

## Multiple STL decomposition in discovering a multi-seasonality of intraday trading volume

Josip Arnerić\*

*Faculty of Economics and Business, University of Zagreb  
Trg J. F. Kennedyja 6, 10000 Zagreb, Croatia  
E-mail: {jarneric@efzg.hr}*

**Abstract.** The seasonal and trend decomposition of a univariate time-series based on Loess (STL) has several advantages over traditional methods. It deals with any periodicity length, enables seasonality change over time, allows missing values, and is robust to outliers. However, it does not handle trading day variation by default. This study offers how to deal with this drawback. By applying multiple STL decompositions of 15-minute trading volume observations, three seasonal patterns were discovered: hourly, daily, and monthly. The research objective was not only to discover if multi-seasonality exists in trading volume by employing high-frequency data but also to determine which seasonal component is most time-varying, and which seasonal components are the strongest or weakest when comparing the variation in the magnitude between them. The results indicate that hourly seasonality is the strongest, while daily seasonality changes the most. A better understanding of trading volume multiple patterns can be very helpful in improving the performance of trading algorithms.

**Keywords:** hourly seasonality, intraday volume, Loess, multiple seasonal patterns, STL decomposition

Received: March 25, 2021; accepted: May 03, 2021; available online: June 29, 2021

DOI: 10.17535/crorr.2021.0006

---

### 1. Introduction

Advances in electronic trading, along with information technology developments, have enabled the recording of ultra-high-frequency data, observed tick-by-tick. The accessibility of such data has created a new niche for financial time-series analysis. In other words, huge data sets are used, and trading dynamics within a single day can be established. Without high-frequency data, practitioners and academics are limited to employing daily data alone or observations at even lower frequencies. Nevertheless, it is challenging to deal with high-frequency data because they have unique characteristics that lower frequency data do not possess [4, 10]. In most of the existing studies, high-frequency data have been widely used in the analysis of realized volatility and related issues considering microstructure noise and price jumps [8, 29]. Stylized regularities of realized volatility and selection of an appropriate sampling frequency to obtain an unbiased estimator of integrated variance are comprehensively documented by [2].

The concept of realized volatility relies only on intraday price dynamics, independent of trading volume dynamics. However, very few studies deal with intraday trading volume analyses [6, 9, 23]. It is well known that high-frequency data exhibit periodic patterns in trading activity, that is, intraday trading volume seasonality [13, 20]. This means that trading is denser at the opening and closing hours of the trading day than at midday. In addition to hourly seasonality, daily or weekly seasonality can be observed within a single time-series, depending

---

\*Corresponding author.

on the seasonal frequency selection. Although multiple studies regarding seasonality in price dynamics and volatility have been conducted [1, 7, 21, 22], there is a lack of studies covering multiple seasonality of trading activity considering intraday volume observations.

Further, great efforts have been made to improve the performance of trading algorithms by incorporating accurate predictions of intraday trading volumes [9, 23]. However, this study does not focus on prediction directly, but on the specification and extraction of the trading volume of multiple seasonal components. A better understanding and detailed insight of multi-seasonality patterns can be very helpful in capturing the dynamics of trading volume and consequently improving the performance of trading algorithms. In recent years, trading algorithms have been developed and used to reduce transaction costs as much as possible [6, 15].

This topic is relevant in the literature, particularly as users employ seasonal adjustments by default and do not adapt existing methods to capture intraday time-series features. Most of the seasonal adjustment methods are not flexible enough to deal with certain issues regarding high-frequency data, such as missing values, non-trading days, calendar variations, seasonality with different periods per cycle, and outliers. By reviewing the existing literature, it can be observed that there are two approaches to the decomposition of multiple seasonal patterns. One approach considers dividing the component with the longest cycle into shorter components and then fitting a Holt-Winters exponential smoothing or Taylor's double seasonal exponential smoothing [5, 17, 24, 25, 26, 30]. Therefore, it is inevitable to point out the study of [12], who introduced the state space concept into exponential smoothing for forecasting multiple seasonal time-series. The authors concluded that the newly developed multiple seasonal model provides more accurate forecasts than the Holt-Winters exponential smoothing and double seasonal exponential smoothing. However, the multiple seasonal model could only use two sub-cycles within hourly observations.

Another approach considers estimating the shortest component first and proceeds to estimate the next components from the residuals, that is, the irregular component. The latter uses the Loess method, which is repeated successively until there is no seasonality left in the residuals [16, 18, 27]. Adjustment of daily time-series within multiple seasonality, based on Loess, has already been studied [11], but there is no consensus about the choice of seasonal frequencies as well as dealing with non-trading days. There are many daily time-series with observations not available during weekends, and the difference in trading and non-trading number of days for such series should not be neglected. This can be partially solved by changing the seasonal frequencies at appropriate levels, as suggested in this study. Despite the flexibility of multiple seasonal adjustments at different frequencies, a user does not have to utilize all possible frequencies, just the ones that are present in the data.

Considering the aforementioned advantages and disadvantages of the commonly used methodology in capturing multiple periodicities of high-frequency time-series, this study contributes to the existing literature in several ways. Foremost, the issue of an unequal number of trading days within a week, month, or year was solved by replacing non-existing weekday data with missing values. Two reasons support this solution: (i) having the same number of observations per sub-cycle, and (ii) employing an STL decomposition that allows missing values. Second, three seasonal components were found by utilizing appropriate frequencies 4, 34, and 739.19 corresponding to hourly, daily, and monthly seasonality, respectively. Third, when comparing the variation in magnitude between extracted components, the hourly seasonality was found to be the strongest with a clear "U"-shaped diurnal pattern, indicating that trading activity at the opening hours was more intense as traders began to accommodate the information accumulated at the closing of the previous day. At the closing hours of the current day, trading is also intensified, which can be explained in terms of some investors attempting to close their open positions. In previous studies, the shape of the diurnal pattern was assumed to be constant for every trading day, while in this study, it was allowed to change over time. Lastly, this study adds a new value to studies dedicated to the investigation of stock market activity

and intensity, outlying price movements. Discovering trading seasonalities at lower frequencies that may occur within volumes observed at the higher frequency contributes to more accurate predictions of how much volume will be traded in a given time, and thus the trading strategy settings can be improved.

Apart from the introduction, this paper has three additional sections. Section 2 presents and discusses multiple STL decompositions. The empirical application, considering the DAX index intraday trading volume, is obtained in Section 3, while Section 4 concludes and provides further research directions.

## 2. Multi-seasonality decomposition by locally weighted polynomial

The abbreviation Loess originates from "locally estimated smoothing". The Lowess term also applies in practical applications denoting "locally weighted scatterplot smoothing". Although both terms represent the same method, Lowess is usually employed when plotting smoothed time-series, while Loess is more common when fitting locally weighted regression for computational reasons. Loess was embedded in time-series decomposition for the first time by [11], resulting in the "Seasonal-Trend decomposition based on Loess" (STL), established for univariate time-series decomposition with a single seasonality. In recent years, data have become accessible and more frequently recorded within a short time interval; thus, multiple seasonal patterns may exist within a single time-series.

STL decomposition is an iterative numerical procedure consisting of two recursive loops, that is, the inner and outer loops. Updating trends and seasonal components are obtained in the inner loop, while robustness weights are determined in the outer loop utilizing the irregular component of a time series from the inner loop. Therefore, the inner loop was nested within the outer loop. Two iterations of the inner loop are most appropriate for achieving convergence, while one iteration of the outer loop is sufficient to obtain satisfactory robust weights and thus reduce the influence of outliers [14].

---

### Seasonal-trend decomposition starting with the shortest periodicity

---

**Start** with the shortest periodicity of length  $k_{short}$

**Inner loop:**

- Step 1. Subtracting the trend from the original series (initial trend estimate is zero)
- Step 2. Disaggregating detrended series into  $k$  sub-series and smoothing each of them by Loess employing  $q$  neighborhood observations
- Step 3. Extending a smoothed detrended sub-series for  $2k$  out-of-sample observations by extrapolation
- Step 4. Filtering a preliminary seasonal component by three MA followed by Loess with  $l$  neighborhood observations
- Step 5. Differences between preliminary seasonal components and filtered seasonal components produce final seasonal components
- Step 6. Subtracting the final seasonal component from the original series
- Step 7. Smoothing seasonally adjusted series by Loess with  $t$  neighborhood observations produces trend component of the series

**Outer loop:**

- Step 8. Subtracting seasonal and trend components to obtain the irregular component
- Step 9. Determine the weights according to outliers from the irregular component
- Step 10. Using the weights obtained in Step 1

**Proceed** with the next periodicity of length  $k_j > k_{short}$  by repeating Steps 1-10

**End** with the longest periodicity  $k_{long}$  by repeating Steps 1-10

---

Table 1: *Multiple STL algorithm.*

The multiple STL algorithm is presented in Table 1 in the context that a single STL starts with the shortest periodicity, repeats successively for all possible periodicities between  $k_{short}$  and  $k_{long}$ , and ends with the longest periodicity.

The following is a detailed description of each step. Steps 1–10 of a single STL decomposition are employed for every length of periodicity. If only one seasonal frequency  $k$  is determined, a multiple STL is reduced to a single STL decomposition in the additive form:

$$y_t = T_t + S_t^k + I_t \quad t = 1, 2, \dots, T. \quad (1)$$

Determination of the seasonal frequency  $k$  is straightforward, for example, if monthly data exhibit yearly periodicity, the seasonal frequency  $k=12$ ; if daily data exhibit weekly periodicity the seasonal frequency  $k=7$  or  $k=5$ , and so on. After subtracting the trend  $T_t$  from the raw time-series  $y_t$  in Step 1, the algorithm continues with the detrended series:

$$y_t^{det} = y_t - T_t = S_t^k + I_t \quad t = 1, 2, \dots, T. \quad (2)$$

In Step 2 the detrended series (2) is disaggregated into an unstacked  $k$  sub-series of the same length  $n = T/k$ . Subsequently, each sub-series  $y_{i,t_j}^{det}$  is smoothed by fitting a locally weighted polynomial of degree  $d=2$  using window bandwidth  $q$  (number of neighborhood observations) and the tricube weight function:

$$w_i(y_{t_j}^{det}) = \left(1 - \left(\frac{|y_{i,t_j}^{det} - y_{t_j}^{det}|}{d_q}\right)^3\right)^3 \quad i = 1, 2, \dots, n \quad j = 1, 2, \dots, k, \quad (3)$$

where  $d_q$  is the distance of the  $q$ th farthest observation from the  $y_{t_j}^{det}$ . Setting a degree of polynomial  $d=2$  is a reasonable choice when data show substantial peaks, such as in the case of intraday trading volume data. The window bandwidth  $q$  usually referred to as a Loess window, is the key parameter that should be carefully selected. According to [11], it should be the smallest odd integer, equal to or greater than 7, for which all detrended sub-series are well smoothed but not too smoothed.

Further, in Step 3, each smoothed detrended sub-series is extended for two observations, one before the first observation and one after the last observation, to avoid losing values in the next step. Extended smoothed detrended sub-series are stacked together, representing a preliminary seasonal component. The same component is filtered in Step 4 using a triple moving averages  $MA(k) \times MA(k) \times MA(3)$ , followed by the Loess with  $l$  neighborhood observations, which is set by default as the smallest odd integer greater than  $k$ . Step 4 ensures that there is no trending behavior left, and thus filtered values are subtracted from the preliminary seasonal values in Step 5 to produce the final seasonal component. In Step 6, the raw series are seasonally adjusted, and in the last step of the inner loop, the final trend component is estimated by smoothing the seasonally adjusted series using Loess with  $t$  neighborhood observations (Step 7). Parameter  $t$  is set by default as the smallest odd integer greater than the following ratio:

$$\frac{1.5k}{1 - \frac{1.5}{q}} \quad (4)$$

Estimates of seasonal and trend components are used to obtain the irregular component in Step 8. If robust estimates are required, observations are additionally weighted with respect to the outlying values using the bisquare function:

$$w_t = \left(1 - \left(\frac{|I_t|}{6Med|I_t|}\right)^2\right)^2 \quad t = 1, 2, \dots, T, \quad (5)$$

where  $Med$  denotes the median. Weights obtained in the outer loop are used to down weight the effect of outliers when the algorithm starts all over again from the inner loop.

The specificity of STL decomposition is the implementation of Loess three times, that is, within Steps 2, 4, and 7. The smoothing parameters with respect to Loess windows  $l$  and  $t$  in Steps 4 and 7 are determined directly from  $k$  and  $q$ . Parameter  $k$  should be set as a single value or as multiple values depending on the pattern(s) present in the data. It is more difficult to determine parameter  $q$  when the detrended sub-series are smoothed in Step 2. The entire decomposition is most sensitive to this parameter, and  $q$  should be chosen carefully. Fortunately, seasonal diagnostic plots can be very helpful in making decisions about  $q$ , whether single or multiple STL decompositions are obtained.

The Loess method is more flexible compared to other smoothing methods as it enables the smoothing time-series with any seasonal period greater than one and allows missing values as well [11]. Univariate STL decomposition has several advantages over traditional methods such as SEATS and X11, that is, it can deal with any type of seasonality, and seasonality is enabled to change over time, missing values are allowed, and it can be robust to outliers [14, 19]. However, it does not automatically handle trading day variations. The purpose of this study is to discover not only which seasonal components exist in trading volume but also which component is most time-varying, and in which seasonal components are the strongest or weakest when comparing the variation in magnitude between them.

It should also be noted that the purpose of intraday trading volume decomposition presented in this paper is beyond seasonally adjusting the time-series. It is more about understanding all the components of variability in the series, which can help improve trading algorithms. The decomposition model proposed in this study is as follows:

$$Y_t = T_t + S_t^{hd} + S_t^{dw} + S_t^{my} + I_t \quad t = 1, 2, \dots, T. \quad (6)$$

Equation (6) defines the intraday time-series decomposition using multi-seasonal components. Three seasonal components are considered with respect to the hour of the day (denoted by superscript  $hd$ ), day of the week (denoted by superscript  $dw$ ), and month of the year (denoted by superscript  $my$ ). These components are most likely to appear in a high-frequency data, which will be presented in the next section considering the DAX index. The DAX index is selected for application purposes as it is heavily traded and exhibits significant intraday market activity. Trading volumes for other stock markets were intermittently available or not available at all, particularly in such short sampling intervals. As per equation (6), three seasonal frequencies should be appropriately determined, which is comprehensively elaborated in the next section. This also means that STL decomposition repeats three times starting with the shortest periodicity and ending with the longest periodicity, that is, hourly seasonality is first extracted, daily seasonality afterward, and monthly seasonality is the last extracted. Following [28], for each extracted seasonal component, a measure of strength is computed:

$$1 - \frac{Var(I_t)}{Var(S_t^k + I_t)}. \quad (7)$$

For a strong seasonal component, detrended data concerning that component should have much variability than the irregular component, that is,  $Var(I_t)/Var(S_t^k + I_t)$  should be relatively small; and thus the value of (7) should be close to one. The strength of the trend component is defined similarly but with respect to the seasonally adjusted data,

$$1 - \frac{Var(I_t)}{Var(T_t + I_t)}. \quad (8)$$

### 3. Empirical application

This study considers a regular time-series of 15 minutes observations concerning the DAX stock index over five years (2014–2018). Raw data were provided by Thomson Reuters Tick History. Prior to the multiple STL decomposition, raw data were examined to check for empty intraday intervals and observations outside the official trading hours. No empty intervals were found because a 15 min trading volume was obtained by summing all recorded volumes within that period. This was expected since the Frankfurt Stock Exchange is highly active, and liquid markets and stocks are traded very frequently. Observations outside official trading hours (09:00-17:30) were omitted. Thus, 34 trading volume observations are left per trading day. More concerning is the unequal number of trading days within a week, month, and year. This issue is solved by replacing the non-existing weekday data with missing values. This solution is supported by two conditions: (i) keeping the number of trading days equal within each cycle and consequently having the same number of observations per sub-cycle, and (ii) employing a seasonal-trend decomposition procedure that allows for missing values.

Having the same number of observations per sub-cycles, allowing for missing values and handling a multiple set of any seasonal frequencies greater than one, makes perfect conditions for applying the STL technique successively, starting from the higher frequency and forwarding to the lower frequency. In this context, a higher frequency implies shorter periodicity, whereas a lower frequency implies longer periodicity.

Seasonal frequencies greater than one can be handled in different ways, as suggested by [19]. Namely, high-frequency observations, for example, carried out every minute, may exhibit an hourly seasonality (frequency=60), daily seasonality (frequency=24x60=1440), weekly seasonality (frequency=24x60x7=10080), or yearly seasonality (frequency=24x60x365.25=525960). In this application, seasonal frequencies  $k$  are set to 4, 34, and 739.19, corresponding to hourly, daily, and monthly seasonality of DAX trading volume, respectively. Yearly seasonality is not included, as it is unlikely to appear in the 15 min trading data, while quarterly seasonality is already nested within monthly seasonality. The monthly seasonality frequency was set to be a non-integer by calculating  $365.25 \times (5/7) \times 34/12 = 739.19$ . Altogether, 44350 observations were used over five years.

As already explained in the previous section, trading activity is expected to depend on the time of the day, day of the week, and month of the year. Informal evidence for this is found by graphical inspection of the data, that is, seasonal plots are used to decide which of the possible seasonal components should be included in a decomposition. Moreover, the selection of appropriate seasonal frequencies was approved by ACF plots.

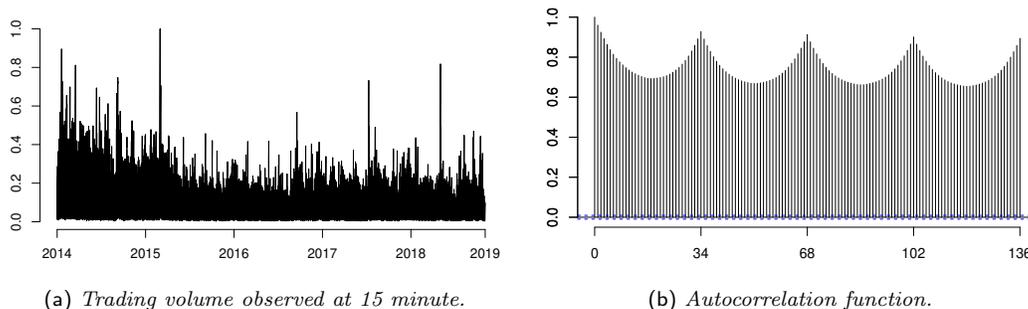
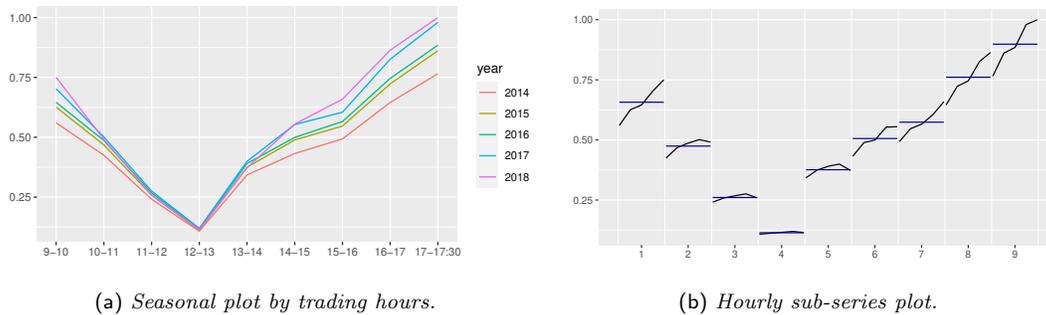


Figure 1: DAX index trading volume at 15 minute frequency from 2014-01-02 09:00 to 2018-12-28 17:30 and autocorrelation function up to 136 lags.

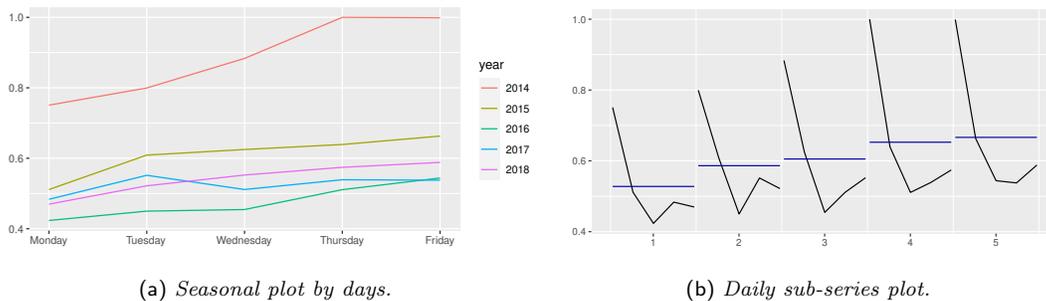
The left panel of **Figure 1** presents the intraday trading volume of the DAX index observed every 15 min. Due to the very large raw time-series scale, intraday trading volume was rescaled to its maximum. A substantial and frequent peak is noticeable over five years period, while the autocorrelation function indicates an intraday periodicity of length 34 (right panel of **Figure 1**). Moreover, "U"-shaped valleys of autocorrelation function support diurnal pattern, which is most common in practice. "U"-shaped diurnal pattern exhibits a much higher volume of trades at the beginning and the end of a trading day than in the lunchtime around midday hours (see [3]).

The diurnal pattern also has a repeating character according to [21], but it is not apparent at lower frequencies. However, by aggregating intraday trading volume daily, a similar periodicity can be discovered. It is not surprising to detect daily seasonality, i.e. accumulated information during weekends affect not only price changes but also trading volume on Mondays as the first weekdays. Additionally, investors are optimistic on Fridays and are prone to sell when prices increase due to the higher demand, as explained by [4]. Along with these two seasonal components, there could be even more components. Multiple seasonal components can also be comprehensively examined using seasonal plots. According to **Figure 2**, hourly seasonality within a trading day indicates "U"-shaped pattern.



**Figure 2:** Seasonal plot of hourly trading volume over years and hourly sub-series plot.

Both seasonal plots in **Figure 3** confirm daily seasonality within a week, that is, trading activity on Monday is always lower compared to other weekdays, and it increases while approaching the end of the week. On Friday, trading volume is the highest. Monthly seasonality within a year can be analogously found by employing the same seasonal plots (**Figure 4**). Monthly seasonality is also apparent, that is, in May, August and December, the trading volume experiences its downturns after increased volume in previous months. Interestingly, trading volume is the highest in January, while it is lowest in December.



**Figure 3:** Seasonal plot of daily trading volume over years and daily sub-series plot.

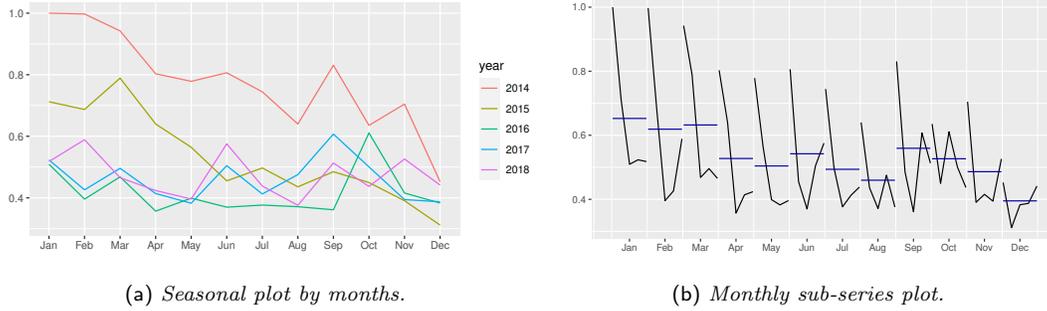


Figure 4: Seasonal plot of monthly trading volume over years and monthly sub-series plot.

|                         |                           |                        |                          |          |                           |
|-------------------------|---------------------------|------------------------|--------------------------|----------|---------------------------|
| $h_2$                   | $-0.155^{***}$<br>(0.002) | $d_2$                  | $0.009^{***}$<br>(0.001) | $m_2$    | $-0.002$<br>(0.002)       |
| $h_3$                   | $-0.183^{***}$<br>(0.002) | $d_3$                  | $0.014^{***}$<br>(0.001) | $m_3$    | $-0.005^{**}$<br>(0.002)  |
| $h_4$                   | $-0.219^{***}$<br>(0.002) | $d_4$                  | $0.024^{***}$<br>(0.001) | $m_4$    | $-0.024^{***}$<br>(0.002) |
| $h_5$                   | $-0.131^{***}$<br>(0.002) | $d_5$                  | $0.030^{***}$<br>(0.001) | $m_5$    | $-0.036^{***}$<br>(0.002) |
| $h_6$                   | $-0.107^{***}$<br>(0.002) |                        |                          | $m_6$    | $-0.027^{***}$<br>(0.002) |
| $h_7$                   | $-0.087^{***}$<br>(0.002) |                        |                          | $m_7$    | $-0.045^{***}$<br>(0.002) |
| $h_8$                   | $-0.061^{***}$<br>(0.002) |                        |                          | $m_8$    | $-0.052^{***}$<br>(0.002) |
| $h_9$                   | $0.028^{***}$<br>(0.002)  |                        |                          | $m_9$    | $-0.024^{***}$<br>(0.002) |
| $c$                     | $0.259^{***}$<br>(0.002)  |                        |                          | $m_{10}$ | $-0.035^{***}$<br>(0.002) |
|                         |                           |                        |                          | $m_{11}$ | $-0.041^{***}$<br>(0.002) |
|                         |                           |                        |                          | $m_{12}$ | $-0.049^{***}$<br>(0.002) |
| 15 min observations     |                           | 42957                  |                          |          |                           |
| R-squared               |                           | 0.513                  |                          |          |                           |
| Residual standard error |                           | 0.066                  |                          |          |                           |
| F statistic             |                           | 1963.37 <sup>***</sup> |                          |          |                           |

Table 2: OLS estimates of multiple seasonality (<sup>\*\*\*</sup>, <sup>\*\*</sup>, <sup>\*</sup> indicate significance of the estimates at 1%, 5%, and 10% respectively, while the standard errors are in parentheses).

Before multiple STL decomposition, OLS estimates are computed to verify whether the results support graphical inspection of the data and if the assumption of three seasonal components is reasonable. Namely, eight seasonal dummy variables were generated to capture hourly seasonality within 15 min observations over the entire sample (opening hour 09:00-10:00 is omitted). To capture daily seasonality and monthly seasonality 4+11 dummy variables were generated while omitting Monday as the first day of the week and January as the first month of the year. In total 23 seasonal dummy variables are included in the OLS regression and 24 parameters are estimated, including the constant term (Table 2).

Estimates with respect to closing hours indicates significant and higher trading levels against omitted trading hours (09:00-10:00), while all other intraday negative estimates with respect to  $h_2, h_3, \dots, h_8$  indicate significant and lower trading levels against opening hours. However, these negative estimates increase toward midday and then decrease toward the closing time of trading. This finding is in line with "U"-shaped intraday pattern. Considering estimates of daily effects concerning  $d_2, d_3, \dots, d_5$ , all parameters are significant and show higher trading levels, which increases almost linearly when approaching the end of the week.

Monthly effect estimates w. r. t.  $m_2, m_3, \dots, m_{12}$  are all negative, implying that trading is the highest in January, while the most negative estimate at 0.049 indicates that trading is the lowest in December. Downturns of negative estimates are noticeable in May and August. Although OLS estimates are taken only as auxiliary results, due to the restriction that seasonality does not change over time, it is surprising to conclude that more than half of the intraday trading volume variations can be solely explained by multiple seasonal patterns ( $R^2 = 0.513$ ).

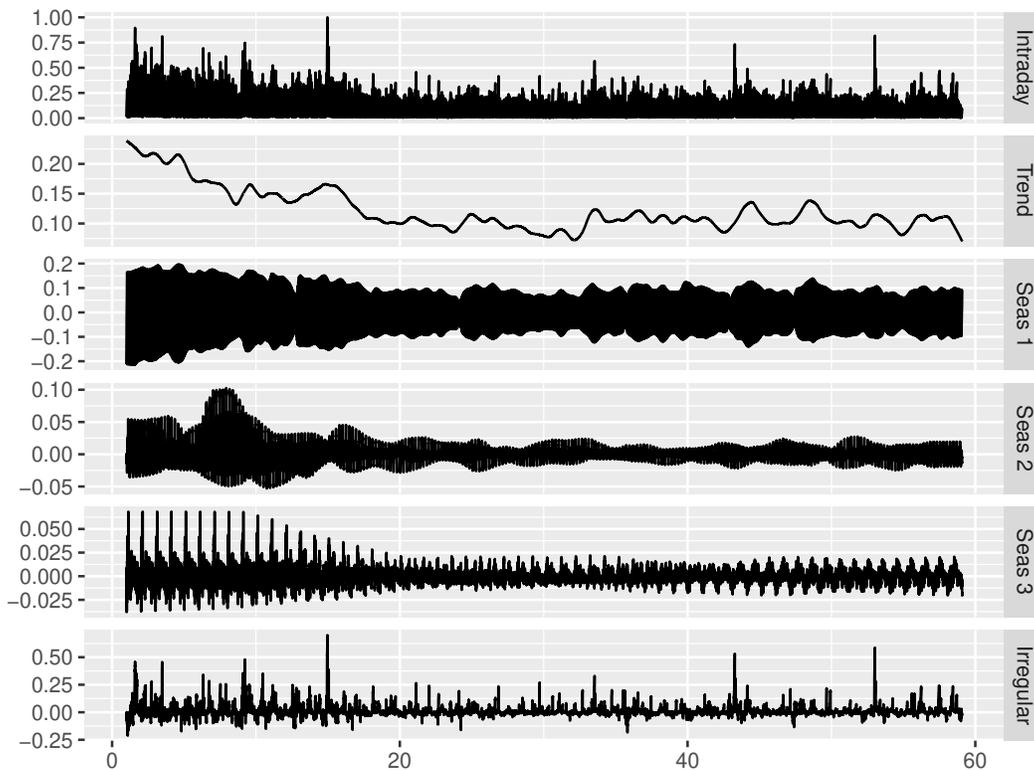


Figure 5: Multiple STL decomposition with 3 seasonal components.

The results of multiple STL decompositions are presented in Figure 5 using the Loess window  $q=39$  and seasonal frequencies  $k=4, 34,$  and  $739.19,$  respectively. Seas 1 is an hourly seasonal component, Seas 2 is a daily seasonal component, and Seas 3 is a monthly seasonal component of the intraday trading volume. The raw time-series, trend component, and irregular component are also presented in the same figure. From Figure 5 it is clear that all seasonal components are time-varying, and consequently, the strength of these components also changes.

The results from the first column of Table 3 suggest that hourly seasonality is the strongest among seasonal effects, that is, the strength of hourly seasonality over a five-year period is around 0.7, which is very close to 1. When analyzing the variation of the first seasonal component toward the remainder component in each year separately, it can be noticed that hourly seasonality is still the strongest, but it does not vary much (it ranges from 0.654 to 0.742). The second strongest seasonal component is the daily seasonal component, with a moderate strength of 0.124. However, daily seasonality changes the most between years, with a range of 0.06. The weakest component was monthly component with a strength below 0.1. After the first strongest seasonal component, the smoothed trend component surprisingly takes a relatively high portion of the variation in magnitude compared to the remainder (0.356 for the entire observation period).

| Seasonal effects  | Period/years |       |       |       |       |       |
|-------------------|--------------|-------|-------|-------|-------|-------|
|                   | 2014-2018    | 2014  | 2015  | 2016  | 2017  | 2018  |
| Hour of the day   | 0.697        | 0.742 | 0.729 | 0.654 | 0.673 | 0.708 |
| Day of the week   | 0.124        | 0.168 | 0.157 | 0.119 | 0.108 | 0.116 |
| Month of the year | 0.085        | 0.095 | 0.094 | 0.081 | 0.083 | 0.086 |
| Trend effect      | 0.356        | 0.346 | 0.320 | 0.312 | 0.324 | 0.338 |

Table 3: *Strength of the seasonal components and the trend component.*

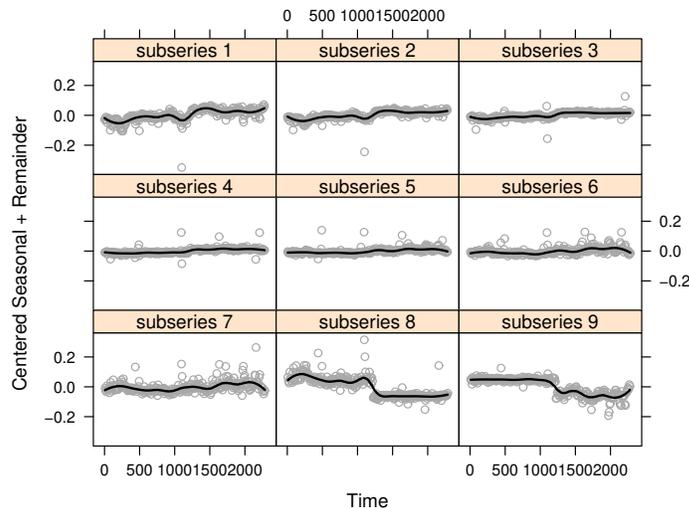


Figure 6: *Seasonal diagnostic plot within hourly sub-series.*

To check the appropriateness of the Loess window  $q=39$ , seasonal diagnostic plots were employed (Figures 6-8). These plots show detrended sub-subseries after subtracting the mean so that each sub-subseries is zero-centered. The sub-series of each seasonal frequency is well smoothed, but not too smoothed, indicating suitable Loess window selection. Moreover,  $q=39$  enables moderate changes in the seasonality, whereas a much higher level of  $q$  would give a constant seasonality.

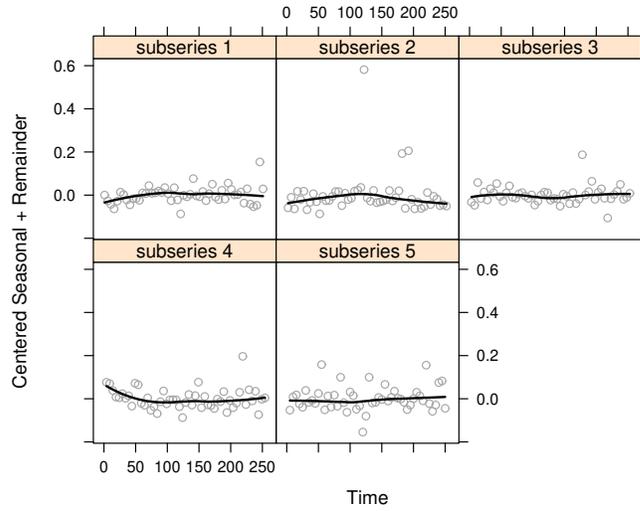


Figure 7: *Seasonal diagnostic plot within daily sub-series.*

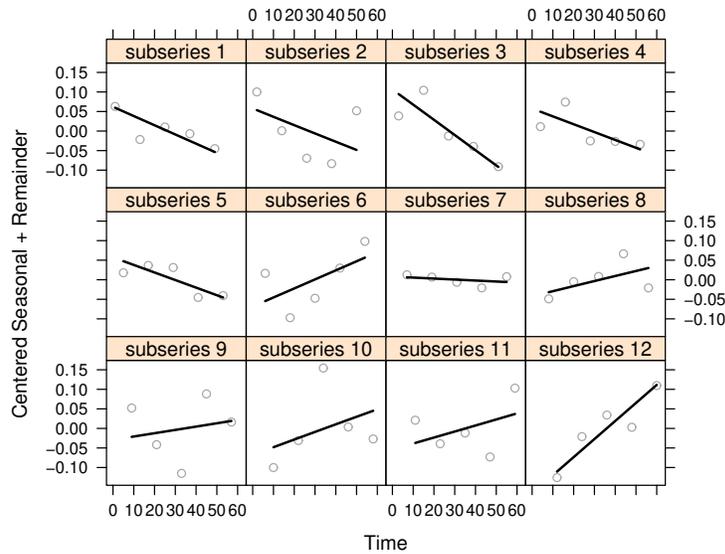


Figure 8: *Seasonal diagnostic plot within monthly sub-series.*

## 4. Conclusion

This study does not focus on intraday trading volume forecasting directly but on the discovery and understanding of multi-seasonal patterns that coexist within a single high-frequency series. Thus, a multiple STL decomposition was employed for DAX intraday trading volume observed every 15 min over five years. The empirical results demonstrated the existence of three seasonal components. Hourly seasonality is the strongest component, indicating a "U"-shaped pattern, and persistent, as it does not vary much over time. However, daily seasonality changes the most, even though it exhibits moderate strength over the years. The daily seasonality indicates an almost linearly increasing trade pattern when approaching the end of the week. Friday volume was the highest, as investors did not want to hold open positions over a weekend because of the increased market risk. For the same reason, trading activity was mostly intensified during the closing hours of the day. Monthly seasonality is also apparent with downturns in May, August, and December after increased trading in the previous months, while the highest trading activity is experienced in January. Multi-seasonality patterns provide better insight into investors' behavior, and consequently, the trading volume becomes a predictable phenomenon that is used in decision making when there is a lack of fundamental information.

What makes this paper distinct from similar existing studies is the suggestion on how to deal with an unequal number of trading days per week, month, and year as well as appropriate specification of seasonal frequencies and Loess window. Having the same number of observations per sub-cycle, allowing for missing values and handling a multiple set of any seasonal frequencies greater than one, makes perfect conditions for applying the STL technique successively, starting with the shortest periodicity, and ending with the longest periodicity. When an appropriate  $q$  is selected, along with  $k$ , all other parameters are determined automatically, as they strongly depend on the time-series periodicity and the length of the Loess window. In the application example, parameter  $q$  is set to 39, while  $k$  considers three seasonal frequencies: 4, 34, and 739.19. In addition, a non-integer seasonal frequency is employed. The appropriateness of these parameter settings was confirmed using seasonal diagnostic plots.

Finally, it should be noted that the auxiliary results indicate that trading volume variations can be explained solely by using more than 50% of multi-seasonal patterns, which is surprisingly a high portion. A comparison of multiple STL decompositions against competing alternatives for forecasting intraday trading volume would be a direction for further research, although the advantage of decomposition by a locally weighted polynomial is well-argued. It can also be beneficial to explore the performance of parametric versions of multiple seasonal decompositions, which may be adopted for real data features.

## Acknowledgements

This work was fully supported by the Croatian Science Foundation under the project "Volatility measurement, modeling and forecasting" (UIP-2013-11-5199).

## References

- [1] Andersen, T. G. and Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*, 4(2–3), 115–158. doi: [10.1016/s0927-5398\(97\)00004-2](https://doi.org/10.1016/s0927-5398(97)00004-2)
- [2] Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71(2), 579–625. doi: [10.1111/1468-0262.00418](https://doi.org/10.1111/1468-0262.00418)
- [3] Arnerić, J. and Matković, M. (2019). Challenges of integrated variance estimation in emerging stock markets. *Proceedings of Rijeka Faculty of Economics: Journal of Economics and Business*, 37(2), 713–739. doi: [10.18045/zbefri.2019.2.713](https://doi.org/10.18045/zbefri.2019.2.713)

- [4] Arnerić, J. and Škrabić Perić, B. (2018). Panel GARCH model with cross-sectional dependence between CEE emerging markets in trading day effects analysis. *Romanian Journal of Economic Forecasting*, 21(4), 71–84. [http://www.ipe.ro/rjef/rjef4\\_18/rjef4\\_2018p71-84.pdf](http://www.ipe.ro/rjef/rjef4_18/rjef4_2018p71-84.pdf)
- [5] Bergmeir, C., Hyndman, R. J. and Benítez, J. M. (2016). Bagging exponential smoothing methods using STL decomposition and Box–Cox transformation. *International Journal of Forecasting*, 32(2), 303–312. doi: 10.1016/j.ijforecast.2015.07.002
- [6] Białkowski, J., Darolles, S. and Le Fol, G. (2008). Improving VWAP strategies: A dynamic volume approach. *Journal of Banking and Finance*, 32(9), 1709–1722. doi: 10.1016/j.jbankfin.2007.09.023
- [7] Boudt, K., Croux, C. and Laurent, S. (2011). Robust estimation of intraweek periodicity in volatility and jump detection. *Journal of Empirical Finance*, 18(2), 353–367. doi: 10.1016/j.jempfin.2010.11.005
- [8] Boudt, K. and Zhang, J. (2015). Jump robust two time scale covariance estimation and realized volatility budgets. *Quantitative Finance*, 15(6), 1041–1054. doi: 10.1080/14697688.2012.741692
- [9] Brownless, C. T., Cipollini, F. and Gallo, G. M. (2011). Intra-daily volume modeling and prediction for algorithmic trading. *Journal of Financial Econometrics*, 9(3), 489–518. doi: 10.1093/jffinec/nbq024
- [10] Brownless, C. T. and Gallo, G. M. (2006). Financial econometric analysis at ultra-high frequency: Data handling concerns. *Computational Statistics & Data Analysis*, 51(4), 2232–2245. doi: 10.1016/j.csda.2006.09.030
- [11] Cleveland, R. B., Cleveland, W. S., McRae, J. E. and Terpenning, I. (1990). STL: A seasonal-trend decomposition procedure based on Loess. *Journal of Official Statistics*, 6(1), 3–73. <https://www.proquest.com/docview/1266805989?pq-origsite=gscholar&cbl=105444&fromopenview=true>
- [12] Gould, P. G., Koehler, A. B., Ord, J. D., Snyder, R. D., Hyndman, R. J. and Vahid-Araghi, F. (2008). Forecasting time series with multiple seasonal patterns. *European Journal of Operational Research*, 191(1), 207–222. doi: 10.1016/j.ejor.2007.08.024
- [13] Gouriéroux, C. S., Jasiak, J. and Le Fol, G. (1999). Intra-day market activity. *Journal of Financial Markets*, 2(3), 193–226. doi: 10.1016/s1386-4181(99)00004-x
- [14] Hafen, R. P. (2010). *Local regression models: Advancements, applications and new methods*. Dissertation submitted to the Faculty of Purdue University, Indiana. [http://ml.stat.purdue.edu/hafen/preprints/Hafen\\_thesis.pdf](http://ml.stat.purdue.edu/hafen/preprints/Hafen_thesis.pdf)
- [15] Härdle, W. K., Hautsch, N. and Mihoci, A. (2012). Modelling and forecasting liquidity supply using semiparametric factor dynamics. *Journal of Empirical Finance*, 19(4), 610–625. doi: 10.1016/j.jempfin.2012.04.002
- [16] Hyndman, R. J., Koehler, A. B., Ord, J. K. and Snyder, R. D. (2008). *Forecasting with exponential smoothing: The state space approach*. Springer. doi: 10.1111/j.1751-5823.2009.00085.17.x
- [17] Hyndman, R. J., Koehler, A. B., Snyder, R. D. and Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, 18(3), 439–454. doi: 10.1016/s0169-2070(01)00110-8
- [18] Hyndman, R. J. and Khandakar, Y. (2008). Automatic time-series forecasting: The forecast package for R. *Journal of Statistical Software*, 27(3), 439–454. doi: 10.18637/jss.v027.i03
- [19] Hyndman, R. J. and Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*. OTexts. <https://otexts.com/fpp2/>
- [20] Ito, R. (2013). Modeling dynamic diurnal patterns in high frequency financial data. *Cambridge Working Papers in Economics*, 1315, 1–35. doi: 10.17863/CAM.5637
- [21] Liu, C. and Maheu, J. M. (2012). Intraday dynamics of volatility and duration: Evidence from Chinese stocks. *Pacific-Basin Finance Journal*, 20(3), 329–348. doi: 10.1016/j.pacfin.2011.11.001
- [22] Proietti, T. and Pedregal, D. J. (2021). Seasonality in high frequency time series. *CEIS Tor Vergata Research Papers*, 19(2), 1–29. doi: 10.2139/ssrn.3802611
- [23] Satish, V., Saxena, A. and Palmer, M. (2018). Predicting intraday trading volume and volume percentages. *The Journal of Trading*, 13(4), 107–116. doi: 10.3905/jot.2018.13.4.107
- [24] Seong, B. (2020). Smoothing and forecasting mixed-frequency time series with vector exponential smoothing models. *Economic Modelling*, 91, 463–468. doi: 10.1016/j.econmod.2020.06.020
- [25] Taylor, J. W. (2003). Short-term electricity demand forecasting using double seasonal exponential smoothing. *Journal of Operational Research Society*, 54(8), 799–805. doi: 10.1057/pal-

- [grave.jors.2601589](#)
- [26] Taylor, J. W. and Snyder, R. D. (2012). Forecasting intraday time-series with multiple seasonal cycles using parsimonious seasonal exponential smoothing. *Omega*, 40(6), 748–757. doi: [10.1016/j.omega.2010.03.004](#)
  - [27] Theodosiou, M. (2011). Forecasting monthly and quarterly time-series using STL decomposition. *International Journal of Forecasting*, 27, 1178–1195. doi: [10.1016/j.ijforecast.2010.11.002](#)
  - [28] Wang, X., Smith, K. and Hyndman, R. J. (2006). Characteristic-based clustering for time-series data. *Data Mining and Knowledge Discovery*, 13(3), 335–364. doi: [10.1007/s10618-005-0039-x](#)
  - [29] Zhang, L., Mykland, P. A. and Aït-Sahalia, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*, 100(472), 1394–1411. doi: [10.1198/016214505000000169](#)
  - [30] Zhang, X. and Wang, J. (2018). A novel decomposition-ensemble model for forecasting short-term load-time series with multiple seasonal patterns. *Applied Soft Computing*, 65, 478–494. doi: [10.1016/j.asoc.2018.01.017](#)