Spectral Domain Approaches to Covariance Estimation with Applications in Asset Pricing and Portfolio Management

Jurgen Wervers - 542079

July 23, 2021

Master Thesis in Quantitative Finance

Erasmus School of Economics

ERASMUS UNIVERSITY ROTTERDAM

Supervisor: dr. Rutger-Jan Lange

Second assessor: dr. Annika Camehl

Abstract

In this thesis, we explore the advantages of estimating covariances of time series in the spectral domain. Estimation of covariance is of key importance in fields such as asset pricing and portfolio management. We propose a novel Fourier transform based covariance measure and show its applications to asset pricing and portfolio management, along with other spectral methods based on wavelet transforms. We perform a simulation study to compare the spectral methods with classical econometric methods and apply the covariance measures to systematic risk and the creation of minimum variance portfolios. We find that spectral methods are beneficial in both portfolio management and asset pricing, the benefits in our application to asset pricing seem to be larger. Using the spectral methods, we conclude that there is evidence for a long-run risk premium as well as a short term volatility feedback or leverage effect in the stock market.

Keywords: spectral domain, covariance, Fourier transform, wavelet transform, long-run, risk premium

Contents

1 Introduction

Correctly estimating covariance is crucial in fields in econometrics such as asset pricing and portfolio management. In asset pricing, one of the most used models is the capital asset pricing model (CAPM) of Sharpe (1964). This model relates systematic risk to returns by estimating the covariance between the individual assets and the market. The CAPM has come under a lot of critique, due to evidence that the relationship between the classical interpretation of systematic risk of the model and returns of assets does not hold (Bali et al., 2017; Frazzini & Pedersen, 2014; Jensen et al., 1972; Reinganum, 1981). However, by using the wavelet transform to expand systematic risk into the spectral domain, Gençay et al. (2005) showed that the relationship between systematic risk and returns became much stronger for greater time horizons. In portfolio management, one of the most used frameworks is the framework of Markowitz (1952), which aims to create optimal mean-variance portfolios. In this framework, correctly estimating covariances between assets is crucial and is currently still one of the main challenges (Ledoit & Wolf, 2020). Low correlations between assets are advantageous in the construction of portfolios. This advantage is caused by the fact that a low correlation implies that diversification is more effective. As Conlon et al. (2018) have shown by using the wavelet transform, assets tend to have a higher correlation in the long-run than what is implied by most classical econometric techniques.

This thesis aims to research the advantages of estimating the covariance of time series in the spectral domain. To answer this question, we propose a covariance measure based on the Fourier transform. This covariance measure is a completely model free estimator for the covariance of a subset of the spectral domain. We also consider the wavelet-based covariance measure that was introduced by Whitcher et al. (2000) as well as the long-run wavelet-based covariance measure of Conlon et al. (2018). The advantage of using a wavelet transform instead of a Fourier transform is that the projections of Fourier transform are global. While when using a wavelet transform, the projections can essentially be seen as local (Ramsey, 1999). This property of local projections allows us to localise specific frequency components in time, since we do not assume that if a mapping exists, it exists everywhere. The advantage of using the

Fourier transform over the wavelet transform is that it allows for more flexibility, since we do not have to work in predefined wavelet scales, but can specify the frequencies of interest ourselves.

To test the performance of the proposed Fourier-based covariance measure, the wavelet-based covariance measure of Whitcher et al. (2000) and the long-run wavelet-based covariance measure of Conlon et al. (2018), we first set up a simulation study. In the first part of the simulation study, we use four different data generating data-generating processes (DGPs) to select which wavelet from the Daubechies family performs best when estimating correlation for the wavelet-based covariance of Whitcher et al. (2000) and the long-run wavelet-based covariance of Conlon et al. (2018). Next, we use the same DGPs with varied parameter settings for the correlation to compare the performances of the Fourier-based covariance, the wavelet-based covariance of Whitcher et al. (2000) and the long-run wavelet-based covariance of Conlon et al. (2018) with the Dynamic Equicorrelation (DECO) algorithm as proposed by Engle and Kelly (2012) and the sample correlation of the generated time series. After the simulation study, we study two applications of the spectral covariance measures. The first application highlights the advantages of using the spectral domain approaches for longer time horizons, while the second application provides evidence for the advantages of using the spectral domain approaches for shorter time horizons. The first application is to systematic risk, or the CAPM beta. In a study similar to Gençay et al. (2005), we show that the relationship between systematic risk and the returns of assets becomes stronger at greater horizons. This strong relationship however, appears at greater horizons than initially shown by Gençay et al. (2005). For the second application, we create minimum variance portfolios by using the Fourier-based covariance, the wavelet-based covariance of Whitcher et al. (2000), the long-run wavelet-based covariance of Conlon et al. (2018), the DECO algorithm of Engle and Kelly (2012), as well as the sample covariance matrix and a naive equal weight strategy. We formally test the difference in variance of the portfolios by making use of the robust variance test of Ledoit and Wolf (2011). This test uses bootstrapped confidence intervals to test the null hypothesis of equal variance between portfolios.

For the first application to systematic risk, we consider all stocks in the S&P500 at the time of writing from 3 January 2005 until 28 May 2021. We retrieve all the price paths of the individual stocks from Google Finance, the time series are sampled with a daily frequency. We

rule out the stocks that had their initial public offering after the start of our sample period, which leaves us with 411 stocks. We linearly interpolate any missing values. We use the S&P500 index over the same time period for the market returns. To construct the minimum variance portfolios, we consider daily return time series from the main indices of 15 countries (Australia, Canada, China, France, Germany, India, Italy, Japan, Netherlands, Russia, Spain, Sweden, Switzerland, United Kingdom, United States). The corresponding indices for each country are given in Table 6 in Appendix A. We retrieve all the time series of the returns of the indices from Bloomberg. The time series run from 10 April 2007 to 7 April 2021. We convert each index from its original currency to EUR. We linearly interpolate missing values, caused by e.g. a mismatch in trading days and holidays.

In the simulation study, we find that within the Daubechies family of wavelets, the wavelet with six vanishing moments tends to perform the best when using the wavelet-based covariance of Whitcher et al. (2000). For the long-run wavelet-based covariance of Conlon et al. (2018), the wavelet with eight vanishing moments tends to perform the best overall. When comparing the performances of the Fourier-based covariance, the wavelet-based covariance of Whitcher et al. (2000), the long-run wavelet-based covariance of Conlon et al. (2018), the DECO algorithm of Engle and Kelly (2012) and the sample correlation, we find that the DECO algorithm and the sample correlation perform quite poorly when the correlation in time series is induced by cyclical behaviour of the time series. These methods however tend to outperform the spectral domain methods when the DGP is a multivariate normal distribution with constant correlation. We see this result as a bias-variance trade-off. This bias-variance tradoff is also present between spectral methods, the less specific a method is, the better its overall performance tends to be, while the more specific methods tend to outperform the more general methods in specific circumstances. We see that the Fourier-based method performs the best when cyclical behaviour is present, while the wavelet-based methods tend to outperform the Fourier-based method in situations where no cyclical behaviour is present. In turn, the wavelet-based method of Whitcher et al. (2000) tends to outperform the more general long-run wavelet-based method of Conlon et al. (2018) when cyclical behaviour is present, while being outperformed by the long-run wavelet-based method of Conlon et al. (2018) when there is no cyclical behaviour.

In the application to systematic risk, we see strong evidence that the relationship between the CAPM beta and the returns of an asset are strongly related in the long-run, when using an appropriate method to estimate the long term beta. We argue that this finding indicates the existence of a long-run risk premium in the stock market. We also find evidence for a short term volatility feedback effect or leverage effect. In the application to minimum variance portfolios, we see that there is some statistically significant evidence that spectral methods can improve minimum variance allocation, although the observed improvement is relatively small. In sum, we conclude that the spectral domain methods for covariance estimation can provide a benefit to applications with short horizons as well as longer horizons. However, the advantage seems to be greater when working with longer time horizons.

2 Spectral methods

2.1 Fourier transform

To compute spectral domain covariance measures, we will analyse time series in the frequency domain. To analyse time series in the frequency domain, we can make use of the Fourier transform. The Fourier transform is essentially an integral transform. Integral transforms map functions from their original function space to a new function space. An integral transform is given as

$$G(s) = \int_a^b g(t)K(s,t)\mathrm{d}t, \tag{1}$$

where $g(t)$ is the original function, $G(s)$ is the transform of $g(t)$, $K(s,t)$ is the kernel of the transform and a and b are the limits of integration. It is possible that $a = -\infty$ and/or $b = \infty$. The Laplace transform of a function $g(t) : \mathbb{R} \to \mathbb{R}$, denoted by $\mathcal{L}\{g(t)\}$ or $G(s) : \mathbb{C} \to \mathbb{C}$, is an integral transform with the kernel $K(s,t) = \mathrm{e}^{-st}$, where $s \in \mathbb{C}$. The Laplace transform has as limits of integration $a = -\infty$ and $b = \infty$ and is thus given as

$$\mathcal{L}\{f(t)\} = G(s) = \int_{-\infty}^{\infty} g(t)\mathrm{e}^{-st}\mathrm{d}t. \tag{2}$$

The Laplace transform is a more general version of the Fourier transform, where the s in the kernel of the Laplace transform is any number in the complex domain, it is fixed at the imaginary

axis for the kernel of the Fourier transform, i.e. the real part is zero. The kernel of the Fourier transform is given by $K(f,t) = \mathrm{e}^{-2\pi ift}$, where $f \in \mathbb{R}$. The Fourier transform of a function $g(t) : \mathbb{R} \to \mathbb{R}$, denoted by $\mathcal{F}\{g(t)\}$ or $G(f) : \mathbb{R} \to \mathbb{C}$, is given as

$$\mathcal{F}\{g(t)\} = G(f) = \int_{-\infty}^{\infty} g(t)\mathrm{e}^{-2\pi ift}\mathrm{d}t.^{1} \tag{3}$$

When $g(t)$ is a real-valued function of time, then $\mathcal{F}\{g(t)\}$ will be a complex-valued function of frequency. The magnitude of the function indicates how strong a sine function is present. The argument of the function indicates the phase offset of the basic sine function. We can explain this property by using Euler's formula, which states that $\mathrm{e}^{ix} = \cos(x) + i\sin(x) \quad \forall x \in \mathbb{R}$. By applying the inverse Fourier transform to the Fourier transform of a function, we get the original function back. The inverse Fourier transform of $G(f)$, denoted by $\mathcal{F}^{-1}\{G(f)\}$, is an integral transform with kernel $K(t,f) = \mathrm{e}^{2\pi itf}$ and is thus given by

$$\mathcal{F}^{-1}\{G(f)\} = g(t) = \int_{-\infty}^{\infty} G(f)\mathrm{e}^{2\pi itf}\mathrm{d}f. \tag{4}$$

The theorem that we can reconstruct the original function from its Fourier transform is also known as the Fourier inversion theorem and a proof can be found in, e.g., Rahman (2011). The intuition behind the theorem is that, by knowing all frequency and phase information of a function, we can reconstruct the function completely.

Since we are not working with a continuous function, the need for the discrete time Fourier transform arises. The discrete time Fourier transform is applicable to sequences instead of continuous functions. The discrete time Fourier transform of a sequence $x[n] : \mathbb{Z} \to \mathbb{R}$ is denoted by $\mathcal{F}\{x[n]\}$, or $X(\omega) : \mathbb{R} \to \mathbb{C}$ and is given as

$$\mathcal{F}\{x[n]\} = X(\omega) = \sum_{n=-\infty}^{\infty} x[n]\mathrm{e}^{-\omega in}, \tag{5}$$

where ω is given in units of radians/sample. Note that $X(\omega)$ is a continuous function, while $x[n]$ is a discrete sequence. $X(\omega)$ is periodic with respect to ω with a period of 2π i.e. $X(\omega + 2\pi a) = X(\omega) \quad \forall a \in \mathbb{Z}$. Due to this periodicity, it suffices to only consider the interval $\omega \in [-\pi, \pi]$.

[1] It can be noted that the Fourier transform is closely related to the characteristic function of a probability density function.

Now let $x[n]$ be a sequence sampled from a continuous function $g(t)$ with a sampling interval of T_s seconds, such that we have $T_s \cdot g(xT_s) = x[n]$. We can then relate the discrete time Fourier transform to frequency by writing

$$X(2\pi f T_s) = \sum_{n=-\infty}^{\infty} x[n] \mathrm{e}^{-2\pi i f T_s n} = \sum_{n=-\infty}^{\infty} T_s \cdot g(nT_s) \mathrm{e}^{-2\pi i f T_s n}, \tag{6}$$

where f is then given in cycles per second, $f \in [-1/(2T_s), 1/(2T_s)]$.

Using the Poisson summation formula, we can rewrite the above expression for $X(2\pi f T_s)$ to

$$X(2\pi f T_s) = \sum_{n=-\infty}^{\infty} T_s \cdot g(nT_s) \mathrm{e}^{-2\pi i f T_s n} = \sum_{n=-\infty}^{\infty} G\left(f - \frac{k}{T_s}\right), \tag{7}$$

where $G(f)$ is the continuous Fourier transform of $g(t)$ as given in Equation 3. From this expression, we can see that the discrete time Fourier transform in this form is essentially a repetition of the regular Fourier transform, shifted with the sampling frequency $f_s = 1/T_s$. Therefore, we require that

$$f_s \geq 2B, \tag{8}$$

where B is the so called bandwidth of the function $g(t)$. B is defined as the difference between the minimum and maximum non-zero frequency domain component of a function, i.e. $B = f_{\max} - f_{\min}$, where $f_{\min} = \min(f)$ such that $G(f) \neq 0$ and $f_{\max} = \max(f)$ such that $G(f) \neq 0$. When we satisfy this condition, we assure that there is no overlap in the repetitions of the Fourier transform. If we do not satisfy this condition, the frequencies that exceed the threshold are 'mapped around' to the lower frequencies. This phenomenon is also know as the Nyquist-Shannon sampling theorem.

The discrete time Fourier transform has its own inverse and is given as

$$\mathcal{F}^{-1}\{X(\omega)\} = x[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(\omega) \mathrm{e}^{\omega i n} \mathrm{d}\omega. \tag{9}$$

When using frequency f instead of radians/sample ω, the inverse discrete time Fourier transform becomes

$$\mathcal{F}^{-1}\{X(f)\} = x[n] = T_s \int_{-\frac{1}{2T_s}}^{\frac{1}{2T_s}} X(f) \mathrm{e}^{2\pi i f T_s n} \mathrm{d}f. \tag{10}$$

We do however not work with infinite sequences, but with a finite amount of data. The necessary finite sample approximation of the discrete time Fourier transform is called the discrete Fourier transform. The discrete Fourier transform of a time series $x[n] : \mathbb{N} \to \mathbb{R}$ is denoted by $X(k) : \mathbb{N} \to \mathbb{C}$ and is given as

$$\mathcal{F}\{x[n]\} = X[k] = \sum_{n=0}^{N-1} x[n] \mathrm{e}^{-2\pi i k \frac{n}{N}}, \tag{11}$$

where $k \in \mathbb{N}$ is the discrete frequency bin. We usually let k run from 0 to $N-1$, implying that we have N discrete frequency bins. When using the discrete Fourier transform, we essentially split up the interval of $\omega \in [-\pi, \pi]$, or $f \in [-1/(2T_s), 1/(2T_s)]$ of the discrete time Fourier transform in the chosen number of frequency bins. The inverse discrete Fourier transform is given by

$$\mathcal{F}^{-1}\{X[k]\} = x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] \cdot \mathrm{e}^{2\pi i k \frac{n}{N}}. \tag{12}$$

2.2 Short-time Fourier transform

Eventhough Fourier transforms are a useful tool for spectral analysis, we would like to capture the time behaviour of the spectral components in order to compute a covariance measure. Fourier transforms are only localised in time. Therefore, it is not possible to determine how certain frequency components evolve over time by using the basic Fourier transform. We can however add a time dimension by making use of the windowed Fourier transform, also known as the short-time Fourier transform (STFT). When computing the STFT of a function, we essentially slide a fixed-sized window function $w(\tau)$ over the time axis of the original function. We then compute the Fourier transform of the multiplication of this window function and the original function, resulting in a time and a frequency dimension. The STFT of a function $g(t) : \mathbb{R} \to \mathbb{R}$, denoted by $\mathrm{STFT}\{g(t)\}$ or $X(\tau, \omega) : \mathbb{R}^2 \to \mathbb{C}$, is given as

$$\mathrm{STFT}\{g(t)\} = X(\tau, f) = \int_{-\infty}^{\infty} g(t) w(t-\tau) \mathrm{e}^{-2\pi i f t} \mathrm{d}t, \tag{13}$$

where $w(\tau)$ is a window function. Using a rectangular window would be the same as constantly changing the limits of integration.

The discrete, finite sample STFT of a function $x[n] : \mathbb{N} \to \mathbb{R}$, denoted by $\text{STFT}\{x[n]\}$ or $X[m,k] : \mathbb{N}^2 \to \mathbb{C}$, is given as

$$\text{STFT}\{x[n]\} = X[m,k] = \sum_{n=0}^{L-1} x[n]w[n-m]\mathrm{e}^{-2\pi i k \frac{n}{L}}, \tag{14}$$

where, similar to the continuous case, $w[n]$ is the window function with fixed length $L \in \mathbb{N}$.

In practise a rectangular window creates a lot of spectral leakage (Harris, 1978). This spectral leakage can essentially be seen as added noise onto our frequency spectrum and is caused by the frequency domain properties of the used window. One of the best performing window functions is the Blackman-Nutall window introduced by Nuttall (1981).

When using the discrete STFT, we have the major drawback that we have to make a trade-off between time and frequency resolution. In the discrete case, we have a time series sampled with sampling frequency f_s and a window of length L. Following the Nyquist-Shannon sampling theorem, we can separate the frequencies from 0 to $f_s/2$ in $L/2$ discrete frequency bins. Implying that all frequency bins are spaced apart by f_s/L Hertz. Increasing the window size will result in frequency bins that are spaced closer together, thus results in a higher frequency resolution. Increasing the window size however implies that we lower the time resolution, due to the fact that we can fit less windows within the original function. This resolution trade-off renders the STFT almost useless in a financial setting.

2.3 Wavelet transform

To overcome the drawback of the STFT, we turn to the wavelet transform. The wavelet transform allows us to capture the spectral behaviour of a time series over time without the resolution trade-off of the STFT. Especially in a financial setting, the use of wavelet transforms instead of Fourier transforms of the STFT makes a lot more sense. Due to localisation in time of the wavelet transform, we can capture time variations and shocks in the time series much better. When applying a wavelet transform, we perform a multiresolution decomposition. Meaning that we decompose the original signal into its constituent multiresolution components. We divide the original function into so called scales. When discussing wavelets, we follow the notation of Crowley (2007).

Wavelets come in pairs, the wavelet function itself, denoted by $\psi(t)$, and a scaling function, denoted by $\phi(t)$. The wavelet function and the scaling function integrate to 0 and 1, respectively, i.e.

$$\int \psi(t)\mathrm{d}t = 0, \tag{15}$$

$$\int \phi(t)\mathrm{d}t = 1. \tag{16}$$

Complex valued wavelet functions do exist. However, for this thesis we will restrict ourselves to the use of real valued functions. As such, we have that $\psi(t) : \mathbb{R} \to \mathbb{R}$ and $\phi(t) : \mathbb{R} \to \mathbb{R}$.

Important properties of the wavelet and scaling functions are those of scaling and shifting. We use these scaled and shifted functions to decompose the series into a sequence of projections. The scaled and shifted wavelet functions and scale functions are given as

$$\psi_{j,k}(t) = \frac{1}{\sqrt{j}}\psi\left(\frac{t-kj}{j}\right) \quad \text{and} \quad \phi_{j,k}(t) = \frac{1}{\sqrt{j}}\phi\left(\frac{t-kj}{j}\right), \tag{17}$$

respectively. In these expressions, j is used as the index of the scale and k is used as the index for the shift. We let the scale j run from 1 to J.

In practise, we usually scale with powers of two. The reason for scaling with the power two is to prevent redundancy as well as that doing so is easily implemented in computationally efficient algorithms, such as done in Mallat's algorithm (Mallat, 1989). The scaled and shifted functions are then given by

$$\psi_{j,k}(t) = \frac{1}{\sqrt{2^j}}\psi\left(\frac{t-k2^j}{2^j}\right) \quad \text{and} \quad \phi_{j,k}(t) = \frac{1}{\sqrt{2^j}}\phi\left(\frac{t-k2^j}{2^j}\right). \tag{18}$$

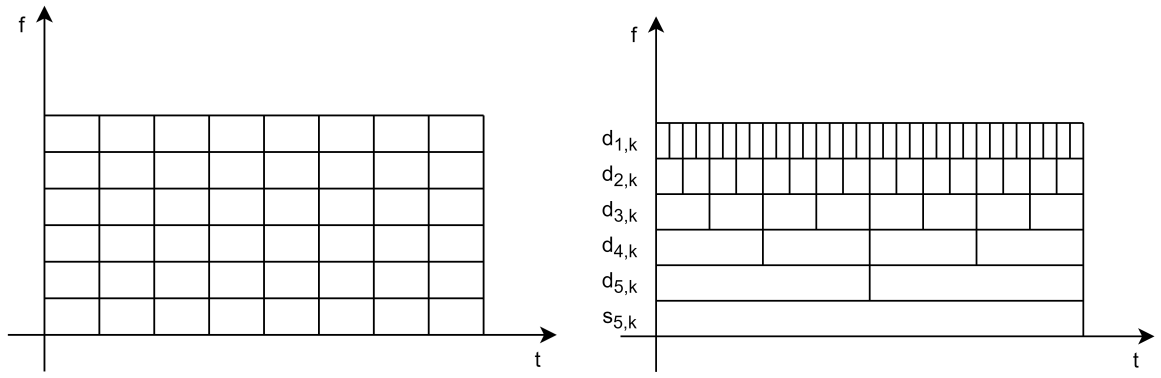
Using these scaled and shifted wavelet and scale functions, we can calculate the wavelet coefficients of a function $g(t) : \mathbb{R} \to \mathbb{R}$ by the integrals

$$d_{j,k} = \int g(t)\psi_{j,k}(t) \quad \text{and} \quad s_{J,k} = \int g(t)\phi_{J,k}(t). \tag{19}$$

Note that these coefficients are obtained by integral transforms, where the used kernel is the wavelet or scaling function.

Figure 1 compares the resolution obtained from such a wavelet transform with the resolution of the STFT. Considering the wavelet transform resolution in Figure 1b, we can see that we have

a high time resolution for the finer scales, while having a lower time resolution for the coarser scales, where such a high resolution is not required. This way we circumvent the resolution trade-off that we have when using the STFT.



(a) (b)

Figure 1: (a) Schematic representation of the time-frequency resolution obtained by applying a STFT, and (b) schematic representation of the resolution obtained by using a wavelet transform by scaling the wavelet with powers of two.

Due to the fact that we scale the wavelet and scaling functions with a power of two for each step, the frequencies contained in scale j are 2^j to 2^{j+1}. Table 7 in Appendix A gives an overview of the frequencies and time horizons corresponding to each scale for different sample frequencies.

Having obtained all wavelet coefficients of all the scales, we can reconstruct the original function using the wavelet coefficients. The original function $g(t)$ is given as

$$\begin{aligned} g(t) = &\sum_k d_{1,k}\psi_{1,k}(t) + \sum_k d_{2,k}\psi_{2,k}(t) + \cdots \\ &+ \sum_k d_{J,k}\psi_{J,k}(t) + \sum_k s_{J,k}\phi_{J,k}(t). \end{aligned} \tag{20}$$

Having obtained this multiresolution decomposition, we accumulate the coefficients in the scale vectors

$$[\tilde{\boldsymbol{d}}_1, \tilde{\boldsymbol{d}}_2, ...\tilde{\boldsymbol{d}}_{J-1}, \tilde{\boldsymbol{d}}_J, \tilde{\boldsymbol{s}}_J], \tag{21}$$

where $\tilde{\boldsymbol{d}}_1$ is the first term of Equation 20, i.e. $\tilde{\boldsymbol{d}}_1 = \sum_k d_{1,k}\psi_{1,k}(t)$. $\tilde{\boldsymbol{d}}_2$ is the second term until

$\tilde{\boldsymbol{s}}_J$, which is the last term.

Since we are decomposing the original time series in these basis functions with finite energy, we do not have to assume that each of the projections exists everywhere such as in the decomposition into sines and cosines as is the case with the Fourier transform. This decomposition into finite energy basis functions allows us to capture typical financial time-series characteristics such as trends, volatility clustering and abrupt shocks much better (Gençay et al., 2002).

The coarsest scales can be accumulated together, i.e.,

$$\tilde{s}_{j-1,k} = \tilde{s}_{J,k} + \tilde{d}_{J,k} + \tilde{d}_{J-1,k} + \cdots + \tilde{d}_{j,k}. \tag{22}$$

This implies that the coarsest (highest numbered and thus lowest frequency) scale contains everything that is not explained with the finer frequency scales, so this coarsest scale can be seen as the trend component.

The discrete wavelet transform is implemented by the aforementioned Mallat's algorithm. In this algorithm, we convolve the original time series $x[n] : \mathbb{N} \to \mathbb{R}$ with two filters, one being the wavelet function filter, the other being the scale function filter. We then downsample the result with a factor two and keep the scale coefficients obtained by the wavelet filter. The scale coefficients obtained from the scale filter are then filtered again. We repeat this procedure until we have obtained all scales coefficients $[d_{1,k}, ..., d_{J,k}]$ and the residual scale coefficients $s_{J,k}$. This procedure results in a pyramid style algorithm. A schematic representation of this algorithm is depicted in Figure 2. In this figure, the transfer function of the series of filter coefficients of the wavelet function is denoted by $F_\psi[i]$ and is a m-length filter, i.e. $1 \leq i \leq m$. The transfer function of the series of filter coefficients of the scale function filter is denoted by $F_\phi[i]$ and is also a m-length filter. $F_\psi[i]$ and $F_\phi[i]$ are given by the Fourier transforms of the series of filter coefficients for the wavelet function and scale function, respectively. These series of filter coefficients are obtained by solving

$$\phi[x] = \sqrt{2}\sum_i f_\phi[i]\phi[2x - i] \qquad \text{and} \qquad \psi[x] = \sqrt{2}\sum_i f_\psi[i]\phi[2x - i] \tag{23}$$

for $f_\psi[i]$ and $f_\phi[i]$, which are the series of filter coefficients for the wavelet function and scale function, respectively. This will result in

$$f_\psi[i] = (-1)^{i+1}\overline{f_\phi[m - i]}, \tag{24}$$

implying that $f_\psi[i]$ is the so called quadrature mirror filter of $f_\phi[i]$.

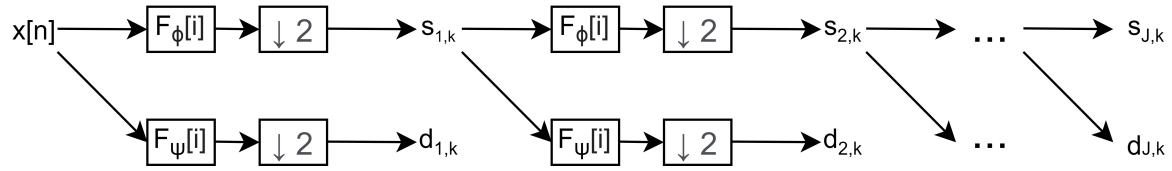


Figure 2: Schematic representation of Mallat's algorithm for the discrete wavelet transform of a given input series $x[n]$.

One drawback of the discrete wavelet transform is that, due to the downsampling of a factor two for each step, we require the the length of the original time series two be equal to a power of two. The maximum overlap discrete wavelet transform (MODWT) as described by Percival and Mofjeld (1997) offers a solution to this problem. Instead of downsampling after each filtering operation, we scale the transfer function of the series of filter coefficients of the next filter operation with a factor of two. A schematic representation for the MODWT algorithm is depicted in Figure 3.

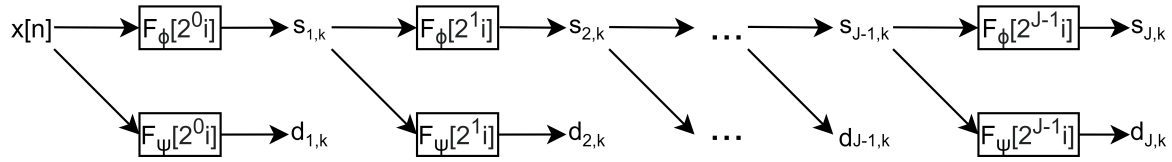


Figure 3: Schematic representation of the MODWT algorithm for a given input series $x[n]$.

The MODWT is highly redundant. We do not gain any more information compared to the regular discrete wavelet transform, while having a lot more coefficients. However, we gain two important features. The first being the fact that we do not require the length of the time series to be equal to a power of two. The second property is the preservation of energy. This preservation of energy is, as with the Fourier transform, the reason we can draw conclusions on the distribution of variance across the different scales. Due to these features, we will use the MODWT as the wavelet transform within this thesis.

When using the wavelet transform, we can choose from a wide variety of wavelet families, as

well as different wavelets within each family. As shown by Gençay et al. (2002), the Daubechies family of wavelets generally has the best performance when the MODWT is performed for financial data. For this reason, we will use the Daubechies family of wavelets when performing the MODWT. The Daubechies family of wavelets was introduced as a family of compactly supported orthonormal wavelets by Daubechies (1992). In her book, Daubechies lists all necessary and sufficient conditions for her family of wavelets and then gives the numerical solutions for the filter coefficients of the scaling function with 2 to 10 vanishing moments. The number of vanishing moments and the domain are the only conditions that are varied across the different wavelets within the family. A wavelet function with m_v vanishing moments satisfies

$$\int_{-\infty}^{\infty} t^k \psi(t) \mathrm{d}t = 0 \quad \text{for } 0 \leqslant k < m_v, \tag{25}$$

where $k \in \mathbb{N} \cup 0$. The domain of the wavelet and scaling functions, also called the support of the wavelet, is given as $[0, 2m_v - 1]$

The filter coefficients for the scaling filters with 2 to 10 vanishing moments can be found in Table 8 in Appendix A. Since the wavelet and scaling filters are orthonormal, the filter coefficients of the wavelet filters are given by the quadrature mirrors of the scaling filters. The filter coefficients of the wavelet filters are given in Table 9 in Appendix A.

Using filter coefficients, we can construct the wavelet and scaling functions themselves by using the cascade algorithm of Daubechies (1992). Given the series of scaling filter coefficients $f_\phi[i]$ we iterate over

$$\phi^{(k+1)}(t) = \sqrt{2} f_\phi[n] * \phi^{(k)}_{extended}[n], \tag{26}$$

where k is the current iteration, $*$ denotes the convolution operator and $\phi^{(k)}_{extended}$ denotes the series of $\phi^{(k)}$ but extended by adding a zero value in between each of the individual values in the series, i.e.

$$\begin{aligned} \phi^{(j)}_{extended}[2j] &= \phi^{(k)}[j], \\ \phi^{(j)}_{extended}[2j-1] &= 0, \end{aligned} \tag{27}$$

where $j \in \mathbb{N}$.

The results of the cascade algorithm are depicted in Figure 9 and Figure 8 in Appendix B for the wavelet functions and scaling functions, respectively.

3 Covariance measures

3.1 Fourier transform based covariance measure

The first proposed covariance measure is based on the simple Fourier transform. The Fourier transform is an energy preserving transform, implying that we can draw conclusions on the variance of the original function after the transform. A proof is provided in Appendix D. We create an expression for the crosscovariance for specific frequency components as a horizon specific covariance measure. We will first show that the cross covariance of two functions $g(t)$ and $h(t)$, denoted by $\gamma_{gh}(\tau)$ $\tau \in \mathbb{R}$ can be written as $\mathcal{F}^{-1}\{\mathcal{F}\{h(t)\}\overline{\mathcal{F}\{g(t)\}}\}$. This property will prove to be useful when evaluating (cross)spectral variance.

Let $\gamma_{gh}(\tau)$ the cross covariance function of two processes $g(t)$ and $h(t)$, where $\tau \in \mathbb{R}$ represents the lag of the cross covariance function. The cross covariance function is then given by

$$\gamma_{gh}(\tau) = \int_{-\infty}^{\infty} \overline{g(t)} h(t+\tau) \mathrm{d}t. \tag{28}$$

We can write the functions $g(t)$ and $h(t)$ as the inverse Fourier transform of its Fourier transforms $G(f)$ and $H(f)$, respectively, resulting in

$$g(t) = \int_{-\infty}^{\infty} G(f) \mathrm{e}^{2\pi i f t} \mathrm{d}f \quad \text{and} \quad h(t) = \int_{-\infty}^{\infty} H(f) \mathrm{e}^{2\pi i f t} \mathrm{d}f. \tag{29}$$

The time shifted by τ function $h(t+\tau)$ is then given by

$$h(t+\tau) = \int_{-\infty}^{\infty} H(f) \mathrm{e}^{2\pi i f (t+\tau)} \mathrm{d}f = \int_{-\infty}^{\infty} H(f) \mathrm{e}^{2\pi i f t} \mathrm{e}^{2\pi i f \tau} \mathrm{d}f. \tag{30}$$

The complex conjugate of $g(t)$, denoted by $\overline{g(t)}$, can also be written as the inverse Fourier transform of its Fourier transform and can thus be written as

$$\overline{g(t)} = \int_{-\infty}^{\infty} \overline{G(f')} \mathrm{e}^{-2\pi i f' t} \mathrm{d}f'. \tag{31}$$

Plugging the expressions of the time shifted function $h(t+\tau)$ and the complex conjugate function $\overline{g(t)}$ into the expression for the cross covariance gives us

$$\gamma_{gh}(\tau) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} H(f) \mathrm{e}^{2\pi i f t} \mathrm{e}^{2\pi i f \tau} \mathrm{d}f \int_{-\infty}^{\infty} \overline{G(f')} \mathrm{e}^{-2\pi i f' t} \mathrm{d}f' \right] \mathrm{d}t. \tag{32}$$

Rewriting this expression yields

$$\gamma_{gh}(\tau) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H(f)\overline{G(f')}\mathrm{e}^{2\pi i t(f-f')}\mathrm{e}^{2\pi i f\tau}\mathrm{d}t\mathrm{d}f\mathrm{d}f'. \tag{33}$$

Which we then rewrite as

$$\begin{aligned}\gamma_{gh}(\tau) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H(f)\overline{G(f')}\int_{-\infty}^{\infty}\mathrm{e}^{2\pi i t(f-f')}\mathrm{d}t\mathrm{e}^{2\pi i f\tau}\mathrm{d}f\mathrm{d}f' \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H(f)\overline{G(f')}\mathrm{e}^{2\pi i f\tau}\int_{-\infty}^{\infty}\mathrm{e}^{-2\pi i t(f'-f)}\mathrm{d}t\mathrm{d}f\mathrm{d}f'.\end{aligned} \tag{34}$$

The inner integral over t is the Fourier transform of a constant and a time shift. Resulting in a time shifted Dirac delta function, denoted by $\delta(f'-f)$. The Fourier transform of the Dirac delta function and some of its properties are given in Appendix C. Filling in the Dirac delta function results in

$$\gamma_{gh}(\tau) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H(f)\overline{G(f')}\mathrm{e}^{2\pi i f\tau}\delta(f'-f)\mathrm{d}f\mathrm{d}f'. \tag{35}$$

We then swap the order of the two integrals to get

$$\gamma_{gh}(\tau) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H(f)\overline{G(f')}\mathrm{e}^{2\pi i f\tau}\delta(f'-f)\mathrm{d}f'\mathrm{d}f. \tag{36}$$

Then by the properties of the Dirac delta function we rewrite the expression of the cross covariance to

$$\gamma_{gh}(\tau) = \int_{-\infty}^{\infty} H(f)\overline{G(f)}\mathrm{e}^{2\pi i f\tau}\mathrm{d}f, \tag{37}$$

Which is the inverse Fourier transform of the product of $H(f)\overline{G(f)}$. Therefore, we get our result that

$$\gamma_{gh}(\tau) = \mathcal{F}^{-1}\{H(f)\overline{G(f)}\} = \mathcal{F}^{-1}\{\mathcal{F}\{h(t)\}\overline{\mathcal{F}\{g(t)\}}\}. \tag{38}$$

This result can be generalised to autocovariance as well, i.e. $\gamma_{gg}(\tau) = \mathcal{F}^{-1}\{\mathcal{F}\{g(t)\}\overline{\mathcal{F}\{g(t)\}}\}$.

We consider the result $\gamma_{gh}(\tau) = \mathcal{F}^{-1}\{\mathcal{F}\{h(t)\}\overline{\mathcal{F}\{g(t)\}}\}$. By filling in the discrete Fourier transforms we get

$$\gamma_{gh}[\tau] = \frac{1}{N}\sum_{k=0}^{N-1} H[k]\overline{G[k]}\mathrm{e}^{2\pi i k\frac{\tau}{N}}, \tag{39}$$

where $G[k] : \mathbb{N} \to \mathbb{C}$ and $H[k] : \mathbb{N} \to \mathbb{C}$ are the discrete Fourier transforms of $g[n] : \mathbb{N} \to \mathbb{R}$ and $h[n] : \mathbb{N} \to \mathbb{R}$, respectively. Since we are only interested in the covariance at lag $\tau = 0$, we get

$$\text{Cov}(g[n], h[n]) = \gamma_{gh}[0] = \frac{1}{N} \sum_{k=0}^{N-1} H[k]\overline{G[k]}\text{e}^0 = \frac{1}{N} \sum_{k=0}^{N-1} H[k]\overline{G[k]}. \tag{40}$$

With this expression, we have obtained a clear decomposition of the covariances across all frequencies. We can now define a subset of interest, $\mathcal{I} \subseteq \{0, ..., N-1\}$, which contains the frequency bins we are interested in. When summing over these specific frequency components, we get a frequency specific covariance measure defined as

$$\text{Cov}_{\mathcal{I}}(g[n], h[n]) = \frac{1}{N} \sum_{k \in \mathcal{I}} H[k]\overline{G[k]}. \tag{41}$$

Computing the covariance in this manner is essentially the same as computing the sample covariance of the, by the subset $\mathcal{I}$ restricted, reconstructed functions. Where we would use the inverse discrete time Fourier transform to reconstruct the functions.

3.2 Wavelet-based covariance measure

The wavelet cross-covariance estimator for scale j of time series $x[n]$ and $y[n]$ for lag τ proposed by Whitcher et al. (2000) is given as

$$\gamma_{j,XY}(\tau) = \frac{1}{N_j} \sum_{l=L_j-1}^{N-\tau-1} d_{j,l}^{(X)} d_{j,l+\tau}^{(Y)}, \tag{42}$$

where $d_{j,l}^{(X)}$ denotes the element l of the scale vector $\boldsymbol{d}_j$ of the MODWT of $x[n]$. L_j denotes the length of the series of the MODWT filter coefficient used for scale j of the MODWT. L_j is given as

$$L_j = 2^{j-1}(L-1) + 1, \tag{43}$$

where L is the length of the series of the MODWT filter coefficients of the first scale and is thus equal to L_1. N is the length of the original time series $x[n]$ and N_j is the number of times we can fit the MODWT filter of scale j within the original time series of length N. N_j is given as

$$N_j = N - L_j + 1. \tag{44}$$

Since we are only interested in the covariance at a certain scale at lag $\tau = 0$, our expression for the MODWT covariance between $x[n]$ and $y[n]$ at scale j becomes

$$\text{Cov}_j(x[n], y[n]) = \frac{1}{N_j} \sum_{l=L_j-1}^{N-1} d_{j,l}^{(X)} d_{j,l}^{(Y)}. \tag{45}$$

3.3 Long-run wavelet-based correlation

We now follow the derivation of the Long-run wavelet-based correlation measure of Conlon et al. (2018). the long-run correlation can be computed from the long-run covariance, which can be seen as the regular sample covariance coefficient with a series of correction terms for the serial and cross-serial correlation.

The vector of wavelet scale coefficients of scale j for a T length vector of returns $\boldsymbol{r}_m$, denoted by $\boldsymbol{s}_{m,j}$, is given by the convolution of the vector containing the level j scaling filter coefficients $\boldsymbol{g}_j$ and $\boldsymbol{r}_m$, i.e.,

$$\boldsymbol{s}_{m,j} = \boldsymbol{g}_j * \boldsymbol{r}_m. \tag{46}$$

Since $\boldsymbol{g}_j$ is the level j scaling filter, the filter coefficients of $\boldsymbol{g}_j$ are given by the convolution of scaling filter 1 until scaling filter j, as represented in Figure 3. They then write the the wavelet scaling covariance between the time series of returns $\boldsymbol{r}_m$ and $\boldsymbol{r}_n$ as

$$\begin{aligned} \text{Cov}(\boldsymbol{s}_{m,j}, \boldsymbol{s}_{n,j}) &= \text{Cov}(\boldsymbol{g}_j * \boldsymbol{r}_m, \boldsymbol{g}_j * \boldsymbol{r}_n) = \boldsymbol{g}_j * \boldsymbol{g}_j * \text{Cov}(\boldsymbol{r}_m, \boldsymbol{r}_n) \\ &= \sum_{k=0}^{L_j-1} \sum_{l=0}^{L_j-1} g_{j,k} g_{j,l} \, \text{Cov}\left(\boldsymbol{r}_{m,t-k \bmod T}, \boldsymbol{r}_{n,t-l \bmod T}\right). \end{aligned} \tag{47}$$

Which they split up for $k = l$ and $k \neq l$. For $k = l$ they write

$$g_{j,k} g_{j,l} \, \text{Cov}\left(\boldsymbol{r}_{m,t-k \bmod T}, \boldsymbol{r}_{n,t-l \bmod T}\right) = g_{j,l}^2 \, \text{Cov}\left(\boldsymbol{r}_m, \boldsymbol{r}_n\right) = \left(g_{1,l}^2\right)^j \, \text{Cov}\left(\boldsymbol{r}_m, \boldsymbol{r}_n\right). \tag{48}$$

For $k \neq l$ they use the relation $\text{Cov}(\boldsymbol{r}_m, \boldsymbol{r}_n) = \rho(\boldsymbol{r}_m, \boldsymbol{r}_n)\sigma(\boldsymbol{r}_m)\sigma(\boldsymbol{r}_n)$ to write

$$\begin{aligned} g_{j,k} g_{j,l} \, \text{Cov}\left(\boldsymbol{r}_{m,t-k \bmod T}, \boldsymbol{r}_{n,t-l \bmod T}\right) &= g_{j,k} g_{j,l} \rho^{k-l}\left(\boldsymbol{r}_m, \boldsymbol{r}_n\right) \sigma\left(\boldsymbol{r}_{m,t-k \bmod T}\right) \sigma\left(\boldsymbol{r}_{n,t-l \bmod T}\right) \\ &= g_{j,k} g_{j,l} \rho^{k-l}\left(\boldsymbol{r}_m, \boldsymbol{r}_n\right) \sigma\left(\boldsymbol{r}_m\right) \sigma\left(\boldsymbol{r}_n\right), \end{aligned} \tag{49}$$

where $\rho^{k-l}(\boldsymbol{r}_m, \boldsymbol{r}_n)$ denotes the cross correlation between $\boldsymbol{r}_m$ and $\boldsymbol{r}_n$ for lag $k-l$. They then combine the expressions and use the MODWT filter property $\sum_{l=0}^{L_1-1} g_{1,l}^2 = 1/2$, where L_1 is the length of $\boldsymbol{g}_1$, to get

$$\begin{aligned}\operatorname{Cov}(\boldsymbol{s}_{m,j}, \boldsymbol{s}_{n,j}) = &\left(\frac{1}{2}\right)^j \operatorname{Cov}(\boldsymbol{r}_m, \boldsymbol{r}_n) \\ &+ \sum_{k=0}^{L_j-2} \sum_{l=k+1}^{L_j-1} g_{j,k} g_{j,l} \left[\rho^{k-l}(\boldsymbol{r}_m, \boldsymbol{r}_n) + \rho^{-(k-l)}(\boldsymbol{r}_m, \boldsymbol{r}_n)\right] \sigma(\boldsymbol{r}_m)\, \sigma(\boldsymbol{r}_n)\end{aligned} \tag{50}$$

Similarly, they express the long-run variance as

$$\sigma^2(\boldsymbol{s}_{m,j}) = \left(\frac{1}{2}\right)^j \sigma^2(\boldsymbol{r}_m) + 2 \sum_{k=0}^{L_j-2} \sum_{l=k+1}^{L_j-1} g_{j,k} g_{j,l} \rho^{(k-l)}(\boldsymbol{r}_m)\, \sigma^2(\boldsymbol{r}_m). \tag{51}$$

They then express the long-run correlation as

$$\rho(\boldsymbol{s}_{m,j}, \boldsymbol{s}_{n,j}) = \frac{\operatorname{cov}(\boldsymbol{s}_{m,j}, \boldsymbol{s}_{n,j})}{\sigma(\boldsymbol{s}_{m,j})\, \sigma(\boldsymbol{s}_{n,j})} = \frac{\left(\frac{1}{2}\right)^j \operatorname{Cov}(\boldsymbol{r}_m, \boldsymbol{r}_n) + \sum_{k=0}^{L_j-2L_j-1} \sum_{l=k+1}^{L_j-2} g_{j,k} g_{j,l} \left[\rho^{k-l}(\boldsymbol{r}_m, \boldsymbol{r}_n) + \rho^{-(k-l)}(\boldsymbol{r}_m, \boldsymbol{r}_n)\right] \sigma(\boldsymbol{r}_m)\, \sigma(\boldsymbol{r}_n)}{\sqrt{\left(\left(\frac{1}{2}\right)^j \sigma^2(\boldsymbol{r}_m) + 2\sum_{k=0}^{L_j-1} \sum_{l=k+1}^{L_j-2} g_{j,k} g_{j,l} \rho^{k-l}(\boldsymbol{r}_m)\, \sigma^2(\boldsymbol{r}_m)\right) \left(\left(\frac{1}{2}\right)^j \sigma^2(\boldsymbol{r}_n) + 2\sum_{k=0}^{L_j-1} \sum_{l=k+1}^{j} g_{j,k} g_{j,l} \rho^{k-l}(\boldsymbol{r}_n)\, \sigma^2(\boldsymbol{r}_n)\right)}}. \tag{52}$$

3.4 Dynamic equicorrelation

We will now consider the DECO model proposed by Engle and Kelly (2012), which we use as a non-spectral benchmark. The DECO model is a multivariate generalised autoregressive conditional heteroskedasticity (GARCH) model. These multivariate GARCH models are used to model the variancec and covariances of time series. We derive the DECO model as a computationally efficient estimator of the Dynamic Conditional Correlation (DCC) model proposed by Engle (2002). The DCC model is thus also a multivariate GARCH model and is a generalisation of the constant conditional correlation (CCC) model proposed by Bollerslev (1990).

The generalisation lies within the property of the DCC model to allow for a time varying correlation matrix, while the CCC model assumes a constant correlation over time. Engle (2002) has shown that the DCC model outperforms the CCC model, as well as other multivariate GARCH models and other covariance estimation methods such as moving averages and

exponential smoothing. This result even holds when the simulated covariance matrix is constant over time. The DCC model is one of most used methods for covariance estimation due to its good performance and the fact that the model is easy to implement and estimate with a two-step estimation procedure. In this estimation procedure, the first step is to estimate univariate GARCH models as described by Bollerslev (1986) for each individual time series. Thereafter, we estimate the covariance relations between the individual time series. In this thesis, when estimating the first step GARCH models, we use GARCH(1,1) models.

Engle (2002) defines the returns of k assets at time t, denoted by $\boldsymbol{r}_t$ given the filtration at time $t-1$, $\mathscr{F}_{t-1}$, to follow a multivariate normal distribution, i.e.,

$$\boldsymbol{r}_t \mid \mathscr{F}_{t-1} \sim N\left(0, \boldsymbol{H}_t\right), \tag{53}$$

where $\boldsymbol{H}_t$ is the covariance matrix at time t. It can be noted that when the returns are not normally distributed, the model still holds. In the case of non-normality, We can simply proceed using the regular estimation method, which will then result in quasi-maximum likelihood estimation. $\boldsymbol{H}_t$ is defined as

$$\boldsymbol{H}_t = \boldsymbol{D}_t \boldsymbol{R}_t \boldsymbol{D}_t, \tag{54}$$

where $\boldsymbol{R}_t$ is the correlation matrix at time t and $\boldsymbol{D}_t$ is a diagonal matrix of dimension $k \times k$ containing the time varying standard deviations from each univariate GARCH model. $\boldsymbol{D}_t$ can thus be expressed as

$$\boldsymbol{D}_t^2 = \operatorname{diag}\left\{\boldsymbol{\omega}\right\} + \operatorname{diag}\left\{\boldsymbol{\kappa}\right\} \odot \boldsymbol{r}_{t-1}\boldsymbol{r}_{t-1}' + \operatorname{diag}\left\{\boldsymbol{\lambda}\right\} \odot \boldsymbol{D}_{t-1}^2, \tag{55}$$

where $\boldsymbol{\omega}$ is a vector containing the estimated offset of the k estimated univariate GARCH models, $\boldsymbol{\kappa}$ is the vector of length k containing the estimated coefficients for the ARCH term of the univariate GARCH models and $\boldsymbol{\lambda}$ is the vector of length k containing the estimated coefficients for the GARCH term of each of the univariate GARCH models. $\operatorname{diag}\{\cdot\}$ denotes the diagonalisation operator and results in a matrix with the input vector on the diagonal. $\odot$ denotes the Hadamard matrix product operator. The vector containing the standardised residuals of the univariate GARCH models at time t is denoted as $\boldsymbol{\varepsilon}_t$ and is given as

$$\boldsymbol{\varepsilon}_t = \boldsymbol{D}_t^{-1}\boldsymbol{r}_t. \tag{56}$$

The modeled correlation structure is then given as

$$\boldsymbol{R}_t = \operatorname{diag}(\boldsymbol{Q}_t)^{-\frac{1}{2}} \boldsymbol{Q}_t \operatorname{diag}(\boldsymbol{Q}_t)^{-\frac{1}{2}}, \tag{57}$$

where $\operatorname{diag}(\boldsymbol{Q}_t)$ is given as

$$\operatorname{diag}(\boldsymbol{Q}_t) = \boldsymbol{I} \odot \boldsymbol{Q}_t. \tag{58}$$

$\boldsymbol{I}$ denotes a k by k identity matrix. $\boldsymbol{Q}$ is then expressed as

$$\boldsymbol{Q}_t = (1 - \alpha - \beta)\,\boldsymbol{S} + \alpha\left(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}'_{t-1}\right) + \beta\boldsymbol{Q}_{t-1}, \tag{59}$$

where $\boldsymbol{S}$ is the covariance matrix of the standardised residuals of the univariate GARCH models from the first estimation step. α and β are the DCC parameters to be estimated. We estimate α and β by maximising the log likelihood of $\boldsymbol{r}_t \mid \mathscr{F}_{t-1}$. The expression for the log likelihood L is given by

$$L = -\frac{1}{2}\sum_{t=1}^{T}\left(k\log(2\pi) + \log\left(|\boldsymbol{H}_t|\right) + \boldsymbol{r}'_t\boldsymbol{H}_t^{-1}\boldsymbol{r}_t\right). \tag{60}$$

We then fill in the expression for $\boldsymbol{H}_t$ to get

$$L = -\frac{1}{2}\sum_{t=1}^{T}\left(k\log(2\pi) + \log\left(|\boldsymbol{D}_t\boldsymbol{R}_t\boldsymbol{D}_t|\right) + \boldsymbol{r}'_t\boldsymbol{D}_t^{-1}\boldsymbol{R}_t^{-1}\boldsymbol{D}_t^{-1}\boldsymbol{r}_t\right). \tag{61}$$

Which we rewrite by using the expression for $\boldsymbol{\varepsilon}$, resulting in

$$L = -\frac{1}{2}\sum_{t=1}^{T}\left(k\log(2\pi) + 2\log|\boldsymbol{D}_t| + \log\left(|\boldsymbol{R}_t|\right) + \boldsymbol{\varepsilon}'_t\boldsymbol{R}_t^{-1}\boldsymbol{\varepsilon}_t\right). \tag{62}$$

Since $\boldsymbol{r}'_t\boldsymbol{D}_t^{-1}\boldsymbol{D}_t^{-1}\boldsymbol{r}_t - \boldsymbol{\varepsilon}'_t\boldsymbol{\varepsilon}_t = 0$, we can add this term to get

$$L = -\frac{1}{2}\sum_{t=1}^{T}\left(k\log(2\pi) + 2\log|\boldsymbol{D}_t| + \boldsymbol{r}'_t\boldsymbol{D}_t^{-1}\boldsymbol{D}_t^{-1}\boldsymbol{r}_t - \boldsymbol{\varepsilon}'_t\boldsymbol{\varepsilon}_t + \log\left(|\boldsymbol{R}_t|\right) + \boldsymbol{\varepsilon}'_t\boldsymbol{R}_t^{-1}\boldsymbol{\varepsilon}_t\right). \tag{63}$$

Which we then separate into two terms

$$L = L_1 + L_2, \tag{64}$$

where

$$L_1 = -\frac{1}{2}\sum_{t=1}^{T}\left(k\log(2\pi) + 2\log|\boldsymbol{D}_t| + \boldsymbol{r}'_t\boldsymbol{D}_t^{-1}\boldsymbol{D}_t^{-1}\boldsymbol{r}_t\right) \tag{65}$$

and

$$L_2 = -\frac{1}{2}\sum_{t=1}^{T}\left(\log\left(|\boldsymbol{R}_t|\right) + \boldsymbol{\varepsilon}_t'\boldsymbol{R}_t^{-1}\boldsymbol{\varepsilon}_t - \boldsymbol{\varepsilon}_t'\boldsymbol{\varepsilon}_t\right). \tag{66}$$

We can now see that L_1 is simply the sum of the univariate GARCH model log likelihood functions. This term is thus maximised by filling in the estimated univariate GARCH parameters of the first estimation step. For convenience, we remove the constant term from L_2 to get

$$L_2 = -\frac{1}{2}\sum_{t=1}^{T}\left(\log\left(|\boldsymbol{R}_t|\right) + \boldsymbol{\varepsilon}_t'\boldsymbol{R}_t^{-1}\boldsymbol{\varepsilon}_t\right). \tag{67}$$

Which we then maximise over α and β under the constraints of $\alpha \geq 0$, $\beta \geq 0$ and $\alpha + \beta < 1$. This maximisation over α and β is possible since $\boldsymbol{R}_t$ is a function of $\boldsymbol{Q}_t$, in which only α and β are unknown parameters.

Engle and Kelly (2012) proposed DECO as a feasible DCC estimator. The idea behind this estimator is imposing a constant correlation structure when maximising the likelihood of L_2. Due to this constant correlation structure, an analytical expression of the inverse of the correlation matrix exists. Replacing a numerically computed inverse of a matrix with an analytical expression reduces the computational time of the maximum likelihood estimation. Engle and Kelly (2012) also found to be the DECO estimator for the DCC model to outperform the regular DCC model, which they explain by claiming that the DECO estimator smooths out more of the estimation noise than it compromises on the true underlying correlation structure. The imposed correlation structure used by DECO is given by

$$\boldsymbol{R}_t^{\text{DECO}} = (1-\rho_t)\,\boldsymbol{I} + \rho_t\boldsymbol{J}, \tag{68}$$

where $\boldsymbol{J}$ denotes a k by k matrix of ones. ρ_t is given as

$$\rho_t = \frac{1}{k(k-1)}\left(\boldsymbol{\iota}'\boldsymbol{R}_t\boldsymbol{\iota} - k\right). \tag{69}$$

By applying the first theorem given in Appendix E, we can write the inverse of $\boldsymbol{R}_t^{\text{DECO}}$ as

$$\boldsymbol{R}_t^{\text{DECO}^{-1}} = \frac{1}{1-\rho_t}\left(\boldsymbol{I} - \frac{\rho_t}{1+(k-1)\rho_t}\boldsymbol{J}\right). \tag{70}$$

Using the second theorem of Appendix E, we rewrite the determinant of $\boldsymbol{R}_t^{\text{DECO}}$ as

$$|\boldsymbol{R}_t^{\text{DECO}}| = (1-\rho_t)^{k-1}(1+(k-1)\rho_t). \tag{71}$$

These expressions allow us to rewrite the log likelihood function L_2 to

$$\begin{aligned} L_2 &= -\frac{1}{2}\sum_{t=1}^{T}\left(\log\left(\left|\boldsymbol{R}_t^{\mathrm{DECO}}\right|\right) + \boldsymbol{\varepsilon}_t'\boldsymbol{R}_t^{\mathrm{DECO}^{-1}}\boldsymbol{\varepsilon}_t\right) \\ &= -\frac{1}{2}\sum_{t=1}^{T}\left(\log\left((1-\rho_t)^{k-1}(1+(k-1)\rho_t)\right) + \boldsymbol{\varepsilon}_t'\left(\frac{1}{1-\rho_t}\left(\boldsymbol{I} - \frac{\rho_t}{1+(k-1)\rho_t}\boldsymbol{J}\right)\right)\boldsymbol{\varepsilon}_t\right) \\ &= -\frac{1}{2}\sum_{t=1}^{T}\left(\log\left((1-\rho_t)^{k-1}(1+(k-1)\rho_t)\right)\right. \\ &\qquad\left. + \frac{1}{1-\rho_t}\left[\sum_{i=1}^{k}(\varepsilon_{i,t}^2) - \frac{\rho_t}{1+(n-1)\rho_t}\left(\sum_{i=1}^{k}\varepsilon_{i,t}\right)^2\right]\right). \end{aligned} \tag{72}$$

Which is the new likelihood function that we maximise to get the DECO estimates for α and β of the DCC model.

4 Simulation study

To evaluate the performance of the proposed covariance measures, we set up two simulation studies. The fist simulation study will be used to determine which of the wavelets within the Daubechies family performs the best for the wavelet-based covariance measure, as well as for the long-run wavelet-based correlation. In the second simulation study, we compare the performances of all proposed covariance measures for four different DGPs.

4.1 DGPs

The first DGP is a set of two discrete Hull-White short rate processes as described by Hull and White (1990), which are given by

$$\begin{aligned} \Delta y_{1,t} &= \kappa(\mu_{1,t} - y_{1,t}) + \sigma\varepsilon_{1,t}, \\ \Delta y_{2,t} &= \kappa(\mu_{2,t} - y_{2,t}) + \sigma\varepsilon_{2,t}, \end{aligned} \tag{73}$$

where $y_{1,t}$ and $y_{2,t}$ are the simulated rates given at time t, κ is the speed of mean reversion, $\mu_{1,t}$ and $\mu_{2,t}$ are the time varying means of the first and second process, respectively. $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ are the values of two independent standard normal random variables and σ is the

volatility of the process. We impose a sine function as the time varying mean to model a cyclical component in the time series. Such a cyclical component represents a financial cycle as described by Borio (2014). We then impose a correlation structure by giving the sine function a phase shift, leading to a lead-lag relationship. There exists strong evidence that such a lead-lag relationship is common in financial time series (Cohen & Frazzini, 2008; Hong et al., 2007; Hou & Moskowitz, 2005; Lo & MacKinlay, 1990). $\mu_{1,t}$ and $\mu_{2,t}$ are given by

$$\begin{aligned} \mu_{1,t} &= \mu_0 + a \sin\left(\frac{2\pi}{T} t\right), \\ \mu_{2,t} &= \mu_0 + a \sin\left(\frac{2\pi}{T} (t+\theta)\right), \end{aligned} \tag{74}$$

respectively. The means of $\mu_{1,t}$ and $\mu_{1,t}$ are given by μ_0, a is the amplitude of the sine function, T is the period of the sine given in number of time steps t and θ is the phase shift of the two sines.

The second DGP is a geometric Brownian motion with a time varying drift as given in Equation 74. The DGP is thus given by

$$\begin{aligned} y_{1,t} &= \exp\left(\left(\mu_{1,t} - \frac{\sigma^2}{2}\right) t + \sigma W_{1,t}\right), \\ y_{2,t} &= \exp\left(\left(\mu_{2,t} - \frac{\sigma^2}{2}\right) t + \sigma W_{2,t}\right), \end{aligned} \tag{75}$$

where $W_{1,t}$ and $W_{2,t}$ are two independent Brownian motions.

The third DGP is a geometric Brownian motion with the lead-lag correlation structure imposed by adding the sine functions given in Equation 74. Resulting in

$$\begin{aligned} y_{1,t} &= \exp\left(\left(\mu_0 - \frac{\sigma^2}{2}\right) t + \sigma W_{1,t}\right)\left(1 + a \sin\left(\frac{2\pi}{T} t\right)\right), \\ y_{2,t} &= \exp\left(\left(\mu_0 - \frac{\sigma^2}{2}\right) t + \sigma W_{2,t}\right)\left(1 + a \sin\left(\frac{2\pi}{T} (t+\theta)\right)\right). \end{aligned} \tag{76}$$

The correlation between the time series of the first three DGPs is given by

$$\cos\left(\frac{2\pi}{T}\theta\right), \tag{77}$$

as shown in Appendix F. The fourth DGP is a series of draws from a multivariate normal random variable,

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} \sim N_2\left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right]. \tag{78}$$

Table 1: MSE$\cdot 10^{-3}$ of the wavelet-based correlation for wavelet 2 to 10 of the Daubechies family for all four DGPs. Simulations are performed 10,000 times for generated time series of length 10,000.

	Number of vanishing moments of Daubechies wavelet								
DGP	2	3	4	5	6	7	8	9	10
1	58.700	52.855	51.094	49.810	48.262	47.514	47.719	47.652	46.793
2	46.533	42.893	41.007	40.633	40.681	40.036	39.242	39.223	39.563
3	8.851	8.421	8.336	7.908	7.629	7.739	7.889	7.665	7.651
4	0.523	0.627	0.702	0.766	0.821	0.875	0.929	0.979	1.028

4.2 Wavelet selection

For both the wavelet-based correlation and the long-run wavelet-based correlation, we test which wavelet performs the best in terms of mean squared error(MSE). This performance is tested by generating 10,000 pairs of time series of length 10,000 for each of the four aforementioned DGPs. For the first DGP we select $\kappa = 0.01$. A yearly volatility of 12.5%, resulting in $\sigma = 0.125/\sqrt{252}$, assuming a year contains 252 trading days. We choose a constant drift of zero, $\mu_0 = 0$. We set $a = 0.175$ and choose a yearly cycle resulting in $T = 252$ and a phase sift of 1 month, resulting in $\theta = 252/12 = 21$. For the second DGP, a yearly drift of 5% is selected, so $\mu_0 = 0.05/252$. the yearly volatility is set to 25%, resulting in $\sigma = 0.25/\sqrt{252}$. a is chosen to be equal to 0.005. We choose a period of one year, so $T = 252$ and a phase sift of 1 month, resulting in $\theta = 252/12 = 21$. For the third DGP, we select a yearly drift of 5%, resulting in $\mu_0 = 0.05/252$, we select a yearly volatility of 25%, resulting in $\sigma = 0.25/\sqrt{252}$. a is chosen to be equal to 0.33. We choose a yearly cycle, so $T = 252$ and a phase sift of 1 month, resulting in $\theta = 252/12 = 21$. For the fourth DGP, we select $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = \sqrt{2}$ and $\rho = 0.85$. We select to use the seventh scale for the wavelet-based correlation, meaning that we examine the correlation corresponding to 128 to 256 days. For the long-run wavelet-based correlation we select the sixth scale, meaning that we examine the correlation or 128 days onward. The resulting MSE for each wavelet when using the wavelet-based correlation is displayed in Table 1. The results for the long-run wavelet-based correlation of Conlon et al. (2018) are displayed in Table 2. From

Table 2: MSE$\cdot 10^{-3}$ of the long-run wavelet-based correlation for wavelet 2 to 10 of the Daubechies family for all four DGPs. Simulations are performed 10,000 times for generated time series of length 10,000.

	Number of vanishing moments of Daubechies wavelet								
DGP	2	3	4	5	6	7	8	9	10
1	244.834	238.511	236.123	235.157	234.777	234.647	234.623	234.644	234.679
2	48.568	45.768	44.722	44.272	44.065	43.967	43.919	43.895	43.882
3	8.999	8.415	8.196	8.100	8.053	8.029	8.016	8.008	8.002
4	0.404	0.427	0.440	0.448	0.453	0.458	0.461	0.464	0.466

Table 1, we observe that the wavelet-based correlation has the lowest MSE, thus performs better when using a higher wavelet for DGP1 and DGP2, while for DGP3 a wavelet in the middle of the range seems more suitable. The wavelet-based correlation performs the best for DGP4 when a low wavelet is selected. Combing these observations, we choose to use the Daubechies wavelet with 6 vanishing moments when working with the wavelet-based correlation from now on. According to Table 2, the long-run wavelet-based correlation performs the best for DGP1, DGP2 and DGP3 for a higher wavelet. the long-run wavelet-based correlation performs the best for DGP4 when using a low wavelet. Therefore, we choose to use Daubechies wavelet with 8 vanishing moments for the long-run wavelet-based correlation.

4.3 Simulation

We now perform a Monte Carlo simulation similar to the simulation used to select the used wavelets. We use the four aforementioned DGPs with the same parameters as for the simulations for the wavelet selection. Next to these simulations, we also simulate all four DGPs with a negative and zero correlation. For the first three DGPs the negative correlation is induced by setting θ to be equal to 5 months, i.e., $\theta = 5/12 \cdot 252 = 105$ days. This value for θ ensures that the underlying cycles of the DGPs are almost in antiphase. For the zero correlation of the first three DGPs, the θ is set to be equal to 3 months, so $\theta = 3/12 \cdot 252 = 63$ days. For the fourth DGP, we set the negative and zero correlation by setting ρ to be equal to -0.85 and 0,

respectively.

We then apply six methods of estimating the correlation, of which the last two are the non-spectral methods.

1. The Fourier-based correlation with frequencies of interest ranging from once in 128 to 256 days. This range is chosen to be equal to range of the seventh wavelet scale. The corresponding frequency bins are 39 until 79.
2. The same Fourier-based correlation, but now for frequency bins more specifically for periods of 1 year. The frequency bins chosen for this method are 38 to 42, corresponding to a frequency of once in 238 to 263 days.
3. The wavelet-based correlation, using the Daubechies wavelet with 6 vanishing moments and selecting the seventh scale.
4. The long-run wavelet-based correlation, with the Daubechies wavelet with 8 vanishing moments and selecting the sixth scale.
5. The DECO model
6. The regular sample correlation coefficient of the series.

We run 1,000 simulations, where the length of each pair of generated time-series is equal to 10,000. The results are shown in Table 3. In Table 3, we see that both the Fourier-based correlation measure, as well as the specific Fourier-based correlation perform very well overall. Where, as expected the specific Fourier-based correlation generally outperforms the Fourier-based correlation when cyclical behaviour is present, while it is the other way around for the fourth DGP, which does not have cyclical behaviour. We attribute this finding to a bias-variance trade-off, by setting the Fourier measure more specific, we see that the estimator will be less biased when the cyclical behaviour is present, while the estimator appears to have a higher variance when no cyclical behaviour is present. Next, we see that the wavelet-based correlation performs very well overall, it outperforms the Fourier correlation when no cyclical behaviour is present, but gets outperformed by the Fourier correlation for the DGPs where cyclical behaviour is present. We attribute this to the fact that the first three DGPs have their correlation induced by sine functions, which are generally better picked up by the Fourier transform, while the

Table 3: MSE of the correlation estimation for all six methods, for all four DGPs. We use each DGP three times with different parameter settings for the correlation. Simulations are performed 1,000 times for generated time series of length 10,000

		method					
DGP	correlation	Fourier	Fourier specific	wavelet	long-run wavelet	DECO	Sample
1	positive	$2.1716 \cdot 10^{-2}$	$3.3043 \cdot 10^{-3}$	$4.7883 \cdot 10^{-2}$	0.23287	1.4965	0.2582
	negative	$2.2575 \cdot 10^{-2}$	$3.5779 \cdot 10^{-3}$	$4.9021 \cdot 10^{-2}$	0.2378	1.4483	0.2628
	zero	$6.1337 \cdot 10^{-3}$	$7.9323 \cdot 10^{-3}$	$5.7416 \cdot 10^{-3}$	$5.7064 \cdot 10^{-3}$	0.7893	$4.8126 \cdot 10^{-3}$
2	positive	$1.6094 \cdot 10^{-2}$	$1.3201 \cdot 10^{-3}$	$4.0526 \cdot 10^{-2}$	$4.3740 \cdot 10^{-2}$	0.6802	0.6810
	negative	$1.7262 \cdot 10^{-2}$	$1.4164 \cdot 10^{-3}$	$4.1346 \cdot 10^{-2}$	$4.5492 \cdot 10^{-2}$	0.6823	0.68102
	zero	$3.3050 \cdot 10^{-2}$	$4.0816 \cdot 10^{-3}$	$3.1333 \cdot 10^{-3}$	$2.6603 \cdot 10^{-3}$	$2.8222 \cdot 10^{-4}$	$9.4309 \cdot 10^{-5}$
3	positive	$2.4937 \cdot 10^{-3}$	$3.9443 \cdot 10^{-4}$	$7.5025 \cdot 10^{-3}$	$7.8643 \cdot 10^{-3}$	0.5794	0.5709
	negative	$2.4488 \cdot 10^{-3}$	$3.7988 \cdot 10^{-4}$	$1.2052 \cdot 10^{-2}$	$1.0152 \cdot 10^{-2}$	0.5702	0.5754
	zero	$1.3626 \cdot 10^{-3}$	$1.4412 \cdot 10^{-3}$	$1.7326 \cdot 10^{-3}$	$1.3000 \cdot 10^{-3}$	$2.0060 \cdot 10^{-3}$	$1.07151 \cdot 10^{-4}$
4	positive	$1.0252 \cdot 10^{-3}$	$1.1484 \cdot 10^{-2}$	$8.4392 \cdot 10^{-4}$	$4.5975 \cdot 10^{-4}$	$1.2114 \cdot 10^{-5}$	$7.4965 \cdot 10^{-6}$
	negative	$1.0511 \cdot 10^{-3}$	$1.1465 \cdot 10^{-2}$	$8.3434 \cdot 10^{-4}$	$4.4329 \cdot 10^{-4}$	$1.2592 \cdot 10^{-5}$	$7.7053 \cdot 10^{-6}$
	zero	$1.2753 \cdot 10^{-2}$	$9.4540 \cdot 10^{-2}$	$1.0374 \cdot 10^{-2}$	$5.9481 \cdot 10^{-3}$	$2.0257 \cdot 10^{-4}$	$9.6551 \cdot 10^{-5}$

wavelet transform picks up all cyclical behaviour and not just sine functions. So, this can also be seen as a bias-variance trade-off. The long-run wavelet correlation seems to perform similar to the wavelet-based correlation, except for DGP1 where it has a substantially larger MSE, from this we conclude that the long-run wavelet correlation generally performs well when the underlying time series contain cyclical behaviour and less when the returns themselves are cyclical. DECO seems to perform extremely well under perfect specification of a multivariate normal distribution, but behaves poorly when the correlation is induced by cyclical behaviour. We even see MSEs of almost 1.5 for DGP1 when using the DECO algorithm, which is very high when estimating correlations. Looking closer at the raw outputs, we find that the absolute estimate for the correlation is often close to the actual value, but with an inverted sign, resulting in large errors. We see similar results for the sample correlation as for DECO. Even though that, unlike DECO, the sample correlation is non-parametric. From these findings, we conclude that the Fourier and wavelet-based methods of correlation estimation perform well in every scenario, while DECO only performs accurately when using the model under perfect specification. This result highlights the advantages of using model free spectral domain techniques such as the wavelet transform and Fourier transform

5 Empirical study

5.1 Application to systematic risk

The first application of our spectral covariance measures is to systematic risk, or the CAPM beta of Sharpe (1964). The CAPM beta of an asset i is given as

$$\beta_i = \frac{\text{Cov}\left(\boldsymbol{r}_i, \boldsymbol{r}_m\right)}{\sigma_m^2}, \tag{79}$$

where $\boldsymbol{r}_i$ denotes the vector of returns of asset i, $\boldsymbol{r}_m$ denotes the vector of market returns and σ_m^2 denotes the variance of the market. According to the CAPM, there should be a linear relationship between the beta of an asset and the expected returns of that asset. Jensen et al. (1972) first showed that the CAPM relationship between beta and returns does not hold by creating beta sorted portfolios and showing that the returns the relationship was inverted, i.e.,

low beta stocks tended to have higher returns compared to high beta stocks. Reinganum (1981) later came to the same conclusion that the CAPM relationship does not hold, but that the beta sorted portfolios take on a parabolic shape in their returns. More recent papers, e.g., Bali et al. (2017) show that the expected relationship still does not hold and that this relation still takes on a parabolic shape as opposed to the expected linear relation. There even exists a strategy called betting against beta, introduced in the influential paper of Frazzini and Pedersen (2014) that exploits this phenomenon. The observation that low beta stocks usually have higher returns compared to the low returns of high beta stocks is referred to as the beta anomaly (Bali et al., 2017). We will conduct a similar experiment by creating beta sorted portfolios, but extend the estimator for beta in the spectral domain by calculating horizon specific betas using the long-run wavelet-based covariance estimator, the wavelet-based covariance estimator and the Fourier-based estimator.

For the analysis, we use the 505 individual stocks in the S&P500 at the time of writing from 3 January 2005 until 28 May 2021, sampled with a daily frequency. We exclude the stocks that had their initial public offering after 3 January 2005 to ensure that all the time series have the sane length. This approach leaves us with 411 assets. All missing values are linearly interpolated. For the market, we use the S&P500 index over the same period.

In an approach similar to Reinganum (1981) and Gençay et al. (2005), we calculate log price differences $r_{i,t}$ of asset i at time t for all assets, i,e. $r_{i,t} = \log(p_{i,t}) - \log(p_{i,t-1})$, where $p_{i,t}$ is the price asset i at time t. We also calculate the log price differences for the market, denoted by $r_{m,t}$. We then apply a moving window of 8 years to the data and determine the beta of each asset over the time period of the window with the preferred method. We then sort the stocks from high beta to low beta and create 15 portfolios based on the beta of the stocks and determine the average beta as well as the average daily returns of the portfolios.[2] We first estimate the beta with the regular sample covariance estimator. The results are displayed in Figure 4. In the figure we clearly see the parabolic relation of the beta anomaly. We now use the long-run wavelet-based covariance measure for scale 1 to 9 to estimate the beta for each of these 9 scales. The results are shown in Figure 5. In this figure we clearly see that as the wavelet scale becomes

[2]For robustness, the moving window length and number of portfolios were varied, yielding similar results.

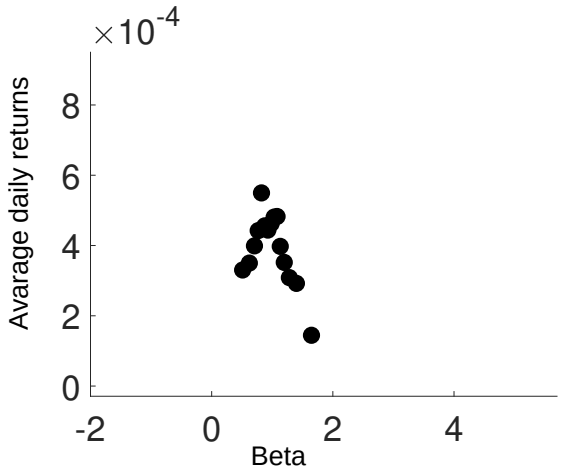


Figure 4: Scatterplot of the beta of the portfolios against their average daily returns, the betas are computed by using the sample covariance function. the x-axis corresponds to the beta of the portfolios, the y-axis corresponds to the average daily returns.

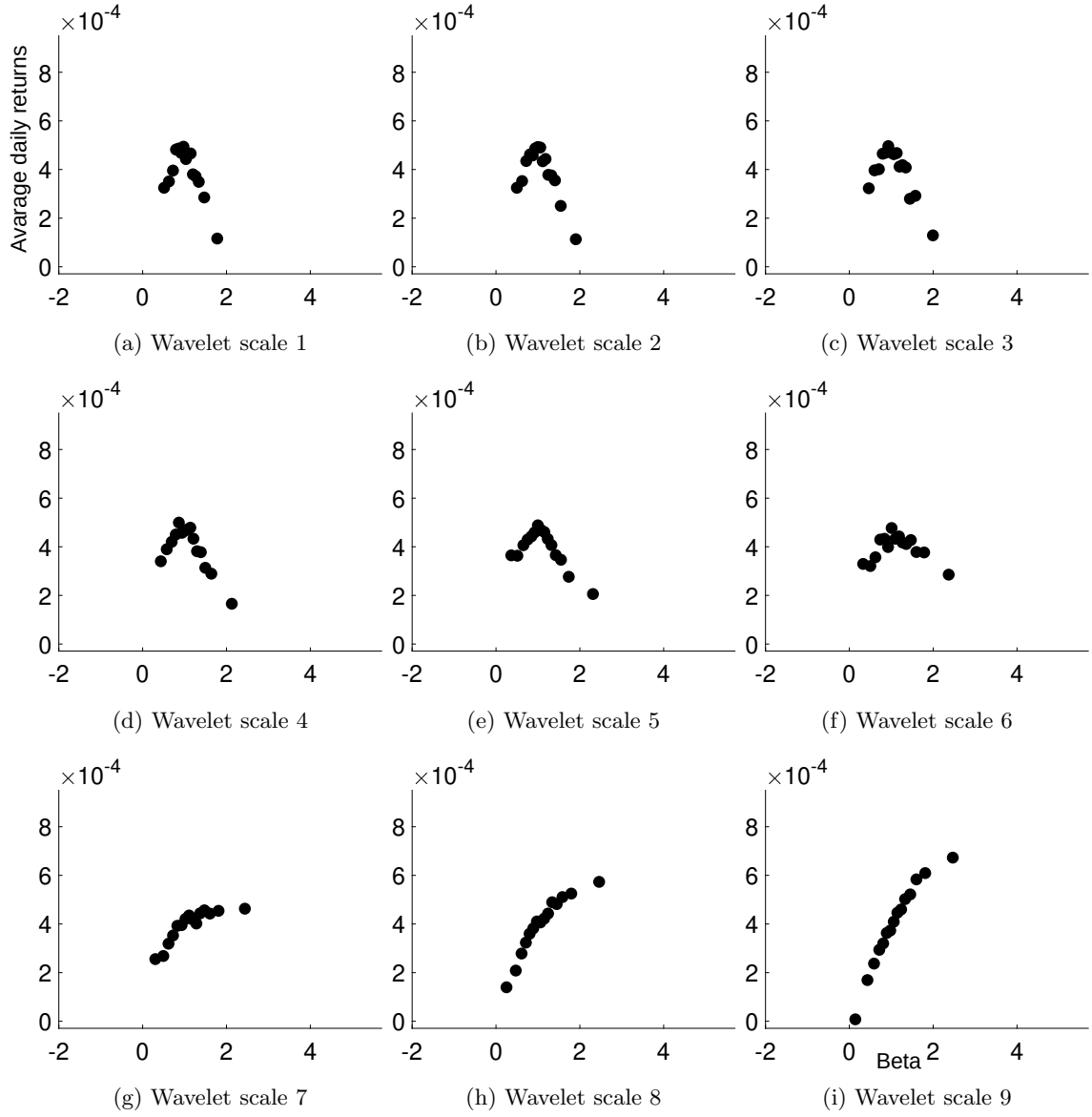
higher, the CAPM relationship between the beta and the returns of a portfolio becomes stronger. Next, we apply the wavelet-based covariance measure. The results are shown in Figure 6, we see that the relationship does become stronger with the increase of the scale. However, the relationship at the higher scales is not quite as strong as seen with the long-run wavelet-based covariance measure. This is caused by the fact that the long-run estimator also takes the scales after the selected scale into account, as opposed to the wavelet-based measure, which only takes into account the selected scale. Therefore, we conclude that the CAPM relationship becomes the strongest after the ninth wavelet scale, corresponding to a horizon of about four years. The Fourier-based covariance measure yields similar results to the wavelet-based covariance measure when we examine the same frequencies of interest in the Fourier-based covariance measure that correspond to the wavelet scales. However, the Fourier-based covariance measure provided more flexibility due to the fact that we can select the frequencies of interest. As such, we are not restricted to specific wavelet scales and have the possibility to split up the examined scales into finer bins. So we now split up the 8 years of the moving window into 16 evenly spaced Fourier frequency bins and compute the Fourier-based betas for each of these frequency bins. This results in Figure 7. From this figure, we can confirm our previous hypothesis that the CAPM relationship becomes the strongest after more than four years. We also see that the relationship

becomes especially strong after around six years.

The finding that the relationship between systematic risk and returns becomes strong after six years indicates that there exists a long term risk premium. This finding is in alignment with the finding of Bandi and Perron (2008), who show that there exists a strong risk premium for horizons of six to ten years by regressing excess market returns on past market variance. In a later study, Bandi et al. (2019) show that the risk premium follows a hump shaped curve, as a function of the time horizon, which is the strongest for a horizon of six to twenty years, but weakens thereafter.

The finding of the negative relation between high systematic risk and returns in the short term also aligns with existing literature, especially with the notions of the volatility feedback effect and the leverage effect. The volatility feedback effect, discussed in, e.g., Bollerslev et al. (2006), describes the idea that there should be a time-varying risk premium if volatility, or another risk measure, is priced. When an increase in volatility is expected, the future expected returns in the long-run should be higher, driving the short term prices down to allow for these higher returns. Thus, an increase in risk decreases the price in the short term implying that an increase in risk leads a decrease in price. The leverage effect, described in, e.g., Aït-Sahalia et al. (2013), describes the observation that when the price of a stock declines, the debt to equity ration of that stock increases. Thus, the stock becomes riskier. So, according to the leverage effect, a decrease in price should lead to an increase in risk implying that the risk increase lags the decrease in price. In this application we do not gain information on the lead-lag relationship between systematic risk and returns, so we can not conclusively say which effect plays a more important role in our analysis. However, due to the fact that, for a high beta, we observe both the increase in returns in the long-run and the decrease of the short term returns, we argue that the volatility feedback effect seems more likely to play an important role in this application.

We conclude that the CAPM relation between beta and returns becomes stronger for greater horizons. The results are very strong for a horizon greater than six years. Implying that there is a long term risk premium and if we are allocating assets for a horizon of more than six years and use an appropriate covariance measure to compute beta, we can actually expect a strong positive relationship between the beta and the returns.

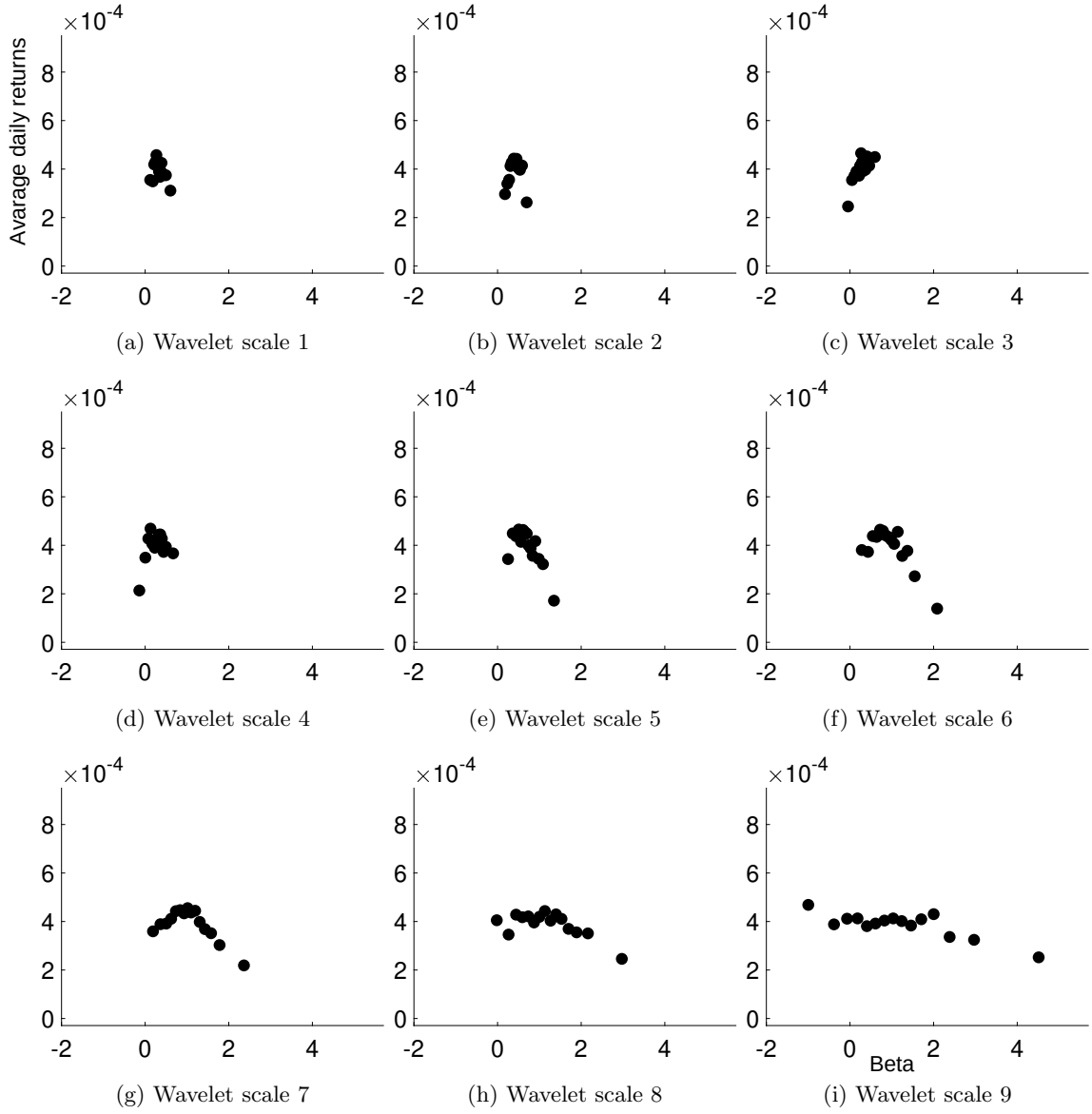


(a) Wavelet scale 1 (b) Wavelet scale 2 (c) Wavelet scale 3

(d) Wavelet scale 4 (e) Wavelet scale 5 (f) Wavelet scale 6

(g) Wavelet scale 7 (h) Wavelet scale 8 (i) Wavelet scale 9

Figure 5: Scatterplots for the portfolio betas of wavelet scale 1 to 9 against the average daily returns, the betas are computed by using the long-run wavelet-based covariance measure. The x-axis corresponds to the beta of the portfolios, the y-axis corresponds to the average daily returns. Each scatterplot correspond to the following wavelet scale: (a) 1, (b) 2, (c) 3, (d) 4, (e) 5, (f) 6, (g), 7, (h) 8, (i) 9.

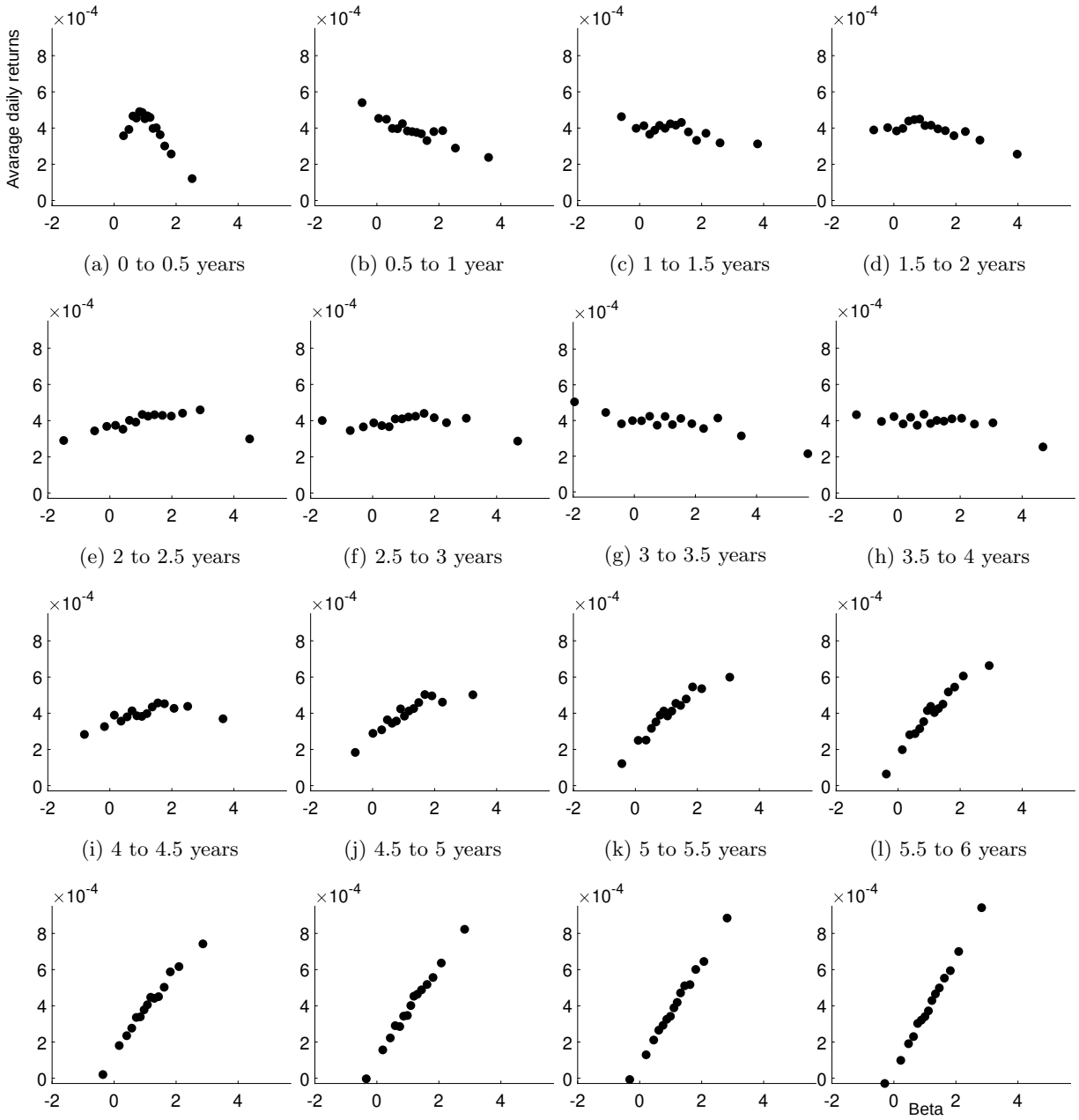


(a) Wavelet scale 1 (b) Wavelet scale 2 (c) Wavelet scale 3

(d) Wavelet scale 4 (e) Wavelet scale 5 (f) Wavelet scale 6

(g) Wavelet scale 7 (h) Wavelet scale 8 (i) Wavelet scale 9

Figure 6: Scatterplots for the portfolio betas of wavelet scale 1 to 9 against the average daily returns, the betas are computed by using the wavelet-based covariance measure. The x-axis corresponds to the beta of the portfolios, the y-axis corresponds to the average daily returns. Each scatterplot correspond to the following wavelet scale: (a) 1, (b) 2, (c) 3, (d) 4, (e) 5, (f) 6, (g), 7, (h) 8, (i) 9.



(a) 0 to 0.5 years (b) 0.5 to 1 year (c) 1 to 1.5 years (d) 1.5 to 2 years

(e) 2 to 2.5 years (f) 2.5 to 3 years (g) 3 to 3.5 years (h) 3.5 to 4 years

(i) 4 to 4.5 years (j) 4.5 to 5 years (k) 5 to 5.5 years (l) 5.5 to 6 years

(m) 6 to 6.5 years (n) 6.5 to 7 years (o) 7 to 7.5 years (p) 7.5 to 8 years

Figure 7: Scatterplots for the Fourier-based betas of the portfolios against the average daily returns, the betas are computed by using the Fourier-based covariance measure and are spaced evenly over 8 years from (a) to (p), i.e. a jump of half a year for each figure. The x-axis corresponds to the beta of the portfolios, the y-axis corresponds to the average daily returns.

5.2 Application to asset allocation

Another application and widely used test for covariance measures is the construction of portfolios based on the covariance matrices. We will construct minimum variance portfolios. For minimum variance portfolios only require to estimate the covariance matrix as opposed to for example mean-variance portfolios, which also require to make an estimate for the future expected returns.

We create minimum variance portfolios by minimising the expression for the variance of the portfolio. The expression for the portfolio variance is given as

$$\sigma_p^2 = \boldsymbol{w}' \boldsymbol{\Sigma} \boldsymbol{w}, \tag{80}$$

where $\boldsymbol{w}$ is the vector of weights and $\boldsymbol{\Sigma}$ is the covariance matrix of the assets. We want all weights of the portfolio to sum up to 1. So we minimise the expression for σ_p^2 over the portfolio weights, $\boldsymbol{w}$, under the constraint that

$$\boldsymbol{w}\boldsymbol{\iota}' = 1, \tag{81}$$

where $\boldsymbol{\iota}$ is a vector of ones. The analytical solution to this minimisation problem is given by

$$\boldsymbol{w} = \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{\iota}}{\boldsymbol{\iota}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\iota}}. \tag{82}$$

The data used are the returns of the main stock indices of 15 countries. The list of countries and the corresponding indices can be found in Table 6 in Appendix A. We convert all indices to EUR before computing returns. The returns are over the period from 10 April 2007 to 7 April 2021 and are sampled with a daily frequency.

For the analysis, we apply a moving window of 10 years to the data and calculate the weights for the minimum variance portfolios using the chosen covariance measure, which we will use as our estimate of the covariance matrix for the next day. We then shift the moving window by one day and estimate the weights again. We continue this estimation procedure until we have covered 1008 days, or 4 years when we assume a year contains 252 trading days. We thus have an out-of-sample period of four years.

For the estimation of the covariance matrices, we use six methods. The first method is by using the Fourier-based covariance measure, where we select the frequency bins corresponding to a period of 1 to 4 days. We chose such a short horizon because we rebalance the portfolio every

day. The second method used is the wavelet-based covariance measure, where we select the first wavelet scale. The third method is the long-run wavelet-based covariance measure, where we also use the first wavelet scale. We use DECO as the fourth estimation method. For the fifth estimation method, we use the regular sample covariance estimator. For the sixth method, we do not use an estimated covariance matrix, but use a naive strategy and create equal weight portfolios. So for our 15 assets, all values of $\boldsymbol{w}$ will be equal to 1/15. The sample variances of the methods in the out-of-sample period are shown in Table 4. The sample variances are computed for the time series of percentage returns of the strategies. In Table 4, we see that

Table 4: Sample variance over the four year out-of-sample period of time series of percentage returns of the constructed portfolios using the six different methods for the estimation of the covariance matrix.

Method	Variance
Fourier	0.2700
wavelet-based	0.2997
Long-run wavelet-based	0.2569
DECO	0.2797
Sample	0.2644
Equal weight	0.2972

only the long-run wavelet-based covariance outperforms the portfolio constructed by the sample covariance matrix. This is a notable result, since the sample covariance matrix can be seen as one of the simpler methods. We also see that the wavelet-based covariance measure performs the worst, closely followed by the most naive strategy of equal weight. The Fourier-based method, as well as DECO seem to perform average compared to the rest.

To formally test whether the variances of the constructed portfolios actually differ from each other, we perform the robust variance test of Ledoit and Wolf (2011). We use this test instead of regular F-test because the returns are correlated and their distribution has heavier tails compared to a normal distribution, which means that two of the assumptions of the F-test do not hold. The p-values of the test for all pairs of portfolio constructions are shown in

Table 5. The test has as null hypothesis that the variances are equal and tests this hypothesis by creating bootstrapped confidence intervals. Surprisingly, Table 5 shows that DECO does have

Table 5: p-values of the robust variance test between the pairs of constructed portfolio.

	Fourier	Wavelet	Long-run wavelet	DECO	Sample	Equal weight
Fourier		0.000	0.000	0.731	0.000	0.009
Wavelet			0.000	0.132	0.000	0.639
Long-run wavelet				0.231	0.001	0.001
DECO					0.540	0.131
Sample						0.001
Equal weight						

a statistically different performance compared to all the other methods. All other pairs, except for the wavelet-based and equal weight pair, seem to perform significantly different from each other on even a 1% level. Therefore, we conclude that the long-run wavelet-based approach performs the best. We conclude that, while there is evidence that using spectral covariance measures is beneficial in asset allocation, the benefits are relatively small.

6 Discussion and Conclusion

In conclusion, we find that spectral domain methods such as the proposed Fourier-based covariance measure, the wavelet-based covariance of Whitcher et al. (2000) and the long-run wavelet-based covariance of Conlon et al. (2018) can contribute greatly to the estimation of covariance. Especially when certain cyclical behaviour is present or when the data does not follow a multivariate normal distribution as under the assumption of the DECO algorithm of Engle and Kelly (2012).

In the empirical application to systematic risk, we find that the relationship between asset returns and the CAPM beta becomes very strong at greater horizons, indicating a long-run risk premium. In the empirical application to minimum variance asset allocation, we find

evidence that the long-run wavelet-based covariance measure of Conlon et al. (2018) marginally outperforms the other methods. The robust variance test of Ledoit and Wolf (2011) indicates that this outperformance is statistically significant to all other methods, except the DECO algorithm of Engle and Kelly (2012).

We argue that the covariance estimates for the application to systematic risk performs the best at greater horizons because we filter out all the short term noise, allowing for the true covariance relationship between the assets and the market to appear. We also argue that long term dynamics seem to be important in the estimation of covariance matrices for minimum variance portfolios, so filtering the long term dynamics out completely does not really improve the minimum variance allocation. Filtering out the noise of the shortest term, while still preserving the long term dynamics, as done with the long-run wavelet-based covariance of Conlon et al. (2018), performs the best and is therefore seen as evidence of the importance of the long term dynamics in the estimation of the covariances.

Suggested future work could be to perform a lead-lag analysis with the phase information of the Fourier transform to discern whether the observed short term effect in the application to systemic risk is due to the volatility feedback effect or the leverage effect. More future work could be to extend the concept of spectral domain betas to other factors, so to extend classical factor models to spectral domain factor models. Another suggestion for future work is to extend upon the minimum variance allocation, by combining the spectral domain estimation procedure for the covariance with commonly used techniques in portfolio management such as shrinkage and applying constraints.

References

Aït-Sahalia, Y., Fan, J., & Li, Y. (2013). The leverage effect puzzle: Disentangling sources of bias at high frequency. *Journal of Financial Economics*, *109*(1), 224–249.

Bali, T., Brown, S., Murray, S., & Tang, Y. (2017). A lottery-demand-based explanation of the beta anomaly. *Journal of Financial and Quantitative Analysis*, *52*(6), 2369–2397.

Bandi, F. M., Perron, B., Tamoni, A., & Tebaldi, C. (2019). The scale of predictability. *Journal of Econometrics*, *208*(1), 120–140.

Bandi, F. M., & Perron, B. (2008). Long-run risk-return trade-offs. *Journal of Econometrics*, *143*(2), 349–374.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. *The Review of Economics and Statistics*, *72*(3), 498–505.

Bollerslev, T., Litvinova, J., & Tauchen, G. (2006). Leverage and volatility feedback effects in high-frequency data. *Journal of Financial Econometrics*, *4*(3), 353–384.

Borio, C. (2014). The financial cycle and macroeconomics: What have we learnt? *Journal of Banking & Finance*, *45*, 182–198.

Cohen, L., & Frazzini, A. (2008). Economic links and predictable returns. *The Journal of Finance*, *63*(4), 1977–2011.

Conlon, T., Cotter, J., & Gençay, R. (2018). Long-run wavelet-based correlation for financial time series. *European Journal of Operational Research*, *271*(2), 676–696.

Crowley, P. M. (2007). A guide to wavelets for economists. *Journal of Economic Surveys*, *21*(2), 207–267.

Daubechies, I. (1992). *Ten lectures on wavelets*. Society for Industrial and Applied Mathematics.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, *20*(3), 339–350.

Engle, R., & Kelly, B. (2012). Dynamic equicorrelation. *Journal of Business & Economic Statistics*, *30*(2), 212–228.

Frazzini, A., & Pedersen, L. H. (2014). Betting against beta. *Journal of Financial Economics*, *111*(1), 1–25.

Gençay, R., Selçuk, F., & Whitcher, B. (2002). *An introduction to wavelets and other filtering methods in finance and economics*. Academic Press.

Gençay, R., Selçuk, F., & Whitcher, B. (2005). Multiscale systematic risk. *Journal of International Money and Finance*, *24*(1), 55–70.

Harris, F. J. (1978). On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE*, *66*(1), 51–83.

Hong, H., Torous, W., & Valkanov, R. (2007). Do industries lead stock markets? *Journal of Financial Economics*, *83*(2), 367–396.

Hou, K., & Moskowitz, T. J. (2005). Market frictions, price delay, and the cross-section of expected returns. *The Review of Financial Studies*, *18*(3), 981–1020.

Hull, J., & White, A. (1990). Pricing interest-rate-derivative securities. *Review of Financial Studies*, *3*(4), 573–592.

Jensen, M. C., Black, F., & Scholes, M. S. (1972). The capital asset pricing model: Some empirical tests. *Capital markets: Asset pricing & valuation*. Praeger Publishers Inc.

Ledoit, O., & Wolf, M. (2011). Robust performances hypothesis testing with the variance. *Wilmott*, *2011*(55), 86–89.

Ledoit, O., & Wolf, M. (2020). The Power of (Non-)Linear Shrinking: A Review and Guide to Covariance Matrix Estimation. *Journal of Financial Econometrics*, Advance online publication.

Lo, A. W., & MacKinlay, A. C. (1990). When are contrarian profits due to stock market overreaction? *The Review of Financial Studies*, *3*(2), 175–205.

Mallat, S. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *11*(7), 674–693.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, *7*(1), 77–91.

Nuttall, A. (1981). Some windows with very good sidelobe behavior. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *29*(1), 84–91.

Percival, D. B., & Mofjeld, H. O. (1997). Analysis of subtidal coastal sea level fluctuations using wavelets. *Journal of the American Statistical Association*, *92*(439), 868–880.

Rahman, M. (2011). *Applications of Fourier transforms to generalized functions*. WIT Press.

Ramsey, J. B. (1999). The contribution of wavelets to the analysis of economic and financial data. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *357*(1760), 2593–2606.

Reinganum, M. R. (1981). A new empirical perspective on the CAPM. *The Journal of Financial and Quantitative Analysis*, *16*(4), 439–462.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, *19*(3), 425–442.

Whitcher, B., Guttorp, P., & Percival, D. B. (2000). Wavelet analysis of covariance with application to atmospheric time series. *Journal of Geophysical Research: Atmospheres*, *105*(D11), 14941–14962.

Appendix A Additional tables

Table 6: Index corresponding with each country.

Country	Index
Australia	S&P/ASX 200
Canada	S&P/TSX 60
China	SSE Composite Index
France	CAC 40
Germany	DAX 30
India	Nifty-50
Italy	FTSE MIB
Japan	TOPIX
Netherlands	AEX
Russia	MOEX
Spain	IBEX-35
Sweden	OMX Stockholm 30
Switzerland	SMI
United Kingdom	FTSE 100
United States	S&P 500

Table 7: Frequencies corresponding to different scales for different sample frequencies. Assuming 252 trading days per year, 21 trading days per month.

	sample frequency		
Scale	yearly	monthly	daily
1	2-4	2-4	2-4
2	4-8	4-8	4-8
3	8-16	8-16 = 8m-1y4m	8–16
4	16-32	16-32 = 1y4m-2y8m	16-32 = 16d-1m11d
5	32-64	32–64 = 2y8m–5y4m	32-64 = 1m11d-3m1d
6	64-128	64–128 = 5y4m–10y8m	64-128 = 3m1d-6m2d
7	128-256	128–256 = 10y8m–21y4m	128-256 = 6m2d-1y4d
8	256-512	256-512 = 21y4m-42y8m	256-512 = 1y4d-2y8d
9	512-1024	512-1024 = 42y8m-83y4m	512-1024 = 2y8d-4y16d

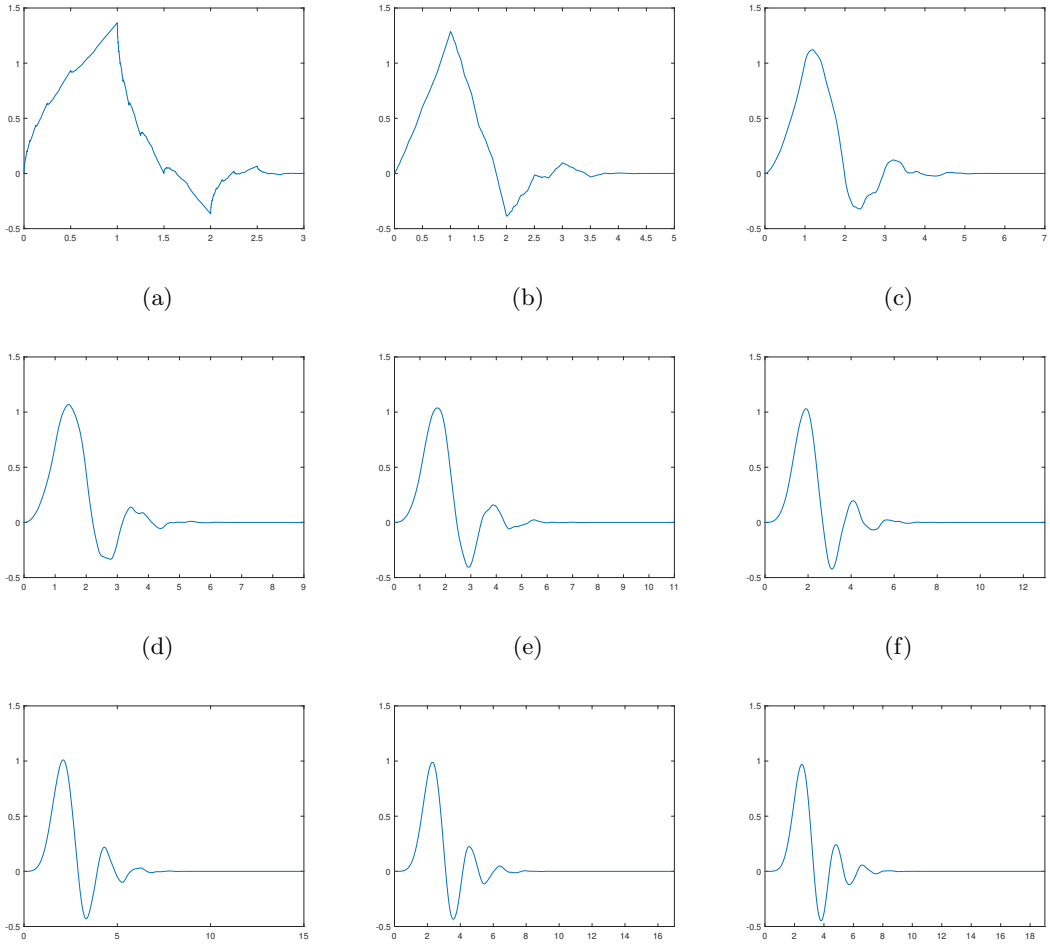
Table 8: Filter coefficients for Daubechies scaling filters with 2 to 10 vanishing moments.

	Number of vanishing moments								
index	2	3	4	5	6	7	8	9	10
1	0.4830	0.3327	0.2304	0.1601	0.1115	0.0779	0.0544	0.0381	0.0267
2	0.8365	0.8069	0.7148	0.6038	0.4946	0.3965	0.3129	0.2438	0.1882
3	0.2241	0.4599	0.6309	0.7243	0.7511	0.7291	0.6756	0.6048	0.5272
4	-0.1294	-0.1350	-0.0280	0.1384	0.3153	0.4698	0.5854	0.6573	0.6885
5		-0.0854	-0.1870	-0.2423	-0.2263	-0.1439	-0.0158	0.1332	0.2812
6		0.0352	0.0308	-0.0322	-0.1298	-0.2240	-0.2840	-0.2933	-0.2498
7			0.0329	0.0776	0.0975	0.0713	4.7248e-04	-0.0968	-0.1959
8			-0.0106	-0.0062	0.0275	0.0806	0.1287	0.1485	0.1274
9				-0.0126	-0.0316	-0.0380	-0.0174	0.0307	0.0931
10				0.0033	5.5384e-04	-0.0166	-0.0441	-0.0676	-0.0714
11					0.0048	0.0126	0.0140	2.5095e-04	-0.0295
12					-0.0011	4.2958e-04	0.0087	0.0224	0.0332
13						-0.0018	-0.0049	-0.0047	0.0036
14						3.5371e-04	-3.9174e-04	-0.0043	-0.0107
15							6.7545e-04	0.0018	0.0014
16							-1.1748e-04	2.3039e-04	0.0020
17								-2.5196e-04	-6.8586e-04
18								3.9347e-05	-1.1647e-04
19									9.3589e-05
20									-1.3264e-05

Table 9: Filter coefficients for Daubechies wavelet filters with 2 to 10 vanishing moments.

	Number of vanishing moments								
index	2	3	4	5	6	7	8	9	10
1	-0.1294	0.0352	-0.0106	0.0033	-0.0011	3.5371e-04	-1.1748e-04	3.9347e-05	-1.3264e-05
2	-0.2241	0.0854	-0.0329	0.0126	-0.0048	0.0018	-6.7545e-04	2.5196e-04	-9.3589e-05
3	0.8365	-0.1350	0.0308	-0.0062	5.5384e-04	4.2958e-04	-3.9174e-04	2.3039e-04	-1.1647e-04
4	-0.4830	-0.4599	0.1870	-0.0776	0.0316	-0.0126	0.0049	-0.0018	6.8586e-04
5		0.8069	-0.0280	-0.0322	0.0275	-0.0166	0.0087	-0.0043	0.0020
6		-0.3327	-0.6309	0.2423	-0.0975	0.0380	-0.0140	0.0047	-0.0014
7			0.7148	0.1384	-0.1298	0.0806	-0.0441	0.0224	-0.0107
8			-0.2304	-0.7243	0.2263	-0.0713	0.0174	-2.5095e-04	-0.0036
9				0.6038	0.3153	-0.2240	0.1287	-0.0676	0.0332
10				-0.1601	-0.7511	0.1439	-4.7248e-04	-0.0307	0.0295
11					0.4946	0.4698	-0.2840	0.1485	-0.0714
12					-0.1115	-0.7291	0.0158	0.0968	-0.0931
13						0.3965	0.5854	-0.2933	0.1274
14						-0.0779	-0.6756	-0.1332	0.1959
15							0.3129	0.6573	-0.2498
16							-0.0544	-0.6048	-0.2812
17								0.2438	0.6885
18								-0.0381	-0.5272
19									0.1882
20									-0.0267

Appendix B Additional figures



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 8: Plots of the scaling functions generated by the cascade algorithm using the coefficients of Table 8. Each scaling function corresponds to the wavelet with the following number of vanishing moments: (a) 2, (b) 3, (c) 4, (d) 5, (e) 6, (f) 7, (g), 8, (h) 9, (i) 10.

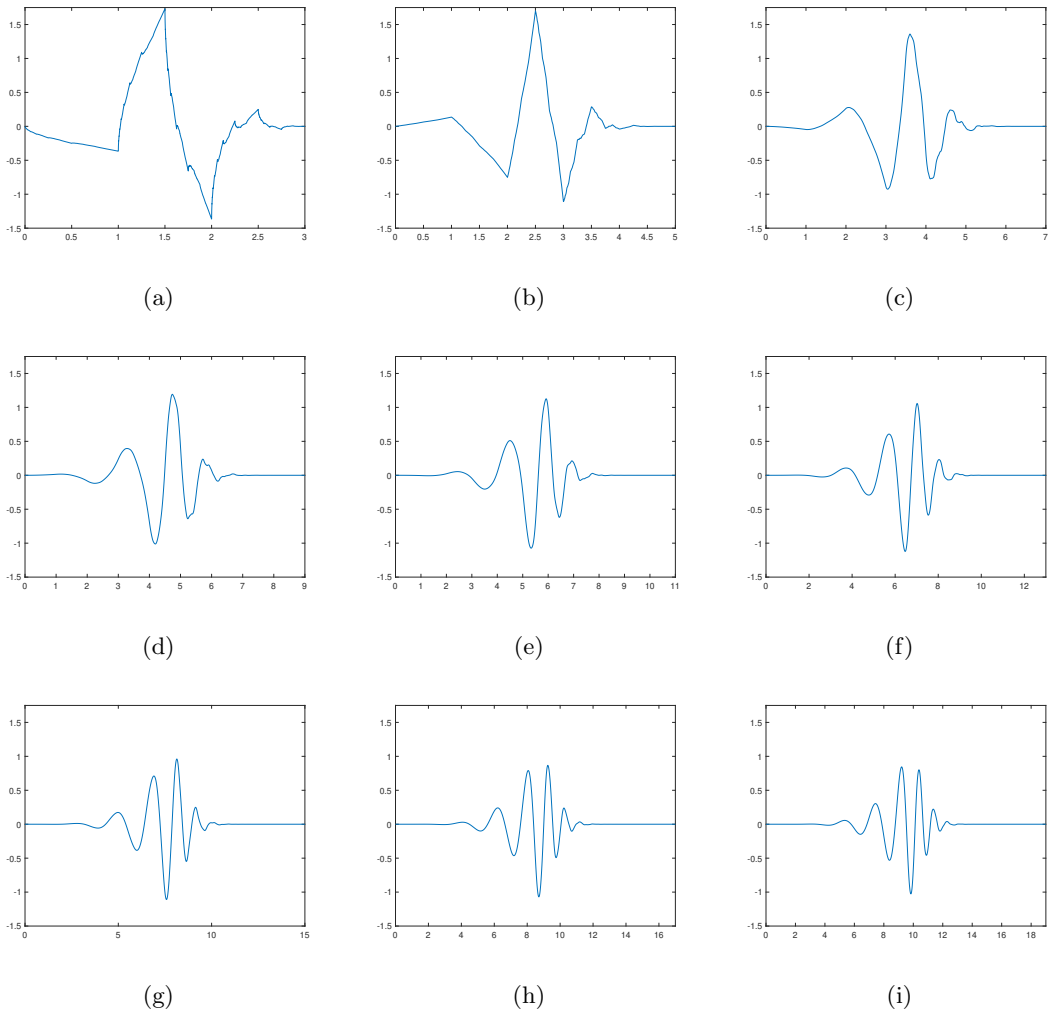


Figure 9: Plots of the wavelet functions generated by the cascade algorithm using the coefficients of Table 9. Each wavelet function has the following number of vanishing moments: (a) 2, (b) 3, (c) 4, (d) 5, (e) 6, (f) 7, (g), 8, (h) 9, (i) 10.

Appendix C Fourier transform and properties of the Dirac delta function

The Dirac delta function, denoted by $\delta(t)$, is formally defined as a distribution. However, in the context of this thesis we can view it as a function. The Dirac delta function can be seen as the derivative of the Heaviside step function. The Heaviside step function is defined as

$$H(t) = \begin{cases} 1, & t \geq 0 \\ 0, & t < 0 \end{cases}, \tag{83}$$

where $t \in \mathbb{R}$. So for the Dirac delta function, we have that

$$\lim_{t \to t_0} \delta\left(t - t_0\right) = \infty, \tag{84}$$

as well as that

$$\delta(t) = 0 \quad \forall t \neq 0. \tag{85}$$

If a function $f(t)$ is continuous at t_0, the Dirac delta function has as property that

$$\int_{t_0-\epsilon}^{t_0+\epsilon} f(t)\delta(t - t_0)\mathrm{d}t = f(t_0) \quad \forall \epsilon > 0, \tag{86}$$

another property of the Dirac delta function is that

$$\int_{-\infty}^{\infty} \delta(t)\mathrm{d}t = 1. \tag{87}$$

Its Fourier transform is given as

$$\mathcal{F}\{\delta(t - t_0)\} = \int_{-\infty}^{\infty} \mathrm{e}^{-2\pi i f t}\delta\left(t - t_0\right)\mathrm{d}t = \mathrm{e}^{-2\pi i f t_0}. \tag{88}$$

In general, shifting a function in time domain by $\pm c$ is equivalent to multiplying the frequency domain function with a factor $\mathrm{e}^{\pm 2\pi f i c}$. To proof this equivalency, let $c \in \mathbb{R}$ be arbitrary and let $x(t)$ be a function with the corresponding Fourier transform $X(f)$. Then

$$\mathcal{F}\{x(t + c)\} = \int_{-\infty}^{\infty} x(t + c)\mathrm{e}^{-2\pi i f t}\mathrm{d}t. \tag{89}$$

We perform a change of variables on $t' = t + c$ to get

$$\int_{-\infty}^{\infty} x(t')\mathrm{e}^{-2\pi i f(t'-c)}\mathrm{d}t' = \mathrm{e}^{2\pi f i c}\int_{-\infty}^{\infty} x(t')\mathrm{e}^{-2\pi i f t'}\mathrm{d}t' = X(f)\mathrm{e}^{2\pi f i c}. \tag{90}$$

Which proves our result. Therefore, when we integrate $\mathrm{e}^{2\pi i f(t-t_0)}$ over f, we get a time shifted Dirac delta function, i.e.

$$\int_{-\infty}^{\infty} \mathrm{e}^{2\pi i f(t-t_0)} \mathrm{d}f = \delta(t-t_0). \tag{91}$$

Appendix D Proof of preservation of energy

One important property of the Fourier transform and its inverse transform is that the transforms are energy preserving. Where energy of a time series $x(t)$ is defined as

$$E_x = \langle x(t), x(t) \rangle = \int_{-\infty}^{\infty} |x(t)|^2 \mathrm{d}t. \tag{92}$$

To prove this property of preservation of energy, Let $X(f)$ be the Fourier transform of $x(t)$ and $\overline{X(f)}$ it's complex conjugate. $X(f)$ and $\overline{X(f)}$ are given by

$$X(f) = \int_{-\infty}^{\infty} x(t) \mathrm{e}^{-2\pi i f t} \mathrm{d}t \quad \text{and} \quad \overline{X(f)} = \int_{-\infty}^{\infty} \overline{x(t')} \mathrm{e}^{2\pi i f t'} \mathrm{d}t' \tag{93}$$

respectively. The energy of the Fourier transform of $x(t)$ is then given as

$$\int_{-\infty}^{\infty} |X(f)|^2 df = \int_{-\infty}^{\infty} X(f) \overline{X(f)} \mathrm{d}f. \tag{94}$$

Filling in the Fourier transforms from Equation 93 results in

$$\int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} x(t) \mathrm{e}^{-2\pi i f t} \mathrm{d}t \int_{-\infty}^{\infty} \overline{x(t')} \mathrm{e}^{2\pi i f t'} \mathrm{d}t' \right] \mathrm{d}f. \tag{95}$$

Rewriting this expression yields

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x(t) \overline{x(t')} \mathrm{e}^{2\pi i f (t'-t)} \mathrm{d}f \mathrm{d}t \mathrm{d}t' = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x(t) \overline{x(t)} \int_{-\infty}^{\infty} \mathrm{e}^{2\pi i f (t'-t)} \mathrm{d}f \mathrm{d}t \mathrm{d}t'. \tag{96}$$

The inner integral over f is the inverse Fourier transform of a constant and the Fourier transform of a time shift. Resulting in a time shifted Dirac delta distribution, denoted by $\delta(t'-t)$. The Fourier transform of the Dirac delta distribution some of its properties are given in Appendix C. Filling in the Dirac delta distribution results in

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \delta(t'-t) x(t) \overline{x(t')} \mathrm{d}t \mathrm{d}t'. \tag{97}$$

Then by the properties of the Dirac delta distribution we rewrite the expression to

$$\int_{-\infty}^{\infty} x(t) \overline{x(t')} \mathrm{d}t = \int_{-\infty}^{\infty} |x(t)|^2 \mathrm{d}t. \tag{98}$$

Which completes our proof of preservation of energy.

Appendix E Inverse and determinant of a matrix with equicorrelated structure

First theorem Let a k by k matrix $\boldsymbol{C}$ be given, such that

$$\boldsymbol{C} = (a-b)\boldsymbol{I} + b\boldsymbol{J}, \tag{99}$$

where a and b are scalars, $\boldsymbol{I}$ is a k by k identity matrix and $\boldsymbol{J}$ is a k by k matrix of ones. Then, the inverse of matrix $\boldsymbol{C}$ exists if and only if $a \neq b$ and $a \neq -(k-1)b$ and is given by

$$\boldsymbol{C}^{-1} = \frac{1}{a-b}\left(\boldsymbol{I} - \frac{b}{a+(k-1)b}\boldsymbol{J}\right). \tag{100}$$

Proof: We will prove the theorem by showing that $\boldsymbol{C}\boldsymbol{C}^{-1} = \boldsymbol{I}$.

$$\begin{aligned}
\boldsymbol{C}\boldsymbol{C}^{-1} &= [(a-b)\boldsymbol{I} + b\boldsymbol{J}]\left[\frac{1}{a-b}\left(\boldsymbol{I} - \frac{b}{a+(k-1)b}\boldsymbol{J}\right)\right] \\
&= \frac{a-b}{a-b}\boldsymbol{I}\left(\boldsymbol{I} - \frac{b}{a+(k-1)b}\boldsymbol{J}\right) + \frac{b}{a-b}\boldsymbol{J}\left(\boldsymbol{I} - \frac{b}{a+(k-1)b}\boldsymbol{J}\right) \\
&= \boldsymbol{I} - \frac{b}{a+(k-1)b}\boldsymbol{J} + \frac{b}{a-b}\boldsymbol{J} - \frac{b^2}{(a+(k-1)b)(a-b)}\boldsymbol{J}\boldsymbol{J}.
\end{aligned} \tag{101}$$

Which we rewrite by noting that $\boldsymbol{J}\boldsymbol{J} = k\boldsymbol{J}$ to get

$$\begin{aligned}
\boldsymbol{C}\boldsymbol{C}^{-1} &= \boldsymbol{I} - \frac{b}{a+(k-1)b}\boldsymbol{J} + \frac{b}{a-b}\boldsymbol{J} - \frac{b^2k}{(a+(k-1)b)(a-b)}\boldsymbol{J} \\
&= \boldsymbol{I} + \boldsymbol{J}\left(\frac{-b}{a+(k-1)b} + \frac{b}{a-b} + \frac{-b^2k}{(a+(k-1)b)(a-b)}\right) \\
&= \boldsymbol{I} + \boldsymbol{J}\left(\frac{-b(a-b) + b(a+(k-1)b) - b^2k}{(a+(k-1)b)(a-b)}\right) \\
&= \boldsymbol{I} + \boldsymbol{J}\left(\frac{-ab + b^2 + ab + b^2(k-1)) - b^2k}{(a+(k-1)b)(a-b)}\right) \\
&= \boldsymbol{I} + \boldsymbol{J}\left(\frac{b^2(k-1)) - b^2k + b^2}{(a+(k-1)b)(a-b)}\right) = \boldsymbol{I} + \boldsymbol{J}\left(\frac{b^2(k-1)) - b^2(k-1)}{(a+(k-1)b)(a-b)}\right) \\
&= \boldsymbol{I},
\end{aligned} \tag{102}$$

which proves the first theorem.

Second theorem Let $\boldsymbol{C}$ be the same matrix as given in the first theorem of this appendix, the determinant of this matrix is then given by $|\boldsymbol{C}| = (a-b)^{k-1}(a+(k-1)b)$.

Proof This matrix can be written as

$$\boldsymbol{C} = \begin{bmatrix} a & b & b & b & \cdots & b \\ b & a & b & b & \cdots & b \\ b & b & a & b & \cdots & b \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ b & b & b & b & \cdots & a \end{bmatrix}. \tag{103}$$

Adding multiples of a row to another row causes the determinant to remain the same. We now define $\boldsymbol{C}'$ where we subtract the second row from the first row. Then subtract the third row from the second row and so on to where we subtract row k from row $k-1$. $\boldsymbol{C}'$ is then given as

$$\boldsymbol{C}' = \begin{bmatrix} a-b & b-a & 0 & 0 & \cdots & 0 \\ 0 & a-b & b-a & 0 & \cdots & 0 \\ 0 & 0 & a-b & b-a & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ b & b & b & b & \cdots & a \end{bmatrix}. \tag{104}$$

Since we have only subtracted rows from each other, we have that $|\boldsymbol{C}| = |\boldsymbol{C}'|$. Adding multiples of a column to another column also leaves the matrix determinant unchanged. We define $\boldsymbol{C}''$ where we take $\boldsymbol{C}'$ and add the first column to the second column. then add the second column to the third column and so on until we add column $k-1$ to column k resulting in

$$\boldsymbol{C}'' = \begin{bmatrix} a-b & 0 & 0 & 0 & \cdots & 0 \\ 0 & a-b & 0 & 0 & \cdots & 0 \\ 0 & 0 & a-b & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ b & 2b & 3b & 4b & \cdots & a+(k-1)b \end{bmatrix}. \tag{105}$$

We still have that $|\boldsymbol{C}| = |\boldsymbol{C}'| = |\boldsymbol{C}''|$. It can be noted that $\boldsymbol{C}''$ is a lower triangular matrix, which implies that $|\boldsymbol{C}''| = (a-b)^{k-1}(a+(k-1)b)$. Therefore, $|\boldsymbol{C}| = (a-b)^{k-1}(a+(k-1)b)$.

Appendix F Correlation of sines

Let

$$\begin{aligned}\mu_{1,t} &= \mu_0 + a\sin\left(\frac{2\pi}{T}t\right),\\ \mu_{2,t} &= \mu_0 + a\sin\left(\frac{2\pi}{T}(t+\theta)\right).\end{aligned} \tag{106}$$

Then

$$\text{Corr} = \frac{\int_0^T(\mu_{1,t}-\overline{\mu_{1,t}})(\mu_{2,t})-\overline{\mu_{2,t}}\mathrm{d}t}{\sqrt{\int_0^T(\mu_{1,t}-\overline{\mu_{1,t}})^2\mathrm{d}t}\sqrt{\int_0^T(\mu_{2,t}-\overline{\mu_{2,t}})^2\mathrm{d}t}}, \tag{107}$$

where $\overline{\mu_{1,t}}$ and $\overline{\mu_{2,t}}$ denote the mean of $\mu_{1,t}$ and $\mu_{2,t}$, respectively. Since the mean of a sine function is equal to zero, $\overline{\mu_{1,t}} = \overline{\mu_{2,t}} = \mu_0$. So

$$\text{Corr} = \frac{\int_0^T a^2\sin\left(\frac{2\pi}{T}t\right)\sin\left(\frac{2\pi}{T}(t+\theta)\right)\mathrm{d}t}{\sqrt{a^2\int_0^T\sin^2\left(\frac{2\pi}{T}t\right)\mathrm{d}t}\sqrt{a^2\int_0^T\sin^2\left(\frac{2\pi}{T}(t+\theta)\right)}} \tag{108}$$

We rewrite this expression using the trigonometric identity
$\sin\phi_1\sin\phi_2 = \frac{1}{2}\left(\cos\left(\phi_1-\phi_2\right)-\cos\left(\phi_1+\phi_2\right)\right)$ and that $\int\sin^2(x)\mathrm{d}x = \frac{x}{2}-\frac{1}{4}\sin(2x)$ to get

$$\begin{aligned}\text{Corr} &= \frac{\int_0^T a^2\frac{1}{2}\left(\cos\left(\frac{2\pi}{T}t-\frac{2\pi}{T}t-\frac{2\pi}{T}\theta\right)\cos\left(\frac{2\pi}{T}t+\frac{2\pi}{T}t+\frac{2\pi}{T}\theta\right)\right)\mathrm{d}t}{\sqrt{a^2\frac{T}{2}}\sqrt{a^2\frac{T}{2}}}\\ &= \frac{\frac{1}{2}a^2\int_0^T\left(\cos\left(-\frac{2\pi}{T}\theta\right)\cos\left(\frac{4\pi}{T}+\theta\right)\right)\mathrm{d}t}{a^2\frac{T}{2}}\\ &= \frac{\frac{1}{2}a^2\left(T\cos\left(-\frac{2\pi}{T}\theta\right)+0\right)}{a^2\frac{T}{2}} = \cos\left(-\frac{2\pi}{T}\theta\right) = \cos\left(\frac{2\pi}{T}\theta\right)\end{aligned} \tag{109}$$