



Article

# Multiscale Partial Correlation Clustering of Stock Market Returns

Antonis A. Michis

Central Bank of Cyprus, 1395 Nicosia, Cyprus; AntonisMichis@centralbank.cy

**Abstract:** This study proposes a wavelet procedure for estimating partial correlation coefficients between stock market returns over different time scales. The estimated partial correlations are subsequently used in a cluster analysis to identify, for each time scale, groups of stocks that exhibit distinct market movement characteristics and are therefore useful for portfolio diversification. The proposed procedure is demonstrated using all the major S&P 500 sector indices as well as precious metals and energy sector futures returns during the last decade. The results suggest cluster formations that vary by time scale, which entails different stock selection strategies for investors differing in terms of their investment horizon orientation.

**Keywords:** stock market returns; wavelets; partial correlation coefficient; cluster analysis; portfolio diversification

## 1. Introduction

The Pearson correlation coefficient is a frequently used statistic in finance, most notably in the area of portfolio analysis, in which it facilitates understanding of the linear associations between various stock market returns. For example, among other applications, it is frequently used for calculating portfolio variance in the well-known mean-variance portfolio theory framework (Cuthbertson and Nitzsche 2004, p. 121).

However, despite its popularity and ease of computation, Kenett et al. (2015) emphasize that the Pearson correlation coefficient is also characterized by an important limitation when used in the context of financial applications. A high correlation value for stock market returns does not necessarily imply a direct relationship between two stocks, but rather a common underlying influence by macroeconomic or psychological factors related to investors.

To properly understand the true direct correlation structure between stocks in such settings, it is important to first account for the common factors that influence stock movements and are most represented by the stock market index in the capital asset pricing model (CAPM) framework. The partial correlation coefficient enables such calculations by quantifying the correlation between two stock returns while conditioning for the effect of a third variable that influences both returns.

In this study, we propose a wavelet-based procedure for calculating the partial correlations between stocks using coefficients generated from maximal overlap discrete wavelet transforms (MODWT) of the stock market returns. Wavelets are classes of orthonormal basis functions (Nason 2008) that enable the decomposition of time series into different frequency bands (time scales). This is a very useful property in finance (In and Kim 2013) since it enables the examination of financial variables over different planning cycles. For example, take the case with investors who have different tolerances related to investment horizons (e.g., portfolio managers, brokers, traders, institutional investors). The need to estimate correlations between assets returns over different time scales was also noted by Wątorok et al. (2019), who proposed using the detrended cross-correlation coefficient (Kwapień et al. 2015) for measuring the degree of cross-correlation between two time series over different time scales.



**Citation:** Michis, Antonis A. 2022. Multiscale Partial Correlation Clustering of Stock Market Returns. *Journal of Risk and Financial Management* 15: 24. <https://doi.org/10.3390/jrfm15010024>

Academic Editor: Stephen Satchell

Received: 3 December 2021

Accepted: 6 January 2022

Published: 9 January 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

In the empirical applications of this study, the proposed method was used to estimate the partial correlations between the main S&P 500 sector indices in the US over different time scales while conditioning for movements in the returns of the general stock market index. The estimated partial correlation coefficients were subsequently used in a clustering procedure to better understand the similarities and differences between the different categories of stock market returns and therefore identify opportunities for portfolio diversification over different planning cycles.

Consequently, in the remaining sections the following research questions are addressed: (i) Can the partial correlation index be used to cluster stock market returns from different economic activity sectors and over different time scales? (ii) How do stock correlations between the different economic activity sectors vary over different time scales, once movements in the general stock market index have been accounted for? (iii) What are the implications for investors that differ in terms of their investment horizon orientation (short-term, medium-term, long-term)?

The rest of the article is organized as follows. Section 2 reviews the relevant literature concerning the application of correlation, clustering, and wavelet methods in finance. Section 3 introduces the data and the estimation procedure for the proposed wavelet-based partial correlation coefficient, which is subsequently used in the clustering applications of Section 4. Section 5 concludes and provides suggestions for further research.

## 2. Literature Review

The Pearson correlation coefficient is most commonly associated with two application areas in finance: portfolio diversification and financial contagion. In portfolio diversification applications, the principal consideration is the correct estimation of the correlation matrices between asset returns. However, in addition to the Pearson correlation coefficient, which is still frequently used in empirical studies (see, for example, [Geertsema and Lu 2020](#)), several other correlation methods have been proposed in the literature, such as time-varying correlation ([Chiang et al. 2007](#)), Fisher correlation ([Krishnan et al. 2009](#)), partial correlation ([Kenett et al. 2015](#)), dynamic conditional correlation ([Engle and Colacito 2006](#)), detrended cross-correlation ([Wątorrek et al. 2019](#)), and partial distance correlation ([Creamer and Lee 2019](#)).

Correlation coefficients have also been used in the financial contagion literature, where the principal aim is to measure cross-market linkages in financial markets following a shock to one or more countries. Two of the earlier studies in this field that used correlation coefficients include [King and Wadhvani \(1990\)](#) and [Lee and Kim \(1993\)](#). In a noteworthy study, [Forbes and Rigobo \(2002\)](#) showed that contagion tests based on the correlation coefficient are biased due to heteroskedasticity since cross-market correlations are conditional on market volatility. Under certain assumptions, the authors demonstrate the magnitude of this bias and explain the necessary adjustment to correct it.

More recently, [Alqaralleh and Canepa \(2021\)](#) proposed a wavelet-copula-GARCH procedure where changes in stock market correlations at higher frequencies are associated with contagion, and changes at lower frequencies are associated with normal interdependence. The authors found strong evidence of contagion between six major stock markets after the start of the COVID-19 pandemic, while before the pandemic the results support the long-run interdependence hypothesis.

Similarly, there is a considerable amount of literature concerning the application of cluster analysis methods in finance. [Bennett and Hugen \(2016\)](#) provide an informative introduction to the application of clustering methods for portfolio analysis with an emphasis on the k-means algorithm, which is also used in this study. [Tola et al. \(2008\)](#) used several clustering algorithms (such as the single and average linkage methods) to show that the reliability of portfolios in terms of the ratio between predicted and realized risk can be improved. In another interesting application, [Ahn et al. \(2009\)](#) used a correlation distance measure in the context of a minimum variance clustering algorithm to construct a set of

basis assets that characterize an investor's opportunity set and that are not susceptible to data-snooping biases.

Other notable applications of clustering methods in finance in recent years include the use of time-series hierarchical clustering for understating the correlation structure of an index and therefore building an optimal tracking portfolio (Focardi and Fabozzi 2004); the application of techniques from network science and linkage algorithms to understand the clustering of exchange-rate time series (Fenn et al. 2012); the incorporation of a distance measure based on variance ratio statistics in the complete-linkage algorithm for clustering international stock market returns (Bastos and Caiado 2014); and the application of a Gaussian mixture model on the voting shares of mutual funds to identify distinctive groups (and philosophies) of corporate governance in the US (Bubb and Catan 2021).

Due to their unique time scale decomposition properties, wavelets have been used in many economic and financial applications. Several studies concentrated on the estimation of the systematic risk and portfolio variance of assets across different time scales, such as the articles by Gençay et al. (2005), Kim and In (2010), and Michis (2014a, 2019). Another stream of research concentrated on the examination of macroeconomic relationships across different time scales, including money growth and inflation (Rua 2012), wage inflation and unemployment (Gallegati et al. 2011), productivity and unemployment (Gallegati et al. 2016), the liquidity effect (Michis 2015a), and the comovements exhibited by European economic sentiment indicators (Michis 2021).

In addition, wavelets have also been used in economic and financial forecasting (see the studies by Michis 2014b, 2015b; Michis and Guy 2017; Rua 2017; Caraianni 2017; Risse 2019), econometric estimation (Michis and Sapatinas 2007; Gençay and Gradojevi 2011; Gilles et al. 2009; Jensen 2000) and testing (Hong and Kao 2004; Fan and Gençay 2010; Gençay and Signori 2015), as well as in filtering-denoising applications (Fleming et al. 2000; Capobianco 2003; Bruzda 2014; Michis 2011).

Rua and Nunes (2009) analyzed international comovements between economic sector returns over different time scales. However, their analysis is based on simple correlations of monthly data in the frequency domain (wavelet coherency) that do not account for movements in the general stock market index, nor are any clustering applications involved. In contrast, Kenett et al. (2015) used partial correlation coefficients to estimate economic activity sector comovements in the US. Even though their authors use daily data, the analysis is not performed over different time scales, the economic activity sectors are not grouped into distinguishable clusters, and the long-run implications for investors are not considered.

Partial correlation coefficients and a cluster analysis were also used by Jung and Chang (2016), who analyzed data on the main stocks listed in the Korean stock exchange. Although interesting, their analysis is based on monthly data, the correlations are not evaluated over different time scales and there are no comparisons with the main US economic activity sectors. A time scale analysis of daily financial data was also suggested by Nava et al. (2018) using the empirical mode decomposition and Pearson's correlation coefficient. However, no cluster analysis was involved, nor any considerations of the comovements between the main US economic activity sectors.

This study provides a synthesis of the research streams by incorporating a partial correlation coefficient distance measure in the k-means clustering algorithm to identify distinct groups of stock market returns by time scale. Time scale decomposition is achieved through the application of a wavelet transform to the actual stock market returns and is important for investors considering improvements in the diversification of their portfolios over different planning cycles, e.g., a long-term buy-and-hold passive strategy vs. a short-term active trading strategy.

### 3. Materials and Methods

#### 3.1. Wavelet Transforms

In this section, we explain the estimation procedure for the wavelet-based partial correlation coefficient. First, we present the basic building blocks of wavelets, followed by the derivation of the MODWT coefficients, which are subsequently used for the estimation of the partial correlation coefficients between the different S&P 500 sector indices.

Wavelets constitute classes of orthonormal basis functions that enable simultaneous time-frequency decompositions of time series into different time scales. A two-dimensional family of wavelet functions can be generated by translating (a shift in the range) and dilating (an expansion in the range) the (real-valued) father and mother wavelets as follows:

$$\phi_{j,k} = 2^{-j/2} \phi\left(\frac{t - 2^j k}{2^j}\right) \text{ and } \psi_{j,k} = 2^{-j/2} \psi\left(\frac{t - 2^j k}{2^j}\right).$$

Lower values of the index  $j$  capture the high frequency, oscillating features of a signal (high resolution), while higher values capture the low frequency (low resolution), smooth components (In and Kim 2013, p. 11).

Wavelets must satisfy the following conditions (Percival and Walden 2000, p. 2):

$$\int_{-\infty}^{\infty} \psi(t) dt = 0 \text{ and } \int_{-\infty}^{\infty} |\psi(t)|^2 dt = 1.$$

While the second condition implies movements of  $\psi(t)$  away from zero, the first condition ensures that these movements cancel out (e.g., movements above zero are cancelled out by movements below zero), thus leading to the formation of small waves. In addition, the following admissibility condition enables the reconstruction of a function from its continuous wavelet transform, where  $\Psi(f)$  is the Fourier transform of the wavelet function:

$$C_\psi = \int_0^\infty \frac{|\Psi(f)|^2}{f} df \text{ and } 0 < C_\psi < \infty.$$

A finite ( $j = 1, 2, \dots, J$ ) wavelet series approximation of a dyadic length ( $2^j$ ) time series  $x_t$  can be represented as follows:

$$x_t \approx \sum_k s_{J,k} \phi_{J,k}(t) + \sum_k d_{J,k} \psi_{J,k}(t) + \sum_k d_{J-1,k} \psi_{J-1,k}(t) + \dots + \sum_k d_{j-1,k} \psi_{j-1,k}(t) + \dots + \sum_k d_{1,k} \psi_{1,k}(t)$$

with wavelet transform coefficients  $s_{J,k} = \int \phi_{J,k}(t) x_t dt$  and  $d_{j,k} = \int \psi_{j,k}(t) x_t dt$  or more compactly in terms of wavelet details ( $D_j = \sum_k d_{j,k} \psi_{j,k}(t)$ ) and smooths ( $S_j = \sum_k s_{j,k} \phi_{j,k}(t)$ ) as:

$$x_t \approx S_J + D_J + D_{J-1} + \dots + D_j + \dots + D_1.$$

The scaling coefficients ( $s_{j,k}$ ) corresponding to father wavelets capture the (low frequency) long-run trend in the series, while the detail coefficients ( $d_{j,k}$ ) corresponding to mother wavelets capture the high frequency, oscillating characteristics of the series.

Consequently, the wavelet series approximation provides a decomposition of  $x_t$  into  $J = 2^j$  time scale components (a multiresolution analysis). Adding progressively smaller time scale components ( $D_j$ ) to the long-run trend ( $S_j$ ) increases the finer detail characteristics of the signal until the original time series  $x_t$  is reconstructed. For a weekly time series of length 512 ( $j = 9$ ), the coefficients at time scale  $D_1$  capture oscillations over cycles of 2–4 weeks. Accordingly, the coefficients at scales  $D_2$  and  $D_3$  capture oscillations over cycles of 4–8 and 8–16 weeks, respectively, and this is the case up to scale 9, which is associated with long-term cyclical variations in  $x_t$ .

Two wavelet transforms frequently used for the analysis of time series with finite length are the discrete wavelet transform (DWT) and the MODWT. In this study, we will analyze stock market data using the MODWT, which is characterized by the following desirable properties (In and Kim 2013, p. 25): it can be used with any sample size (not only dyadic length series), it is shift invariant (circularly shifting the series does not influence the multiresolution analysis), and its wavelet details and smooths are associated with zero-phase filtering (e.g., extreme values in the actual data can be aligned with corresponding values in the multiresolution analysis).

There are two additional characteristics of the MODWT that are particularly relevant for estimating partial correlation coefficients: (i) it provides an asymptotically more efficient variance estimator than the DWT, and (ii) by considering all integer translations of the series at each frequency resolution, it generates coefficient vectors that have the same length as the original time series, which facilitates computations. In applied work, the MODWT is computed using a pyramid algorithm with discrete high-pass and  $(h_0, \dots, h_{L-1})$  low-pass filters  $(g_0, \dots, g_{L-1})$ .

The first level of the algorithm filters the actual data with the rescaled filters  $\tilde{h}_j = h_j/2^j$  and  $\tilde{g}_j = g_j/2^j$ :

$$\tilde{d}_{1,t} = \sum_{l=0}^{L-1} \tilde{h} x_{t-l \bmod N} \text{ and } \tilde{v}_{1,t} = \sum_{l=0}^{L-1} \tilde{g} x_{t-l \bmod N}.$$

The second level proceeds by applying the same filtering procedure to the vector of scaling coefficients produced in level 1 ( $\tilde{v}_1$ ), which gives rise to the second level of wavelet ( $\tilde{d}_2$ ) and scaling coefficient vectors ( $\tilde{v}_2$ ). Level 3 proceeds in the same way by applying the filtering procedure on the scaling vector  $\tilde{v}_2$ . The same filtering procedure is repeated for  $J$  levels, each time using as input the scaling vector of the previous level, which gives rise to the matrix of wavelet coefficient vectors  $\tilde{d} = [\tilde{d}_1, \dots, \tilde{d}_J, \tilde{v}_J]'$  that will be used for the calculation of the wavelet-based partial correlation coefficients in the next section.

### 3.2. Wavelet Based Partial Correlation Coefficients

To estimate the partial correlation coefficient between two time series  $x_t$  and  $y_t$  while conditioning on a third time series  $z_t$ , it is necessary to first estimate the variances and covariances associated with the three variables. An unbiased estimator of the wavelet variance for  $x_t$  at time scale  $\lambda_j$ , using the MODWT coefficients, is given by:

$$\sigma_x^2(\lambda_j) = \frac{1}{\tilde{T}_j} \sum_{t=L_j-1}^{T-1} (\tilde{d}_{j,t}^x)^2$$

where  $L_j = (2^j - 1)(L + 1) + 1$  and  $\tilde{T} = T - L_j + 1$  refer to the length of the wavelet filter and the number of coefficients unaffected by the boundary, respectively (Gençay et al., p. 241). Accordingly, the wavelet covariance between time series  $x_t$  and  $y_t$  can be expressed as:

$$\sigma_{x,y}(\lambda_j) = \frac{1}{\tilde{T}_j} \sum_{t=L_j-1}^{T-1} (\tilde{d}_{j,t}^x)(\tilde{d}_{j,t}^y).$$

It is then straightforward to obtain an expression for the wavelet correlation at scale  $\lambda_j$  through normalization of the wavelet covariance with the individual variances:

$$\rho_{x,y}(\lambda_j) \equiv \frac{\sigma_{x,y}(\lambda_j)}{\sigma_x(\lambda_j) \sigma_y(\lambda_j)}.$$

The wavelet correlation coefficient bares the well-known property:  $|\hat{\rho}_{x,y}(\lambda_j)| < 1$ . In and Kim (2013, pp. 33–34) provided formulas for the computation of confidence intervals for both the wavelet covariance and wavelet correlation estimators presented above.

When working with stock market returns, a high correlation value is not always indicative of a direct association between two stocks. Kenett et al. (2015) emphasize that movements in both stocks can be associated with movements in a third variable that influences both stocks, as is, for example, the case with the general stock market index. In this case, it is necessary to use the partial correlation coefficient, which calculates the correlation between two variables ( $x_t$  and  $y_t$ ), while controlling for the effect of a third variable ( $z_t$ ).

In line with the partial correlation formula for continuous observations (Chen and Popovich 2002, p. 73), the wavelet partial correlation coefficient can be calculated as follows, where  $|\hat{\rho}_{x,y,z}(\lambda_j)| < 1$ :

$$\rho_{x,y,z}(\lambda_j) \equiv \frac{\rho_{x,y}(\lambda_j) - \rho_{x,z}(\lambda_j)\rho_{y,z}(\lambda_j)}{\sqrt{[1 - \rho_{x,z}^2(\lambda_j)][1 - \rho_{y,z}^2(\lambda_j)]}} \tag{1}$$

### 3.3. Data: S&P 500 Sectors

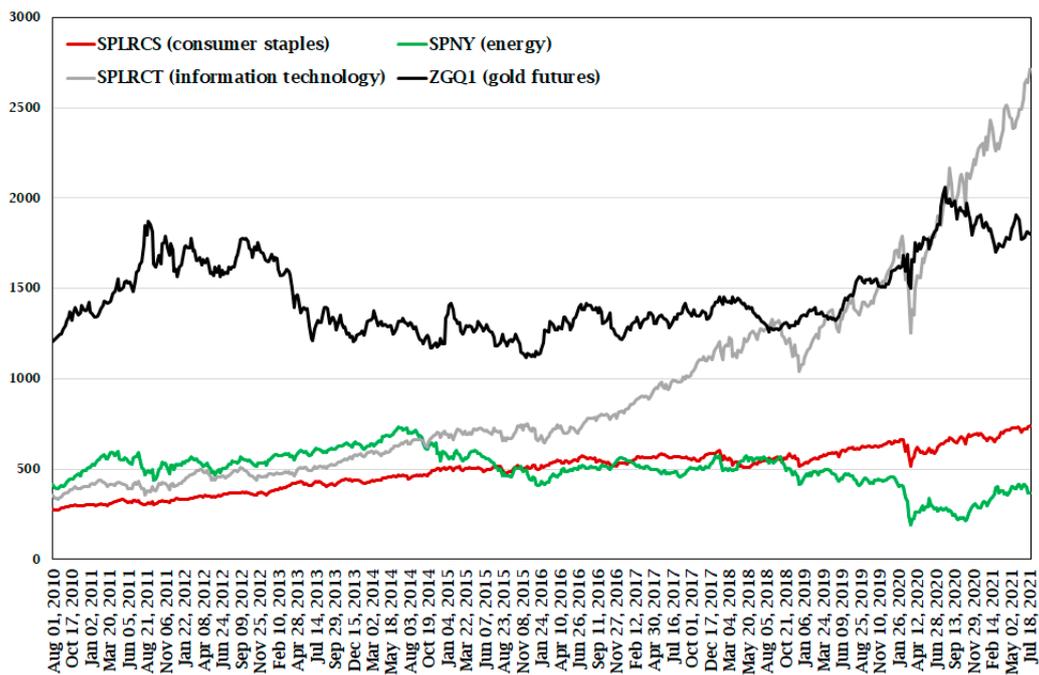
The wavelet partial correlation coefficient in Equation (1) was used to estimate the partial correlation matrices between the following S&P 500 sector returns on a scale-by-scale basis: telecom services (SPLRCL), consumer discretionary (SPLRCD), consumer staples (SPLRCS), energy (SPNY), financials (SPSY), health care (SPXHC), industrials (SPLRCI), information technology (SPLRCT), materials (SPLRCM), real estate (SPLRCREC), and utilities (SPLRCU).

The analysis was also augmented with the following precious metals and energy futures returns that likely exhibit distinct market movement characteristics and are therefore useful for portfolio diversification purposes: gold, silver, copper, Brent oil, and natural gas. For all indices, the data were downloaded from the financial platform [www.investing.com](http://www.investing.com) (accessed on 9 September 2021) and concern weekly closing values for the period 1 August 2010–18 July 2021, in the US. Table 1 includes summary statistics for all the indices, and Figure 1 illustrates the time development of four indices that are representative of the cyclical variations inherent in the data.

**Table 1.** Summary statistics of weekly closing values.

Index	Mean	St. Dev.	Min	Max
S&P 500	2231.031	768.763	1064.590	4411.800
Telecom services	159.849	27.740	111.340	275.960
Consumer discretionary	668.039	295.562	241.460	1476.620
Consumer staples	496.603	115.494	275.320	737.090
Energy	502.264	100.888	193.930	734.160
Financials	343.244	110.181	157.480	637.750
Health care	806.849	295.128	330.080	1531.580
Industrials	499.735	149.601	248.360	890.440
Informat. technology	946.028	560.319	334.000	2713.400
Materials	310.808	72.275	184.550	551.920
Real estate	183.537	37.984	106.450	289.260
Utilities	238.648	52.496	153.810	356.330
Gold futures	1454.874	209.623	1114.500	2061.400
Silver futures	21.922	6.650	12.523	48.584
Copper futures	31.226	5.982	19.940	47.485
Brent Oil futures	76.140	26.998	21.440	126.650
Natural gas futures	3.106	0.781	1.495	6.135

Note: 573 observations sample size per index.



**Figure 1.** Consumer staples, energy, information technology and gold futures.

For example, the information technology index exhibits a rising exponential trend, while the energy sector index has a slightly declining trend after 2014, a further drop in 2020, and a small rebound in 2021. In contrast, consumer staples are mostly associated with a slightly increasing trend and a small increase in volatility after 2018. Furthermore, even though all indices exhibit short-term fluctuations around their long-term trends, it can be observed that the gold futures index also exhibits several medium-term cycles, most notably during the periods August 2010–May 2013, December 2015–December 2016, and January 2020–February 2021.

Wavelets are particularly well suited for decomposing the cyclical characteristics of nonstationary time series (Nason 2008, p. 60), such as those frequently encountered in stock market data. They are also very effective in handling time series with extreme values, such as spikes and shifts in trends (Michis 2015b), which, in the case of Figure 1, are most evident in the gold futures and energy sector values of 2 August 2020 and 15 March 2020, respectively.

For all pairs of stock and futures returns, the wavelet coefficients necessary for estimating the partial correlation coefficient in Equation (1) were generated using the MODWT with a Daubechies least asymmetric wavelet filter of length 8. This is a frequently used filter for financial time series (see, for example, Michis 2014a and Kim and In 2010), which also provided good resolutions of the financial returns analyzed in this study.

The partial correlation matrices for time scales 3, 6, and 8 are presented in Appendix A. We concentrate on the results for these three time scales mainly because they are characteristic of the investment horizons typically encountered in financial markets (short-term, medium-term and long-term), and they are also representative of the correlation matrices estimated for the remaining time scales in the analysis. In several cases, the partial correlations between two financial returns are negative or close to zero, which is consistent with the findings of Kenett et al. (2015).

As is common for correlation estimates (Obilor and Amadi 2018), Appendix B includes  $p$ -values for the estimated partial correlation coefficients of time scales 3, 6 and 8 that were generated using the testing procedure proposed by Kim (2015) based on the Student's  $t$  distribution. Only 14% of the estimated coefficients were not found to be statistically significant at the 90% confidence level.

Jung and Chang (2016) note that due to conditioning on general market index returns, partial correlation coefficients tend to be smaller than Pearson correlation coefficients of the same stock returns. In addition, not accounting for common factors that have an influence on stock returns can result in biased Pearson correlation estimates, which are not useful for investors seeking to diversify risk away from their portfolios. Furthermore, when the common factor in partial correlation estimates is chosen to be the general market index, the partial correlation coefficient resembles the CAPM.

#### 4. Results: Clustering of Stock Market Returns

Jung and Chang (2016) were among the first to propose the use of partial correlation coefficients for clustering stock returns using an agglomerative hierarchical clustering procedure. In this section, we extend these clustering applications by proposing the use of a wavelet-based partial correlation dissimilarity measure in the context of the k-means clustering algorithm to identify groups of stocks that exhibit distinct market movement characteristics.

In addition to being useful for portfolio diversification purposes (Cuthbertson and Nitzsche 2004), such an analysis enables the exploration of differences between stocks on a time scale basis, such as variations over short-term, medium-term, and long-term cycles, and it is therefore particularly well suited for investors that differ in terms of their investment horizon orientations.

Can the partial correlation coefficient be used to cluster stock market returns from different economic activity sectors and over different time scales?

For the clustering applications of this section, we will use the following partial correlation dissimilarity measure between stocks  $x_t$  and  $y_t$ , which is based on the wavelet partial correlation coefficient formula presented in Section 3:

$$\delta_{x,y;z}(\lambda_j) = [1 - \rho_{x,y;z}(\lambda_j)] / 2 \tag{2}$$

This measure resembles the Pearson correlation dissimilarity measure that is frequently used for clustering continuous data observations (Everitt et al. 2011, p. 50). However, it differs in two important aspects: (i) it is based on the partial correlation coefficient that measures the association between two stocks ( $x_t$  and  $y_t$ ), while controlling for the effect of a third variable (the stock market index), and (ii) it can be calculated and used in clustering applications separately for each time scale.

Consequently, using the dissimilarity measure  $\delta_{x,y;z}(\lambda_j)$ , the k-means clustering algorithm can be applied separately for each time scale. To see this, first consider a  $p$ -dimensional dataset consisting of the level  $j$  wavelet coefficient vectors generated from the returns series of  $p$  separate stocks, as described in Section 3.1. If there are  $K$  separate clusters inherent in the data, total dispersion is provided by the following matrix:

$$T_j = \sum_{g=1}^K \sum_{i=1}^{T_g} (\mathbf{d}_{g:i,j} - \bar{\mathbf{d}}_j)(\mathbf{d}_{g:i,j} - \bar{\mathbf{d}}_j)'$$

where  $\mathbf{d}_{g:i,j}$  refers to the  $i$ th item in cluster  $g$  and  $\bar{\mathbf{d}}_j = \frac{1}{N} \sum_{g=1}^K \sum_{i=1}^{T_g} \mathbf{d}_{g:i,j}$  is the overall mean. Total dispersion can then be decomposed into within-cluster dispersion  $W$  and between-cluster dispersion  $B$ ,

$$T_j = W_j + B_j$$

where  $W_j = \sum_{g=1}^K \sum_{i=1}^{T_g} (\mathbf{d}_{g:i,j} - \bar{\mathbf{d}}_{g:j})(\mathbf{d}_{g:i,j} - \bar{\mathbf{d}}_{g:j})'$ ,  $B_j = \sum_{g=1}^K \sum_{i=1}^{T_g} (\mathbf{d}_{g:j} - \bar{\mathbf{d}}_j)(\mathbf{d}_{g:j} - \bar{\mathbf{d}}_j)'$  and

$\bar{\mathbf{d}}_{g:j} = \frac{1}{T_g} \sum_{i=1}^{T_g} \mathbf{d}_{g:i,j}$  is the cluster  $g$  mean value.

The k-means algorithm used in this study proceeds to group items in a way that minimizes  $W_j$  and therefore most of the variation in the data are accounted for by between cluster variation  $B_j$ . Consequently, it attempts to minimize the following objective function:

$$W_j = \sum_{g=1}^K \sum_{i=1}^{T_g} u_{gi;j} \delta_{gi;j} \tag{3}$$

where  $u_{gi;j}$  is the cluster membership indicator ( $u_{gi;j} = 1$  when item  $i$  is assigned to cluster  $g$ ) and  $\delta_{gi;j}$  is the distance between item  $i$  and cluster  $g$  at time scale  $j$ . The classic k-means algorithm is based on Euclidean distances as used in the total dispersion decomposition above. However, when working with time series data, the Pearson correlation distance  $[1 - \rho_{Pearson}]$  is frequently preferred. Berthold and Höppner (2016) showed that for normalized series with zero mean and unit variance, the Euclidean distance is equal to  $2T$  times the correlation distance:  $\delta_{Euclidian} = 2T \delta_{Pearson}$ .

In the empirical applications of this study, we will minimize the objective function (3) using the partial correlation dissimilarity measure presented in Equation (2), which is more suitable for analyzing stock returns. The k-means algorithm in this case proceeds in four steps:

1. Define an initial partition of the data into  $g$  groups (randomly or using, e.g., a hierarchical method with a dendrogram).
2. For this initial partition, compute the score of  $|W_j|$ .
3. Then, reallocate each item  $d_{gi;j}$  separately into other groups until a partition is found that reduces the criterion  $|W_j|$  the most.
4. Repeat the procedure from step 2 until the score of  $|W_j|$  cannot be reduced further.

Even though this procedure is successful at finding a local minimum, several repetitions with different initial partitions can lead to solutions that are equal or close to the global minimum (Cox 2005, p. 93).

In applied work, a frequently used method for choosing the number of clusters suitable for partitioning a dataset first proceeds by calculating and plotting the within-group sum of squares generated by each separate k-means group solution. Increasing the number of clusters leads to a reduction in the sum of squares. Therefore, a noticeable “elbow” in the sum of squares plot provides an indication for the most suitable number of clusters to form (Everitt and Hothorn 2011, p.180). Using the partial correlation matrices of time scales 3, 6, and 8 in Appendix A, Figure 2 plots the within-group sum of squares for seven different cluster numbers, generated using the dissimilarity measure (2) in the context of the k-means algorithm. The results suggest the following partitions: five clusters for time scale 3, four clusters for time scale 6, and three clusters for time scale 8.

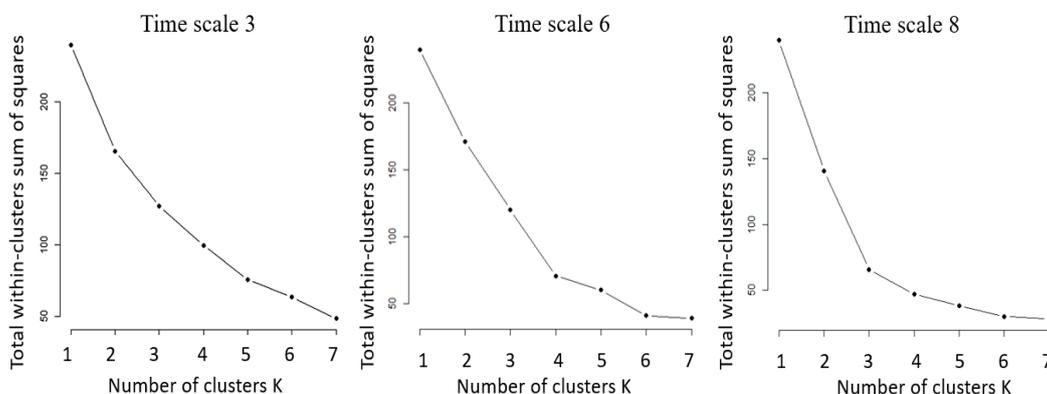


Figure 2. Total within-cluster sums of squares for time scales 3, 6 and 8.

To evaluate the consistency of the resulting cluster formations for time scales 3, 6, and 8, we performed a silhouette analysis, which is a frequently used method for validating cluster formations. For each object, an index is defined that compares the object’s separation from its cluster against the heterogeneity of the cluster (Everitt et al. 2011). Consequently, it provides an indication of how similar an observation is with the cluster to which it has been assigned relative to other clusters.

Values of the index close to 1 suggest that the object is well classified since the cluster’s heterogeneity is smaller than its separation, while values close to  $-1$  suggest misclassification. Borderline cases where the object is not clearly assigned between two neighbouring clusters are indicated with values close to zero.

In silhouette plots, the values of the index for each cluster are displayed with horizontal bars, as shown in Figure 3 for the three clusters of time scale 8. Kaufman and Rousseeuw (1990), who first proposed the silhouette method, consider silhouette width values higher than 0.5 to be indicative of a good cluster formation, as is the case with the results of time scale 8. However, more useful in applied work is the average silhouette width curve for different cluster sizes, since it enables identification of the optimal group sizes that maximize silhouette width.

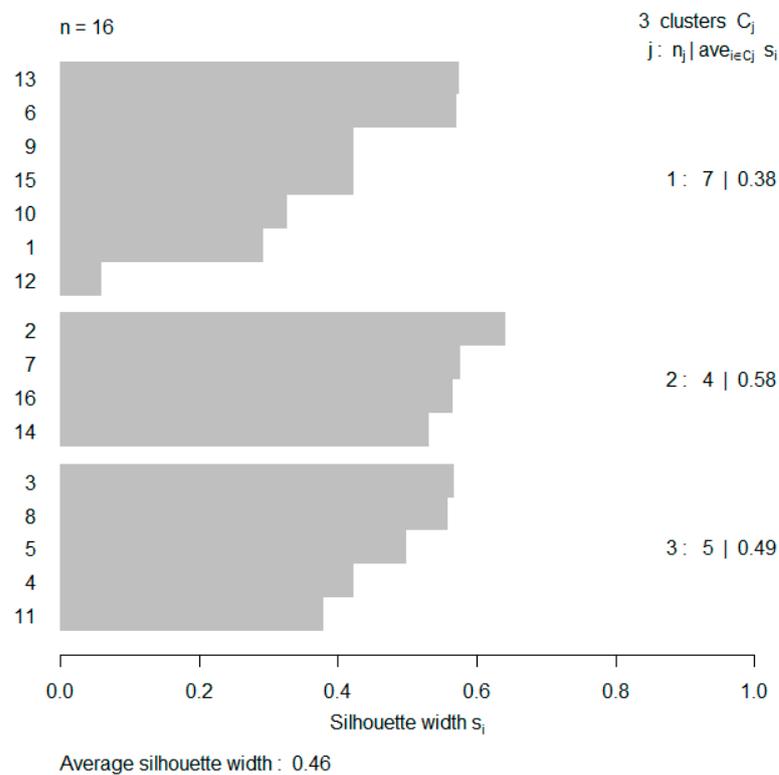


Figure 3. Silhouette analysis for time scale 8.

Appendix C depicts the silhouette width curves for the clustering solutions of time scales 3, 6, and 8. In each case, the cluster formations were generated using a random starting point of two clusters, followed by a 100 iteration convergence limit. The maximum point of the average silhouette width curve for timescale 3 suggests that the optimal cluster size in this case is five. For time scales 6 and 8, the respective optimal cluster sizes are four and three respectively.

How do stock correlations between the different economic activity sectors vary over different time scales once movements in the general stock market index have been accounted for?

The results of the k-means clustering procedure using the abovementioned cluster partitions are presented in Figures 4–6. All data were analyzed using the wavelsim, cluster,

factoextra, ppcor, and ggplot2 packages for R available from the CRAN archive. The horizontal and vertical axes correspond to the first two principal components that account for most of the variability in the partial correlation distance matrix, as represented by the respective percentage values. This dimension reduction method is commonly used in applied work for graphically illustrating clustering results in the space of the first two principal components (Everitt and Hothorn 2011). For the purposes of this study, we used the visualisation procedure suggested by Kassambara and Mundt (2020).

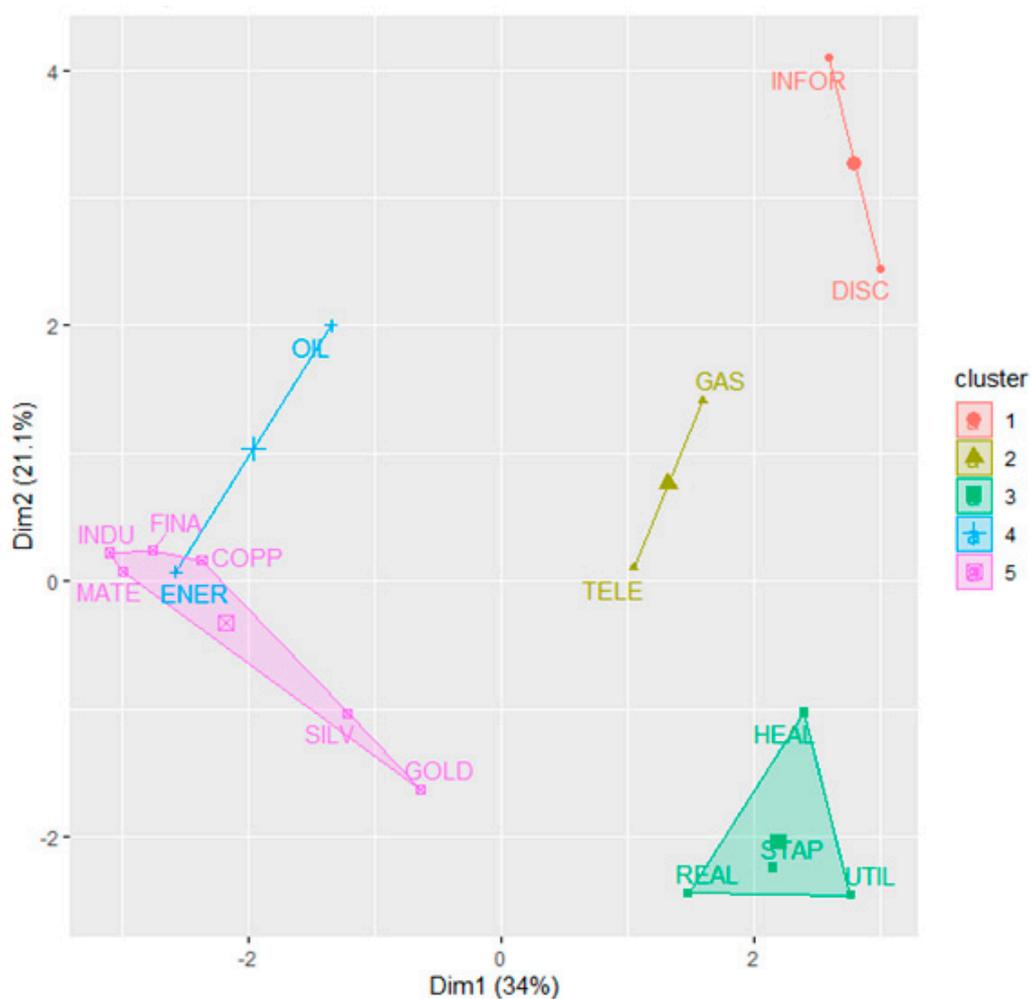


Figure 4. Cluster analysis results for time scale 3.

It can also be observed that the precious metal (gold, silver, and copper) futures returns are always grouped together, while some indices are always grouped separately from precious metals, e.g., “Consumer discretionary” and “Consumer staples”. In contrast, the returns associated with the energy futures (Brent oil and natural gas) are grouped together only when considering movements over long-term cycles (time scale 8). The “Energy” and “Financial” indices are grouped together when considering medium-term (time scale 6) and long-term cycles, and the “Real estate” and “Utilities” indices are grouped separately only when considering long-term cycles.

Furthermore, the abovementioned cluster analysis results provide some interesting insights. First, the number of cluster formations necessary for grouping the indices is reduced as one moves progressively to higher time scales that represent longer duration cycles in the data. Consequently, in the long term, the cluster formations are larger, consist of more homogeneous stock returns, and are also more distinct, as represented by the distances

between them in the configuration for time scale 8 in Figure 6. In such environments, frequent trading and portfolio changes become less important.

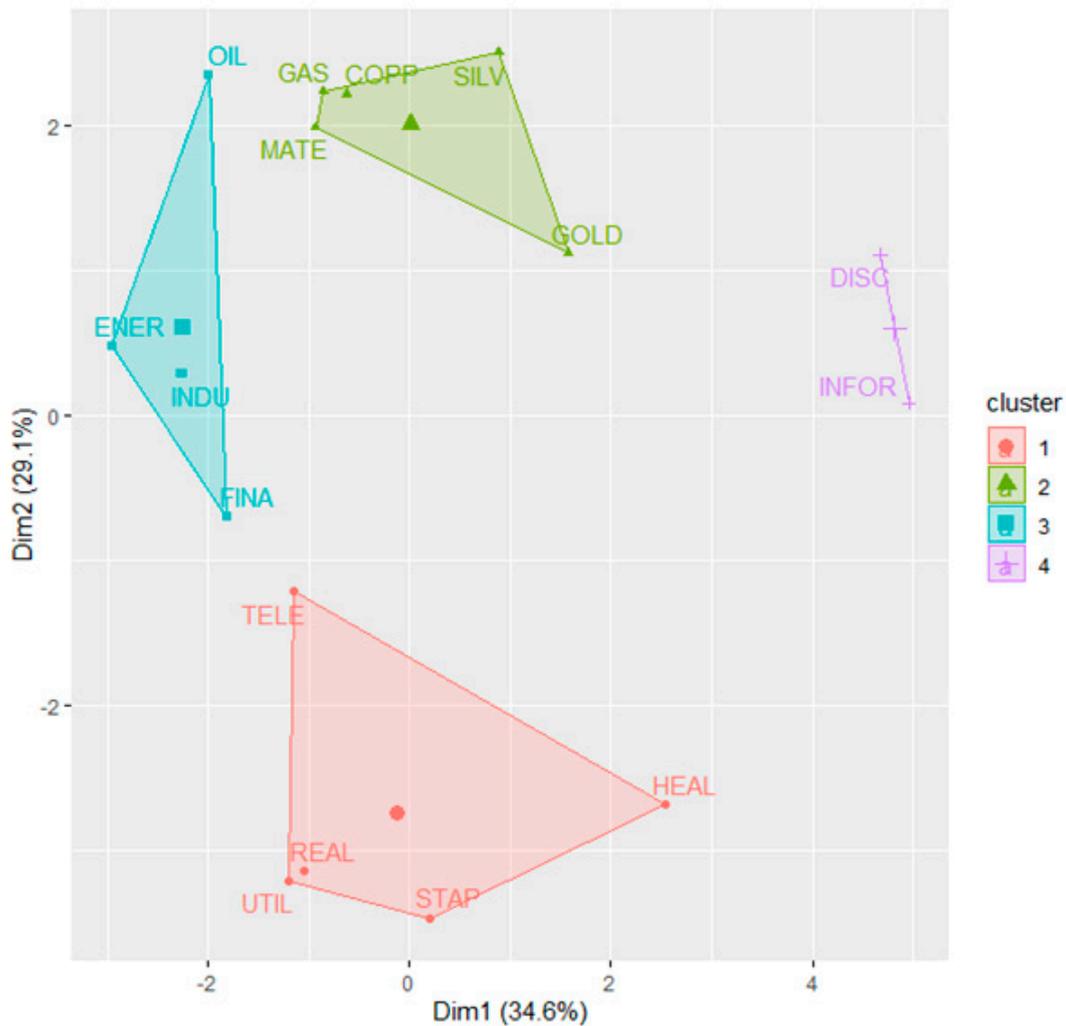


Figure 5. Cluster analysis results for time scale 6.

What are the implications for investors that differ in terms of their investment horizon orientation (short-term, medium-term, long-term)?

Second, cluster participation varies by time scale. For example, “Information technology” and “Materials” belong to Cluster 1 at time scale 8, but to different clusters at time scales 3 and 6. In this case, a different portfolio selection strategy is required for investors adopting a short-term investment horizon strategy (since the two sectors can add diversification benefits in the short term, as shown by their negative partial correlation coefficient) compared to investors adopting a long-term investment horizon strategy.

Third, the partial correlation indices and the degree of homogeneity between the sector indices in each cluster also vary by time scale. This finding suggests that the traditional approach to determining the similarity of stocks based on their official industrial classification and irrespective of the investment horizon under consideration might not be the most appropriate way to consider adjustments in investment portfolios. Our results suggest that the similarities between sector indices differ by time scale. Therefore, a data-driven approach to identifying distinct market movements between sector indices (and stocks) on a scale-by-scale basis could be more beneficial for diversifying risk away from a portfolio.

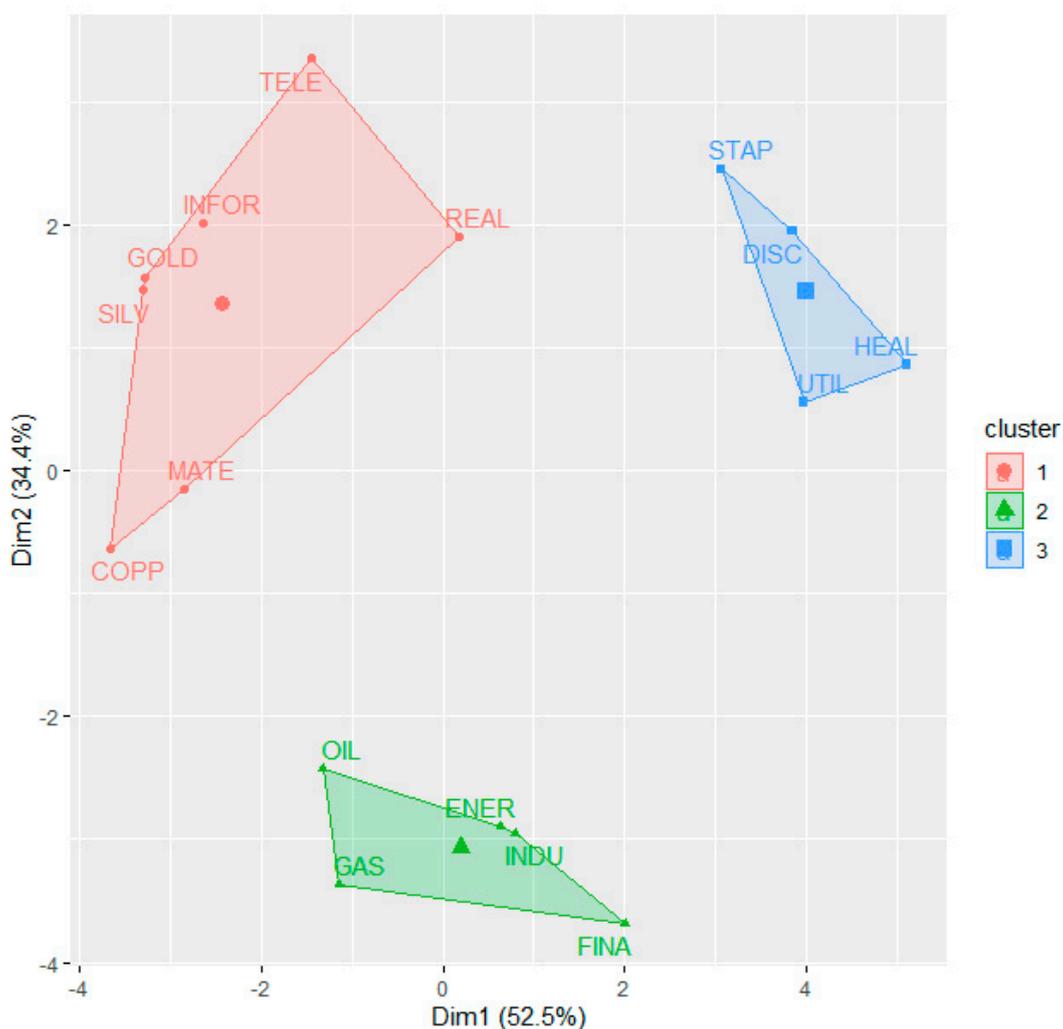


Figure 6. Cluster analysis results for time scale 8.

Wavelets can also be used for analyzing shorter time scales, using daily observations and concentrating on smaller periods of increased volatility in financial markets, as is the case with the recent COVID-19 pandemic. Consequently, for all the indices considered in our study, we collected daily observations (closing values) for the period 1 May 2019–8 July 2021 and performed the cluster analysis described in Section 4. Appendix D includes figures of the resulting cluster formations for time scales 3, 4, and 5 that cover cyclical movements of length 8–16, 16–32, and 32–64 days respectively.

The results for time scale 3 suggest that the precious metal (gold, silver, and copper) and energy futures (oil and gas) returns form two distinct clusters, which are also closely located to each other. In the case of energy futures, a possible explanation concerns the unique market movements of energy fuels during the COVID-19 pandemic where an initial drop in prices was followed by sharp increases due to limitations on production and increased demand (Camp et al. 2020). Precious metals on the other hand are usually considered as safe haven assets that many investors prefer during times of financial distress (Michis 2019). In a recent study, Sifat et al. (2021) found evidence that speculations in the energy and precious metals futures markets have in fact increased during the COVID-19 pandemic, which could also be related with the distinct behavior of these indices in the cluster analysis results for time scale 3.

This conclusion (speculative behavior confined to the short-term cycles of time scale 3) is further supported by the clustering results for time scales 4 and 5 in Appendix D that concern longer cycles. It can be observed that the precious metal and energy futures returns

become less distinct and progressively move closer to some other sectors when evaluated over these longer time scales. In addition, the cluster formations in these time scales are more similar to the clustering results of Figure 4 that were generated using weekly data. For example, the “Real estate”, “Consumer staples”, “Health care”, and “Utilities” sectors are grouped together in these time scales, as is the case for the “Materials”, “Industrials”, “Energy”, and “Financials” sectors. The same is also true for the “Information technology” and “Consumer discretionary” sectors.

It is also worth mentioning that, in the daily results for time scale 5, the “Gas futures” and “Telecom services” indices are grouped together, as is the case with the weekly data results for time scale 3 in Figure 4. This unique behavior of the “Telecom services” sector may be explained by the high increase in demand for these services after the beginning of the COVID-19 pandemic, which is also reflected in the general trend of the index in Figure 7. However, it is important to emphasize at this stage that a more in-depth analysis of the sectors that exhibit distinct market behavior during periods of increased volatility should be combined with an analysis of individual stock returns, since heterogeneity could exist within the aggregated sector movements. The analysis by Jung and Chang (2016) provides several useful suggestions in this direction.

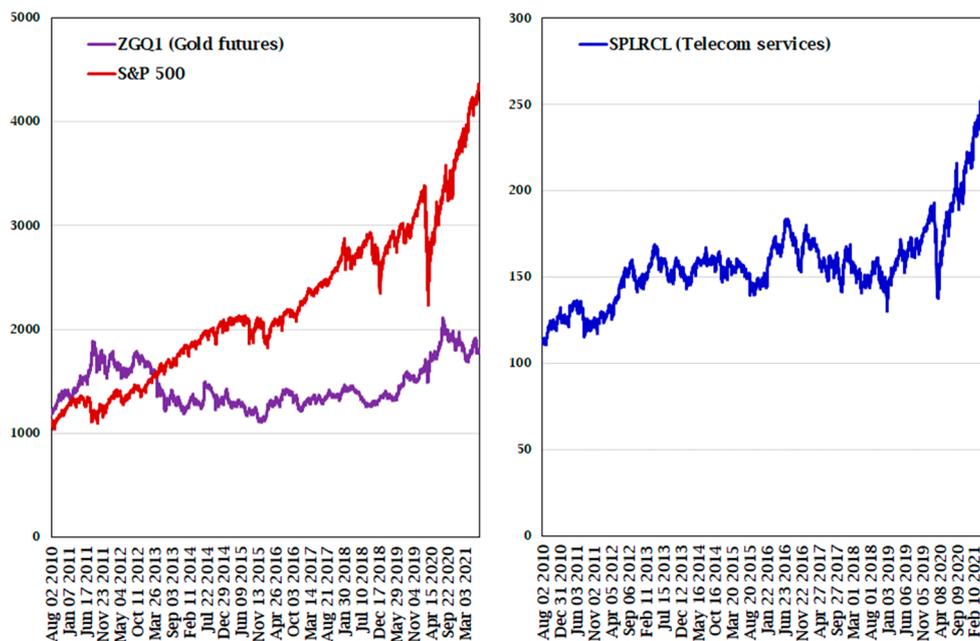


Figure 7. S&P 500, gold futures and telecom services indices (daily obs.).

## 5. Conclusions

This study proposed a data-driven procedure for analyzing associations between stock market returns using a wavelet-based estimator of the partial correlation coefficient. A key feature of this coefficient is the ability to estimate the associations between stock market returns over different time scales (short-term, medium-term, long-term), while conditioning for the general stock market movements that tend to inflate the Pearson correlation estimates.

The estimated partial correlation coefficients can also be used for constructing a dissimilarity measure between stock returns, which, when incorporated in the k-means clustering algorithm, generates separate cluster formations by time scale. This is an important distinction for investors differing in terms of their investment horizon orientations.

The results of an empirical study using all the major S&P 500 sector indices in the US, as well as precious metals and energy sector futures returns during the last decade, suggest that cluster formations (number of clusters, synthesis, and degree of homogeneity) do vary

by time scale, which entails different implications for investors wishing to diversify risk away from their portfolios but differ in terms of their investment horizon orientation.

Similarly to [Kenett et al. \(2015\)](#), our findings also suggest that stocks can be correlated outside their primary sector classification and the correlation levels vary by time scale. These findings have implications for the stock selection strategies of investors relying on sectoral diversification in their efforts to diversify risk away from their portfolios. The degree of comovement between the different economic activity sectors is higher when considering long-term cycles. Therefore, diversification is less important in the long-run, a finding also reported by [Rua and Nunes \(2009\)](#). In contrast, more clusters were found to exist over short-term cycles, suggesting the existence of sectors with distinct market movements and therefore more opportunities for diversification.

Furthermore, accounting for movements in the general stock market index changes the correlation levels between the different economic activity sectors, which suggests that relying on simple Pearson correlation estimates might not be adequate for constructing well-diversified portfolios. Our proposed method can therefore be used as a supplementary tool that can reveal previously unobserved correlations between stocks, not dominated by movements in the general stock market index.

Our research can be extended in several directions that can improve some of the limitations associated with our proposed method. First, it would be useful to consider the incorporation of alternative wavelet-based dissimilarity measures in the clustering algorithm that can potentially improve the within-cluster dispersions when working with stock market returns. Moreover, the areas of application can be extended to include additional asset classes such as fixed-income securities. Second, it is worth considering the use of alternative wavelet transforms for calculating the partial correlations between stock market returns, such as the discreet wavelet transform and the maximal overlap discreet wavelet packet transform ([Percival and Walden 2000](#)). Similarly, using alternative wavelet families can potentially improve the resolutions of stock market returns at higher frequencies. These are fruitful areas for further research, particularly with the aid of simulation studies that can evaluate the performance of the different methods with varying sample sizes.

Third, financial markets constitute complex environments influenced by the activities of multiple agents interacting over different cycles. In such environments, it is highly likely that correlations between stock market returns are influenced by many common factors and not just the general stock market index. Developing alternative methods for clustering stock returns over different time scales, while conditioning for multiple common factors in the analysis, is an interesting area for further research.

Finally, the partial correlation coefficient between stock returns, much like the Pearson correlation coefficient, is not static and could change periodically, particularly during times of extreme market volatility. Therefore, it is important to update the estimated coefficients frequently and in line with changes in the market conditions, especially when working with individual stock returns, which are more directly influenced by changes in the individual strategies of companies.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data are available upon request.

**Acknowledgments:** The author would like to thank Guy Nason and three anonymous referees for useful comments and suggestions.

**Conflicts of Interest:** The author declares no conflict of interest.

Appendix A. Partial Correlation Matrices for Time Scales 3, 6 and 8

Time scale 3

	HGU1	SPLRCD	SPNY	SPSY	NGU1	ZGQ1	SPXHC	SPLRCI	SPLRCT	SPLRCM	BV1	SPLRCREC	ZIU1	SPLRCS	SPLRCL	SPLRCU
HGU1	1.000	-0.151	0.190	0.200	-0.024	0.324	-0.092	0.221	-0.188	0.411	0.077	0.014	0.385	-0.119	0.017	-0.238
SPLRCD	-0.151	1.000	-0.330	-0.414	-0.042	-0.084	0.026	-0.410	0.194	-0.162	-0.048	-0.069	-0.102	-0.090	-0.015	-0.105
SPNY	0.190	-0.330	1.000	0.262	0.075	0.118	-0.244	0.361	-0.440	0.240	0.543	0.045	0.204	-0.034	0.169	-0.049
SPSY	0.200	-0.414	0.262	1.000	-0.194	-0.142	-0.131	0.562	-0.482	0.349	0.013	-0.021	-0.050	-0.138	-0.126	-0.209
NGU1	-0.024	-0.042	0.075	-0.194	1.000	-0.076	0.083	-0.155	0.008	-0.229	0.143	-0.012	0.047	-0.052	0.130	0.023
ZGQ1	0.324	-0.084	0.118	-0.142	-0.076	1.000	0.064	-0.002	-0.146	0.323	-0.037	0.180	0.790	0.173	-0.100	0.190
SPXHC	-0.092	0.026	-0.244	-0.131	0.083	0.064	1.000	-0.162	-0.331	-0.066	-0.416	0.129	-0.096	0.150	-0.099	0.217
SPLRCI	0.221	-0.410	0.361	0.562	-0.155	-0.002	-0.162	1.000	-0.418	0.422	0.115	-0.052	0.053	-0.145	-0.030	-0.200
SPLRCT	-0.188	0.194	-0.440	-0.482	0.008	-0.146	-0.331	-0.418	1.000	-0.334	-0.035	-0.303	-0.134	-0.309	-0.255	-0.256
SPLRCM	0.411	-0.162	0.240	0.349	-0.229	0.323	-0.066	0.422	-0.334	1.000	0.087	-0.062	0.270	-0.152	-0.136	-0.252
BV1	0.077	-0.048	0.543	0.013	0.143	-0.037	-0.416	0.115	-0.035	0.087	1.000	-0.113	0.120	-0.060	0.114	-0.156
SPLRCREC	0.014	-0.069	0.045	-0.021	-0.012	0.180	0.129	-0.052	-0.303	-0.062	-0.113	1.000	0.138	0.353	-0.011	0.551
ZIU1	0.385	-0.102	0.204	-0.050	0.047	0.790	-0.096	0.053	-0.134	0.270	0.120	0.138	1.000	0.057	-0.030	0.081
SPLRCS	-0.119	-0.090	-0.034	-0.138	-0.052	0.173	0.150	-0.145	-0.309	-0.152	-0.060	0.353	0.057	1.000	0.250	0.525
SPLRCL	0.017	-0.015	0.169	-0.126	0.130	-0.100	-0.099	-0.030	-0.255	-0.136	0.114	-0.011	-0.030	0.250	1.000	0.168
SPLRCU	-0.238	-0.105	-0.049	-0.209	0.023	0.190	0.217	-0.200	-0.256	-0.252	-0.156	0.551	0.081	0.525	0.168	1.000

Time scale 6

	HGU1	SPLRCD	SPNY	SPSY	NGU1	ZGQ1	SPXHC	SPLRCI	SPLRCT	SPLRCM	BV1	SPLRCREC	ZIU1	SPLRCS	SPLRCL	SPLRCU
HGU1	1.000	-0.137	0.089	0.168	0.443	0.252	-0.215	0.259	-0.158	0.516	0.205	0.038	0.449	-0.146	0.194	-0.154
SPLRCD	-0.137	1.000	-0.617	-0.522	-0.072	0.064	-0.109	-0.219	0.377	-0.048	-0.226	-0.602	0.186	-0.195	-0.399	-0.523
SPNY	0.089	-0.617	1.000	0.602	0.359	-0.127	-0.292	0.446	-0.563	0.217	0.803	0.180	-0.019	-0.214	0.282	0.161
SPSY	0.168	-0.522	0.602	1.000	0.029	-0.567	-0.192	0.425	-0.502	-0.243	0.330	0.046	-0.481	-0.191	-0.095	-0.026
NGU1	0.443	-0.072	0.359	0.029	1.000	0.202	0.011	0.093	-0.413	0.388	0.510	-0.010	0.571	-0.190	0.165	-0.092
ZGQ1	0.252	0.064	-0.127	-0.567	0.202	1.000	0.043	-0.159	0.170	0.680	0.068	0.171	0.734	0.133	0.295	0.108
SPXHC	-0.215	-0.109	-0.292	-0.192	0.011	0.043	1.000	-0.408	-0.196	-0.235	-0.446	0.191	-0.123	0.252	-0.141	0.158
SPLRCI	0.259	-0.219	0.446	0.425	0.093	-0.159	-0.408	1.000	-0.498	0.467	0.446	0.019	0.002	0.233	0.029	0.388
SPLRCT	-0.158	0.377	-0.563	-0.502	-0.413	0.170	-0.196	-0.498	1.000	-0.190	-0.339	-0.365	-0.023	-0.332	-0.228	-0.363
SPLRCM	0.516	-0.048	0.217	-0.243	0.388	0.680	-0.235	0.467	-0.190	1.000	0.466	0.104	0.664	0.159	0.316	0.282
BV1	0.205	-0.226	0.803	0.330	0.510	0.068	-0.446	0.446	-0.339	0.466	1.000	-0.176	0.248	-0.364	0.099	-0.049
SPLRCREC	0.038	-0.602	0.180	0.046	-0.010	0.171	0.191	0.019	-0.365	0.104	-0.176	1.000	-0.006	0.675	0.694	0.742
ZIU1	0.449	0.186	-0.019	-0.481	0.571	0.734	-0.123	0.002	-0.023	0.664	0.248	-0.006	1.000	-0.011	0.232	-0.076
SPLRCS	-0.146	-0.195	-0.214	-0.191	-0.190	0.133	0.252	0.233	-0.332	0.159	-0.364	0.675	-0.011	1.000	0.395	0.848
SPLRCL	0.194	-0.399	0.282	-0.095	0.165	0.295	-0.141	0.029	-0.228	0.316	0.099	0.694	0.232	0.395	1.000	0.459
SPLRCU	-0.154	-0.523	0.161	-0.026	-0.092	0.108	0.158	0.388	-0.363	0.282	-0.049	0.742	-0.076	0.848	0.459	1.000

Time scale 8

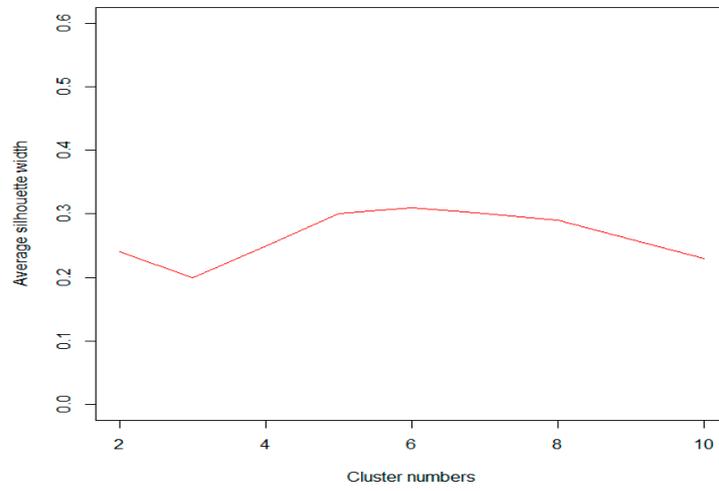
	HGU1	SPLRCD	SPNY	SPSY	NGU1	ZGQ1	SPXHC	SPLRCI	SPLRCT	SPLRCM	BV1	SPLRCREC	ZIU1	SPLRCS	SPLRCL	SPLRCU
HGU1	1.000	-0.635	0.286	-0.159	0.553	0.856	-0.979	0.139	0.517	0.699	0.672	0.219	0.848	-0.512	0.393	-0.592
SPLRCD	-0.635	1.000	-0.083	-0.274	-0.416	-0.336	0.706	-0.374	-0.492	-0.807	-0.236	0.216	-0.430	0.607	0.060	0.492
SPNY	0.286	-0.083	1.000	0.583	0.664	-0.013	-0.156	0.723	-0.639	-0.033	0.809	0.317	-0.003	-0.087	-0.308	0.226
SPSY	-0.159	-0.274	0.583	1.000	0.584	-0.618	0.166	0.796	-0.589	-0.018	0.279	-0.444	-0.544	-0.248	-0.801	0.152
NGU1	0.553	-0.416	0.664	0.584	1.000	0.131	-0.496	0.557	-0.158	0.295	0.779	-0.251	0.123	-0.559	-0.339	-0.356
ZGQ1	0.856	-0.336	-0.013	-0.618	0.131	1.000	-0.847	-0.223	0.625	0.575	0.394	0.494	0.967	-0.163	0.781	-0.457
SPXHC	-0.979	0.706	-0.156	0.166	-0.496	-0.847	1.000	-0.111	-0.609	-0.769	-0.560	-0.131	-0.849	0.514	-0.425	0.621
SPLRCI	0.139	-0.374	0.723	0.796	0.557	-0.223	-0.111	1.000	-0.564	0.360	0.357	0.075	-0.062	0.054	-0.338	0.178
SPLRCT	0.517	-0.492	-0.639	-0.589	-0.158	0.625	-0.609	-0.564	1.000	0.510	-0.152	-0.177	0.591	-0.470	0.439	-0.677
SPLRCM	0.699	-0.807	-0.033	-0.018	0.295	0.575	-0.769	0.360	0.510	1.000	0.092	0.086	0.726	-0.207	0.393	-0.497
BV1	0.672	-0.236	0.809	0.279	0.779	0.394	-0.560	0.357	-0.152	0.092	1.000	0.186	0.307	-0.477	-0.115	-0.159
SPLRCREC	0.219	0.216	0.317	-0.444	-0.251	0.494	-0.131	0.075	-0.177	0.086	0.186	1.000	0.530	0.585	0.649	0.351
ZIU1	0.848	-0.430	-0.003	-0.544	0.123	0.967	-0.849	-0.062	0.591	0.726	0.307	0.530	1.000	-0.074	0.795	-0.459
SPLRCS	-0.512	0.607	-0.087	-0.248	-0.559	-0.163	0.514	0.054	-0.470	-0.207	-0.477	0.585	-0.074	1.000	0.431	0.622
SPLRCL	0.393	0.060	-0.308	-0.801	-0.339	0.781	-0.425	-0.338	0.439	0.393	-0.115	0.649	0.795	0.431	1.000	-0.127
SPLRCU	-0.592	0.492	0.226	0.152	-0.356	-0.457	0.621	0.178	-0.677	-0.497	-0.159	0.351	-0.459	0.622	-0.127	1.000

HGU1—Copper futures; SPLRCD—Consumer discretionary; SPNY—Energy; SPSY—Financials; NGU1—Natural gas futures; ZGQ1—Gold futures; SPXHC—Health care; SPLRCI—Industrials; SPLRCT—Information Technology; SPLRCM—Materials; BV1—Brent oil futures; SPLRCREC—Real estate; ZIU1—Silver futures; SPLRCS—Consumer staples; SPLRCL—Telecom services; SPLRCU—Utilities

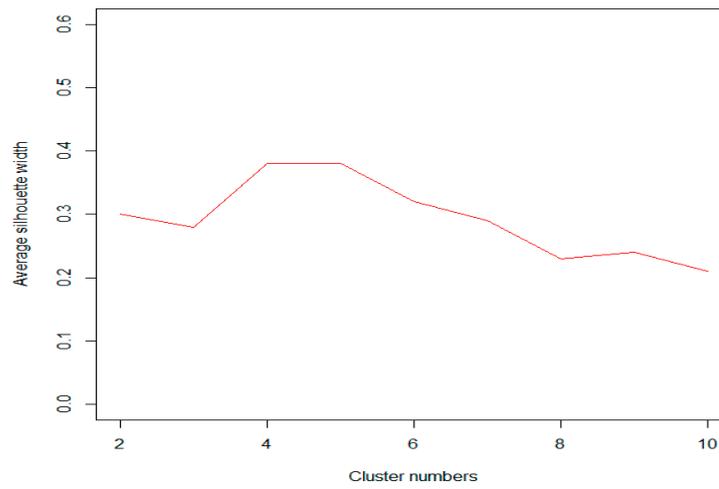


### Appendix C. Silhouette width Analysis

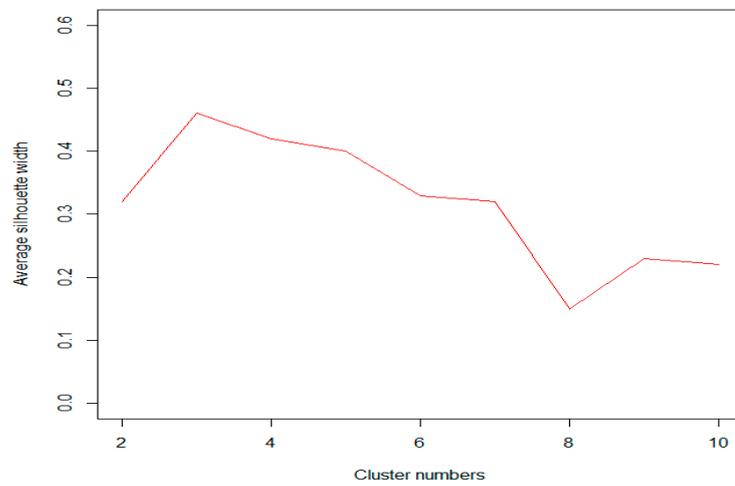
**Time scale 3**



**Time scale 6**

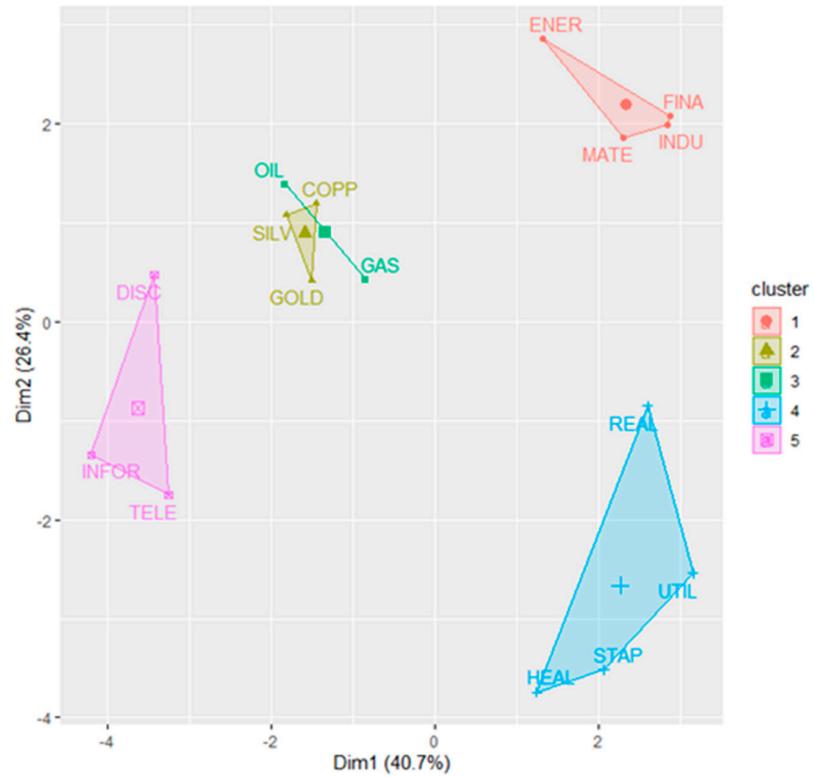


**Time scale 8**

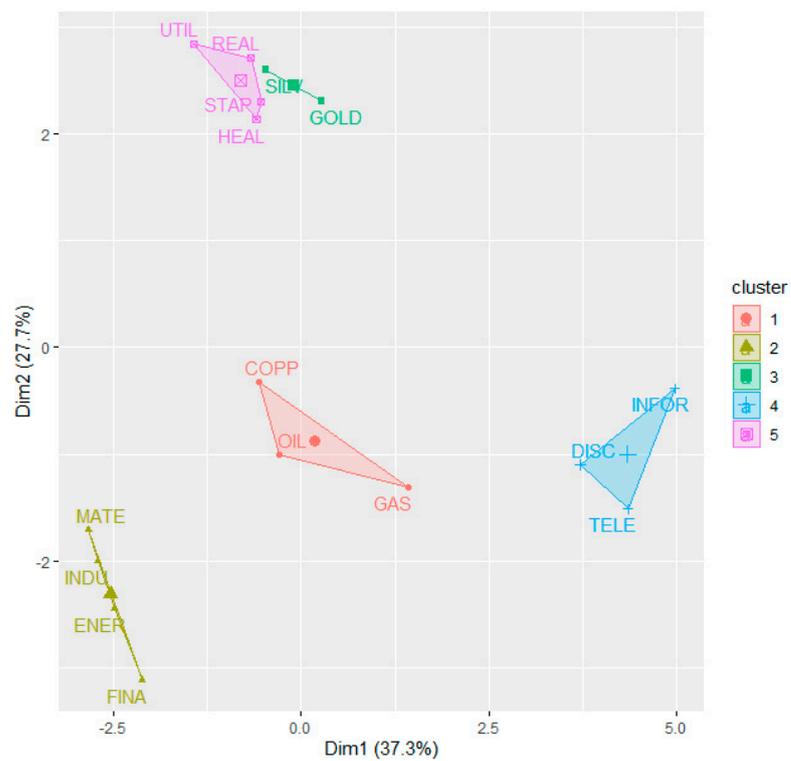


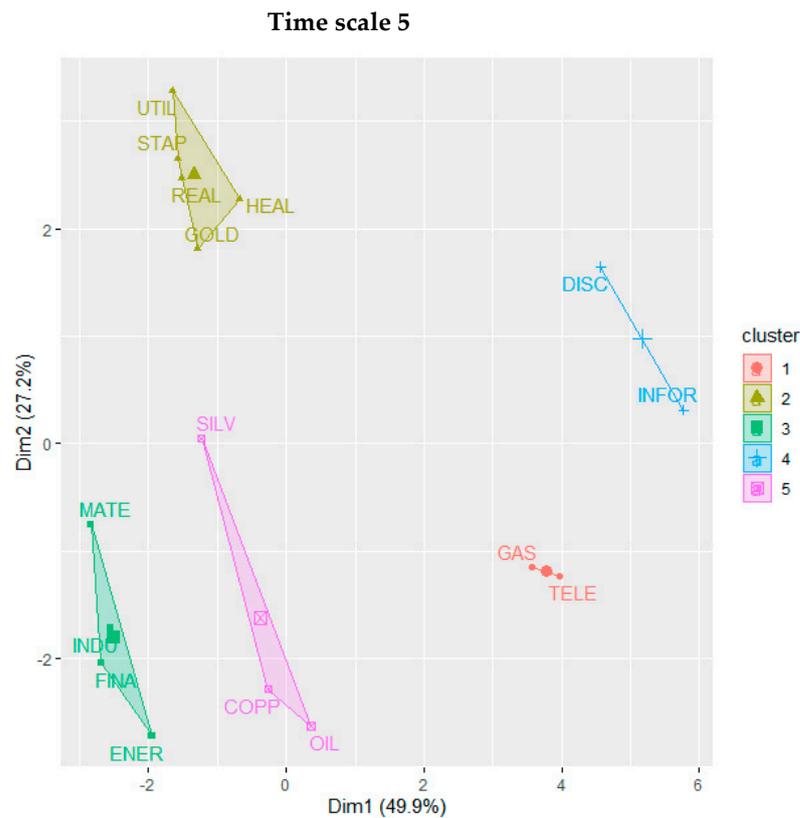
**Appendix D. Cluster Formations for Time Scales 3, 4 and 5 Using Daily Observations for the Period: 1 May 2019–8 July 2021**

**Time scale 3**



**Time scale 4**





## References

- Ahn, Dong-Hyun, Jennifer Conrad, and Robert F. Dittmar. 2009. Basis assets. *The Review of Financial Studies* 22: 5133–74. [\[CrossRef\]](#)
- Alqaralleh, Huthaifa, and Alessandra Canepa. 2021. Evidence of Stock Market Contagion during the COVID-19 Pandemic: A Wavelet-Copula-GARCH Approach. *Journal of Risk and Financial Management* 14: 329. [\[CrossRef\]](#)
- Bastos, João A., and Jorge Caiado. 2014. Clustering financial time series with variance ratio statistics. *Quantitative Finance* 14: 2121–33. [\[CrossRef\]](#)
- Bennett, Mark J., and Dirk L. Hugen. 2016. *Financial Analytics with R*. Cambridge: Cambridge University Press.
- Berthold, Michael R., and Frank Höppner. 2016. *On Clustering Time Series Using Euclidean Distance and Pearson Correlation*. Working Paper. Konstanz: University of Konstanz.
- Bruzda, Joanna. 2014. Forecasting via Wavelet Denoising: The Random Signal Case. In *Wavelet Applications in Economics and Finance, Dynamic Modeling and Econometrics in Economics and Finance*. Edited by Marco Gallegati and Willi Semmler. Cham: Springer, vol. 20, pp. 187–225.
- Bubb, Ryan, and Emiliano M. Catan. 2021. The Party Structure of Mutual Funds. *The Review of Financial Studies*. forthcoming. [\[CrossRef\]](#)
- Camp, Kevin M., David Mead, Stephen B. Reed, Christopher Sitter, and Derek Wasilewski. 2020. From the barrel to the pump: The impact of the COVID-19 pandemic on prices for petroleum products. In *Monthly Labor Review*. Washington: U.S. Bureau of Labor Statistics. [\[CrossRef\]](#)
- Capobianco, Enrico. 2003. Empirical volatility analysis: Feature detection and signal extraction with function dictionaries. *Physica A* 319: 495–518. [\[CrossRef\]](#)
- Caraiani, Petre. 2017. Evaluating exchange rate forecasts along time and frequency. *International Review of Economics & Finance* 51: 60–81.
- Chen, Peter Y., and Paula M. Popovich. 2002. *Correlation: Parametric and Nonparametric Measures*. Thousand Oaks: Sage Publications.
- Chiang, Thomas C., Lin Tan, and Huimin Li. 2007. Empirical analysis of dynamic correlations of stock returns: Evidence from Chinese A-share and B-share markets. *Quantitative Finance* 7: 651–67. [\[CrossRef\]](#)
- Cox, Trevor F. 2005. *An Introduction to Multivariate Statistical Analysis*. London: Hodder Arnold.
- Creamer, Germán G., and Chihoon Lee. 2019. A multivariate distance nonlinear causality test based on partial distance correlation: A machine learning application to energy futures. *Quantitative Finance* 19: 1531–42. [\[CrossRef\]](#)
- Cuthbertson, Keith, and Dirk Nitzsche. 2004. *Quantitative Financial Economics: Stocks, Bonds and Foreign Exchange*. Chichester: John Wiley and Sons.
- Engle, Robert, and Riccardo Colacito. 2006. Testing and Valuing Dynamic Correlations for Asset Allocation. *Journal of Business & Economic Statistics* 24: 238–53.
- Everitt, Brian, and Torsten Hothorn. 2011. *An Introduction to Applied Multivariate Analysis with R*. New York: Springer.

- Everitt, Brian S., Sabine Landau, Morven Leese, and Daniel Stahl. 2011. *Cluster Analysis*, 5th ed. Chichester: John Wiley and Sons.
- Fan, Yanqin, and Ramazan Gençay. 2010. Unit root tests with wavelets. *Econometric Theory* 26: 1305–31. [CrossRef]
- Fenn, Daniel J., Mason A. Porte, Peter J. Mucha, Mark McDonald, Stacy Williams, Neil F. Johnson, and Nick S. Jones. 2012. Dynamical clustering of exchange rates. *Quantitative Finance* 12: 1493–520. [CrossRef]
- Fleming, Brian J. W., Dejin Yu, Robert G. Harrison, and David Jubb. 2000. Analysis of effect of detrending of time scale structure of financial data using discrete wavelet transform. *International Journal of Theoretical and Applied Finance* 3: 357–79. [CrossRef]
- Focardi, Sergio M., and Frank J. Fabozzi. 2004. A methodology for index tracking based on time-series clustering. *Quantitative Finance* 4: 417–25. [CrossRef]
- Forbes, Kristin J., and Roberto Rigobo. 2002. Non contagion, only interdependence: Measuring stock market comovements. *Journal of Finance* 57: 2223–61. [CrossRef]
- Gallegati, Marco, Mauro Gallegati, James Bernard Ramsey, and Willi Semmler. 2011. The US Wage Phillips Curve across Frequencies and over Time. *Oxford Bulletin of Economics and Statistics* 73: 489–508. [CrossRef]
- Gallegati, Marco, Mauro Gallegati, James Bernard Ramsey, and Willi Semmler. 2016. Productivity and unemployment: A scale-by-scale panel data analysis for the G7 countries. *Studies in Nonlinear Dynamics & Econometrics* 20: 4.
- Geertsema, Paul, and Helen Lu. 2020. The correlation structure of anomaly strategies. *Journal of Banking and Finance* 119: 105934. [CrossRef]
- Gençay, Ramazan, and Nikola Gradojevi. 2011. Errors-in-variables estimation with wavelets. *Journal of Statistical Computation and Simulation* 81: 1545–64.
- Gençay, Ramazan, and Daniele Signori. 2015. Multi-scale tests for serial correlation. *Journal of Econometrics* 184: 62–80.
- Gençay, Ramazan, Faruk Selçuk, and Brandon Whitcher. *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics*. New York: Academic Press.
- Gençay, Ramazan, Selçuk Faruk, and Brandon Whitcher. 2005. Multiscale systematic risk. *Journal of International Money and Finance* 24: 55–70.
- Gilles, Fay, Moulines Eric, François Roueff, and Murad S. Taqqu. 2009. Estimators of long-memory: Fourier versus wavelets. *Journal of Econometrics* 151: 159–77.
- Hong, Yongmiao, and Chihwa Kao. 2004. Wavelet-based testing for serial correlation of unknown form in panel models. *Econometrica* 72: 1519–63. [CrossRef]
- In, Francis, and Sangbae Kim. 2013. *An Introduction to Wavelet Theory in Finance: A Wavelet Multiscale Approach*. Singapore: World Scientific.
- Jensen, Mark J. 2000. An alternative maximum likelihood estimator of long-memory processes using compactly supported wavelets. *Journal of Economic Dynamics & Control* 24: 361–87.
- Jung, Sean S., and Woojin Chang. 2016. Clustering stocks using partial correlation coefficients. *Physica A* 462: 410–20. [CrossRef]
- Kassambara, Alboukadel, and Fabian Mundt. 2020. Factoextra: Extract and Visualize the Results of Multivariate Data Analyses. Available online: <https://CRAN.R-project.org/package=factoextra> (accessed on 2 December 2021).
- Kaufman, Leonard, and Peter J. Rousseeuw. 1990. *Finding Groups in Data. An Introduction to Cluster Analysis*. New York: John Wiley & Sons, Inc.
- Kenett, Dror Y., Xuqing Huang, Irena Vodenska, Shlomo Havlin, and Eugene H. Stanley. 2015. Partial correlation analysis: Applications for financial markets. *Quantitative Finance* 15: 569–78. [CrossRef]
- Kim, Seongho. 2015. ppcor: An R Package for a Fast Calculation to Semi-partial Correlation Coefficients. *Communications for Statistical Applications and Methods* 22: 665–74. [CrossRef] [PubMed]
- Kim, Sangbae, and Francis In. 2010. Portfolio allocation and the investment horizon: A multiscale approach. *Quantitative Finance* 10: 443–53. [CrossRef]
- King, Mervyn A., and Sushil Wadhvani. 1990. Transmission of volatility between stock markets. *The Review of Financial Studies* 3: 5–33. [CrossRef]
- Krishnan, C. N. V., Ralitsa Petkova, and Peter Ritchken. 2009. Correlation risk. *Journal of Empirical Finance* 16: 353–67. [CrossRef]
- Kwapien, Jarosław, Paweł Oświęcimka, and Stanisław Drożdż. 2015. Detrended fluctuation analysis made flexible to detect range of cross-correlated fluctuations. *Physical Review E* 92: 052815. [CrossRef]
- Lee, Sang Bin, and Kwang Jung Kim. 1993. Does the October 1987 crash strengthen the co-movements among national stock markets? *Review of Financial Economics* 3L: 89–102. [CrossRef]
- Michis, Antonis A. 2011. Denoised Least Squares Forecasting of GDP Changes Using Indexes of Consumer and Business Sentiment. *IFC Bulletin* 25: 383–92.
- Michis, Antonis A. 2014a. Investing in gold: Individual asset risk in the long-run. *Finance Research Letters* 11: 369–74. [CrossRef]
- Michis, Antonis A. 2014b. Time Scale Evaluation of Economic Forecasts. *Economics Letters* 123: 279–81. [CrossRef]
- Michis, Antonis A. 2015a. Multiscale analysis of the liquidity effect in the UK economy. *Computational Economics* 45: 615–33. [CrossRef]
- Michis, Antonis A. 2015b. A Wavelet Smoothing Method to Improve Conditional Sales Forecasting. *Journal of the Operational Research Society* 66: 832–44. [CrossRef]
- Michis, Antonis A. 2019. The systematic risk of gold at different time scales. *Economics Bulletin* 39: 1215–27.
- Michis, Antonis A. 2021. Wavelet Multidimensional Scaling Analysis of European Economic Sentiment Indicators. *Journal of Classification* 38: 443–480. [CrossRef]

- Michis, Antonis A., and Nason Guy. 2017. Case Study: Shipping Trend Estimation and Prediction via Multiscale Variance Stabilisation. *Journal of Applied Statistics* 44: 2672–84. [[CrossRef](#)]
- Michis, Antonis A., and Theofanis Sapatinas. 2007. Wavelet Instruments for Efficiency Gains in Generalized Method of Moment Models. *Studies in Nonlinear Dynamics & Econometrics* 11: 4.
- Nason, Guy. 2008. *Wavelet Methods in Statistics with R*. New York: Springer.
- Nava, Noemi, T. Di Matteo, and Tomaso Aste. 2018. Dynamic correlations at different time-scales with empirical mode decomposition. *Physica A* 502: 534–44. [[CrossRef](#)]
- Obilor, Esezi Isaac, and Eric Chikweru Amadi. 2018. Test for Significance of Pearson's Correlation Coefficient ( $r$ ). *International Journal of Innovative Mathematics, Statistics & Energy Policies* 6: 11–23.
- Percival, Donald B., and Andrew T. Walden. 2000. *Wavelet Methods for Time Series Analysis*. Cambridge: Cambridge University Press.
- Risse, Marian. 2019. Combining wavelet decomposition with machine learning to forecast gold returns. *International Journal of Forecasting* 35: 601–15. [[CrossRef](#)]
- Rua, António. 2012. Money Growth and Inflation in the Euro Area: A Time-Frequency View. *Oxford Bulletin of Economics and Statistics* 74: 875–85. [[CrossRef](#)]
- Rua, António. 2017. Wavelet-based multivariate multiscale approach for forecasting. *International Journal of Forecasting* 33: 581–90.
- Rua, António, and Luís C. Nunes. 2009. International comovement of stock market returns: A wavelet analysis. *Journal of Empirical Finance* 16: 632–39. [[CrossRef](#)]
- Sifat, Imtiaz, Ghafoor Abdul, and Abdollah Ah Mand. 2021. The COVID-19 pandemic and speculation in energy, precious metals, and agricultural futures. *Journal of Behavioral and Experimental Finance* 30: 100498. [[CrossRef](#)]
- Tola, Vincenzo, Fabrizio Lillo, Mauro Gallegati, and Rosario N. Mantegna. 2008. Cluster analysis for portfolio optimization. *Journal of Economic Dynamics & Control* 32: 235–58.
- Wątarek, Marcin, Stanisław Drożdż, Paweł Oświęcimka, and Marek Stanuszek. 2019. Multifractal cross-correlations between the world oil and other financial markets in 2012–17. *Energy Economics* 81: 874–85. [[CrossRef](#)]