# Long-term dynamics of the VIX index and its tradable counterpart VXX

forthcoming in Journal of Futures Markets

Milan Bašta

Faculty of Informatics and Statistics, Department of Statistics and Probability, University of Economics, Prague, Czech Republic

Peter Molnár

University of Stavanger, UiS Business School, Stavanger, Norway

Faculty of Finance and Accounting, Department of Monetary Theory and Policy, University of Economics, Prague, Czech Republic

**Abstract**

We study the relationship of the VIX index and the exchange traded note VXX on various time scales. We find that changes of VIX and VXX are correlated only contemporaneously on time scales of days, but VIX leads VXX on time scales of months. Next, we construct a simple joint model for VXX and VIX which replicates all the key characteristics of these two time series, but in which VIX and VXX are related only via a correlated error term. Therefore, VIX cannot be used as a predictor of VXX and there is no apparent trading profit opportunity.

**Keywords**: implied volatility, VIX, VXX, predictability, comovement

# 1 Introduction

This paper studies the long-term relationship between the VIX index (volatility implied by options on the S&P 500 index), and a derivative instrument VXX, which seeks to provide investors with exposure to the VIX index. The exchange traded note VXX is practically a 30-day futures on the VIX index. Most financial futures have a very close relationship with their underlying asset due to the no-arbitrage condition. For example, the S&P 500 index and futures on the S&P 500 index are moving closely together. Moreover, there is basically no lead-lag relationship between the S&P 500 index and futures on the S&P 500, or such a lead-lag relationship exists only on very short time scales and can be found only in high-frequency data.

On the other hand, the situation is very different for the VIX index and VIX futures (or alternatively, the exchange traded note VXX). The no-arbitrage condition cannot be applied, because the VIX index is not tradable (cannot be bought or sold). Therefore, the relationship between the non-tradable VIX index and VXX, which is tradable as easily as a simple stock, is not governed by any immediate arbitrage, but by risk perceptions and expectations about the future of the financial markets. As a result, it is not obvious what the joint dynamics between VIX and VXX should look like.

The CBOE Market Volatility Index with ticker VIX was launched by the Chicago Board Options Exchange back in 1993. The VIX index is nowadays one of the most followed financial indicators in the world. Following the huge success of the VIX index, many other volatility indices have been introduced not only for equity markets (Bugge et al., 2016), but also for various commodities, including electricity (Birkelund et al., 2015).

Figure 1 illustrates the relation between the VIX index and its underlying S&P 500 index in the period from January 2007 to July 2016. As this figure illustrates, implied volatility was high during the Financial crisis of 2007-2009. Moreover, there is a strong negative correlation between changes in the S&P 500 index and changes in the VIX index.

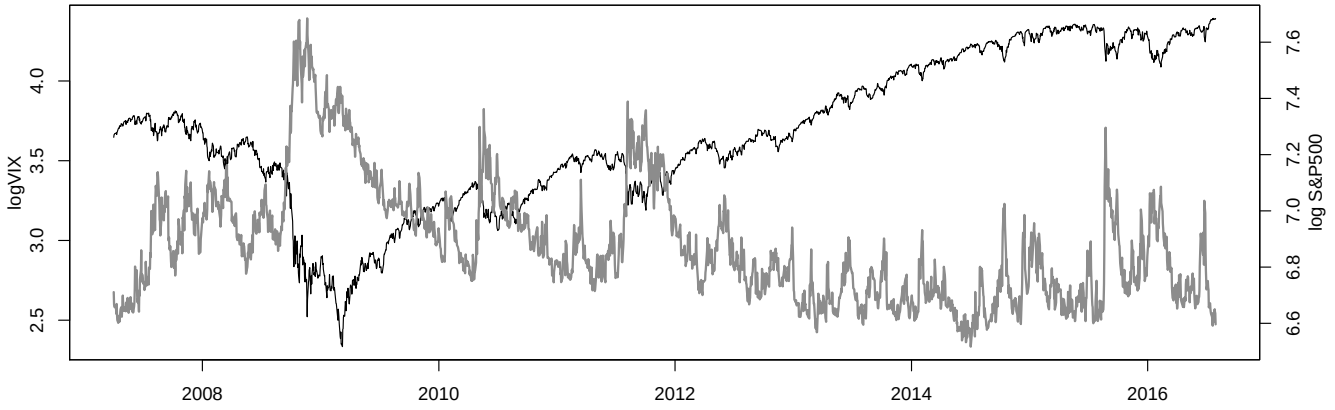


Figure 1: The natural logarithm of VIX (logVIX, gray, left vertical axis) and the natural logarithm of the S&P 500 close price (log S&P500, black, right vertical axis).

Due to its strong negative correlation (see Figure 1) with the overall stock market, the VIX index would be a very valuable instrument for portfolio diversification if it was a traded instrument. However, it is not traded. Direct trading of volatility started in 2004 when the CBOE introduced futures contracts on the VIX index. These futures allow investors to trade volatility directly. However, futures

contracts were not suitable for small investors. In 2009, futures indices were introduced. These indices are constructed by rolling futures contracts in such a way that the average maturity of all the contracts in the index remains constant. Most importantly, futures indices enabled the creation of exchange traded products that replicate the performance of these indices. The first VIX exchange traded product, VXX, was introduced in 2009. It replicates the performance of the S&P 500 VIX Short-Term Futures Index, which has a constant 30-day maturity. The popularity of the VIX exchange traded products, and VXX in particular, has been increasing tremendously - see Figure 2, where the natural logarithm of the traded volume of SPDR S&P 500 ETF (SPY) and the natural logarithm of the traded volume of VXX are depicted for the period from January 2009 to July 2016. This Figure illustrates that VXX experienced almost exponential growth in its traded volume during the period shown (the traded volume has increased by about 20 thousand times) whereas the traded volume of the most popular exchange traded fund, the SPY, replicating the S&P 500 index, has not changed much.

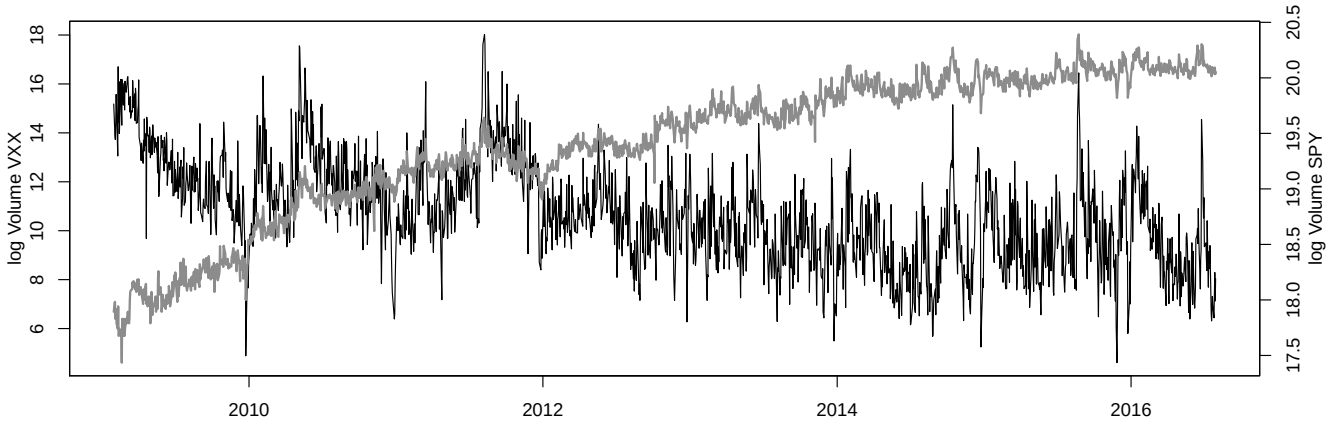


Figure 2: Natural logarithm of the traded volume of VXX (gray, left vertical axis) and the natural logarithm of the traded volume of SPY (black, right vertical axis) in the period from January 2009 to July 2016.

Recently, VIX exchange traded products have been studied from the perspective of risk-return tradeoff, diversification and trading strategies in several academic studies. Most of these studies have concluded that VIX exchange traded products are not suitable for diversification, because of the high negative expected return of the VIX short-term futures index (Alexander and Korovilas, 2013; Whaley, 2013; Alexander et al., 2016). On the other hand, this high negative expected return offers a possibility for highly profitable trading strategies (Bordonado et al., 2017). An excellent overview of trading and investment in volatility products is provided in Alexander et al. (2015).

The relationship between VIX and VIX futures is another important topic that has recently attracted increasing attention. One strand of literature (Lu and Zhu, 2010; Zhang et al., 2010) focuses on the relationship between VIX and its derivatives from the perspective of derivatives pricing. The goal of these models is to explain the price of VIX derivatives for a given level of the VIX index. Another strand of the literature studies whether VIX or its derivatives react faster to new information and whether VIX leads its derivatives, or vice versa. This relationship has been studied at daily and high-frequency (15-second or 1-minute) time scales. At the daily scale, changes of the VIX index and returns of VXX (or equivalently returns of VIX futures) are highly correlated (Zhang et al., 2010). On the other hand, at

15-second scale contemporaneous correlation in the changes of VIX and changes of its futures is close to zero (Frijns et al., 2016).

Regarding the lead-lag relationship, there is a strong evidence that VIX futures play dominant role at high frequencies (Frijns et al., 2016; Chen and Tsai, 2017). However, the literature is ambiguous about the lead-lag relationship between the VIX index and its derivatives at daily scale. Shu and Zhang (2012) find that even though simple cointegration tests suggest that VIX futures lead the VIX index, a more sophisticated test reveals that both VIX futures and VIX react simultaneously to new information.

Since the relationship between VIX and its derivatives might depend on the time scale and since the classical approaches do not explicitly assume that the relationship is scale-specific[1], we complement the existing literature by investigating the relationship between the relative changes of the VIX index and relative changes of VXX (note that changes of VXX are almost identical to changes of futures prices) on various time scales from one day to several months. To the best of our knowledge, no paper has yet studied the relationship between VIX and VXX as a function of scale.

Specifically, we employ wavelet analysis to understand the properties of VIX and VXX and the relationship between them as a function of scale, utilizing several scales ranging from the shortest (associated with 1 and 2 trading days) through medium term (4, 8, 16 trading days) to the longest time scales (32 and 64 trading days). This provides us with an insight into the nature of both the time series and their relationship. Specifically, we show that VIX leads VXX on long time scales of several months in the sense that the dynamics of VIX in the present is strongly correlated with the dynamics of VXX in the future. We propose a model which includes scale-specific terms and which can be considered a generalization of the traditional vector autoregression model. The proposed model can explain the observed properties of VIX and VXX as well as the leading behavior of VIX on long time scales. It also reveals that the fact that VIX leads VXX on long time scales cannot be exploited for improving the prediction of VXX.

The rest of this paper is organized as follows. Section 2 presents the data. Section 3 introduces wavelet analysis. Section 4 analyzes univariate characteristics of VXX and VIX. Section 5 studies the relationship between VXX and VIX. A model for the joint dynamics of VXX and VIX is presented in Section 6. Section 7 concludes.

## 2 Data

We use the close of S&P 500 VIX ST Futures ETN (hereafter VXX) and the close of Chicago Board Options Exchange (CBOE) Volatility Index (hereafter VIX). Both the time series were downloaded from Yahoo! Finance[2] and span the period from 30 January 2009 to 29 July 2016. The natural logarithm of VXX and the natural logarithm of VIX will be denoted as logVXX and logVIX, respectively. Both logVXX and logVIX are plotted in the top plot of Figure 3. LogVXX exhibits a well-known sharp decrease in level during the analyzed period, whereas the level of logVIX, which indicates the perceived volatility of the market, is mean-reverting, being high in high-volatility periods and low in low-volatility periods.

Since VXX is an instrument created to provide exposure to the non-traded VIX index, VIX and VXX could be expected to be cointegrated. However, as can be observed in Figure 3, this is clearly not the

[1] An exception is the framework of cointegration, which distinguishes between short-term and long-term relationships.
[2] http://finance.yahoo.com/

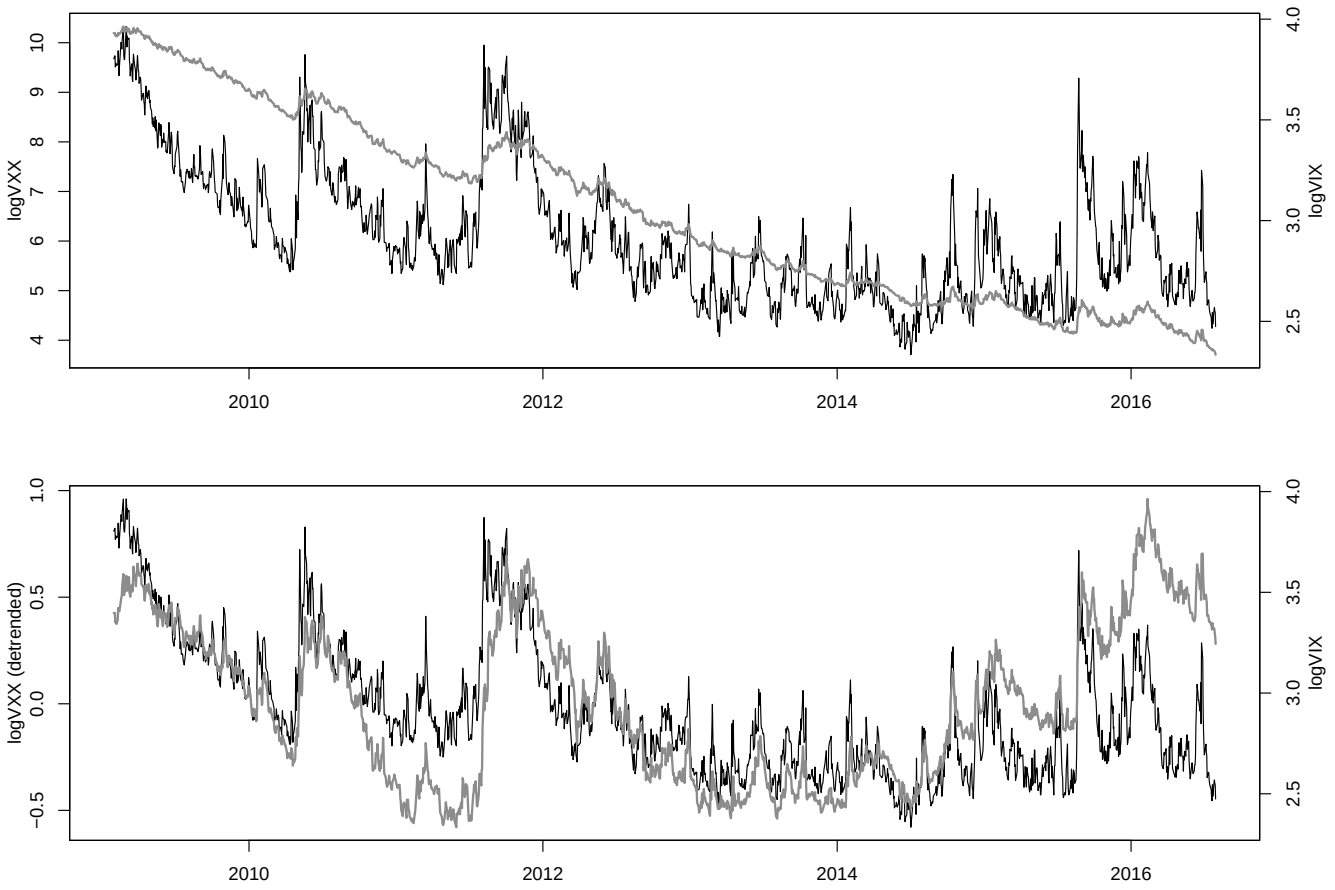


Figure 3: *Top plot*: logVXX (gray, left vertical axis) and logVIX (black, right vertical axis) during the period from 30 January 2009 to 29 July 2016. *Bottom plot*: The same time series, but logVXX is detrended.

case. VIX is a long-memory rather than a unit-root process, because it measures implied volatility, and volatility as such has a lower and an upper bound. Since standard cointegration between two processes can occur only if both the processes have unit roots, it cannot occur between VIX and VXX.

VXX replicates the S&P 500 VIX Short-Term Futures Index, and replicates it very well, as returns of VXX and of the VIX Short-Term Futures Index are almost identical (Whaley, 2013). The return on the VIX Short-Term Futures Index is a return from a trading strategy investing in VIX futures in such a way that the maturity of the futures portfolio is 30 days. When there is a futures with a 30-day maturity, we can look at VXX as at a basket of many 30-day futures. At this moment, a return of VXX will be almost the same as a return of the 30-day futures. However, after one day, the basket will now consist of several 29-day futures. In order to maintain 30-day maturity across the whole basket, some futures with longer maturity must be bought and this is financed by selling some of the 29-day futures. In other words, keeping the maturity of the futures portfolio constant at 30 days requires regular selling of futures with shorter maturity and buying of futures with longer maturity. Since futures with longer maturity are usually more expensive than futures with shorter maturity, this strategy delivers negative returns on average. Moreover, the slope of the VIX term structure is on average very high, which causes the negative average return of VXX to be high in magnitude. Further discussion of the cause of the negative average return of VXX can be found in Alexander and Korovilas (2013) and Whaley (2013).

VXX is not the only exchange traded product based on the VIX Short-Term Futures index. There are other products with almost the same purpose, for example ProShares VIX Short-Term Futures ETF (VIXY), whereas other products provide investors with a leveraged or inverse exposure to the VIX Short-Term Futures index. The reason why we choose VXX in our analysis is that it has the longest history and is by far the most traded VIX derivative. Moreover, we study the comovement at long time scales, and the various VIX products based on the VIX Short-Term Futures index move very closely at long horizons. For example, Bordonado et al. (2017) document that correlation between the VXX and VIXY is 0.723 for 1-minute returns, but 0.999 for daily returns. Therefore, the choice of a particular VIX product would have only a minor impact on our results but could be crucial for an analysis conducted on high-frequency data. Further, liquidity considerations also favor the choice of VXX as it is by far the most liquid VIX product. Since the early VXX market was relatively illiquid (Chen and Tsai, 2017), analysis similar to ours, but based on high-frequency data, could be problematic. However, illiquidity issues are unlikely to cause price misalignment at daily frequency which we employ in our analysis. To sum up, for our purposes VXX is a very good representative of the VIX Short-Term Futures index and related products[3].

In the bottom plot of Figure 3, we plot logVIX and also logVXX after subtracting the linear trend from the dynamics of logVXX. This makes it easier to compare the dynamics of logVXX with those of logVIX. This comparison suggests that both the time series are strongly interconnected and that the turning points (local maxima and minima) in logVIX occur before the turning points in logVXX. Making use of wavelet analysis, we show that this is indeed the case, but that this aspect of the joint dynamics of logVXX and logVIX does not directly imply that logVIX could be used to improve the forecasts of logVXX.

## 3 Wavelet analysis

There are several types of wavelet transforms such as the discrete wavelet transform (DWT), maximal overlap discrete wavelet transform (MODWT) and the continuous wavelet transform (CWT). This nomenclature of wavelet transforms corresponds to the one used in Percival and Walden (2006) where an introduction to these transforms is given. An interested reader can also be referred to Chakrabarty et al. (2015) where a short introduction to wavelet transform is provided together with a literature review of the use of wavelet theory for the description/explanation of horizon heterogeneity in financial markets.

Each of the transforms given above has somewhat different properties, captures the characteristics of an input time series in a slightly different way and can be more or less suitable for a practical task at hand. The CWT-based characteristics such as wavelet coherence and phase difference provide an interesting approach to studying scale-specific and time-varying strength of the relationship between time series as well as lead-lag effects. Aguiar-Conraria et al. (2014) studied the relation between oil price changes and macroeconomy. Aguiar-Conraria and Soares (2011) analyzed the time-frequency relationship between oil prices and macroeconomic variables. Graham et al. (2013) studied the time-frequency relationship between the returns of S&P 500 and S&P GSCI commodity index. Vacha and Barunik (2012) examined the relationship between commodities (crude oil, gasoline, heating oil and natural gas). Rua and Nunes

[3]It would be possible to study the comovement between the VIX index and a product based on the VIX Medium-Term Futures index. However, liquidity, data availability and popularity make VXX a preferred choice.

(2009) examined the comovement of international (Germany, Japan, United Kingdom and United States) stock markets. In and Kim (2006) explored the relationship between the Australian stock and futures market. Bašta and Molnár (2018) examined the relationship between oil and stock market volatility.

The CWT produces a redundant amount of information which can be unsuitable for further statistical modeling in some situations. The MODWT removes the redundancy of the CWT in the scale dimension, which allows for a better manipulation and processing of the analysis outputs in some applications. At the same time, the MODWT-based wavelet cross-covariance and cross-correlation still provide information about scale-specific (and potentially also time-varying) strength of the relationship between time series and lead-lag effects. The MODWT-based tools are also favorite measures used in economic and financial research. Gallegati (2008) studied the relationship between stock market and economic activity on a scale-by-scale basis. Kim and In (2005) examined the relationship between nominal stock returns and inflation on a scale-by-scale basis. Naccache (2011) examined the relationship between oil prices and macroeconomy. Fernández-Macho (2012) introduced MODWT-based wavelet multiple cross-correlation and wavelet multiple correlation and analyzed the comovement of eleven Eurozone stock markets. Crowley and Mayes (2009) analyzed the comovement of growth cycles of Germany, France and Italy. Crowley and Lee (2005) explored the relationship of business cycles of European countries.

The DWT continues in removing the remaining redundancy by downsampling the MODWT coefficients in time. While this results in some appealing properties of the DWT such as orthogonality of the DWT, approximate uncorrelatedness or independence of the DWT coefficients, which can be useful for some analyses such as denoising, signal reconstruction or some types of statistical inference, the DWT coefficients do not generally allow – due to the downsampling issue – for a straightforward examination of lead-lag effects. This can partly be overcome by using the DWT-based details and smooth which have some undesirable properties however (Percival and Walden (2006), Section 5). Consequently, the literature gives preferences to either the CWT or the MODWT while studying lead-lag aspects of a relationship between time series.

We have decided to use the MODWT-based tools in our analysis. We introduce the MODWT wavelet coefficients in Section 3.1. Based on these coefficients, we define wavelet variance in Section 3.2 and wavelet cross-covariance and cross-correlation in Section 3.3.

### 3.1 MODWT wavelet coefficients

To start with, let us assume a stochastic process $\{X_t : t = 0, ..., N-1\}$, where $N$ is the length of the process. Let $J \geq 1$ be an integer. The maximal overlap discrete wavelet transform (MODWT) wavelet coefficients $\{W_{X,j,t}\}$ of level $j$ (for $j = 1, 2, ..., J$) for $\{X_t : t = 0, ..., N-1\}$ are defined as ( Percival and Walden, 2006, Ch. 5)

$$W_{X,j,t} \equiv \sum_{l=0}^{L_j-1} h_{j,l} X_{t-l}, \quad t = L_j - 1, ...., N-1, \tag{1}$$

where $\{h_{j,l} : l = 0, ..., L_j - 1\}$ are the $j$th level MODWT wavelet filters ( Percival and Walden, 2006, Ch. 5) of length $L_j$, where ( Percival and Walden, 2006, Ch. 4)

$$L_j = (2^j - 1)(L_1 - 1) + 1. \tag{2}$$

Further, let $M_j$ be the length of the sequence of wavelet coefficients of level $j$ (we assume that $N \geq L_J$), i.e.

$$M_j \equiv N - L_j + 1, \quad j = 1, ..., J. \tag{3}$$

Percival and Walden (2006) show that the $j$th level MODWT wavelet filter, $\{h_{j,l}\}$, is an approximate ideal filter with the nominal passband $[1/2^{j+1}, 1/2^j]$. There are various families (sets) of filters, the most commonly used being the Haar, D(4) or LA(8) family. For example, the $j$th level *Haar* wavelet filters are linear filters of length $L_j = 2^j$ defined as (Percival and Walden, 2006)

$$h_{j,l} \equiv \begin{cases} 1/2^j, & \text{for } l = 0, ..., 2^{j-1} - 1, \\ -1/2^j, & \text{for } l = 2^{j-1}, ..., 2^j - 1. \end{cases} \tag{4}$$

The D(4) filters are filters of higher length $L_j$ than the corresponding Haar filters. For example, $L_1 = 4$ for D(4) filters and $L_1 = 2$ for Haar filters. As a result, D(4) wavelet filters $\{h_{j,l}\}$ provide a better approximation to ideal filters with nominal passbands $[1/2^{j+1}, 1/2^j]$ than Haar filters do (Percival and Walden, 2006).

Percival and Walden (2006) show that the $j$th level MODWT wavelet coefficients defined in Equation 1 are closely related to the (generalized) differences between the adjacent weighted averages of $\{X_t : t = 0, ..., N-1\}$, the averages being calculated effectively on scale $2^{j-1}$ (this can be clearly seen for the case of Haar wavelet filters defined in Equation 4 above).

Table 1 provides the connection between level $j$ (for $j = 1, ..., 7$) of wavelet coefficients, the passband of the corresponding ideal filter and the corresponding scale of analysis. For example, level 5 wavelet coefficients are obtained by filtering $\{X_t : t = 0, ..., N-1\}$ with a linear filter which approximates an ideal filter with a nominal passband $[1/64, 1/32]$ and these coefficients are associated with changes between adjacent weighted averages calculated on scale $2^{5-1} = 16$.

Table 1: The level of wavelet coefficients, the passband of the corresponding ideal filter and the corresponding scale of analysis.

| level | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| passband | $[\frac{1}{4}, \frac{1}{2}]$ | $[\frac{1}{8}, \frac{1}{4}]$ | $[\frac{1}{16}, \frac{1}{8}]$ | $[\frac{1}{32}, \frac{1}{16}]$ | $[\frac{1}{64}, \frac{1}{32}]$ | $[\frac{1}{128}, \frac{1}{64}]$ | $[\frac{1}{256}, \frac{1}{128}]$ |
| scale | 1 | 2 | 4 | 8 | 16 | 32 | 64 |

Since the wavelet coefficients defined in Equation 1 are obtained by linear filtering $\{X_t : t = 0, ..., N-1\}$ with *causal* linear filters, these coefficients lag behind $\{X_t : t = 0, ..., N-1\}$. However, it is possible to synchronize the coefficients with $\{X_t : t = 0, ..., N-1\}$ by shifting them ahead appropriately. For the values of these shifts see e.g. Percival and Walden (2006).

### 3.2 Wavelet variance

Since wavelet coefficients are obtained by linear filtering of $\{X_t : t = 0, ..., N-1\}$ with wavelet filters, the properties of the coefficients can be easily derived from the properties of the filters. For example, if $\{X_t : t = 0, ..., N-1\}$ is stationary, wavelet coefficients $\{W_{X,j,t} : t = L_j - 1, ..., N-1\}$ are also stationary, having a zero mean (Percival and Walden, 2006). For processes integrated of order $d$, wavelet coefficients

are stationary provided $L_1 \geq 2d$, the mean of the wavelet coefficients being zero if $L_1 > 2d$ (Percival and Walden, 2006).

Further, let us assume that wavelet coefficients $\{W_{X,j,t} : t = L_j - 1, ..., N-1\}$ are stationary. The variance of $\{W_{X,j,t} : t = L_j - 1, ..., N-1\}$, denoted as $\nu^2_{X,j}$, is called the wavelet variance and is associated with the variability of the stochastic process $\{X_t : t = 0, ..., N-1\}$ (be the process stationary or not) over scale $2^{j-1}$ ( Percival and Walden, 2006, Ch. 8). If the mean of the wavelet coefficients is zero, the following estimator can be used to estimate $\nu^2_{X,j}$ ( Percival and Walden, 2006, Ch. 8.3)

$$\hat{\nu}^2_{X,j} \equiv \frac{1}{M_j} \sum_{t=L_j-1}^{N-1} W^2_{X,j,t}. \tag{5}$$

If the mean of the wavelet coefficients is not zero, a different estimator of wavelet variance has to be used which takes into account the fact that the mean of the wavelet coefficients is unknown and needs to be estimated.

Approximate 95% confidence limits for $\nu^2_{X,j}$ are given as ( Percival and Walden, 2006, Ch. 8)

$$\left[ \frac{\eta \hat{\nu}^2_{X,j}}{\chi^2_{0.975}[\eta]}, \frac{\eta \hat{\nu}^2_{X,j}}{\chi^2_{0.025}[\eta]} \right], \tag{6}$$

where $\chi^2_{0.975}[\eta]$ and $\chi^2_{0.025}[\eta]$ are the 97.5th and 2.5th percentiles of a chi square distribution with $\eta$ degrees of freedom, where

$$\eta = max\{M_j/2^j, 1\}. \tag{7}$$

The $j$th level wavelet variance of a white noise process with variance $\sigma^2_X$ is given as

$$\nu^2_{X,j} \equiv \text{var}\{W_{X,j,t}\} = \text{var}\left\{ \sum_{l=0}^{L_j-1} h_{j,l} X_{t-l} \right\} = \sum_{l=0}^{L_j-1} h^2_{j,l} \text{var}\{X_{t-l}\} = \frac{1}{2^j} \sigma^2_X, \tag{8}$$

where we have made use of the fact that $\sum_{l=0}^{L_j-1} h^2_{j,l} = \frac{1}{2^j}$ (Percival and Walden, 2006).

### 3.3 Wavelet cross-covariance and cross-correlation functions

Let us assume a bivariate stationary stochastic process whose components are the $\{X_t : t = 0, ..., N-1\}$ and $\{Y_t : t = 0, ..., N-1\}$ processes. The classical cross-covariance and cross-correlation functions at lag $\tau$ between the processes (components) are defined as

$$\Gamma_{XY,\tau} \equiv \text{cov}\left(X_t, Y_{t+\tau}\right), \quad \tau = -(N-1), ..., 0, ..., N-1, \tag{9}$$

$$R_{XY,\tau} \equiv \frac{\Gamma_{XY,\tau}}{\left[\text{var}(X_t)\text{var}(Y_t)\right]^{1/2}}, \quad \tau = -(N-1), ..., 0, ..., N-1. \tag{10}$$

Now, we drop the assumption that $\{X_t : t = 0, ..., N-1\}$ and $\{Y_t : t = 0, ..., N-1\}$ constitute a bivariate stationary process and only require that wavelet coefficients $\{W_{X,j,t} : t = L_j - 1, ..., N-1\}$ and $\{W_{Y,j,t} : t = L_j - 1, ..., N-1\}$ are components of a bivariate stationary process. The wavelet

cross-covariance and wavelet cross-correlation function of level $j$ (for $j = 1, ..., J$) at lag $\tau$ are defined as (Whitcher et al., 2000)

$$\gamma_{XY,j,\tau} \equiv \text{cov}\left(W_{X,j,t}, W_{Y,j,t+\tau}\right), \quad \tau = -(M_j - 1), ..., 0, ..., M_j - 1, \tag{11}$$

$$\rho_{XY,j,\tau} \equiv \frac{\gamma_{XY,j,\tau}}{\left[\nu^2_{X,j}\nu^2_{Y,j}\right]^{1/2}}, \quad \tau = -(M_j - 1), ..., 0, ..., M_j - 1, \tag{12}$$

where $\nu^2_{X,j}$ and $\nu^2_{Y,j}$ are $j$th level wavelet variances for $\{X_t : t = 0, ..., N-1\}$ and $\{Y_t : t = 0, ..., N-1\}$. The wavelet cross-covariance and wavelet cross-correlation function of level $j$ inform us about the strength of the comovement and the lead/lag relationship between $\{X_t : t = 0, 1, ..., N-1\}$ and $\{Y_t : t = 0, 1, ..., N-1\}$ on scale $2^{j-1}$. It is important to understand that $\{W_{X,j,t} : t = L_j - 1, ..., N-1\}$ and $\{W_{Y,j,t} : t = L_j - 1, ..., N-1\}$ are often strongly autocorrelated (especially for large values of $j$). This has to be taken into account when interpreting the wavelet cross-covariance or wavelet cross-correlation functions which can consequently exhibit non-zero values for lags $\tau \neq 0$ even in such a situation where the original processes $\{X_t : t = 0, ..., N-1\}$ and $\{Y_t : t = 0, ..., N-1\}$ are components of a bivariate white noise process.

If the mean of $\{W_{X,j,t} : t = L_j - 1, ..., N-1\}$ and $\{W_{Y,j,t} : t = L_j - 1, ..., N-1\}$ is zero, $\gamma_{XY,j,\tau}$ and $\rho_{XY,j,\tau}$ (for $j = 1, ..., J$) can be estimated as follows (Whitcher et al., 2000)

$$\hat{\gamma}_{XY,j,\tau} \equiv \begin{cases} \frac{1}{M_j}\sum_{t=L_j-1}^{N-1-\tau} W_{X,j,t}W_{Y,j,t+\tau}, & \tau = 0, ..., M_j - 1, \\ \frac{1}{M_j}\sum_{t=L_j-\tau-1}^{N-1} W_{X,j,t}W_{Y,j,t+\tau}, & \tau = -(M_j - 1), ..., -1, \end{cases} \tag{13}$$

$$\hat{\rho}_{XY,j,\tau} \equiv \frac{\hat{\gamma}_{XY,j,\tau}}{\left(\hat{\nu}^2_{X,j}\right)^{1/2}\left(\hat{\nu}^2_{Y,j}\right)^{1/2}}, \quad \tau = -(M_j - 1), ..., 0, ..., M_j - 1, \tag{14}$$

where $\hat{\nu}^2_{X,j}$ and $\hat{\nu}^2_{Y,j}$ are the estimators of $\nu^2_{X,j}$ and $\nu^2_{Y,j}$ (see Equation 5). If the means of the wavelet coefficients are not zero, the estimators given in Equations 13 and 14 have to be adjusted correspondingly to take into account the fact that the means need to be estimated.

An approximate 95% confidence interval for $\rho_{XY,j,\tau}$ is given as (Whitcher et al., 2000)

$$\left[\tanh\left\{\tanh^{-1}\left(\hat{\rho}_{XY,j,\tau}\right) - \frac{u_{0.975}}{\sqrt{\frac{M_j}{2^j} - 3}}\right\}, \tanh\left\{\tanh^{-1}\left(\hat{\rho}_{XY,j,\tau}\right) + \frac{u_{0.975}}{\sqrt{\frac{M_j}{2^j} - 3}}\right\}\right], \tag{15}$$

where $u_{0.975}$ is the 97.5th percentile of a standard normal distribution.

Let us calculate $\rho_{XY,j,\tau}$ for a bivariate white noise process $(\{X_t\}, \{Y_t\})$, where $\sigma_{XY}$ and $\rho_{XY}$ is the classical covariance and correlation between $\{X_t\}$ and $\{Y_t\}$, and $\sigma^2_X$ and $\sigma^2_Y$ the classical variance of $\{X_t\}$ and $\{Y_t\}$. Making use of the definition of Equation 12 and the result of Equation 8, we can write

$$\rho_{XY,j,\tau} \equiv \frac{\text{cov}\left(W_{X,j,t}, W_{Y,j,t+\tau}\right)}{\sqrt{\text{var}(W_{X,j,t})}\sqrt{\text{var}(W_{Y,j,t})}} = \frac{\text{cov}\left(\sum_{l=0}^{L_j-1} h_{j,l}X_{t-l}\sum_{k=0}^{L_j-1} h_{j,k}Y_{t+\tau-k}\right)}{\sqrt{\frac{1}{2^j}\sigma^2_X}\sqrt{\frac{1}{2^j}\sigma^2_Y}} =$$

$$= \frac{2^j \sum_{l=0}^{L_j-1} \sum_{k=0}^{L_j-1} h_{j,l} h_{j,k} \mathrm{cov}\left(X_{t-l}, X_{t+\tau-k}\right)}{\sigma_X \sigma_Y}.$$

For $|\tau| < L_j - 1$, we get

$$\rho_{XY,j,\tau} = \frac{2^j \sum_{l=0}^{L_j-1-|\tau|} h_{j,l} h_{j,l+|\tau|} \sigma_{XY}}{\sigma_X \sigma_Y} = 2^j \rho_{XY} \sum_{l=0}^{L_j-1-|\tau|} h_{j,l} h_{j,l+|\tau|}. \tag{16}$$

Otherwise, for $|\tau| > L_j - 1$, we get

$$\rho_{XY,j,\tau} = 0. \tag{17}$$

We can see that for a bivariate white noise process the wavelet cross-correlation function is symmetric around zero (for any $j$), is non-zero for $|\tau| < L_j - 1$ and is equal to $\rho_{XY}$ at lag zero (for any $j$).

# 4 Univariate characteristics of logVXX and logVIX

In this section we provide the univariate characteristics of logVXX and logVIX. The traditional characteristics are presented in Section 4.1, whereas the wavelet characteristics are presented in Section 4.2.

## 4.1 Traditional characteristics

It follows from the nature of logVXX and logVIX (see Section 2) as well as from unit-root tests that logVXX is a unit-root process and logVIX is a near unit-root process. Therefore, no sensible cointegrating relationship can exist between logVXX and logVIX. Consequently, while exploring the processes as well as the relationship between them by *classical* approaches we use the first difference of logVXX and the first difference of logVIX.

Table 2: Summary statistics for the first difference of logVXX and the first difference of logVIX.

| | min | max | mean | median | std. dev |
|---|---|---|---|---|---|
| first diff. of logVXX | -0.15 | 0.21 | -0.0034 | -0.0062 | 0.039 |
| first diff. of logVIX | -0.35 | 0.41 | -0.0007 | -0.0070 | 0.073 |

The elementary characteristics of the first difference of logVXX and the first difference of logVIX are presented in Table 2. The mean of the first difference of logVXX is negative and corresponds to an approximately 0.34% decrease in VXX per day (which corresponds to an approximately 7% decrease in VXX per month and an approximately 60% decrease in VXX per year, which is in agreement with the sharp decrease in logVXX documented in Figure 3). The distribution of the first difference of logVXX has a slightly longer right tail (compared to its left tail), which makes the mean larger than the median and the maximum larger than the absolute value of the minimum. The standard deviation of the first difference of logVXX is 0.039 which corresponds to a monthly value of 0.18 and an annualized value of 0.62. Since the analyzed period starts in 2009, when market volatility and thus logVIX were high, and ends in 2016, when market volatility and logVIX were lower, the mean of logVIX is slightly negative.

The classical sample autocorrelation functions of the first difference of logVXX and the first difference of logVIX are presented in Figure 4. The autocorrelation in the time series of the first difference of logVXX

is very weak, suggesting that the corresponding process can be approximately considered to be white noise. On the other hand, the time series of the first difference of logVIX is slightly autocorrelated, suggesting that the corresponding process deviates slightly from a white noise process.

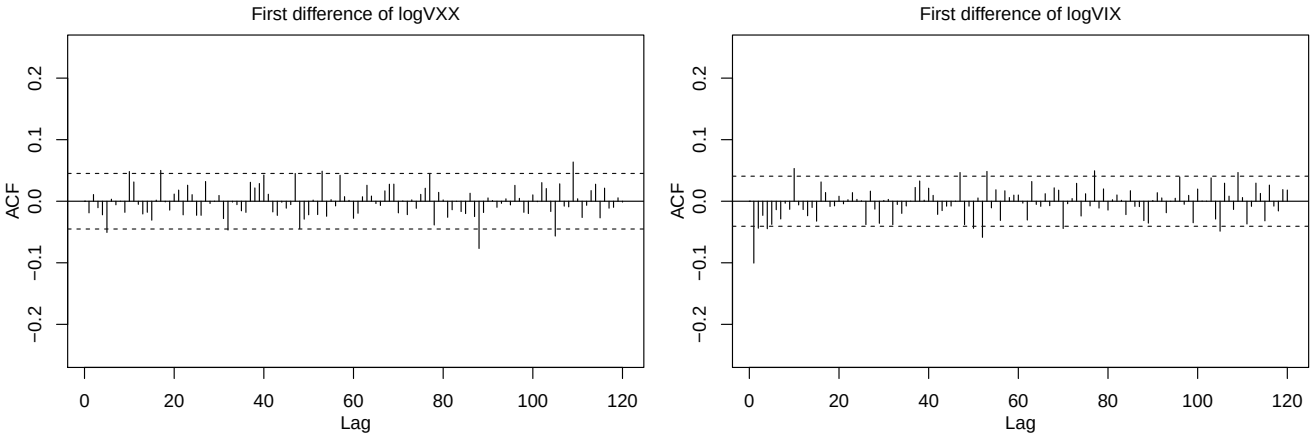


Figure 4: Classical sample autocorrelation functions of the first difference of logVXX (left) and the first difference of logVIX (right). The dashed lines depict the critical values for the individual tests of zero autocorrelation at the given lag assuming a 5% significance level.

## 4.2 Wavelet characteristics

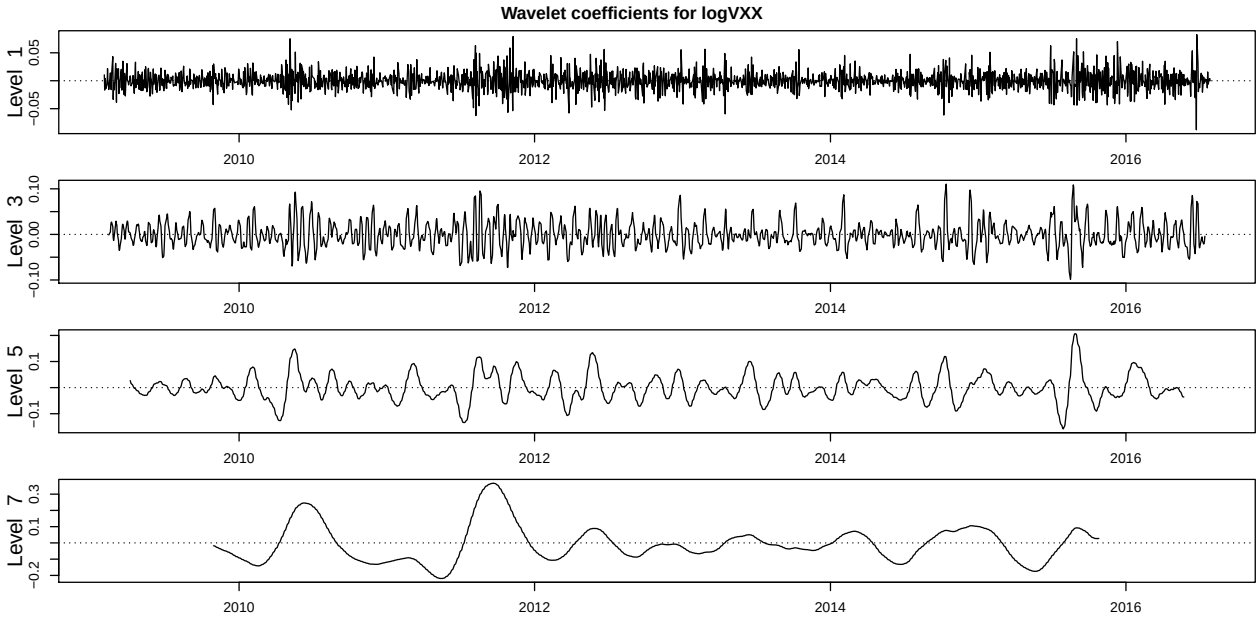


Figure 5: D(4) wavelet coefficients for logVXX of levels 1, 3, 5 and 7. The coefficients have been synchronized with logVXX.

In our wavelet analysis, D(4) filters will be used. We have chosen D(4) filters over Haar filters since D(4) filters provide better approximations to the corresponding ideal filters and are thus capable of ex-

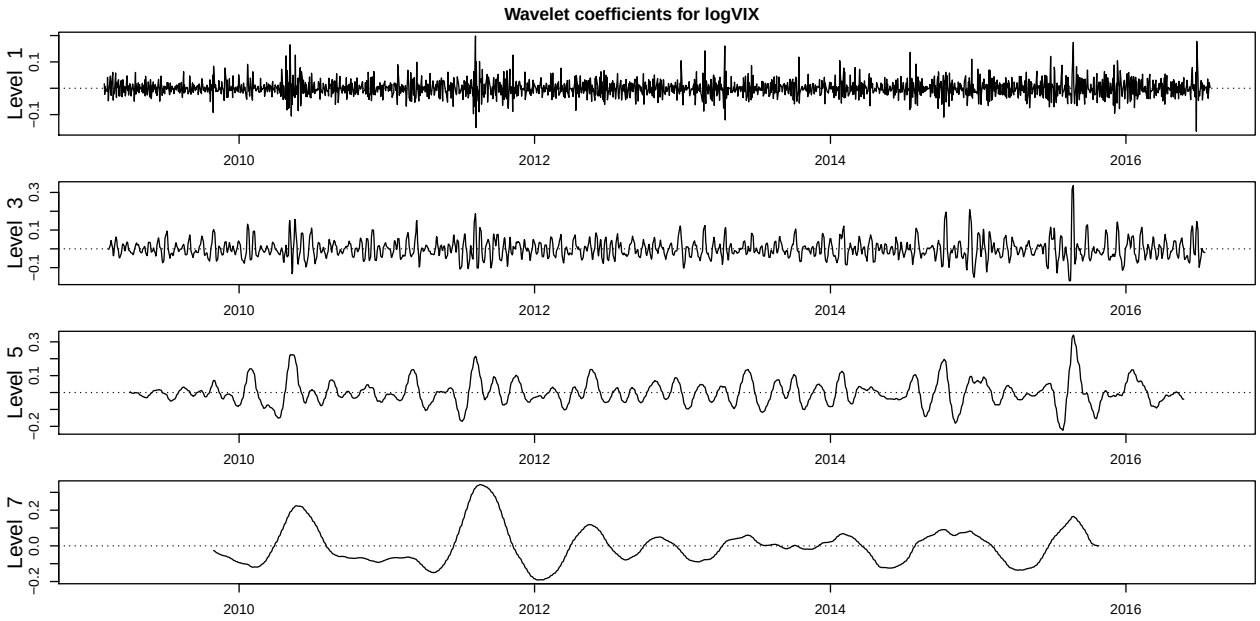


Figure 6: D(4) wavelet coefficients for logVIX of levels 1, 3, 5 and 7. The coefficients have been synchronized with logVIX.

tracting the required frequency content in a better way[4].

Since the D(4) wavelet coefficients for logVXX and logVIX are stationary and have zero means (see also Section 3), it is not necessary to take the first difference of logVXX and logVIX prior to wavelet analysis. The D(4) wavelet coefficients for levels 1, 3, 5 and 7 are plotted in Figure 5 for logVXX and in Figure 6 for logVIX . We do not explore levels $j$ higher than 7 since $L_j$ is a large number for $j > 7$, which leads to either only a few (relative to the length of the input time series) or no wavelet coefficients being available for $j > 7$.

The estimated wavelet variances (for $j = 1, ..., 7$) are depicted in Figure 7. The shape of the dependence that would correspond to a random walk process is depicted by a dotted line, the corresponding first level wavelet variance of the random walk process being equal to the first level wavelet variance estimated from the data. In agreement with the classical analysis described in the previous paragraphs, we see that logVXX can be well considered a random walk (with drift) and the first difference of logVXX can be considered white noise, while the dynamics of logVIX deviates from the dynamics of a random walk process. Specifically, logVIX is deficient in variability at high levels given the variability at low levels.

[4]Of course, we could have chosen other filters (such as LA(8)) that would provide even better approximations to the ideal filters, but this would be at the expense of having fewer wavelet coefficients, since the other filters would have larger values of $L_j$. The choice of D(4) filters is thus a compromise between the two aspects – the quality of the approximation and the length of the filter.

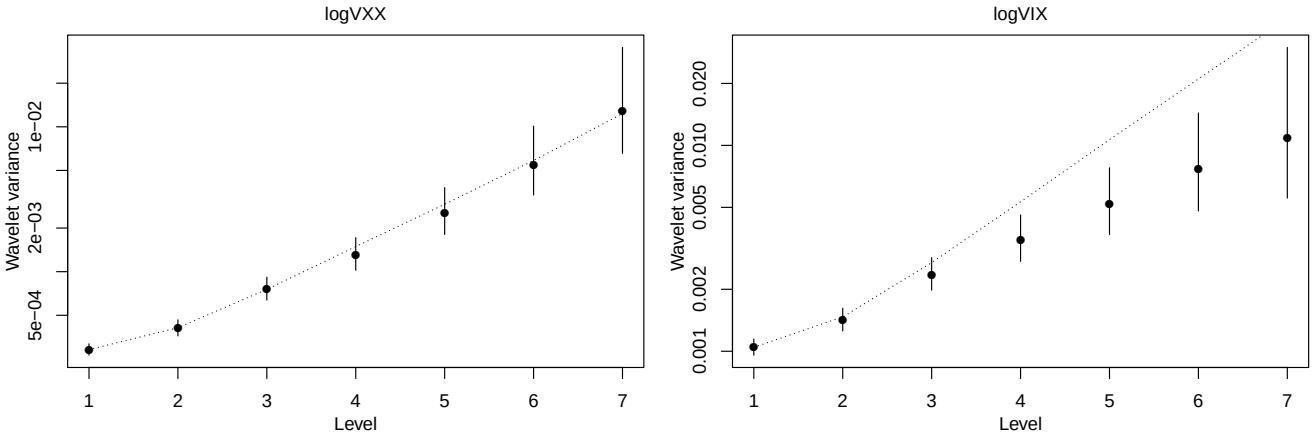


Figure 7: Point estimates of D(4) wavelet variances for logVXX and logVIX (without differencing) and the corresponding 95% confidence intervals (vertical segments) as a function of level $j$ (for $j = 1, ..., 7$). The point estimates have been obtained making use of Equation 5, the confidence intervals being obtained as dictated by Equation 6. To help assess to what extent the processes deviate from a random walk process, dotted lines are provided that capture the wavelet variances of a random walk process with level 1 wavelet variance equal to the point estimate of wavelet variance for level 1 obtained from the data (see text).

# 5 Bivariate characteristics of logVXX and logVIX

In this section we explore the bivariate characteristics of logVXX and logVIX. Section 5.1 deals with the traditional characteristics, whereas Section 5.2 looks at the wavelet characteristics.

## 5.1 Traditional characteristics

In Figure 8 we provide an estimate of the classical cross-correlation function (see Equation 10) between the processes of the first difference of logVXX ($Y_t$) and the first difference of logVIX ($X_t$). The correlation attains its maximum (0.87) at lag zero and is very weak at other lags. To aid the assessment of the form and strength of the relationship at lag zero, the scatterplot of the first difference of logVXX against the first difference of logVIX is also provided.

## 5.2 Wavelet characteristics

A different and more interesting insight into the relationship between the variables can be obtained by wavelet analysis. Whereas on short time scales the cross-correlation function between logVXX ($Y_t$) and logVIX ($X_t$) is approximately symmetric around lag zero with maximum correlation occurring at lag zero, on long time scales the cross-correlation function is approximately symmetric around a positive lag (i.e. lag larger than zero), the maximum correlation occurring also at this positive lag (see the top plot in Figure 9). In this sense, changes in logVIX on long time scales lead changes in logVXX on long time scales. The strength of the relationship (i.e. the maximum value of the correlation) seems to increase slightly with increasing scale, although this effect is statistically insignificant due to a relatively large error in the estimation of wavelet cross-correlation on large scales (see Figure 9). We checked that similar

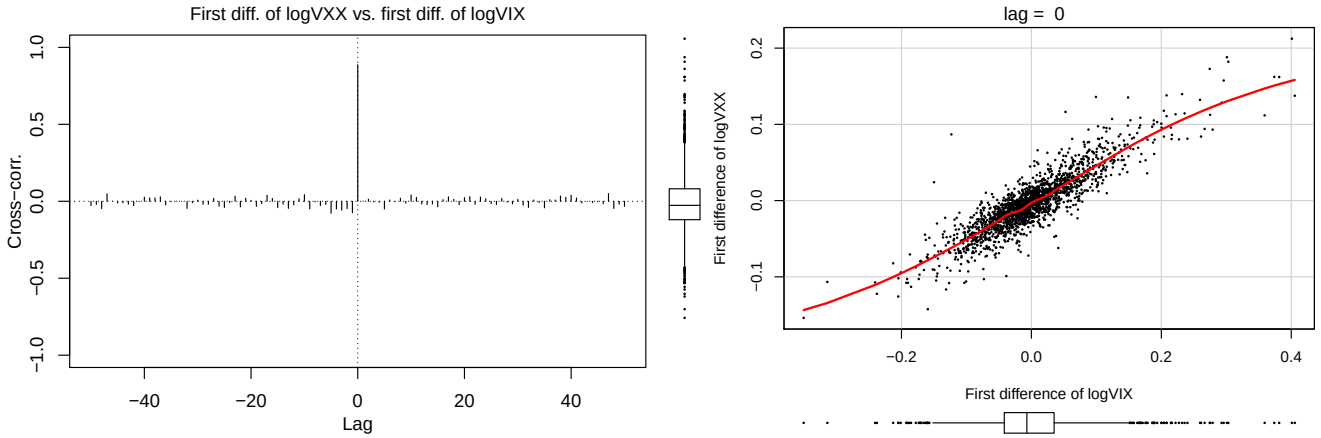


Figure 8: *Left plot*: Classical cross-correlation function between between the time series of the first difference of logVXX ($Y_t$) and the first difference of logVIX ($X_t$), and the scatterplot of the first difference of logVXX against the first difference of logVIX (with no lag).

results would have been obtained if we had used Spearman correlations instead of Pearson ones, and that the results do not depend on the choice of the wavelet filter – if Haar filters were used instead of D(4) filters, similar results would have been obtained.

To demonstrate the lead/lag relationship on large scales explicitly, level 7 wavelet coefficients for logVXX and level 7 wavelet coefficients for logVIX are plotted in the middle plot of Figure 9 (both series of coefficients have been synchronized in the figure with the original time series logVXX and logVIX). Observing the turning points in both series, it is obvious that level 7 wavelet coefficients for logVIX lead level 7 wavelet coefficients for logVXX, and that the lead is particularly pronounced in those periods where the coefficients series exhibit higher variability (years 2010, 2011, 2014 and 2015). This is in agreement with the bottom plots of Figure 9 which depict level 5 (level 7) wavelet coefficients for logVXX against level 5 (level 7) wavelet coefficients delayed by 2 days (15 days) for logVIX. It is obvious that there is an upper tail dependence between the variables, i.e. large values of one variable are "accompanied" by large values of the second variable.

In Table 3 we summarize the results of the wavelet cross-correlation analysis. Namely, the lag where the maximum of the sample wavelet cross-correlation occurs is reported together with the respective value of the sample cross-correlation. We can note that for levels larger than three the lag becomes positive and increases. The respective sample cross-correlation at this lag also increases with increasing level.

Table 3: The lag where the maximum of the sample wavelet cross-correlation occurs is reported together with the respective value of the sample cross-correlation. The results are given both for Pearson (in the first portion of the table) and for Spearman correlation.

| Level ($j$) | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| Pearson, max. cross-corr. at lag | 0 | 0 | 0 | 1 | 2 | 6 | 15 |
| Pearson, value of cross-corr. | 0.887 | 0.899 | 0.898 | 0.916 | 0.921 | 0.929 | 0.945 |
| Spearman, max. cross-corr. at lag | 0 | 0 | 0 | 1 | 2 | 6 | 14 |
| Spearman, value of cross-cor. | 0.866 | 0.885 | 0.897 | 0.915 | 0.886 | 0.901 | 0.933 |

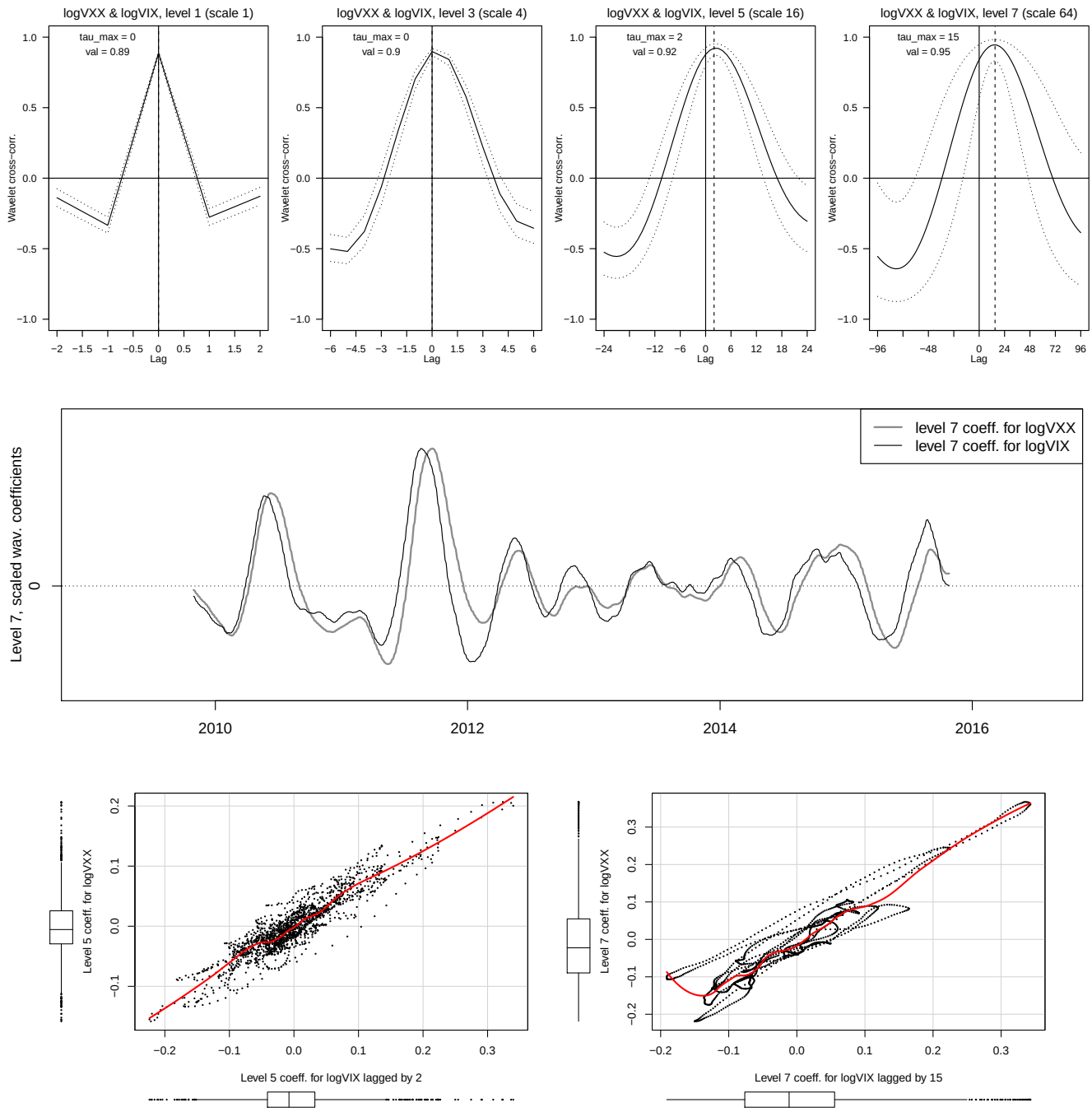


Figure 9: *Top plot*: Wavelet cross-correlation functions between logVXX ($Y_t$) and logVIX ($X_t$) for levels 1, 3, 5 and 7 (solid line). The respective 95% confidence intervals are depicted by dotted lines. The vertical dashed lines denote the lag (see tau_max in the plot) where the maximum wavelet cross-correlation occurs (see val in the plot for this maximum cross-correlation). *Middle plot*: Time series of scaled level 7 wavelet coefficients for logVXX (gray) and scaled level 7 wavelet coefficients for logVIX (black); both the series have been synchronized with the original logVXX and logVIX series; scaling is accomplished by dividing the coefficients by their standard deviations. *Bottom plot*: Scatter plots of wavelet coefficients of level 5 (level 7, right plot) for logVXX against wavelet coefficients of level 5 (level 7) delayed by 2 days (15 days) for logVIX. A smooth is overplotted (red) to capture the mean relationship.

A question arises as to what the relationship looks like for levels higher than 7. Level 7 scaling coefficients for logVXX and logVIX are depicted (for D(4) wavelets) in Figure 10 and carry information about the relationship between the time series associated with the range of frequencies $\left[0, 2^{-8}\right]$. From the figure it follows that the time series are interrelated at the right boundary of the frequency range (i.e. near frequency $2^{-8}$), since the level 7 scaling coefficients for logVXX and logVIX have a few "bumpy patterns" in common, but are not interrelated as we approach zero frequency: logVIX and logVXX are not cointegrated and thus neither are the scaling coefficients.

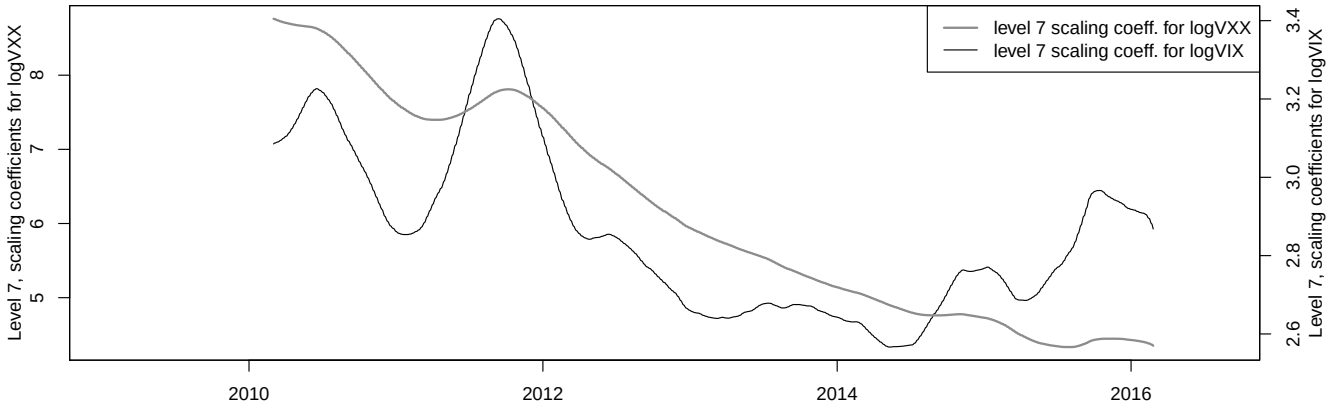


Figure 10: Level 7 scaling coefficients for logVXX (gray, left vertical axis) and level 7 scaling coefficients for logVIX (black, right vertical axis). The coefficients have been synchronized with the original time series logVXX and logVIX.

### 5.3 Interpretation of wavelet coefficients

To interpret the series of wavelet coefficients (and thus the lead/lag behavior between logVXX and logVIX) in an understandable way, we will discuss these series for the case of Haar wavelet coefficients. As mentioned above the Haar wavelet cross-correlation functions are very similar to the D(4) wavelet cross-correlation functions presented above. However, the interpretation is easier for the Haar wavelet coefficients. Specifically, let the mean level of time series $\{X_t\}$ over the period of length $k$ from time $t-k+1$ till time $t$ be denoted as

$$m_{X,t}^{(k)} = \frac{1}{k} \sum_{i=t-k+1}^{t} X_i. \tag{18}$$

Subsequently, the Haar wavelet coefficients of level $j$ for $\{X_t\}$ can be rewritten as

$$\begin{aligned} W_{X,j,t}^{(Haar)} &= \frac{1}{2}\left[m_{X,t}^{(2^{j-1})} - m_{X,t-2^{j-1}}^{(2^{j-1})}\right] = \\ &= \frac{1}{2}\left[\frac{1}{2^{j-1}}\left(X_t + ... + X_{t-2^{j-1}+1}\right) - \frac{1}{2^{j-1}}\left(X_{t-2^{j-1}} + ... + X_{t-2^j+1}\right)\right]. \\ &= \frac{1}{2}\left[\frac{1}{2^{j-1}}\left((X_t - X_{t-2^{j-1}}) + ... + (X_{t-2^{j-1}+1} - X_{t-2^j+1})\right)\right] = \\ &= \frac{1}{2}\left[\frac{1}{2^{j-1}}\left(r_{X,t}^{(2^{j-1})} + ... + r_{X,t-2^{j-1}+1}^{(2^{j-1})}\right)\right] = \end{aligned}$$

$$= \frac{1}{2^j}\left[q^{(2^j-1)} * r_{X,t}^{(1)}\right],$$

where

$$r_{X,t}^{(k)} = X_t - X_{t-k} \tag{19}$$

and $*$ stands for convolution and $q^{(2^j-1)}$ is a causal triangular filter of length $2^j - 1$ with weights given as $[1, 2, 3, ..., 2^{j-1}, ..., 3, 2, 1]$. Consequently, if $\{X_t\}$ is logVXX, we see that the corresponding Haar wavelet coefficient of level $j$, $W_{logVXX,j,t}^{(Haar)}$, is proportional:

- to the difference between mean log levels $m_{logVXX,t}^{(2^{j-1})}$ and $m_{logVXX,t-2^{j-1}}^{(2^j-1)}$,
- to the mean of $2^{j-1}$-period log returns of VXX ($r_{logVXX,t}^{(2^{j-1})}$) calculated over the period of length $2^{j-1}$ from time $t - 2^{j-1} + 1$ through time $t$,
- to one-period log returns of VXX ($r_{logVXX,t}^{(1)}$) filtered from time $t - 2^j - 2$ till time $t$ with a triangular filter of length $2^j - 1$.

An analogous interpretation applies to the Haar wavelet coefficients for logVIX. Consequently, if we plot Haar wavelet coefficients of level 7 for logVXX and for logVIX (see Figure 11), these coefficients have one of the three possible interpretations given above. The D(4) wavelet coefficients given in the middle plot of Figure 9 can be interpreted in a qualitatively analogous manner. Specifically, they are associated with changes in weighted averages (of log level), the weighted averages being calculated effectively on scale $2^{j-1}$. Since D(4) wavelet filters are better approximations to the ideal filter for the respective passbands than Haar wavelet filters, the time series of D(4) wavelet coefficients are smoother and less noisy compared to the Haar filters. This makes it easier for the D(4) coefficients to visually identify the lead/lag behavior in the turning points.

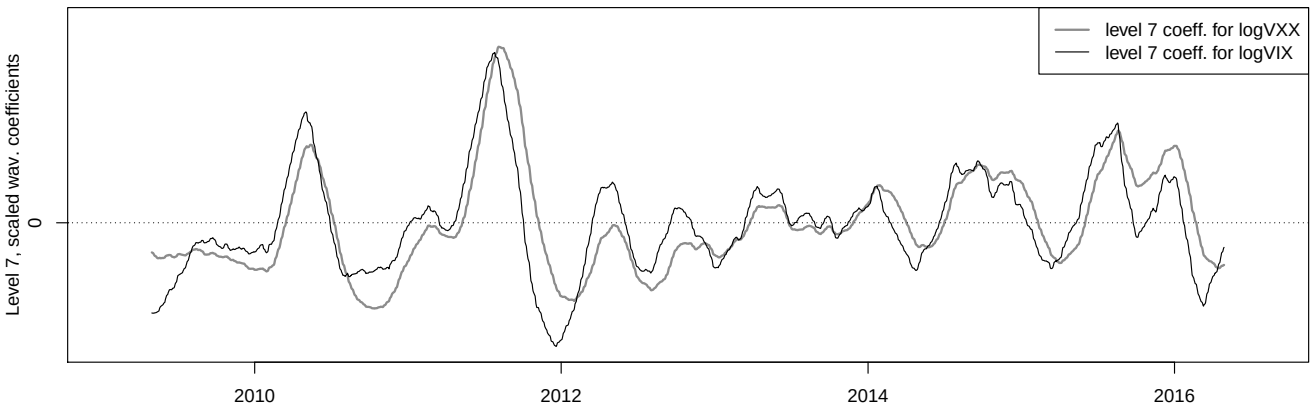


Figure 11: Haar wavelet coefficients of level 7 for logVXX (gray) and logVIX (black). Both sets of coefficients have been centered to zero and divided by the respective standard deviation to allow for better comparison of the coefficients.

# 6 Modeling the joint dynamics of logVXX and logVIX

In Section 6.1 we propose a bivariate model for logVXX and logVIX which is capable of reproducing the observed characteristics of logVXX and logVIX as well as the observed joint dynamics between the time series. In Section 6.2 the model is estimated and further refined.

## 6.1 Model

We suggest that the properties of logVXX and logVIX and their joint dynamics can be explained by a bivariate vector autoregression model (VAR) for log returns of VXX ($r_{logVXX,t}$) and level of logVIX ($logVIX_t$) where specific regressors are used on the right-hand side of the model equations. Specifically, let us write $r_{logVXX,t}$ and $logVIX_t$ as follows

$$r_{logVXX,t} = \tilde{\alpha}_0 + \sum_{k\in\{1,3,5,7,9\}} \tilde{\alpha}_k logVIX_{t-1}^{(2^{k-1})} + \sum_{k\in\{1,3,5,7,9\}} \tilde{\beta}_k r_{logVXX,t-1}^{(2^{k-1})} + \epsilon_{1,t}, \tag{20}$$

$$logVIX_t = \alpha_0 + \sum_{k\in\{1,3,5,7,9\}} \alpha_k logVIX_{t-1}^{(2^{k-1})} + \sum_{k\in\{1,3,5,7,9\}} \beta_k r_{logVXX,t-1}^{(2^{k-1})} + \epsilon_{2,t}, \tag{21}$$

where $\{\epsilon_{1,t}\}$ and $\{\epsilon_{2,t}\}$ are components of a bivariate white noise process and

$$r_{logVXX,t-1}^{(2^{k-1})} = \frac{1}{2^{k-1}} \sum_{i=t-2^{k-1}}^{t-1} r_{logVXX,i}, \tag{22}$$

$$logVIX_{t-1}^{(2^{k-1})} = \frac{1}{2^{k-1}} \sum_{i=t-2^{k-1}}^{t-1} logVIX_i. \tag{23}$$

The model resembles the standard VAR model but explicitly takes into account the fact that the dependence between the variables operates over various scales. In this sense, it is also a generalization of the univariate heterogeneous autoregressive (HAR) model of Corsi (2009). The HAR model has become increasingly popular in finance because, despite its simple structure, it reproduces the main empirical features of volatility very well. Moreover, the HAR model also exhibits very good forecasting performance. As a result, the HAR model and its various extensions have been applied not only to equity markets (Lyócsa and Molnár, 2017), but also to other financial assets such as currencies (Lyócsa et al., 2016) and commodities (Haugom et al., 2014; Lyócsa and Molnár, 2016). The HAR model is based on the heterogeneous market hypothesis, which assumes that various agents in financial markets have various investment or trading horizons, and therefore contribute differently to volatility. As a result, the model consists of volatility components realized over various time scales, see Corsi (2009).

The various scales used in Equations 20 and 21 are of the form $2^{k-1}$ so that $k$ can be loosely interpreted as the corresponding level of wavelet analysis. We have assumed only odd values of $k$ in order not to over-parametrize the model and to include large scales at the same time.

Since logVXX has a unit root, it is used in the first difference in the model, i.e. as $r_{logVXX,t} = logVXX_t - logVXX_{t-1}$, whereas logVIX is used in levels in the model since it does not have a unit root and modeling logVIX in levels better corresponds to the standard approaches to modeling volatility time series.

### 6.2 Parameter estimation and model refinement

We have estimated the parameters of both the equations 20 and 21 separately by ordinary least squares; the estimates are reported in Table 4.

Table 4: Estimated parameters of the model of Equation 20 and 21, estimated standard errors and p-values.

| | Model for $r_{logVXX}$ | | | | Model for $logVIX$ | | |
|---|---|---|---|---|---|---|---|
| | Estimate | Stand. error | p-value | | Estimate | Stand. error | p-value |
| $\tilde{\alpha}_0$ | 0.022 | 0.014 | 0.13 | $\alpha_0$ | 0.045 | 0.027 | 0.09 |
| $\tilde{\alpha}_1$ | 0.048 | 0.027 | 0.08 | $\alpha_1$ | 0.895 | 0.051 | <0.001 |
| $\tilde{\alpha}_3$ | -0.033 | 0.029 | 0.25 | $\alpha_3$ | 0.014 | 0.053 | 0.80 |
| $\tilde{\alpha}_5$ | -0.005 | 0.017 | 0.79 | $\alpha_5$ | 0.050 | 0.031 | 0.11 |
| $\tilde{\alpha}_7$ | -0.008 | 0.012 | 0.53 | $\alpha_7$ | 0.016 | 0.022 | 0.47 |
| $\tilde{\alpha}_9$ | -0.013 | 0.012 | 0.28 | $\alpha_9$ | 0.008 | 0.022 | 0.74 |
| $\tilde{\beta}_1$ | -0.053 | 0.039 | 0.17 | $\beta_1$ | -0.008 | 0.071 | 0.91 |
| $\tilde{\beta}_3$ | -0.075 | 0.087 | 0.39 | $\beta_3$ | 0.086 | 0.160 | 0.59 |
| $\tilde{\beta}_5$ | -0.224 | 0.198 | 0.26 | $\beta_5$ | 0.082 | 0.365 | 0.82 |
| $\tilde{\beta}_7$ | 0.052 | 0.416 | 0.90 | $\beta_7$ | 0.177 | 0.766 | 0.82 |
| $\tilde{\beta}_9$ | -1.051 | 0.627 | 0.09 | $\beta_9$ | -1.721 | 1.154 | 0.14 |

Using the two sets of residuals from the estimated equations and the results of equations 8, 16 and 17 we have checked that errors $\{\epsilon_{1,t}\}$ and $\{\epsilon_{2,t}\}$ can be considered components of a bivariate white noise process. Moreover, $\{\epsilon_{1,t}\}$ and $\{\epsilon_{2,t}\}$ exhibit slight non-normality and heteroskedasticity. This slight non-normality and heteroskedasticity will be neglected in our further analysis.

The above given model can be further simplified using hypothesis testing. At first, we test the following hypotheses

$$A.)\ H_0 : \tilde{\alpha}_1 = \tilde{\alpha}_3 = ... = \tilde{\alpha}_9 = \tilde{\beta}_1 = \tilde{\beta}_3 ... = \tilde{\beta}_9 = 0, \quad H_1 : \text{non } H_0. \tag{24}$$

$$B.)\ H_0 : \alpha_1 = \alpha_3 = ... = \alpha_9 = \beta_1 = \beta_3 = ... = \beta_9 = 0, \quad H_1 : \text{non } H_0, \tag{25}$$

Since the first null hypothesis (A.) is not rejected at the 5% significance level (p-value = 0.31), we postulate that

$$r_{logVXX,t} = \tilde{\alpha}_0 + \epsilon_{1,t}. \tag{26}$$

On the other hand, since the second null hypothesis (B.) is rejected (p-value < 0.001), we continue with the following tests

$$C.)\ H_0 : \alpha_1 = \alpha_3 = ... = \alpha_9 = 0, \quad H_1 : \text{non } H_0, \tag{27}$$

$$D.)\ H_0 : \beta_1 = \beta_3 = ... = \beta_9 = 0, \quad H_1 : \text{non } H_0. \tag{28}$$

Since the C.) null hypothesis is rejected (p-value < 0.001), whereas hypothesis D.) is not (p-value = 0.76),

we postulate that

$$logVIX_t = \alpha_0 + \sum_{k\in\{1,3,5,7,9\}} \alpha_k logVIX_{t-1}^{(2^{k-1})} + \epsilon_{2,t}. \tag{29}$$

We estimate Equations 26 and 29 separately by ordinary least squares. We assess whether the model for logVIX (see Equation 29) can be further refined, while the model for logVXX remains unchanged as dictated by Equation 26. To refine the model for logVIX, we start with Equation 29 and remove the regressor with the largest p-value provided that this value is larger than 0.05. The model with the removed regressor is then re-estimated and the procedure of removing regressors is repeated until all the p-values in the re-estimated model are less than or equal to 0.05. The sequence of models in this elimination process is presented in the first four rows of Table 5 (see models 1 through 4 in the table) and ends with a model that has only two regressors (not counting the intercept) both of which are statistically significant (model 4 in the table).

For comparison, a standard AR(1) model is also included (model 5 in the table) as well as a model (model 6 in the table) that is similar to model 4 but uses the more familiar scales of 1 day and 1 month (= 22 trading days) instead of the dyadic scales of 1 and 16 days employed in model 4. Model 7 with the structure of the traditional HAR volatility model of Corsi (2009) is also examined. The model for logVXX remains as given in Equation 26 for all the models for logVIX given in the table. The parameters of the equations for logVXX and logVIX are estimated separately by ordinary least squares, the response variable covering the same time period in all the seven cases.

We provide a check of the univariate properties for each model in the table (column "Univar." in Table 5). Specifically, using wavelet variance, the residuals (from the equations for logVXX and logVIX) are assessed for whether they can be considered white noise processes (among other things, we also make use of Equation 8 to perform this assessment), and the time series generated by the estimated equations are assessed for whether they exhibit wavelet variances similar to those observed empirically for logVXX and logVIX. The result of the evaluation of the univariate properties may be either "Good" (for a good/perfect agreement) or "Bad" (otherwise). We conclude that all the models in the table except model 5 provide white noise residuals and are capable of reproducing the univariate characteristics observed for logVXX and logVIX well.

Further, we perform a check of the bivariate properties for each model (column "Bivariate" in Table 5). Specifically, the residuals from the equations for logVXX and logVIX are assessed – using the results of Equations 16 and 17 – for whether they can be considered bivariate white noise, and the wavelet cross-correlation functions implied by the model are assessed for whether they coincide well with the empirical cross-correlation functions between logVXX and logVIX. Again, the results of the evaluation for each model can be either "Good" or "Bad". All the models except model 5 can be considered good at capturing the observed bivariate characteristics. Model 5 can approximately reproduce the empirical lead-lag relationship between logVXX and logVIX that is observed empirically, but fails to reproduce the magnitude of the delay and the strength of the relationship at different scales.

Further, we assess whether the estimated parameters of the models imply a unit root (see column "unit-root"). If no unit root is implied by the estimated parameters of the model (which is the case for all the models), the long-term mean of the corresponding process is estimated (column "mean").

Since models 4 and 6 for logVIX require the lowest number of parameters, they can be considered the

best. Their estimated parameters are reported in Table 6 and Table 7, the estimated variance of $\{\epsilon_1\}$, of $\{\epsilon_2\}$ and the estimated correlation between $\{\epsilon_1\}$ and $\{\epsilon_2\}$ being 0.0016, 0.0055 and 0.90 both for model 4 and model 6.

Table 5: Various explored models for logVIX. See text for the description.

| Model # | Model | unit-root | mean | Univar. | Bivariate |
|---|---|---|---|---|---|
| 1 | $logVIX_t = \alpha_0 + \sum_{k\in\{1,3,5,7,9\}} \alpha_k logVIX_{t-1}^{(2^{k-1})} + \epsilon_{2,t}$ | NO | 2.79 | Good | Good |
| 2 | $logVIX_t = \alpha_0 + \sum_{k\in\{1,3,5,9\}} \alpha_k logVIX_{t-1}^{(2^{k-1})} + \epsilon_{2,t}$ | NO | 2.79 | Good | Good |
| 3 | $logVIX_t = \alpha_0 + \sum_{k\in\{1,5,9\}} \alpha_k logVIX_{t-1}^{(2^{k-1})} + \epsilon_{2,t}$ | NO | 2.79 | Good | Good |
| 4 | $logVIX_t = \alpha_0 + \sum_{k\in\{1,5\}} \alpha_k logVIX_{t-1}^{(2^{k-1})} + \epsilon_{2,t}$ | NO | 2.83 | Good | Good |
| 5 | $logVIX_t = \alpha_0 + \alpha_1 logVIX_{t-1} + \epsilon_{2,t}$ | NO | 2.84 | Bad | Bad |
| 6 | $logVIX_t = \alpha_0 + \sum_{z\in\{1,22\}} \gamma_z logVIX_{t-1}^{(z)} + \epsilon_{2,t}$ | NO | 2.83 | Good | Good |
| 7 | $logVIX_t = \alpha_0 + \sum_{z\in\{1,5,22\}} \gamma_z logVIX_{t-1}^{(z)} + \epsilon_{2,t}$ | NO | 2.83 | Good | Good |

Table 6: Estimated parameters of the model of Equation 26 and of Model 4 from Table 5, estimated standard errors and p-values.

| Model for $r_{logVXX}$ | | | | Model 4 for $logVIX$ | | | |
|---|---|---|---|---|---|---|---|
| | Estimate | Stand. error | p-value | | Estimate | Stand. error | p-value |
| $\tilde{\alpha}_0$ | -0.003 | 0.001 | 0.001 | $\alpha_0$ | 0.072 | 0.020 | <0.001 |
| | | | | $\alpha_1$ | 0.916 | 0.014 | <0.001 |
| | | | | $\alpha_5$ | 0.059 | 0.015 | <0.001 |

Table 7: Estimated parameters of the model of Equation 26 and of Model 6 from Table 5, estimated standard errors and p-values.

| Model for $r_{logVXX}$ | | | | Model 6 for $logVIX$ | | | |
|---|---|---|---|---|---|---|---|
| | Estimate | Stand. error | p-value | | Estimate | Stand. error | p-value |
| $\tilde{\alpha}_0$ | -0.003 | 0.001 | 0.001 | $\alpha_0$ | 0.069 | 0.021 | 0.001 |
| | | | | $\gamma_1$ | 0.924 | 0.013 | <0.001 |
| | | | | $\gamma_{22}$ | 0.051 | 0.014 | <0.001 |

### 6.3 Model verification with simulated data

In order to verify that model 6 is capable of reproducing the observed bivariate characteristics of logVXX and logVIX, we have also simulated a pair of time series of logVXX and logVIX from model 6 (see Figure 12). For the simulated pair we have calculated wavelet cross-correlations and plotted level 7 wavelet coefficients (see Figure 13). We see that model 6 is indeed capable of capturing the observed bivariate characteristics of the original time series. More specifically, the simulated data of Figure 12 resemble those of Figure 3. The peaks and troughs in the VIX index seem to be leading the peaks and troughs in VXX, for both the real and the simulated data. Further, the wavelet cross-correlation functions calculated for the simulated pair resemble those for the original data in the sense that no lead/lag relationship is observed on short scales for the simulated pair, whereas VIX leads VXX on longer scales (compare Figure 9 with Figure 13). Moreover, the strength of the relationship and the amount by which VIX leads

VXX on longer scales for the simulated pair (16 days) are in qualitative agreement with those observed for the original data (15 days). Similar agreement between the bivariate characteristics of the simulated pair and the bivariate characteristics of the original data also occurs for other simulated pairs of time series from model 6, but for the sake of brevity we do not report the results from these other simulations.

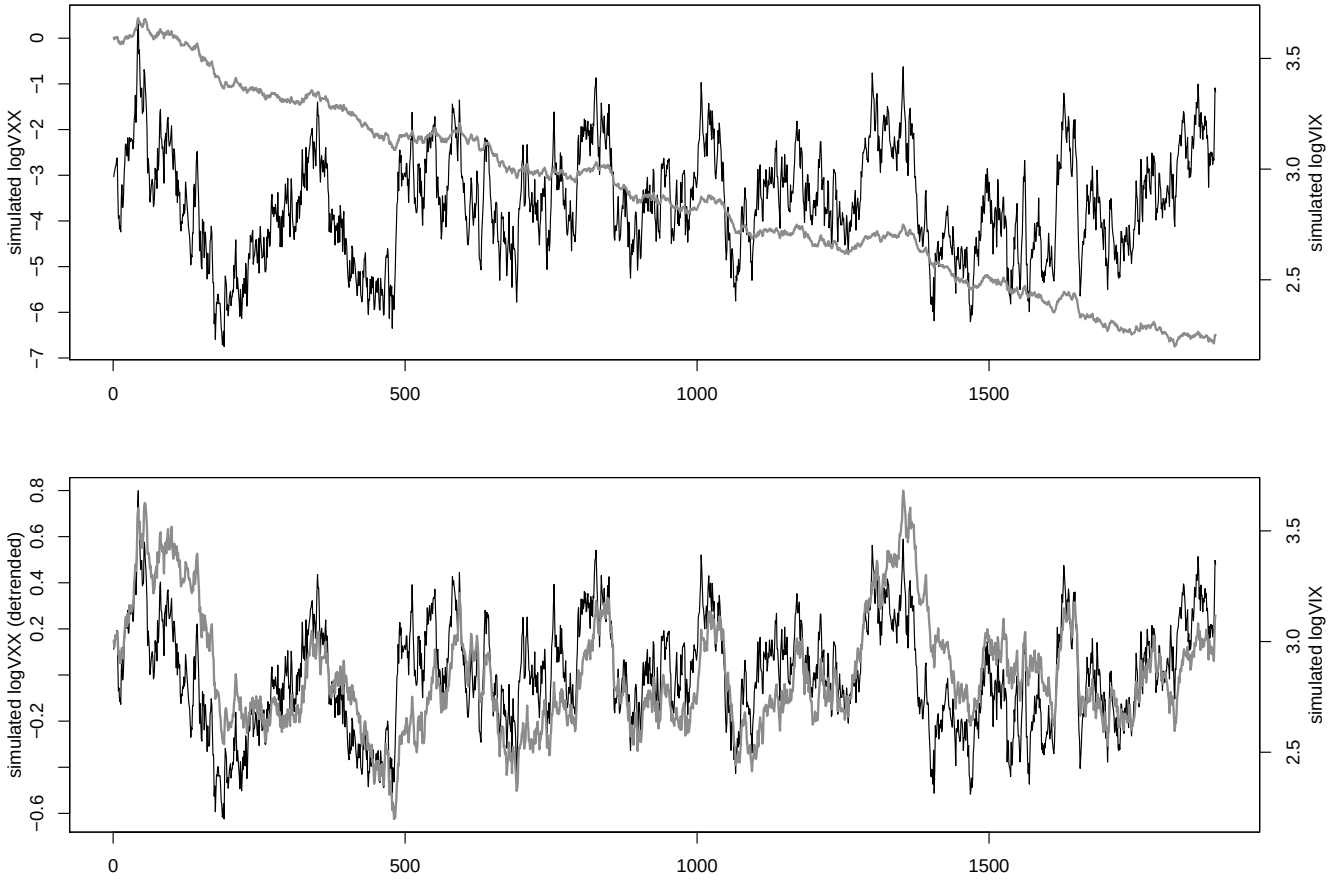


Figure 12: *Top plot*: Simulated logVXX (gray, left vertical axis) and simulated logVIX (black, right vertical axis) of the same lengths as the original time series. *Bottom plot*: The same as the top plot, but simulated logVXX is detrended.

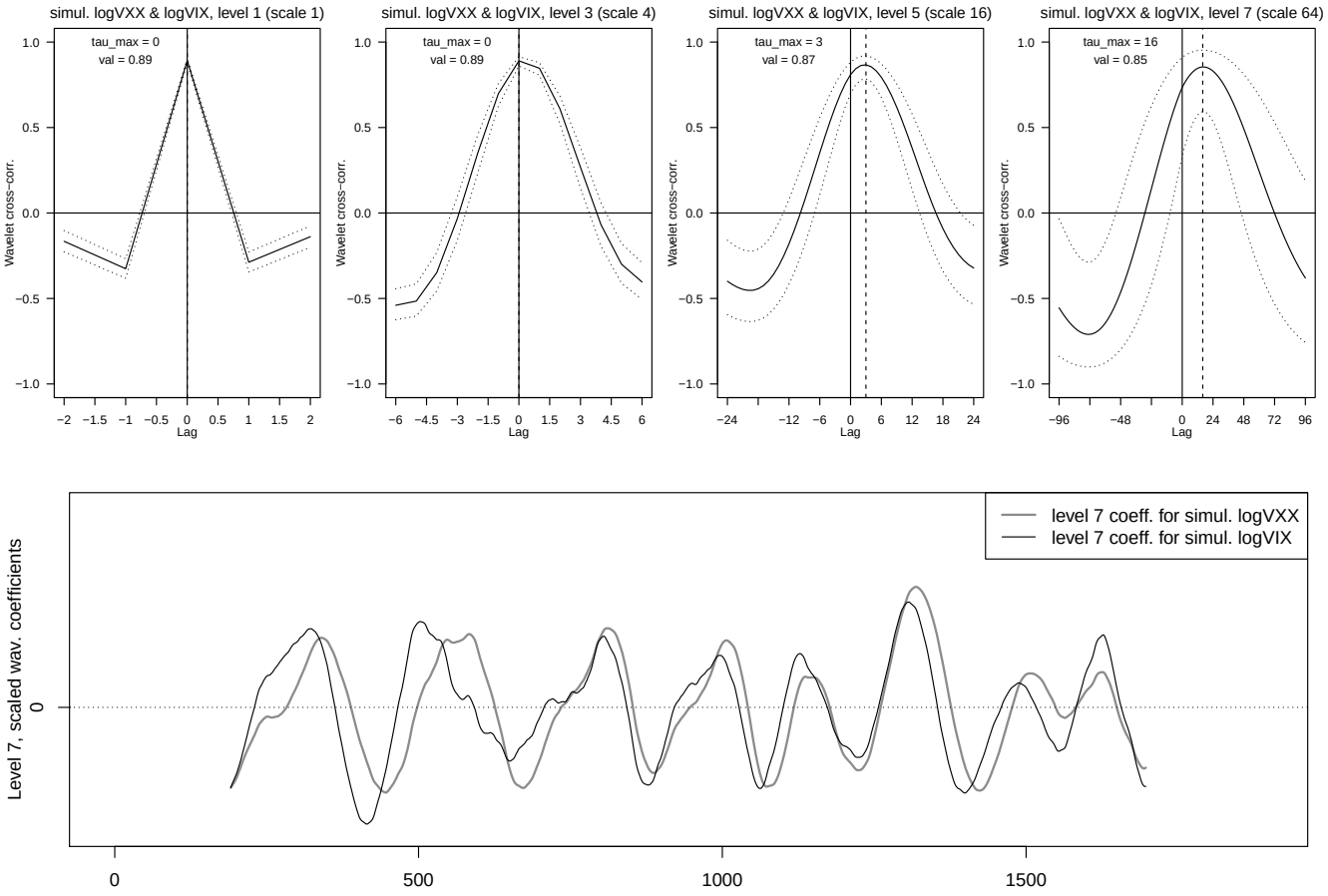


Figure 13: *Top plot*: D(4) wavelet cross-correlation functions between simulated logVXX ($Y_t$) and logVIX ($X_t$) for levels 1, 3, 5 and 7 (solid line) with the respective 95% confidence intervals depicted by dotted lines. *Middle plot*: Time series of D(4) scaled level 7 wavelet coefficients for simulated logVXX (gray) and D(4) scaled level 7 wavelet coefficients for logVIX (black); both the series have been synchronized.

# 7 Conclusion

The VIX index, the volatility implied by S&P 500 option prices, has existed for several decades and has become one of the most followed financial indices in the world. However, the VIX index itself is not tradable. Therefore, until recently, it was impossible to trade volatility directly. In 2004, futures on the VIX index emerged. Since then, direct trading of volatility has become increasingly popular and other volatility products have been emerging. The most popular of these, the exchange traded note VXX, was introduced in February 2009. The VIX futures and VXX move very closely together. However, unlike VIX futures, which are accessible only to larger investors, VXX can be traded as easily as a simple stock.

We study the comovement between VIX and VXX on time scales varying from days to months. Wavelet analysis reveals that on time scales of days no lead/lag relationship between VIX and VXX is present. However, the VIX index leads VXX on time scales of several months. This can be observed even directly from the plots of VIX and VXX, where peaks and troughs in VIX tend to occur before the peaks and troughs in VXX.

Next we construct a joint model for VIX and VXX that is in accordance with the observed charac-

teristics of these two time series. Surprisingly, we find a rather simple model, in which VIX and VXX are related only via a correlated error term. In this joint model, the logarithm of the VIX index has a structure similar to that of the HAR model of Corsi (2009), whereas the return of VXX is modeled as a white noise with a constant. These results imply that despite the VIX index is leading VXX, there is no apparent trading opportunity, since information from VIX index cannot be utilized for improving forecasts of VXX.

## References


Aguiar-Conraria, L., Rodrigues, T. M., and Soares, M. J. (2014). Oil shocks and the euro as an optimum currency area. In *Wavelet applications in economics and finance*, pages 143–156. Springer.

Aguiar-Conraria, L. and Soares, M. J. (2011). Oil and the macroeconomy: using wavelets to analyze old issues. *Empirical Economics*, 40(3):645–655.

Alexander, C., Kapraun, J., and Korovilas, D. (2015). Trading and investing in volatility products. *Financial Markets, Institutions & Instruments*, 24(4):313–347.

Alexander, C. and Korovilas, D. (2013). Volatility exchange-traded notes: curse or cure? *The Journal of Alternative Investments*, 16(2):52–70.

Alexander, C., Korovilas, D., and Kapraun, J. (2016). Diversification with volatility products. *Journal of International Money and Finance*, 65:213–235.

Bašta, M. and Molnár, P. (2018). Oil market volatility and stock market volatility. *Finance Research Letters*.

Birkelund, O. H., Haugom, E., Molnár, P., Opdal, M., and Westgaard, S. (2015). A comparison of implied and realized volatility in the Nordic power forward market. *Energy Economics*, 48:288–294.

Bordonado, C., Molnár, P., and Samdal, S. R. (2017). VIX exchange traded products: Price discovery, hedging, and trading strategy. *Journal of Futures Markets*, 37(2):164–183.

Bugge, S. A., Guttormsen, H. J., Molnár, P., and Ringdal, M. (2016). Implied volatility index for the Norwegian equity market. *International Review of Financial Analysis*, 47:133–141.

Chakrabarty, A., De, A., Gunasekaran, A., and Dubey, R. (2015). Investment horizon heterogeneity and wavelet: Overview and further research directions. *Physica A: Statistical Mechanics and its Applications*, 429:45–61.

Chen, Y.-L. and Tsai, W.-C. (2017). Determinants of price discovery in the VIX futures market. *Journal of Empirical Finance*, pages 59–73.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196.

Crowley, P. M. and Lee, J. (2005). Decomposing the co-movement of the business cycle: a time-frequency analysis of growth cycles in the euro area.

Crowley, P. M. and Mayes, D. G. (2009). How fused is the euro area core? *OECD Journal: Journal of Business Cycle Measurement and Analysis*, 2008(1):63–95.

Fernández-Macho, J. (2012). Wavelet multiple correlation and cross-correlation: a multiscale analysis of eurozone stock markets. *Physica A: Statistical Mechanics and its Applications*, 391(4):1097–1104.

Frijns, B., Tourani-Rad, A., and Webb, R. I. (2016). On the intraday relation between the VIX and its futures. *Journal of Futures Markets*, 36(9):870–886.

Gallegati, M. (2008). Wavelet analysis of stock returns and aggregate economic activity. *Computational Statistics & Data Analysis*, 52(6):3061–3074.

Graham, M., Kiviaho, J., and Nikkinen, J. (2013). Short-term and long-term dependencies of the S&P 500 index and commodity prices. *Quantitative Finance*, 13(4):583–592.

Haugom, E., Langeland, H., Molnár, P., and Westgaard, S. (2014). Forecasting volatility of the US oil market. *Journal of Banking & Finance*, 47:1–14.

In, F. and Kim, S. (2006). Multiscale hedge ratio between the Australian stock and futures markets: Evidence from wavelet analysis. *Journal of Multinational Financial Management*, 16(4):411–423.

Kim, S. and In, F. (2005). The relationship between stock returns and inflation: new evidence from wavelet analysis. *Journal of Empirical Finance*, 12(3):435–444.

Lu, Z. and Zhu, Y. (2010). Volatility components: The term structure dynamics of VIX futures. *Journal of Futures Markets*, 30(3):230–256.

Lyócsa, Š. and Molnár, P. (2016). Volatility forecasting of strategically linked commodity ETFs: gold-silver. *Quantitative Finance*, 16(12):1809–1822.

Lyócsa, Š. and Molnár, P. (2017). The effect of non-trading days on volatility forecasts in equity markets. *Finance Research Letters*.

Lyócsa, S., Molnár, P., and Fedorko, I. (2016). Forecasting exchange rate volatility: The case of the Czech Republic, Hungary and Poland. *Finance a Uver*, 66(5):453.

Naccache, T. (2011). Oil price cycles and wavelets. *Energy Economics*, 33(2):338–352.

Percival, D. B. and Walden, A. T. (2006). *Wavelet methods for time series analysis*. Cambridge university press.

Rua, A. and Nunes, L. C. (2009). International comovement of stock market returns: A wavelet analysis. *Journal of Empirical Finance*, 16(4):632–639.

Shu, J. and Zhang, J. E. (2012). Causality in the VIX futures market. *Journal of Futures Markets*, 32(1):24–46.

Vacha, L. and Barunik, J. (2012). Co-movement of energy commodities revisited: Evidence from wavelet coherence analysis. *Energy Economics*, 34(1):241–247.

Whaley, R. E. (2013). Trading volatility: At what cost? *The Journal of Portfolio Management*, 40(1):95–108.

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

Zhang, J. E., Shu, J., and Brenner, M. (2010). The new market for volatility trading. *Journal of Futures Markets*, 30(9):809–833.