARIMA to indicate the presence of trading day effects after modeling or adjustment. In *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, pages 144–149.

Spurr, W. A. (1937). A graphic method of measuring seasonal variation. *Journal of the American Statistical Association*, 32(198):281–289.

Spurr, W. A. (1940). A graphic short cut to the moving average method of measuring seasonality. *Journal of the American Statistical Association*, 35(212):667–670.

Stachura, M. (2014). Detecting seasonality via wavelet methods. *Studia Ekonomiczne*, 207:223–232.

Starica, C. and Granger, C. (2005). Nonstationarities in stock returns. *The Review of Economics and Statistics*, 87(3):503–522.

Stock, J. H. and Watson, M. (2009). Forecasting in dynamic factor models subject to structural instability. *The Methodology and Practice of Econometrics. A Festschrift in Honour of David F. Hendry*, 173:205.

Stock, J. H. and Watson, M. W. (1988). A probability model of the coincident economic indicators. Working Paper 2772, National Bureau of Economic Research.

Su, L. and Wang, X. (2017). On time-varying factor models: Estimation and testing. *Journal of Econometrics*, 198(1):84 – 101.

Svetunkov, I. and Kourentzes, N. (2016). Complex exponential smoothing. Working paper, Lancaster University Management School.

Sweldens, W. (1998). The lifting scheme: A construction of second generation wavelets. *SIAM J. Math. Anal.*, 29(2):511–546.

Szatmary, K., Gal, J., and Kiss, L. L. (1996). Application of wavelet analysis in variable star research. II. The semiregular star V Bootis. *Astronomy and Astrophysics*, 308:791–798.

Szatmary, K., Vinko, J., and Gal, J. (1994). Application of wavelet analysis in variable star research. I. Properties of the wavelet map of simulated variable star light curves. *Astronomy and Astrophysics Suppl.*, 108:377–394.

To, A. C., Moore, J. R., and Glaser, S. D. (2009). Wavelet denoising techniques with applications to experimental geophysical data. *Signal Processing*, 89(2):144 – 160.

Tripodis, Y. and Penzer, J. (2004). Periodic time series models: a structural approach.

Veenstra, J. Q. (2012). *Persistence and Anti-persistence: Theory and Software*. PhD thesis, Western University.

Vidakovic, B. (2009). *Statistical modeling by wavelets*, volume 503. John Wiley & Sons.

Vlachos, M., Lin, J., Keogh, E., and Gunopulos, D. (2003). A wavelet-based anytime algorithm for k-means clustering of time series. In *In Proc. Workshop on Clustering High Dimensionality Data and Its Applications*, pages 23–30.

Vlachos, M., Yu, P., and Castelli, V. (2005). *On Periodicity Detection and Structural Periodic Similarity*, pages 449–460. Society for Industrial and Applied Mathematics.

Watson, M. W. and Engle, R. F. (1983). Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models. *Journal of Econometrics*, 23(3):385 – 400.

Webel, K. and Ollech, D. (2018). An overall seasonality test based on recursive feature elimination in conditional random forests. In *International Conference on Time Series and Forecasting Proceedings of Papers*, pages 20–31.

Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, 6(3):324–342.

Yang, R. (2019). R implementation of arser algorithm. `https://github.com/gangwug/MetaCycle/blob/master/R/ARS.R`.

Yang, R. and Su, Z. (2010). Analyzing circadian expression data by harmonic regression based on autoregressive spectral estimation. *Bioinformatics*, 26(12):i168–i174.

Yau, C. Y. and Davis, R. A. (2012). Likelihood inference for discriminating between long-memory and change-point models. *Journal of Time Series Analysis*, 33(4):649–664.

Young, T. (1992). Seasonal adjusting of flow of fund matrices. Internal Mimeo.

Zhang, H., Nussbaum, M. A., and Agnew, M. J. (2015). A time-frequency approach to estimate critical time intervals in postural control. *Computer Methods in Biomechanics and Biomedical Engineering*, 18(15):1693–1703. PMID: 25105745.