Charles University in Prague

Faculty of Social Sciences
Institute of Economic Studies

MASTER'S THESIS

Application of band spectrum regression in economic problems

Author: **Bc. Andrej Zubaľ**

Supervisor: **PhDr. Jozef Baruník, Ph.D.**

Academic Year: **2014/2015**

Declaration of Authorship

The author hereby declares that he compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain a different or the same degree.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis document in whole or in part.

Prague, May 15, 2015

Signature

Acknowledgments

The author is grateful to his supervisor for introducing him to this interesting topic and always providing prompt and helpful assistance.

Abstract

In recent years, there has been a rise of interest in the use of various spectral methods in economics and econometrics. These methods have their theoretical background in mathematics, particularly in Fourier analysis. The less traditional and relatively new branch of methods stems from the so-called wavelet analysis. Wavelet methods are believed to have a wide applicability in the analysis of economic time series. The motivation for this thesis is to introduce these methods and apply them in the analysis of economic problems, thereby showing their usefulness within the economic context. Particular attention is paid to band spectrum regression, which allows for decomposition of economic relationships into different frequency components. In this work, we use wavelet band spectrum regression, among other wavelet methods, to analyze the relationship between realized and implied volatilities for the price of crude oil. Second application is from the field of macroeconomics. We analyze the relationship between unemployment and labor productivity growth for four major European economies.

JEL Classification	C01, C02, C58, C69,
Keywords	Wavelets, Band Spectrum Regression, DWT, Applications in Economics

Author's e-mail	`andrej.zubal@gmail.com`
Supervisor's e-mail	`barunik@fsv.cuni.cz`

Abstrakt

V posledných rokoch došlo k nárastu záujmu o používanie rôznych spektrálnych metód v ekonómii a ekonometrii. Tieto metódy majú svoje teoretické zázemie v matematike, najmä vo Fourierovej analýze. Menej tradičné a relatívne nové metódy vychádzajú z takzvanej vlnkovej (waveletovej) analýzy. Vlnkové metódy majú širokú použiteľnosť pri analýze ekonomických časových radov. Motiváciou pre túto prácu je predstaviť vlnkové metódy a aplikovať ich v analýze ekonomických problémov, dokazujúc ich užitočnosť v ekonomickom kontexte. Zvláštna pozornosť je venovaná takzvanej pásovej spektrálnej regresii (band spectrum regression), ktorá nám umožňuje rozložiť ekonomické vzťahy do rôznych frekvenčných komponentov. V tejto práci, pomocou vlnkovej pásovej spektrálnej regresie ale aj iných vlnkových metód, najprv skúmame vzťah medzi zrealizovanou a implikovanou volatilitou pre cenu ropy. Druhá aplikácia je makroekonomického charakteru. Analyzujeme vzťah medzi nezamestnanosťou a rastom pracovnej produktivity pre štyri veľké európske ekonomiky.

Klasifikace JEL	C01, C02, C58, C69,
Klíčová slova	Vlnky, Spektrálna regresia, DWT, Aplikácie v ekonómii

E-mail autora	andrej.zubal@gmail.com
E-mail vedoucího práce	barunik@fsv.cuni.cz

Contents

List of Tables

List of Figures

Acronyms

BPV Bipower Variation

BSR Band Spectrum Regression

CBOE Chicago Board Options Exchange

CWT Continuous Wavelet Transform

DWT Discrete Wavelet Transform

JWTSRV Jump Wavelet Two Scale Realized Variance

MODWT Maximum Overlap Discrete Wavelet Transform

OLS Ordinary Least Squares

RV Realized Variance

TSRV Two Scale Realized Variance

WBSR Wavelet Band Spectrum Regression

Master's Thesis Proposal

Author	Bc. Andrej Zubaľ
Supervisor	PhDr. Jozef Baruník, Ph.D.
Proposed topic	Application of band spectrum regression in economic problems

Topic characteristics I will try to use the so-called band spectrum regression (possibly among other wavelet methods). This is a method introduced by Engle (1972). The basic idea is to estimate 'traditional' linear regression coefficients by calculating spectra and co-spectra of corresponding time series using wavelet decompositions of the series. The nature of the wavelet transforms then allows us to look at the underlying relationships from the perspective of different time scales e.g. to distinguish short-term fluctuations from long-term trends etc. This property can be very useful within the economic context.

In Baruník & Baruníková (2013), the authors found band spectrum regression to be useful in modelling the relationship between implied and realized volatility of assets. In this thesis, I would like to apply their methods and findings in the analysis of some concrete market e.g the market for crude oil for which the data on implied volatility is easily available. Focus will be on discovering long-term relationships on the crude oil market. Secondly, I will briefly analyze the relationship between unemployment and labor productivity growth for four major European economies using wavelet methods.

Hypotheses 1. Hypothesis: The oil VIX index is an unbiased estimator of realized volatility of crude oil.

2. Hypothesis: The oil VIX index shows significant correlation with realized volatility only for time horizons of 32 days and more like it was in Baruník & Baruníková (2013).

3. Hypothesis: Short-term, medium-term and long-term relationship be-

tween labor productivity growth and unemployment in Europe since the 90's is the same for France, Germany, Italy and UK.

Methodology The first step will be to use the established literature on the wavelet methods such as Percival & Walden (2000) or Gençay *et al.* (2009) to get a firm understanding of them. The theory should be briefly outlined in the thesis with not too much focus on technical details. The main ideas and mechanisms behind the wavelet methods (especially band spectrum regression) will be explained.

Next I shall focus on the application of the model described in Baruník & Baruníková (2013). For this purpose I shall choose asset/commodity (most probably crude oil) for which the so-called VIX index is being listed on the stock markets. This is reasonable because the VIX index is thought to be a good measure of implied volatility. With the data on implied volatility it will be possible to estimate the model for the realized-implied volatility relationship. Furthermore, with the wavelet methodology it will be possible to decompose this relationship into different time scales, hopefully yielding some interesting insights. Comparison of the wavelet estimation to different estimations of realized volatility shall also be done. The methodology for the testing of the third hypothesis should also be band spectrum regression with looking specifically at the short-term, medium-term and long-term relationships.

Outline

1. Introduction: Brief outline of the topic and structure of the thesis.
2. Theoretical part: Main ideas behind the methods will be outlined here. Focus will be on ideas and mechanisms.
3. Applications: Basically two parts. One should be the analysis of volatility on the crude oil market. Second one the analysis of unemployment and labor productivity growth. Possibly some more applications.
4. Concluding remarks: I will summarize my findings and their implications for future research.

Core bibliography

1. Baruník, J. & M. Baruníková (2013): "Revisiting the long memory dynamics of implied-realized volatility relation:A new evidence from wavelet band spectrum regression." *IES working paper 2013.*

2. Engle, R. (1972): "Band Spectrum Regressions." *Working papers 96*,Massachusetts Institute of Technology (MIT), Department of Economics, 1972

3. Gençay, R., F. Selçuk, B. Whitcher, (2002): *"An Introduction to Wavelets and Other Filtering Methods in Finance and Economics"* San Diego, California; London: Academic Press, 2002.

4. Percival, D. B.,& A. T. Walden (2000): "Wavelet Methods for Time Series Analysis" , Cambridge: Cambridge University Press, 2000.

Chapter 1

Introduction

When studying economic relationships we often encounter the need to "break them up" into different time horizons. After all, economic literature is frequently using terms such as the 'short-run', the 'medium-run' etc. Suitable analytical tools for this "breaking up" of economic relationships are the various methods known collectively as the spectral methods. They enable us to look at the given phenomena from the perspective of different frequencies. Spectral methods were first used in economics in around the 1960's in the works of notable economists such as Granger, Morgenstern etc.[1] Later, they have been overshadowed by more traditional time domain methods, but are becoming increasingly popular again. Potentially useful method is the so-called band spectrum regression, which is a spectral modification of the classical linear regression. It was first introduced into economics by Engle (1972), who used it to study the relationship between consumption and income. It allows us to conduct regression analysis on a scale-by-scale basis, providing an excellent tool for decomposition of economic relationships into different time horizons (scales). Spectral methods have been traditionally based upon various Fourier methods, especially the discrete Fourier transform. Nowadays, a new branch of spectral methods (called wavelet methods) is becoming increasingly popular in the academic world.[2] The main advantage of wavelets is that they provide localized frequency content of the time series instead of the global picture one gets when using Fourier methods. This localization makes wavelets especially useful when dealing with non-stationary time series (i.e. time series that are not behaving uniformly over time), which are not uncommon in economics.

[1] See, for example, Granger & Morgenstern (1963) or Granger *et al.* (1964).

[2] See Gallegati & Semmler (2014) for many examples in the economic context.

Recently, Baruník & Baruníková (2013) introduced a wavelet-based band spectrum regression that we intend to use in the thesis.

In this thesis, we shall provide basic theoretical background to wavelet spectral methods including wavelet transforms and the wavelet band spectrum regression. The main objective is to demonstrate the usefulness of these methods in the analysis of concrete economic time series. Applications of these methods will include the analysis of the relationship between realized and implied volatility on the crude oil market as well as the analysis of the relationship between unemployment and labor productivity growth for four major European economies.

The thesis has a fairly simple layout and is structured as follows. In Chapter 2, there will be basic theoretical outline of the 'arsenal' that will be used in subsequent chapters. We should gradually define what we mean by the term wavelet, various wavelet transforms, wavelet variance/covariance and also wavelet band spectrum regression. The emphasis should be put on basic ideas and intuition rather than full rigor. In Chapter 3, we shall use the techniques outlined in Chapter 2 to analyze economic time series. For this purpose, we will analyze crude oil market, for which the so-called VIX index is being listed on the stock markets. This is reasonable because the VIX index is thought to be a good measure of implied volatility. With the data on implied volatility it will be possible to estimate the model for the realized-implied volatility relationship. Furthermore, with the wavelet methodology we will decompose this relationship into different time scales, yielding some interesting insights. Next, we will analyze the relationship between unemployment and labor productivity growth for Germany, United Kingdom, France and Italy. The focus will be again on breaking down the relationship into different time scales. In Chapter 4, we will simply summarize our thesis and provide concluding remarks to our findings.

Chapter 2

Theoretical Background

2.1 Introduction to Wavelets

Wavelets or 'small waves' can be thought of (somewhat simplistically) as locally oscillating functions on a real line. The counterpart to them are globally oscillating functions or 'big waves', for example the sine and cosine functions. This notion of local oscillation is too vague for analytical purposes, therefore we need more rigorous definition of wavelets. Below we will provide standard mathematical definition of wavelets and also a few examples of them.

Definition 2.1 (Wavelet). *A* ***wavelet*** *is any real-valued*[1] *function* $\boldsymbol{w}$ *defined on a real line satisfying two conditions:*

1. *The oscillation condition i.e. the integral of* $\boldsymbol{w}$ *is equal to zero:*

$$\int_{-\infty}^{\infty} \boldsymbol{w}(x)dx = 0 \tag{2.1}$$

2. *The finite energy condition:*

$$\int_{-\infty}^{\infty} |\boldsymbol{w}(x)|^2 dx = 1 \tag{2.2}$$

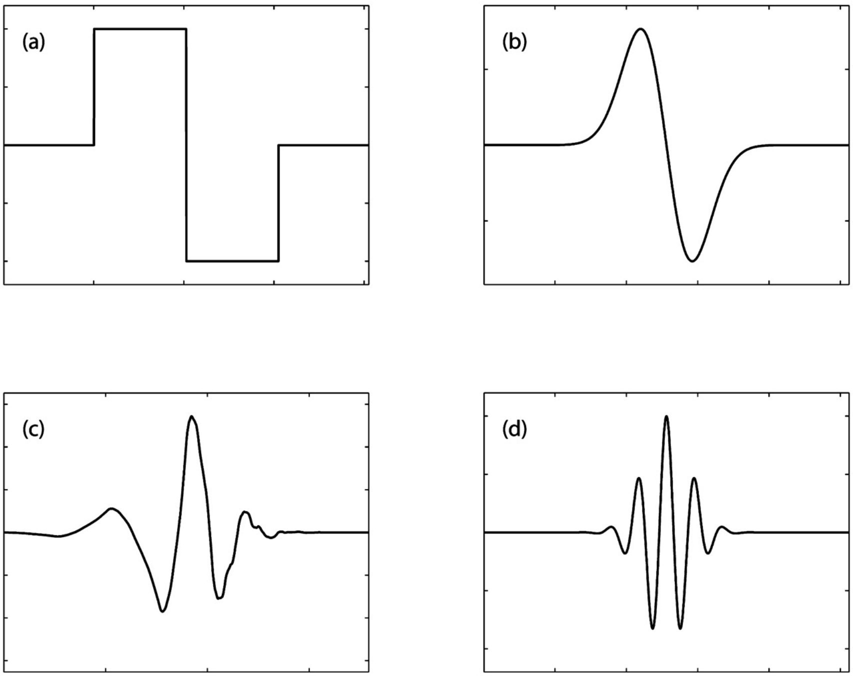
Equation (2.1) is intuitively telling us that $\boldsymbol{w}$ needs to be, on average, evenly distributed between both positive and negative y-values, a condition any oscillating function should satisfy. On the other hand, equation (2.2) can be intuitively interpreted as restricting $\boldsymbol{w}$ from oscillating too much, while at the same time ensuring that $\boldsymbol{w}$ is not a constant zero. Equation (2.2) precisely

[1] Wavelets can actually be complex-valued as well. The definition would not change.

captures the local nature of a wavelet. In fact, as a consequence of this equation, $\boldsymbol{w}$ needs to be practically equal to zero outside of some fixed interval [-T,T]. In other words, the non-zero activity of $\boldsymbol{w}$ is locally confined to some finite interval.

Needless to say, there is an infinite number of functions satisfying definition 2.1. Here we shall only mention a few examples of wavelets for illustrative purposes.

Figure 2.1: Examples of wavelets



In figure 2.1 there are four basic examples of wavelets. As we can see, all of them are locally oscillating functions. In figure 2.1 a) there is the simplest possible wavelet called the Haar wavelet. In figure 2.1 b) we have a wavelet that is formed by differentiating the bell-shaped Gaussian normal density function. The rest of the pictures show more complicated wavelets.

One should notice that the axes are not depicted in figure 2.1. This is not a coincidence. As it turns out, what fully characterizes a wavelet is its basic shape. In fact, we can take any wavelet and shift it and/or stretch it in a certain way with the resulting function still being a wavelet. To be more specific, if $w(u) = w_{1,0}(u)$ is any wavelet function satisfying equations (2.1) and (2.2), then the function defined as

$$w_{\lambda,t}(u) = \frac{1}{\sqrt{\lambda}} w(\frac{u-t}{\lambda})$$

is also satisfying equations (2.1) and (2.2) for any choice of λ and t, therefore it is also a wavelet.[2] Hence, by defining a mother wavelet $w_{1,0}(u)$ we automatically have access to a whole family of wavelets $w_{\lambda,t}(u)$. Parameter t is called the time parameter and is in essence a shifting parameter determining the position of the wavelet , whereas parameter λ is called the scale parameter and is essentially a stretching parameter determining the width of a wavelet.

Mother wavelets are usually positioned at the origin and have a scale equal to 1. In our examples from figure 2.1 a) and 2.1 b) the mother wavelets are defined as follows.

$$w_{1,0}^{(H)} \equiv \begin{cases} 1/\sqrt{2} & if -1 < u \leq 0 \\ -1/\sqrt{2} & if \quad 0 < u \leq 1 \\ 0 & otherwise \end{cases}$$

is the mother Haar wavelet, and

$$w_{1,0}^{(fdG)} \equiv -\frac{\sqrt{2}ue^{\frac{-u^2}{2\sigma^2}}}{\sigma^{3/2}\pi^{1/4}}$$

is the mother 'first-differentiated Gaussian' wavelet. Choice of a proper mother wavelet largely depends on the nature of analysis we want to undertake.

2.2 Continuous Wavelet Transform

So far we have only defined wavelets, but we have not seen what they are good for. Without a doubt, most applications of wavelets have their roots in the Continuous Wavelet Transform (CWT). The CWT, in its generality, was originally designed to overcome some shortcomings of the more traditional Fourier transform. The main problem was that the Fourier transform provides us only with information concerning whether a frequency is present in a signal/function, it does not provide information about where it is present in the signal i.e. it does not give any information about the location of a frequency. This might be a problem if a signal is not stationary and has different frequency content at different time periods. The local nature of wavelets serves very well to overcome this shortcoming. An analysis based on the use of wavelets does not only tell

[2] This fact can be easily checked by methods of substitution in the integrals defining a wavelet.

us about the frequency content of a signal,[3] via the time parameter t it also provides us with information about the location of each frequency. Wavelet analysis of a continuous signal/function is essentially done through the CWT.

Definition 2.2 (Continuous Wavelet Transform). *Given a real function $f(u)$ and a mother wavelet $w(u) = w_{1,0}(u)$ we define the Continuous Wavelet Transform (provided it is well-defined) to be the function of scale λ and time t as follows:*

$$W(\lambda, t) = \int_{-\infty}^{\infty} w_{\lambda,t}(u) f(u) du. \tag{2.3}$$

It is useful to note that the CWT is defined as a scalar product of two functions. Perhaps a good intuitive interpretation of the CWT is that with the help of wavelets we examine our signal/function $f(u)$ at different times (time parameter) looking for different frequencies (represented by the scale parameter). The value of the CWT then tells us how strong is the presence of a particular frequency at a particular time (relative to other times and frequencies). However, the CWT can have different interpretations, also depending on the particular mother wavelet that is used.[4]

For a transform such as the CWT to be truly useful, the existence of an inverse transform is usually desired. This ensures the possibility of recovering original data from the transformed data. In the case of the CWT, there is one more fundamental requirement on the wavelet functions that needs to be satisfied for the inverse transform to exist.

Definition 2.3 (Admissibility condition). *A wavelet $w(u)$ satisfies the admissibility condition, if for its Fourier transform defined as*

$$F(z) = \int_{-\infty}^{\infty} w(u) e^{-i2\pi zu} du \tag{2.4}$$

the number

$$C_w = \int_{0}^{\infty} \frac{|F(z)|^2}{z} dz \tag{2.5}$$

belongs to the interval $(0, \infty)$.

[3]Frequency is in fact captured in the scale parameter λ. The higher the scale λ the more stretched out the wavelets are, therefore we can think of them as representing lower frequencies. In this sense, there is an inverse relationship between the notion of scale and frequency.

[4]See Percival & Walden (2000) for interpretation using weighted averages.

With this condition in place, the following fundamental theorem about the CWT holds.

Proposition 2.1. *If the mother wavelet $w(u) = w_{1,0}(u)$ satisfies the admissibility condition and if the function $f(t)$ is such that*

$$\int_{-\infty}^{\infty} f^2(t)dt < \infty,$$

then there exists an inverse CWT that is defined as

$$f(t) = \frac{1}{C_w}\int_0^{\infty}\int_{-\infty}^{\infty} W(\lambda,t)\frac{1}{\sqrt{\lambda}}w\left(\tfrac{t-u}{\lambda}\right)du\frac{d\lambda}{\lambda^2}. \tag{2.6}$$

Furthermore, we also have

$$\int_{-\infty}^{\infty} f^2(t)dt = \frac{1}{C_w}\int_0^{\infty}\int_{-\infty}^{\infty} W^2(\lambda,t)dt\frac{d\lambda}{\lambda^2}. \tag{2.7}$$

Proof. The proof can be found in Mallat (1999), pp. 122. □

Equation (2.7) is particularly useful because it provides easy interpretation of the CWT. We can see from this equation that $W(\lambda, t)$ essentially decomposes the energy present in the original signal into different times and scales. This is in fact very similar to the Parseval's equality in Fourier analysis, which is also a very useful result used in band spectrum regressions.

The CWT has a wide spectrum of applications. It can be found in fields such as image compression, acoustics processing, filter design etc. Concrete examples can be found in almost any standard textbook dealing with the CWT and wavelet analysis.[5]

2.3 Discrete Wavelet Transforms

Although the CWT turns out to be a useful analytical tool, it is rarely used in its pure form defined in the previous section. One of its shortcomings is that it transforms a function of one variable into a function of two variables. In this sense it is largely redundant i.e. provides unnecessary information. Furthermore, in real-life applications we rarely deal with continuous signals, especially in economics where we mostly work with discrete time series. Also, modern computers can store only a finite amount of information, thus cannot

[5] For example, see Percival & Walden (2000) or Mallat (1999).

really work with ideal continuous functions. All of these reasons call for some sort of discrete wavelet transform that is more compact and succinct than the CWT and, of course, can be easily applied and computed on discrete time series.

In this section, two commonly used discrete wavelet transforms will be introduced. These are the Discrete Wavelet Transform (DWT) and the Maximum Overlap Discrete Wavelet Transform (MODWT). The theory behind these transforms is fairly extensive and is beyond the scope of this work. For this reason, only the main features of these transforms will be outlined. For more complete treatment one should confront Percival & Walden (2000) or Gençay *et al.* (2009).

2.3.1 The Discrete Wavelet Transform

The DWT is the basic tool for discrete wavelet analysis and plays a similar role to the Discrete Fourier Transform in Fourier analysis. It is usually defined only for time series of length $N = 2^J$ where J is a positive integer. It is always constructed in such a way, that it is equivalent to a multiplication by an orthogonal square $N \times N$ matrix Q.[6] If we denote our time series as $\mathbf{X} = (\mathbf{X_1}, \mathbf{X_2}, \ldots, \mathbf{X_N})$, then we can express the DWT in matrix form as

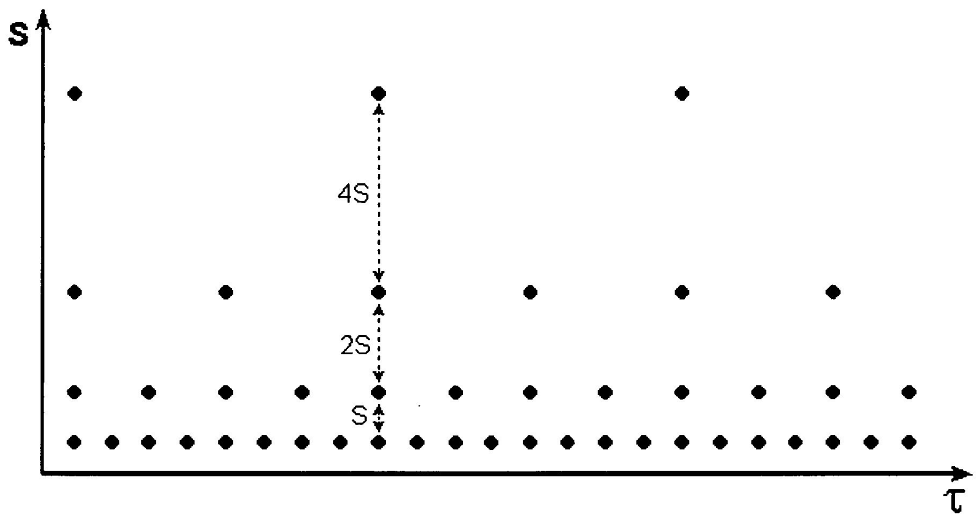
$$\mathbf{W} = Q\mathbf{X} \tag{2.8}$$

From here, $\mathbf{W} = (\mathbf{W_1}, \mathbf{W_2}, \ldots, \mathbf{W_N})$ where $\mathbf{W_i}$ is the i-th wavelet coefficient. The exact content of the orthogonal transform matrix Q is determined essentially by the mother wavelet (or mother wavelet filter) we want to use for the analysis. Each row of the matrix Q is in fact a discretized version of some wavelet $w_{\lambda,t}$.

Qualitatively, the DWT is an approximation to the CWT only at certain points i.e. at certain scales λ and certain times t. The scale and time grid for the DWT is depicted in figure 2.2.

[6] Orthogonal matrix is such that $Q^T Q = QQ^T = I$.

Figure 2.2: The DWT grid for approximation of the CWT



Source: http://www.google.com/patents/US20050069197

Here **S** denotes the scale (λ) and τ the time (t). This is the so-called dyadic grid. Lowest scales are represented very densely. As we double the scales, the number of approximations decreases by one half. This is logical in a way. We can think of it as if we were looking at a map. The higher the scale of a map, the less images we need to cover certain area at a particular scale. Each point in the grid corresponds to one DWT coefficient $\mathbf{W_i}$, which is essentially an approximation of the CWT at that point. The important thing is that, despite approximating the CWT at such a sparse grid as the one depicted in figure 2.2, there is essentially no information loss because the inverse DWT is very simply defined by matrix equation

$$\mathbf{X} = Q^{\mathbf{T}}\mathbf{W}. \tag{2.9}$$

This is a simple consequence of multiplying equation (2.8) by Q^T from the left. In this way, the DWT disposes of the redundancy problem of the CWT. Additionally, there are many more convenient properties of the DWT that are ensured by the way DWT is constructed. We shall mention those properties that are important to us, as well as some drawbacks, but not go into the details of precisely constructing the DWT.

It follows from the orthogonality of matrix Q that

$$\| \mathbf{W} \|^{\mathbf{2}} = \| \mathbf{X} \|^{\mathbf{2}} .$$

This is again a modification of the Parseval's identity and often has useful interpretations. The DWT with $N = 2^J$ is constructed in such a way that the first $N/2$ coefficients belong to the lowest scale, the next $N/2^2$ coefficients belong to the second lowest scale and so on until only two last coefficients are left. The second to last coefficient belongs to the largest scale and the last coefficient is usually just the sample mean of the time series multiplied by $\sqrt{N}$. This presupposes a natural grouping of the DWT coefficients into J+1 groups

$$\bar{\mathbf{W}}_{\mathbf{1}}, \bar{\mathbf{W}}_{\mathbf{2}}, \ldots, \bar{\mathbf{W}}_{\mathbf{J}}, \bar{\mathbf{V}}_{\mathbf{J}},$$

where $\bar{\mathbf{W}}_{\mathbf{j}}$ subvector contains all the coefficients belonging to the j-th lowest scale. $\bar{\mathbf{W}}_{\mathbf{1}}$ contains $N/2$ elements and each subsequent $\bar{\mathbf{W}}_{\mathbf{j}}$ group has only half elements of the preceding group. $\bar{\mathbf{V}}_{\mathbf{J}}$ is simply the sample mean multiplied by $\sqrt{N}$ i.e. $\bar{\mathbf{V}}_{\mathbf{J}} = \sqrt{\mathbf{N}}\,\overline{\mathbf{X}}$. The DFT is then equal to

$$\mathbf{W} = Q\mathbf{X} = \begin{pmatrix} \bar{\mathbf{W}}_{\mathbf{1}} \\ \bar{\mathbf{W}}_{\mathbf{2}} \\ \vdots \\ \bar{\mathbf{W}}_{\mathbf{J}} \\ \bar{\mathbf{V}}_{\mathbf{J}} \end{pmatrix} . \tag{2.10}$$

It turns out this is a very useful representation since now the sample variance of the time series can be expressed in the following way [7]

$$\hat{\sigma}_X^2 = \frac{1}{N} \sum_{j=1}^{J} \| \bar{\mathbf{W}}_{\mathbf{j}} \|^{\mathbf{2}} \tag{2.11}$$

This equation provides a natural analysis of sample variance into different scales, where $\| \mathbf{W}_{\mathbf{j}} \|^{\mathbf{2}} / \mathbf{N}$ represents the contribution of the j-th lowest scale to the sample variance.

Another interesting and useful property of the DWT coefficients is that the j-th lowest scale coefficients $\bar{\mathbf{W}}_{\mathbf{j}}$ capture the frequency content of the original

[7] For details, please see Percival & Walden (2000), pp.62.

time series at the frequency band $f \in [1/2^{j+1}, 1/2^j]$.[8] Another big advantage of the DWT is its relative computational efficiency that is achieved through the so-called Pyramid algorithm introduced by Mallat (1999). This algorithm needs only operations of the degree $O(N)$ to compute the DWT and, in this sense, is even more efficient than the more famous FFT algorithm.

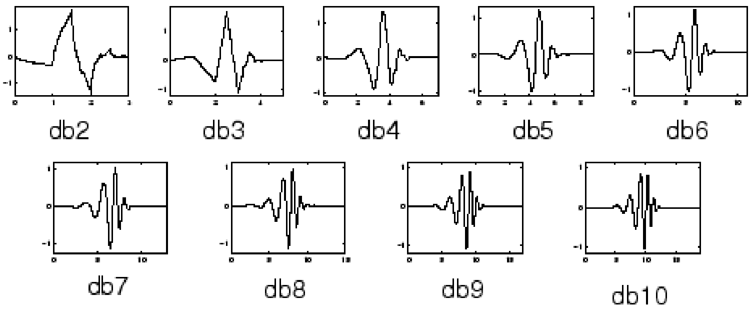
For all the nice things said so far about the DWT, there also comes a price in form of a few drawbacks of the DWT. One of the drawbacks of the DWT is that for all the nice properties that were mentioned in the previous paragraph to hold, we cannot choose the mother wavelets (or mother wavelet filters) arbitrarily (and go on to discretize them and then use them to construct the Q matrix that defines the DWT). Instead there are only certain classes of mother wavelets (or mother wavelet filters) that will preserve all the nice properties of the DWT. Again there is a whole extensive theory behind this.[9] Here we will mention only the most popular classes of wavelets used for the construction of the DWT. These are the so-called Daubechies wavelets (or daublets). Daubechies(1) wavelet is the Haar wavelet already mentioned above. The first few other Daubechies wavelets are depicted in figure 2.3. Of course for the use in the DWT these continuous wavelets are properly discretized and scaled. Other popular DWT wavelets include the symmlets. It should be said that the different families of suitable wavelets for the DFT is usually broad enough for practical purposes, therefore this drawback does not pose such a problem.

The main drawbacks of the DWT for an economist, as pointed out by Baruník & Baruníková (2013), is the restriction on the number of observations $N = 2^J$, great sensitivity of the DWT to shifts of the time series, as well as sensitivity with respect to the boundary values. These drawbacks are largely a consequence of the downsampling scheme defined by the grid for the DWT in figure 2.2.

[8] The exact meaning and derivation of this property can be found in Percival & Walden (2000) or Gençay *et al.* (2009).

[9] For more details, one can refer to Percival & Walden (2000).

Figure 2.3: First few Daubechies wavelets



Source: http://www.mathworks.com/help/wavelet/gs/introduction-to-the-wavelet-families.html

2.3.2 The Maximal Overlap Discrete Wavelet Transform

The Maximal Overlap Discrete Wavelet Transform (MODWT) is an attempt to modify the DWT in such a way, as to get rid of the drawbacks mentioned in the previous section, while retaining most of the analytical power that the DWT possesses. Again, we will not provide a full mathematical treatment of the MODWT here,[10] but rather provide a brief and basic outline of the properties of the MODWT, along with a comparison with the DWT.

The main difference between the DWT and the MODWT is that the MODWT is highly redundant transform when compared to the DWT. As we have seen in previous sections, the DWT can be thought of as an approximation to the CWT at certain points given by the grid in figure 2.2. The grid is getting more dense as we move from higher scales to lower scales. In the case of the MODWT this is not so and the grid is equally dense at all scales. Given the size N of a sample $\mathbf{X}$, the MODWT's output consists of J+1 vectors (with J

[10] For full mathematical treatment please confront Percival & Walden (2000), Mallat (1999) or Gençay *et al.* (2009).

chosen beforehand, indicating the number of scales)

$$\hat{\mathbf{W}}_{\mathbf{1}}, \hat{\mathbf{W}}_{\mathbf{2}}, \ldots, \hat{\mathbf{W}}_{\mathbf{J}}, \hat{\mathbf{V}}_{\mathbf{J}}.$$

Each of these vectors has length N and, as in the case of the DWT, the vector $\hat{\mathbf{W}}_{\mathbf{j}}$ corresponds to the j-th lowest scale approximations (each scale is again twice the previous scale). The MODWT is constructed similarly as the DFT through an efficient pyramid algorithm (although it is less efficient than in the case of the DWT). While for the DWT the length of the sample needed to be a power of two, this is no longer true for the MODWT and the length can be arbitrary.

Most importantly for our subsequent analysis is that the MODWT keeps the convenient properties of energy and variance decomposition with respect to different scales. These are captured in the following equations.

$$\| \mathbf{X} \|^{\mathbf{2}} = \sum_{\mathbf{j=1}}^{\mathbf{J}} \| \hat{\mathbf{W}}_{\mathbf{j}} \|^{\mathbf{2}} + \| \hat{\mathbf{V}}_{\mathbf{J}} \|^{\mathbf{2}}, \tag{2.12}$$

and

$$\hat{\sigma}_X^2 = \frac{1}{N} \sum_{j=1}^{J} \| \hat{\mathbf{W}}_{\mathbf{j}} \|^{\mathbf{2}} + \frac{\mathbf{1}}{\mathbf{N}} \| \hat{\mathbf{V}}_{\mathbf{J}} \|^{\mathbf{2}} - \overline{\mathbf{X}}^{\mathbf{2}}. \tag{2.13}$$

Furthermore, the MODWT also preserves the frequency interpretation i.e. the coefficients in vector $\hat{\mathbf{W}}_{\mathbf{j}}$ are associated with frequencies $f \in [1/2^{j+1}, 1/2^j]$.

2.4 Wavelet Variance

The MODWT is very useful in describing a theoretical concept named wavelet variance. Wavelet variance is in turn a very useful tool in analyzing the variance of stationary stochastic processes (time series). In this section we will give a brief outline of wavelet variance, this will be useful later on when we will define wavelet band spectrum regression.

When we defined the DWT we did it by means of matrix multiplication. It turns out that there is a different way of defining the DWT and the MODWT. This is through the concept of filtering. In fact, the highly efficient pyramid algorithm to compute the DWT and the MODWT uses this filtering approach.[11]

[11] More detailed exposition is provided in Percival & Walden (2000) and Gençay *et al.* (2009).

In case of the MODWT, this essentially boils down to assigning to the j-th lowest scale (denoted τ_j) a finite series of numbers $\left\{h_{j,l}\right\}_{l=0}^{L_j}$ called the j-th level MODWT filter.[12] With the use of such filter, the t-th coefficient $\hat{W}_{j,t}$ of the MODWT output vector $\mathbf{\hat{W}_j}$ is defined as

$$\hat{W}_{j,t} = \sum_{l=0}^{L_j-1} h_{j,l} X_{t-l\ modN} \qquad t \in \{1, \ldots, N\}. \tag{2.14}$$

Let us now consider some stochastic process $\left\{\mathbf{X_t}\right\}_{t=-\infty}^{\infty}$ and filter it with our j-th level MODWT wavelet filter to get a new stochastic process

$$\overline{W}_{j,t} \equiv \sum_{l=0}^{L_j-1} h_{j,l} \mathbf{X_{t-l}} \qquad \mathbf{t} \in \mathbb{Z}. \tag{2.15}$$

Variance of this process essentially defines the wavelet variance.

Definition 2.4 (Wavelet variance). *If the variance of the stochastic process defined by equation (2.15) exists, is finite and time-invariant, then we define the wavelet variance at scale τ_j as*

$$\boldsymbol{\upsilon}^{\mathbf{2}}_{\mathbf{X}}(\tau_j) \equiv var\left\{\overline{W}_{j,t}\right\} \tag{2.16}$$

The main reason the wavelet variance is useful to us is that it can decompose the variance of a stationary stochastic process as summarized in the following proposition.

Proposition 2.2 (Decomposition of variance). *Let $\left\{\mathbf{X_t}\right\}_{t=-\infty}^{\infty}$ be a stationary stochastic process with finite variance σ_X^2, and let the wavelet variance $\boldsymbol{\upsilon}^{\mathbf{2}}_{\mathbf{X}}(\tau_j)$ be defined for all scales $\tau_j, \quad j \in \mathbb{N}$. Then the following equality defining a decomposition of variance σ_X^2 across different scales holds:*

$$\sigma_X^2 = \mathbf{var}(\mathbf{X_t}) = \sum_{\mathbf{j=1}}^{\infty} \boldsymbol{\upsilon}^{\mathbf{2}}_{\mathbf{X}}(\tau_{\mathbf{j}}). \tag{2.17}$$

Proof. The proof can be found in Percival & Walden (2000), pp. 301. □

Furthermore, given a sample $\mathbf{X} = (\mathbf{X_1}, \mathbf{X_2}, \mathbf{X_3}, \ldots, \mathbf{X_N})$ it turns out that

[12] Here $L_j = (2^j - 1)(L - 1) - 1$, with $L = L_1$ being the starting length of the filter. The filter $\left\{h_{j,l}\right\}_{l=0}^{L_j}$ can also be thought of as the discretized wavelet function we use for our analysis.

the wavelet variance can be easily estimated using the MODWT as

$$\hat{\boldsymbol{\upsilon}}^2_{\mathbf{X}}(\tau_j) \equiv \frac{1}{N}\sum_{t=1}^{N} \hat{W}^2_{j,t}, \tag{2.18}$$

where $\hat{W}_{j,t}$ is the t-th coefficient of the MODWT output vector $\hat{\mathbf{W}}_{\mathbf{j}}$ (as before). This in fact is a biased estimator of wavelet variance. An unbiased estimator is given by a slightly more complicated formula

$$\tilde{\boldsymbol{\upsilon}}^2_{\mathbf{X}}(\tau_j) \equiv \frac{1}{M_j}\sum_{t=L_j}^{N} \hat{W}^2_{j,t}, \tag{2.19}$$

where L_j is the length of the j-th wavelet filter and $M_j = N - L_j$. For more details see Percival & Walden (2000), chapter 8.

In a similar manner to wavelet variance we can define wavelet covariance of two stochastic processes. Here we will just state the basics without proofs or formal definitions. Let $\left\{\mathbf{X_t}\right\}$ and $\left\{\mathbf{Y_t}\right\}$ be two stochastic processes for which we define two new stochastic processes $\overline{W}_{X,j,t}$ and $\overline{W}_{Y,j,t}$ in line with equation (2.15). Then, wavelet covariance of processes $\left\{\mathbf{X_t}\right\}$ and $\left\{\mathbf{Y_t}\right\}$ at scale τ_j is simply defined as

$$\boldsymbol{\upsilon}_{\boldsymbol{XY}}(\tau_j) \equiv cov\left\{\overline{W}_{X,j,t}, \overline{W}_{Y,j,t}\right\}. \tag{2.20}$$

Under some further assumptions on the processes $\left\{\mathbf{X_t}\right\}$ and $\left\{\mathbf{Y_t}\right\}$,[13] similar proposition to proposition 2.2 holds also in the case of wavelet covariance i.e.

$$\sum_{j=1}^{\infty} \boldsymbol{\upsilon}_{\boldsymbol{XY}}(\tau_j) = cov\left\{\mathbf{X_t}, \mathbf{Y_t}\right\}. \tag{2.21}$$

Of course, we require that the right hand side of this equation does not depend on t. This result will be useful in the definition of wavelet band spectrum regression in the next sections.

Again, there is a way to estimate the wavelet covariance using the MODWT coefficients. The estimator is analogous to the estimator of the wavelet variance and has the form

$$\tilde{\boldsymbol{\upsilon}}_{\boldsymbol{XY}}(\tau_j) \equiv \frac{1}{M_j}\sum_{t=L_j}^{N} \hat{W}_{X,j,t}\hat{W}_{Y,j,t},$$

[13] For more details, see Whitcher *et al.* (2000).

where L_j is the length of the j-th wavelet filter and $M_j = N - L_j$. An interested reader can refer to Whitcher *et al.* (2000) for further details.

2.5 Band Spectrum Regressions

We are now ready to describe the ideas and principles behind the main tool that we will use in the analysis of concrete economic data, the band spectrum regression, especially the band spectrum regression using wavelets.

2.5.1 Band Spectrum Regression using Fourier Methods

The idea of band spectrum regression was first proposed by Engle (1972). Let us imagine the simplest linear regression setting where we are given two vectors $\mathbf{y} = (\mathbf{y_1}, \mathbf{y_2}, \dots, \mathbf{y_n})$ and $\mathbf{x} = (\mathbf{x_1}, \mathbf{x_2}, \dots, \mathbf{x_n})$ which constitute the observed data of some quantities Y (the dependent variable) and X (the fixed regressor). We know/assume that the data were generated from a basic linear model

$$\mathbf{y_i} = \beta \mathbf{x_i} + \epsilon_\mathbf{i} \qquad \forall \mathbf{i} \in \{\mathbf{1}, \dots, \mathbf{n}\}\,,$$ [14]

where $\epsilon = (\epsilon_\mathbf{1}, \epsilon_\mathbf{2}, \dots, \epsilon_\mathbf{n})$ is a multivariate normal random vector with zero mean and $n \times n$ identity covariance matrix i.e. $\mathbf{N}(\mathbf{0}, \mathbf{I_{n\times n}})$. Thus in vector form the model can be written as

$$\mathbf{y} = \beta \mathbf{x} + \epsilon \tag{2.22}$$

and our concrete observed data $\mathbf{y_i}$ come in fact from one realization of the random vector ϵ.

The main problem, of course, is the unknown parameter β, which we want to estimate. The traditional way of estimating β is using the OLS estimator $\hat{\beta}^{OLS} = (\mathbf{x^H x})^{-\mathbf{1}}\mathbf{x^H y}$,[15], which turns out to be the best linear unbiased estimator according to standard theory.

The basic idea of the band spectrum regression is to take the discrete Fourier transform (DFT) of both sides of equation 2.22 to get a new equation of the form

$$\tilde{\mathbf{y}} = \beta \tilde{\mathbf{x}} + \tilde{\epsilon} \tag{2.23}$$

[14] The x's are fixed in the model, random part comes only from epsilon.

[15] Operator $\mathbf{H}$ is the conjugate transpose. In case of real vectors it is equivalent to the transpose operator $\mathbf{T}$.

and then estimate β using these transformed vectors. Taking the DFT is equivalent to multiplying by a DFT transform matrix $\mathbf{F}$, which is orthogonal,[16] therefore $\tilde{\mathbf{y}} = \mathbf{F}\mathbf{y}$, $\tilde{\mathbf{x}} = \mathbf{F}\mathbf{x}$ and $\tilde{\epsilon} = \mathbf{F}\epsilon$ in the above equation. Constructing the OLS estimator from the transformed equation 2.23 yields

$$\hat{\beta}^{\mathbf{OLS2}} = (\tilde{\mathbf{x}}^{\mathbf{H}}\tilde{\mathbf{x}})^{-1}\tilde{\mathbf{x}}^{\mathbf{H}}\tilde{\mathbf{y}} = ((\mathbf{F}\mathbf{x})^{\mathbf{H}}(\mathbf{F}\mathbf{x}))^{-\mathbf{1}}(\mathbf{F}\mathbf{x})^{\mathbf{H}}(\mathbf{F}\mathbf{y}) =$$
$$= (\mathbf{x}^{\mathbf{H}}\mathbf{F}^{\mathbf{H}}\mathbf{F}\mathbf{x})^{-\mathbf{1}}\mathbf{x}^{\mathbf{H}}\mathbf{F}^{\mathbf{H}}\mathbf{F}\mathbf{y} = (\mathbf{x}^{\mathbf{H}}\mathbf{I}\mathbf{x})^{-\mathbf{1}}\mathbf{x}^{\mathbf{H}}\mathbf{I}\mathbf{y} = (\mathbf{x}^{\mathbf{H}}\mathbf{x})^{-\mathbf{1}}\mathbf{x}^{\mathbf{H}}\mathbf{y} = \hat{\beta}^{\mathbf{OLS}},$$

which is the same estimator as before. The key thing to realize, however, is that the equation can be rewritten as

$$\hat{\beta}^{\mathbf{OLS2}} = (\tilde{\mathbf{x}}^{\mathbf{H}}\tilde{\mathbf{x}})^{-1}\tilde{\mathbf{x}}^{\mathbf{H}}\tilde{\mathbf{y}} = (\sum_{\mathbf{i=1}}^{\mathbf{n}} |\tilde{\mathbf{x}}_{\mathbf{i}}|^{\mathbf{2}})^{-\mathbf{1}}(\sum_{\mathbf{i=1}}^{\mathbf{n}} \tilde{\mathbf{x}}_{\mathbf{i}}^{\mathbf{H}}\tilde{\mathbf{y}}_{\mathbf{i}}),$$

where $|\tilde{\mathbf{x}}_{\mathbf{i}}|^{\mathbf{2}}$ is the i-th member of the periodogram of $\mathbf{x}$ and $\tilde{\mathbf{x}}_{\mathbf{i}}^{\mathbf{H}}\tilde{\mathbf{y}}_{\mathbf{i}}$ is the i-th member of the cross periodogram of $\mathbf{x}$ and $\mathbf{y}$. In other words, they represent the frequency content at the i-th lowest frequency. This turns out to be very useful because if some part of the spectrum is not interesting to us i.e. we are trying to estimate β only for some specific frequencies denoted as A,[17] then we can simply forget all the other frequencies and estimate β as

$$\hat{\beta}^{\mathbf{BSR}} = (\sum_{\mathbf{i\in A}} |\tilde{\mathbf{x}}_{\mathbf{i}}|^{\mathbf{2}})^{-\mathbf{1}}(\sum_{\mathbf{i\in A}} \tilde{\mathbf{x}}_{\mathbf{i}}^{\mathbf{H}}\tilde{\mathbf{y}}_{\mathbf{i}}). \tag{2.24}$$

Needless to say, this kind of approach should be very interesting to economists. It is in the very nature of economics to establish relationships between variables at different frequencies, the easiest example that comes to mind is the distinction that is often made between short-term and long-term relationships. Band spectrum regression provides a tool that is very well-suited for exactly this kind of problems.

It should also be noted that the same approach also works in the case of multivariate regression where $\mathbf{x}$ is a $n \times k$ matrix. The special case of $\mathbf{x}$ being a vector was chosen for illustrative purposes only. In the next section, we will

[16] This means that $\mathbf{F}^{\mathbf{H}}\mathbf{F} = \mathbf{F}\mathbf{F}^{\mathbf{H}} = \mathbf{I}$.

[17] For instance, for only long-term β we would get rid of the high frequencies and vice versa for the short-term β. A can also be only some band of the spectrum, thus the name Band Spectrum Regression.

describe a slightly different approach towards band spectrum regression using wavelets.

2.5.2 Wavelet Band Spectrum Regression

In this section, we will describe the main tool that will be used in our analysis, the Wavelet Band Spectrum Regression (WBSR). The exposition will be purely heuristic, giving just the main ideas instead of a fully rigorous treatment.

When defining wavelet band spectrum regression, we will need to work with a model that is little bit different than the model in the previous section. We still have the observed data $\mathbf{y} = (\mathbf{y_1}, \mathbf{y_2}, \ldots, \mathbf{y_n})$ and $\mathbf{x} = (\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n})$, but this time we will assume that $\mathbf{x_i}$'s, instead of being fixed, come from a sample of i.i.d.[18] random variables $\mathbf{X_1}, \mathbf{X_2}, \ldots, \mathbf{X_n}$. Furthermore, we will add an intercept term α to the model.[19] Then the underlying model looks like

$$\mathbf{Y_i} = \alpha + \beta \mathbf{X_i} + \epsilon_\mathbf{i} \qquad \forall \mathbf{i} \in \{\mathbf{1}, \ldots, \mathbf{n}\}, \tag{2.25}$$

where ϵ_i's are defined as in the previous section and are also uncorrelated with $\mathbf{X_i}$'s. It is not difficult to see that in this model the unknown coefficient β can be expressed as

$$\beta = \frac{\mathbf{cov}(\mathbf{X_i}, \mathbf{Y_i})}{\mathbf{var}(\mathbf{X_i})} \qquad \forall \mathbf{i} \in \{\mathbf{1}, \ldots, \mathbf{n}\}. \tag{2.26}$$

The OLS estimate of β is also easily calculated and is equal to

$$\hat{\beta}^{\mathbf{OLS}} = \frac{\frac{1}{\mathbf{n}} \sum_{\mathbf{i=1}}^{\mathbf{n}} (\mathbf{x_i y_i} - \bar{\mathbf{x}}\bar{\mathbf{y}})}{\frac{1}{\mathbf{n}} \sum_{\mathbf{i=1}}^{\mathbf{n}} (\mathbf{x_i}^\mathbf{2} - \bar{\mathbf{x}}^\mathbf{2})} = \frac{\widehat{\mathbf{cov}(\mathbf{X_i}, \mathbf{Y_i})}}{\widehat{\mathbf{var}(\mathbf{X_i})}}. \tag{2.27}$$

Thus, we can see that to estimate β means estimating the covariance of $\mathbf{X_i}$ and $\mathbf{Y_i}$ and dividing it by an estimate of the variance of $\mathbf{X_i}$. But this can also be done using the wavelet methods that were described in section 2.4. In fact, using equations (2.17) and (2.21) we can express β as the sum of wavelet covariances divided by the sum of wavelet variances i.e.

$$\beta \equiv \beta_{(\mathbf{1},\infty)} = \frac{\mathbf{cov}(\mathbf{X_i}, \mathbf{Y_i})}{\mathbf{var}(\mathbf{X_i})} = \left(\sum_{\mathbf{j=1}}^{\infty} \boldsymbol{\upsilon}_{\boldsymbol{X}}^{\mathbf{2}}(\tau_\mathbf{j}) \right)^{-1} \left(\sum_{\mathbf{j=1}}^{\infty} \boldsymbol{\upsilon}_{\boldsymbol{XY}}(\tau_\mathbf{j}) \right). \tag{2.28}$$

[18] Identically and independently distributed.

[19] The intercept term is added because in the applications part of the thesis we will usually use a model with an intercept.

Here, the interval $(1, \infty)$ indicates that the β is for all scales j from 1 to infinity. In a similar manner, we can define β only for some band of scales (K, L) as

$$\beta_{(\mathbf{K},\mathbf{L})} \equiv \left(\sum_{\mathbf{j}=\mathbf{K}}^{\mathbf{L}} \boldsymbol{\upsilon}_{\boldsymbol{X}}^{\mathbf{2}}(\tau_{\mathbf{j}}) \right)^{-\mathbf{1}} \left(\sum_{\mathbf{j}=\mathbf{K}}^{\mathbf{L}} \boldsymbol{\upsilon}_{\boldsymbol{XY}}(\tau_{\mathbf{j}}) \right) \qquad \mathbf{K}, \mathbf{L} \in \mathbb{N} \cup \infty, \mathbf{K} \leq \mathbf{L}. \quad (2.29)$$

The interpretation should be straightforward since we know from previous sections that the band of scales $j \in (K, L)$ is associated with the frequency band $f \in [1/2^{L+1}, 1/2^{K+1}]$. Thus, similarly as in the case of the band spectrum regression in the previous section, we are dealing with β only for some specified frequencies f.[20]

For purposes of estimation, we simply use the estimators of wavelet variance and covariance mentioned in section 2.4.

$$\hat{\beta}^{\mathbf{WBSR}}_{(\mathbf{K},\mathbf{L})} \equiv \left(\sum_{\mathbf{j}=\mathbf{K}}^{\mathbf{L}} \tilde{\boldsymbol{\upsilon}}_{\boldsymbol{X}}^{\mathbf{2}}(\tau_{\mathbf{j}}) \right)^{-\mathbf{1}} \left(\sum_{\mathbf{j}=\mathbf{K}}^{\mathbf{L}} \tilde{\boldsymbol{\upsilon}}_{\boldsymbol{XY}}(\tau_{\mathbf{j}}) \right) \qquad \mathbf{K}, \mathbf{L} \in \mathbb{N} \cup \infty, \mathbf{K} \leq \mathbf{L}. \quad (2.30)$$

We will not discuss the theoretical properties of this estimator here. For more details, the reader can, for example, refer to Fadili & Bullmore (2002).

[20] In the case of WBSR, f always belongs to some 'band' in the form $[1/2^{L+1}, 1/2^{K+1}]$ $\qquad K, L \in \mathbb{N} \cup \infty, K \leq L$.

Chapter 3

Applications

In this chapter, we will use the theoretical knowledge from the previous chapter to analyze two problems that can be related to economics, econometrics or finance. Although the main tool in the analysis should be wavelet band spectrum regression, this will not be the only way we will apply wavelets in the upcoming chapter.

In section 3.1, we will estimate the so-called realized-implied volatility relationship for the price of crude oil. At the beginning, we shall describe the theoretical framework underpinning our analysis, giving detailed account of realized volatility estimation techniques used. Next, we conduct an analysis of realized-implied volatility regression for the crude oil using a useful wavelet-based tool called wavelet coherence. At the end, coefficients of the regression are estimated on different scales using wavelet band spectrum regression. A brief discussion of the results concludes the section.

In section 3.2, we use similar techniques to analyze the relationship between labor productivity and unemployment for four major European countries during the period 1990-2014. We begin by a brief discussion of the theoretical knowledge behind this relationship. Then, we proceed to the wavelet coherence analysis for each country and also wavelet band spectrum regression analysis. The section ends by a discussion of the results.

3.1 Estimating Implied-Realized Volatility Relation in the Oil Market

In this section, we will introduce some basics of volatility modelling in modern financial econometrics. The key aim is to carry out an analysis of the so-called implied-realized volatility relationship, which should be similar to the analysis that can be found in Baruník & Baruníková (2013). While in Baruník & Baruníková (2013), the volatilities of S&P 500 and DAX indices are examined, we should instead focus on the volatility of the oil returns.[1] Throughout the whole analysis we will rely on the wavelet methods heavily, making the subject a very good example of the use of wavelets in modern economic practice.

3.1.1 Volatility (the Model)

One of the key areas of interest of financial economics is the study of asset prices. The underlying assets can be anything ranging from simple commodities such as oil or gold up to highly complex financial instruments such as various derivatives, indices etc. With studying the price process, which is undoubtedly interesting in itself, there naturally arises the need to study the returns i.e. the gain or loss an asset provides in some time period. After all, for any investor only the returns should carry important information, not the actual nominal prices. Particularly interesting from the point of view of economics is then the variability of prices during a certain period, a question closely connected to the behavior of returns as is explained below. Volatility, as a generic term, tries to capture or measure this variability of assets' prices over time. It is not difficult to see the importance of this matter for the economists. For instance, one of the reasons volatility might be important are the economic crises. We know from experience that economic crises are usually accompanied by high volatility in the financial markets (almost by definition), whereas in periods of prosperity and growth one would probably expect the volatilities to be lower. Therefore the study of volatility has rightfully an important place in financial economics.

Needless to say, since volatility is not readily observable, one usually needs to adopt some sort of mathematical model to be able to describe volatility precisely. Delving into the details of volatility modelling is far beyond the scope of this work.[2] Here we shall only describe the model that we are using in our analysis. The model and its theoretical properties were described in Andersen *et al.* (2001) and is also used by Baruník & Baruníková (2013), Baruník &

[1] One of the reasons that we chose the oil market is the availability of the so-called VIX index for the price of oil. As will be explained below, the VIX index is an estimator of implied volatility.

[2] Comprehensive overview of different ways to model volatility is provided by e.g. Bauwens *et al.* (2011).

Vácha (2013), or Baruník *et al.* (2015).[3] It is based on the so called "theory of continuous-time arbitrage-free price processes and the theory of quadratic variation". Its main advantage is that it can facilitate the use of high-frequency intraday data (returns) into volatility modelling, making the model suitable for contemporary needs.

First of all, we shall assume that the logarithm of price of some asset at time $t \in [0, T]$, denoted by p_t, is a stochastic process defined on a complete probability space $(\Omega, \Im, \mathbb{P})$.[4] Furthermore, the price process p_t is behaving according to the following stochastic differential equation

$$dp_t = \mu_t dt + \sigma_t dW_t + \xi_t dq_t, \tag{3.1}$$

where μ_t is predictable mean, σ_t is strictly positive 'volatility process', W_t is the standard Wiener process, q_t is a constant intensity (λ) Poisson process and ξ_t is a jump process. In fact, equation (3.1) is describing a Wiener process, with drift μ_t and diffusion parameter σ_t, that is occasionally (with intensity λ) jumping according to the jump process ξ_t. Following Baruník & Baruníková (2013), we further assume that the price process p_t is contaminated by microstructure *i.i.d.* noise ϵ_t with variance η^2, thus the observed log prices y_t have the form

$$y_t = p_t + \epsilon_t \qquad \forall t \in [0, T]. \tag{3.2}$$

With the description of the price process we can now easily define the continuously compounded asset return $r_{t,h}$ over a time interval $[t-h, t]$, where $0 \leq h \leq t \leq T$, as simply the difference of the log prices at times $t-h$ and t, i.e.

$$r_{t,h} = p_t - p_{t-h}. \tag{3.3}$$

What is of main interest to us is how to correctly capture the variability of the logarithmic price process defined above. A natural mathematical tool for this is the quadratic variation (denoted $QV_{t,h}$) of the stochastic process p_t defined as

$$QV_{t,h} = \lim_{||\Delta|| \to 0} \sum_{i=1}^{n} (p_{z_i} - p_{z_{i-1}})^2, \tag{3.4}$$

[3]Indeed, when describing the model we heavily rely on the last three sources and also Křehlík (2013).

[4]For our purposes we do not need to consider more complicated structure using filtrations. See Andersen *et al.* (2001) for more details.

where $\Delta = \Big\{ z_0, z_1, \ldots, z_n : t-h = z_0 < z_1 < z_2 < \ldots < z_n = t \Big\}$ is some partition of the interval $[t-h, t]$ and $||\Delta|| = \max_i(z_i - z_{i-1})$ is its norm. The limit sign in the equation (3.4) in this case refers to convergence in probability. Using equation (3.3), we can immediately see that the quadratic variation can in our case be written as

$$QV_{t,h} = \lim_{||\Delta|| \to 0} \sum_{i=1}^{n} r^2_{z_i,(z_i - z_{i-1})}, \tag{3.5}$$

thus we realize that variability of the price process is closely connected to the sum of squared returns of the process.

The quadratic variation defined by equation (3.4) measures the variability of the price process as a whole. However, according to equation (3.1) our price process contains random jumps and these jumps naturally contribute to the overall variation (as measured by $QV_{t,h}$) of the process. Sometimes we might be interested in filtering out the part of quadratic variation that is due to jumps, since jumps are random and 'unsystematic' in nature. In this way, we can obtain a measure of 'systematic' process variation that is not biased by jumps. It turns out that a natural measure of this 'systematic' variation is the so-called integrated variance (denoted $IV_{t,h}$). It is, quite intuitively, the integral of the squared 'volatility process' σ_t^2 from equation (3.1), i.e.

$$IV_{t,h} = \int_{t-h}^{t} \sigma_s^2 ds. \tag{3.6}$$

Under some further technical assumptions on the price process p_t, Back (1991) was able to show that p_t is a special kind of semi-martingale process and therefore can be decomposed into a continuous part and a discrete jump part.[5] This means that the quadratic variation can also be decomposed into two parts

$$QV_{t,h} = \int_{t-h}^{h} \sigma_s^2 ds \; + \sum_{t-h \le l \le t} J_l^2 \equiv IV_{t,h} + JV_{t,h}, \tag{3.7}$$

where $\sum_{t-h \le l \le t} J_l^2 \equiv JV_{t,h}$ denotes the sum of squares of jumps that occurred in the time interval $[t-h, t]$. Equation (3.7) provides a natural decomposition of the quadratic variation into the 'systematic' integrated variance part and 'unsystematic' jumps part.

[5] For more details on this see Back (1991). For an overview of the theory of semi-martingales see, for example, Kallenberg (2002).

In reality, we are then interested in finding suitable estimators for $QV_{t,h}$ and/or $IV_{t,h}$.

3.1.2 Realized Volatility

It is important to realize that both $QV_{t,h}$ and $IV_{t,h}$ from the previous section are actually random variables defined on our original probability space $(\Omega, \Im, \mathbb{P})$. Realized volatility is a generic term that simply refers to the actual realized values of the random variables $QV_{t,h}$ and/or $IV_{t,h}$. To put it more simply, when we observe some historical prices (i.e. realized values of the price process) over an interval $[0, T]$, the realized volatility is simply the concrete values of $QV_{t,h}$ and/or $IV_{t,h}$ corresponding to the prices that we have observed.

The problem that remains is that both $QV_{t,h}$ and $IV_{t,h}$ are not directly observable, which is equivalent to saying that realized volatility is not directly observable. For this reason, various methods have been devised that try to estimate realized volatility (i.e. realized $QV_{t,h}$ and/or $IV_{t,h}$) as a function of the observed prices (returns). In this section, we will define and briefly describe three different estimators of realized volatility. These are the RV, BPV and JWTSRV estimators. The latter of these is heavily relying on the use of wavelets, therefore we will allocate the most attention to it. We will use all of these three estimators later for the estimation of realized-implied volatility relationship in the oil market.

RV Estimator

The RV (realized variance) estimator is the simplest of all estimators. It stems directly from equation (3.5) and was first proposed and used by Andersen & Bollerslev (1998).

Definition 3.1 (RV estimator). *The RV estimator of quadratic variation $QV_{t,h}$ of a stochastic price process p_t is defined as*

$$RV_{t,h} \equiv \sum_{i=1}^{n} r^2_{z_i,(z_i - z_{i-1})}, \tag{3.8}$$

where all the notation is in line with section 3.1.1 above. The partition

$$\Big\{ z_0, z_1, \ldots, z_n : t - h = z_0 < z_1 < z_2 < \ldots < z_n = t \Big\}$$

of the interval $[t-h, t]$ *is assumed to be equidistant and corresponds to sampling points of the price process.*

From equation (3.5) it is immediately clear that the RV estimator converges in probability to the quadratic variation,[6] i.e

$$RV_{t,h} \xrightarrow[n\to\infty]{P} QV_{t,h}. \tag{3.9}$$

Therefore, with the absence of microstructure noise and jumps, the RV estimator is a consistent estimator of quadratic variation $QV_{t,h}$, which is in this case equal to the integrated variance $IV_{t,h}$. Barndorff-Nielsen & Shephard (2001) were able to show asymptotic normality for the RV estimator in the form

$$\sqrt{n}\left(\frac{RV_{t,h}-QV_{t,h}}{\sqrt{2IQ_{t,h}}}\right) \xrightarrow[n\to\infty]{D} N(0,1), \tag{3.10}$$

where $IQ_{t,h} \equiv \int_t^{t-h} \sigma_s^4 ds$ is the integrated quarticity. We can see that all the nice properties of the RV estimator are asymptotic in nature, therefore large number of observations is desirable. The RV estimator is therefore well-suited for the high-frequency intraday data that tend to be very numerous.

The downside of the RV estimator is that it only estimates the $QV_{t,h}$ and is unable to filter out the $IV_{t,h}$ in the presence of jumps in the process. Furthermore, the RV estimator becomes biased when the process is subject to microstructure noise as is demonstrated, for example, in Zhang *et al.* (2005). For more interested reader, a nice overview of RV estimator is provided by Andersen & Benzoni (2008).

BPV Estimator

The RV estimator always estimates the quadratic variation $QV_{t,h}$ and therefore is not robust to jumps or microstruccture noise (from the point of view of estimating $IV_{t,h}$). It is no wonder then that various more sophisticated methods of estimating integrated variance $IV_{t,h}$ have been proposed. One of the most successful approaches is the one devised by Barndorff-Nielsen & Shephard (2003). The method is collectively known as the realized bipower variation. It was originally devised to overcome the problems with jumps in the price process,

[6] Existence of quadratic variation is guaranteed by the theory of semi-martingales, see for example Kallenberg (2002).

but as we will see below there exists a minor modification that deals also with microstructure noise. The original definition of the bipower variation estimator is below.

Definition 3.2 (Bipower variation estimator). *The bipower variation estimator of integrated variance $IV_{t,h}$ of a stochastic price process p_t is defined as*

$$BV_{t,h} \equiv \frac{\pi}{2} \sum_{i=1}^{n-1} |r_{z_i,(z_i - z_{i-1})}||r_{z_{i+1},(z_{i+1}-z_i)}|, \tag{3.11}$$

where all the notation is in line with section 3.1.1 above. The partition

$$\Big\{ z_0, z_1, \ldots, z_n : t-h = z_0 < z_1 < z_2 < \ldots < z_n = t \Big\}$$

of the interval $[t-h, t]$ is assumed to be equidistant and corresponds to sampling points of the price process.

The theoretical properties of this estimator were studied by Barndorff-Nielsen & Shephard (2004) or Barndorff-Nielsen *et al.* (2005) and they turn out to be quite favorable. Most importantly, it allows for the separation of the quadratic variation into the integrated variance part and the jump part. In fact, the bipower variation converges in probability to the integrated variance i.e.

$$BV_{t,h} \xrightarrow[n\to\infty]{P} IV_{t,h} = \int_{t-h}^{t} \sigma_s^2 ds. \tag{3.12}$$

Thus, when estimating $IV_{t,h}$, bipower variation estimator is robust to jumps in the process. Moreover, it can also be useful in testing for the presence of jumps.[7]

We have dealt with the problem of jumps, but there still remains the problem of microstructure noise that can introduce bias into the estimator. For this reason, a slightly modified version of the bipower estimator has been introduced.

Definition 3.3 (BPV estimator). *The BPV estimator of integrated variance $IV_{t,h}$ of a stochastic price process p_t is defined as*

$$BPV_{t,h} \equiv \frac{\pi}{2} \frac{n}{n-2} \sum_{i=1}^{n-2} |r_{z_i,(z_i - z_{i-1})}||r_{z_{i+2},(z_{i+2}-z_{i+1})}|, \tag{3.13}$$

[7] See, for example, Andersen *et al.* (2007).

where all the notation is in line with section 3.1.1. above. The partition

$$\Big\{z_0, z_1, \ldots, z_n : t-h = z_0 < z_1 < z_2 < \ldots < z_n = t\Big\}$$

of the interval $[t-h, t]$ *is assumed to be equidistant and corresponds to sampling points of the price process.*

The BPV estimator from equation (3.13) was proposed by Andersen *et al.* (2003) and its properties were studied by Huang & Tauchen (2005). The main point is that the BPV estimator still keeps its convergence in probability to the integrated variance i.e.

$$BPV_{t,h} \xrightarrow[n\to\infty]{P} IV_{t,h} = \int_{t-h}^{t} \sigma_s^2 ds, \tag{3.14}$$

while at the same time it should also diminish the bias introduced by potential microstructure noise.[8] In our analysis below, when referring to a bipower variation estimator, we shall always mean and always use the BPV estimator from equation (3.13) .

JWTSRV Estimator

Another method that estimates integrated variance in the presence of jumps and microstructure noise is the recently proposed 'jump adjusted wavelet two scale realized variance' (JWTSRV) estimator. It was introduced by Baruník & Vácha (2013) and is relying on the use of wavelets in two stages. In the first stage, following the work of Fan & Wang (2007), the structure of the most basic Haar wavelet is used to detect and filter out the jumps in the price process. On the other hand, in the second stage a noise robust 'two scale realized variance' (TSRV) estimator developed by Zhang *et al.* (2005) is 'waveletized' using energy-preserving properties of wavelets.

First of all, we shall describe how the JWTSRV estimator deals with jumps. The idea behind this is fairly intuitive and uses the inherent structure of the wavelets. The oscillating property of wavelets (equation (2.1)) turns out to be very useful in detecting jumps. We can see from figure 2.1.a) and 2.1.b) that the basic wavelets, such as the Haar wavelet, can actually be thought of as 'jumps' when the scale parameter is low enough. We remember that the basic idea of the wavelet transform such as the MODWT is to decompose the time

[8] See Huang & Tauchen (2005) p. 30 (485) for the main idea behind this.

series into the weighted sum of such 'jumps'. The weights are then given by the corresponding wavelet coefficients. If a wavelet coefficient is high, it can be interpreted as a high 'jump'. Therefore, a relatively high low-scale wavelet coefficient is a good indication of a jump in the process (or time series). Of course, this idea is further formalized and refined in the literature. For more details, the reader can refer to Baruník & Vácha (2013) and Fan & Wang (2007). Here we shall restrict ourselves to defining the procedure of jump detection and stating some theoretical results as in Baruník (2011).

Definition 3.4 (Wavelet jump detection). *Let* $\hat{W}_{1,k}, \quad k \in \{1, 2, \ldots, n\}$*, be the* 1^{st} *level Haar MODWT wavelet coefficients of some time series* p_t *over* $[t-h, t]$ *of length* n*. If for some* $\hat{W}_{1,k}$

$$|\hat{W}_{1,k}| > \frac{median\{|\hat{W}_{1,k}|, \quad k \in \{1, 2, \ldots, n\}\}}{0.6745} \sqrt{2 \log n}, \tag{3.15}$$

then $\hat{\tau}_l = \{k\}$ *is the estimated jump location with size* $\bar{p}_{\hat{\tau}_{l+}} - \bar{p}_{\hat{\tau}_{l-}}$ *(averages over* $[\hat{\tau}_l, \hat{\tau}_l + \delta_n]$ *and* $[\hat{\tau}_l, \hat{\tau}_l - \delta_n]$*, respectively, with* δ_n *being the small neighborhood of the estimated jump location).*

The jump variation $\widehat{WJV}$ *is then estimated by the sum of the squares of all the estimated jump sizes:*

$$\widehat{WJV} = \sum_{jumps} (\bar{p}_{\hat{\tau}_{l+}} - \bar{p}_{\hat{\tau}_{l-}})^2. \tag{3.16}$$

It turns out that the $\widehat{WJV}$ estimator is converging in probability to the jump variation $JV_{t,h}$ (see equation (3.7)) i.e.

$$\widehat{WJV} \xrightarrow[n \to \infty]{P} JV_{t,h}. \tag{3.17}$$

Furthermore, the quadratic variation of the filtered process $p^{(J)} = p_t - \widehat{WJV}$ is converging to the integrated variance $IV_{t,h}$. In this way, by filtering out $\widehat{WJV}$ we are able to get rid of the jump problem and this is what constitutes the first stage of the JWTSRV estimator.

In the second stage of the JWTSRV estimation, we work with the already filtered series $p^{(J)}$. The idea is to use a wavelet version of TSRV estimator defined by Zhang *et al.* (2005). In this work, a few ways of tackling the problem of microstructure noise are presented. One of the key methods of dealing with the noise problem is to subsample the data. This generally boils down to taking every k-th ($k > 1$) observation instead of every observation. This makes the data more widespread and should generally help to reduce the microstructure noise.[9] Zhang *et al.* (2005) were able to use this technique to find an estimator that deals with microstructure noise very well.

Definition 3.5 (TSRV estimator). *Let us have a time series $p^{(J)}$ of length n corresponding to a filtered price process over an interval $[t-h, t]$. Then the TSRV estimator of quadratic variation of the price process $p^{(J)}$ is given by*

$$RV_{t,h}^{(TSRV)} \equiv \frac{1}{K}\sum_{k=1}^{K} RV_{t,h}^{(k)} - \frac{\bar{n}}{n} RV_{t,h}^{(all)}, \tag{3.18}$$

where $K > 1$ (but is small relatively to n),[10] $RV_{t,h}^{(k)}$ is the RV estimator from definition (3.1) applied to a subsample $(p_k^{(J)}, p_{k+K}^{(J)}, p_{k+2K}^{(J)}, \ldots, p_{n_k}^{(J)})$.[11] On the other hand, $RV_{t,h}^{(all)}$ is the RV estimator applied to the whole time series $p^{(J)}$. Lastly, $\bar{n} = \frac{1}{K}\sum_{k=1}^{K} n_k$.

Theoretical properties of the TSRV estimator are, again, studied in detail in Zhang *et al.* (2005). For our purposes it suffices to say that the TSRV estimator $RV_{t,h}^{(TSRV)}$ is a consistent estimator of the integrated variance $IV_{t,h}$ of a price process p_t (or $p^{(J)}$ in our case) in the presence of microstructure noise. Thus, applying TSRV to the filtered data $p^{(J)}$ would provide a consistent estimator of integrated variance in the presence of both noise and jumps. To get the full JWTSRV estimator we still need to add a wavelet adjustment to the estimator.

In Baruník (2011) and Baruník & Vácha (2013), the $RV_{t,h}^{(k)}$ and $RV_{t,h}^{(all)}$ from equation (3.18) are essentially substituted with sums of squared MODWT wavelet coefficients. This is possible due to the decomposition relationship from equation (2.12) where on the right-hand side we substitute our RV estimator from definition (3.1) and on the left-hand side we have the MODWT wavelet coefficients of subsampled returns. By changing the order of the summation so that $\sum_{j=1}^{J+1}$ comes to the front in all expressions we obtain the JWTSRV estimator.

Definition 3.6 (JWTSRV estimator). *Let us have a time series $p^{(J)}$ of length n corresponding to a filtered price process over an interval $[t-h, t]$. Then the JWTSRV estimator of quadratic variation of the price process $p^{(J)}$ is given by*

[9] After all, the BPV estimator from equation (3.13) also uses this technique.

[10] Zhang *et al.* (2005) give an optimal K to be of order $n^{2/3}$. In their work, even more precise formula for K can be found.

[11] Here, n_k is maximum possible given the sample length.

$$RV_{t,h}^{(JWTSRV)} \equiv \sum_{j=1}^{J+1} RV_{j,t,h}^{(JWTSRV)} \equiv \sum_{j=1}^{J+1} \left(\frac{1}{K} \sum_{k=1}^{K} ||\hat{\mathbf{W}}_{\mathbf{j}}^{(\mathbf{k})}||^2 - \frac{\bar{n}}{n} ||\hat{\mathbf{W}}_{\mathbf{j}}^{(\mathbf{all})}||^2 \right), \tag{3.19}$$

where $K > 1$ (but is small relatively to n) as before, $\hat{\mathbf{W}}_{\mathbf{j}}^{(\mathbf{k})}$ are the j-th level MODWT wavelet coefficients that are obtained by transforming the returns of a subsample $(p_k^{(J)}, p_{k+K}^{(J)}, p_{k+2K}^{(J)}, \ldots, p_{n_k}^{(J)})$. On the other hand, $\hat{\mathbf{W}}_{\mathbf{j}}^{(\mathbf{all})}$ are the j-th level MODWT wavelet coefficients that are obtained by transforming the returns of the whole sample $p^{(J)}$. In all places we assume that the MODWT uses J scales and that $\hat{\mathbf{W}}_{\mathbf{J+1}}^{()} = \hat{\mathbf{V}}_{\mathbf{J}}^{()}$ Lastly, $\bar{n} = \frac{1}{K}\sum_{k=1}^{K} n_k$ as in the definition of TSRV.

We can see from the definition that the JWTSRV can naturally decompose the quadratic variation estimator into J+1 different scales, thus assigning to each scale its contributing weight of the total quadratic variation. This can again be a useful property guaranteed by the use of wavelets. Moreover, as was shown in Baruník (2011), the JWTSRV still retains most of the nice properties of the TSRV estimator,[12] namely it is a consistent estimator of the integrated variance $IV_{t,h}$ in the presence of both microstructure noise and jumps.

3.1.3 Implied Volatility and the VIX Index

In the previous section, we provided a fairly extensive outline of the methods we shall use to measure realized volatility. Realized volatility is in fact a realization of a random variable[13] and, as such, can only be measured using historical prices. The whole concept is inherently connected with the past and does not tell us much about the future.

On the other hand, implied volatility is concerned with forecasting the future volatility of underlying prices of assets i.e. looking into the future instead of the past. The straightforward approach would normally be to use the historical realized volatilities to produce forecasts. Indeed, this is a legitimate approach, but modern research and methods are more focused on the so-called option implied volatility. This approach is based on the widespread belief that, in an efficient market, option prices carry useful information about the future realized volatility (as measured by $QV_{t,h}$ and/or $IV_{t,h}$) of a price process (see the

[12] After all, it is a modification of the TSRV estimator.

[13] See the brief discussion at the beginning of the previous section.

quote below). The term 'implied volatility' then usually refers to the volatility implied by the option prices.

"It is widely believed that the volatility implied in an options price is the option markets forecast of future return volatility over the remaining life of this option. Under a rational expectations assumption, the market uses all the information available to form its expectations about future volatility, and hence the market option price reveals the markets true volatility estimate. Furthermore, if the market is efficient, the markets estimate, the implied volatility, is the best possible forecast given the currently available information . That is, all information necessary to explain future realized volatility generated by all other explanatory variables in the market information set should be subsumed in the implied volatility."[14]

Traditionally, implied volatility models had been based on some specific option pricing model, for example, the famous Black-Scholes model introduced by Black & Scholes (1973). The evidence on model-based implied volatility being an efficient forecast of realized volatility had been mixed.[15] One of the first research papers that were able to support the quote above were the authors of it, Christensen & Prabhala (1998) and Christensen & Hansen (2002). They found that implied volatility produces an efficient and unbiased forecasts of the future realized volatility and also revealed some deficiencies of previous works.

More recently, new ways of modelling implied volatility have been proposed. Britten-Jones & Neuberger (2000) developed a characterization of "all continuous price processes that are consistent with current option prices." Using this characterization, a new implied volatility forecast is presented. This implied volatility does not assume any underlying pricing model i.e. it is a model-free implied volatility (MFIV). It uses a complete set of call options with continuum of strikes (K) and continuum of maturities (T). The current price of the option with strike K and maturity T is denoted as $C(T, K)$. Under some additional assumptions,[16] they were able to show that the sum of squared returns between two future periods T_1 and T_2 (i.e. realized volatility between T_1 and T_2) has a risk-free expected value completely determined by the current prices of call options. The equation that formalizes this statement is taken from Jiang &

[14] Christensen & Hansen (2002) page 1 (187).

[15] See Christensen & Prabhala (1998) and Christensen & Hansen (2002) for an overview and also some more about the model-based methodology.

[16] See Proposition 2 in Britten-Jones & Neuberger (2000).

Tian (2005) and has the form

$$E_0^Q\left(\int_{T_1}^{T_2}\left(\frac{dp_t}{p_t}\right)^2\right) = 2\int_0^\infty \frac{C(T_2,K)-C(T_1,K)}{K^2}dK, \qquad (3.20)$$

where on the left-hand side we have the expected value (under risk-free measure Q) of the sum of squared returns (expressed as an integral) and on the right-hand side there is an expression completely determined by the current option prices. Britten-Jones & Neuberger (2000) derived MFIV only for continuous price process, excluding, for example, price processes with jumps. Jiang & Tian (2005) were able to extend the MFIV model to include price processes with jumps. Furthermore, they provided useful practical modifications to equation (3.20) at the same time showing that MFIV is an efficient and unbiased estimator of future realized volatility and subsumes all the information present in the Black-Scholes implied volatility and historical volatility.

MFIV can be simply estimated from the option prices (by estimating the integral on the right-hand side of equation (3.20) or some modification of it). Another of its strengths is that, for the MFIV to be valid, it does not need any underlying pricing model to hold. Thus, to calculate a valid implied volatility measure for some asset one only needs sufficient amount of current option prices for the underlying asset. One of the first regularly-calculated measures of implied volatility was the Chicago Board Option Exchange (CBOE) Volatility Index (VIX), known simply as the VIX index.

The VIX index was introduced by CBOE in 1993. It was designed to be a measure of market's expectation of future 30-day volatility of the S&P 100 index. It is calculated and disseminated to the trading markets in real time throughout each trading day. At the beginning, the CBOE used Black-Scholes model-based methodology to compute the VIX index.[17] On the 22^{nd} of September 2003 a new revamped version of the VIX index was introduced as a joint effort of CBOE and Goldman Sachs. The biggest change was that the new VIX index no longer uses model-based methodology for calculating the implied volatility.[18] Instead, a model-free framework is inherited from the work of Demeterfi *et al.* (1999). The framework is based on the concept of

[17] The old VIX index is nowadays also referred to as the VXO index. The details on its calculation can be found on the CBOE's website (http://www.cboe.com/micro/vxo/).

[18] Also, the new VIX index calculates implied volatility for the S&P 500 index as opposed to the older S&P 100 index. The exact calculation method for the new VIX index will not be presented here. For the details one can consult CBOE (2003).

'fair value of future variance' and is similar to the MFIV framework introduced above. In fact, Jiang & Tian (2006) were able to prove that that the new framework is theoretically equivalent to the MFIV methodology of Britten-Jones & Neuberger (2000). Thus, the VIX index can be perceived as an MFIV implied volatility estimator for the S&P 500 index.

The CBOE, using the same methodology as for the new VIX index, introduced implied volatility estimators (or forecasts) also for several other tradable assets.[19] Later in the thesis, we will be using the so-called oil VIX (officially listed as OVX index). According to CBOE it "measures the market's expectation of 30-day volatility of crude oil prices by applying the VIX methodology to United States Oil Fund, LP (Ticker - USO) options spanning a wide range of strike prices."[20] The OVX was introduced in 2007 and started to be traded on the 10^{th} of May 2007. In our analysis, the OVX will be used as a proxy to the 30-day model-free implied volatility of the crude oil i.e. as a 30-day model-free option based forecast of the future realized volatility of the crude oil's price.

3.1.4 Realized-Implied Volatility Regression

Given some measures of realized volatility and implied volatility, one is mostly interested in how well the implied volatility forecasts the future realized volatility. In other words, we would like to assess the information content (about future realized volatility) that is present in our implied volatility estimator. Ideally, implied volatility should be highly correlated with future realized volatility, providing an efficient forecast.

The simplest way to measure the information content of implied volatility is by the means of univariate linear regression. This is conventionally done by using an equation with the following form.

Definition 3.7 (Realized-Implied Volatility Regression). *For some underlying asset price process p_t, let us assume that we possess estimates of ex-post realized volatility $RV_{t+h,h}$ over an interval $[t, t+h]$ and over sufficient span of finite times $t \in T$. Furthermore, let us also assume that for the same span of times $t \in T$ we obtained a h-period ex-ante estimator of implied volatility IV_t. By*

[19] Among others these are the CBOE Nasdaq-100 Volatility Index (VXN), the CBOE Russell 2000 Volatility Index (RVX), the CBOE Gold ETF Volatility Index (GVZ) or the CBOE EuroCurrency ETF Volatility Index (EVZ).

[20] Cited from (http://www.cboe.com/micro/oilvix/introduction.aspx).

the Realized-Implied Volatility Regression we mean a regression equation in the form

$$RV_{t+h,h} = \alpha + \beta IV_t + \epsilon_t \quad \forall t \in T, \tag{3.21}$$

where ϵ_t *is the usual i.i.d. white noise with zero mean and finite variance.*

Of course, the main interest is then the estimation of the coefficients α and β from equation (3.21). Needless to say, these coefficients will generally depend on the concrete estimators chosen for $RV_{t+h,h}$ and IV_t.

Christensen & Prabhala (1998) identify at least three testable hypotheses connected to equation (3.21). Firstly, we can test for β being zero (i.e. no information content in implied volatility estimator). Secondly, $\alpha = 0$ and $\beta = 1$ would mean that IV_t is an unbiased estimator of future realized volatility $RV_{t+h,h}$. Thirdly, to test for the efficiency of the IV_t estimator we would test for the residuals being serially uncorrelated and white noise with zero mean and constant variance.

One would normally use classical OLS estimate for β in equation (3.21). However, there might be some drawbacks or pitfalls with using the OLS estimate. First of all, as we have discussed in chapter 2, economic relationships have often complicated structure making it reasonable to break them down into different frequencies or scales (e.g. short-term and long-term). A standard OLS estimate does not provide this differentiation and, thus, might not give us the full picture of the relationship. For example, studying the S&P 500 and DAX indices Baruník & Baruníková (2013), using wavelet band spectrum regression, found that IV_t contains much more information about the $RV_{t+h,h}$ in the time scales higher than 32 days, while at the shorter time scales there might be very little correlation between the two. This is the case where the band spectrum regression methods described in section 2.5 turn out to be very useful providing a much richer view of the complicated relationships.

Second issue, as discussed in Baruník & Baruníková (2013), is the problem of fractional cointegration.[21] This would be the case if both $RV_{t+h,h}$ and IV_t from equation (3.21) were $I(d)$ processes (i.e. integrated of order d) and the series $\epsilon_t = RV_{t+h,h} - \alpha - \beta IV_t$ would be $I(d_u)$ fractionally integrated long-memory process with $d_u < d$. It is a stylized fact that the volatilities of asset price processes usually show long-memory behavior (i.e. the are $I(d)$ with

[21] For details on time series integration and its terminology see for example Engle & Granger (1987).

$0 < d < 0.5$). Kellard *et al.* (2010) even suggested that they lie in the non-stationary region $0.5 < d < 1$. Furthermore, few researchers found that there really might be a fractional cointegration relation between $RV_{t+h,h}$ and IV_t.[22] As mentioned by Bandi & Perron (2003), fractional cointegration is associated by long-term co-movements and , as a result, the classical OLS estimate will not provide us with accurate information about unbiasedness of the implied volatility as a forecast for the future realized volatility. Instead, a tool that is able to differentiate between the long-term and short-term would be useful for solving this problem. Again, the standard solution for part of the literature is to use one of the band spectrum regression methods described in section 2.5.[23] Baruník & Baruníková (2013) also note that the specific use of wavelet spectral methods might be especially useful in the case of non-stationarity of the volatilities. This is because wavelets are generally very handy when dealing with non-stationary series.[24]

Here we shall not burden our work with more detailed theoretical analysis of the fractional cointegration problem. An interested reader can confront one of the sources from the preceding paragraph. Instead, we shall proceed to the practical analysis of the realized-implied volatility relationship from equation (3.21) for the crude oil.

3.1.5 Data and Procedure

In our analysis, we use two different datasets to compute the left-hand side and the right-hand side of equation (3.21).

For the left-hand side we use daily (except from Saturdays, Sundays and bank holidays) 24-hour intraday prices of crude oil as traded on the New York Mercantile Exchange (NYMEX); the data were obtained from Tick Data, Inc. We use both 5 minute prices and 1 minute prices. From this raw high-frequency data we extract the logarithmic returns. The data are spanning the period between the 10^{th} of May 2007 and the 12^{th} of February 2014, summing to 1707 trading days overall.

We try to use different methods from section 3.1.2 in order to compute daily realized volatility estimates. In this sense, the interval $[t-h, t]$ from definitions 3.1, 3.2, 3.3 and 3.6 is meant to represent one 24-hour trading day. For

[22] See, for example, Bandi & Perron (2003) or Kellard *et al.* (2010).

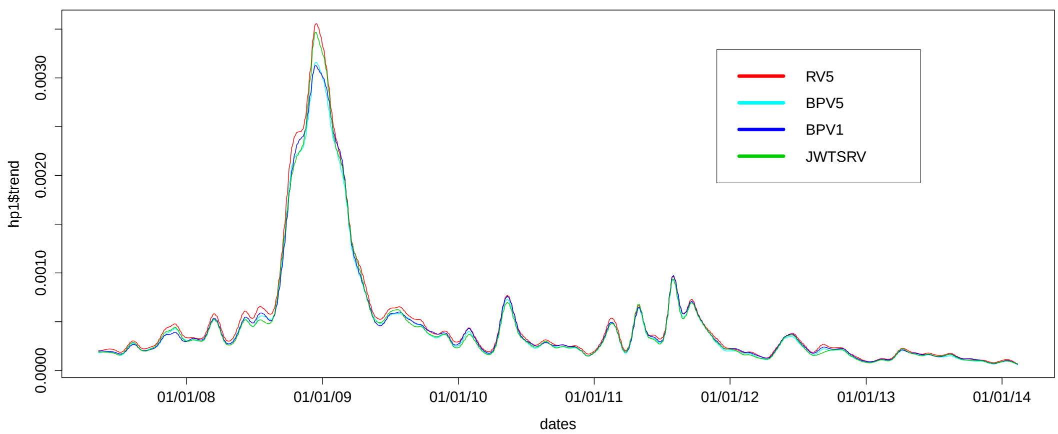
[23] See, for example, Nielsen & Frederiksen (2010) for more details.

[24] One of the main rationale behind wavelet analysis is to be able to locally analyze non-stationary time series.

each trading day we compute 4 different estimates of daily realized volatility. First of all, we compute the simplest RV estimator from definition 3.1 using 5 minute data (denoted RV5). Then, we compute both 5 minute and 1 minute BPV estimators from definition 3.3, which should be robust to both jumps and microstructure noise (denoted BPV5 and BPV1). Finally, we compute the two-stage JWTSRV estimator using the 1 minute data (denoted JWTSRV). For the first stage of the computation (i.e. jump detection), we use the simplest Haar wavelet. For the second stage of the computation, we use the Dauchechies D4 wavelet using four scales (i.e. $J = 4$), meaning scales of 1-2 minutes, 2-4 minutes, 4-8 minutes and 8-16 minutes. We also perform the decomposition of the realized volatility into $J + 1$ different components, first four corresponding to the four scales and the fifth corresponding to the average vector **V5**. The results of these procedures are presented in the Appendix A. Realized volatility was by far the highest around January 2009, shortly after the oil price peak of 147.3 US dollars per barrel in July 2008. We can see from figure A.6 from the appendix that, generally speaking, price variations on the first two time scales (1-2 minutes and 2-4 minutes) contribute the most towards realized volatility.

For comparison of the different estimators we plot them into one figure (figure A.5). Since the raw data are too coarse for meaningful graphical comparison, we smooth the data using the well established Hodrick-Prescott filter and plot the smoothed data instead.

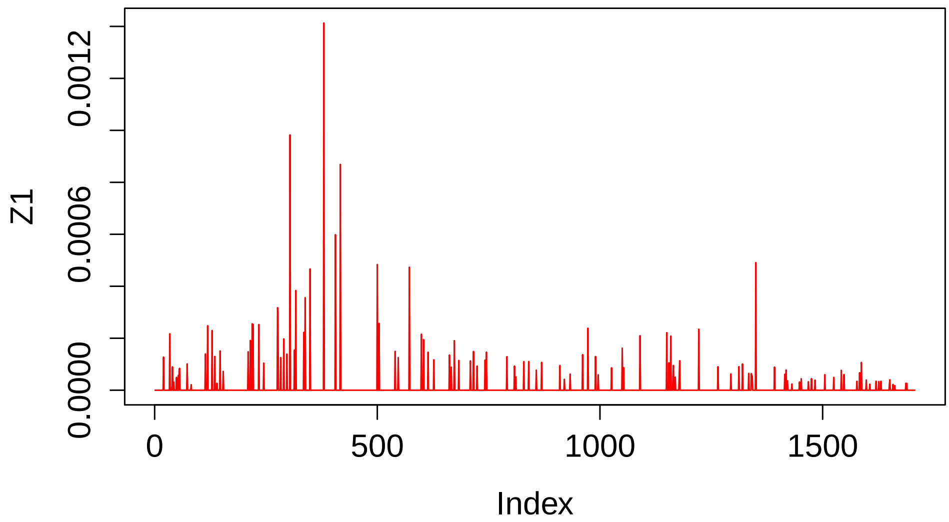
Figure 3.1: Smoothed daily realized volatility estimators



From figure 3.1 (and A.5) we immediately see that all four estimators produce fairly similar results. This result might suggest that in our observed price process there were not many jumps and not much microstructure noise. Upon closer look, though, we notice that the RV5 estimator tends to be slightly higher than the rest, especially during the peak period around January 2009. This is to be expected, since the RV5 estimator is an estimator of the quadratic variation including jumps, whereas the other three are estimators for integrated variance only.

In fact, we carried out a brief jumps analysis comparing the RV5 estimator to the BPV5 estimator. We tested the data for the presence of jumps using a standard methodology based on theoretical works on bipower variation mostly due to Barndorff-Nielsen and Shephard.[25] It turns out there are some jumps present in the data.

Figure 3.2: Jumps present in the price process (1% confidence level)



In figure 3.2, we depict the observations that can be regarded as jumps as tested on the 1% confidence level. They constitute about 7 percent of the total number of observations. However, their contribution to the total sum of quadratic variation (i.e. analogous to $JV_{t,h}$ for all periods) is only about 2 percent.[26] Thus, the total quadratic variation is mainly made up of the integrated variance part.

Turning to the right-hand side of equation (3.21), we use the aforementioned VIX index for oil (OVX) as a measure of implied volatility. The data on the

[25]See, for example, Barndorff-Nielsen & Shephard (2001), Barndorff-Nielsen & Shephard (2004) or Barndorff-Nielsen *et al.* (2005).

[26]Compared to, respectively, 3 and 6 percent for 5% and 10% confidence levels.

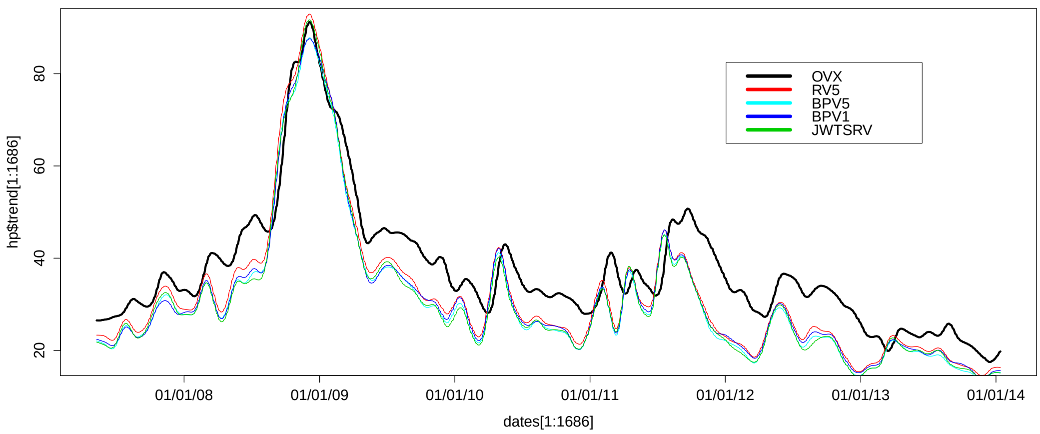
OVX index is publicly available.[27] We use daily (excluding Saturdays and Sundays) closing prices for the OVX index starting from the 10^{th} of May 2007 until the 15^{th} of January 2014 summing to 1686 observations overall.

As we already know, the OVX index is a forecast of 30-day future realized volatility. This means that our daily realized volatilities, which we calculated above, are not yet directly comparable to it. To make them comparable to the OVX index, we need to transform them into the 30-day realized volatilities first. The correct methodology for this is presented by Carr & Wu (2006). We use a slightly amended formula

$$RV_{t+h,h} = 100 \times \sqrt{\frac{252}{21} \sum_{i=0}^{20} r_{t+i}^2}, \tag{3.22}$$

where r_t stands for daily realized volatility on day t. Number 252 roughly corresponds to the number of trading days (no weekends) in a year. Likewise, 21 is the number of trading days in an average 30-day period. Thus, we simply applied formula (3.22) on our four estimators of daily realized volatility to obtain the left-hand side of equation (3.21). Again, the smoothed results of this procedure are depicted in figure 3.3.[28] Of course, for further analysis we only use raw data (can be found in the appendix).

Figure 3.3: Smoothed daily OVX index compared with different realized volatility estimators



From the smoothed data, we can already see that the VIX index tends to, perhaps, overestimate realized volatility. Furthermore, for some time periods

[27] We downloaded the data from (www.yahoo.finance.com).

[28] We smoothed the data for strictly illustrative purposes.

there is a visible phase shift where the OVX index is somewhat lagging behind the realized volatility estimators. However, these observations need to be further analyzed.

3.1.6 Band Selection (Wavelet Coherence)

To get a better picture of the relationship between realized and implied volatilities, we would like to use wavelet band spectrum regression techniques described in section 2.5. One problem we need to solve before doing the analysis is the selection of bands (see equations (2.29) and (2.30)). Ideally, we would like to select bands, for which the relationship between realized and implied volatilities will be as strong as possible.[29] This is not easy to do without any apriori knowledge.

An indispensable tool in such cases is the so-called wavelet coherence. Wavelet coherence between two time series X and Y is, to put it simply, a normalized smoothed version of the product of two continuous wavelet transforms $W_X(\lambda, t)$ and $W_Y(\lambda, t)$ i.e. a version of

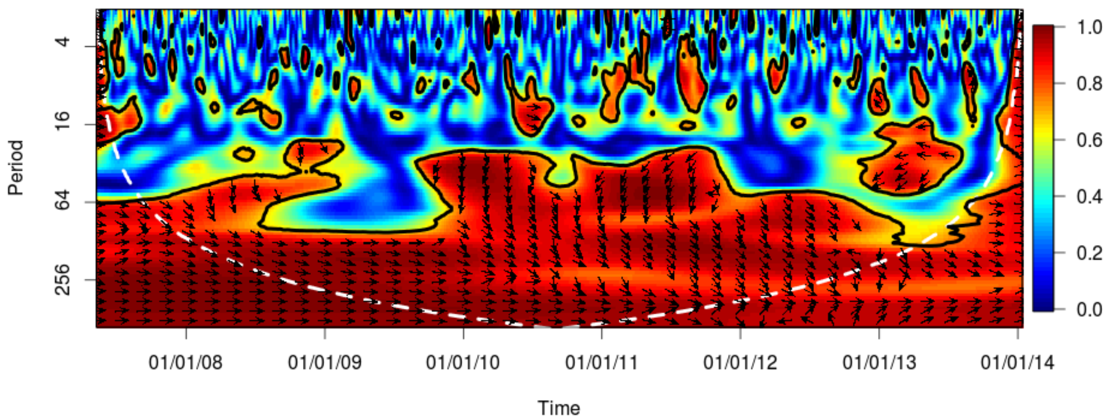
$$W_{XY}(\lambda) = W_X(\lambda, t)\overline{W_Y(\lambda, t)}, \tag{3.23}$$

where $\overline{W_Y(\lambda, t)}$ is the complex conjugate of $W_Y(\lambda, t)$. For each point (λ, t) in the time-frequency (or time-scale) domain the wavelet coherence is a value between zero and one. This essentially provides a localized time-frequency measure of correlation (co-movement) between two time series. It serves as a very powerful tool in discovering complex relationships or patterns in the data. The complete theoretical framework behind wavelet coherence is rather technical and will be omitted here. For technical details, the reader can refer to Torrence & Compo (1998),Torrence & Webster (1999) and Grinsted *et al.* (2004).

Probably the best way to demonstrate the power of wavelet coherence is by showing a concrete example. In figure 3.4, there is a plot of wavelet coherence between the OVX index and the RV5 estimator of realized volatility.

[29] In the sense that the estimate of β should be close to one.

Figure 3.4: Wavelet coherence between the OVX index and the RV5 estimator



The red areas, marked by the thick black contours,[30] show the time-scale points with the highest interactions between the two time series. Precisely in these areas we can expect the estimates of β from equation (3.21) to be the highest. In our case, most of strong interaction takes place in the region where the scale is 32 days and more (i.e. for scale levels starting from five and higher). Thus, for the wavelet band spectrum regression it would be sensible to select bands starting with level 5. This is a similar result to the one obtained by Baruník & Baruníková (2013). Comparing their wavelet coherence plots with figure 3.4, we conclude that, for the given time periods, realized and implied volatilities of crude oil exhibit similar wavelet coherence structure as in the case of S&P 500 and DAX indices. On the other hand, low values of wavelet coherence for scales lower than 32 days indicate that the realized-implied relationship is not very strong in the short-run. In other words, the OVX index does not carry much information about future realized volatility when we consider relatively short time spans.

Wavelet coherence plot can also give us localized information about the leading/lagging relationships of two time series. This information is coded via the orientation of the arrows seen in figure 3.4. The angle at which the arrow is pointing is meant to quantify the phase difference between the first and the second time series at a specific point in the time-scale space. An arrow

[30] Outside of these contours the coherence is not significantly different from Brownian noise. See Torrence & Compo (1998) and Grinsted *et al.* (2004) for more details about the technical side.

pointing to the right (left) means that the two series are in phase (anti-phase). An upward pointing arrow indicates that the first time series leads the second one by $\pi/2$ and vice versa for a downward sloping arrow.

In our case, the arrows are mostly pointing to the right, especially for the first half of the sample, meaning that the OVX index and RV5 estimator are in phase. This is generally a desired property for the OVX index. However, there are certain time intervals in the second half of the sample, for which the arrows are pointing downwards. This is indicating that the RV5 estimator is leading the OVX index at these points. In other words, the OVX is adapting to already observed realized volatility, which is probably not a good sign regarding the performance of the OVX index since it should forecast future volatility and not merely follow historical realized volatility.

Wavelet coherences between the OVX and the other estimators of realized volatility are plotted in figure A.8 in the appendix. Since the estimators do not differ much among themselves, there is no significant difference among the wavelet coherences either and the wavelet coherence from figure 3.4 can be taken as a representative for all.

3.1.7 Results

Armed with the plot of wavelet coherence from figure 3.4, we proceed to the estimation of β from equation (3.21). After little bit of experimenting, we decided to use the Haar MODWT with $J = 7$ for the WBSR estimation. This is a larger J than in Baruník & Baruníková (2013), but we also have a larger dataset. We compute the WBSR estimates of β for bands $[1, 2]$, $[3, 4]$ and $[5, 7]$ using formula (2.30). For the calculation of standard errors and intercepts we run standard OLS regressions between wavelet coefficients,[31] for which the OLS estimate of β is equivalent to formula (2.30). We also provide classical OLS estimates with and without intercept. The results are presented in table 3.1 below.

We can see that for scales 1 and 2 (1-2 days and 2-4 days) there is very little information content (about 3%) in the OVX index for all considered estimators of realized volatility. This improves a little bit for scales 3 and 4, but it is

[31] Regressions have the form $Y \sim X$, where Y are wavelet coefficients of realized volatility corresponding to the scales of the chosen band. Similarly, X stands for wavelet coefficients of the OVX index for the chosen band. It is not difficult to verify that the OLS estimate for β for such regression is equivalent to formula (2.30) when using formula (2.18) for the estimation of wavelet variance and covariance.

Table 3.1: Table showing the results of estimating β with various methods and for various realized volatility estimators. Standard errors are shown below the estimates.

	RV5		BPV5		BPV1		JWTSRV	
	α	β	α	β	α	β	α	β
WBSR(1,2)	0.000	0.035	0.000	0.032	0.000	0.031	0.000	0.030
	(0.007)	(0.006)	(0.007)	(0.006)	(0.007)	(0.006)	(0.007)	(0.006)
WBSR(3,4)	-0.000	0.166	-0.000	0.149	-0.000	0.154	-0.000	0.144
	(0.021)	(0.013)	(0.021)	(0.013)	(0.020)	(0.013)	(0.021)	(0.013)
WBSR(5,7)	0.000	0.927	0.000	0.890	-0.000	0.906	-0.000	0.931
	(0.033)	(0.009)	(0.033)	(0.009)	(0.033)	(0.009)	(0.034)	(0.010)
OLS	-5.866	1.015	-6.015	0.978	-5.612	0.980	-6.700	(0.999)
	(0.406)	(0.010)	(0.398)	(0.009)	(0.392)	(0.009)	(0.410)	(0.010)
OLS(no intercept)	-	0.879	-	0.838	-	0.850	-	0.844
	-	(0.003)	-	(0.004)	-	(0.004)	-	

only for the higher scales (5-7) that we can see strong co-movement of realized volatility and the OVX index. The highest β is achieved by using the JWT-SRV estimator, followed closely by the RV5 estimator. The BPV estimators produced slightly lower β.

This result is not surprising, given the wavelet coherence graph above. It is, perhaps, also a natural result because the OVX index is designed to capture 30-day volatility and, on a scale of 32 days and higher, it does so fairly well. For the small scales, there might be an effect of market frictions and other noise that distorts the realized-implied volatility relationship.

The OLS estimates show OVX to be biased estimator of realized volatility. Interestingly enough, the bias comes solely from the intercept term with β being non-significantly different from one for all estimators. This is in contrast with the result from Baruník & Baruníková (2013) where α's were close to zero and the bias came mostly from the β coefficients. In this respect, the OVX index behaves more like the Nasdaq 100 volatility index (VXN) between 1995 and 2002, for which the OLS results are similar as is documented in Corrado & Miller Jr (2005). However, when running an OLS regression without the intercept term, the estimates of β drop below the WBSR(5,7) estimates as was the case in Baruník & Baruníková (2013).

3.2 Analysis of the Relationship between Labor Productivity and Unemployment

In this section, we shall use wavelet coherence and wavelet band spectrum regression techniques to analyze the relationship between unemployment and labor productivity growth for four largest European economies. These are Germany, United Kingdom, France and Italy. As before, the key focus of the analysis is to break down the relationship into different scales, which should provide interesting insights into the phenomena. We shall reflect on the similarities and differences among the four analyzed countries, also comparing them to the US economy.

3.2.1 Labor Productivity and Unemployment

Although the relationship between labor productivity and unemployment has been subject to intensive research, there is no clear-cut consensus on its nature. Most of the research is focused on analyzing the effect of labor productivity growth on unemployment, however. Since the times of David Ricardo there has been fear of productivity-induced unemployment through the process of 'creative destruction'. On the other hand, over the passing decades technological progress has brought increased prosperity and well-being, with positive long-term effects on employment.

Theoretical works on the relationship have produced ambiguous results. As mentioned by Gallegati *et al.* (2015), there is a contrast between real business cycle models and new Keynesian sticky-prices models. The former predict a negative short-run effects of productivity on unemployment, whereas the latter imply positive short-run effects and negative long-run effects. Another branch of theory trying to explain the relationship are the search models studied by, for example, Aghion & Howitt (1994) or Pissarides (1990). They suggest that productivity growth acts on unemployment through two different opposing channels. One was the 'capitalization' channel, where increased productivity "raises the capitalized returns from creating jobs and consequently reduces the equilibrium rate of unemployment."[32] On the other hand, there is the already mentioned 'creative destruction' channel, which tends to increase unemployment rate through substituting labor for capital. It is then the relative strength of each channel that determines the overall effect of productivity

[32] From Aghion & Howitt (1994), page 1.

growth on unemployment. Interestingly, this relative strength is then determined by the nature of productivity growth (or technological change), with embodied technological change favoring 'creative destruction' and disembodied change favoring 'capitalization'.[33] They also mentioned the importance of possibly different time horizons for these channels to take effect.

One of the first papers that stressed the importance of time horizons when modelling the relationship between labor productivity and unemployment was Blanchard *et al.* (1995), distinguishing between short-run, medium-run and long-run. Similar ideas were expressed by Solow (2000) and/or Landmann *et al.* (2004):

> "The nature of the mechanism that link [unemployment and productivity growth] changes with the time frame adopted" because one needs "to distinguish between an analysis of the forces shaping long-term equilibrium paths of output, employment and productivity on the one hand and the forces causing temporary deviations from these equilibrium paths on the other hand"[34]

This makes the relationship between productivity and unemployment well-suited for analysis by spectral methods, which are able to perform the decomposition into different frequencies/scales. In Tripier (2005), the relationship for the US is studied using Fourier spectral methods. They found strong co-movements between labor productivity growth and unemployment for the medium- and long-term time horizons. Moreover, the relationship was positive for short- and long-term and negative for medium term (business cycle frequency).

More recently, Gallegati *et al.* (2014) used the wavelet methodology[35] in the analysis of the relationship for the US economy for the post-war period. Later, the same analysis was also done for the G7 countries by Gallegati *et al.* (2015). In both cases, authors found a positive relationship in the short- and medium-run and a negative relationship in the long-run. They also suggested (implicitly) that, for the European countries during the 90's, the relationship might have been weak for all considered scales since there was a period of sustained productivity growth accompanied with stagnating unemployment.

[33] Embodied technological change is more tangible and one that can be taken advantage of only via investing e.g. new machinery. Disembodied technological change is less tangible and can be enjoyed throughout the society e.g. the internet.

[34] From Landmann *et al.* (2004), page 35.

[35] Including wavelet coherence and a modification of wavelet band spectrum regression.

Here we try to confirm or disprove this statement by analyzing the relationship between labor productivity and unemployment for four largest European economies. Again, in our analysis we make use of wavelet techniques, namely wavelet coherence and wavelet band spectrum regression.

3.2.2 Data

For the analysis, we use quarterly data from the OECD website.[36] The period covered is from the first quarter of 1991 until the last quarter of 2014. The countries are Germany, UK, France and Italy.

As a measure of labor productivity we use seasonally adjusted index of GDP per person employed. The data is indexed with base year being 2010. We further calculate the quarterly growth of this index and express it in percents. For the unemployment we use harmonized unemployment rate (in percents) as can be found on the OECD website.[37] Although the traditional way to analyze the relationship is to compare labor productivity growth to unemployment rate, we also try to compare labor productivity growth with *growth* of unemployment rate. Accordingly, we also compute the *growth* rate of unemployment for all four countries measured in percentage points.

3.2.3 Wavelet Coherence

As in the previous application, we proceed by calculating the wavelet coherence between unemployment and labor productivity growth (in this order) for all four countries in consideration. The coherences are shown in figure 3.5. More detailed pictures of the coherences can be found in the appendix. We give a brief summary for each country.

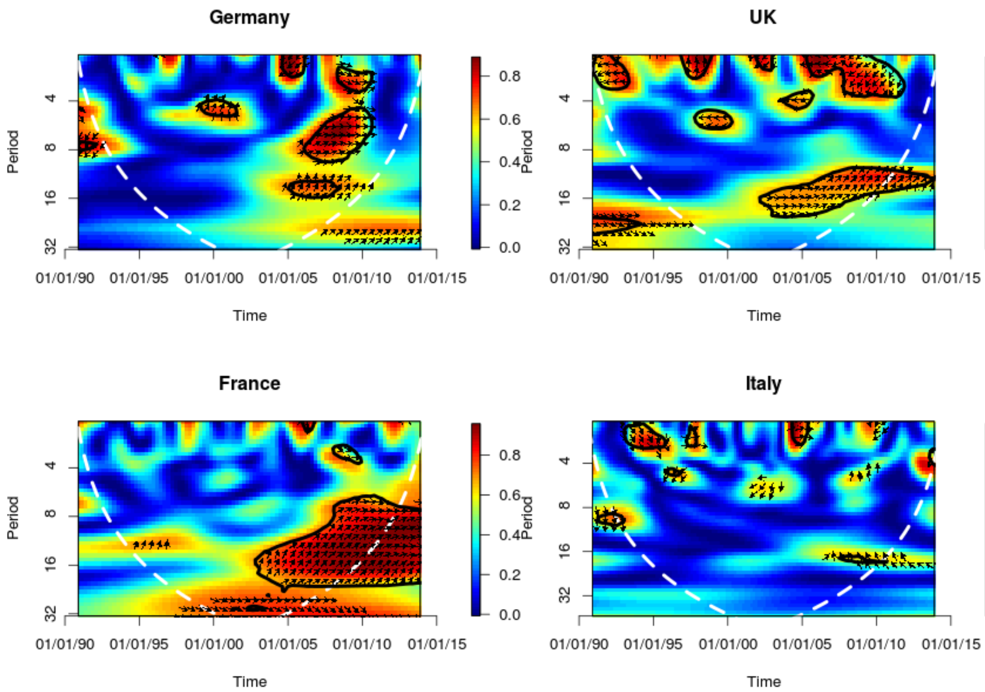
For Germany, there is no consistent overall (for all times) co-movement of the two variables at any scales. We can observe just little patches of co-movements in the 2000's. In the 90's, there is almost no relationship found at any scales. Also, the phase difference arrows are pointing predominantly to the right for all time scales, which hints that there is a positive relationship (variables are more in phase than out of phase).[38]

[36] The web page is (stats.oecd.org).

[37] For details on the methodology, please use the OECD website.

[38] In our observations, we exclude the data that is outside of the cone of influence marked by the white dashed line. These observations are influenced by zero padding, which is a technical issue in the calculation of wavelet coherence. See the section on wavelet coherence above for relevant literature.

Figure 3.5: Wavelet coherences between unemployment and labor productivity growth



For the UK, we do not observe a consistent overall co-movement over any scale. However, there is much more interaction in the smaller scales than in the larger scales. Surprisingly, the arrows are indicating more of a negative relationship in the short-run and positive one in the medium-run or long-run. For the period of the 90's, there is no long-run interaction between the two variables.

In the case of France, the 90's again do not show any significant co-movements at any scales. There is a significant interaction for the medium-run (8-16 quarters) starting from the year 2000. The arrows are leaning upwards and are fairly inconclusive for the positive/negative judgment of the relationship.

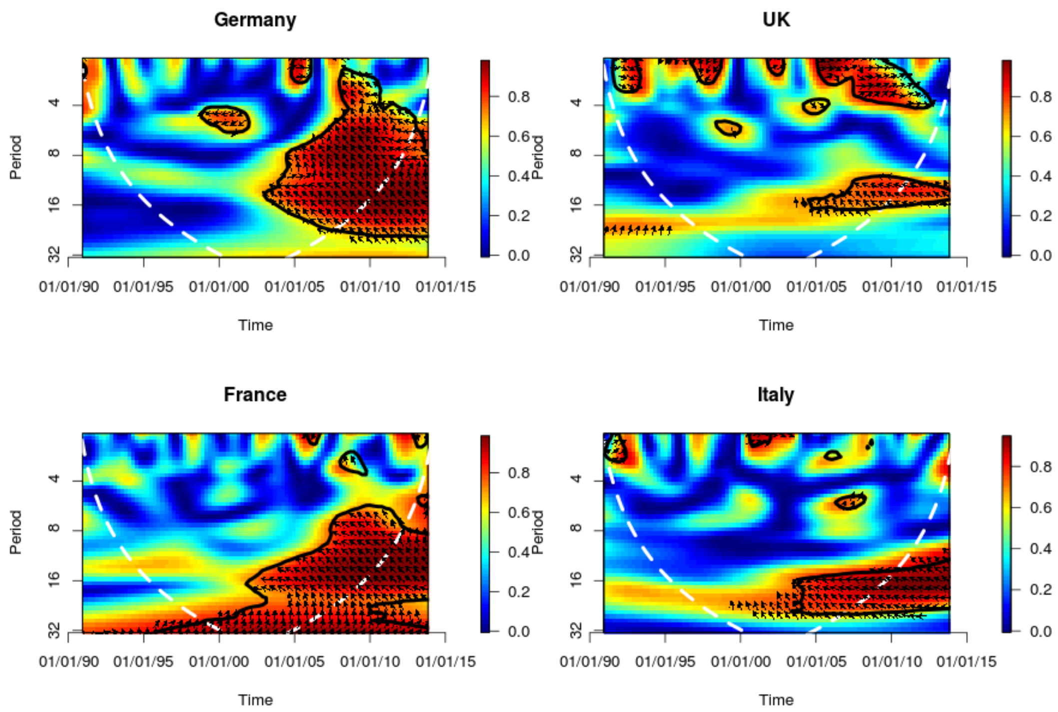
Finally, Italy shows very little coherence overall, except from few small patches on the smallest scale. It is not possible to make any conclusive observation, except from the fact that there is probably no co-movement of the two variables for any scale and any time, including the 90's.

If we had to summarize the findings, we would probably say that wavelet coherences provide an inconclusive picture of the relationship between unemployment and labor productivity growth for the considered countries. However, one common feature for all four countries is the lack of interaction during the

90's, which is consistent with the observation made in Gallegati *et al.* (2014).

Since the wavelet coherences do not provide much information about the relationship, we decided to also model the relationship between labor productivity growth and the *growth* of unemployment. The rationale behind this is that the unemployment rate itself can be influenced by a plethora of external influences that are not connected with labor productivity. This might then blur the relationship considerably. However, labor productivity growth could still directly influence the *growth* rate of unemployment (the slope), independently of external factors. We computed the wavelet coherences between labor productivity growth and the *growth* of unemployment. They are depicted in figure 3.6 below. We again make some brief observations.

Figure 3.6: Wavelet coherences between growth of unemployment and labor productivity growth



Germany shows little interaction for the 90's. However, there is a big patch covering short-term and medium-term scales for the period between 2005 and 2010. The plot for the UK is similar to the one before, except that the arrows are now pointing in different directions, which is less surprising than before. France also shows a similar structure than before, but arrows are now leaning more to the left, indicating a shift towards anti phase in the medium-run (negative relationship). For Italy, there starts to be an interaction in the medium-run for the period 2005 to 2015. The arrows are leaning more towards a negative relationship. Overall, there is again little interaction in the 90's.

3.2.4 Regression Results

The wavelet coherences serve only as a 'lens' into the relationship and do not provide any quantitative analysis. Therefore, we again use the wavelet band spectrum regression technique to quantify the relationship at different scales. Unemployment is taken to be the dependent variable. We use the Haar wavelet as before, but this time the number of levels is only four because of the smaller dataset we have. We calculate WBSR(1,2) and WBSR(3,4) estimates for each country. This is corresponding to short-run and medium-run estimates of the relationship. For the long-run estimate, we use the OLS estimate for the V5 vectors of the two series,[39] corresponding to scales $J = 5$ and higher. Finally, we also compute the classical OLS estimator. The results for the two relationships are presented in tables below. The standard errors are in parentheses and bold font indicates estimates that are significantly different from zero.

Table 3.2: Table showing the results of estimating relationship between unemployment and labor productivity growth

	Germany		**UK**		**France**		**Italy**	
	α	β	α	β	α	β	α	β
WBSR(1,2)	0.000	**0.195**	0.000	**0.402**	0.000	**0.460**	0.000	**0.446**
	(0.012)	(0.030)	(0.013)	(0.045)	(0.010)	(0.045)	(0.019)	(0.052)
WBSR(3,4)	0.000	**0.835**	-0.000	**0.924**	-0.000	**0.961**	-0.000	**0.881**
	(0.022)	(0.042)	(0.020)	(0.044)	(0.017)	(0.044)	(0.029)	(0.051)
V5	7.657	1.317	**6.171**	**1.66**	**8.362**	**6.889**	**9.127**	**2.873**
	(0.257)	(0.974)	(0.250)	(0.506)	(0.280)	(1.073)	(0.034)	(0.473)
OLS	**7.889**	0.228	**6.590**	**0.625**	**9.795**	**0.855**	**9.339**	0.507
	(0.178)	(0.215)	(0.201)	(0.271)	(0.174)	(0.376)	(0.182)	(0.258)

From table 3.2 above we see, similarly to the case of the US[40], that the relationship between unemployment and productivity growth is significantly positive in the short-run with β increasing for the-medium run. In this respect, the four European countries do not differ with the USA. For the long run estimate V5 there is a stark difference though, with all estimates being positive even in the long-run. This suggests that for European countries the relationship might be positive for all time scales meaning that the 'creative destruction' effect prevails uniformly over the 'capitalization' effect (from the perspective of

[39] See Gallegati *et al.* (2014) on how this is done.
[40] See Gallegati *et al.* (2014).

theory). Interestingly, all four countries show similar structure. OLS estimates are marginally positive as well.

Table 3.3: Table showing the results of estimating relationship between the growth of unemployment and labor productivity growth

	Germany		**UK**		**France**		**Italy**	
	α	β	α	β	α	β	α	β
WBSR(1,2)	0.000	**0.760**	0.000	**1.025**	0.000	**0.907**	0.000	**0.986**
	(0.056)	(0.067)	(0.075)	(0.068)	(0.063)	(0.083)	(0.095)	(0.095)
WBSR(3,4)	0.000	**0.876**	-0.000	**0.924**	-0.000	**0.891**	-0.000	**0.854**
	(0.070)	(0.068)	(0.095)	(0.089)	(0.068)	(0.090)	(0.069)	(0.084)
V5	**-1.453**	**6.714**	**0.807**	**-2.692**	0.550	-1.646	**9.127**	0.124
	(0.249)	(0.943)	(0.266)	(0.538)	(0.293)	(1.129)	(0.177)	(0.618)
OLS	0.085	-0.485	-0.147	-0.331	0.520	**-1.518**	0.495	-0.117
	(0.312)	(0.377)	(0.440)	(0.594)	(0.300)	(0.646)	(0.342)	(0.486)

When comparing the *growth* of unemployment and labor productivity growth we observe a few differences (see table 3.2). First of all, the positive relationship is more pronounced in the short-run than before with the medium-run results being similar. More importantly, with the exception of Germany, the long-run positive relationship is either less pronounced or changes to a negative one. This is consistent with the belief that productivity growth has a positive influence on employment (i.e. negative on unemployment) in the long-run. The case of Germany is clearly an outlier and needs further explanation.[41] The OLS estimates are also all negative (even for Germany) compared to the positive ones before.

Overall, it is interesting that there is a difference between long-run relationship when measuring unemployment rate and *growth* of unemployment rate. This might indicate a possible strong long-run external factor acting positively on the rate of unemployment. This factor is then able to offset the influence of labor productivity. However, it is a task for further research to find this factor (or other explanation) and explain the observed difference.

[41] Germany was quite special because of the unification in the beginning of the 90's. A possible explanation could come from this fact.

Chapter 4

Conclusion

In this thesis, we introduce wavelets and wavelet-based spectral techniques in a manner that should be understandable (we hope) to a wide range of readers. The most important technique is the wavelet band spectrum regression, which allows us to decompose economic relationships into different time horizons. We demonstrate the usefulness of wavelet methods through the analysis of the realized-implied volatility relationship for the crude oil and also the analysis of the relationship between unemployment and labor productivity for four major European economies during the period 1991-2015.

In the first analysis, we compare different measures of realized volatility to the VIX index of oil, which is a measure of implied volatility. With the help of wavelet coherence, we find that the two series show significant co-movements only for scales of 32 days and higher, a result that is further confirmed by wavelet band spectrum regression analysis. This is consistent with the findings of Baruník & Baruníková (2013), where a similar analysis is done for the S&P 500 and DAX indices. All results show that the VIX (OVX) index is a biased estimator of realized volatility. The results of the classical OLS estimation are different to the ones in Baruník & Baruníková (2013), but more similar to the results for the NASDAQ 100 volatility index between 1995 and 2002.

For the second analysis, we use the same techniques to examine the relationship between unemployment and labor productivity growth of Germany, UK, France and Italy. We compare the results with the case of USA and the G7 area studied by Gallegati *et al.* (2014) and Gallegati *et al.* (2015). Wavelet coherences show mixed results, with the common feature of low co-movements of the two variables during the 1990's, a result consistent with remarks made in Gallegati *et al.* (2014). Wavelet band spectrum regressions show similarity with the USA only for lower scales where unemployment is positively correlated with labor productivity growth. Unlike the USA and the G7 area, the relationship is also positive for the long-term scales, indicating a possibility of strong 'creative destruction' element in Europe for the given period. With the exception of Germany, the results begin to show similar structure to the USA and G7 area only when we use growth rate of unemployment for the analysis. An economic explanation of these findings is, perhaps, a good motivation for further research.

Bibliography

Aghion, P. & P. Howitt (1994): "Growth and unemployment." *Review of Economic Studies* **61(3)**: pp. 477–94.

Andersen, T. & L. Benzoni (2008): "Realized volatility." *Working Paper Series WP-08-14*, Federal Reserve Bank of Chicago.

Andersen, T. & T. Bollerslev (1998): "Answering the skeptics: Yes, standard volatility models do provide accurate forecasts." *International Economic Review* **39(4)**: pp. 885–905.

Andersen, T., T. Bollerslev, & F. Diebold (2003): "Some like it smooth, and some like it rough: Untangling continuous and jump components in measuring, modeling, and forecasting asset return volatility." *Pier working paper archive*, Penn Institute for Economic Research, Department of Economics, University of Pennsylvania.

Andersen, T., T. Bollerslev, F. Diebold, & P. Labys (2001): "Modeling and forecasting realized volatility." *NBER Working Papers 8160*, National Bureau of Economic Research, Inc.

Andersen, T., T. Bollerslev, & X. Huang (2007): "A reduced form framework for modeling volatility of speculative prices based on realized variation measures." *Creates research papers*, School of Economics and Management, University of Aarhus.

Back, K. (1991): "Revisiting the long memory dynamics of implied-realized volatility relation:A new evidence from wavelet band spectrum regression." *Journal of Mathematical Economics 20, p. 317-395* .

Bandi, F. & B. Perron (2003): "Long memory and the relation between implied and realized volatility." *Econometrics*, EconWPA.

Barndorff-Nielsen, O. E., S. E. Graversen, J. Jacod, & N. Shephard (2005): "Limit theorems for bipower variation in financial econometrics." *Ofrc working papers series*, Oxford Financial Research Centre.

Barndorff-Nielsen, O. E. & N. Shephard (2001): "Econometric analysis of realised volatility and its use in estimating stochastic volatility models." *Economics Series Working Papers 71*, University of Oxford, Department of Economics.

Barndorff-Nielsen, O. E. & N. Shephard (2003): "Power and bipower variation with stochastic volatility and jumps." *Economics Papers 2003-W17*, Economics Group, Nuffield College, University of Oxford.

Barndorff-Nielsen, O. E. & N. Shephard (2004): "Econometrics of testing for jumps in financial economics using bipower variation." *Ofrc working papers series*, Oxford Financial Research Centre.

Baruník, J. (2011): *Wavelet-based Realized Variation and Covariation Theory.* Ph.D. thesis, Charles University in Prague, Institute of Economic Studies.

Baruník, J. & M. Baruníková (2013): "Asset pricing for general processes." *IES working paper* .

Baruník, J., T. Křehlík, & L. Vácha (2015): "Modeling and forecasting exchange rate volatility in time-frequency domain." *Papers*, arXiv.org.

Baruník, J. & L. Vácha (2013): "Realized wavelet-based estimation of integrated variance and jumps in the presence of noise." *Papers*, arXiv.org.

Bauwens, L., C. Hafner, & S. Laurent (2011): "Volatility models." *CORE Discussion Papers 2011058*, Université catholique de Louvain, Center for Operations Research and Econometrics (CORE).

Black, F. & M. Scholes (1973): "The pricing of options and corporate liabilities." *Journal of Political Economy* **81(3)**: pp. 637–54.

Blanchard, O., R. Solow, & B. A. Wilson (1995): "Productivity and unemployment." *Massachusetts Institute of Technology. Mimeo* .

Britten-Jones, M. & A. Neuberger (2000): "Option prices, implied price processes, and stochastic volatility." *Journal of Finance* **55(2)**: pp. 839–866.

CARR, P. & L. WU (2006): "A Tale of Two Indices." *Journal of Derivatives* **13**: pp. 13–29.

CBOE (2003): "The vix white paper." `http://www.cboe.com/micro/vix/vixwhite.pdf`.

CHRISTENSEN, B. J. & C. HANSEN (2002): "New evidence on the implied-realized volatility relation." *The European Journal of Finance* **8(2)**: pp. 187–205.

CHRISTENSEN, B. J. & N. R. PRABHALA (1998): "The relation between implied and realized volatility." *Journal of Financial Economics* **50(2)**: pp. 125–150.

CORRADO, C. J. & T. W. MILLER JR (2005): "The forecast quality of cboe implied volatility indexes." *Journal of Futures Markets* **25(4)**: pp. 339–373.

DEMETERFI, K., E. DERMAN, M. KAMAL, & J. ZOU (1999): "More than you ever wanted to know about volatility swaps." *Goldman Sachs Quantitative Strategies Research Notes* .

ENGLE, R. & C. GRANGER (1987): "Co-integration and error correction: Representation, estimation, and testing." *Econometrica* **55(2)**: pp. 251–76.

ENGLE, R. F. (1972): "Band Spectrum Regressions." *Working papers 96*, Massachusetts Institute of Technology (MIT), Department of Economics.

FADILI, J. M. & E. T. BULLMORE (2002): "Wavelet-Generalized Least Squares: A New BLU Estimator of Linear Regression Models with 1/f Errors." *NeuroImage 15, p. 217–232* .

FAN, J. & Y. WANG (2007): "Multi-scale jump and volatility analysis for high-frequency financial data." *Journal of the American Statistical Association* **102**: pp. 1349–1362.

GALLEGATI, M., M. GALLEGATI, J. RAMSEY, & W. SEMMLER (2014): "Does productivity affect unemployment? a time-frequency analysis for the us." In M. GALLEGATI & W. SEMMLER (editors), "Wavelet Applications in Economics and Finance," volume 20 of *Dynamic Modeling and Econometrics in Economics and Finance*, pp. 23–46. Springer International Publishing.

GALLEGATI, M., M. GALLEGATI, J. B. RAMSEY, & W. SEMMLER (2015): "Productivity and unemployment: a scale-by-scale panel data analysis for the g7 countries." *Studies in Nonlinear Dynamics & Econometrics* .

Gallegati, M. & W. Semmler (2014): *Wavelet Applications in Economics and Finance*. Dynamic modeling and econometrics in economics and finance. Springer.

Gençay, R., F. Selçuk, & B. Whitcher (2009): *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics*. San Diego: Academic Press, Elsevier.

Granger, C. W. & O. Morgenstern (1963): "Spectral analysis of new york stock market prices." *Kyklos* **16(1)**: pp. 1–27.

Granger, C. W. J., M. Hatanaka *et al.* (1964): "Spectral analysis of economic time series." *Spectral analysis of economic time series.* .

Grinsted, A., J. C. Moore, & S. Jevrejeva (2004): "Application of the cross wavelet transform and wavelet coherence to geophysical time series." *Nonlinear Processes in Geophysics* **11(5/6)**: pp. 561–566.

Huang, X. & G. Tauchen (2005): "The relative contribution of jumps to total price variance." *Journal of Financial Econometrics* **3(4)**: pp. 456–499.

Jiang, G. J. & Y. S. Tian (2005): "The model-free implied volatility and its information content." *Review of Financial Studies* **18(4)**: pp. 1305–1342.

Jiang, G. J. & Y. S. Tian (2006): "Extracting Model-Free volatility from option prices: An examination of the vix index." *Social Science Research Network Working Paper Series* .

Kallenberg, O. (2002): *Foundations of Modern Probability*. Applied probability. Springer.

Kellard, N., C. Dunis, & N. Sarantis (2010): "Foreign exchange, fractional cointegration and the implied-realized volatility relation." *Journal of Banking & Finance* **34(4)**: pp. 882–891.

Křehlík, T. (2013): *Does wavelet decomposition and neural networks help to improve predictability of realized volatility?* Master's thesis, Charles University in Prague, Institute of Economic Studies.

Landmann, O. *et al.* (2004): "Employment, productivity and output growth." *Employment Strategy Papers* **17**: pp. 1–61.

Mallat, S. (1999): *A Wavelet Tour of Signal Processing.* Academic Press.

Nielsen, M. & P. Frederiksen (2010): "Fully modified narrow-band least squares estimation of weak fractional cointegration." *Creates research papers*, School of Economics and Management, University of Aarhus.

Percival, D. B. & A. T. Walden (2000): *Wavelet Methods for Time Series Analysis.* Cambridge University Press.

Pissarides, C. A. (1990): *Equilibrium unemployment theory.* MIT press.

Solow, R. M. (2000): "Toward a macroeconomics of the medium run." *The Journal of Economic Perspectives* pp. 151–158.

Torrence, C. & G. P. Compo (1998): "A practical guide to wavelet analysis." *Bulletin of the American Meteorological Society* **79(1)**: pp. 61–78.

Torrence, C. & P. J. Webster (1999): "Interdecadal changes in the enso-monsoon system." *Journal of Climate* **12(8)**: pp. 2679–2690.

Tripier, F. (2005): "Sticky prices, fair wages, and the co-movements of unemployment and labor productivity growth." *Macroeconomics 0510015*, EconWPA.

Whitcher, B., P. Guttorp, & D. Percival (2000): "Wavelet analysis of covariance with application to atmospheric time series." *NRCSE Technical Report Series* .

Zhang, L., P. A. Mykland, & Y. Ait-Sahalia (2005): "A tale of two time scales: Determining integrated volatility with noisy high-frequency data." *Journal of the American Statistical Association* **100**: pp. 1394–1411.

Appendix A

Additional figures

Figure A.1: RV5 estimator computed for each day

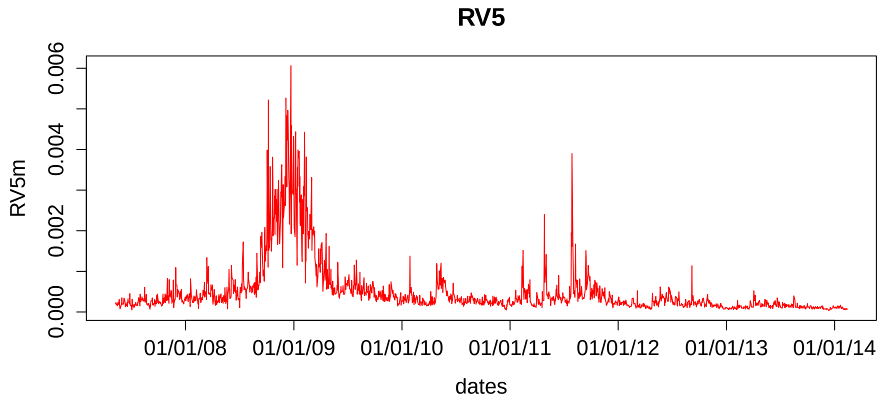


Figure A.2: BPV5 estimator computed for each day

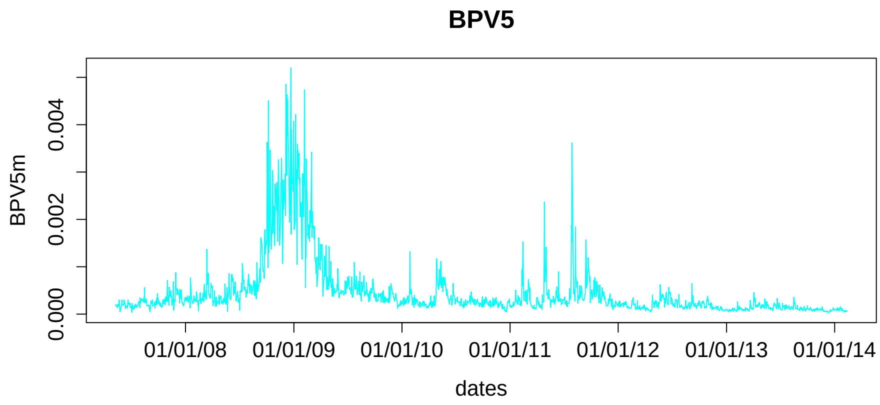


Figure A.3: BPV1 estimator computed for each day

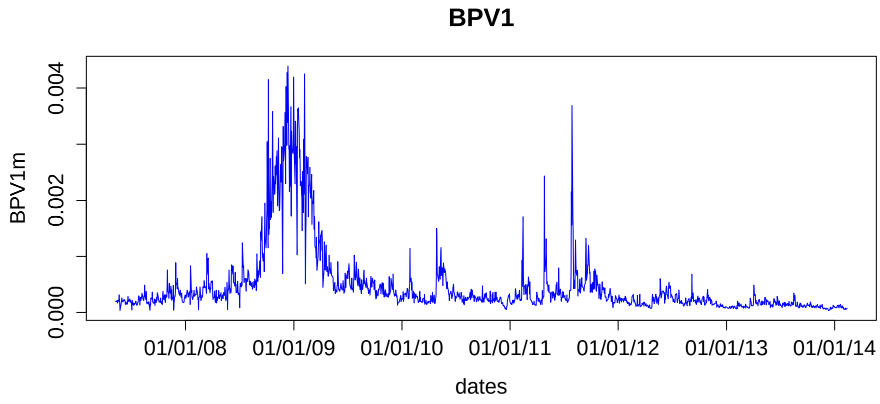


Figure A.4: JWTSRV estimator computed for each day

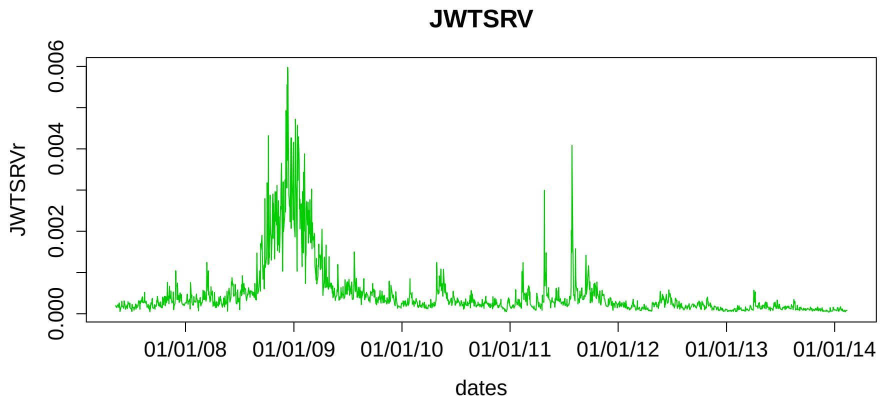


Figure A.5: All estimators in one graph

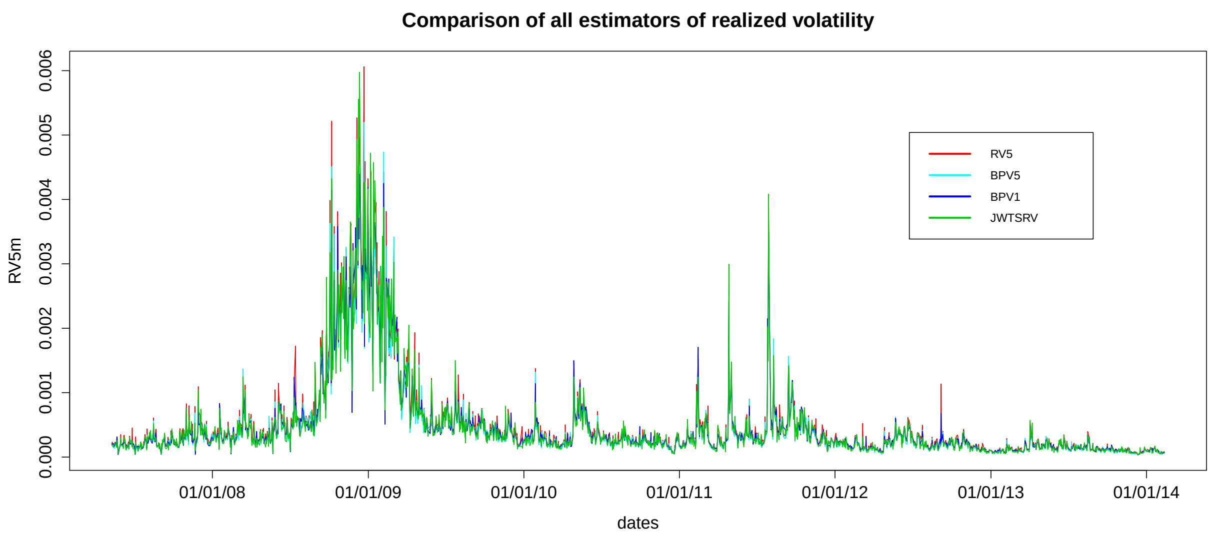


Figure A.6: Decomposition of realized volatility using wavelets

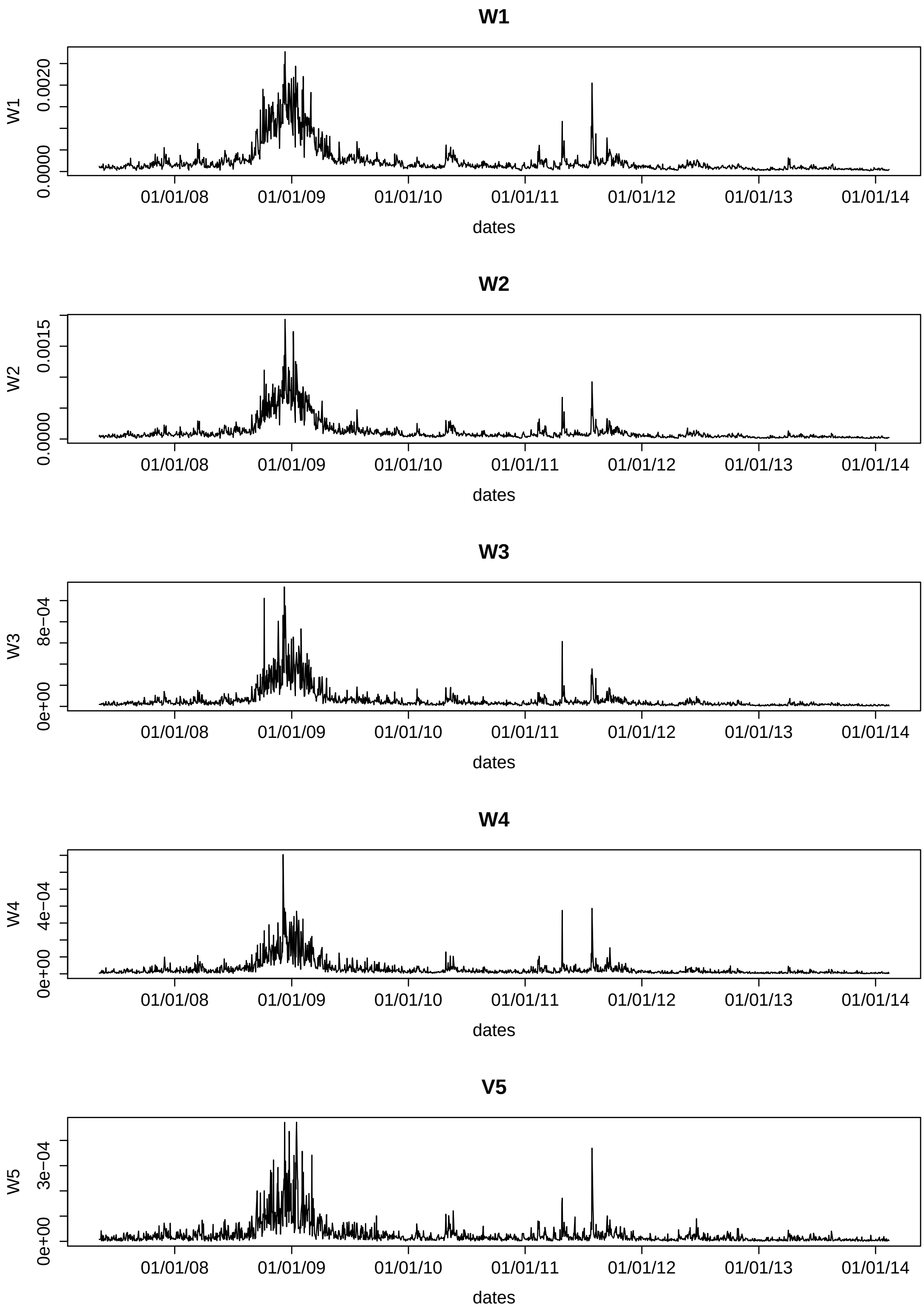


Figure A.7: Raw data on the OVX index and realized volatility estimators

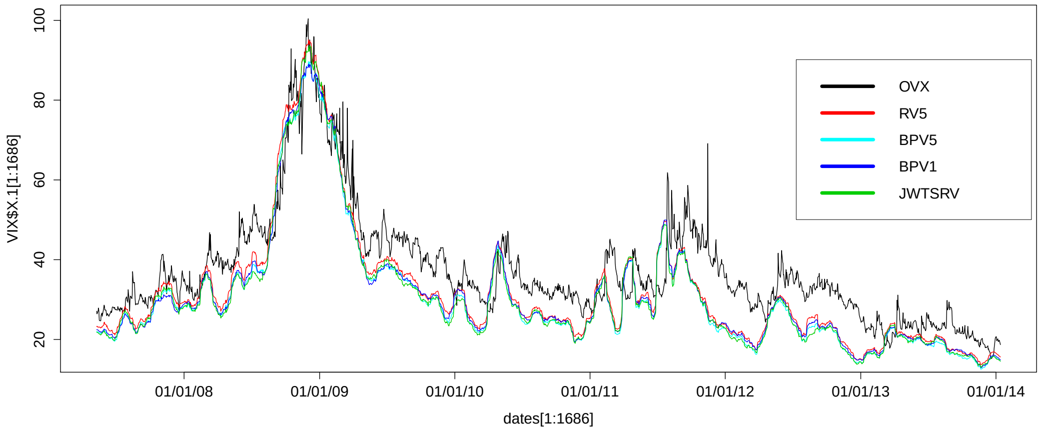


Figure A.8: Wavelet coherences between OVX index and all four estimators of realized volatility

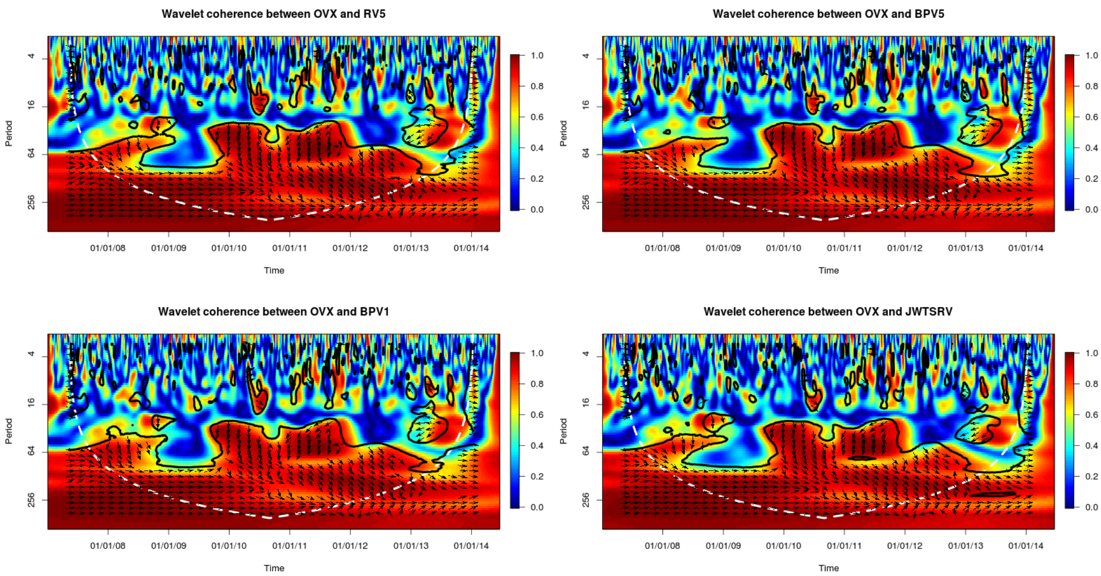


Figure A.9: Wavelet coherence between unemployment and labor productivity growth for Germany

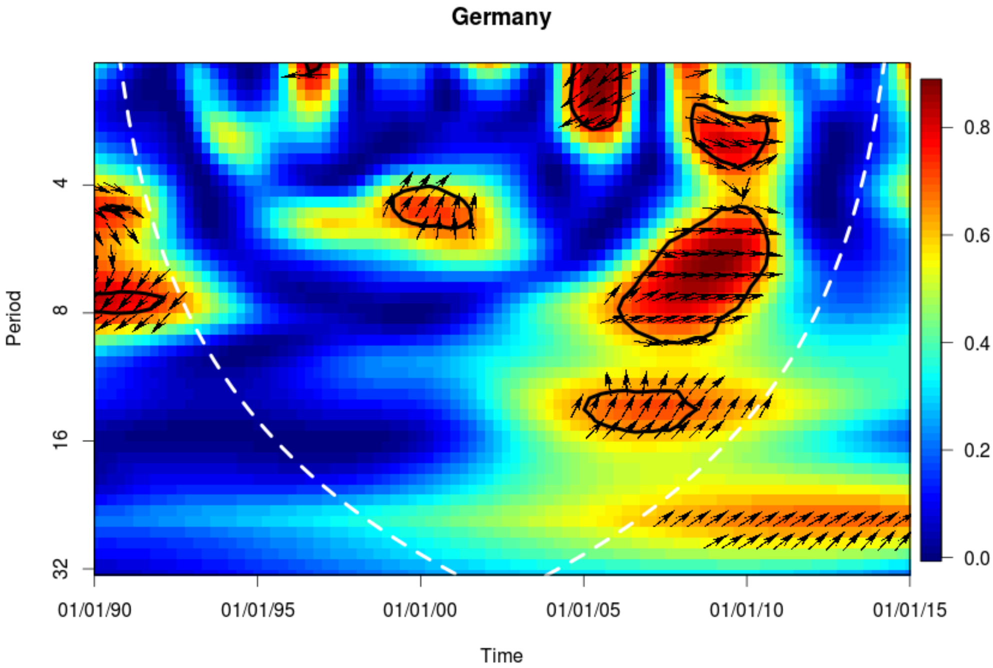


Figure A.10: Wavelet coherence between unemployment and labor productivity growth for the UK

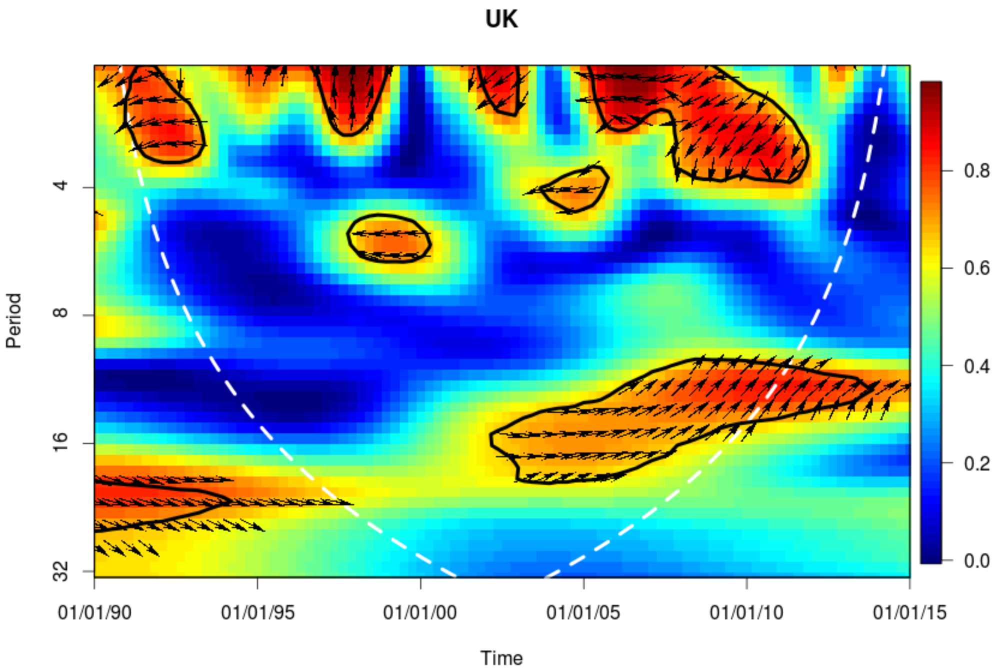


Figure A.11: Wavelet coherence between unemployment and labor productivity growth for France

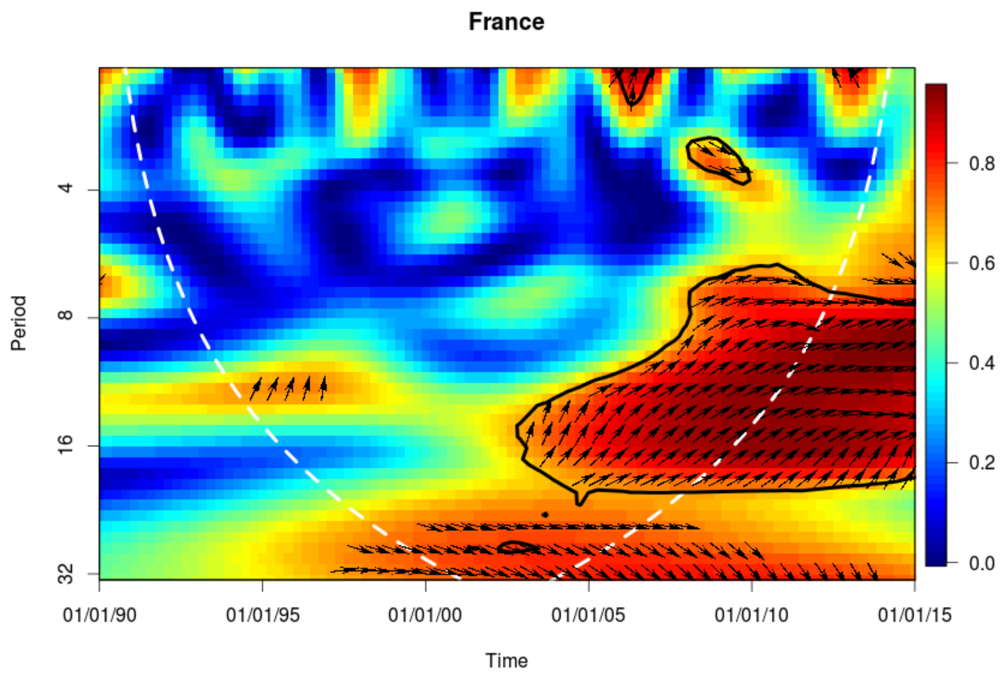


Figure A.12: Wavelet coherence between unemployment and labor productivity growth for Italy

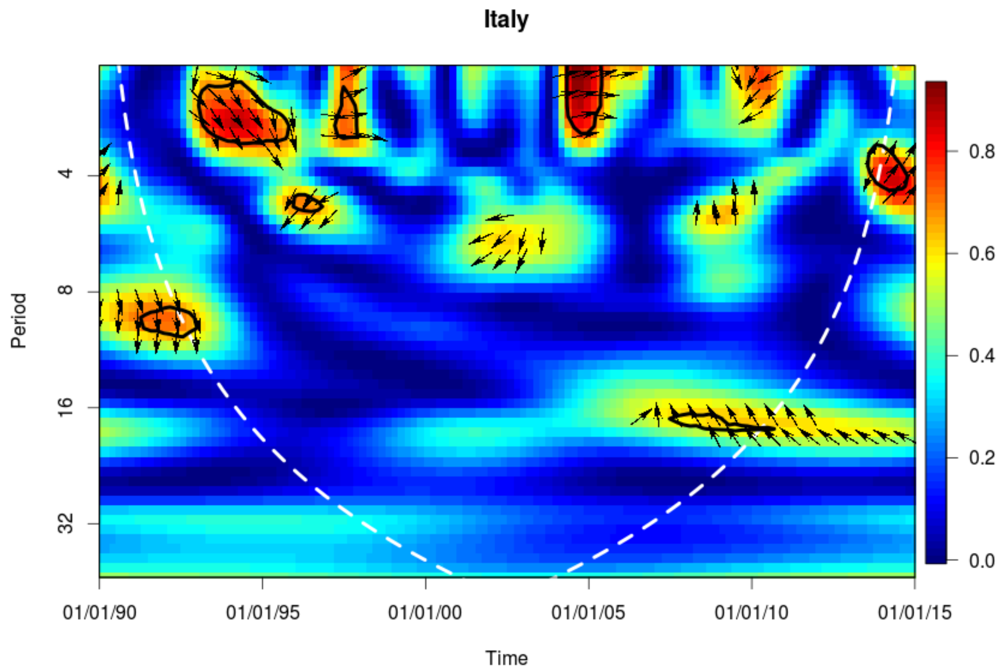


Figure A.13: Wavelet coherence between growth of unemployment and labor productivity growth for Germany

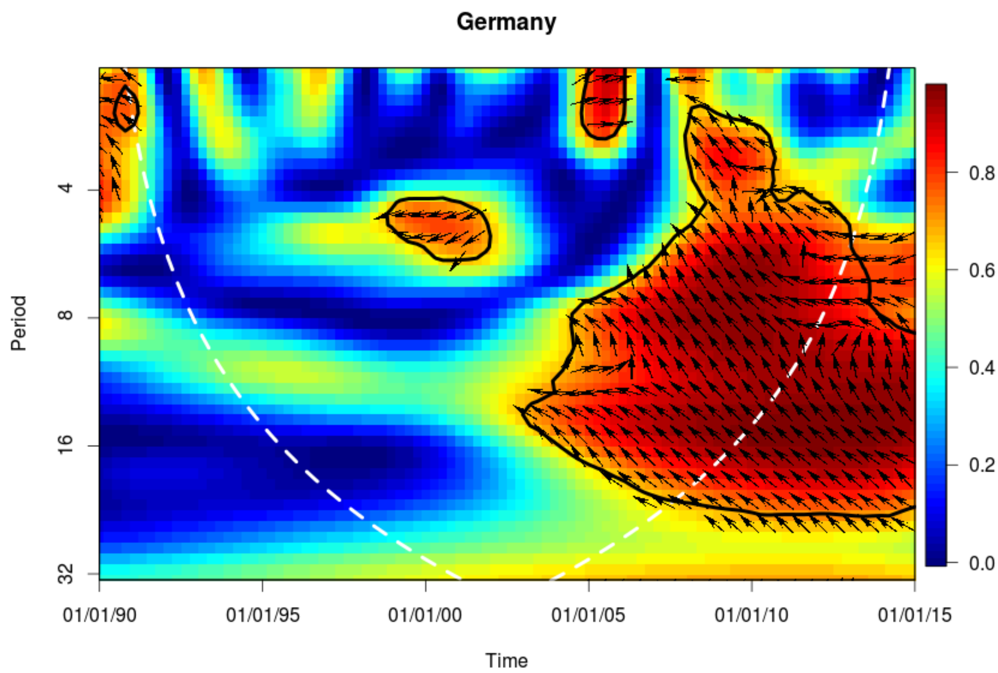


Figure A.14: Wavelet coherence between growth of unemployment and labor productivity growth for the UK

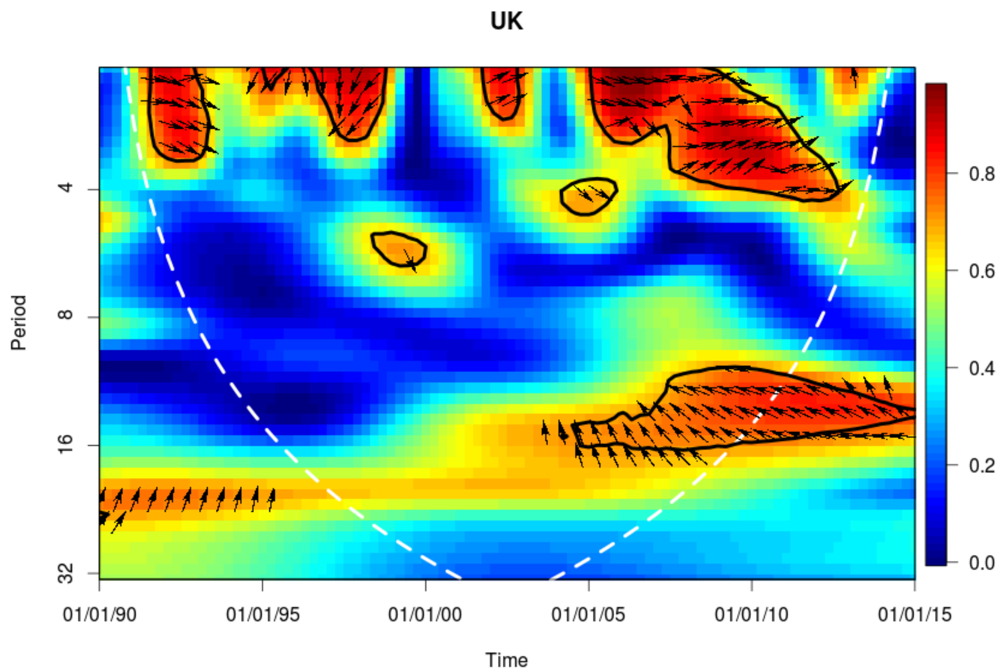


Figure A.15: Wavelet coherence between growth of unemployment and labor productivity growth for France

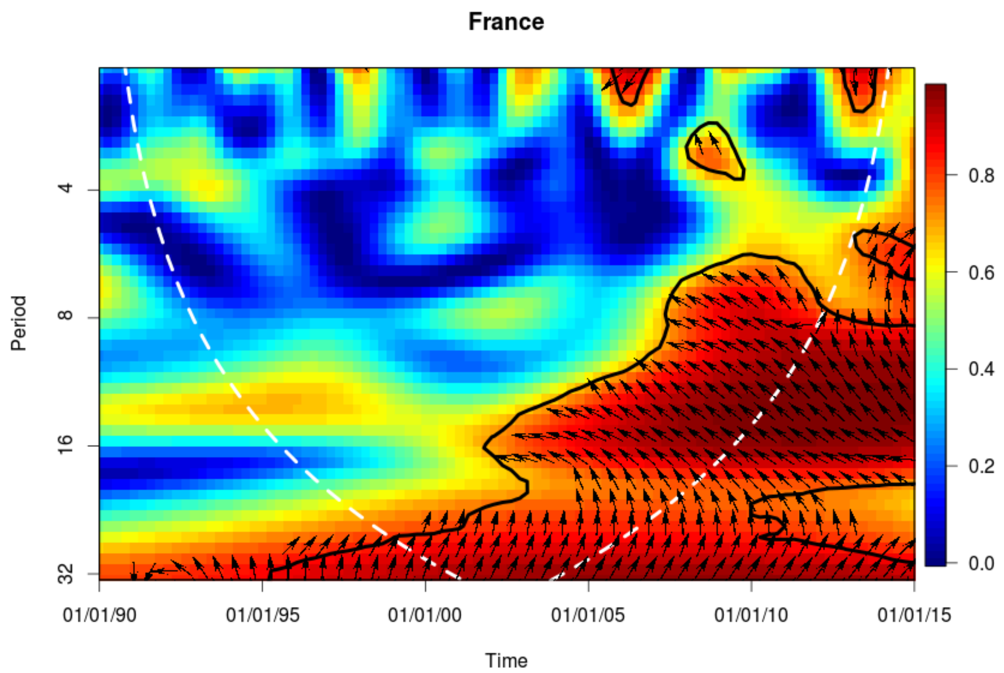


Figure A.16: Wavelet coherence between growth of unemployment and labor productivity growth for Italy