Long Seasonal Cycle Modeling: the Case of Realized Volatility

Jiří Procházka [1] | *University of Economics, Prague, Czech Republic*
Milan Bašta [2] | *University of Economics, Prague, Czech Republic*
Samuel Flimmel [3] | *University of Economics, Prague, Czech Republic*
Matej Čamaj [4] | *University of Economics, Prague, Czech Republic*

Abstract

Time series with long seasonal periods are very common. Several methods have been proposed for modeling of long seasonal cycles, the most commonly used ones being those based on basis expansion. In this paper, we present and discuss these methods. We also use them to model seasonality in realized volatility of several major stock market indices and find evidence for the existence of yearly as well as weekly seasonality. The presented approaches can potentially be used for modeling of any seasonal time series with a long seasonal period.

Keywords	JEL code
Time series, seasonality, basis expansion, volatility	*C58, C22*

INTRODUCTION

While seasonality of returns has been studied thoroughly, seasonality in volatility of stock markets has received much less attention in the literature. Back et al. (2013) showed that including a seasonal component in volatility into an option pricing model greatly improves the model performance. This was further studied by Arismendi et al. (2016) who proposed a seasonally varying long-run mean variance process. Day-of-the-week effect in volatility of stock markets during the period from 1988 through 2002 was studied in Kiymaz and Berument (2003) who found that the highest volatility occurs on Mondays for Germany and Japan, on Fridays for Canada and the United States, and on Thursdays for the United Kingdom. Furthermore, Giovanis (2009) studied the calendar effects (turn-of-the-month, day-of-the-week, month-of-the-year and semi-month effect) in the volatility of 55 stock market indices and

[1] Department of Statistics and Probability, Faculty of Informatics and Statistics, University of Economics, Prague, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic.

[2] Department of Statistics and Probability, Faculty of Informatics and Statistics, University of Economics, Prague, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic. Corresponding author: e-mail: milan.basta@vse.cz.

[3] Department of Statistics and Probability, Faculty of Informatics and Statistics, University of Economics, Prague, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic.

[4] Department of Statistics and Probability, Faculty of Informatics and Statistics, University of Economics, Prague, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic.

concluded that there is a day-of-the-week and month-of-the-year effect in volatility. Another seasonal anomaly in volatility was reported by Seyyed et al. (2005) who studied data from Saudi Arabian stock market during Ramadan and found a significant and predictable decrease of volatility during this period.

In our paper, we model the *yearly* seasonal cycle in *daily* time series of realized volatility of 19 stock market indices in the period from 2000 till 2017. Such a seasonal cycle is an example of seasonality with a long seasonal period, namely 365 days. We investigate various approaches to modeling of such a seasonal cycle that are based on basis expansion. We demonstrate that the choice of the basis functions is a crucial step in the modeling process. The results we obtain may play an important role in volatility modeling. Moreover, the approaches we present are more general since they can potentially be used for modeling of any time series with a long seasonal period.

The paper is organized as follows. Section 1 introduces the approaches to seasonality modeling based on basis expansion methods. Section 2 performs the analysis of the 19 realized volatility time series. Last section concludes.

1 BASIS EXPANSION METHODS FOR MODELING SEASONALITY

In the following text a daily time series of realized volatility of length N will be denoted as $\{X_t : t = 1, \ldots N\}$, where time t corresponds to *trading* days. We will assume the following model for $\{X_t\}$:

$$X_t = S_{1,m(t)} + S_{2,n(t)} + E_t \qquad t = 1, \ldots, N, \tag{1}$$

where $\{S_{1,m(t)}\}$ is a yearly deterministic seasonal component *including an intercept* and $\{S_{2,n(t)}\}$ a weekly deterministic seasonal component. $m(t)$, being a function of t, is the *calendar day of the year* and can take any value from 1 to $L = 365$. Leap days have been removed from the analysis for simplicity. Analogously, $n(t)$ is the *day of the week*, where 1 = Monday, 2 = Tuesday, 3 = Wednesday, 4 = Thursday and 5 = Friday. $\{E_t : t = 1, \ldots N\}$ is a stationary ARMA(p, q) term given as:

$$E_t = \phi_1 E_{t-1} + \ldots + \phi_p E_{t-p} + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q} + \epsilon_t, \qquad t = 1, \ldots, N, \tag{2}$$

where $\{\phi_i : i = 1, \ldots, p\}$ and $\{\theta_i : i = 1, \ldots, q\}$ are parameters and $\{\epsilon_t : t = 1, \ldots, N\}$ is Gaussian white noise.

In the paragraphs below, we introduce several approaches to modeling $\{S_{1,m(t)}\}$ based on basis expansion. By basis expansion we mean rewriting $\{S_{1,m(t)}\}$ as a linear combination of "basis functions".

1.1 Dummy variables

The most common approach to modeling yearly seasonal cycle in daily time series is rewriting $S_{1,m(t)}$ as a sum of an intercept term and 11 dummy variables. Specifically, we can write:[5]

$$S_{1,m} = c + \sum_{k=2}^{12} \alpha_k \psi_k, \qquad m = 1, \ldots, L, \tag{3}$$

where c is an intercept and α_k, for $k = 2, \ldots 12$, are parameters and ψ_k, for $k = 2, \ldots, 12$, is a dummy variable which is equal to 1 if calendar day m is part of month k, and to 0 if day m is not part of month k.[6]

---

[5] For simplicity, we write $S_{1,m}$ instead of $S_{1,m(t)}$.
[6] k = 2 is February, k = 3 is March, … , k = 12 is December.

1.2 Fourier basis

Now, we introduce the approach of representing $\{S_{1,m(t)}\}$ by means of Fourier basis functions, which amounts to rewriting $\{S_{1,m(t)}\}$ as a sum of sines and cosines of different frequencies and amplitudes. More specifically, we can write:[7]

$$S_{1,m} = c + \sum_{k=1}^{K} \alpha_k \sin\left(2\pi \frac{k}{L} m\right) + \sum_{k=1}^{K} \beta_k \cos\left(2\pi \frac{k}{L} m\right), \quad m = 1, \ldots, L, \tag{4}$$

where c is an intercept and α_k and β_k, for $k = 1, \ldots, K$, are parameters. K can potentially take any value from 1 to 182, assuming that $L = 365$. The number of parameters used in the expansion (including the intercept) is $1 + 2K$.

The choice of K has a direct impact on the smoothness of the seasonal cycle. A large value of K results in low as well as high-frequency features being present in the seasonal cycle. On the other hand, a small value of K results only in low-frequency features being present in the cycle.

1.3 Periodic linear spline

A potential disadvantage of using the decomposition of $\{S_{1,m(t)}\}$ into sines and cosines (see Section 1.2) is the fact that sines and cosines are *not* localized in time, which can make *local* features in the seasonal component difficult to model. Consequently, we consider another approach to modeling the seasonal component. Specifically, we represent it using a regression *linear* spline with K knots.

Specifically, let the linear spline function $c(m)$, now considered as a function of a *continuous* variable m, be defined on the interval $[0, L]$ as follows (see Friedman, Hastie and Tibshirani, 2001; Ramsay and Silverman, 2002; or Ramsay and Silverman, 2005):

$$c(m) = \beta_0 + \beta_1 m + \sum_{k=1}^{K} \delta_k (m - \xi_k)_+, \quad 0 \le m \le L, \tag{5}$$

where β_0, β_1, and δ_k, for $k = 1, \ldots, K$, are parameters and $(.)_+$ denotes the positive part of the expression inside the brackets. ξ_k, for $k = 1, \ldots, K$, are (real or whole) numbers called *knots* which satisfy $0 \le \xi_1 < \xi_2 < \ldots < \xi_K < L$. From the definition, it follows that the linear spline function is continuous on the interval $[0, L]$. Further, it is piecewise linear since it is a linear function on each of the intervals $[0, \xi_1), [\xi_1, \xi_2), \ldots, [\xi_L, L]$ separately. The first derivative does not exist at the knots.

One potential disadvantage of using the linear spline function (compared to sines and cosines of Section 1.2) is the fact that the linear spline function is not "periodic" with period equal to L. We suggest that periodicity be imposed by assuming appropriate constrains. Specifically, we require that:

$$c(0) = c(L), \tag{6}$$

$$c(0)^{(1)} = c(L)^{(1)}, \tag{7}$$

where $c(0)^{(1)}$ denotes the first right derivative at point 0 and $c(L)^{(1)}$ the first left derivate at point L. These constraints ensure that $c(m)$ returns to the same value after period L (the first constraint) and that the "connection of the two ends of the cycle" is smooth (the second constraint).[8] The two constraints reduce the number of parameters of the linear spline function by 2. Specifically, a linear spline function satisfying the constraints will be called a periodic linear spline, has $2 + K - 2 = K$ parameters and can be written as a linear combination of K basis functions. Figure 1 presents an example of a periodic linear spline.

---

[7] For simplicity, we write $S_{1,m}$ instead of $S_{1,m(t)}$.
[8] If the first knot ξ_1 is placed at 0, then the first derivative of $c(m)$ does not exist at 0.

If the periodic linear spline is sampled at discrete values of $m = 1, 2, ..., L$, it can be used as a representation of the seasonal component $\{S_{1,m(t)}\}$ of Formula (1). An appropriate choice of the position of the knots can help us to capture local features in the seasonal component.

Figure 1 An example of a periodic linear spline (solid line) with $L = 365$ and three knots placed at 73, 110, 329 (dashed vertical lines)
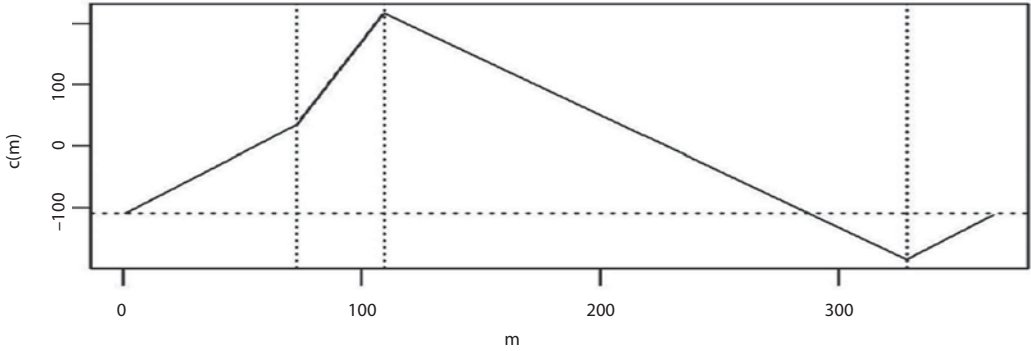
Source: Own construction

1.4 Periodic cubic spline

To ensure a more flexible shape of the function on the intervals $[0, \xi_1), [\xi_1, \xi_2), ..., [\xi_L, L]$ and more smoothness at the knots, i.e. the existence of (at least some) derivatives at the knots, we can consider using cubic spline functions instead of linear ones.

Using an analogous notation as in Section 1.3, a cubic spline function with K knots $0 \le \xi_1 < \xi_2 < ... < \xi_K < L$ can be defined on the interval $[0, L]$ as follows (see Friedman, Hastie and Tibshirani, 2001; Ramsay and Silverman, 2002; or Ramsay and Silverman, 2005):

$$c(m) = \beta_0 + \beta_1 m + \beta_2 m^2 + \beta_3 m^3 + \sum_{k=1}^{K} \delta_k \left(m - \xi_k\right)_+^3, \quad 0 \le m \le L, \tag{8}$$

where $\beta_0, \beta_1, \beta_2, \beta_3$ and δ_k, for $k = 1, ..., K$, are parameters. From the definition, it follows that the cubic spline function is continuous on the interval $[0, L]$. It is a piecewise cubic polynomial since it is a cubic polynomial on each of the intervals $[0, \xi_1), [\xi_1, \xi_2), ..., [\xi_L, L]$ separately. The first and second derivatives exist at the knots, whereas the third derivative does not.

Further, we require that:

$$c(0) = c(L), \tag{9}$$

$$c(0)^{(1)} = c(L)^{(1)}, \tag{10}$$

$$c(0)^{(2)} = c(L)^{(2)}, \tag{11}$$

$$c(0)^{(3)} = c(L)^{(3)}, \tag{12}$$

where $c(0)^{(1)}, c(0)^{(2)}, c(0)^{(3)}$ denote the first, second and third right derivative at point 0, and $c(L)^{(1)}, c(L)^{(2)}, c(L)^{(3)}$ denote the first, second and third left derivate at point L. Analogously to Section 1.3, the constraints ensure that $c(m)$ returns to the same value after period L (the first constraint) and that "the connection of the two ends of the cycle" is smooth (the second, third and fourth constraint).[9] Using these constraints,

---

[9] If the first knot ξ_1 is placed at 0, then the third derivative of c(m) does not exist at 0.

the number of parameters is reduced by 4. The cubic spline function which fulfills these constraints will be called a periodic cubic spline, has $4 + K - 4 = K$ parameters and can be written as a linear combination of K basis functions. Figure 2 presents an example of a periodic cubic spline.

If the periodic cubic spline is sampled at discrete values of $m = 1, 2, …, L$, it can be used as a representation of the seasonal component $\{S_{1,m(t)}\}$ of Formula (1).

If local features (sudden changes and bumps) are a priori expected to be present in some regions of the seasonal cycle, more knots can be placed in that region to allow for a more flexible representation of the seasonal cycle in that region (Ramsay and Silverman, 2005).

Figure 2 An example of a periodic cubic spline (solid line) with $L = 365$ and three knots placed at 73, 110, 329 (dashed vertical lines)
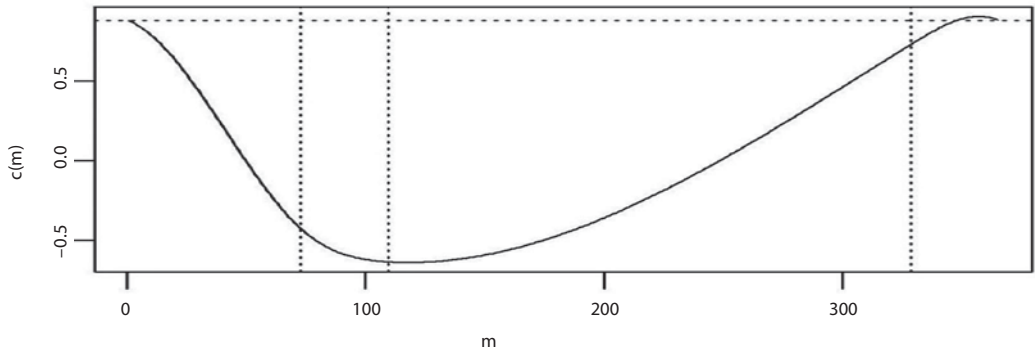
Source: Own construction

1.5 Model for weekly seasonality

To model the weekly seasonal cycle $\{S_{2,n(t)}\}$, we consider only the most common approach based on dummy variables. Specifically, we write:[10]

$$S_{2,n} = \sum_{k=2}^{5} \alpha_k \lambda_k, \quad n = 1,…,5, \tag{13}$$

where α_k, for $k = 2, …, 5$, are parameters and λ_k, for $k = 2, …, 5$, is a dummy variable which is equal to 1 if $k = n$, and to 0 otherwise.

1.6 Model selection

To find an optimal model for $\{X_t\}$, various representations of the yearly seasonal cycle (dummy variables, Fourier basis, periodic linear splines or periodic cubic splines) as well as various numbers of parameters for the yearly seasonal cycle (i.e. various numbers of sines/cosines or various numbers of knots) can be explored. The selection of the best model from the candidate models can be based on information criteria such as the Akaike information criterion (AIC) or the Bayesian information criterion (BIC) which are defined as:

$$AIC = 2p - 2 \cdot \hat{l}, \tag{14}$$

$$BIC = \log(N)\,p - 2 \cdot \hat{l}, \tag{15}$$

---

[10] For simplicity, we write $S_{2,n}$ instead of $S_{2,n(t)}$.

where p is the total number of estimated parameters in the model of Formula (1) and $\hat{l}$ is the natural logarithm of the maximized likelihood function. Models with the lowest values of AIC or BIC are considered as optimal models.

The terms $2p$ and $\log(N)\,p$ in Formulas (14) and (15) can be regarded as penalties. Consequently, BIC penalizes a model with a larger number of parameters more heavily than AIC provided that $\log(N) > 2$, which is the case of our data. AIC may lead to selection of too big a model. On the other hand, provided N is large enough, BIC may easily select too small a model. Consequently, we will work with both the criteria while selecting the optimal model, giving a little bit more preference to AIC.

A question arises as to what extent a model with the lowest value of AIC is better compared to another model M with a higher value of AIC. If the model with the lowest value of AIC is assigned "relevance" (weight) equal to 1, then the relevance of model M could be given as (Claeskens and Hjort, 2008):

$$\exp\left(-\frac{\Delta AIC_M}{2}\right), \tag{16}$$

where $\Delta AIC_M \geq 0$ is the difference between AIC of model M and AIC of the model with the lowest value of AIC. An analogous formula applies to comparing models by their BIC values. Namely, the "relevance" (weight) of model M is given as (Claeskens and Hjort, 2008):

$$\exp\left(-\frac{\Delta BIC_M}{2}\right), \tag{17}$$

where $\Delta BIC_M \geq 0$ is the difference between BIC of model M and BIC of the model with the lowest value of BIC.

2 EMPIRICAL ANALYSIS

In this section, we will use the above mentioned approaches to model yearly seasonality in the daily time series of realized volatility of several stock market indices:

- S&P 500, DJIA and NASDAQ 100 represent the American stock market.
- FTSE 100 represents the London Exchange. DAX, Euro STOXX 50 (both Germany), CAC 40 (France), AEX Index (Netherlands), Swiss Market Index (Switzerland), IBEX 35 (Spain) and FTSE MIB (Italy) represent the Continental European stock exchange.
- Nikkei 225 (Japan), Hang Seng (Hong Kong), KOSPI Composite Index (South Korea), FT Straits Time (Singapore) represent the Asian stock exchange.
- IPC Mexico and Bovespa Index represent the Mexican and Brazilian stock exchange.
- All Ordinaries represent the Australian stock exchange and the S&P/TSX Composite Index the Canadian one.

The data cover the period from January 3, 2000 till July 30, 2017[11] and have been obtained from Heber et al. (2009). Specifically, realized volatility (from the open to the close of the trading day) is available for each of the 19 indices being given as the square root of the sum of squared 5-minute log returns.[12] Details about the calculation of the realized volatility can be found in Liu, Patton and Sheppard (2012).

In Figure 3, we present the realized volatility time series for S&P 500, the logarithm of the time series and the annualized version of the realized volatility time series given in percentages obtained by multiplying the realized volatility time series by the square root of 252 and by 100.

---

[11] For some indices data are available only for shorter time spans.
[12] A trading day, which has, for example, 6.5 hours, has a total of 78 5-minute log returns.

We can see that periods of high realized volatility are present during the dot-com bubble (years 2000–2001), during the stock market downturn in 2002 and especially during the U.S. subprime mortgage crisis (2007–2009), which overlaps with the global financial crisis (2007–2008) and the Great Recession (2007–2009). Increased levels of volatility can also be observed during the recovery from the Great Recession (2010–2012), during the Chinese stock market crisis (2015–2016) and the Russian financial crisis (2014–2017).

The phenomena of volatility clustering and mean-reversion can be clearly discerned from the figure, clusters of high volatility being followed by clusters of low volatility. No obvious yearly seasonal pattern in realized volatility can be seen in the figure.

In Table 1, descriptive statistics (mean, standard deviation, coefficient of variation, minimum, maximum, skewness and kurtosis) are given for the realized volatility time series of S&P 500, FTSE 100, KOSPI Composite Index 100 and Bovespa Index. We can see that the average values are quite similar for the four indices, the coefficients of variation ranging from 0.49 to 0.69. The unconditional distributions of the realized volatility time series are highly skewed and leptokurtic.

Analogous descriptive statistics for the *natural logarithm* of the four realized volatility time series are presented in Table 2. We can see that the logarithm of the realized volatility time series is more Gaussian. This is also in agreement with Figure 3. Consequently, in further analysis, the natural logarithm of the realized volatility time series will be used. This is no drawback since the natural logarithm has a straightforward interpretation, its change being directly related to percentage change in the original volatility time series.

Table 1 Descriptive statistics for four realized volatility time series

|	Mean	Standard deviation	Coefficient of variation	Min.	Max.	Skewness	Kurtosis
S&P 500	0.009	0.006	0.689	0.001	0.088	3.135	17.838
FTSE 100	0.008	0.005	0.623	0.002	0.068	2.910	16.168
KOSPI CI	0.010	0.006	0.616	0.002	0.077	2.461	11.753
Bovespa Index	0.013	0.006	0.486	0.003	0.082	3.492	21.714

Source: Own construction

Table 2 Descriptive statistics for the natural logarithm of four realized volatility time series

	Mean	Standard deviation	Coefficient of variation	Min.	Max.	Skewness	Kurtosis
S&P 500	−4.886	0.553	0.113	−6.667	−2.430	0.410	0.313
FTSE 100	−4.985	0.506	0.101	−6.242	−2.688	0.514	0.194
KOSPI CI	−4.778	0.526	0.110	−6.040	−2.563	0.362	-0.194
Bovespa Index	−4.398	0.385	0.087	−5.756	−2.499	0.619	1.697

Source: Own construction

Figure 3 Time series of realized volatility of S&P 500 based on 5-minute log returns (top plot), the natural logarithm of the time series (middle plot) and the annualized version of the time series given in percentages (bottom plot). The ticks on the x-axis correspond to the start of calendar years
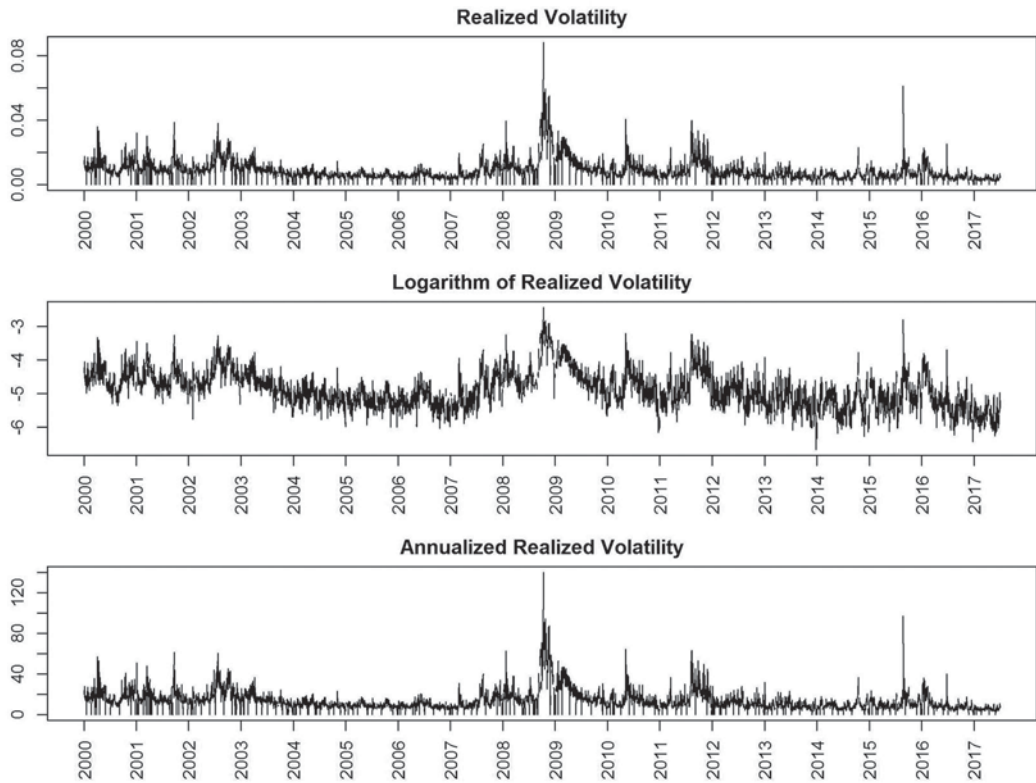
2.1 Seasonality models

We assume the model of Formula (1) and work with the natural logarithm of realized volatility based on 5-minute log returns. For each of the 19 time series, each time series being associated with one stock market index, six different approaches to yearly seasonal cycle modeling will be assumed:

1.) Intercept only, which corresponds to no yearly seasonality. This approach to yearly seasonality modeling will be denoted as N.
2.) Intercept + 11 dummy variables, which amounts to 12 parameters. Denoted as D.
3.) Fourier basis functions with $K = 2, 3, 5, 8$ and 11 (see Formula (4)), which corresponds to 5, 7, 11, 17 and 23 parameters. Denoted as F5, F7, F11, F17 and F23.
4.) Periodic linear splines with the following number of knots (parameters): $K = 5, 7, 11, 17$ and 23. The knots are placed equidistantly, see below. Denoted as L5, L7, L11, L17 and L23.
5.) Periodic cubic splines with the following number of knots (parameters): $K = 5, 7, 11, 17$ and 23. The knots are placed equidistantly, see below. Denoted as C5, C7, C11, C17 and C23.
6.) Periodic cubic splines with the following number of knots (parameters): $K = 5, 7, 11, 17$ and 23. *Expert* placement of knots is used, see below. Denoted as E5, E7, E11, E17 and E23.

In approaches L and C, the knots will be placed equidistantly throughout the year so that the distance between two neighboring knots is constant, the first knot being placed at the beginning of the calendar year.

Concerning approach E, the first knot is placed at the beginning of the calendar year and the second one at the start of December. Every further *odd* knot is placed within the interval from the start of the calendar year to the start of December (11-month interval) and every *even* knot is placed within the interval from the start of December to the start of the calendar year (1-month interval). In both these intervals the knots are placed equidistantly. Approach E thus places more knots to December than to any other month of the year. This is an example of an expert placement of the knots which can be supported by the fact that the end of the calendar year is a very special period overlapping with Christmas in many cultures and countries, and with an upcoming new year.[13]

Concerning the weekly seasonal cycle, two approaches will be assumed for its modeling:

1.) No weekly seasonal cycle. Denoted as N.

2.) 4 dummy variables. Denoted as D.

The six approaches to modeling the yearly seasonal cycle will be explored in combination with the two possible approaches to modeling the weekly seasonal cycle. Table 3 provides a summary of the approaches explored. The approaches used for yearly seasonal cycle modeling are given in columns, whereas the approaches used for weekly seasonal cycle modeling are given in rows. The number in each cell is the total number of models explored for the given combination of approaches, the models in the cell differing in the number of parameters used for the yearly seasonal cycle model (5, 7, 11, 17 or 23).

A given combination of a model for yearly seasonality and a model for weekly seasonality will by denoted, for example, as F11 + D, which means that Fourier basis functions with eleven parameters (including intercept) were used to model the yearly seasonal cycle and dummy variables were used to model the weekly cycle.

A total of 44 different models for the seasonal parts will be explored for each of the 19 time series of the natural logarithm of realized volatility. For each of the 44 models, the following ARMA model for $\{E_t\}$ (see Formulas (1) and (2)) will be considered:

$$E_t = \phi_1 E_{t-1} + \phi_2 E_{t-2} + \phi_6 E_{t-6} + \theta_1 \epsilon_{t-1} + \epsilon_t, \quad t = 1,\ldots,N , \tag{18}$$

since it provides a good approximation to the HAR model (Corsi, 2009) commonly used for realized volatility modeling.

R software (R Core Team, 2017) and the following contributed R packages have been used in the analysis: *forecast* (Hyndman, 2017; Hyndman and Khandakar, 2008) and *pbs* (Wang, 2013).

Employing the *Arima()* function from the *forecast* package (Hyndman, 2017) Gaussian maximum likelihood estimation will be used to estimate the parameters of Formula (1). Further, the best combination of the yearly and weekly seasonal cycle (i.e. the best model) will be selected according to the value of AIC or BIC. The residuals of the selected model are checked whether they are uncorrelated, homoskedastic, normal and whether no outliers are present (see Section 2.3).

Since it is impossible to simply interpret the individual parameters of the Fourier or spline basis expansion, we will neither interpret the individual parameter estimates nor will we assess the accuracy of the estimates in our analysis. On the other hand, we will interpret the estimated seasonal cycle (i.e. a linear combination of parameters) and will assess the accuracy of the estimate of the seasonal cycle, se further.

---

[13] Ramsay and Silverman (2005) state that a researcher may want to have more knots over regions where the function to be estimated exhibits the most complex variations. We presume that the knowledge of such regions is often a question of an expert understanding of the problem at hand. Consequently, different placements of the knots are expected for different applications and are subjective to some extent. We believe that December is a reasonable choice for the placement of more knots in our application.

Table 3 Number of models used for the various combinations of approaches to yearly and weekly seasonal cycle modeling. The models in each cell differ by the number of parameters used for yearly seasonal cycle modeling

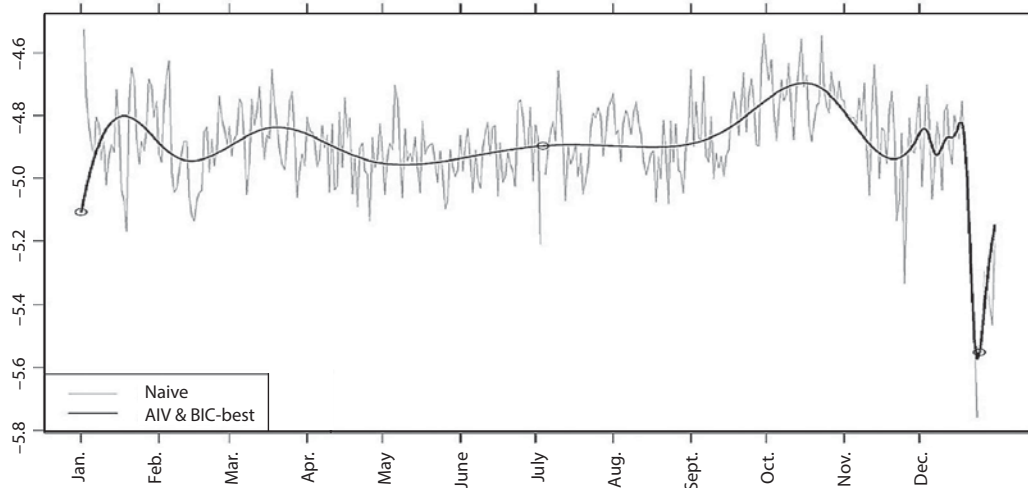|		Yearly seasonal cycle					
		N	D	F	L	C	E
Weekly seasonal cycle	N	1	1	5	5	5	5
	D	1	1	5	5	5	5

Source: Own construction

2.2 Results for S&P 500

At first, results for the natural logarithm of realized volatility of S&P 500 will be presented. The best model among the 44 models of Table 3 according to both AIC and BIC is model E17 + D, i.e. model E17 for yearly seasonality combined with model D for weekly seasonality.

Figure 4 presents a naïve estimate of the yearly seasonal cycle[14] and the estimated yearly seasonal cycle for the best model (E17 + D). The estimated cycle (for model E17 + D) has a minimum on December 24 and a maximum on October 16. The shift from the minimum to the maximum value corresponds to a 140 percent increase of volatility. The estimated seasonal cycle exhibits a deep trough from December 18 till the start of January. Another, higher and flatter, bottom in volatility occurs during May. Typical variations of the estimated seasonal cycle throughout the year correspond to an approximately 10 percent change in volatility.

Figure 4 Naïve estimate of the yearly seasonal cycle (gray) in the logarithm of realized volatility of S&P 500 together with the estimated cycle from the best model (E17 + D) according to AIC and BIC (black). The three points on the estimated cycle correspond to calendar days (Jan 1, Jul 4 and Dec 25) where in fact no trading ever happens due to public holidays
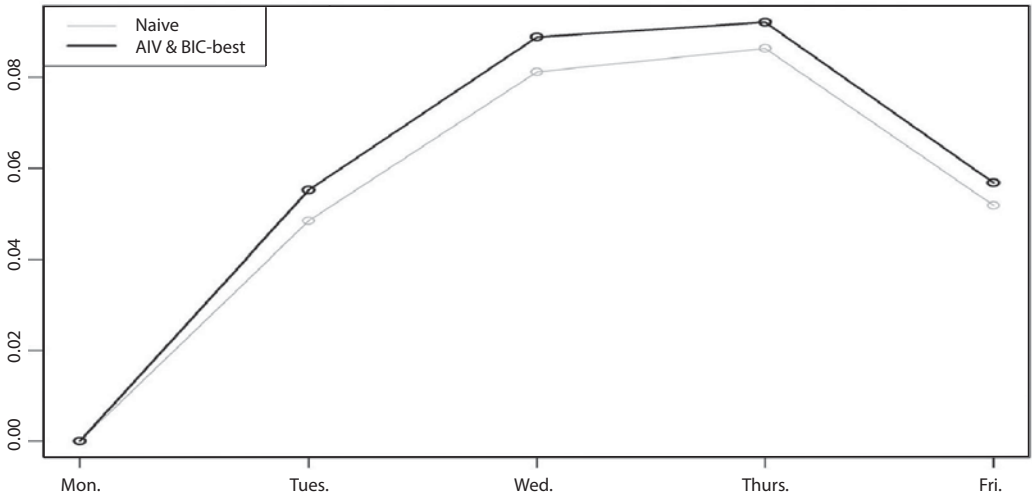

Source: Own construction

---

[14] A naïve estimate of the yearly seasonal cycle (including the intercept/level) is obtained by calculating the average of the time series separately for each calendar day of the year. In this way we get 365 average values, one average value being associated with one day of the calendar year.

Figure 5 presents a naïve estimate of the weekly seasonal cycle[15] and the estimated weekly cycle from the best model (E17 + D). The highest level occurs on Thursday and the lowest on Monday, the shift from the minimum to the maximum value corresponding to an approximately 10 percent change in volatility.

Figure 5 Naïve estimate of the weekly seasonal cycle (gray) in the logarithm of realized volatility of S&P 500 together with the estimated cycle from the best model (E17 + D) according to AIC and BIC (black)

Source: Own construction

Table 4 presents the "relevance" (weight) for the *best* model according to AIC (BIC) in the given *group of models* associated with the cell of the table calculated according to Formula (16) (Formula (17)) with respect to the *global best* model according to AIC (BIC). In other words, the *best* model in *the group of models* is compared to the *global best* model (E17 + D) using Formula (16) (Formula (17)). The relevance for the cell which contains the global best model (E17 + D) is 1, the values of relevance in other cells being rather small. This suggests that the combination of approach E used to model the yearly seasonal cycle and approach D to model the weekly seasonal cycle performs considerably better than any other explored combination of approaches.

Table 4 "Relevance" for the best model according to AIC or BIC in the given group of models associated with the cell of the table calculated according to Formula (16) (AIC, the first number) and Formula (17) (BIC, the second number) with respect to the global best model according to AIC or BIC

		Yearly seasonal cycle					
		N	D	F	L	C	E
Weekly seasonal cycle	N	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0
	D	0, 0.03	0, 0	0, 0	0, 0	0, 0	1, 1

Source: Own construction

---

[15] A naïve estimate of the weekly seasonal cycle is obtained by calculating the average of the time series of the natural logarithm of realized volatility separately for different days of the week (Monday, Tuesday, Wednesday, Thursday and Friday). Further, the average for Monday is subtracted from the five averages (for Monday, Tuesday, Wednesday, Thursday and Friday).

To compare approaches to modeling the yearly seasonal cycle based on smooth functions, i.e. approaches F, C and E, in Figure 6, we present the estimated yearly seasonal cycle for the best F approach according to AIC, the best C approach according to AIC and the best E approach according to AIC. We can clearly observe that the advantage of the E approach is the expert placement of the knots which allowed for more flexibility of the seasonal cycle in December. On the other hand, in order to capture the trough in December with the use of Fourier basis functions, it would have been necessary to use a much larger value of K.

Figure 6 Comparison of the estimates of the yearly seasonal cycle in the natural logarithm of realized volatility of S&P 500 for the AIC-best approach among the F approaches (black), the AIC-best approach among the C approaches (dashed) and the AIC-best approach among the E approaches (dotted). The naïve estimate is presented in gray
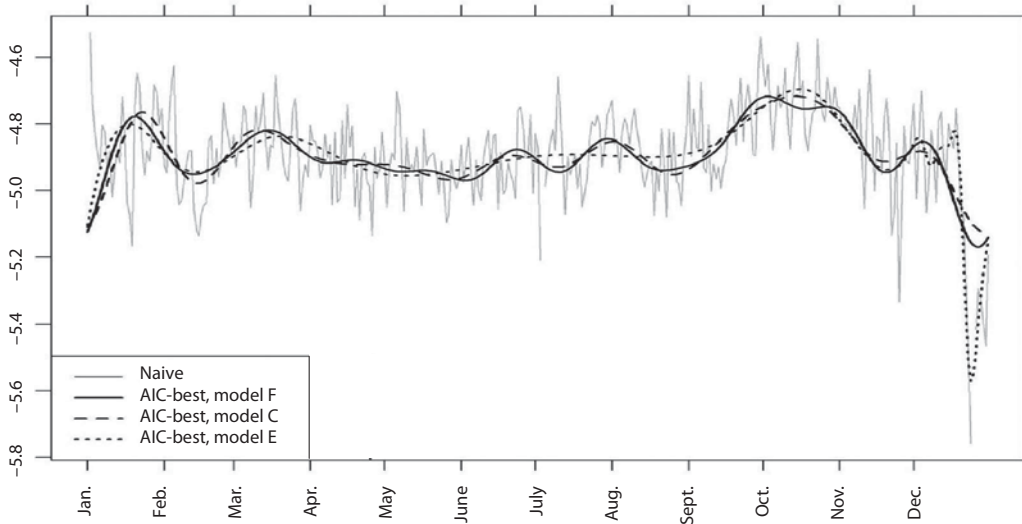
2.3 Results for all the indices

Table 5 presents the results for the 19 indices and the corresponding time series of the natural logarithm of realized volatility based on 5-minute log returns. Specifically, the number of time series is given for which the corresponding combination of the yearly and weekly seasonal cycle model resulted in the *lowest* value of AIC or BIC.

Table 6 is constructed as follows. 19 tables analogous to Table 4 are constructed containing the values of "relevance" according to AIC and BIC. Each of the 19 tables corresponds to one specific time series. Further, for each cell of the table, we calculate (separately for AIC and BIC) the average value of the cell across the 19 tables, i.e. we obtain the average value of "relevance" based on AIC and the average value of "relevance" based on BIC. These average values are reported in Table 6.

The results from Tables 5 suggest that approach E is presumably the most appropriate approach for modeling the yearly seasonal cycle. It wins 17 times (out of 19) according to AIC and 6 times (out of 19) according to BIC. Even in terms of the "average relevance" (see Table 6) it performs very well. Model N for yearly seasonality also performs rather well as assessed by BIC. This suggests that if we use a very heavy penalty on the number of parameters, as is the case in BIC, no yearly seasonality is suggested as an appropriate model in some instances. The other approaches to modeling the yearly seasonality (D, F, L, C) do not perform very well, the reason presumably being that they are not capable of capturing the deep trough in the yearly seasonal cycle which occurs in the second half of December.

Table 5 Number of time series of the natural logarithm of realized volatility for which the corresponding combination of the yearly seasonal cycle model and the weekly seasonal cycle model resulted in the lowest value of AIC (the first number in the cell) or BIC (the second number)

		Yearly seasonal cycle					
		N	D	F	L	C	E
Weekly seasonal cycle	N	0, 3	0, 0	0, 0	0, 0	0, 0	0, 0
	D	1, 9	1, 1	0, 0	0, 0	0, 0	17, 6

Source: Own construction

Table 6 Average value of "relevance" according to AIC (first number) and BIC (second number); see text for details

		Yearly seasonal cycle					
		N	D	F	L	C	E
Weekly seasonal cycle	N	0, 0.16	0, 0	0, 0	0, 0	0, 0	0.02, 0
	D	0.05, 0.49	0.05, 0.06	0, 0	0, 0	0, 0	0.89, 0.32

Source: Own construction

For each of the 19 time series of the natural logarithm of realized volatility the following is reported in Table 7:

- **Best**: The best model according to *AIC*.
- **Day of year, min. and max.**: The estimated day of the year (in the range 1–365) where the minimum and maximum value of the *yearly* seasonal cycle occurs according to the best model selected with the use of AIC.
- **Day of week, min. and max.**: The estimated day of the week (1 = Mon, 2 = Tue, 3 = Wed, 4 = Thu, 5 = Fri) where the minimum and maximum value of the *weekly* seasonal cycle occurs according to the best model selected with the use of AIC.
- **sd:** The standard deviation of the estimated yearly cycle from the best model according to AIC.
- **R**: The "relevance" calculated according to Equation 16 of the AIC-best model where *no yearly* seasonality is present with respect to the global AIC-best model. Low values of "relevance" suggest that yearly seasonality is present in the time series.[16]
- **aver. std. err.**: Average of 365 standard errors of the estimated yearly seasonal cycle for different days of the year.
- **resid. diagn.:** The assumption of no autocorrelation, homoskedasticity and normality of error was assessed in the AIC-best model using the residuals of the model. Formal hypothesis tests were *not* used since they would have large power due to the length of the time series and would lead to the rejection of the null hypotheses even for minor and practically irrelevant deviations from the null hypothesis. Instead, qualitative assessment with focus given on practically important deviations from the null hypothesis was used, employing autocorrelation function of residuals, autocorrelation function of squared residuals, QQ plots and time series of residuals. An "OK" in the column stands for satisfied assumptions. "SNN" stands for slight non-normality, "NN" for non-normality, "SH" for slight heteroskedasticity and "OUT" for outliers.

---

[16] The X symbol in the R column implies that the estimation procedure did not converge while fitting the model with no yearly seasonality.

Table 7 See text for details

|	Best	Day of year, min. and max.	Day of week, min. and max.	sd	R	aver. std. error	resid. diagn.
S&P 500	E17 + D	358, 289	1, 4	0.11	0	0.12	OK
DJIA	E17 + D	358, 289	1, 4	0.10	0	0.11	OK
NASDAQ 100	D + D	-	1, 3	0.09	X	0.13	SNN
FTSE 100	E17 + D	358, 289	1, 5	0.09	0	0.12	SNN
DAX	E17 + D	360, 289	1, 4	0.09	0	0.12	SNN
Euro STOXX 50	E17 + D	359, 290	1, 5	0.11	0	0.10	NN, OUT
CAC 40	E17 + D	358, 288	1, 5	0.10	0	0.11	SH, SNN
AEX	E23 + D	359, 283	1, 5	0.11	0	0.11	SNN
Swiss Index	E11 + D	360, 27	1, 4	0.06	0	0.09	SNN
IBEX 35	E17 + D	358, 337	1, 4	0.08	X	0.12	SNN
FTSE MIB	E17 + D	359, 337	1, 5	0.08	0	0.11	SH, SNN
Nikkei 225	E17 + D	359, 18	1, 5	0.07	0	0.09	SH, NN
Hang Seng	E17 + D	358, 302	5, 4	0.07	0	0.10	SNN
KOSPI	N + D	-	1, 4	-	1	0.16	OUT
FT Straits Time	E23 + D	360, 286	1, 3	0.06	0	0.09	SNN
IPC Mexico	E17 + D	359, 19	1, 3	0.08	0	0.10	SNN
Bovespa	E23 + D	360, 293	1, 4	0.10	0	0.07	SNN
All Ordinaries	E11 + D	361, 277	2, 3	0.05	0.03	0.10	SNN
S&P/TSX	E17 + D	358, 292	1, 3	0.09	0	0.13	OK

Source: Own construction

We can clearly see that according to AIC the model E17 + D performs mostly the best. The minimum of the yearly seasonal cycle occurs at the end of the calendar year in all the time series, the maximum occurring mostly in October or at the beginning of December or during January. Figure 7 presents the estimated seasonal cycles (disregarding the level of the cycle) in the best model according to AIC for 7 different indices.

Concerning the weekly seasonal cycle, the lowest volatility occurs mostly on Mondays, while the highest occurs during the second part of the week (Wednesday, Thursday or Friday).

The assumptions of no autocorrelation, heteroskedasticity and normality of errors and the presence of no outliers were reasonably satisfied[17] in most of the AIC-best models, except for Euro Stoxx 50 (where more severe non-normality and outliers are present), Nikkei (where more severe non-normality is present) and KOSPI (which exhibits outliers).

---

[17] In the sense that no deviations or only slight deviations from the assumptions were detected.

Figure 7 Comparison of the estimated yearly seasonal cycles for the best models according to AIC for 7 indices (S&P 500, FTSE 100, DAX, IBEX, NIKKEI, BOVESPA, ALL ORDINARIES). The levels of the cycles should not be compared since the estimated cycles are intentionally plotted at different levels to allow for a better comparison of their shapes
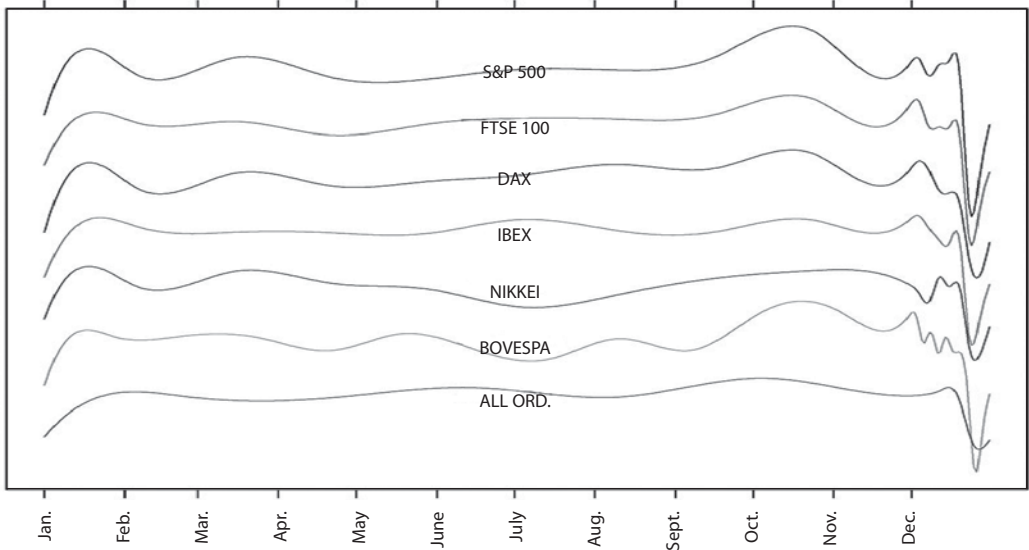
Source: Own construction

CONCLUSION

We have demonstrated the usefulness of various basis expansion methods in representing the yearly seasonal cycle in the natural logarithm of realized volatility based on 5-minute log returns of 19 stock market indices. Cubic splines with an expert placement of knots seem to be an appealing approach which suggests that yearly seasonality is present in most of the time series. These findings are important for future research since they play a crucial role in decisions regarding asset allocation and risk management, and should also be taken into the account while pricing options.

The presented approaches can potentially be used for modeling of other time series containing long seasonal periods such as daily time series of number of products sold, car accidents, electricity consumption and traffic volume.

ACKNOWLEDGEMENT

References

ARISMENDI, J. C., BACK, J., PROKOPCZUK, M., PASCHKE, R., RUDOLF, M. Seasonal stochastic volatility: implications for the pricing of commodity options. *Journal of Banking & Finance*, 2016, 66, pp. 53–65.

BACK, J., PROKOPCZUK, M., RUDOLF, M. Seasonality and the valuation of commodity options. *Journal of Banking & Finance*, 2013, 37(2), pp. 273–290.

CLAESKENS, G. AND HJORT, N. L. *Model Selection and Model Averaging*. Cambridge University Press, 2008.

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

FRIEDMAN, J., HASTIE, T., TIBSHIRANI, R. *The Elements of Statistical Learning*. New York: Springer series in statistics, 2001.

GIOVANIS, E. Calendar Effects and Seasonality on Returns and Volatility [online]. *Munich Personal RePEc Archive*, 2009. [cit. 3.1.2018]. <https://mpra.ub.uni-muenchen.de/64404/>.

HEBER, G., LUNDE, A., SHEPHARD, N., SHEPPARD, K. *Oxford-Man Institute's realized library.* Oxford-Man Institute, University of Oxford, 2009, library version number: 0.2.

HYNDMAN R. J. Forecast: Forecasting functions for time series and linear models. *R package*, 2017, version 8.2.

HYNDMAN, R. J. AND KHANDAKAR, Y. Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, 2008, 26(3), pp. 1–22.

KIYMAZ, H. AND BERUMENT, H. The day of the week effect on stock market volatility and volume: International evidence. *Review of financial Economics*, 2003, 12(4), pp. 363–380.

LIU, L., PATTON, A., J., SHEPPARD, K. Does Anything Beat 5-Minute RV? A Comparison of Realized Measures Across Multiple Asset Classes. *Journal of Econometrics*, 2012, 187(1), pp. 293–311.

R CORE TEAM. *R: A Language and Environment for Statistical Computing*, 2017.

RAMSAY, J. O. AND SILVERMAN, B. W. *Applied functional data analysis: methods and case studies*. New York: Springer, 2002.

RAMSAY, J. O. AND SILVERMAN, B. W. *Functional Data Analysis.* 2nd Ed. Springer, 2005.

SEYYED, F. J., ABRAHAM, A., AL-HAJJI, M. Seasonality in stock returns and volatility: The Ramadan effect. *Research in International Business and Finance*, 2005, 19(3), pp. 374–383.