Computers & Industrial Engineering 139 (2020) 106169

Contents lists available at ScienceDirect

Computers & Industrial Engineering

journal homepage: www.elsevier.com/locate/caie

A leading macroeconomic indicators' based framework to automatically generate tactical sales forecasts

Gylian Verstraete[a,c,*], El-Houssaine Aghezzaf[a,d], Bram Desmet[b,c]

[a] Department of Industrial Systems Engineering and Product Design, Ghent University, Technologiepark 903, 9052 Zwijnaarde (Gent), Belgium
[b] Operations & Supply Chain Management, Vlerick Business School, Reep 1, 9000 Gent, Belgium
[c] Solventure, Sluisweg 1 Bus 18, 9000 Gent, Belgium
[d] Flanders Make, Belgium

ARTICLE INFO

Keywords:
Forecasting
Sales forecasting
Tactical sales forecasting
Macroeconomic indicators
Decomposition
LASSO regression

ABSTRACT

Tactical sales forecasting is fundamental to production, transportation and personnel decisions at all levels of a supply chain. Traditional forecasting methods extrapolate historical sales information to predict future sales. As a result, these methods are not capable of anticipating macroeconomic changes in the business environment that often have a significant impact on the demand. To account for these macroeconomic changes, companies adjust either their statistical forecast manually or rely on an expert forecast. However, both approaches are notoriously biased and expensive. This paper investigates the use of leading macroeconomic indicators in the tactical sales forecasting process. A forecasting framework is established that automatically selects the relevant variables and predicts future sales. Next, the seasonal component is predicted by the seasonal naive method and the long-term trend using a LASSO regression method with macroeconomic indicators, while keeping the size of the indicator's set as small as possible. Finally, the accuracy of the proposed framework is evaluated by quantifying the impact of each individual component. The carried out analysis has shown that the proposed framework achieves a reduction of 54.5% in mean absolute percentage error when compared to the naive forecasting method. Moreover, compared to the best performing conventional methods, a reduction of 25.6% is achieved in the tactical time window over three different real-life case studies from different geographical areas.

1. Introduction

Forecasting is one of the key aspects of operations management (Oliva & Watson, 2009). Sales forecasting plays a major role in the allocation of corporate resources (Stein, 1997), marketing (Crittenden, Gardiner, & Stam, 1993), and impacts decisions on production, transportation and personnel at all kinds of horizons in the supply chain (Hyndman & Athanasopoulos, 2014). Historically, forecasting research attempted to find the best model for the used data set (De Gooijer & Hyndman, 2006). With the rapid expansion of the internet, a lot of external data has become available. IBM estimates that in 2020 43 trillion GB of data will be created, which is 300 times the volume produced in 2008. This growth in data availability is causing a shift from finding the best model to finding the right data (causal method forecasting).

Traditional statistical forecasting methods only extrapolate historical trends and seasonal influences to predict future sales. As a consequence, these methods are not capable of anticipating macroeconomic changes in the business environment, which often significantly impact the demand. To account for these future changes, companies either adjust their statistical forecast manually or rely on expert forecasts. However, both approaches are notoriously biased, as humans are generally bad in making these adjustments, and are time consuming.

Back in 1988, studies stressed the need for research on multivariate methods (Chatfield, 1988; Ord, 1988). Interestingly, 18 years later, De Gooijer and Hyndman (2006) stated that multivariate time series forecasting was still not widely empirically investigated, citing easy to use software as suspected reasons. Nowadays the widespread availability of data and statistical software has made forecasting with multivariate models more common. This has resulted in several studies applying multivariate techniques with a wide variety of independent variables (e.g. installed base information in Kim, Dekker, & Heij (2017), macro-economic variables in Li & Chen (2014), Kim & Swanson (2016),

* Corresponding author at: Department of Industrial Systems Engineering and Product Design, Ghent University, Technologiepark 903, 9052 Zwijnaarde (Gent), Belgium.

E-mail addresses: Gylian.Verstraete@UGent.be (G. Verstraete), ElHoussaine.Aghezzaf@UGent.be (E.-H. Aghezzaf), Bram_Desmet@solventure.eu (B. Desmet).

https://doi.org/10.1016/j.cie.2019.106169
Received 5 March 2019; Received in revised form 28 August 2019; Accepted 8 November 2019
Available online 13 November 2019

Ding (2018), Xia, Wang, & Liu (2019), promotional data in Guidolina, Guseoa, & Mortarinoa (2019), weather in Verstraete, Aghezzaf, & Desmet (2019), social media in Cui, Gallino, Moreno, & Zhang (2017), customer information in Murray, Agard, & Barajas (2018), the historical sales pattern Yan & Tu (2012) and a combination of different types in Qu et al. (2017), Yan, Wu, & Tu (2013), Yan, Tu, & Wu (2018)). For tactical sales forecasting in business to business environments using macroeconomic indicators few studies have been done. At the time of writing, there is only one framework described in Sagaert, Aghezzaf, Kourentzes, and Desmet (2017, 2018), which combines seasonal dummy variables, autoregressed sales and macroeconomic variables in a LASSO regression model to predict tactical sales. This is in sharp contrast with the known advantages of an improved tactical sales forecasting process on other parts of supply chain management. An overview of the advantages of improving sales forecast accuracy can be found in Sagaert, Kourentzes, De Vuyst, Aghezzaf, and Desmet (2019), Babai, Ali, Boylan, and Syntetos (2013), Fildes and Kingsman (2011), Özer and Wei (2004) and Wang and Petropoulos (2016).

Using leading indicators in sales forecasting is not completely new. Surveys (Dalrymple, 1987; Klassen & Flores, 2001) report that respectively 36% and 44% of industrial firms are currently including leading indicators in their forecast process. Another survey (Sanders & Manrodt, 2003) shows that up to 65% of respondent companies think that macroeconomic indicators are important factors for the forecast process. Based on our experiences, we assume that currently macroeconomic indicators are mainly used as a qualitative aid in judgmental forecasts or adjustments of statistical baselines.

In this paper, we propose a methodology that automatically generates tactical sales forecasts based on large groups of macroeconomic indicators. We evaluate the framework by quantifying the accuracy influence of the individual components. We analyze the use of (1) the statistical model, (2) the indicator data set size, (3) the sales decomposition and macroeconomic indicators preprocessing by STL-decomposition and principal component analysis. For these components, we quantify the implemented techniques by comparing the out-of-sample forecast accuracy, using the mean absolute percentage of error. TThe remainder of the paper is organized as follows: Section 2 is dedicated to reviewing previous literature related to forecasting using macroeconomic variables. In Section 3 presents the proposed methodology and the framework. Section 4 discusses the used data and provides some insights on the accuracy impact of the framework. Section 5 presents the major conclusions.

2. Literature review

The use of macroeconomic indicators for sales forecasting in a tactical time window presents two main challenges. The first challenge, which is a common challenge in sales forecasting, is the limited sample size of the dependent variables. Most companies cannot easily access historical data. Even if companies have the required history of data available, it is often not representative anymore due to changes in product portfolio and changes in consumption pattern of consumers. The amount of data is further abridged because macroeconomic data is typically available on a monthly basis or at higher aggregation levels.

The second challenge originates from the large quantity of available macroeconomic indicators. Examples of publicly available data sources are the Federal Reserve of Economic Data (FRED), the Eurostat and Organization for Economic Cooperation and Development (OECD). All together, these data sources represent over a million macroeconomic time series. Selecting variables from such data sets is an open research topic. The combination of both challenges results in the so-called 'short and fat data problem': there are few observations for a large number of independent variables. Traditional regression techniques, such as ordinary least squares regression, are known to perform poorly on these problems. To overcome this shortcoming, Sagaert et al. (2018) suggests the use of LASSO regression models. Strong similarities with the second challenge can be found in the literature on forecasting macro-economic variables. Therefore, we draw heavily from this topic, where techniques such as LASSO, dynamic factor models and decomposition are widely used.

LASSO regression (Tibshirani, 1996) is a technique that penalizes the absolute value of the coefficients of variables. This penalization forces many coefficients to be equal to exactly zero and hence produces sparse and interpretable models. LASSO has been applied successfully in various research areas. For example, Li and Chen (2014) used LASSO regression to forecast 20 macroeconomic time series using a publicly available data set containing 107 macroeconomic indicators. The study showed that LASSO regression outperformed dynamic factor models for most of the variables under investigation. Ludwig, Feuerriegel, and Neumann (2015) benchmarked LASSO regression to ARMA models to forecast energy prices. The authors achieved a 17% improvement in forecast accuracy. Plakandaras, Gogas, Papadimitriou, and Gupta (2017) investigated the prediction of the US inflation using macroeconomic indicators. The authors benchmarked the LASSO on (nonlinear) machine learning approaches but found no evidence of decreased performance due to being limited to linear relations with LASSO. Smeekes and Wijler (2018) demonstrated the applicability of penalized regression methods on a large dataset with US based macroeconomic indicators. Uematsu and Tanaka (2019) reached similar conclusions in forecasting GDP in quarterly time buckets using a high-dimensional monthly data set. Panagiotelis, Athanasopoulos, Hyndman, Jiang, and Vahid (2019) did a similar analysis on Australian macroeconomic variables. The study showed that Australian key macroeconomic data are more difficult to predict. Futhermore, the study states there is no evidence that an indicator set over 20–40 variables leads to increased prediction accuracy.

Another approach in literature on predicting retail sales and macroeconomic variables is dynamic factor models. Dynamic factor models are introduced by Stock and Watson (2002a) and tested by forecasting eight monthly U.S. macroeconomic time series using 215 predictors (Stock & Watson, 2002b). The authors found that just six factors account for most of the variance of the 215 predictors, and few of those factors are required to forecast accurately. Li and Chen (2014) benchmarked the dynamic factor models to common penalized regression techniques (such as LASSO) by forecasting macroeconomic variables. The authors discovered that penalized regression models provide better accuracy, but suggest that LASSO could be used as an additional information source to dynamic factor models. Kim and Swanson (2016) tested factor based models by performing a forecast "horse race". In this race, 11 macroeconomic variables are forecasted using 144 other macroeconomic indicators. Factor based models often were the best performing model, suggesting their usefulness in forecasting. The authors suggest that dimension reduction by using factors, but also by using machine learning and shrinkage methods, are useful for forecasting macroeconomic variables. Boivin and Ng (2006) investigated the effect of the size of the data set on the forecast accuracy. They showed that reducing the data set from 147 to 40 prescreened variables generally leads to equally good or better results. Kim and Swanson (2018) combined dynamic factor models with the Mixed Data Sampling (MIDAS) framework which allows factors and variables of different frequencies to be included in the MIDAS framework to predict the Korean GDP. The authors determined that very sparse models (1 or 2 used factors) are most useful when the uncertainty is highest. Cepni, Güney, and Swanson (2019) confirmed that dynamic factor models yield good prediction results in nowcasting and forecasting GDP.

Forecasting by decomposition has been proven to predict accurately in various situations. An example is the theta method (Assimakopoulos & Nikolopoulos, 2000) performing well in the M3 competition. Literature has shown that decomposition can fit in a forecasting framework by using other techniques to predict the sub components. Theodosiou (2011) applied decomposition on the NN3 and M1 competition time series outperforming the standard statistical forecasting methods. The proposed framework consists of decomposing a time series into seasonal, trend and error components using Seasonal and Trend decomposition using Loess (STL decomposition) and predicting each component separately using traditional statistical methods. The author shows that the decomposition method results in greater relative improvements in accuracy as the forecast window increases. Sakai, Nakajima, Higashihara, Yasuda, and Oosumi (1999) applied decomposition to split vending machine sales into a macro (trend) and micro (time series) factor. The authors then proceed to predict each component separately. Bergmeir, Hyndman, and Benítez (2016) used STL decomposition to allow bootstrapped aggregated forecasts with traditional methods by bagging the error component of the decomposition. Xiong, Li, and Bao (2018) used STL decomposition to decompose vegetable prices. The authors proceed by predicting each component independently using the seasonal naive method for the seasonality component and extreme learning machines (ELMs) for the trend and the error components. These contributions all assume the decomposed components are independent and thus are influenced by different types of factors. One of those factors for the tactical window in sales forecasting are macroeconomic variables.

Aforementioned techniques are known to improve forecasting accuracy separately. Our proposed framework combines the techniques to predict sales in a widely applicable way. The decomposition method splits the time series in a long-term trend, a seasonal and a noise pattern. We propose to forecast the long-term trend using macroeconomic indicators under the assumption that this information drives the long-term trend of the sales. For the seasonal and noise signals other techniques and data can be used if applicable. Previous studies in forecasting with macroeconomic data did not tackle this problem. Advantages of this new approach is that the macroeconomic trend can be forecasted without the requirement of taking into account other types of variables (in case they are unknown or have no predictive value on the tactical sales level). Moreover, interference between the macroeconomic and other types of variables is automatically handled by the decomposition.

3. Proposed framework

In this section, we describe the proposed forecasting framework to automatically predict sales in the tactical time frame using (leading) macroeconomic indicators. The framework consists of several methods that are known to improve accuracy separately, and are combined in a way that exploits the structure of the data. Fig. 1 offers an overview of the proposed framework. The different steps of the framework are clarified in this section.

3.1. Sales data transformation

The first step is to decompose the sales data into independant signals. In the proposed framework, we use the popular STL decomposition (proposed by Cleveland, Cleveland, McRae, & Terpenning (1990)). STL decomposition disaggregates a time series x_t into three components, namely the trend (m_t), the seasonal (s_t) and error/remainder (e_t) components, so that $x_t = m_t + s_t + e_t$ for all periods t. STL decomposition works iteratively by applying two loops. Schematically, the inner loop includes six steps for the (k + 1) th iteration of m_t^{k+1} and s_t^{k+1}:

1. Detrending: the original series x_t is detrended with the estimated trend component m_t^k obtained at the kth pass $x_t^{detrend} = x_t - m_t^k$.
2. Cycle sub series smoothing: the temporary seasonal component $\hat{s}_t^{k+1}$ is obtained by applying a LOESS smoother to the sub-cycle series $x_t^{detrend}$.
3. Low-pass filtering of the smoothed cycle sub series: $\hat{s}_t^{k+1}$ from step 2 is processed using a low-pass filter, followed by a LOESS smoother, in order to identify the remaining trend $\hat{m}_t^{k+1}$.
4. Detrending of the smoothed cycle sub series: the additive seasonal component s_t^{k+1} is computed by subtracting the low-pass values from the temporary seasonal component: $s_t^{k+1} = \hat{s}_t^{k+1} - \hat{m}_t^{k+1}$.
5. Deseasonalizing: The additive seasonal component s_t^{k+1} is subtracted from the original time series x_t to obtain a seasonally adjusted series $x_t^{deseason}$.
6. Trend smoothing: the seasonally adjusted series $x_t^{deseason}$ is smoothed by a LOESS smoother to obtain the trend component m_t^{k+1}.

Afterwards the outer loop will calculate the remainder for iteration k + 1 using . Then, the inner loop is repeated with the remainder e_t^{k+1} as an outlier indication, scaling down the influence of these potential outliers. This procedure is repeated until the predetermined maximum amount of iterations is reached.

We choose STL decomposition because of its robustness to both outliers and in the endpoints of the time series, and its versatility, as seasonal components do not need to be constant over different time periods and smoothness of the trend is controllable. For this framework, we argue to use constant seasonality in order to achieve robust estimates. For the smoothness of the trend, we use the default rule proposed by Cleveland et al. (1990).

3.2. Indicator data transformation

On the indicator data, we perform standardization of the variables (centering by subtracting the mean and scaling by dividing with the standard deviation). Then, we potentially combine two data transformations. First, we apply principal component analysis to the data, which filters the main indicator movements from the data set. Principal component analysis (PCA) is a statistical technique that is used to transform a group of (potentially) correlated variables into a smaller group of principal components. The technique performs an orthogonal (linear) transformation of the data. The components are created so that the first contains the largest amount of variance of the variable set. As each of the following components has a smaller explained variance, data reduction can be achieved by selecting the first number of components. The explained variance for each component is conveniently represented by the eigenvalues of the covariance matrix of the data. We note that principal component analysis requires a standardized variable set. Without standardization, the eigenvectors will all be of different lengths. The resulting principal components will be dominated by the variables with the highest variance (which is a scale dependent metric). As this is an unwanted feature (we want variables to be grouped independent of the scale) we standardize the variables. The factors are calculated as:

$$X_t = \Lambda_t F_t + e_t$$

With X_t the matrix of predictor variables, F_t the matrix of dynamic factors, Λ_t the factor loadings. The factor loadings can be seen as the rotation of the variable space that allows each component the highest possible remaining explained variance. e_t are the idiosyncratic errors, which arise from measurement errors or special features (e.g. local shocks) in individual series of the predictor variables. The number of extracted factors can be determined by:

- ordering the factors on explained variance and (visually) assess when the marginal contribution of the cumulative explained variance drops.
- using an information criteria, such as the Bai and Ng information criterion (Bai & Ng, 2002).
- letting the described statistical learning models determine which factors to include.

Furthermore, we apply STL decomposition to smooth the data. Similarly to the sales transformation, we only use the trend component of the indicator in the data set. This results in four different indicator transformations being tested: using no transformation at all, using the

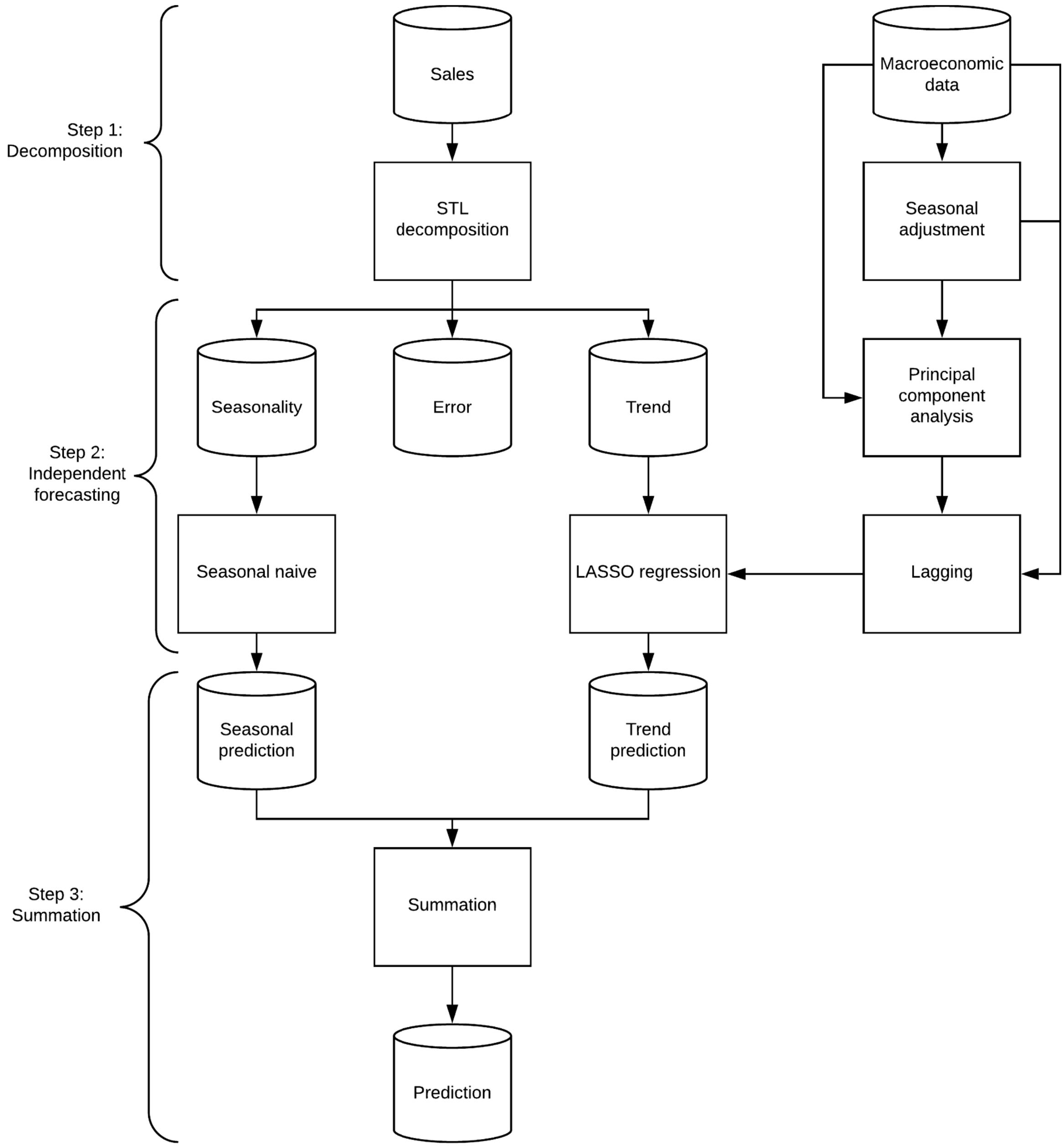

Fig. 1. Overview of the forecasting framework.

STL decomposed trend of the indicators, using the principal components of the indicators and using the combination.

Finally, we lag the data in time (time shifting). By shifting the indicators forward in time (as we are only interested in leading variables) between a predetermined number of variable lags, the number of variables used in the analysis is multiplied by the number of lags.

3.3. Statistical methods for predicting sales

In this section, we describe the methods used for predicting the sales. In Section 3.3.1 we handle the seasonal naive method that is used to predict the seasonal component. Afterwards, in Sections 3.3.2 and 3.3.3, we discuss the ordinary least squares method and its LASSO regression extension we use to predict the macroeconomic trend.

3.3.1. Seasonal naive method

An extension of the standard naive method that is useful for highly seasonal data. The seasonal naive method sets each forecast to be equal to the last observed value of that seasonal time window. In the case of monthly time windows this is the last available observation of that month. Mathematically the method is written as

$$y_{t+h} = y_{t+h-km}$$

with m the seasonal period and $k = \lfloor (h-1)/m \rfloor + 1$, the factor allowing to predict more than one seasonal cycle in the future.

The seasonal naive method is used to predict the seasonal decomposed pattern and will merely extrapolate the last seasonal observation. The seasonal window parameter of the decomposition will determine how much emphasis is placed on recent observations.

3.3.2. Ordinary least squares linear regression

Linear regression is a statistical method of the form $\hat{y}_{t+h} = \beta_0 + \beta_j x_{t+h,j}$. The ordinary least squares (OLS) estimation will minimize the sum squared deviance between the observed dependent variable in the given data set and those predicted by the independent variables by the linear function. The OLS estimation does not perform variable selection; the method will estimate all parameters β and thus use all given variables.

While in general the ordinary least squares method works well, it has some limitations in predictions using a large set of indicators:

- Causality: the method will attempt to create a best fit for a given set of data using all the variables. However, this says nothing about the real influence between the observed and the dependent variable(s). The method does not determine whether there is an influence between those variables.
- Multicollinearity: When collinear variables are introduced to an ordinary least squares method, the coefficients will be incorrectly estimated.
- Overfitting: The method will select as many variables as possible to minimize the in-sample error. This can easily lead to overfitting issues.

The stepwise regression method extends OLS regression with a variable selection procedure, by repeatedly adding or removing the variable that improves a selection metric. However, the method is criticized as it does not guarantee selecting the best set of variables. Stepwise regression was not used because of computational limitations, as it requires to calculate the selection criterion for all variables in each selection step.

3.3.3. LASSO regression

The least absolute shrinkage and selection operator is a regression technique that performs both variable selection and fitting. LASSO complements the ordinary least squares method by penalizing the absolute value of the coefficients, which leads to the coefficients of variables with a weak influence being shrunk towards zero (and thus performing variable selection) (Tibshirani, 1996). Mathematically this equals

$$minimize \sum_t (y_t - \beta_0 - \sum_j \beta_j x_{tj})^2 + \lambda \sum_j |\beta_j|, \quad j > 0.$$

With λ being the penalty factor that is determined by performing a cross-validation on the training set. Literature suggests two values for λ:

- The value of λ that has the smallest error (λ_{min})
- The parsimonious method whose error is no more than one standard error above the error of the smallest error λ_{min} (λ_{1se}).

Following the inventors' guidelines (Hastie, Tibshirani, & Friedman, 2009), we chose the parsimonious λ_{1se} approach, as the risk curves are estimated with error (Friedman, Hastie, & Tibshirani, 2010) and there is a distinction between overfitting during variable selection and model fitting (Cawley & Talbot, 2010).

The method is family of the regularized regression methods. These methods include additional metrics in the variable fitting process to avoid the final model from overfitting on the training set. Also, penalization is known to solve multicollinearity issues present in indicator sets. This makes regularization an ideal method for predicting the long-term trend. LASSO regression was chosen over other regularized regression methods because the L1-penalization (penalizing the absolute value of the coefficients) creates sparser models which are more explainable to users. This type of models are known to have a better performance in selecting causal variables.

3.4. Prediction methodology

After applying the STL decomposition on the sales, we consider the extracted signals as independent and we forecast each component separately (decomposed forecast). Each of the extracted components are predicted as follows:

- we assume that the trend of the sales (m_t in the STL decomposition) is determined by the macroeconomic environment. Considering the size of macroeconomic databases, selecting the right variables for the underlying macroeconomic trend is not a trivial problem. Sagaert et al. (2018) suggests using LASSO regression to select and forecast using a large number of macroeconomic indicators. As the reasons for applying LASSO regression also address our research challenges, we opt to use this technique.
- we use the seasonal naive method to predict the seasonal component (s_t in the STL decomposition) as proposed by Xiong et al. (2018). This allows the user to adjust the STL decomposition to put more emphasis on the recent occurring seasonal patterns. The seasonal naive method will merely extrapolate the latest seasonal observation.
- we assume the error component (e_t in the STL decomposition) is driven by other factors than macroeconomic indicators. We assume the term is driven by factors such as social media, promotions and weather or is a result of random noise. However, as the predictive power of these factors is out of the tactical time window, predicting the error component is considered out-of-scope.

After predicting the individual components, the final forecast is calculated by combining the patterns:

$$\hat{x}_t = \hat{m}_t + \hat{s}_t + \hat{e}_t$$

Additional advantages of splitting the signals into different components, is that we can fit macroeconomic variables to the trend without interference of other factors that are captured in the seasonal or the noise component. Moreover, the macroeconomic trend can be predicted without the requirement of taking into account other types of variables. These variables are potentially unknown or have no predictive power in the tactical sales window.

4. Data and forecasting results

In this section, we compare the different components of our forecasting framework using real-life case studies. First, we describe the sales and indicator data used in this case study. Afterwards, we discuss the design and evaluation of the components of the framework. In Section 4.3, we analyze the different components of the prediction methodology. While validating the impact of one dimension, the effect of the other components is averaged out. Then, after determining the best configuration, we quantify the accuracy improvement of the methodology on three different cases in Section 4.3.5. In the final section, we discuss the impact one of the case studies experienced using the framework.

4.1. Used data

4.1.1. Sales data

We determine the impact of each component of the forecasting framework using sales information of a second-tier supplier to the automotive sector, that is also a first-tier supplier to the tire industry. The data consists of monthly aggregated sales information ranging from January 2012 to February 2017. This data is split into four different time series. We divide the data by geographical location (Europe and North America) and by application (passenger and truck tires).

After quantifying the impact of each component of the forecast methodology, we validate it by using two other case studies (a global producer of steel and a global producer of composite building materials). Both are splitted geographically between Europe and North America. The producer of building materials supplied sales information from January 2012 to November 2016, whereas the producer of steel supplied data from January 2010 to November 2016. This ensures that our methodology is validated on various companies in various sectors over multiple geographical locations.

Table 1
The used data sets.

Data set	# indicators	Selection method
Keyword 1	±2500	Keywords
Keyword 2	±300	Keywords
Expert Selection	±10	Manual

4.1.2. Macroeconomic indicator data

The macroeconomic indicator pool comprises all monthly aggregated time series of the Federal Reserve of Economic Data (FRED) of the Federal Reserve Bank of St. Louis (US). At the time of this study, we counted 66420 unique monthly time series. We split the data set into three different data sets based on market intelligence, with the aim of making the data sets more related to the sector of the case as the data set becomes smaller. Two of the data sets are based on keywords, and one is an expert selection of time series that are hand picked by the strategic marketing department of the companies. An overview of the different data sets is given in Table 1. The content of the data set is unavailable due to confidentiality. As we lag the indicator data in time, the actual number of variables used by the method is multiplied by the number of time lags used (here 12 lags).

4.2. Experiment design and evaluation

In this experiment, we will test how the different components of the framework influence the forecast accuracy. We investigate the influence of the following components:

- The used statistical method
- The number of variables in the indicator set
- The decomposition of the sales (sales transformation)
- The indicator processing

To validate the accuracy of the different components of the framework, we perform an out-of-sample evaluation, using a rolling origin approach with a minimal training window of 24 months, while predicting the next 12 months. Fig. 2 visualizes this approach schematically. We stop forecasting 12 months before the last available period to ensure an equal amount of observations for each future time horizon.

Then, the dimensions are assessed by performing a full factorial experiment design. The full experiment design consists of 24 unique combinations. For each of those combinations we have created a total of 108 forecasts over the four different used time series of the supplier to the automotive. This results in 2592 predictions.

To measure the forecasting accuracy, we use the mean absolute percentage of error (MAPE), which is denoted as

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\frac{|y_t - \hat{y}_t|}{y_t}$$

The advantage of this metric is that it gives a good representation of the quality of the forecast, and its impact on other aspects of the supply chain, such as safety stock. Hyndman and Koehler (2006) show that measures based on percentage errors have a disadvantage of being infinite or undefined if $y_t = 0$ for any t in the period of interest, and having skewed results when any value of y_t is close to zero. As the used data do not have these properties MAPE can be used to evaluate the proposed framework.

4.3. Results

4.3.1. Statistical model

The results of the out-of-sample MAPE test for the different statistical methods is presented in Table 2 and Fig. 3. We see the proposed LASSO model outperforming the traditional techniques from horizon 1–6. We see the traditional univariate techniques, such as ETS, ARIMA and theta, surpassing LASSO regression in the later time windows. Sagaert et al. (2018) determined similar findings, citing the performance loss due to fewer available variables in further horizons. Our study did not limit the time lags of the variables, and as we used the same data set, we conclude that this is not the case here. As indicators are selected by optimizing fit and the near future pattern will most likely not differ too much, the short term prediction of the variables will often lead to accurate predictions even if they are not causally related. However, on the further horizons, selecting non causally related indicators will lead to large prediction errors.

Furthermore, we see LASSO regression outperforming the OLS regression with one variable. This shows that using an intelligent selection and fitting technique that takes into account multiple variables adds value to the tactical forecasting process.

4.3.2. Data set size

Table 3 and Fig. 4 show the influence of the size of the data set on the accuracy of the prediction methods. The accuracy consistently increases as the used data set decreases in size. The explanation for this behaviour can be found in the data set containing a higher density of causal variables, which will lead to better predictive ability. This effect has been reported many times in literature.

Consistently with our previous analysis, the LASSO model outperforms the OLS regression method. The difference in performance increases as the indicator set becomes larger, as the OLS method selects only one variable from the large set, which results in a 'hit or miss' effect when selecting variables.

4.3.3. Sales transformation

Table 4 and Fig. 5 show that decomposing the sales to predict each component of the time series individually leads to a large improvement in prediction accuracy. Over the entire time horizon, we see a relative improvement in MAPE of 44%. This shows that decomposition is very effective in splitting the time series in different independent components, allowing independent and pattern specific modelling for each part. Additionally, it allows to specify separate variable groups for each sub signal. Table 4 confirms this, as not performing the decomposition leads to a worse forecast accuracy than using a naive method.

4.3.4. Indicator transformation

Table 5 and Fig. 6 show the influence of the indicator transformations on the accuracy of the framework over the forecast horizon. The results show that not transforming offers the best prediction accuracy.

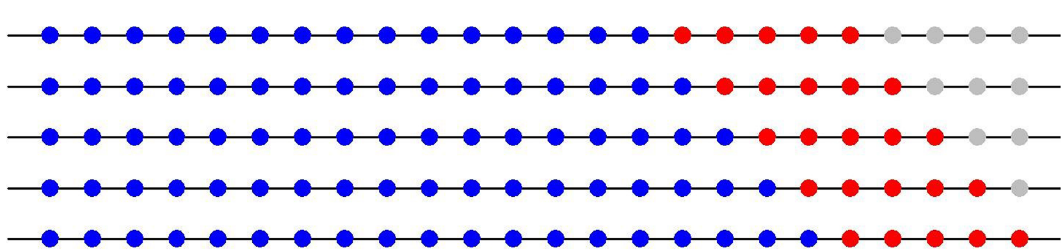

Fig. 2. Example of a rolling window approach.

Table 2
The MAPE (%) comparison of the statistical model.

Horizon	1	2	3	4	5	6	7	8	9	10	11	12
LASSO	11.6	11.9	12.5	13.3	14.4	15.3	16.0	16.8	17.9	18.8	19.7	20.8
Naive	13.5	16.3	17.2	15.6	16.8	19.0	17.2	17.9	18.7	20.0	18.6	15.9
Single	11.7	12.8	13.7	14.2	15.7	16.8	17.8	18.4	19.4	20.6	21.4	22.4
sNaive	17.1	16.9	16.7	16.5	16.6	16.4	16.5	16.1	15.7	15.9	16.1	15.9
Theta	13.0	14.5	14.7	14.1	15.3	16.0	15.4	15.6	16.5	17.5	17.5	17.9
ARIMA	12.3	12.9	14.0	14.3	16.4	16.4	16.7	16.0	16.4	17.2	17.0	17.5
ETS	12.6	14.0	14.7	14.5	15.6	15.9	16.3	16.5	17.4	17.8	17.7	18.8

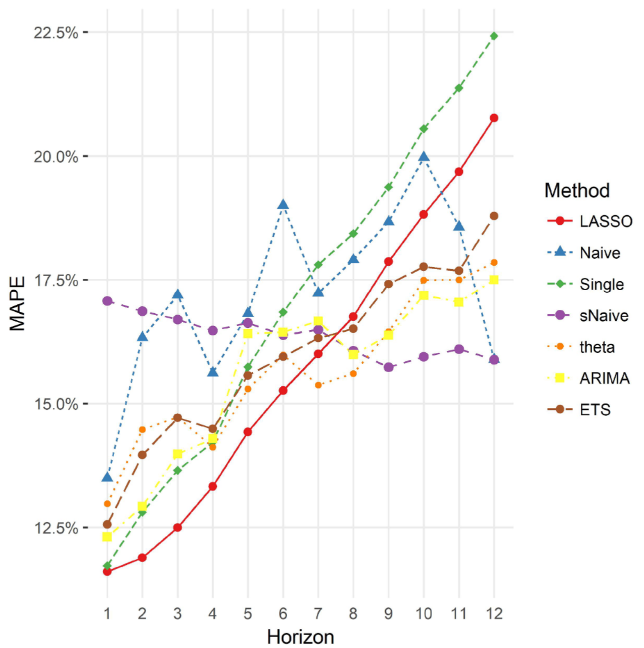

Fig. 3. The influence of the statistical model on the prediction accuracy over the forecast horizon.

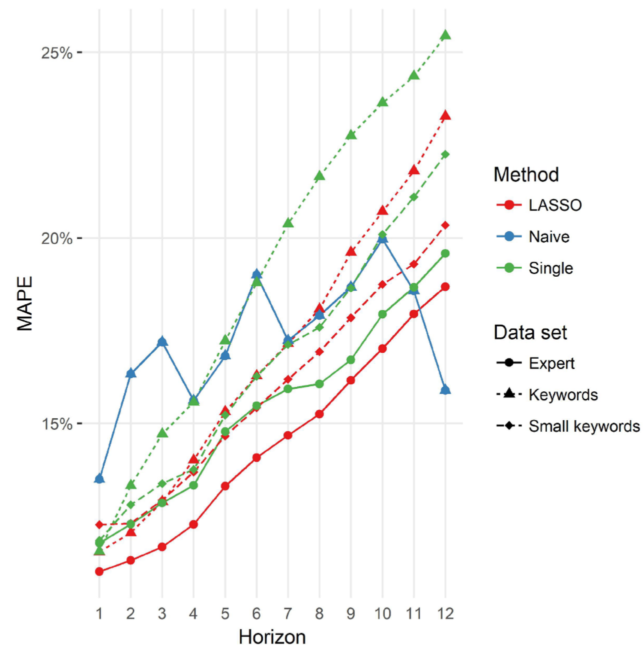

Fig. 4. The influence of the data set on the prediction accuracy over the forecast horizon.

Overall we see that decomposing the variables and using the trend component decreases the prediction accuracy. This is a result of the smoothing removing the peculiarities of the time series, which leads to increased multicollinearity between variables. This makes the regression methods more likely to select non-causal indicators, which leads to bad out-of-sample performance.

Furthermore, we see that the LASSO regression methodology does not have an advantage of using principal component analysis. For the OLS method with one variable (Single) using principal components performs on par with the regular variable set. This is explained by the principal components containing information of multiple indicators, making the OLS approach select the influence of multiple variables. The LASSO method is able to select multiple variables, which reduces the value of performing principal component analysis.

4.3.5. Summary of the proposed methodology

Previous sections have shown that the best performing techniques for the forecasting framework are:

- Use a LASSO model;
- Use the expert data set;
- Decompose the sales data;
- Use raw indicator data.

In this section, we validate the methodology on three case studies that are based in different sectors, and span over multiple geographical locations (Europe and North America).

Table 6, Table 7 and Table 8 show the prediction performance of the methodology. We observe that in most time windows of the forecast horizon a reduction in MAPE is achieved for the proposed methodology. For the supplier to the tire sector and the producer of composite

Table 3
The MAPE (%) comparison of the data set size of the statistical models over the forecast horizon.

Method	Data set	1	2	3	4	5	6	7	8	9	10	11	12
LASSO	Keywords	11.5	12.1	12.9	14.0	15.3	16.3	17.1	18.1	19.6	20.7	21.8	23.3
LASSO	Small key.	12.3	12.3	12.9	13.7	14.7	15.4	16.2	16.9	17.8	18.7	19.3	20.3
LASSO	Expert	11.0	11.3	11.7	12.3	13.3	14.1	14.7	15.2	16.2	17.0	17.9	18.7
Naive	N/A	13.5	16.3	17.2	15.6	16.8	19.0	17.2	17.9	18.7	20.0	18.6	15.9
Single	Keywords	11.6	13.3	14.7	15.6	17.2	18.8	20.4	21.6	22.8	23.6	24.4	25.4
Single	Small key.	11.9	12.8	13.4	13.8	15.2	16.3	17.1	17.6	18.6	20.1	21.1	22.3
Single	Expert	11.8	12.3	12.9	13.3	14.8	15.5	15.9	16.1	16.7	17.9	18.7	19.6

Table 4
The MAPE (%) comparison of the sales transformation of the statistical models over the forecast horizon.

Method	Transformation	1	2	3	4	5	6	7	8	9	10	11	12
LASSO	Transformed	7.6	7.7	8.0	8.6	9.6	10.6	11.6	12.4	13.3	14.4	15.3	16.1
Single	Transformed	8.6	8.9	9.3	10.0	11.1	12.0	13.0	13.5	14.3	15.4	16.4	17.4
LASSO	Raw	15.6	16.1	17.0	18.1	19.2	19.9	20.4	21.1	22.4	23.3	24.1	25.4
Single	Raw	14.8	16.7	18.0	18.4	20.4	21.7	22.6	23.4	24.5	25.7	26.4	27.4
Naive	N/A	13.5	16.3	17.2	15.6	16.8	19.0	17.2	17.9	18.7	20.0	18.6	15.9

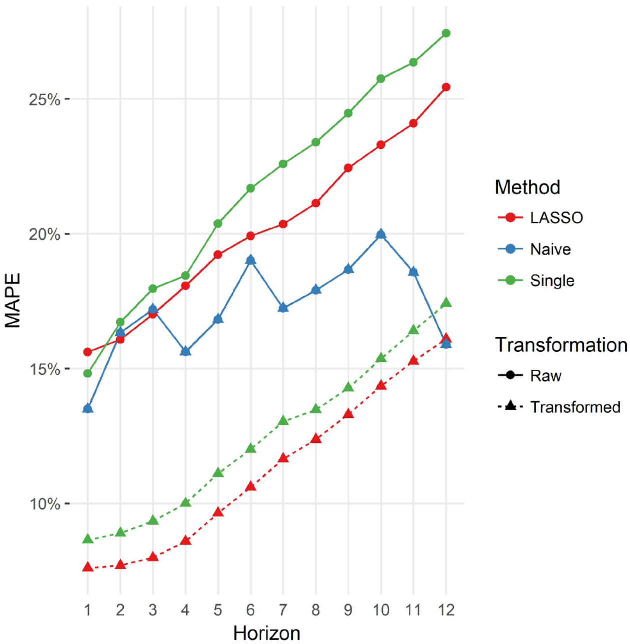

Fig. 5. The influence of the sales transformation on the prediction accuracy over the forecast horizon.

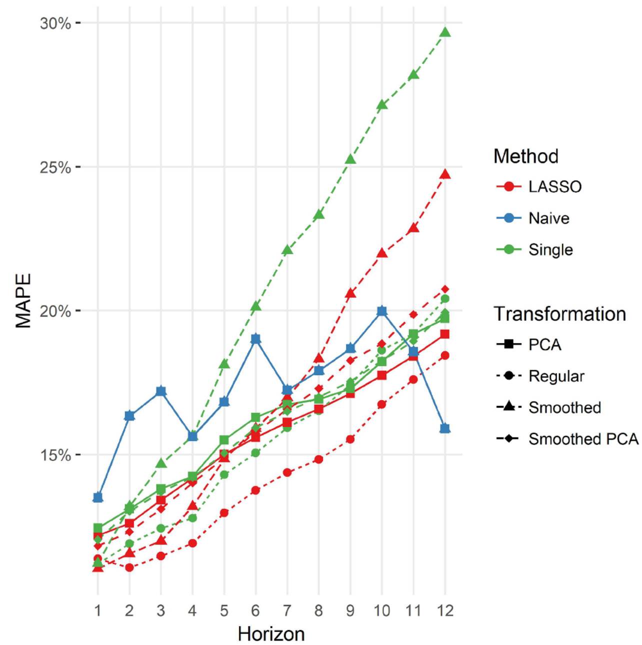

Fig. 6. The influence of the indicator transformation on the prediction accuracy over the forecast horizon.

building materials an improvement is seen for all horizons. For the producer of steel, we see the proposed methodology out-performing the traditional methods for the first 7 time windows. Afterwards, we see methods that assume do not necessarily use a trend (seasonal naive, ETS) taking over. This is a result of the characteristics of the time series that contain a weak trend over the test horizon, which is a difficult pattern to select indicators on as there are no inflection points.

Over the different case studies in multiple geographical locations, we see several patterns. First, we considerably outperform the naive method (the average relative MAPE improvement over all time horizons is 54.5%). Secondly, we consistently outperform the case-wise best performing traditional methods with an average relative accuracy improvement of 25.6%. The smallest improvement comes from the data provided by the steel producer (14.4%) where the method is outperformed by the traditional techniques on the further time horizons.

Comparing the proposed methodology with the OLS variant, which contains the same indicator set and transformations, we see a consistent improvement of the LASSO regression method of 10.9% over using ordinary least squares regression with one variable. This shows the added value of using LASSO that selects multiple variables intelligently.

4.3.6. Managerial impact

The proposed sales forecasting framework resulted in many benefits for the supplier to the tire sector. Before using the macroeconomic forecasting framework, the company created a bottom-up sales forecast for the next three months. This project allowed the firm to forecast accurately further in time, which has led to several advantages.

First, they can better assign their staff based on the macroeconomic environment of their business. The company can anticipate on higher demand levels by hiring new people while avoiding overtime. Secondly, the leading indicator project enables them to buy resources based on the more accurate leading indicator forecast. The macroeconomic

Table 5
The MAPE (%) comparison of the indicator transformation of the statistical models over the forecast horizon.

Method	Transformation	1	2	3	4	5	6	7	8	9	10	11	12
LASSO	Raw	11.4	11.1	11.5	11.9	13.0	13.8	14.4	14.8	15.5	16.7	17.6	18.4
Single	Raw	11.2	11.9	12.4	12.8	14.3	15.1	15.9	16.5	17.4	18.6	19.2	20.4
LASSO	Smoothing	11.0	11.6	12.0	13.2	14.8	15.9	17.0	18.3	20.6	22.0	22.9	24.7
Single	Smoothing	11.2	13.2	14.7	15.6	18.1	20.1	22.1	23.3	25.2	27.1	28.2	29.6
LASSO	PCA	12.2	12.6	13.4	14.2	15.0	15.6	16.1	16.6	17.1	17.7	18.4	19.2
Single	PCA	12.4	13.1	13.8	14.2	15.5	16.3	16.7	16.9	17.3	18.2	19.2	19.7
LASSO	Smoothing + PCA	11.8	12.3	13.1	14.0	14.9	15.8	16.6	17.3	18.3	18.8	19.9	20.7
Single	Smoothing + PCA	12.0	13.0	13.7	14.2	15.0	15.9	16.5	17.0	17.5	18.3	18.9	19.9
Naive	N/A	13.5	16.3	17.2	15.6	16.8	19.0	17.2	17.9	18.7	20.0	18.6	15.9

Table 6
The MAPE (%) comparison of the proposed method over the forecast horizon for the first-tier supplier to the tire sector.

Method	1	2	3	4	5	6	7	8	9	10	11	12
Proposed methodology	7.3	7.3	7.5	8.1	8.9	9.7	10.6	11.0	11.6	12.7	13.6	14.5
Naive	13.5	16.3	17.2	15.6	16.8	19.0	17.2	17.9	18.7	20.0	18.6	15.9
Single	7.3	7.4	7.7	8.6	10.0	10.8	11.8	12.7	13.5	14.6	15.3	16.2
sNaive	17.1	16.9	16.7	16.5	16.6	16.4	16.5	16.1	15.7	15.9	16.1	15.9
theta	13.0	14.5	14.7	14.1	15.3	16.0	15.4	15.6	16.5	17.5	17.5	17.9
ARIMA	12.3	12.9	14.0	14.3	16.4	16.4	16.7	16.0	16.4	17.2	17.0	17.5
ETS	12.6	14.0	14.7	14.5	15.6	15.9	16.3	16.5	17.4	17.8	17.7	18.8

Table 7
The MAPE (%) comparison of the proposed method over the forecast horizon for the producer of building materials.

Method	1	2	3	4	5	6	7	8	9	10	11	12
Proposed methodology	4.0	4.1	4.1	4.2	4.3	4.5	4.8	5.0	5.4	5.6	5.9	6.3
Naive	10.5	13.9	13.7	13.6	15.5	15.5	16.0	14.8	14.8	14.9	12.2	7.1
Single	4.4	4.7	4.8	4.9	5.1	5.2	5.2	5.4	5.8	6.0	6.2	6.3
sNaive	6.4	6.4	6.4	6.5	6.5	6.3	6.5	6.7	6.9	7.0	7.1	7.1
theta	7.1	8.8	9.6	10.7	10.4	10.0	10.6	10.4	9.6	9.6	8.4	8.0
ARIMA	6.0	7.1	7.8	7.5	7.5	6.7	6.5	7.2	7.6	8.6	8.9	10.2
ETS	7.9	9.3	10.5	11.5	10.6	10.1	10.3	9.4	8.2	8.4	7.9	8.4

Table 8
The MAPE (%) comparison of the proposed method over the forecast horizon for the producer of steel.

Method	1	2	3	4	5	6	7	8	9	10	11	12
Proposed methodology	6.4	7.0	7.7	8.9	10.1	11.2	12.3	13.0	13.0	13.8	14.1	16.0
Naive	34.6	32.4	35.1	32.7	36.4	35.5	32.6	26.6	35.6	37.9	35.7	13.5
Single	7.9	9.0	9.9	10.6	12.1	13.2	14.0	14.8	15.3	15.3	16.2	16.6
sNaive	12.4	12.7	12.5	12.7	12.8	12.8	12.9	12.6	13.3	13.3	13.3	13.5
theta	10.2	11.5	12.2	12.4	14.8	15.3	16.2	16.2	15.4	15.4	15.3	15.8
ARIMA	11.5	11.2	12.6	14.5	14.9	15.5	15.8	18.6	18.4	19.0	19.0	20.6
ETS	9.8	10.6	11.7	15.5	15.3	15.2	16.0	16.4	13.7	13.5	13.2	13.8

business environment can increase demand for raw materials, which leads to higher prices. Insights in these drivers allow the company to anticipate for price increases. Thirdly, an inventory reduction was observed. The company works for the most part with a make to order production approach. However, as the raw materials need to be ordered and shipped several weeks beforehand, a significant drop in inventory of raw materials was observed. In other steps of the supply chain, improvements were also observed. The firm improved its service level, because of the better anticipation on demand changes.

5. Conclusions

In this paper we propose an automatic framework that automatically selects macroeconomic indicators for the tactical sales time frame. The proposed framework is based on techniques that are known to improve accuracy when used separately, but have to the best of our knowledge not been combined for tactical sales forecasting so far. The framework we propose consists of decomposing the sales using STL decomposition and predicting each component separately. We suggest to predict the seasonal pattern using the seasonal naive method and the macroeconomic trend using a LASSO regression model.

With this research work we have shown that decomposing the sales data, and forecasting each component independently reduces the forecast error. We ascertain that LASSO regression using multiple external variables should be used to predict the trend extracted from the sales. We show that reducing the indicator set to a smaller, but with a higher density of relatable indicators improves the forecast accuracy. However, this comes with risk, as unidentified business influences might be left out. Literature suggested that variable transformation techniques can improve the prediction performance. We tested the influence of principal component analysis and STL decomposition on the macroeconomic variables, but found no evidence of improved accuracy. Therefore, we suggest to use the unprocessed macro-economic data in the prediction framework.

We quantified accuracy improvements of the methodology on three case studies, consisting of different B2B companies over different geographical areas. We observed a relative reduction in mean absolute percentage of error of on average 54.5% (in comparison to the naive forecasting method) and 25.6% (the case wise best performing conventional forecasting technique) in the tactical time window. We benchmarked our methodology on ordinary least squares regression selecting 1 variable, but with the same framework and variables, and achieved an increase in accuracy of 10.9%, suggesting the added value of LASSO combined with multiple variables.

In this set-up we only tested macroeconomic indicators of the same aggregation level (monthly) as the dependant variable. Future research could include indicators that are reported in higher aggregation levels, so that more variables could be added.

We focused on selecting macroeconomic variables to predict the extracted (long-term) trend of the sales. We suspect prediction performance could be improved by adding non-macroeconomic variables to the study. However, this is not a trivial issue, as such information (weather, price, promotions, etc.) is usually defined on a different aggregation level and is not available for the (entire) tactical horizon. We do believe the impact on the selection of the macroeconomic variables is limited as these sudden swings are captured in the noise component (and not in the trend) of the decomposition method.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. We thank the reviewers for their comment and remarks. These helped us improve the paper.

References

Assimakopoulos, V., & Nikolopoulos, K. (2000). The theta model: A decomposition approach to forecasting. *International Journal of Forecasting, 16*(4), 521–530.

Babai, M. Z., Ali, M. M., Boylan, J. E., & Syntetos, A. A. (2013). Forecasting and inventory performance in a two-stage supply chain with arima (0, 1, 1) demand: Theory and empirical analysis. *International Journal of Production Economics, 143*(2), 463–471.

Bai, J., & Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica, 70*(1), 191–221.

Bergmeir, C., Hyndman, R. J., & Benítez, J. M. (2016). Bagging exponential smoothing methods using stl decomposition and box–cox transformation. *International Journal of Forecasting, 32*(2), 303–312.

Boivin, J., & Ng, S. (2006). Are more data always better for factor analysis? *Journal of Econometrics, 132*(1), 169–194.

Cawley, G. C., & Talbot, N. L. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. *Journal of Machine Learning Research, 11*(Jul), 2079–2107.

Cepni, O., Güney, I. E., & Swanson, N. R. (2019). Nowcasting and forecasting gdp in emerging markets using global financial and macroeconomic diffusion indexes. *International Journal of Forecasting, 35*(2), 555–572.

Chatfield, C. (1988). The future of the time-series forecasting. *International Journal of Forecasting, 4*(3), 411–419.

Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). Stl: A seasonal-trend decomposition. *Journal of Official Statistics, 6*(1), 3–73.

Crittenden, V. L., Gardiner, L. R., & Stam, A. (1993). Reducing conflict between marketing and manufacturing. *Industrial Marketing Management, 22*(4), 299–309.

Cui, R., Gallino, S., Moreno, A., & Zhang, D. J. (2017). The operational value of social media information. *Production and Operations Management*.

Dalrymple, D. J. (1987). Sales forecasting practices: Results from a united states survey. *International Journal of Forecasting, 3*(3–4), 379–391.

De Gooijer, J. G., & Hyndman, R. J. (2006). 25 years of time series forecasting. *International journal of forecasting, 22*(3), 443–473.

Ding, S. (2018). A novel discrete grey multivariable model and its application in forecasting the output value of china's high-tech industries. *Computers & Industrial Engineering*.

Fildes, R., & Kingsman, B. (2011). Incorporating demand uncertainty and forecast error in supply chain planning models. *Journal of the Operational Research Society, 62*(3), 483–500.

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software, 33*(1), 1.

Guidolina, M., Guseoa, R., & Mortarinoa, C. (2019). Regular and promotional sales in new product life cycles: Competition and forecasting. *Computers & Industrial Engineering*.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference and prediction* (2nd ed.). Springer.

Hyndman, R. J. & Athanasopoulos, G. (2014). Forecasting: Principles and practice. OTexts.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting, 22*(4), 679–688.

Kim, H. H., & Swanson, N. R. (2016). Mining big data using parsimonious factor, machine learning, variable selection and shrinkage methods. *International Journal of Forecasting*.

Kim, H. H., & Swanson, N. R. (2018). Methods for backcasting, nowcasting and forecasting using factor-midas: With an application to korean gdp. *Journal of Forecasting, 37*(3), 281–302.

Kim, T. Y., Dekker, R., & Heij, C. (2017). Spare part demand forecasting for consumer goods using installed base information. *Computers & Industrial Engineering, 103*, 201–215.

Klassen, R. D., & Flores, B. E. (2001). Forecasting practices of canadian firms: Survey results and comparisons. *International Journal of Production Economics, 70*(2), 163–174.

Li, J., & Chen, W. (2014). Forecasting macroeconomic time series: Lasso-based approaches and their forecast combinations with dynamic factor models. *International Journal of Forecasting, 30*(4), 996–1015.

Ludwig, N., Feuerriegel, S., & Neumann, D. (2015). Putting big data analytics to work: Feature selection for forecasting electricity prices using the lasso and random forests. *Journal of Decision Systems, 24*(1), 19–36.

Murray, P. W., Agard, B., & Barajas, M. A. (2018). Forecast of individual customer's demand from a large and noisy dataset. *Computers & Industrial Engineering, 118*, 33–43.

Oliva, R., & Watson, N. (2009). Managing functional biases in organizational forecasts: A case study of consensus forecasting in supply chain planning. *Production and operations Management, 18*(2), 138–151.

Ord, J. K. (1988). Future developments in forecasting: The time series connexion. *International Journal of Forecasting, 4*(3), 389–401.

Özer, Ö., & Wei, W. (2004). Inventory control with limited capacity and advance demand information. *Operations Research, 52*(6), 988–1000.

Panagiotelis, A., Athanasopoulos, G., Hyndman, R. J., Jiang, B., & Vahid, F. (2019). Macroeconomic forecasting for australia using a large number of predictors. *International Journal of Forecasting, 35*(2), 616–633.

Plakandaras, V., Gogas, P., Papadimitriou, T., & Gupta, R. (2017). The informational content of the term spread in forecasting the us inflation rate: A nonlinear approach. *Journal of Forecasting, 36*(2), 109–121.

Qu, T., Zhang, J., Chan, F. T., Srivastava, R., Tiwari, M., & Park, W.-Y. (2017). Demand prediction and price optimization for semi-luxury supermarket segment. *Computers & Industrial Engineering, 113*, 91–102.

Sagaert, Y. R., Aghezzaf, E.-H., Kourentzes, N., & Desmet, B. (2017). Temporal big data for tire industry tactical sales forecasting. *Interfaces*.

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.

Sagaert, Y. R., Kourentzes, N., De Vuyst, S., Aghezzaf, E.-H., & Desmet, B. (2019). Incorporating macroeconomic leading indicators in tactical capacity planning. *International Journal of Production Economics, 209*, 12–19.

Sakai, H., Nakajima, H., Higashihara, M., Yasuda, M., & Oosumi, M. (1999). Development of a fuzzy sales forecasting system for vending machines. *Computers & industrial engineering, 36*(2), 427–449.

Sanders, N. R., & Manrodt, K. B. (2003). The efficacy of using judgmental versus quantitative forecasting methods in practice. *Omega, 31*(6), 511–522.

Smeekes, S., & Wijler, E. (2018). Macroeconomic forecasting using penalized regression methods. *International Journal of Forecasting, 34*(3), 408–430.

Stein, J. C. (1997). Internal capital markets and the competition for corporate resources. *The Journal of Finance, 52*(1), 111–133.

Stock, J. H., & Watson, M. W. (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American statistical association, 97*(460), 1167–1179.

Stock, J. H., & Watson, M. W. (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics, 20*(2), 147–162.

Theodosiou, M. (2011). Forecasting monthly and quarterly time series using stl decomposition. *International Journal of Forecasting, 27*(4), 1178–1195.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological),* 267–288.

Uematsu, Y., & Tanaka, S. (2019). High-dimensional macroeconomic forecasting and variable selection via penalized regression. *The Econometrics Journal, 22*(1), 34–56.

Verstraete, G., Aghezzaf, E.-H., & Desmet, B. (2019). A data-driven framework for predicting weather impact on high-volume low-margin retail products. *Journal of Retailing and Consumer Services, 48*, 169–177.

Wang, X., & Petropoulos, F. (2016). To select or to combine? The inventory performance of model and expert forecasts. *International Journal of Production Research, 54*(17), 5271–5282.

Xia, Y., Wang, H., & Liu, W. (2019). The indirect carbon emission from household consumption in china between 1995–2009 and 2010–2030: A decomposition and prediction analysis. *Computers & Industrial Engineering, 128*, 264–276.

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.

Yan, H.-S., & Tu, X. (2012). Short-term sales forecasting with change-point evaluation and pattern matching algorithms. *Expert Systems with Applications, 39*(5), 5426–5439.

Yan, H.-S., Tu, X., & Wu, Q. (2018). Robust v-support vector machine and its application to sale forecasting. *International Journal of Industrial Engineering: Theory, Applications and Practice, 25*(3), 387–401.

Yan, H.-S., Wu, Q., & Tu, X. (2013). A short-term forecasting model with inhibiting normal distribution noise of sale series. *Applied Artificial Intelligence, 27*(6), 496–519.