

# Identifying Leading Indicators for Tactical Truck Parts Sales Predictions using LASSO

Dylan Gerritsen, Vahideh Reshadat\*

Eindhoven University of Technology, Eindhoven, Netherlands  
`d.m.gerritsen@student.tue.nl`, `v.reshadat@tue.nl`

**Abstract.** This paper aimed to identify leading indicators for a case company that supplies truck parts to the European truck aftersales market. We used LASSO to extract relevant information from a collected pool of business, economic, and market indicators. We propose the efficient one-standard error rule, as an alternative to the default one-standard error rule, to reduce the influence of sampling variation on the LASSO tuning parameter value. We found that applying the efficient one-standard error rule over the default one, improved forecasting performance with an average of 0.73%. Next to that, we found that, for our case study, applying forecast combination yielded the best forecasting performance, outperforming all other considered models, with an average improvement of 2.38%. Thus, including leading context information did lead to more accurate parts sales predictions for the case company. Also, due to the transparency of LASSO, using LASSO provided business intelligence about relevant predictors and lead effects. Finally, from a pool of 34 indicators, 7 indicators appeared to have clear lead effects for the case company.

**Keywords:** LASSO, Sales Forecasting, Leading Indicators

## 1 Introduction

Sales forecasting plays a significant role in business strategies nowadays. In particular, tactical (i.e. up to 12 months) forecasting often supports short-term decision-making in supply chain management as it serves as a basis for raw material purchase, inventory planning, and production scheduling. Alternatively, strategic forecasting is often referred to as long-term forecasting, and principally supports decision-making in the development of overall strategies and capacity planning. Both forecasting strategies commonly use observed values of the past and available knowledge to predict the future as accurately as possible [2]. However, including external information could enhance the performance of a sales forecasting model [1]. According to Currie and Rowley [1] using additional information can enhance forecasting performance in volatile environments. The main focus of previous research has been enhancing operational forecasts (i.e. up to 48 hours ahead). For example, Williams et.al. [3] successfully integrated supply chain information into a forecasting model, whereas Ma et.al., [4] used additional price and promotional data to improve forecast accuracy. Conversely,

the dynamics of tactical forecasts can be different due to the relevant horizons and business models. Moreover, leading indicators, such as macroeconomics, can contain leading context information in terms of changing economic conditions [2]. These indicators are mainly published on a monthly or quarterly basis and are therefore useless for operational forecasting purposes. However, for medium to long-term horizons (i.e. 3 to 12 months ahead), macroeconomic information is relevant and could enhance forecast performance [5].

This paper aims to identify leading indicators for a case company that supplies truck parts to the European truck aftersales market. Also, this paper explores whether including leading context information leads to more accurate sales predictions in comparison with traditional time series methods. Additionally, this paper has several contributions to the existing academic literature:

- We propose the efficient one-standard error rule, as an alternative to the default one-standard error rule, by combining efficient CV, proposed in Jung [20], with the commonly used one-standard error rule, described in Hastie et al. [16]. The purpose of the efficient one-standard error rule is to reduce the influence of sampling variation on the LASSO tuning parameter value.
- The studies of Sagaert et al. [2] and Verstraete et al. [6] reported forecasting performance losses of LASSO on the longer horizons, compared to traditional methods. These studies solely considered and compared the forecasting performance of individual models, whereas we found that, for our case study, applying forecast combination resulted in improved forecasting performance over almost all horizons.

The rest of this paper is organized as follows. The next section discusses relevant literature related to identifying leading indicators in a tactical sales forecasting environment. Section 3 introduces the case study, the modeling characteristics we are dealing with, and the defined experimental setup. Finally, the results are presented in Section 4, followed by a brief conclusion in Section 5.

## 2 Related Work

### 2.1 Tactical Sales Forecasting Problem

Forecasting methods have large influence on the development of different artificial intelligent branches consists of Fuzzy Systems [7], Natural Language Processing [8][9][10][11], Expert Systems [12] etc.

According to Verstraete et.al.,[6] using macroeconomic indicators as input variables for tactical sales forecasting introduces two major challenges. The first challenge is the limited sample size of available sales data, which is considered a typical challenge in sales forecasting in general. It is often the case that companies lack having effective data management practices and therefore cannot access historical data. On the other hand, even if companies do have effective data management practices to a certain extent, it is often the case that historical data is not representative anymore due to changing product portfolios and customer behaviors. Moreover, as macroeconomic data is mainly reported

on monthly or higher aggregation levels, it could be that the amount of usable data becomes even more limited. The second challenge is the large number of available macroeconomic indicators across multiple publicly available data sources. For example, the best known economic database sources such as Organisation for Economic Cooperation and Development (OECD), Federal Reserve Economic Data (FRED), and Eurostat provide access to thousands of macroeconomic time-series. As a result, selecting potential macroeconomic indicators could become very time consuming and quite complex. According to Sagaert et al. [2], these two challenges together create a distinct tactical sales forecasting problem. Hence, the tactical sales forecasting problem consists of a considerably large set of predictors ( $p$ ) with limited sales data sample sizes ( $n$ ).

## 2.2 Forecasting Frameworks

Sagaert et al. [13] proposed a framework to improve tactical sales forecasting using macroeconomic leading indicators. The proposed framework automatically identified key leading indicators from an enormous set of macroeconomic indicators for a supplier to the tire industry. In many studies, the gross domestic product (GDP) is used to represent the ongoing economic activity at some point in time. For clarity, GDP represents the value of all finished goods and services produced within a country in a specific time period. However, Sagaert et al. [2] indicated that GDP is principally an aggregate variable and therefore does not provide detailed changes in the various sectors and economic activities. Moreover, they mention that using more detailed macroeconomic indicators could provide relevant information and therefore the number of potentially relevant indicators will increase extensively, especially for supply chains across multiple markets and countries. The case company of Sagaert et al. [2] has a global supply chain and supplies numerous tire manufacturers across multiple markets and as a result they initially selected 67,851 monthly macroeconomic variables from several sections of the FRED database. To model any indicator leading effects to the sales variable, each input variable is lagged in time up to a maximum considered time lag. A maximum leading effect of 12 months is assumed and therefore the number of input variables increased to a total of  $67,851 * 12 = 814,212$  predictors. Due to the extensive number of predictors causal-regression modeling becomes highly complex and truly impossible. For this reason, [13] proposed using LASSO regression. Moreover, they opted for the use of LASSO regression as “the LASSO forecast is transparent, and provides insights into the selected leading indicators. Experts can benefit by gaining a better understanding of their market and can thus improve their understanding of market dynamics and interactions” [13, p. 127]. Additionally, since each input variable is lagged in time multiple times, they highlight that multicollinearity may be present among the input variables. Due to the shrinkage properties of LASSO, authors mention that LASSO is capable of effectively dealing with multicollinearity.

Verstraete et al. [6] proposed a comparable methodology that automatically generates tactical sales forecasts by using a large group of macroeconomic indicators. A noticeable difference is they assumed that macroeconomic conditions

determine the trend of sales. Therefore, they opted to use the LASSO regression technique of Sagaert et al. [2] to forecast the trend component. The sales data is decomposed into a trend, a seasonal, and a remainder component using the STL decomposition proposed by Cleveland et.al., [14]. Additionally, they motivate their choice for STL decomposition because it can be robust to outliers, the seasonal component may alter across different periods, and the smoothness of the trend component is controllable. Furthermore, as the data is divided into three independent components, each component is forecasted separately. Verstraete et al. [6] used the seasonal naive method as proposed by Xiong et.al., [15] to forecast the seasonal component. For clarity, the seasonal naive method uses the latest seasonal observation as a forecast for the consecutive seasonal period. Verstraete et al. [6] assumed that the remainder component is determined by other factors than macroeconomic indicators. For instance, they mention social media, promotions, weather, and random noise as factors that will mainly determine the remainder component. However, they considered predicting the remainder component out-of-scope as they assumed that the predictive power of these factors is out of the tactical time window.

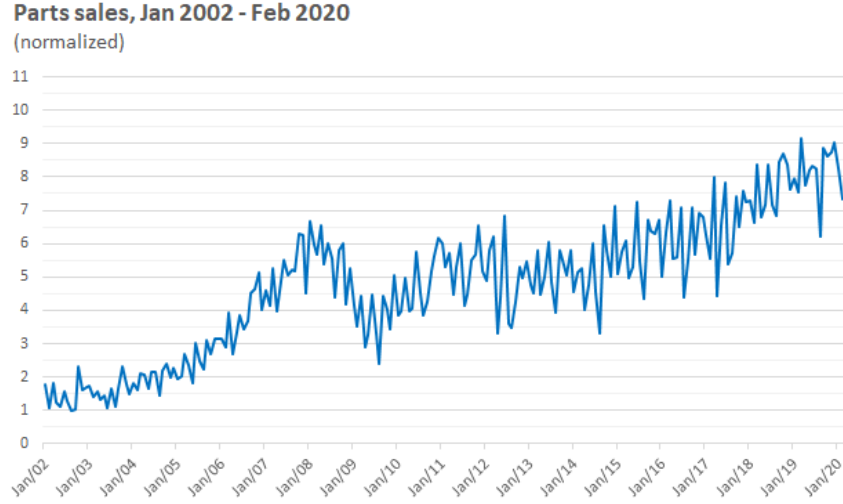
### 3 Case Study

#### 3.1 Target Variable

Leading indicators are defined as variables that contain predictive information and ideally can predict a certain movement for a target variable in advance. Hence, to identify any leading indicators that are relevant for the case company's parts sales, it is necessary to specify a target variable. For the case company, we specified the target variable as the total monthly truck parts sales, reported by the entire European dealer network. Figure 1 shows the specified target variable in this case study. It should be noted that the y-axis is normalized due to data confidentiality. As can be seen, observed sales data is available from January 2002 to February 2020 which is equivalent to 218 monthly observations. Moreover, it becomes clear that the sales data is following a certain trend and that it contains a seasonal pattern that repeats every 12 months.

#### 3.2 Leading Indicators

The studies of Sagaert et al. [13] and Verstraete et al. [6] only included publicly available macroeconomic indicators as potential leading indicators. Concerning this field project, it was possible to access non-publicly available databases with data specifically related to the case company and the European road freight market. Therefore, a distinction is made between three different types of potential leading indicators: business, economic, and market indicators. We collected several indicators that have the potential of being a leading indicator for the case company's parts sales. To collect indicators that have the potential of being a leading indicator for the case company's parts sales, different data sources



**Fig. 1.** Parts sales across the entire European dealer network.

were used. First of all, internal departments of the case company served as data sources for several business indicators that covered the case company's business activities. Secondly, the Eurostat and OECD publicly available economic databases provided access to thousands of macroeconomic time series related to European territories and served as data sources for several economic indicators that covered Europe's overall economic climate. Finally, Rementum Research & Management, further denoted as Rementum, is a market research and advisory firm specialized in both the European road freight market and commercial vehicles with a gross vehicle weight over 6 tonnes. Rementum collects data from a broad array of sources relevant for heavy commercial road transport such as road carriers, transport equipment OEMs, and OE & aftermarket component suppliers to analyze the European road transport market conditions. The expertise of Rementum was used to collect several market indicators that covered the ongoing activities in Europe's road transport sector. Accordingly, Table 1 presents an overview of all included business, economic and market indicators in this case study. Noticeably, all indicators have data available from January 2005, and thus we have January 2005 to February 2020 available for data preparation and modeling, which is equivalent to 182 observations.

Table 1: Collected indicators that have the potential of being a leading indicator.

|       | Indicator description                   | Unit   | Source     |
|-------|---|--------|------------|
| $Y$   | <b>Autoregressive information:</b> NSA  | €      | Case comp. |
| $X_1$ | <b>Truck deliveries:</b> NSA            | Trucks | Case comp. |
| $X_2$ | <b>Business climate:</b> SA             | Index  | Rementum   |
| $X_3$ | <b>Economic sentiment:</b> SA           | Index  | Eurostat   |
| $X_4$ | <b>Passenger car registrations:</b> NSA | Cars   | Rementum   |

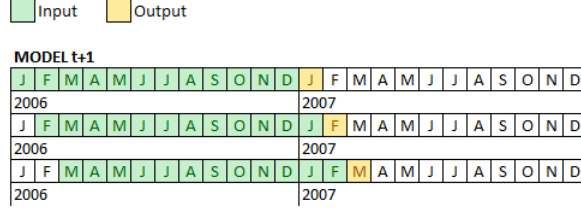
|          |  |                |          |
|----------|--|----------------|----------|
| $X_5$    | <b>Retail confidence:</b> SA                                   | Balance        | Eurostat |
| $X_6$    | <b>Industrial confidence:</b> SA                               | Balance        | Eurostat |
| $X_7$    | <b>Industrial production:</b> SA                               | Index          | Eurostat |
| $X_8$    | <b>Gross domestic product:</b> SA                              | Index          | OECD     |
| $X_9$    | <b>Producer price index:</b> SA                                | Index          | Eurostat |
| $X_{10}$ | <b>Construction confidence:</b> SA                             | Balance        | Eurostat |
| $X_{11}$ | <b>Construction spending:</b> NSA                              | Index          | Rementum |
| $X_{12}$ | <b>Construction activity:</b> SA                               | Index          | Eurostat |
| $X_{13}$ | <b>Construction and mining equipment sales:</b> MA (3mos), NSA | Index          | Rementum |
| $X_{14}$ | <b>Replacement truck tire sales (ST):</b> MA (2mos), NSA       | Index          | Rementum |
| $X_{15}$ | <b>Replacement truck tire sales (LT):</b> MA (12mos), NSA      | Index          | Rementum |
| $X_{16}$ | <b>OE truck tire sales (ST):</b> MA (2mos), NSA                | Index          | Rementum |
| $X_{17}$ | <b>OE truck tire sales (LT):</b> MA (12mos), NSA               | Index          | Rementum |
| $X_{18}$ | <b>Aftermarket truck tire deliveries:</b> NSA                  | Index          | Rementum |
| $X_{19}$ | <b>Diesel consumption growth (LT):</b> MA (12mos), NSA         | %              | Rementum |
| $X_{20}$ | <b>Diesel consumption growth (ST):</b> MA (2mos), NSA          | %              | Rementum |
| $X_{21}$ | <b>Retail diesel price:</b> NSA                                | €              | Rementum |
| $X_{22}$ | <b>Automotive diesel deliveries:</b> NSA                       | m <sup>3</sup> | Rementum |
| $X_{23}$ | <b>OEM truck orders growth (LT):</b> MA (12mos), NSA           | %              | Rementum |
| $X_{24}$ | <b>OEM truck orders (ST):</b> MA (3mos), NSA                   | Trucks         | Rementum |
| $X_{25}$ | <b>OEM order intake expectations:</b> MA (6mos), NSA           | Balance        | Rementum |
| $X_{26}$ | <b>OEM production expectations:</b> MA (6mos), NSA             | Balance        | Rementum |
| $X_{27}$ | <b>Road transport activity:</b> NSA                            | Index          | Rementum |
| $X_{28}$ | <b>Road transport capacity:</b> SA                             | Index          | Rementum |
| $X_{29}$ | <b>Freight volume index:</b> MA (3mos), NSA                    | Index          | Rementum |
| $X_{30}$ | <b>Carrier confidence:</b> NSA                                 | Balance        | Rementum |
| $X_{31}$ | <b>Carrier demand expectations:</b> NSA                        | Balance        | Rementum |
| $X_{32}$ | <b>Carrier hiring expectations:</b> NSA                        | Balance        | Rementum |
| $X_{33}$ | <b>Carrier pricing expectations:</b> NSA                       | Balance        | Rementum |

---

ST = Short-term; LT = Long-term; SA = Seasonally adjusted; NSA = Not seasonally adjusted; MA = Moving average; mos = months;

### 3.3 Modeling of Lead Effects

The previous sections have shown that the identification of leading indicators in a sales forecasting environment results in a high-dimensional problem with the presence of multicollinearity among the predictors. Accordingly, LASSO has already been found useful in the identification of leading indicators, due to its shrinkage properties and transparency [2, 6, 13]. As a result, for the same reasoning, LASSO is chosen as the modeling technique for this case study. In order to use regression techniques on a forecasting problem with temporal dependencies, the data must be restructured to a supervised learning task. Given a sequence of numbers for a time series, the data can be restructured as a supervised learning task by using previous time steps as input variables and the next time step as the output variable. Sagaert et al. [13] indicated that macroeconomic indicators can contain leading context information up to a maximum horizon of 12 months, and thus this case study includes results up to 12 months ahead. Next to that, to model any lead effects, the decision was made to include all 12 previous time steps as input variables. This means that every potential leading indicator is lagged in time 12 times, and therefore the number of predictors increases significantly from 34 business, economic and market indicators to a total number of  $34 * 12 = 408$  predictors. Hence, we are dealing with a high-dimensional problem since the number of predictors exceeds the number of observations ( $p > n$ ). Figure 2 shows how the data is restructured to a supervised learning task for the first three observations when predicting  $(t + 1)$  ahead.



**Fig. 2.** Time series data to supervised learning setting.

### 3.4 Proposed Method for Identifying Lead Effects

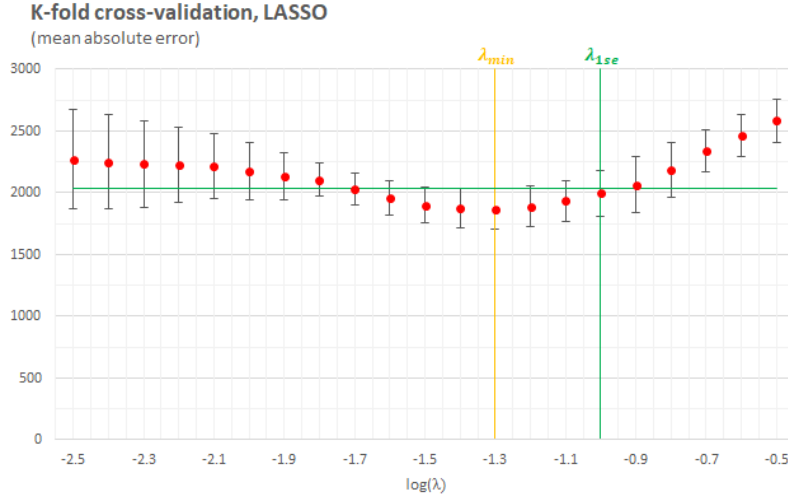
LASSO is a linear regression analysis method that performs both variable selection and regularization in order to prevent overfitting of high-dimensional data, and in order to enhance both prediction accuracy and model interpretability [18]. The LASSO solution minimizes a penalized residual sum of squares, yielding coefficients that are shrunk to zero:

$$\hat{\beta}^{lasso} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^n \left( y_i - \beta_0 - \sum_{p=1}^p \beta_p x_{ip} \right) + \lambda \sum_{p=1}^p |\beta_p| \quad (1)$$

The solution and thus the  $\hat{\beta}^{lasso}$  estimator highly depends on the magnitude of regularization, which equals a value between 0 and 1, and is represented by tuning parameter  $\lambda$ . Hastie et.al., [16] proposes to determine  $\lambda$  based on the cross-validation estimate of the prediction error. Typically,  $K$ -fold cross-validation (KCV) randomly splits the data into  $K$ -folds and subsequently fits a model using  $K - 1$  folds and uses the  $K^{th}$  fold for testing. As the data is partitioned randomly, using KCV in a time series environment does not seem applicable as temporal dependencies are ignored. Nevertheless, using CV in a time series environment was extensively studied by Bergmeir and Benítez [17], and they did not find any practical problems with standard cross-validation. Moreover, they suggest to use standard KCV or blocked CV together with stationary data, as this uses all available information for training and testing. Accordingly, we used 10-fold CV with stationary data to determine the CV estimate on the in-sample data. Additionally, we selected  $\lambda$  corresponding to the most regularized model within one-standard error ( $\lambda_{1se}$ ) of the minimum CV error estimate ( $\lambda_{min}$ ), also known as the one-standard error rule [16]. Figure 3 shows an example of how tuning parameter values  $\lambda_{min}$  and  $\lambda_{1se}$  are determined based on the CV error curve with vertical standard error bars.

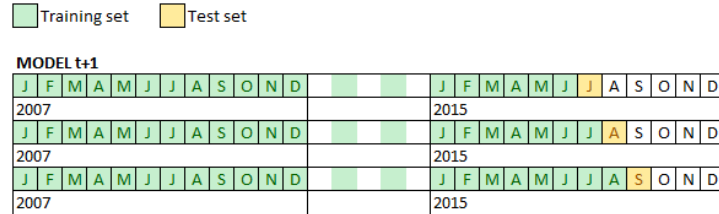
### 3.5 Experimental Setup

To conduct any experiments, the observations available for modeling must be split into two subsets: in-sample observations (i.e. training data) and out-of-sample observations (i.e. test data). A common approach to split training and



**Fig. 3.** Tuning parameter values  $\lambda_{min}$  and  $\lambda_{1se}$ .

test sets when dealing with temporal dependencies, is time series cross-validation [19]. This approach uses a series of test sets, with each test set consisting of a single observation. The corresponding training set consists only of observations prior to the observation in the test set. Figure 4 shows how the training and test sets are defined, when predicting  $(t + 1)$  ahead. We set the initial size of the in-sample data to 70% of the available sample size, after which, every time new observations become available, the in-sample data is updated. The smallest



**Fig. 4.** Time series cross-validation data split when predicting  $(t + 1)$  ahead.

sample size, after data preparation activities, equals 147 observations for model  $M_{t+12}$  and as a result, the initial size of the training set was set to 102 observations for all 12 models. Accordingly, Table 2 presents an overview of how many observations are used as a series of test sets in order to evaluate model performance. When aggregating all observations across all horizons, a total number of 606 observations are available for model performance evaluation. On the whole,



**Table 2.** Data available for model performance evaluation.

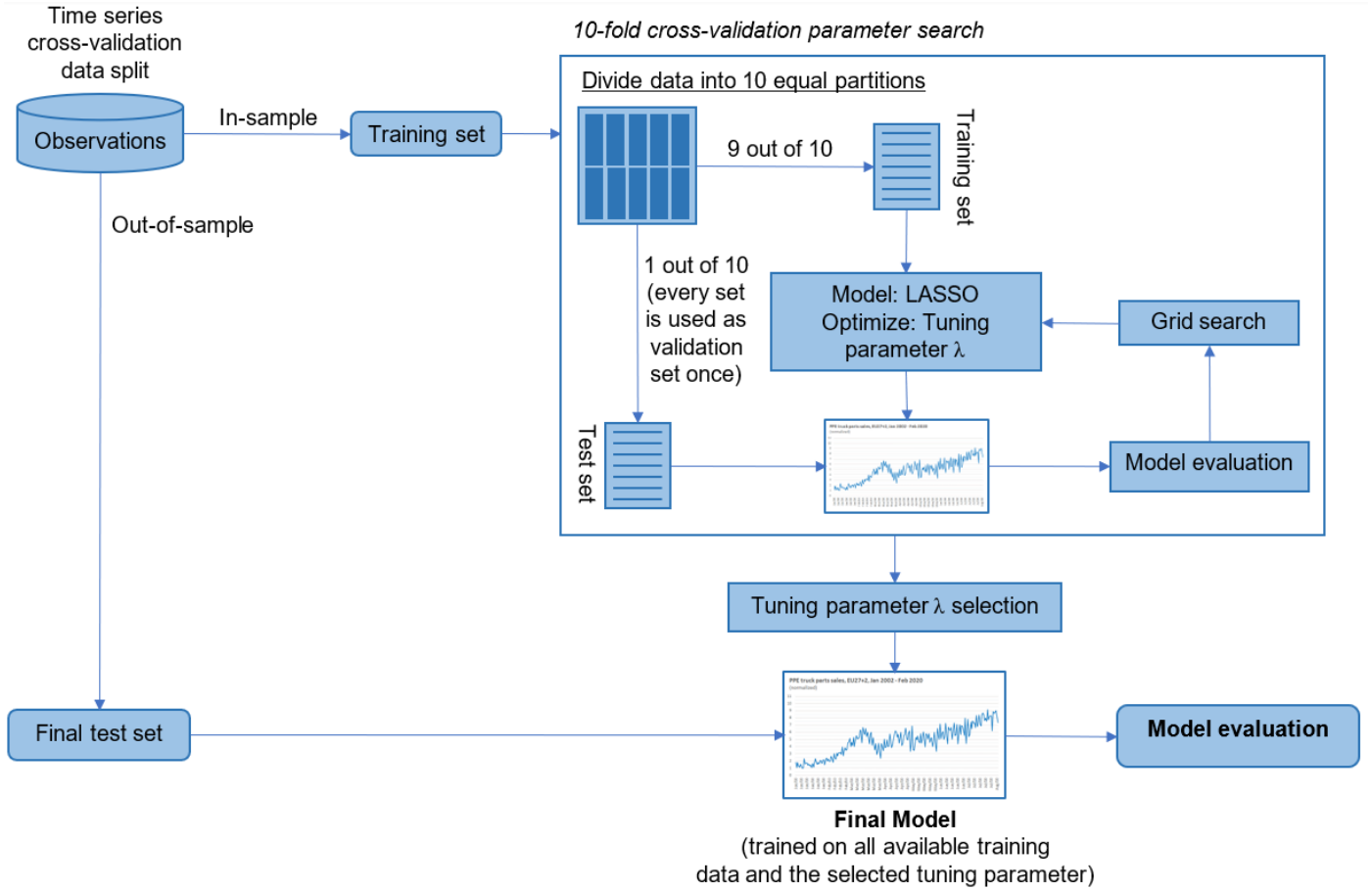
| Model Period used as test sets |                      | Number of test sets |
|--------------------------------|----------------------|---------------------|
| $M_{t+1}$                      | Jul 2015 to Feb 2020 | 56                  |
| $M_{t+2}$                      | Aug 2015 to Feb 2020 | 55                  |
| $M_{t+3}$                      | Sep 2015 to Feb 2020 | 54                  |
| $M_{t+4}$                      | Oct 2015 to Feb 2020 | 53                  |
| $M_{t+5}$                      | Nov 2015 to Feb 2020 | 52                  |
| $M_{t+6}$                      | Dec 2015 to Feb 2020 | 51                  |
| $M_{t+7}$                      | Jan 2016 to Feb 2020 | 50                  |
| $M_{t+8}$                      | Feb 2016 to Feb 2020 | 49                  |
| $M_{t+9}$                      | Mar 2016 to Feb 2020 | 48                  |
| $M_{t+10}$                     | Apr 2016 to Feb 2020 | 47                  |
| $M_{t+11}$                     | May 2016 to Feb 2020 | 46                  |
| $M_{t+12}$                     | Jun 2016 to Feb 2020 | 45                  |

Figure 5 shows the experiment design that will be used at every time step, across all forecast horizons. First of all, we determine the CV estimate of the prediction error using the in-sample data and a 10-fold CV grid search. Thereafter, using this grid search, we determine the value of tuning parameter  $\lambda$  with the commonly used one-standard error rule. Then, we fit a  $\hat{\beta}^{lasso}$  estimator using all available in-sample data and the selected tuning parameter  $\lambda$ , after which the  $\hat{\beta}^{lasso}$  estimator is used to predict the observation in the final test set.

## 4 Results

### 4.1 Case Study

The purpose of extracting information from leading indicators is to ideally improve the parts sales predictions by including information in terms of changing economic and market conditions. In order to assess whether including leading indicators improves forecasting performance, the performance of LASSO is benchmarked to commonly used univariate methods that are unable to respond to these changing conditions. Since we are dealing with a very small sample size, complex machine learning techniques, such as the recurrent neural network, are considered out of scope as these techniques require a large sample size for training purposes. The methods used as a benchmark are additive Holt-Winters (AHW), multiplicative Holt-Winters (MHW), and SARIMA. With regard to the benchmark models, both the additive (AHW) and multiplicative Holt-Winters (MHW) methods were implemented with optimal smoothing parameters  $\alpha = 0.2$ ,  $\beta = 0.1$ ,  $\gamma = 0.2$ , whereas a SARIMA model with  $(p, d, q)(P, D, Q)_m$  equal to  $(2, 1, 0)(1, 1, 1)_{12}$  was selected using the Akaike information criterion (AIC). Overall, Table 3 shows the mean absolute prediction error across all considered models and forecast horizons.



**Fig. 5.** Experimental setup.

**Table 3.** Mean absolute prediction error across all models and forecast horizons.

| Forecast horizon | LASSO           | SARIMA          | AHW             | MHW      |
|------------------|-----------------|-----------------|-----------------|----------|
| 1-month          | 1,865.93        | 1,876.55        | <b>1,732.56</b> | 1,994.59 |
| 2-months         | 1,883.23        | 1,778.85        | <b>1,636.93</b> | 1,935.19 |
| 3-months         | 1,895.85        | 1,955.72        | <b>1,669.06</b> | 1,999.18 |
| 4-months         | 1,808.90        | 1,778.09        | <b>1,722.89</b> | 2,043.26 |
| 5-months         | 1,803.76        | 1,820.68        | <b>1,726.69</b> | 2,087.26 |
| 6-months         | <b>1,785.07</b> | 1,872.53        | 1,887.68        | 2,140.95 |
| 7-months         | 2,067.32        | <b>1,968.04</b> | 2,125.34        | 2,313.06 |
| 8-months         | 2,092.85        | <b>1,939.43</b> | 2,038.74        | 2,248.73 |
| 9-months         | 2,242.76        | <b>2,021.24</b> | 2,120.01        | 2,291.26 |
| 10-months        | <b>2,138.72</b> | 2,232.01        | 2,342.59        | 2,547.03 |
| 11-months        | 2,377.19        | <b>2,363.90</b> | 2,405.85        | 2,680.01 |
| 12-months        | 2,349.52        | <b>2,316.47</b> | 2,378.72        | 2,614.66 |

As can be seen in Table 3, when comparing AHW with MHW, AHW consistently outperforms MHW and thus the seasonal variations could be considered additive. Moreover, when comparing LASSO, SARIMA and AHW, their model performances seem more competitive as no model consistently outperforms the other. In particular, AHW seems to predict more accurate on the shorter horizons, whereas on the longer horizons, SARIMA seems to predict more accurate. Thus, despite the fact that LASSO uses information from external indicators, forecasting performance has not improved compared to traditional time series forecasting methods. Accordingly, Section 4 elaborates on two experiments that have been conducted in order to explore, investigate and analyze whether forecasting performance can be enhanced by applying efficient tuning parameter selection or forecast combination.

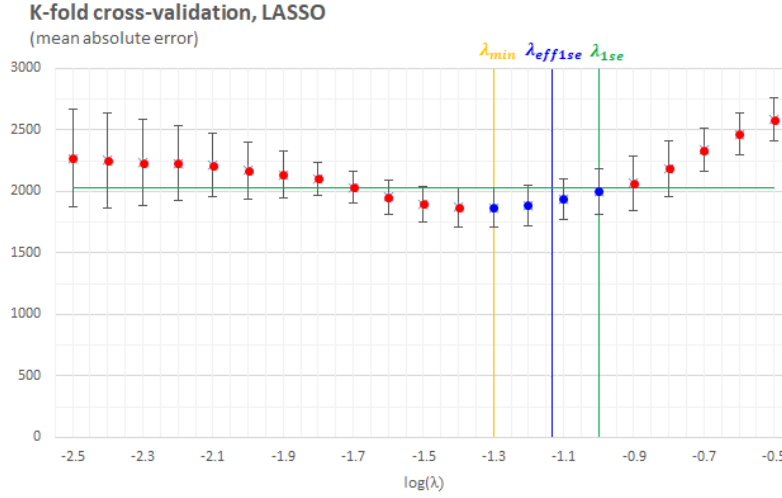
#### 4.2 Effect of Efficient Tuning Parameter Selection

With regard to the case study, the commonly used one-standard error rule was used for choosing the value of  $\lambda$  [2, 6]. Jung [20] stated that tuning parameter selection is often one of the crucial parts in high-dimensional modeling and hence using CV to select a single value as optimal value for the tuning parameter can be unstable due to the sampling variation. A possible solution to account for these sampling variations is to apply repeated CV. Nevertheless, applying repeated CV significantly increases computational costs when predicting multi-steps ahead and as a result Jung [20] proposed the use of efficient CV. Efficient CV selects multiple candidates of parameter values and calculates an average based on different weights depending on their performance without significant additional computational costs. As a criterion to select  $C$  candidates, Jung [20] opts to select the top  $C$  best performing parameter values. This experiment explores and analyzes an extension that combines efficient CV with the one-standard error rule. Thus, instead of choosing the top  $C$  best performing parameter values as candidates, all parameter values which are considered

by the one-standard error rule  $\{\lambda_{min}, \dots, \lambda_{1se}\}$  are selected as candidates. The combination of efficient CV with the one-standard error rule will be further denoted as the efficient one-standard error rule. The efficient one-standard error rule will calculate a weighted average of all candidates with different weights depending on the CV error estimates as proposed by Jung [20]. The estimates of the weights are designed in such a way that candidate values with lower CV errors are assigned a greater weight. Additionally, the weights are normalized and thus the weights of all candidate models add up to 1 [20]. The tuning parameter corresponding to the efficient one-standard error rule is obtained by:

$$\hat{\lambda}_{eff1se} = \sum_{c=1}^C w_c \lambda_c \quad \text{with} \quad w_c = \frac{\left(\frac{1}{CV(\lambda_c)}\right)}{\sum_{c=1}^C \left(\frac{1}{CV(\lambda_c)}\right)} \quad (2)$$

Figure 6 shows an example of which parameter values are selected as candidate values by the efficient one-standard error rule. As can be seen in Figure 6, the



**Fig. 6.** Tuning parameter values  $\lambda_{min}$ ,  $\lambda_{1se}$  and  $\lambda_{eff1se}$ .

efficient one-standard error rule selects a total number of 4 candidate models:  $\log(\lambda) = -1.3$ ,  $\log(\lambda) = -1.2$ ,  $\log(\lambda) = -1.1$ ,  $\log(\lambda) = -1.0$ . After determining the weights  $w_c$ , the efficient one-standard error tuning parameter is calculated at  $\log(\lambda_{eff1se}) = -1.14$ . It should be noted that the value of  $\lambda_{eff1se}$  is not a value on the grid used for the parameter search. Hence, the efficient one-standard error rule is capable of finding parameter values on a finer grid without any additional

computational costs [20]. Accordingly, Table 4 presents an overview of the model performances when using both the default and the efficient one-standard error rule. As can be seen in Table 4, the proposed efficient one-standard error rule

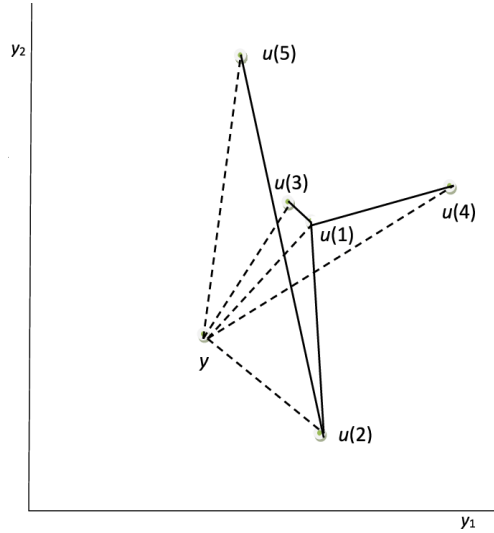
**Table 4.** Mean absolute prediction error when using the default and efficient one-standard error rule.

|            | <b>Model One-standard error rule (<math>\lambda_{lse}</math>)</b> | <b>Efficient one-standard error rule (<math>\lambda_{efflse}</math>)</b> |
|------------|---|--|
| $M_{t+1}$  | 1,865.93  | <b>1,855.95</b>  |
| $M_{t+2}$  | 1,883.23  | <b>1,855.71</b>  |
| $M_{t+3}$  | 1,895.85  | <b>1,891.97</b>  |
| $M_{t+4}$  | 1,808.90  | <b>1,788.00</b>  |
| $M_{t+5}$  | 1,803.76  | <b>1,786.83</b>  |
| $M_{t+6}$  | <b>1,785.07</b>   | 1,818.96   |
| $M_{t+7}$  | 2,067.32  | <b>2,063.71</b>  |
| $M_{t+8}$  | 2,092.85  | <b>2,082.58</b>  |
| $M_{t+9}$  | 2,242.76  | <b>2,190.58</b>  |
| $M_{t+10}$ | <b>2,138.72</b>   | 2,150.93   |
| $M_{t+11}$ | 2,377.19  | <b>2,335.49</b>  |
| $M_{t+12}$ | 2,349.52  | <b>2,297.18</b>  |

outperforms the default one-standard error rule for 10 out of 12 models, with an average improvement of 0.73%. Hence, choosing multiple candidate values in order to reduce the influence of sampling variation on the tuning parameter value, instead of choosing one optimal value, does seem to cause improvements in both the tuning parameter selection process and forecasting performance.

### 4.3 Effect of Forecast Combination

The case study has shown that LASSO did not outperform traditional time series forecasting methods, whereas the studies of Sagaert et al. [2] and Verstraete et al. [6] reported forecasting performance losses on the longer horizons compared to traditional methods. It should be noted that these studies solely compared the forecasting performance of individual models. Bates and Granger [21] noted that combining sets of forecasts can lead to improvements if each set contains independent information. Moreover, Bates and Granger [21] indicated that this independent information could be of two types: (1) forecasts are based on variables or information that other forecasts have not considered, and (2) forecasts make different assumptions about the form of relationships between variables. For clarity, Figure 7 illustrates whenever forecast combinations are superior to individual forecasts. As can be seen in Figure 7, forecasts  $u(1)$  and  $u(3)$  are highly correlated and therefore combining these forecasts will not improve the forecasting performance significantly. Moreover, forecast  $u(5)$  is a considerably poor forecast as the distance to  $y$  is large. However, combining forecasts  $u(5)$



**Fig. 7.** Forecast combinations considering five forecast vectors  $u(1)$ ,  $u(2)$ ,  $u(3)$ ,  $u(4)$  and  $u(5)$  and two steps ahead  $y_1$  and  $y_2$ . The solid lines represent the forecast combination in pairs of two, whereas the dotted lines to  $y$  represent the corresponding error of the forecast vectors [22].

with  $u(2)$  will improve the forecasting performance significantly as the distance between  $y$  and the solid line between  $u(5)$  with  $u(2)$  is reduced. Clearly, the performance improvement is due to the diversity of both models [22]. With regard to LASSO, SARIMA and AHW, each individual model creates forecasts based on independent information. For example, LASSO extracts information from leading indicators, whereas SARIMA extracts information from autocorrelations and AHW extracts information from level, trend and seasonal variations. Hence, in order to quantitatively assess their model diversities, correlation between the individual forecast errors are presented in Table 5. As can be seen in

**Table 5.** Correlation of individual forecast errors.

|        | LASSO AHW SARIMA |       |   |
|--------|------------------|-------|---|
| LASSO  | 1                | -     | - |
| AHW    | 0.758            | 1     | - |
| SARIMA | 0.825            | 0.944 | 1 |

Table 5, the least correlation exists between the forecast errors of LASSO and AHW. Thus, combining the individual forecasts of LASSO and AHW will have the highest potential for enhanced forecasting performance. In order to obtain combined forecasts, weights must be allocated to the individual forecasts. Ac-

cordingly, Bates and Granger [21] introduced numerous methods for determining the weights of each individual forecast as it is preferred to assign a greater weight to an individual forecast with higher accuracy. However, Armstrong [23] mentioned that applying weights is only beneficial if there is strong evidence that particular forecasting models are likely to predict better than others. Otherwise, the use of equal weights is likely to perform better under almost all other circumstances [24]. In our case study, there is no strong evidence that LASSO outperforms AHW or vice versa and thus the decision was made to allocate equal weights to the individual forecasts, i.e. the individual forecasts of LASSO and AHW are averaged. Overall, Table 6 shows the forecasting performance of the combined LASSO and AHW forecasts (LASSO-AHW) in comparison to all other individual models. As can be seen in Table 6, after combining the individual

**Table 6.** Mean absolute prediction error across all forecast horizons.

| Forecast horizon | LASSO           | SARIMA          | AHW      | LASSO-AHW       |
|------------------|-----------------|-----------------|----------|-----------------|
| 1-month          | 1,855.95        | 1,876.55        | 1,732.56 | <b>1,618.09</b> |
| 2-months         | 1,855.71        | 1,778.85        | 1,636.93 | <b>1,581.07</b> |
| 3-months         | 1,891.97        | 1,955.72        | 1,669.06 | <b>1,650.04</b> |
| 4-months         | 1,788.00        | 1,778.09        | 1,722.89 | <b>1,648.04</b> |
| 5-months         | 1,786.83        | 1,820.68        | 1,726.69 | <b>1,642.26</b> |
| 6-months         | 1,818.96        | 1,872.53        | 1,887.68 | <b>1,736.84</b> |
| 7-months         | 2,063.71        | 1,968.04        | 2,125.34 | <b>1,928.84</b> |
| 8-months         | 2,082.58        | 1,939.43        | 2,038.74 | <b>1,939.34</b> |
| 9-months         | 2,190.58        | <b>2,021.24</b> | 2,120.01 | 2,047.29        |
| 10-months        | 2,150.93        | 2,232.01        | 2,342.59 | <b>2,120.66</b> |
| 11-months        | 2,335.49        | 2,363.90        | 2,405.85 | <b>2,279.48</b> |
| 12-months        | <b>2,297.18</b> | 2,316.47        | 2,378.72 | 2,316.44        |

forecasts of LASSO and AHW, LASSO-AHW outperforms all other individual models for almost all forecast horizons. Thus, it seems that both the LASSO and AHW models are so diverse, that combining the predictions of these models results into enhanced forecasting performance, with an average improvement of 2.38%. Apparently, LASSO extracted valuable information from leading indicators, whereas Holt-Winter extracted valuable information from level, trend and seasonal variations, and ultimately, combining all of this information resulted in forecasting performance improvements. Hence, with regard to this case study, the inclusion of information extracted from leading indicators actually did lead to more accurate parts sales predictions.

## 5 Conclusion

This paper aimed to identify leading indicators for a case company that supplies truck parts to the European truck aftersales market. Next to that, it was

explored whether including leading context information leads to more accurate predictions in comparison with traditional time series methods, often used in businesses. LASSO was used to extract relevant information from a collected pool of business, economic, and market indicators. It was found that combining predictions of LASSO and the traditional Holt-Winters method yielded the best forecasting performance, outperforming all other considered Holt-Winters and SARIMA models. Thus, including leading context information did improve forecasting performance for the case company. Also, due to the transparency of LASSO, using LASSO provided business intelligence about relevant predictors and any lead effects. Finally, from a pool of 34 indicators, 7 indicators appeared to have clear lead effects for the case company. The exact indicators and lead effects are not revealed due to confidentiality. Additionally, this research has several contributions to the existing academic literature. First of all, we proposed the efficient one-standard error rule, as an alternative to the default one-standard error rule, by combining efficient CV, proposed in Jung [20], with the commonly used one-standard error rule, described in Hastie et al. [16]. The purpose of the efficient one-standard error rule is to reduce the influence of sampling variation on the actual tuning parameter value. As stated earlier, we found that applying the efficient one-standard error rule over the default one, improved forecasting performance in 10 out of 12 models. With regard to future research purposes, there is a need to explore and analyze whether the efficient one-standard error rule improves performance, compared to the default one-standard error rule, when applied on multiple and larger data sets. Secondly, the studies of Sagaert et al. [13] and Verstraete et al. [6] reported forecasting performance losses of LASSO on the longer horizons, compared to traditional methods. These studies solely considered and compared the forecasting performance of individual models, whereas we found that, for our case study, applying forecast combination resulted in improved forecasting performance over almost all horizons.

## References

1. Currie, C. S., & Rowley, I. T. (2010). Consumer behaviour and sales forecast accuracy: What's going on and how should revenue managers respond. *Journal of Revenue and Pricing Management*, 9(4), 374–376. doi:10.1057/rpm.2010.22
2. Sagaert, Y. R., Aghezzaf, E. H., Kourentzes, N., & Desmet, B. (2018). Tactical sales forecasting using a very large set of macroeconomic indicators. *European Journal of Operational Research*, 264(2), 558–569. doi:10.1016/j.ejor.2017.06.054
3. Williams, B. D., Waller, M. A., Ahire, S., & Ferrier, G. D. (2014). Predicting retailer orders with POS and order data: The inventory balance effect. *European Journal of Operational Research*, 232(3), 593–600. doi:10.1016/j.ejor.2013.07.016
4. Ma, S., Fildes, R., & Huang, T. (2016). Demand forecasting with high dimensional data: The case of SKU retail sales forecasting with intra- and inter-category promotional information. *European Journal of Operational Research*, 249(1), 245–257. doi:10.1016/j.ejor.2015.08.029
5. Fildes, R., Goodwin, P., Lawrence, M., & Nikolopoulos, K. (2009). Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for im-



- provement in supply-chain planning. *International Journal of Forecasting*, 25(1), 3–23. doi:10.1016/j.ijforecast.2008.11.010
6. Verstraete, G., Aghezzaf, E. H., & Desmet, B. (2020). A leading macroeconomic indicators' based framework to automatically generate tactical sales forecasts. *Computers and Industrial Engineering*, 139(August 2019), 106169. doi:10.1016/j.cie.2019.106169
  7. Atsalakis, G. S., Protopapadakis, E. E., & Valavanis, K. P. (2016). Stock trend forecasting in turbulent market periods using neuro-fuzzy systems. *Operational Research*, 16(2), 245–269. doi:10.1007/s12351-015-0197-6
  8. RESHADAT, V., HOORALI, M., & FAILI, H. (2016). A Hybrid Method for Open Information Extraction Based on Shallow and Deep Linguistic Analysis. *Interdisciplinary Information Sciences*, 22(1), 87–100. doi:10.4036/iis.2016.R.03
  9. Nourani, E., & Reshadat, V. (2020). Association extraction from biomedical literature based on representation and transfer learning. *Journal of theoretical biology*, 488(), 110112. doi:10.1016/j.jtbi.2019.110112
  10. Reshadat, V., & Faily, H. (2019) A new open information extraction system using sentence difficulty estimation. *Computing and Informatics*, 38(4), 986–1008. doi:10.31577/cai\_2019\_4\_986
  11. Reshadat, V., & Feizi-Derakhshi, M. R. (2012). Studying of semantic similarity methods in ontology. *Research Journal of Applied Sciences, Engineering and Technology*, 4(12), 1815–1821.
  12. Collopy, F., & Armstrong, J. S. (1992). Rule-based forecasting: Development and validation of an expert systems approach to combining time series extrapolations. *Management science*, 38(10), 1394–1414.
  13. Sagaert, Y. R., Aghezzaf, E. H., Kourentzes, N., & Desmet, B. (2017). Temporal big data for tactical sales forecasting in the tire industry. *Interfaces*, 48(2), 121–129. doi:10.1287/inte.2017.0901
  14. Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A Seasonal-Trend Decomposition Procedure Based on Loess. *Journal of Official Statistics*, 6(1), 3–73.
  15. Xiong, T., Li, C., & Bao, Y. (2018). Seasonal forecasting of agricultural commodity price using a hybrid STL and ELM method: Evidence from the vegetable market in China. *Neurocomputing*, 275, 2831–2844. doi:10.1016/j.neucom.2017.11.053
  16. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. *Data Mining, Inference, and Prediction*. (Second ed.). Springer.
  17. Bergmeir, C., & Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences*, 191, 192–213. doi:10.1016/j.ins.2011.12.028
  18. Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1), 267–288.
  19. Hyndman, R., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2nd ed.). Australia: OTexts.
  20. Jung, Y. (2016). Efficient Tuning Parameter Selection by Cross-Validated Score in High Dimensional Models. *World Academy of Science, Engineering and Technology*, 10.
  21. Bates, A. J. M., & Granger, C. W. J. (1969). The Combination of Forecasts. 20(4), 451–468. doi:10.2307/3008764
  22. Atiya, A. F. (2020). Why does forecast combination work so well? *International Journal of Forecasting*, 36(1), 197–200. doi:10.1016/j.ijforecast.2019.03.010
  23. Armstrong, J. S. (2001). Combining Forecasts. *Principles of forecasting: a handbook for researchers and practitioners*, 417–439. doi:10.4018/jncr.2012070103

24. Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, 5(4), 559–583. doi:10.1016/0169-2070(89)90012-5