BRNO UNIVERSITY OF TECHNOLOGY

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

FAKULTA ELEKTROTECHNIKY
A KOMUNIKAČNÍCH TECHNOLOGIÍ

DEPARTMENT OF RADIO ELECTRONICS

ÚSTAV RADIOELEKTRONIKY

SELECTED ASPECTS OF STATISTICAL SIGNIFICANCE TESTING IN TIME-FREQUENCY ANALYSIS

VYBRANÉ ASPEKTY TESTOVÁNÍ STATISTICKÉ VÝZNAMNOSTI V ČASOVĚ-FREKVENČNÍ ANALÝZE

SHORT VERSION OF DOCTORAL THESIS

TEZE DIZERTAČNÍ PRÁCE

AUTHOR Ing. Eva Klejmová

AUTOR PRÁCE

SUPERVISOR doc. RNDr. Jitka Poměnková, Ph.D.

ŠKOLITEL

BRNO 2022

ABSTRACT

This doctoral thesis is focused on analyses and assessment of the quality of the frequency and time-frequency transform of the data and the formulation of recommendations for working with such methods. When using these methods, the question arises of how to evaluate which components of the spectrogram are statistically significant and which are not. In this thesis, we analyze the properties of standard statistical significance tests. We discuss their advantages and disadvantages taking into account the heteroskedastic character of data. Based on our experiments we propose two types of improved testing methods that reduce the negatives standard tests. The final step is creating a framework for data filtering using our proposed methods.

KEYWORDS

spectrogram, time-frequency analysis, wavelet transform, Fourier transform, autoregressive process, significance testing, co-movement filtering

ABSTRAKT

Přeložená dizertační práce se zabývá analýzou a posouzením kvality odhadu frekvenční a časově-frekvenční transformace dat a formulaci doporučení pro práci s metodami. Při použití těchto metod vyvstává otázka, jak vyhodnotit, které složky spektrogramu jsou statisticky významné a které nikoli. V této práci analyzujeme vlastnosti standardních testů statistické významnosti. Diskutujeme o jejich výhodách a nevýhodách s ohledem na heteroskedastický charakter dat. Na základě našich experimentů jsou v práci navrženy dva typy testovacích metod, které snižují negativní aspekty standardních testů. Práce jen zakončena vytvořením rámce pro filtrování dat pomocí námi navržených metod.

KLÍČOVÁ SLOVA

spektrogram, časově frekvenční analýza, vlnková transformace, Fourierova transformace, autoregresivní proces, testování významnosti, filtrování společného pohybu

Contents

Introduction

The need to analyze the data can be found across a variety of scientific disciplines. Despite the diversity of disciplines, it is a common goal to obtain the maximum information from data analysis to help solve the tasks set. Concerning the scientific area, such data are given as observations in the form of time series of input signals. The standard analytical instruments are given in the time or frequency domain. Linking of both approaches giving us a more compact view can be done via time-frequency techniques. The combination of time and frequency tools provides a more efficient means of data analysis, allowing us a deep look into the signal structure.

The graphical representation of time-frequency analysis is a spectrogram. Its estimation can be done via several parametric or nonparametric methods. The most used are short-time Fourier transform, estimation via the time-frequency varying autoregressive process, and wavelet transform. While the periodogram is the classic estimator for stationary signals, multiple windows or short-time Fourier transformation can be useful for non-stationary signals. The time-frequency varying autoregressive process is a simplification of the general autoregressive moving average model. The signals can be corrupted by noise which can affect the precision of instantaneous frequency; therefore is good to investigate several types of analyses methods to reach the required precision.

The key aspect of time-frequency analysis is the precision of the estimated spectrogram. For further processing and filtering of data, it is appropriate to specify which components of the spectrogram are statistically significant and which are not. There are several test methods for this purpose. One of the most used methods is based on the identified distribution of background noise, and several requirements need to be met for its appropriate usage. Other methods work with the use of geometric and topological changes or simulations of background noise. The obtained selection of significant regions can then be used for further description and filtering of the data, taking into account the objectives of the analysis, either in time, frequency, or time-frequency domain.

Considering the current progress in the field and the gap in current research, this work is focused on describing the framework of analysis from the use of individual methods, through statistical significance testing of their results, and finally filtering of data based on these significance tests. The following chapter contains the core of the dissertation, detailed analyses are included only in the full version of the dissection.

1 State of the Art

The need to describe and analyze the input signal for further use occurs across all scientific disciplines, from technical to social sciences. In terms of approach, we can define the analysis in the time domain (TD), frequency domain (FD), and time-frequency domain (TFD). Fundamental analysis can be performed in the time domain. Such an analysis deals with the changes in a signal over a span of time, i.e. variation of the amplitude of the signal with time. In contrast, the frequency domain describes the behavior of the signal across a given frequency band concerning a range of frequencies and can include information on phase shift. It is possible to use time-domain techniques or frequency domain techniques separately; however, their ability to capture the frequency behavior of the analyzed time series with respect to the time is somewhat limited. The combination of time and frequency tools provides a more efficient means of statistical analysis, reflecting the fact that the time-frequency analysis of input signal is an instrument that has been used in interdisciplinary analysis for a long time.

Time-frequency (TF) techniques are an instrumental approach, reflecting both the time and frequency behavior of input time series. These approaches predominates in the last decade in many fields of science. It is a useful instrument in natural sciences [1–4], engineering [5, 6], biology or medicine [7–9] or social and economic sciences.

The time-frequency representation of the signal can be estimated via several approaches. The most common method is Short Time Fourier Transform (STFT). The periodogram or its modification, such as the multiple window method using Slepian sequences [9] can also be used. We can also use estimation via the time-frequency varying Autoregressive Process (TFAR) [10], wavelet analysis (CWT) [11, 12] or alternatively Modified empirical mode decomposition method [13]. While the periodogram belongs to the group of the classic estimator for stationary signals, multiple windows or STFT can be a valuable instrument for non-stationary signals [9, 14–16]. As Jiang and Mahadevan [17] wrote, the advantage of the wavelet analysis is that it can capture the features of non-stationarity signal due to the simultaneous time-frequency decomposition of inputs. The TFAR process is a simplification of the general AutoRegressive Moving Average (ARMA) model.

Among the advantages of Fourier transform and its derivatives, we can include low computational complexity and a wide range of software and hardware implementations with a selection of optimal parameters that provide satisfactory results. AR process used for estimation of signal spectrum representation provides fair results, especially in very short signals when STFT tends to fail. For longer signals, it provides good results [18]. In such cases, the variance of insignificant cyclical components that usually take the character of noise has a lower level than in the case of STFT. This advantage can be useful when we investigate thresholding such as in [19]. The time-varying representation of the AR process provides a more complex view compared to a simple spectrum estimate in the frequency domain only. It has time and frequency resolution corresponding to the size of the window and the size of window overlap, which must be selected. In such a way, it is similar to the STFT method. Unfortunately, the disadvantage of the method is its accuracy which strongly depends on the selection of optimal lag order. Therefore, it is good to investigate various optimization criteria for its optimal selection. Another disadvantage is that there are not many existing implementations; most are only on a software level. Continuous wavelet transform is a relatively new method compared to Fourier transform. As pointed out in [20] or [17], the wavelets have several advantages. It is applicable to non-stationary data. It also has the ability to uncover the latent process with changing cyclical patterns. Such features are typical for an economic time series. Additionally, the wavelet analysis has very good time resolution, and there is no need to optimize the parameters. There is only discussion about the mother wavelet and the scale selection.

The need to validate the estimated model arises with the application of TF methods on real values with respect to the application area (engineering, medicine, etc.). This leads to the significance testing [21, 22]. The fundamental work in this field can be found in Torrence and Compo [23]. This paper presents the comparison of the windowed Fourier transform to the wavelets. The authors also focus on the relationship between wavelet scale and Fourier frequency and the choice of an appropriate wavelet basis function. The proposed statistical significance test is given for wavelet power spectra and is based on theoretical derivation for white and red noise processes.

Motivated by the work of Torrence and Compo [23], Ge [24] proposes significance testing of wavelet power and wavelet power spectrum. He derived the sampling distributions for the power spectrum of a Gaussian White Noise (GWN). And also for the wavelet power of GWN. He proved that the results given by [23] are numerically accurate when if the sampling period factor is incorporated. Ge [25] uses the same methodological approach for wavelet cross-spectrum and linear coherence. However, one of the disadvantages of this test is that it takes into account the variance of the entire signal. In specific cases where the data exhibit highly variable volatility, the variance of the whole signal may not be sufficiently descriptive. The question is then how this affects the accuracy of the test and how to interpret the results.

A similar approach to Torrence and Campo can be found in the work of Schulte et al. [21]. They use geometric and topological methods for assigning contiguous significance regions of significant wavelet coefficient with respect to selected noise models with application to climatic data. Also, James and Fleming [22] use the Torrence and Compo approach to identify significant spatial scales of pattern and spatial boundaries in geo-science.

Model validation of structural dynamics example is proposed by Jiang and Mahadevan [17]. They investigate simulation-based predictions of structural response on the virtually generated data. The authors use testing with the help of Monte Carlo simulations to infer whether the model prediction and experimental observation represent two coherent processes. Wang et al. [26] present another point of view by introducing the general sequential Monte Carlo method to estimate the probability density function and to optimize wavelet transform for extracting bearing fault features.

Given the above methods, we found that the literature insufficiently describes several areas. One of these areas is the effect of the character of the data on significance testing of time-frequency methods. This character can manifest itself through structural changes in the data that lead to changes in volatility. This raises the question of how to interpret the results of standard tests in this case.

A related insufficiently described area is how TFA can be used to provide additional information that will contribute to subsequent signal analysis.

Most standard methods have been designed to work with technical data. The physical nature of these types of data is usually known; their description is available, their behavior and their content are known, and there is knowledge of what their deterministic components may look like. However, there are scientific areas where factors, that are often unpredictable, may affect or change the character of the data. This problem is typical, for example, for economic data, where due to various events (economic shocks, crises, pandemics, etc.) diverse structural changes may arise, such as in trend, volatility, growth, etc.

Given the above observations, this work will focus on the analysis of economic, technical, and simulated data.

2 Dissertation objectives

This dissertation thesis deals with analyses and assessment of the quality of the frequency and time-frequency transform and with the formulation of recommendations for working with such methods. We take into account how much a priori information will help to obtain maximum information about the data.

We researched literature and resources and evaluated current progress and gaps in this field. We found out that the literature does not deal with the influence of the data character on the significance testing of the time-frequency methods. Most of the literature focuses on the technical area, where the application of methods and interpretation of results is facilitated by knowledge or information about the data character. In some cases, such as the selected photonic Doppler velocimetry data set, these are rather experimental data, therefore, a priori information may be more general.

The different types of data in terms of nature are economic data, which are less informative in terms of technical analysis. The process and mechanism of this data generation are influenced by factors such as unexpected events, economic shocks, psychological factors, etc., which are difficult to predict and simulate. This may appear as structural changes in the data, to which standard methods may not respond correctly in all cases. For the TFD application and subsequent testing, the question then arises as to how data with structural changes can be analyzed to obtain relevant results. The third type of data is simulated data used to verify standard and designed methods.

We focus on the issue of statistical testing of data mentioned above in order to verify the standard methods and to propose methods for cases where the data volatility is changing in time. Based on these we defined the following objectives of the thesis.

Objective I. *Is it possible to use and combine different characteristics of individual TFA methods to obtain relevant spectrogram?*
In Chapter 3 we propose an approach to incorporate advantages and suppress disadvantages of individual methods in order to bring out significant components and suppress noise.

Objective II. *How can we modify standard tests to eliminate/reduce their disadvantages and shortcomings in case of data with changing volatility?*
In Chapter 4 we propose adaptive testing approach and recommendation for their usage in case of data with changing volatility.

Objective III. *How can we use these modified test methods for subsequent data filtering?*

In some cases, it is useful to work selected spectral components represented in the time domain. Especially in the case of evaluation of time-series co-movement. Therefore, in Chapter 5 we propose co-movement filtering as an instrument for obtaining co-movement indicator. In its construction and application, the expertise gained from Objective II. is used.

3 Enhanced TF Representation

Based on analyses of selected TFA methods we analyze and assess the quality of selected methods and formulate the recommendation for working with such methods. We formulate recommendations for AR process optimization using the Monte Carlo method. We list the advantages and disadvantages of selected parametric and non-parametric TF methods taking into account data character.

To highlight important spectral components we propose combination of several TF methods. In each method background noise is depict with different characteristics. However significant spectral components should be captured in most cases. Based on such assumption we should be able to suppress the noise and highlight required components by using their combination. The procedure of this method is shown in Fig. 3.1. This procedure can be perceived as an alternative to significance tests, which we will discuss in the Chapter 4. The characteristics of input data must be taken into account. Therefore, we will show the application on different types of data.

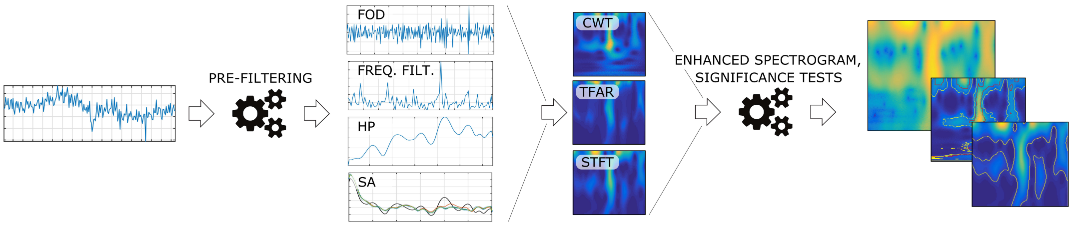


Fig. 3.1: Enhanced modelling of TF spectrograms.

3.1 Combination of TF methods

To obtain the best possible TF representation we combined results from the CWT, TFAR and STFT approach. Since the main focus was on the amplitude part of the spectra we have omitted phase part of complex spectra S_{CWT} and S_{STFT}. In case of focus on amplitude and phase components whole signal can be used for subsequent processing.

Firstly we align time axis (time resolution) of spectral representations S_{CWT}, S_{TFAR} and S_{STFT} so each spectrum would correspond to one another. All three time vectors have linearly increasing trend so for the time axis alignment the only requirement was to adjust starting and ending point for each method. We omitted first and last 15 columns of S_{CWT}, we denoted this remaining matrix as S'_{CWT}. By doing tis we ensured corresponding time axis for all three methods.

Secondly we needed to align the frequency/scale axis of S'_{CWT}, S_{TFAR} and S_{STFT}. The frequency range of S_{TFAR} and S_{STFT} was cropped to correspond the range of S'_{CWT} which was 1 year to 10 years cycles. Resulting frequency/business cycles vectors $\overline{f_{\mathrm{TFAR}}}$ and $\overline{f_{\mathrm{STFT}}}$ had a linearly increasing trend however trend of $\overline{f_{\mathrm{CWT}}}$ was non linear. To obtain corresponding vectors we matched each point of $\overline{f_{\mathrm{CWT}}}$ with one value of $\overline{f_{\mathrm{TFAR}}}/\overline{f_{\mathrm{STFT}}}$ with 1.4% tolerance:

$$\begin{aligned} |f_{\mathrm{CWT}} - f_{\mathrm{STFT}}| &\leq 0.014 \left|\max(\overline{f_{\mathrm{CWT}}}; \overline{f_{\mathrm{STFT}}})\right|, \\ |f_{\mathrm{CWT}} - f_{\mathrm{TFAR}}| &\leq 0.014 \left|\max(\overline{f_{\mathrm{CWT}}}; \overline{f_{\mathrm{TFAR}}})\right|. \end{aligned} \tag{3.1}$$

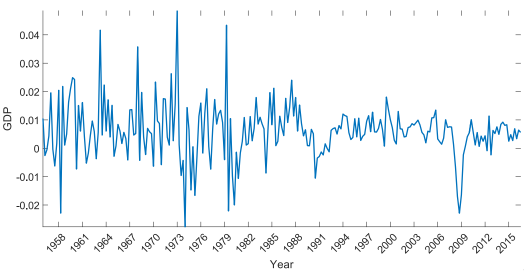
With this step we have gained adjusted TF matrices S'_{TFAR} and S'_{STFT} making all three methods aligned. For the methods combination we selected simple multiplication. We used combination of CWT and TFAR ($S_{\mathrm{CWT,TFAR}}$) and combination of CWT, TFAR and STFT ($S_{\mathrm{CWT,AR,STFT}}$):

$$\begin{aligned} S_{\mathrm{CWT,TFAR}} &= S'_{\mathrm{CWT}} S'_{\mathrm{TFAR}}, \\ S_{\mathrm{CWT,TFAR,STFT}} &= S'_{\mathrm{CWT}} S'_{\mathrm{TFAR}} S'_{\mathrm{STFT}}. \end{aligned} \tag{3.2}$$

3.2 Application on Economic Data

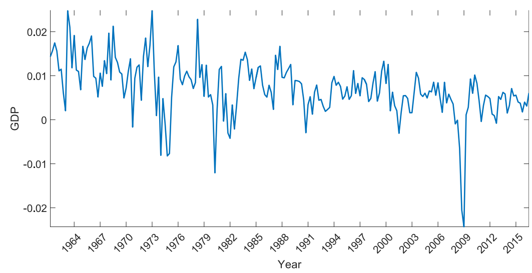
3.2.1 Data Describtion

As a representative of economic data with not clearly descriptive background noise, we use seasonally adjusted quarterly data of GDP. We selected volume index in OECD reference year 2010 [27] of the United Kingdom (UK) in 1956/01-2016/03 and Group of 7 (G7) in 1961/02-2016/03. All variables are in FODLOG (Fig. 3.2). G7 countries are: Canada, France, Germany, Italy, Japan, the United Kingdom, and the United States.



,

(a) UK



(b) G7

Fig. 3.2: GDP of UK and G7 in time domain.

Our analyses consists from several steps. In the first, we analyse data using CWT. We set scales to correspond range of 1 year to 10 years, with 257 individual scales. As mother wavelet we selected complex Morlet with center frequency $f_{\mathrm{b}} = 1.5$. The complex Morlet wavelet is based on standard Morlet with the advantage of providing complex results making it possible to obtain phase part (quadrature) of spectrum. In case of TF estimation via TFAR process we used Burg approach for coefficient estimates on 30 samples with 29 samples overlay and Hann window. Optimal value of lag order was based on AIC criteria. Parameters of STFT were set to correspond TFAR settings (30 samples, 29 samples overlay, Hann window) to simplify the process of combination of methods.

The data and results for UK and G7 are presented graphically in Fig. 3.2a-b, in Fig. 3.3a-f and Fig. 3.4a-d. There are four types of figures. Namely time representation of GDP for UK and G7 (Fig. 3.2a-b), TF transform via CWT (Fig. 3.3a-b), TF transform via AR (Fig. 3.3c-d), transformation via STFT (Fig. 3.3e-f) and adjustment of CWT picture with the help of AR (Fig. 3.4a-b) and with the help of TFAR+STFT (Fig. 3.4c-d).

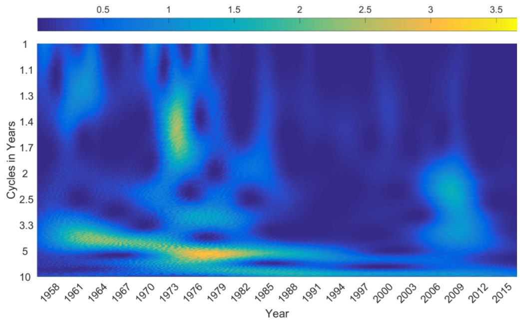
3.2.2 Results

After a short analyses of time representation of the data we apply TF approaches. Firstly we modelled CWT (Fig. 3.3a–b), consequently TFAR (Fig. 3.3c–d) and STFT (Fig. 3.3e–f). As we expected CWT provides results with very good time resolution. We can see several important areas across time and frequency. Focusing on TFAR representation the results are not so clear from time perspective as CWT, but they give us better information from frequency perspective similarly as STFT. Therefore, we decided to do adjustment of CWT picture with the help of TFAR and TFAR+STFT according to the calculation (eq. 3.1,3.2) described in Combination of TF methods.

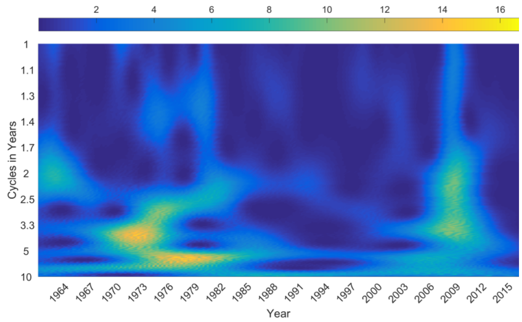
3.3 Application on Engineering Data

3.3.1 Data Describtion

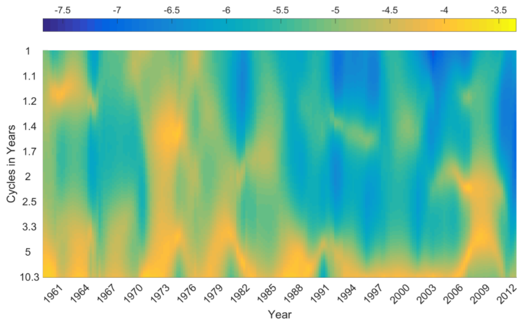
Another selected data type was Photonic Doppler Velocimetry Data (PDV) data. Their parameters are based on physical nature and their structure is thus different from economic data. As input data we took STFT, TFAR and wavelet spectral representation of the data. Equations (3.1) and (3.2) were then used. Because the main focus was on the amplitude part of the spectra we used only the amplitude part of STFT and wavelets.



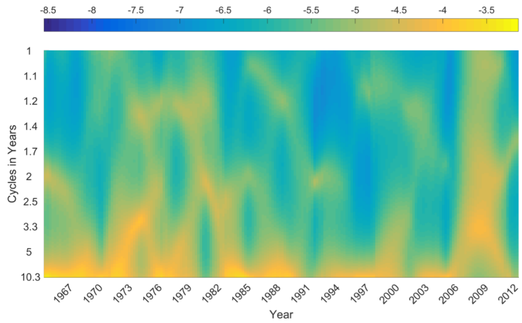
(a) CWT UK



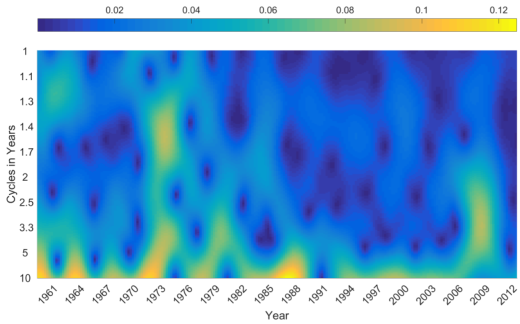
(b) CWT G7



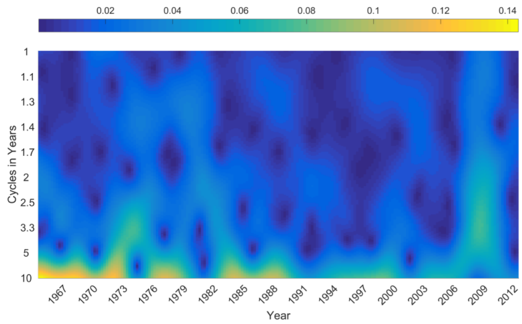
(c) TFAR UK



(d) TFAR G7



(e) STFT UK



(f) STFT G7

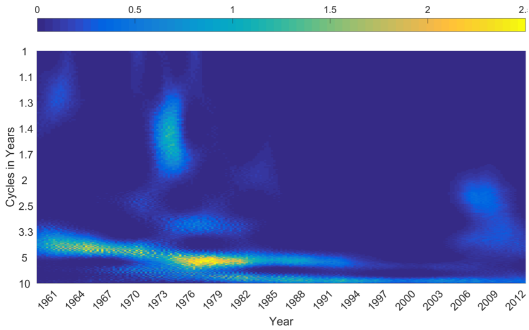
Fig. 3.3: Spectrum of GDP of UK and G7.

3.3.2 Results

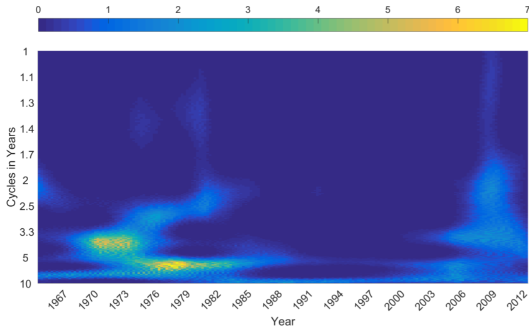
Resulting modified spectrogram is in Fig. 3.5. We can see that the scatter of background noise is smoothed and the data signal is more clearly visible. Even on this type of input data, it was confirmed that the method can highlight the required components in the spectrogram and therefore provides required advantages for further processing.

3.4 Chapter Conclusion

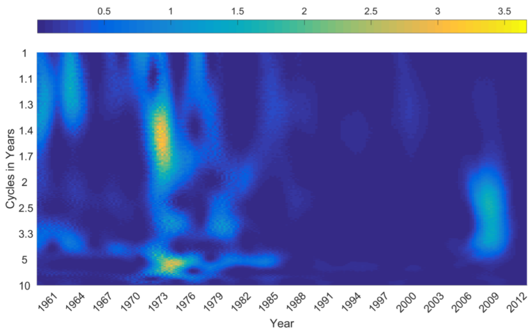
If we review results, by combining several TF approaches we were successful in background noise suppression. Consequently, events of interest became more visible



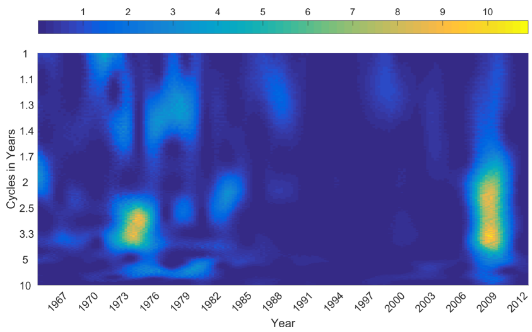
(a) Enhanced UK



(b) Enhanced G7

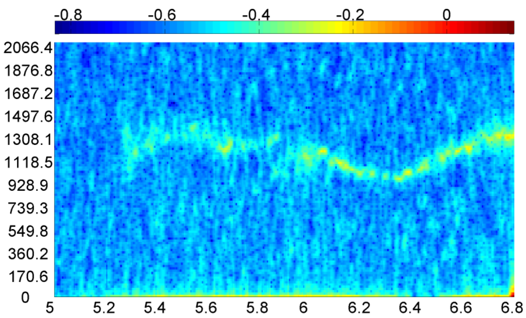


(c) Enhanced UK

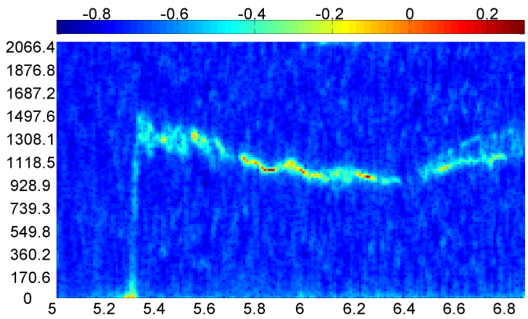


(d) Enhanced G7

Fig. 3.4: Adjustment of TF methods.



(a) signal A



(b) signal B

Fig. 3.5: Enhanced PDV data ($x-$axis: time, $y-$axis: frequencies).

and their identification in time, as well as in frequency was easier. An example of the possible use of this identification is the trend detection in a spectrogram. This approach can also be taken as a supplement to the significance testing with the investigation of background noise description, which will be described in Chapter 4.

4 Adaptive Significance Tests

Time-frequency transform can give reasonable results of both perspectives, time and frequency, in one moment. In some branches such engineering the physical nature of inputs is obvious and give valuable information. We can assume existence of several harmonic components corresponding to the specific frequency during all time of given input. Unfortunately, in others such in economy or sociology it may not be so simple. Applications of TF analyses have been so far limited by the fact that it was impossible to draw any implications on the statistical significance. Thus, significance testing of obtained results are welcomed. The original contribution in the spectrogram testing was provided by Torrence and Compo [23], followed by Ge [24, 28]. Both provided a framework for testing individual spectrograms as well as testing of co-movement representation. We denote them both as standard testing approach (STA). Both testing statistics presented in [23] and [24, 28] are formed as the power value of the spectrogram of a noise signal normalized by the signal variance in the time domain. In the case of an input signal with strongly localized fluctuations of the signal strength, the total variance may not sufficiently describe the character of the data. It is, therefore, not surprising that high amplitude events, may have a strong impact causing a suppression of other events. To avoid this problem, we propose an adaptive form of STA testing named a local-adaptive-based testing (LAB) and segmentation-adaptive-based testing (SAB). In the case when the data does not have such problem, the STA, LAB and SAB testing produce same results.

4.1 Standard Testing

STA significant testing which follows the work of Torrence and Compo [23] for the special case if the background spectra is Gaussian white noise and power wavelet cospectra (PWCS) is the W_2 distribution (see also Ge [25, 29, 30]). The significance level $Z(1-\alpha)$ for the risk α can be deduced from $1-\alpha$ percentile of the W_2 distribution [23, 25].

Thus, the mask M (see eq. (5.1)) for the HT approach is given by

$$M(a,b) = \begin{cases} 1 & |W_{xy}(a,b)|^2 \geq thr \\ \\ 0 & |W_{xy}(a,b)|^2 < thr \end{cases} \tag{4.1}$$

where the threshold thr (given by STA according to TC98)

$$thr = \frac{1}{4}\sigma_x^2\sigma_y^2 Z(1-\alpha)$$

is a fix scalar number for the risk α for all PWCS coefficients. The value of $Z(1-\alpha)$ is calculated by STA ([23]).

4.2 Segmentation-Adaptive-Based Testing

The SAB testing is suitable if we are able to identify the sub-segments with different volatility in the data. Firstly, for each time series $x(t)$ and $y(t)$, we identify the moments of the change of the time series variance. It can be done by expert estimate (usually in the case if the data are filtered by the long-term component and take the form of fluctuation around x-axis) or by statistical testing [31]. Secondly, we arrange all moments for both time series in the ascending time-order and we split the time range into the segments (SG) reflecting volatility changes in $x(t)$ and $y(t)$. Consequently, we identify the critical value for the significance testing in each segment by STA.

The proposed SAB masking method is designed as follows:

1. Identify the moments of the variance change of the time series $x(t), y(t)$ via expert estimate or statistical testing.
2. Arrange all identified moments for both time series in the ascending order and split the time range $t = 1, \ldots, n$ into the segments $SG_j, j = 1, \ldots J$ reflecting volatility changes in $x(t)$ and $y(t)$.
3. Estimate the PWCS and split it into the segments $PWCS_j, j = 1, \ldots J$ according to segments SG_j.
4. Construct the segments of the mask $M_j(a,b)$ corresponding to the segment SG_j, i.e. calculate $M_j(a,b)$ in each segment SG_j according to eq. (4.1) with respect to the variances of the time series in $j-$th segment. That is,

$$M_j(a,b) = \begin{cases} 1 & |W_{xy,j}(a,b)|^2 \geq thr_j \\ \\ 0 & |W_{xy,j}(a,b)|^2 < thr_j \end{cases} \tag{4.2}$$

Here, in the SG_j segment, the threshold

$$thr_j = \sigma^2_{x,j}\sigma^2_{y,j} 0.25 Z(1-\alpha)$$

is a fixed scalar number, α is a risk, $\sigma^2_{x,j}\sigma^2_{y,j}$ are variances for time series x, y in the time segment SG_j and $|W_{xy,j}(a,b)|^2$ is the corresponding part of PWCS. The value of $Z(1-\alpha)$ is calculated by STA test [23]. Compared to the STA, in the case of SAB masking the threshold is the vector $thr = (thr_1, \ldots thr_J)$ adaptively changing with respect to the variances in segment.

5. Construct the mask $M(a,b)$ as the composition of the $M_j(a,b), j=1,\ldots,J$ mask segments, i.e.

$$M(a,b) = (M_1(a,\tau),\ldots M_J(a,b)) \tag{4.3}$$

where the $j-$th ($j=1,\ldots,J$) segment of the mask $M(a,b)$ corresponding to the time segment SG_j as described in eq. (4.2).

4.3 Local-Adaptive-Based Testing

The LAB testing is suitable if the variability of the data slowly increases or/and decreases, once or several times during the time range of the series x or/and y. Before starting the LAB algorithm, we have to set the value of l - the length of a sliding window. Consequently, we can identify the critical value for a significance testing in each segment by STA.

The proposed SAB masking method is designed as follows:

1. Select the length l of the sliding window as an odd number.
2. Estimate the PWCS for the time series $x(t), y(t)$.
3. Calculate the mask $M_t(a,b)$ in each time $t=1,\ldots,n$ according to (4.4) with the variances as described in (4.5), i.e.:

$$M_t(a,b) = \begin{cases} 1 & |W_{xy,t}(a,b)|^2 \geq thr_t \\ \\ 0 & |W_{xy,j}(a,b)|^2 < thr_t \end{cases} \tag{4.4}$$

and the threshold

$$thr_t = \sigma^2_{x,t}\sigma^2_{y,t}0.25Z(1-\alpha)$$

is the fixed scalar number in the $t-$th sliding window, α is a risk. The value of $Z(1-\alpha)$ is calculated by STA test, ([23]). The variance $\sigma^2_{x,t}$ is calculated as follows

$$\sigma^2_{x,t} = \begin{cases} \frac{1}{l-1}\sum_{i=1}^{l}(x(i)-\bar{x})^2 & t\in 1,\ldots(l-1)/2 \\ & t\in n-(l-1)/2+1,\ldots,n \\ \frac{1}{l-1}\sum_{i=t-(l-1)/2}^{t+(l-1)/2}(x(i)-\bar{x})^2 & t\in(l-1)/2+1,\ldots,n-(l-1)/2 \end{cases} \tag{4.5}$$

where l is the odd number representing the sliding window length, $\bar{x}$ is the mean value of the time series x in the sliding window. The variance $\sigma^2_{y,t}$ is calculated accordingly. Compared to the STA testing, in the case of LAB masking the threshold is the vector $thr = (thr_1,\ldots thr_n)$ adaptively changing with respect to the variances in the sliding window.

4. Construct the mask $M(a, b)$ as the composition of the $M_t(a, b), t = 1, \ldots, n$, i.e.

$$M(a, b) = (M_1(a, b), \ldots, M_n(a, b)) \tag{4.6}$$

where the $t-$th ($t = 1, \ldots, n$) part of the mask $M(a, b)$ is calculated according to eq. (4.5).

4.4 Experimental Results

To demonstrate the methods described above we present them both on simulated signals and on real economic data. In practice, across various non-technical disciplines, there are signals or time series for which the exact description of its character is not as clear as in technical signals [32]. While in the case of engineering, signals can be simulated as the simplification of the composition of several harmonic component, in the case of economic data, their structure is more complicated. Usually, it contains structural trend-breaks, outliers, cyclical components of close frequencies which can occur or diminish in different time sub-periods (not during the whole time), or nested cycles with different frequency limited in time [33–36]. Moreover, the nature of economic indicators play an important role and can influence the character of the frequency structure, e.g. business cycles, financial cycles etc. Then, it is quite difficult to simulate the universal behaviour of the economic series and its noising with a generalized artificial signal.

Therefore, we decided to model an artificial signal as a simplification of basic features in economic data. That is, a composition of signals which have co-movement in time-limited long-term trend (low frequency component), short time-limited middle-term co-movement with the high amplitude, middle-term co-movement during the whole time period (i.e. cyclical fluctuations in BC frequencies) and short-term trend (cyclical fluctuations of high frequency, such as seasonality) in the first half of time-period.

The quality of the identification of significant co-movement part in co-spectra (i.e. TF components) is evaluated via two metrics [37]. The first one evaluates how many TF components were significant and were not identified as significant by the test, i.e. relevant parts were not identified:

$$\mathrm{M1} = \frac{FN}{TP}. \tag{4.7}$$

The second metrics evaluates how many TF components were not significant and were identified as significant by test, i.e. irrelevant parts were identified:

$$\mathrm{M2} = \frac{FP}{TN}. \tag{4.8}$$

Here, TP (True Positives) is the number of correctly identified TF components in co-spectra; FN is the number of TF components in co-spectra which were significant but were not identified as significant; FP (False Positives) is the number of TF components in co-spectra which were insignificant but were identified as significant; and TN is the number of TF components in co-spectra which were insignificant and were identified as insignificant.

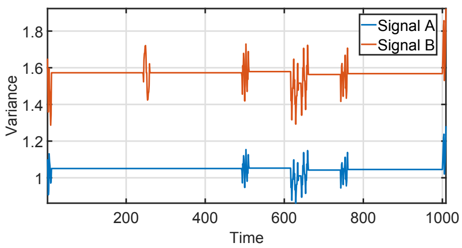
4.4.1 Simulated Data Description

For testing purposes, we have created four artificial signals, each of the length 1000 samples. Figure 4.1a) illustrates their time domain representation with constant variance, Figure 4.1b) with segmented variance (changing volatility), each for two signals. All signals were noised with signal-to-noise (SNR) ratio $SNR = 10, 3.16, 2$. In the simulation, we tried to approach the behavior of economic time series in the field of business cycle and synchrony analysis, therefore we selected $SNR = 10$ and 3.16 as stated above.

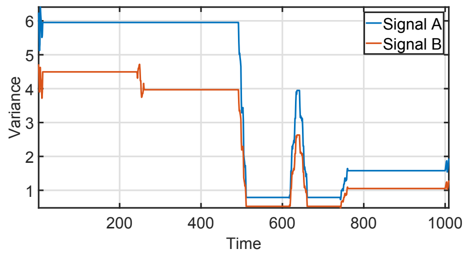
To be in correspondence with the real data analyses, we use the following settings during the analyses of artificial signals. For the PWCS, we set the scales in the range 1–10 years divided into 257 individual scales. Further, we use the complex Morlet with the center frequency $f_\mathrm{b} = 1.5$ as a mother wavelet. The LAB testing is done according to Sec. 4.3. The sliding window length used in the eq. (4.5) is the same for both signals (A and B) and is set to $l = 36$ samples, which corresponds to 3 years. The SAB testing is done according to Sec. 4.2. As the scope of this paper is not to investigate the optimal method for variance segmentation, the number of SG_j segments is set to match the number of segments in the artificial signals. Thus, $j = 5$ for the signal A and $j = 6$ for the signal B.

To be able to quantify the accuracy of the proposed SAB and LAB methods, we have created the so called benchmark figure of ideal PWCS $\pm$ *spread* for co-spectral components (see Fig. 4.2a. This was done to include energy spread in frequency for each individual wavelet. The size of the frequency spread is set as $\pm 15\%$ from the maximum in the center frequency of each individual wavelet. As the result, we can see a wider co-spectral components represented as yellow blocks in the figures (e.g. Fig. 4.2a). In the next step, we calculate the metrics M1 and M2 using the benchmark representation and the masked PWCS. Further, we noise all signals as mentioned in the first paragraph of this subsection and then estimate PWCS of signals with constant variance and PWCS of signals with segmented variance.

In the following figures (Figs. 4.2–4.5), the x-axis represents time, the y-axis represents specific periods (cycles in years) and the z-axis represents the values of spectrogram. The figures show a two-dimensional projection of three-dimensional



(a) Signals with constant variance.



(b) Signals with changing (segmented) variance.

Fig. 4.1: Behaviour of a variance of simulated signals in the time.

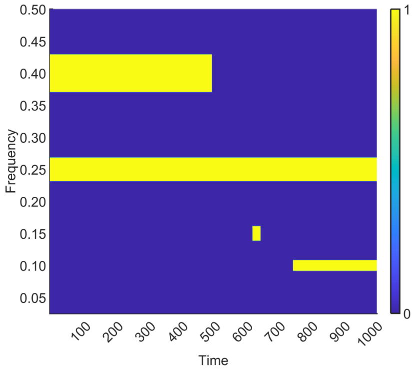
charts. The intensity of each contour represents the relative importance of the different periodicities and time, i.e. from dark blue (low amplitude) to yellow (high amplitude) colour. The yellow curve denotes the mask in all figures.

4.4.2 Results for Constant Variance

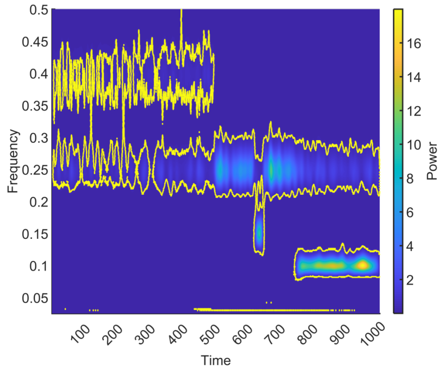
Next, we identify the significant co-movement of PWCS via HT, LAB, SAB masking for $SNR = 10$ (Figs. 4.2b,c,d) and for $SNR = 3.16$ (Figs. 4.3b,c,d). Then we calculate the metrics $FN, FP, M1, M2, \Delta M1, \Delta M2$ (Tab. 4.1). The metrics $\Delta M1, \Delta M2$ describe how the metrics changed for SAB, LAB with respect to HT. As we can see in Figs. 4.2, 4.3 and Tab. 4.1, there are no big differences between the results for HT, SAB and LAB masking when the variance of signals is constant. That is, the metrics $\Delta M1, \Delta M2$ are mostly lower than 1%; in the case of LAB ($SNR = 10$) the metric $\Delta M1 = -1.34$ is a little higher than 1%.

4.4.3 Results for Segmented Variance

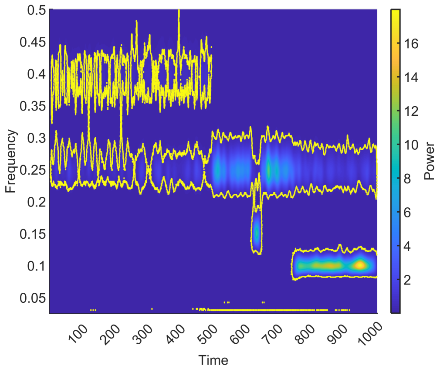
As for the signals with segmented variance, we also identify the significant co-movement of estimated PWCS via HT, LAB, SAB masking for $SNR = 10$ (Figs. 4.4b,c,d) and for $SNR = 3.16$ (Figs. 4.5b,c,d). Then we calculate the metrics $FN, FP, M1, M2$, $\Delta M1, \Delta M2$ (Tab. 4.2). Comparing Figs. 4.4b–d with Fig. 4.4a we can see that HT masking of PWCS was not able to identify well the frequency component corresponding to frequency 0.25 in the second half of the time. Moreover, in the case of 0.40, the frequency component PWCS HT masking covered a wider range of surrounding components than SAB, LAB masking. This fact is also documented in Tab. 4.2. The metrics $\Delta M1, \Delta M2$ describe how the metrics changed for SAB, LAB with respect to the HT. As we can see, there are differences between the results for HT, SAB and LAB masking for segmented variance compared to the



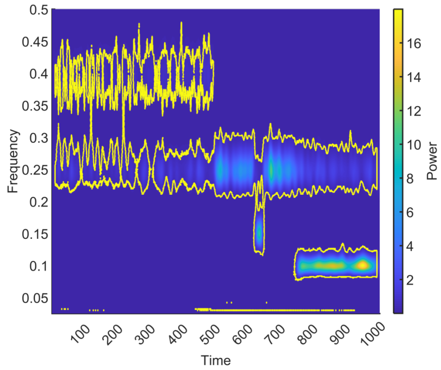
(a) Ideal case.



(b) PWCS HT masking.



(c) PWCS SAB masking.

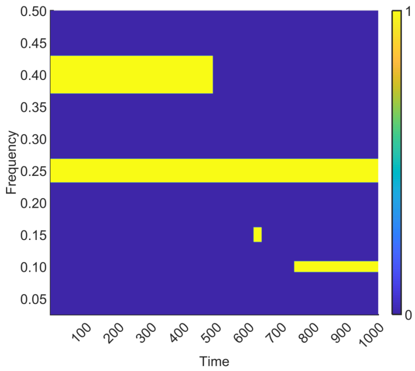


(d) PWCS LAB masking.

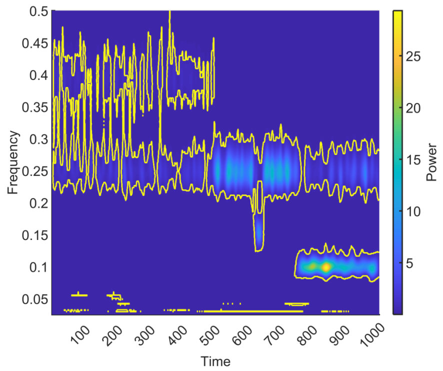
Fig. 4.2: PWCS and its estimate for constant variance and $SNR = 10$.

SNR = 10						
	FN	**M1** [%]	Δ **M1** [%]	**FP**	**M2** [%]	Δ **M2** [%]
HT	$6.44 \cdot 10^3$	12.62	–	$5.83 \cdot 10^4$	13.58	–
SAB	$6.74 \cdot 10^3$	13.31	−0.60	$5.76 \cdot 10^4$	13.42	0.16
LAB	$7.12 \cdot 10^3$	13.95	−1.34	$5.46 \cdot 10^4$	12.72	0.86
SNR = 3.16						
	FN	**M1** [%]	Δ **M1** [%]	**FP**	**M2** [%]	Δ **M2** [%]
HT	$1.25 \cdot 10^4$	32.66	–	$5.25 \cdot 10^4$	13.94	–
SAB	$1.27 \cdot 10^4$	33.10	−0.43	$5.21 \cdot 10^4$	13.80	0.14
LAB	$1.28 \cdot 10^4$	33.60	−0.94	$5.16 \cdot 10^4$	13.66	0.28
SNR = 2						
	FN	**M1** [%]	Δ **M1** [%]	**FP**	**M2** [%]	Δ **M2** [%]
HT	$1.89 \cdot 10^4$	59.13	–	$4.65 \cdot 10^4$	12.15	–
SAB	$1.89 \cdot 10^4$	58.81	0.32	$4.65 \cdot 10^4$	12.14	0.01
LAB	$1.89 \cdot 10^4$	58.93	0.19	$4.61 \cdot 10^4$	12.02	0.13

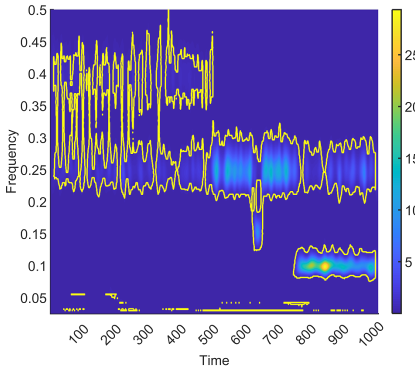
Tab. 4.1: Metrics for constant variance- averages of 1000 MC simulations.



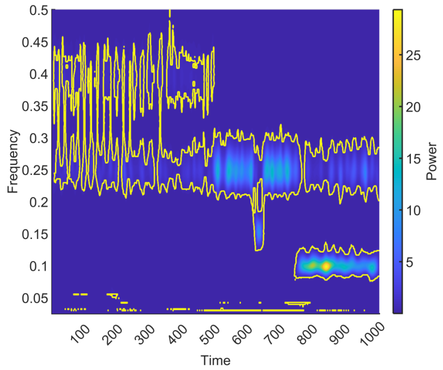
(a) Ideal case.



(b) PWCS HT masking.



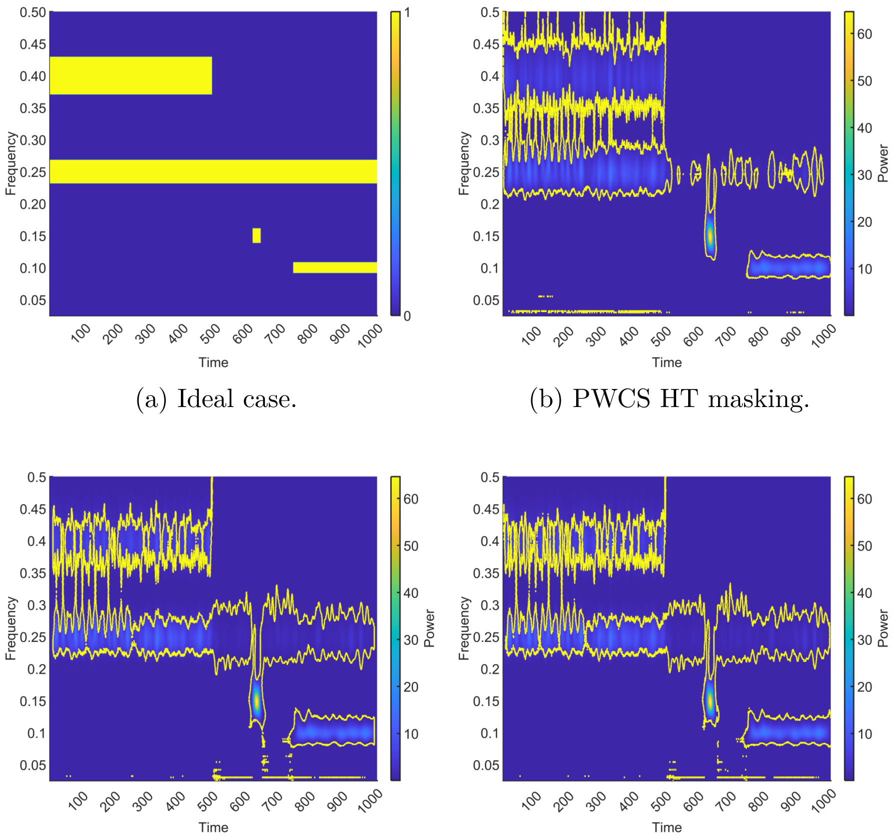
(c) PWCS SAB masking.



(d) PWCS LAB masking.

Fig. 4.3: PWCS and its estimate for constant variance and $SNR = 3.16$.

constant variance. That is, for the $SNR = 10$, the metric $\Delta M1$ is mostly 11.5% higher; the metric $\Delta M2$ is mostly 3% higher. For the $SNR = 3.16$, the metric $\Delta M1$ is mostly 0.5-1% higher; the metric $\Delta M2$ is again mostly 3% higher. We can see that the metric $\Delta M1$ is more sensitive to the noise level. An additional simulation for SNR=2 confirmed the sensitivity of $\Delta M1$ and the slow decrease of $\Delta M2$. That is, the growing noise level causes the increase of false negative components and thus the decrease of the level of improvements measured by $\Delta M1$ in HT, SAB and LAB. The level of $\Delta M2$ keeps a roughly the same level, which means that the level of spurious significance given by HT is corrected. Thus we can conclude that in the case of segmented variance, HT masking can produce worse results and should be replaced by adaptive masking (SAB, LAB). The graphical comparison is visualized in Figs. 4.4 and 4.5.



(a) Ideal case.

(b) PWCS HT masking.

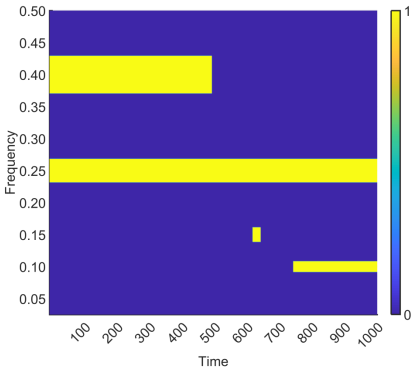
(c) PWCS SAB masking.

(d) PWCS LAB masking.

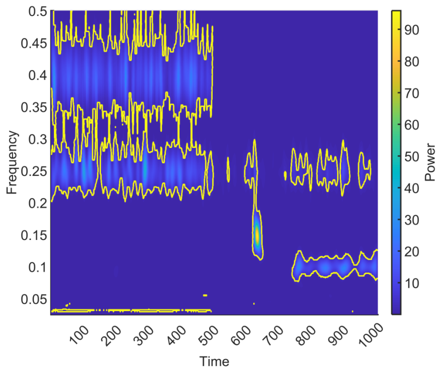
Fig. 4.4: PWCS and its estimate for changing variance and $SNR = 10$.

SNR = 10						
	FN	**M1** [%]	Δ **M1** [%]	**FP**	**M2** [%]	Δ **M2** [%]
HT	$9.01 \cdot 10^3$	21.75	–	$6.49 \cdot 10^4$	17.80	–
SAB	$4.73 \cdot 10^3$	10.24	11.50	$5.53 \cdot 10^4$	14.79	3.00
LAB	$4.70 \cdot 10^3$	10.15	11.60	$5.39 \cdot 10^4$	14.36	3.44
SNR = 3.16						
	FN	**M1** [%]	Δ **M1** [%]	**FP**	**M2** [%]	Δ **M2** [%]
HT	$1.07 \cdot 10^4$	26.60	–	$5.77 \cdot 10^4$	15.54	–
SAB	$1.58 \cdot 10^4$	26.23	0.37	$4.86 \cdot 10^4$	12.77	2.77
LAB	$1.50 \cdot 10^4$	25.85	0.75	$4.71 \cdot 10^4$	12.35	3.19
SNR = 2						
	FN	**M1** [%]	Δ **M1** [%]	**FP**	**M2** [%]	Δ **M2** [%]
HT	$1.49 \cdot 10^4$	41.50	–	$4.85 \cdot 10^4$	12.75	–
SAB	$1.85 \cdot 10^4$	57.09	−15.58	$4.33 \cdot 10^4$	11.23	1.52
LAB	$1.85 \cdot 10^4$	56.87	−15.37	$4.20 \cdot 10^4$	10.84	1.91

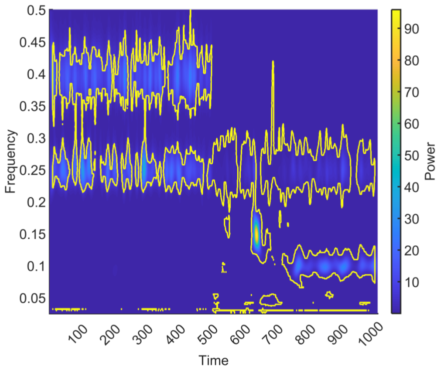
Tab. 4.2: Metrics for changing variance - averages of 1000 MC simulations.



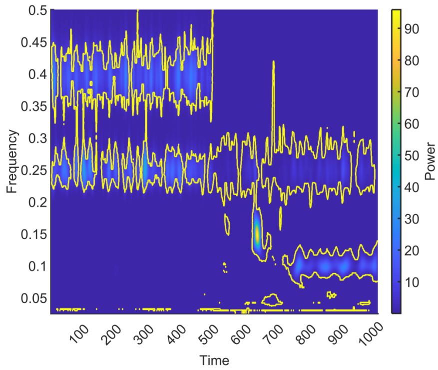
(a) Ideal case.



(b) PWCS HT masking.



(c) PWCS SAB masking.



(d) PWCS LAB masking.

Fig. 4.5: PWCS and its estimate for changing variance and $SNR = 3.16$.

4.5 Chapter Conclusion

To summarize the results from the simulation, we can give the following recommendation. Before the significance testing of co-movement measure, an analyst should identify the behaviour of the time series volatility. If the time series have a constant variance, then HT masking is a plausible instrument. If one time series, or both, indicates a changing variance during the time (heteroscedasticity), then adaptive masking is a proper way how to obtain relevant information. This recommendation may be particularly useful for time series for which heteroskedasticity is expected, as in the case of economic time series.

5 Co-movement Selective Detection Filter

A large number of econometric analysts use filtering when processing data. They are usually interested in a decomposition into the long-term trend and oscillations. After that, the filtered time series is taken as an input to further econometric analyses. Filtering techniques are viewed not only from the perspective of removing the trend component, but also from the perspective of identifying trend-breaks, outliers, or its removing ability [38, 39].

For a long time, the filtering of time series prevailed in the time-domain represented by deterministic [40,41] or stochastic methods [42], or their combinations [43]. The time series processing in the time-domain is simple, but is not able to remove a specific frequency range and, in some cases, is inflexible in the long-term trend modelling. Unfortunately, time domain approaches are weak in capturing a cyclical character of the time series and need parameter optimization. Therefore, analysts began to use methods in the frequency-domain, i.e. low-, high- or band-pass filters [44–46], or a windowed filter as proposed by [47]. Such approaches are more flexible and are widely used in the economic area for the business cycle (BC) analyses [48]. An alternative point of view is filtering via eigenvalue-based decomposition [49–51], which allows a decomposition of the time series into the number of components. Then, some of the components can be removed and the others can be reconstructed into the time domain. The drawback is an arbitrariness of the decision what will be removed. This deficiency can be solved by the TF selective filtering [52].

The use of hard, soft and adaptive threshoding applications in engineering great popularity of wavelets among economists motivates us to improve our earlier study [53]. This previous research employed the hard threshold for the co-movement-selective filter which was applied to filtering out symmetric macroeconomic shocks from individual time series. There, the hard thresholding was based on the analyst's experience, while the current study proposes a more sophisticated approach based on statistical testing. We apply the approach according to Torrence and Compo (TC98) [23] who were the first to propose algorithms for significance testing of power wavelet spectrum, the cross-spectrum and the linear coherence. An improvement of their work was provided by Ge [25, 29]. The algorithm presents several assumptions to test the significance of the power wavelet cross-spectrum. That is, inputs are two independent GWN and thus the power wavelet co-spectrum is the product of two χ^2-distributed random variables. Further, using the Bessel function, we can test whether the power wavelet cross-spectrum coefficients are significant with respect to the variance of each time series. Notice that the authors of the mentioned publications work with a constant variance of input time series.

5.1 The Wavelet Transform, Co-movement Measures and its Testing

The algorithm for the co-movement selective filtering presented here is designed for the WT because of its advantages, especially very good time resolution and usability for non-stationary time series [20, 32, 54], and for the power wavelet co-spectrum, as the co-movement measure [23, 46]. It can be modified for the Short Term Fourier transform and for different co-movement measures such as coherence [23, 25, 29]. As these methods are well known, we do not describe them in this paper. Instead, we focus on the description of testing approaches for the co-movement measure.

5.2 An Algorithm for the Co-movement Selective Detection Filter

An algorithm for the co-movement selective detection filter is based on two processes, i.e. transformation plus analysis and reconstruction. The transformation process consists of the TF modelling (WS and PWCS analyses) and masking. The reconstruction process inversely transforms time series from the TF to the time domain. Figure 5.1 proposes a block diagram describing this algorithm.

5.2.1 An Algorithm for the Co-movement Selective Detection Filter

The proposed method for the identification of the time series co-movement indicator or for filtering out a symmetric behaviour is designed as follows:

1. **Time-frequency transform**
   Transform time series $x(t)$ and $y(t)$ using CWT resulting in wavelet spectrograms $W_x(a,b), W_y(a,b)$ and wavelet cross-spectrum $W_{xy}(a,b)$ respectively. Alternatively, another co-movement measure as coherence could be used.
2. **Decision about type of thresholding**
   Decide the method for identifying the threshold of wavelet cross-spectrum coefficients WCS and find the threshold thr, i.e. decide for the method of the mask M design via standard or adaptive thresholding (see Sec. 5.3).
3. **Mask design**
   Divide the wavelet cross-spectrum in power form, i.e. PWCS $|W_{xy}(a,b)|^2$, into regions with significant and insignificant co-movement of $x(t)$ and $y(t)$ based

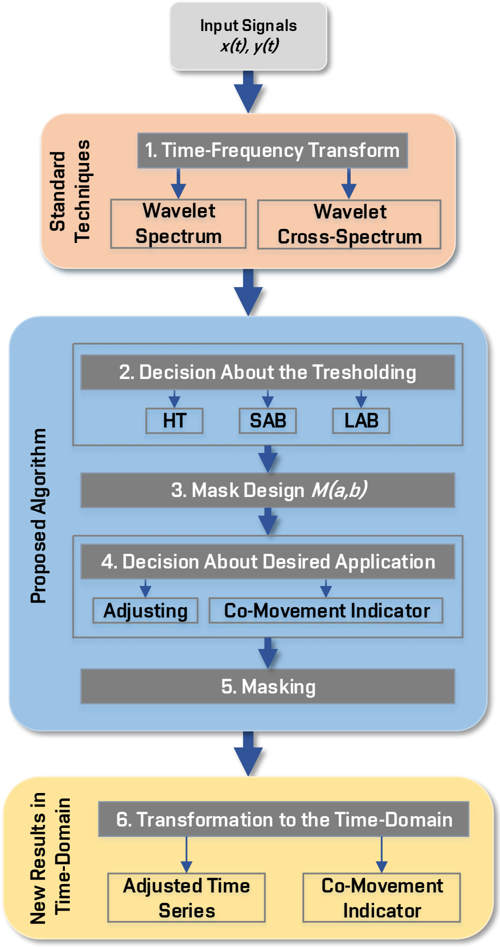


Fig. 5.1: Block diagram of co-movement detection filter algorithm.We use the following abbreviations: WS denotes the wavelet spectrogram coeffcient, PWCS denotes the power wavelet cross-spectrum, HT denotes Hard-Threshold Masking (Sec. 4.1), SAB denotes Segmentation-Adaptive-Based Masking (Sec. 4.2) and LAB denotes Local-Adaptive-Based Masking (Sec. 4.3).

on the threshold thr. That is, we construct the mask M:

$$M(a,b) = \begin{cases} 1 \text{ (significant co-movement)}, & |W_{xy}(a,b)|^2 \geq thr \\ 0 \text{ (insignificant co-movement)}, & |W_{xy}(a,b)|^2 < thr \end{cases} \tag{5.1}$$

The term "significant co-movement" denotes statistically significant PWCS values which will be identified via statistical testing. See Sec. 5.3 for a detailed description of this testing.

The PWCS coefficients are used due to the complex valued cross-spectrum for the majority of practically used mother wavelet functions.

4. **Decision about desired application**

Decide what the desired application is. We investigate two cases: i) if the application is to remove the co-moved part from the time series, or ii) the inverse task, i.e. keeping the co-moved part of the time series and remove the part with distinct cyclical behaviour from the time series.

5. **Masking**
   i) Create a modified wavelet spectrogram (MWS) by masking i.e. co-movement selective detection filtering (adjusted time-frequency transform)
   $$MWS_x(a,b) = (1 - M(a,b)) * W_x(a,b).$$
   Analogously for the time series $y(t)$.
   ii) Create a modified PWCS (MPWCS) by masking, i.e. selecting the distinct cyclical behaviour leading to the time-frequency transform of a co-movement indicator
   $$MPWCS(a,b) = |W_{xy}(a,b)|^2 * M(a,b).$$
6. **Transforming to the time domain**
   Transform inversely product of masking from the previous step via inverse continuous wavelet transform (ICWT). We can obtain: i) adjusted time series in the time domain)
   $$\widetilde{x}_{\text{adj}}(t) = ICWT\{MWS_x(a,b)\}.$$
   Analogously for the time series $y(t)$, we can get $\widetilde{y}_{\text{adj}}(t)$.
   ii) the co-movement indicator represented in the time domain
   $$\widetilde{x}_c(t) = ICWT\{MPWCS(a,b)\}.$$
   That is, we can construct the time representation of the co-movement indicator. Or, we inversely transform the pre-defined frequency region (e.g. a sub-part of MPWCS in BC frequencies) of this product into the time domain to construct the time representation of the co-movement sub-indicator corresponding to the pre-defined frequency region.

5.3 Mask Design

Let us consider the PWCS coefficients for the time series $x(t)$ and $y(t)$. Based on the TC98 significance testing of PWCS we are going to design the mask. We follow two basic approaches. One is based on the hard threshold given by the testing via STA leading to the so called hard-threshold (HT) masking. The other is based on the adaptive threshold identification leading to two possibilities, i.e. local-adaptive-based (LAB) threshold and segmentation-adaptive-based (SAB) threshold. This adaptivity is in time. In both adaptive cases, we propose an improvement of STA in the adaptive form.

If an analyst focuses on the time series adjustment about the co-moved part, or on the construction of the co-movement indicator with respect to the full time range in order to identify the most important events in the time series, we recommend the use of HT masking, i.e. STA.

The idea of SAB and LAB testing considers the situation when the variance of the time series x or/and y in the TD may vary for some sub-period, even for a short duration. In this case, the adaptive masking may be more suitable, because there may exist events (such as the financial crisis in 2008) having a higher level of amplitude in the co-spectrum, which may suppress the significance of other events. These events can be usually visible in the time representation of the data (structural breaks, outlier or cause changes in the volatility of the data).

5.3.1 Real Data Description

To demonstrate the proposed methodology we use the seasonally adjusted monthly data of IPI from the OECD [27] database which are commonly used among economists for business cycle modelling as the macroeconomic indicator of country economy. With respect to the globalisation of economies we focused on the EA and selected G8 countries [27]: the US, Japan, Russia, and the UK. The sample period starts with July 1975 and ends in December 2017 for all countries except Russia. In the case of Russia, the available data are in the range from January 1993 to December 2017. This selection was motivated by the following facts: the US was, for a long time, the leading world economy causing the crisis in 2008; the EA is taken as a representative of 19 European economies; The UK was preselected because of Brexit; Japan is an Asia Pacific representative economy; and Russian is taken as an East European Asia Country. We examine TF selective filtering based on co-movement between the growth cycles of the US and the selected countries. The data were transformed to FODLOG values which represent the growth business cycles [55] of selected countries.

Figure 5.2 displayes these business cycles (data in levels) in the time domain. Further, the data are transformed into the growth business cycles (i.e. fluctuation around a potential product) and are used for the synchrony analysis via wavelets as usual by economists [20, 33, 36, 56, 57].

As a preliminary analysis, assuming the existence of synchrony between the US and selected countries, we calculate the correlation coefficients of business cycles of selected countries. The synchronization among countries during the economic crisis in 2008 is also illustrated in Fig. 5.2. Here we can see the tendency of the curves to

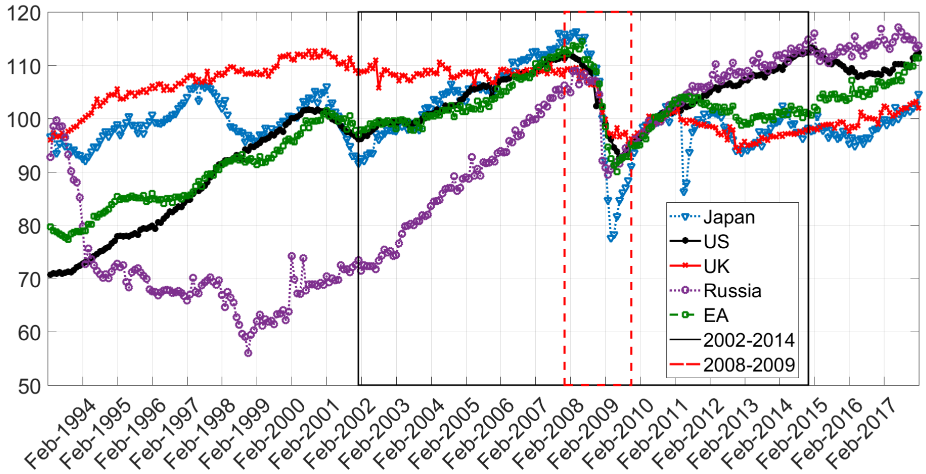


Fig. 5.2: Industrial production index of selected countries (in the levels).

US	Levels			FOD transform		
	1978-2017	1993-2017	2002-2014	1993-2017	2002-2014	2008-2009
Japan	0.7546***	0.3459***	0.6255***	0.1825***	0.1904**	0.3599**
UK	0.7605***	−0.0365	0.0607	0.1390**	0.1540*	0.2131
EA19	0.9721***	0.9312***	0.7268***	0.2963***	0.3732***	0.4090**
Russia		0.6483***	0.6669***	0.0781	0.1621**	0.

Note: statistically significant at: ***1%, **5%, *10%

Tab. 5.1: Correlation coefficients between US and selected countries.

converge especially arround 2008–2009 time window, which is caused by a structural symmetric economic shock, i.e. the global financial crisis. Comparing correlation coefficients in the overview given in Tab. 5.1, we can see the influence of the sample size on levels of the data and on the FOD transform (growth business cycle). We focused on the difference between the time period around the crisis, i.e. 2002–2014, and the available sample size. As we can see from the FOD transform, the selected countries represent situations with a generally stable correlation during the crisis time in 12-year window (Japan, EA), with an increase (Russia) as well as a decrease (UK) in the correlation significance. The table also briefly compares the correlations for the data without any transform (in the levels).

5.3.2 Settings for Implementation

During the analyses, we use the following settings. For the PWCS calculation, we set wavelet scales corresponding to the range of 1 year to 10 years divided into 257 individual scales. We select the complex Morlet wavelet with the center frequency

$f_b = 1.5$ as mother wavelet. For LAB testing we set a sliding window for 3 years, i.e. 36 samples, with 1 sample step ahead. The choice of Morlet wavelet was motivated by the fact that it is a widely used wavelet for the co-movement analysis by economists. For the SAB testing, we will describe the number of segments during the presentation of particular results. After masking the country wavelet spectrogram, we use its inverse transform to obtain filtered time series. In such a way, the inverse transform of the whole wavelet spectrogram works as a band-pass filter with respect to the scales setting.

5.3.3 Demonstration of the Proposed Mask Construction

For the demonstration of the proposed approaches, we choose the US and EA countries. We concentrate on the removal of all co-movements with the US from the EA data. The TF representation re-calculated into the time-scale form is given in the PWCS figures (Fig. 5.3(a),(b),(c)). After the testing of the obtained PWCS via STA (Fig. 5.3a)), via SAB (Fig. 5.3b)) and LAB testing (Fig. 5.3c)) we construct the masks. Then, we partition the TF plane into two regions, with and without significant co-movement. In all figures, the border is highlighted as a yellow solid-line curve. After the EA spectrogram masking, we inversely transform wavelet coefficients which correspond to the part without a co-movement. The obtained time series, for all three masks, are depicted in Fig. 5.4. The dotted black line is the original time series $\widetilde{x}$ (i.e. growth business cycle) obtained via ICWT, the green line represents the adjustment via HT, the red line represents the adjustment via LAB, the blue line represents the adjustment via SAB. To ensure a better visibility of the detected areas, we zoomed the figure to a shorter time range i.e. 2006–2012 (the shape of the curves are the same before 2006 and after 2012). In this way we obtain time series adjusted for significant co-movement parts with the reference (the US) country.

As presented in Fig. 5.3, we identify the mask via three approaches leading to three adjusted time series. The HT masking produce the mask covering cycles of the range 5–0.7 years. The most energy of co-spectrum is concentrated in the cycles 5–1.5 years; thus, the adjustment via the HT masking removes mainly long and business cycles, as well as part of short and very short cycles, from the original EA data. As a result, the EA's time series will reduce the fluctuations in the time around the crisis. In other words, due to the fact that the mutual movement of the countries manifested itself in many different periods then the removal of these components from the time series results in its greater smoothness with respect to the temporal localization of the co-movement. Next, the HT masking is constructed as a selection of a co-movement of countries relative to the full time range. Thus, if a significant event occurs, such as the 2008 crisis (which is reflected by a significantly higher amplitude in the spectral and co-spectral component) and if we evaluate co-movement in the whole time range, the significance of other spectral components will be significantly lower.

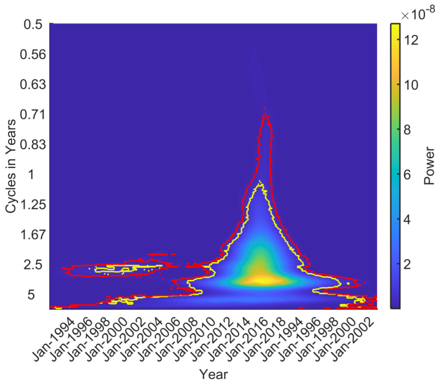
If we use the LAB or SAB masking, we evaluate the significance in the window that adaptively calculates the variance. Thus, it marks the components that contribute to the overall variance as significant. Because of this, some spectral components in other areas are not suppressed, and thus gain in significance. In the case of the US and EA, the LAB and SAB masking concentrates most of the PWCS energy into long cycles and shows that the most important co-movement is in long and business cycles, while the HT masking shows also very short cycles. As the result of the adaptability, the high-frequency components are not removed from the EA, which results in a lower volatility reduction in the crisis period.

To validate the results of SAB, LAB masking, we provide MC simulations (the red line in Fig. 5.3). We can see the proximity of SAB, LAB masking with MC simulations for presented volatile data. Thus, the difference in the significant region identified via STA testing (i.e. HT masking) and MC simulations confirmed the influence of the volatility on the testing. The detailed description of MC simulation can be found in [17, 26].

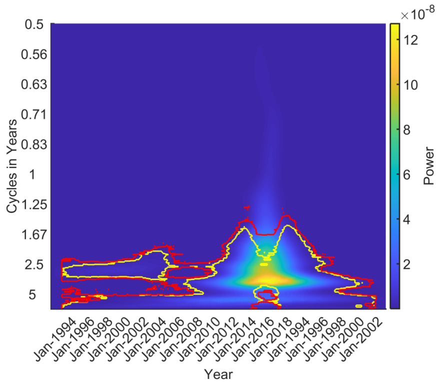
If we assess the EA data, adjusted by the HT masking about the co-movement with the US, then we can see that the fluctuation in industrial production without a linking to the US is smaller. Conversely, the local effect of interdependence during the crisis period (using adaptive masking) results in a greater fall in the index value than in the long-term time horizon (using HT masking). Furthermore, the adjustment of the local co-movement in relation to the unadjusted data points to a larger drop in values, i.e. a larger structural break. Thus, the interdependence with the US economy led to a deepening of the structural breakdown of the crisis (a high correlation has given a reason to believe that there will be a significant reaction and deepening).

5.4 Chapter Conclusion

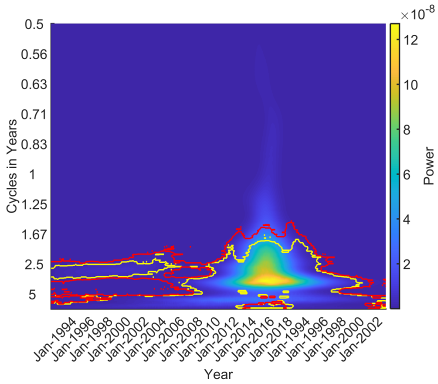
This chapter presents time-frequency selective filtering for the time series adjustment based on the time series co-movement measure. We propose a mask which can be used for selective filtering (adjustment) on statistical basis. The adjustment means removing common components from time series with respect to the reference time series. We investigate two approaches, via hard thresholding based on the χ^2 testing and via adaptive thresholding considering the data character. As the result of the co-movement selective filtering (which includes masking and inverse transform) we



(a) PWCS of EA: HT masking.



(b) PWCS of EA: SAB masking.



(c) PWCS of EA: LAB masking.

Fig. 5.3: Power wavelet co-spectra. The yellow curve denotes the respective mask and the red curve denotes MC simulations.

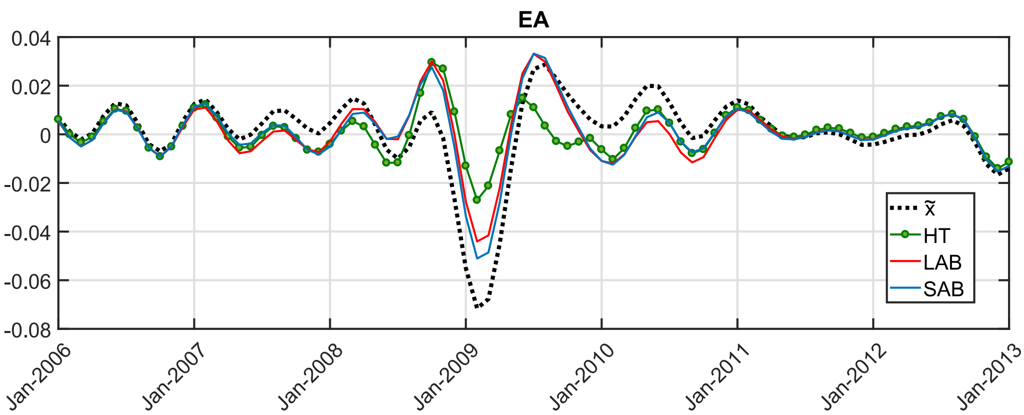


Fig. 5.4: IPI of EA adjusted of co-movement with US.

obtain two time series, i.e. the time series adjusted about the co-moved part and the time series containing just the co-moved part. The adjusted time series can be then used for consequent econometric analyses. The validation of the proposed method is done in MATLAB on the application of symmetric shock removal from selected G8 countries with the US as the reference country.

Considering the type of mask construction, our research leads to the following recommendations. If an analyst focuses on the time series adjustment about the co-moved part with respect to the full time range in order to identify the most important events in the time series, we recommend the use of hard-threshold masking, i.e. STA. If an analyst is interested in the adjustment of economic event with respect to its lead/lag influence, especially in the case when this event causes changes in the data volatility, then the adaptive masking (SAB or LAB) is a valuable instrument. The choice between the LAB and SAB approaches depends on the evolution of data volatility.

The presented approaches are able to provide an in-depth analysis of the time series. This can be done via adjustment for the significant symmetric shocks measured in the TF domain. In this way, we can investigate the global and regional country specific cyclical behaviour. It can be also done via investigating the time series representation of adjusted part as an inverse transform of significant co-movement regions leading to the construction of a co-movement indicator.

As can be seen in literature, a large number of economic studies uses TF domain, especially wavelets, for an individual time series analysis as well as for a co-movement analysis. Thus, the presented approach for the time-frequency selective filtering, or for the construction of the co-movement indicator, can be applicable and can reveal additional information about the investigated problem focusing on adjusted time series.

Conclusion

The doctoral thesis is focused on the current problems and shortcomings of time-frequency analysis and subsequent significance testing. The presented literature review shows current progress and gaps in this field. We found that the literature does not deal with the influence of the data character, i.e. data volatility, on the significance testing of the time-frequency methods. To include and encompass this required data character we selected three types of data. The technical data were chosen as a signal with known parameters, economic data as a signal where factors that are often unpredictable may affect or change the character of the data, and simulated data for verification. Using these types of data we focus on the issue of statistical testing of selected data in order to verify the standard methods and to propose methods for cases where the data volatility is changing over time. Chapters 3–5 represent the core of the dissertation and each of the chapters deals with a corresponding dissertation objective. Additional analyses necessary for the proper design of individual methods are included only in the full version of the dissection.

In Chapter 3, we propose an approach for the enhancement of TF representation leading to background noise suppression. We denote this approach as "Enhanced TF representation". The core of this method is the combination/multiplication of several TF approaches. Thus, based on this, we can easily identify important areas in the TF representation. In specific cases, such as economic data the application of the designed methodology allows a more straightforward interpretation from time and frequency perspectives. Moreover, it can also be taken as a supplement to the significance testing or simulations of background noise levels.

By evaluating standard tests, we find that the standard method may fail in some cases (detailed analyses are included only in the full version of the dissection). In the case of an input signal with strongly localized fluctuations of the signal strength, the total variance may not sufficiently describe the character of the data. In Chapter 4, we propose two modified methods of significance testing. We denote them as segmentation adaptive based (SAB) and local adaptive based (LAB) testing. Both of these methods take into account the possibility of changing the volatility of input data and adapting to it. The SAB method proposes segmentation of the data according to its levels of variance and thus providing better results when the changes in data variance have step character. The LAB method uses a sliding window and is, therefore, better when the variance change is gradual. We also confirm that in the case of different volatility levels in inputs, the significance testing needs a more careful interpretation of the results.

Chapter 5 uses significance tests for subsequent data filtering. We use the statistically significant part of the power wavelet co-spectrum to construct a co-movement selective detection filter suitable for assessing the synchrony between two signals. We propose a mask construction that can be used for selective filtering, i.e. adjustment, on a statistical basis. The adjustment means removing common components from the time series with respect to the reference time series. We investigate approaches based on standard and newly proposed SAB and LAB testing. The advantage of the proposed co-movement selective detection filter is no loss of observations (such as correlation). Moreover, it is possible to construct sub-indicators that correspond to the predefined frequency range. In such a way, we can obtain a decomposition of the (total) co-movement indicator, which covers the full range of frequencies, into the required range.

Bibliography

[1] B. R. Gupta and V. Kumar, “Time-frequency analysis of asymmetric triaxial galaxy model including effect of spherical dark halo component,” *International Journal of Astronomy and Astrophysics*, vol. 05, no. 02, pp. 106–115, 2015. [Online]. Available: https://doi.org/10.4236%2Fijaa.2015.52014

[2] K. Vida, K. Olah, and R. Szabo, “Looking for activity cycles in late-type kepler stars using time-frequency analysis,” *Monthly Notices of the Royal Astronomical Society*, vol. 441, no. 3, pp. 2744–2753, may 2014. [Online]. Available: https://doi.org/10.1093%2Fmnras%2Fstu760

[3] P. Wang, J. Gao, and Z. Wang, “Time-frequency analysis of seismic data using synchrosqueezing transform,” *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 12, pp. 2042–2044, dec 2014. [Online]. Available: https://doi.org/10.1109%2Flgrs.2014.2317578

[4] H. Ding and B. F. Chao, “Detecting harmonic signals in a noisy time-series: the z-domain autoregressive (AR-z) spectrum,” *Geophysical Journal International*, vol. 201, no. 3, pp. 1287–1296, apr 2015. [Online]. Available: https://doi.org/10.1093%2Fgji%2Fggv077

[5] S. Stanković, I. Orović, and V. Sucic, “Averaged multiple l-spectrogram for analysis of noisy nonstationary signals,” *Signal Processing*, vol. 92, no. 12, pp. 3068 – 3074, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0165168412001892

[6] S. Liu, D. Wang, T. Li, G. Chen, Z. Li, and Q. Peng, “Analysis of photonic doppler velocimetry data based on the continuous wavelet transform,” *Review of Scientific Instruments*, vol. 82, no. 2, p. 023103, feb 2011. [Online]. Available: https://doi.org/10.1063%2F1.3534011

[7] O. Faust, U. R. Acharya, H. Adeli, and A. Adeli, “Wavelet-based EEG processing for computer-aided seizure detection and epilepsy diagnosis,” *Seizure*, vol. 26, pp. 56–64, mar 2015. [Online]. Available: https://doi.org/10.1016%2Fj.seizure.2015.01.012

[8] M. R. Canal, “Comparison of wavelet and short time fourier transform methods in the analysis of EMG signals,” *Journal of Medical Systems*, vol. 34, no. 1, pp. 91–94, oct 2010. [Online]. Available: https://doi.org/10.1007%2Fs10916-008-9219-8

[9] Y. Xu, S. Haykin, and R. Racine, “Multiple window time-frequency distribution and coherence of EEG using slepian sequences and hermite functions,” *IEEE Transactions on Biomedical Engineering*, vol. 46, no. 7, pp. 861–866, jul 1999. [Online]. Available: https://doi.org/10.1109%2F10.771197

[10] J. G. Proakis, C. L. Nikias, C. M. Rader, F. Ling, M. Moonen, and I. K. Proudler, *Algorithms for Statistical Signal Processing*, 1st ed. Upper Saddle River, NJ, USA: Prentice Hall PTR, 2001.

[11] D. Walnut, *An Introduction to Wavelet Analysis*, ser. Applied and Numerical Harmonic Analysis. Birkhäuser Boston, 2002. [Online]. Available: https://books.google.cz/books?id=atOStnsp9a0C

[12] H. Resnikoff and R. Wells, *Wavelet Analysis: The Scalable Structure of Information*. Springer New York, 2002. [Online]. Available: https://books.google.cz/books?id=poMzLbTn6JUC

[13] V. Sebesta, R. Marsalek, and J. Pomenkova, “The modified empirical mode decomposition method for analysing the cyclical behavior of time series,” in *ECMS 2013 Proceedings edited by: Webjorn Rekdalsbakken, Robin T. Bye, Houxiang Zhang*. ECMS, may 2013. [Online]. Available: https://doi.org/10.7148%2F2013-0288

[14] K. Gröchenig, *Foundations of Time-Frequency Analysis*. Birkhäuser Boston, 2001. [Online]. Available: https://doi.org/10.1007%2F978-1-4612-0003-1

[15] L. Huang, Q. Kemao, B. Pan, and A. K. Asundi, “Comparison of fourier transform, windowed fourier transform, and wavelet transform methods for phase extraction from a single fringe pattern in fringe projection profilometry,” *Optics and Lasers in Engineering*, vol. 48, no. 2, pp. 141–148, feb 2010. [Online]. Available: https://doi.org/10.1016%2Fj.optlaseng.2009.04.003

[16] J. Zhong and Y. Huang, “Time-frequency representation based on an adaptive short-time fourier transform,” *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5118–5128, oct 2010. [Online]. Available: https://doi.org/10.1109%2Ftsp.2010.2053028

[17] X. Jiang and S. Mahadevan, “Wavelet spectrum analysis approach to model validation of dynamic systems,” *Mechanical Systems and Signal Processing*, vol. 25, no. 2, pp. 575–590, feb 2011. [Online]. Available: https://doi.org/10.1016%2Fj.ymssp.2010.05.012

[18] D. Pollock, D. Pollock, R. Green, and T. Nguyen, *A Handbook of Time-series Analysis, Signal Processing and Dynamics*, ser. Signal processing and its applications. Academic, 1999. [Online]. Available: https://books.google.cz/books?id=sf5xEq1N8MwC

[19] F. Abramovich, Y. Benjamini, D. L. Donoho, and I. M. Johnstone, "Adapting to unknown sparsity by controlling the false discovery rate," *The Annals of Statistics*, vol. 34, no. 2, pp. 584–653, apr 2006. [Online]. Available: https://doi.org/10.1214%2F009053606000000074

[20] C. Aloui, B. Hkiri, and D. K. Nguyen, "Real growth co-movements and business cycle synchronization in the GCC countries: Evidence from time-frequency analysis," *Economic Modelling*, vol. 52, pp. 322–331, jan 2016. [Online]. Available: https://doi.org/10.1016%2Fj.econmod.2015.09.009

[21] J. A. Schulte, C. Duffy, and R. G. Najjar, "Geometric and topological approaches to significance testing in wavelet analysis," *Nonlinear Processes in Geophysics Discussions*, vol. 1, no. 2, pp. 1331–1363, aug 2014. [Online]. Available: https://doi.org/10.5194%2Fnpgd-1-1331-2014

[22] P. M. A. James, R. A. Fleming, and M.-J. Fortin, "Identifying significant scale-specific spatial boundaries using wavelets and null models: spruce budworm defoliation in ontario, canada as a case study," *Landscape Ecology*, vol. 25, no. 6, pp. 873–887, mar 2010. [Online]. Available: https://doi.org/10.1007%2Fs10980-010-9465-2

[23] C. Torrence and G. P. Compo, "A practical guide to wavelet analysis," *Bulletin of the American Meteorological society*, vol. 79, no. 1, pp. 61–78, 1998.

[24] Z. Ge, "Significance tests for the wavelet power and the wavelet power spectrum," *Annales Geophysicae*, vol. 25, no. 11, pp. 2259–2269, nov 2007. [Online]. Available: https://doi.org/10.5194%2Fangeo-25-2259-2007

[25] Z. Ge, "Significance tests for the wavelet cross spectrum and wavelet linear coherence," *Annales Geophysicae*, vol. 26, no. 12, pp. 3819–3829, dec 2008. [Online]. Available: https://doi.org/10.5194%2Fangeo-26-3819-2008

[26] D. Wang, S. Sun, and P. W. Tse, "A general sequential monte carlo method based optimal wavelet filter: A bayesian approach for extracting bearing fault features," *Mechanical Systems and Signal Processing*, vol. 52-53, pp. 293–308, feb 2015. [Online]. Available: https://doi.org/10.1016%2Fj.ymssp.2014.07.005

[27] "Organisation for economic co-operation and development (OECD)." [Online]. Available: http://stats.oecd.org/Index.aspx?DatasetCode=SNA_TABLE1

[28] Z. Ge, "Corrigendum to" significance tests for the wavelet power and the wavelet power spectrum" published in ann. geophys., 25, 2259–2269, 2007," in *Annales Geophysicae*, vol. 31, no. 2. Copernicus GmbH, 2013, pp. 315–315.

[29] ——, "Corrigendum to "significance tests for the wavelet cross spectrum and wavelet linear coherence" published in ann. geophys., 26, 3819–3829, 2008," *Annales Geophysicae*, vol. 31, no. 2, pp. 317–317, feb 2013. [Online]. Available: https://doi.org/10.5194%2Fangeo-31-317-2013

[30] W. T. Wells, R. L. Anderson, and J. W. Cell, "The distribution of the product of two central or non-central chi-square variates," *The Annals of Mathematical Statistics*, vol. 33, no. 3, pp. 1016–1020, sep 1962. [Online]. Available: https://doi.org/10.1214%2Faoms%2F1177704469

[31] D. N. Gujarati, D. C. Porter, and S. Gunasekar, *Basic econometrics.* Tata McGraw-Hill Education, 2012.

[32] A. Bhattacharyya, L. Singh, and R. B. Pachori, "Fourier–bessel series expansion based empirical wavelet transform for analysis of non-stationary signals," *Digital Signal Processing*, vol. 78, pp. 185–196, 2018.

[33] A. N. Berdiev and C.-P. Chang, "Business cycle synchronization in asia-pacific: New evidence from wavelet analysis," *Journal of Asian Economics*, vol. 37, pp. 20–33, apr 2015. [Online]. Available: https://doi.org/10.1016%2Fj.asieco.2015.01.004

[34] Y. Funashima, "The fed-induced political business cycle: Empirical evidence from a time–frequency view," *Economic Modelling*, vol. 54, pp. 402–411, 2016.

[35] C. Su, X. Yin, R. Tao, O.-R. Lobonţ, and N.-C. Moldovan, "Are there significant linkages between two series of housing prices, money supply and short-term international capital?–evidence from china," *Digital Signal Processing*, vol. 83, pp. 148–156, 2018.

[36] A. K. Tiwari, M. I. Mutascu, and C. T. Albulescu, "Continuous wavelet transform and rolling correlation of european stock markets," *International Review of Economics & Finance*, vol. 42, pp. 237–256, mar 2016. [Online]. Available: https://doi.org/10.1016%2Fj.iref.2015.12.002

[37] L. K. Westin. (2015) Receiver operating characteristic (roc) analysis. evaluating discriminance effects among decision support systems. [Online]. Available: http://nutkin.cs.umu.se/research/reports/2001/018/part1.pdf

[38] L. P. Dutrieux, C. C. Jakovac, S. H. Latifah, and L. Kooistra, "Reconstructing land use history from landsat time-series: Case study of a swidden agriculture system in brazil," *International Journal of Applied Earth Observation and Geoinformation*, vol. 47, pp. 112–124, 2016.

[39] R. Metz, "Filter-design and model-based analysis of trends and cycles in the presence of outliers and structural breaks," *Cliometrica*, vol. 4, no. 1, pp. 51–73, 2010.

[40] I. Korhonen and A. Peresetsky, "What influences stock market behavior in russia and other emerging countries?" *Emerging Markets Finance and Trade*, vol. 52, no. 5, pp. 1210–1225, sep 2015. [Online]. Available: https://doi.org/10.1080%2F1540496x.2015.1037200

[41] L. Fredette, J. T. Dreyer, T. E. Rook, and R. Singh, "Harmonic amplitude dependent dynamic stiffness of hydraulic bushings: Alternate nonlinear models and experimental validation," *Mechanical Systems and Signal Processing*, vol. 75, pp. 589–606, 2016.

[42] T. Proietti, "The multistep beveridge–nelson decomposition," *Econometric Reviews*, vol. 35, no. 3, pp. 373–395, 2016.

[43] E. N. Salachas, N. T. Laopodis, and G. P. Kouretas, "The bank-lending channel and monetary policy during pre-and post-2007 crisis," *Journal of International Financial Markets, Institutions and Money*, vol. 47, pp. 176–187, 2017.

[44] J. Valle e Azevedo and A. Pereira, "Macroeconomic forecasting using low-frequency filters," *Oxford Bulletin of Economics and Statistics*, vol. 80, no. 1, pp. 39–64, 2018.

[45] D. S. G. Pollock, "Econometric filters," *Computational Economics*, vol. 48, no. 4, pp. 669–691, 2016.

[46] C. Croux, M. Forni, and L. Reichlin, "A measure of comovement for economic variables: Theory and empirics," *Review of Economics and Statistics*, vol. 83, no. 2, pp. 232–241, may 2001. [Online]. Available: https://doi.org/10.1162%2F00346530151143770

[47] A. Iacobucci and A. Noullez, "A frequency selective filter for short-length time series," *Computational economics*, vol. 25, no. 1-2, pp. 75–102, 2005.

[48] A. F. Burns and W. C. Mitchell, *Measuring business cycles.* National bureau of economic research, 1946, no. burn46-1.

[49] K. Chowdhary and H. N. Najm, "Bayesian estimation of karhunen–loeve expansions; a random subspace approach," *Journal of Computational Physics*, vol. 319, pp. 280–293, 2016.

[50] M. de Carvalho, P. C. Rodrigues, and A. Rua, "Tracking the US business cycle with a singular spectrum analysis," *Economics Letters*, vol. 114, no. 1, pp. 32–35, jan 2012. [Online]. Available: https://doi.org/10.1016%2Fj.econlet.2011.09.007

[51] M. C. Leles, J. P. H. Sansão, L. A. Mozelli, and H. N. Guimarães, "Improving reconstruction of time-series based in singular spectrum analysis: A segmentation approach," *Digital Signal Processing*, vol. 77, pp. 63–76, 2018.

[52] W.-H. Yang, S. H. Holan, C. K. Wikle *et al.*, "Bayesian lattice filters for time-varying autoregression and time–frequency analysis," *Bayesian Analysis*, vol. 11, no. 4, pp. 977–1003, 2016.

[53] R. Marsalek, J. Pomenkova, and S. Kapounek, "A wavelet-based approach to filter out symmetric macroeconomic shocks," *Computational Economics*, vol. 44, no. 4, pp. 477–488, oct 2013. [Online]. Available: https://doi.org/10.1007%2Fs10614-013-9403-x

[54] D. F. Walnut, *An introduction to wavelet analysis.* Springer Science & Business Media, 2013.

[55] J. Bonenkamp, J. Jacobs, G. H. Kuper *et al.*, *Measuring business cycles in The Netherlands, 1815-1913: a comparison of business cycle dating methods.* Graduate School/Research Institute Systems, Organisation and Management, 2001.

[56] Z. Ftiti, A. Tiwari, A. Belanès, and K. Guesmi, "Tests of financial market contagion: Evolutionary cospectral analysis versus wavelet analysis," *Computational Economics*, vol. 46, no. 4, pp. 575–611, sep 2014. [Online]. Available: https://doi.org/10.1007%2Fs10614-014-9461-8

[57] L. Aguiar-Conraria, P. Brinca, H. Gudjonsson, and J. Soares, "Optimal currency area and business cycle synchronization across us states." 2015.

List of Abbreviations

BC	Business Cycle
CWT	Continuous Wavelet Analysis
FODLOG	First-Order Difference of Natural Logarithms
GDP	Gross Domestic Product
GWN	Gaussian White Noise
HT	Hard Threshold
IPI	Industrial Production Index
LAB	Local Adaptive Based Testing
MSE	Mean Square Error
PDV	Photonic Doppler Velocimetry
PWCS	Power Wavelet Cross-Spectrum
SAB	Segmentation Adaptive Based Testing
SNR	Signal-to-Noise Ratio
STA	Standard Testing Approach
STFT	Short Time Fourier Transformation
TF	Time-Frequency
TFA	Time-Frequency Analysis
TFAR	Time-Frequency Varying Autoregressive Process