

# A multivariate approach for multi-step demand forecasting in assembly industries: Empirical evidence from an automotive supply chain

João N.C. Gonçalves<sup>a,\*</sup>, Paulo Cortez<sup>b</sup>, M. Sameiro Carvalho<sup>a</sup>, Nuno M. Frazão<sup>c</sup>

<sup>a</sup>*ALGORITMI Research Centre, Department of Production and Systems, University of Minho, 4710-057 Braga, Portugal*

<sup>b</sup>*ALGORITMI Research Centre, Department of Information Systems, University of Minho, 4804-533 Guimarães, Portugal*

<sup>c</sup>*Robert Bosch GmbH, Automotive Electronics Division, Logistics Innovation Section, 4701-970 Braga, Portugal*

---

## Abstract

Demand forecasting works as a basis for operating, business and production planning decisions in many supply chain contexts. Yet, how to accurately predict the manufacturer's demand for components in the presence of end-customer demand uncertainty remains poorly understood. Assigning the proper order quantities of components to suppliers thus becomes a nontrivial task, with a significant impact on planning, capacity and inventory-related costs. This paper introduces a multivariate approach to predict manufacturer's demand for components throughout multiple forecast horizons using different leading indicators of demand shifts. We compare the autoregressive integrated moving average model with exogenous inputs (ARIMAX) with Machine Learning (ML) models. Using a real case study, we empirically evaluate the forecasting and supply chain performance of the multivariate regression models over the component's life-cycle. The experiments show that the proposed multivariate approach provides superior forecasting and inventory performance compared with traditional univariate benchmarks. Moreover, it reveals applicable throughout the component's life-cycle, not just to a single stage. Particularly, we found that demand signals at the beginning of the life-cycle are predicted better by the ARIMAX model, but it is outperformed by ML-based models in later life-cycle stages.

*Keywords:* Forecasting, Supply chain management, Decision support, Machine learning, ARIMAX.

---

\*Correspondence: João N.C. Gonçalves (e-mail address: jncostagoncalves@gmail.com)

## 1. Introduction

Demand forecasting is a fundamental aspect in supply chain management (SCM) with a significant impact on planning, capacity and inventory control decisions [1]. In the inventory control context, underestimated and overestimated forecasts impose an increase in backlog and holding costs. The resulting demand volatility is therefore one of the main causes of order amplification from downstream to upstream supply chain (SC) players [2].

Time series forecasting has been the focus of extensive research studies across several fields, including energy, transportation, fashion retailing, finance and SCM [see, e.g., 3, 4, 5, 6, 7], but the generation of accurate SC demand forecasts still represents a challenging task for both researchers and practitioners [8]. One possible explanation for this may be the fact that sales forecasts are often prone to errors primarily given their dependence on exogenous and nonlinear variables (e.g., retail price and advertising) that, all together, potentially hamper the development of effective forecasting models. While it is well-known that demand variability may lead to large cost implications to upstream SC echelons (even when it is not amplified but simply transmitted) [9], understanding of the processes that lead to improvements of demand forecasting at the manufacturer stage is limited. It is interesting to note here that demand forecasting at the manufacturer level is quite challenging, mainly for two reasons: first, it is in practice very difficult to accurately predict end-customer demand based on market information, due to its erratic behavior. As a corollary, demand signals from end-customers to manufacturers can be increasingly distorted. Second, even the minutest shift in the end-customer demand signals can lead to significant levels of uncertainty at the upstream levels of the SC network. Indeed, this holds true mainly when the levels of cooperation between SC members are low. Such information asymmetry makes the estimation of the quantities to be ordered for various components<sup>1</sup> to different suppliers a challenging task, especially when the market information that feeds end-customer demand signals is highly erratic. This gives rise to an interesting research question we aim to answer in this paper: How can we promote accurate manufacturer's demand for components in the presence of end-customer demand forecast uncertainty while maintaining suitable service levels and decreased inventory-related costs? The objective of this research is to explore a new multivariate approach to forecast demand for compo-

---

<sup>1</sup>A *component* is here defined as an inventory item, other than a finished product, that goes into higher level items in the bill of materials.

nents that takes advantage of different leading indicators of manufacturer’s demand shifts. Following an unconditional (*ex ante*) forecasting setup, we explore several Machine Learning (ML) models together with the statistical ARIMAX model to assess the viability of the proposed approach over the component’s life-cycle. Interestingly, note that the ARIMAX model has recently achieved good forecasting results on series with different levels of volatility [10]. To empirically evaluate our forecasting approach, we joined forces with a major automotive electronics organization – Bosch Automotive Electronics, Portugal – considered to be one of the largest business units belonging to the Bosch group. The performance of the multivariate regression models is then compared to that of traditional univariate statistical ones and evaluated according to the forecasting errors as well as to the overall inventory-related costs derived therefrom. The proposed approach could act as an intelligent decision support system (DSS) for logistics planners at the upstream-end of automotive electronics SCs. In a nutshell, the contribution of this research to the extant literature on demand forecasting is threefold:

- i. By resorting to ARIMAX and ML-based models, we explore a simple yet effective multivariate approach for multi-step forecasting of manufacturer’s demand based on suitable leading indicators. Using a realistic rolling origin forecasting setup, the stability of the proposed approach is assessed across the component’s life-cycle stages, and the forecasting results obtained are compared with those derived from traditional pure statistical methods, including Naïve benchmarks.
- ii. We highlight the usefulness of the proposed forecasting approach by addressing a real-life problem faced by a major make-to-order (MTO) automotive electronics organization (Bosch Automotive Electronics, Portugal), that operates with multiple suppliers and end-customers worldwide.
- iii. We consider the accuracy-cost trade-off when comparing the performance of the forecasting models such that company managers can assess what kind of impact can be expected from forecast variations on the overall inventory performance.

This paper is structured as follows. In the next Section, we provide a literature overview on SC demand forecasting models. In Section 3, we describe the problem at stake. Section 4 provides details on the proposed framework, including the multivariate regression models considered. Section 5 outlines the experimental design and the model evaluation criteria. We further present and discuss the empirical results in Section 6. Here, the implications of

our work for decision support are also provided. Finally, conclusions are drawn in Section 7.

## 2. Related work

Typically, SC demand forecasting studies carried out to date have been proposed to: 1) forecast intermittent demand patterns [11, 12, 13, 14, 15]; 2) predict new product demands over its life-cycle using product differentiation information [16]; or even 3) forecast spare parts demand for regular products [17]. Table 1 provides an overview of the discussed literature on SC demand forecasting. In the overwhelming majority of the cases, the presented models rely on statistical-based techniques to grasp important forecast implications in several contexts. ML-based forecasting approaches have, however, been explored to capture the complex dynamics of SC demand. Neural networks (NNs) are the most popular ML models concerning time series forecasting [18], and have proven to be effective in forecasting very challenging contexts associated with lumpy demand [19] and demand with incomplete data [20], sometimes outperforming traditional univariate methods such as exponential smoothing and multiple regression [21]. Oftentimes, NN-based models are also exploited in combination with other methods to form hybrid approaches. Aburto and Weber [22] introduce a hybrid intelligent demand forecasting system that combines traditional statistical ARIMA models and NNs. The authors show that this combination provides more accurate forecasts than the methods individually. Jaipuria and Mahapatra [23] present an integrated approach, based on the combination of discrete wavelet transforms and NNs, applicable to linear, nonlinear, stationary or non-stationary data series. Though NNs have been successfully applied to diverse SC contexts, Abolghasemi et al. [10] provide evidences that simple statistical forecasting models can outperform some ML approaches when demand series is highly volatile. Nikolopoulos et al. [24] also show that forecast accuracy performance can be deteriorated when using nearest neighbor approaches to forecast demand series with high levels of intermittence.

In summary, a notorious progress has been made in proposing novel or improved models to forecast SC demand under different circumstances. Yet, there is no clear evidence that ML models outperform traditional statistical forecasting methods, and previous research studies [25, 26] corroborate these findings. This is not surprising as the performance of forecasting models strongly depends on the nature of the time series and no forecasting method is a panacea for all forecasting problems. There are also other factors that come

into play here, such as the length of the series and the forecast horizon.

Apart from the fact that the overwhelming majority of the works listed in Table 1 consider univariate information extrapolating from the past when forecasting SC demand, we further observe that the academic literature is rather scarce in terms of demand forecasting applications relying on the manufacturer level (hereafter referred to as upstream-end side). For instance, Carbonneau et al. [27] studied the effectiveness of ML strategies in forecasting distorted demand signals as they move through upstream SC echelons. It is shown that despite NN models reveal to be more accurate than linear regression, the differences are not statistically significantly different. This supports the claim that simple baseline benchmark models should be considered to attest the superiority of ML techniques. A second example is the work of Mukattash and Samhuri [28], who proposed a forecasting framework to improve supply planning using a multivariate regression model, without having to resort to the bill of materials (BOM).

Depending on the nature and complexity of the SC, many different variables may impact on the dynamics of demand. Recent published research has found improvements in forecast accuracy by including leading indicators in the forecasting process [29, 30]. Unfortunately, understand how logistics leading indicators may improve manufacturer’s demand for components remains an important unresolved question demanding appropriate answers, especially in the context of inventory management in assembly industries. In this context, to the best of our knowledge, this is the first paper combining different leading indicators, encompassing product structure information and end-customer demand behavior, into a multivariate forecasting framework able to predict manufacturer’s demand. In this paper, we contribute to the demand forecasting literature by proposing a multivariate framework able to forecast manufacturer’s demand throughout the entire component’s life-cycle. Furthermore, we study the dynamics of the proposed framework using both forecast accuracy and inventory performance measures.

Table 1: Overview on supply chain demand forecasting literature.

Authors	Forecast approach	Forecast technique(s) <sup>a</sup>	Target frequency <sup>b</sup>	Empirical data
Willemain et al. [11]	Univariate	Bootstrap, Croston, ES	M	Yes
Syntetos and Boylan [12]	Univariate	SBA, SMA, SES, Croston	M	Yes
Carbonneau et al. [27]	Univariate	MLR, NN, RNN, SVM Naïve, MA, Trend	M	Yes
Gutierrez et al. [19]	Univariate	NN, Croston, SES, SBA	D	Yes
Efendigil et al. [20]	Univariate	NN	M	Yes
Ferbar et al. [31]	Univariate	Wavelet denoising, ES	–	No
Sayed et al. [32]	Univariate	SES, HW, GA	M	Yes
Yelland [33]	Univariate	Bayesian method, SSM	Q	Yes
Mukattash and Samhouri [28]	Multivariate	MLR, VAR	–	No
Petropoulos et al. [13]	Univariate	Croston-Theta, SBA, SMA Naïve, SES	M	Yes
Lau et al. [21]	Univariate	NN, ES; MLR	M	Yes
Kourentzes [14]	Univariate	NN, Croston <sup>c</sup> , Naïve, MA, SES	M	Yes
Jaipuria and Mahapatra [23]	Univariate	DWT, NN, ARIMA	M	Yes
Rego and Mesquita [17]	Univariate	Bootstrap, SMA, SBA	W, M	Yes
Nikolopoulos et al. [24]	Univariate	$k$ -NN, SES, SBA, TSB, Croston	M	Yes
Huber et al. [34]	Uni/Multivariate	ARIMAX, ARIMA	D	Yes
Afrin et al. [16]	Univariate	DDI-EWMA	Y	Yes
Murray et al. [35]	Univariate	ARIMA	W	Yes
Fu and Chien [15]	Univariate	MA, ARIMA, Croston, $k$ -NN, SBA, TSB, SVM, RNN, MAPA	W	Yes
Abolghasemi et al. [10]	Uni/Multivariate	ETS, ETSX, ARIMA, SVM ARIMAX, NN, DLR, Theta	W	Yes
Current study	Uni/Multivariate	SVM, NN, RF, ARIMAX AutoML, ERNN, ARIMA Theta, Naïve	W	Yes

<sup>a</sup> ES: Exponential Smoothing, SMA: Simple Moving Average, SES: Single Exponential Smoothing, MLR: Multiple Linear Regression, NN: Neural Networks, RNN: Recurrent Neural Networks, SVM: Support Vector Machines, MA: Moving Average, SBA: Syntetos-Boylan Approximation, HW: Holt Winter's, GA: Genetic Algorithm, SSM: State Space Model, VAR: Vector Autoregression, DWT: Discrete Wavelet Transforms, ARIMA: Autoregressive Integrated Moving Average, TSB: Teunter-Syntetos-Babai approximation, ARIMAX: ARIMA with exogenous information, DDI-EWMA: Demand Differentiation Index-Exponential Weighted Moving Average,  $k$ -NN: Nearest Neighbor, MAPA: Multiple Aggregation Prediction Algorithm, ETSX: Exponential Smoothing in the state space framework, DLR: Dynamic Linear Regression, RF: Random Forest, AutoML: Automated Machine Learning, ERNN: Elman Recurrent Neural Network.

<sup>b</sup> Target frequency legend: D=Daily, W=Weekly, M=Monthly, Q=Quarterly, Y=Yearly, –=Not reported.

<sup>c</sup> Different Croston-based methods are used as modeling approaches.

### 3. Problem formulation

Consider a classical SC topology consisting of a single manufacturer linked with different suppliers and end-customers (Fig. 1), in which there exists an information flow (dashed line) from the market to the different suppliers and a material flow (solid line) from the suppliers to the end-customers. We will speak of end-customers as the original equipment manufacturing (OEM) buyers. Forecasted demands for finished products are used to determine the component order sizes to suppliers. The supplied components  $\{c_1, c_2, \dots, c_j\}$  are then assembled by the manufacturer to further produce a set of finished products  $\{p_1, p_2, \dots, p_r\}$ , with  $r \leq j$ , and fulfill end-customer requests. We consider that the production of finished products is MTO and happens continuously over time. Under this topology, demand forecasts for finished products may serve as input for the development of daily production plans.

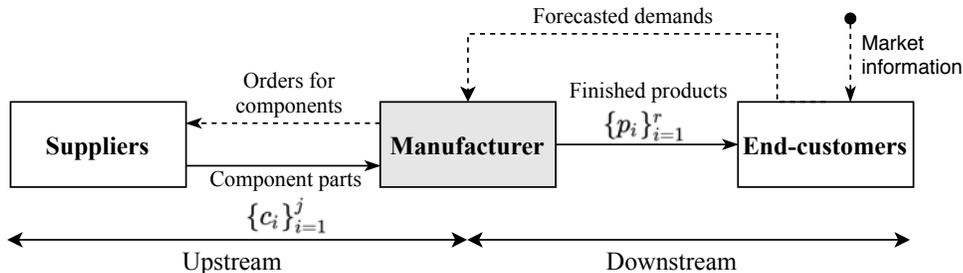


Figure 1: Information (- -) and material flow (—) in a classical supply chain (SC) topology.

Typically, the manufacturer determines the component's order sizes by considering product structure information (BOM), safety stocks, inventory levels and transportation lead times. From a modeling perspective, there are two classical approaches for forecasting manufacturer's demand for components. The first consists of using univariate time series forecasting models directly on historical records of manufacturer's demand for components, but such strategy can be biased since it uses no information from the customer(s) demand behavior. The second approach is aligned with the material requirements planning (MRP) methodology [36] and takes advantage of the BOM to provide the component requirements for future time periods based on finished product forecasts [37]. Nevertheless, if a given component is used to produce a large set of finished products, it would be necessary to use forecasted demands of all these products in order to further provide the component requirements via BOM-explosions. In the end, this procedure would lead to

both significant cumulative forecasting errors and inventory-related costs (especially if the component is used more than once in the different products), which impact strongly on sourcing commitments and transportation decisions. Besides, erratic market information may also easily generate distorted end-customer demand signals to manufacturer, entailing several implications for management. In particular, if the real end-customer demand is higher than the end-customer demand forecast, then the manufacturer may need to resort to last minute emergency shipments (called premium freights) in a bid to avoid stock-outs, delay risks and/or production line stoppages. Conversely, if the end-customer demand forecast is higher than the real end-customer demand, the manufacturer incurs in additional holding costs. To face demand uncertainty, organizations often tend to increase the safety stocks for components at early life-cycle stages.

Hence, because end-customer demand signals are naturally error-prone, our goal in this study is to enhance manufacturer's demand forecasting, as in [27], but without considering merely past information relating to component consumptions/orders.

Let  $c$  be a given component and let  $F = \{p_i\}_{i=1}^r$  be a finite and unordered set of different finished products that make use of  $c$  in their BOM-structures. Hereinafter, we will speak of finished products as only those that make use of  $c$  in their product structure. In addition to previous component's demand, we propose to include three leading indicators that can capture the fundamental dynamics of manufacturer's demand for  $c$  over the course of its life-cycle. The first is the cardinality of  $F$  ( $\#F$ ). The manufacturer's demand for  $c$  tends to increase whenever the demand for finished products  $p_i \in F$  (of which  $c$  is a component part) increases, especially when  $\#F$  is large. As both new assignments of components to a given BOM and major engineering change requests (ECR) in the finished product structure are typically planned and created weeks ahead of the realization of the actual manufacturer's demand,  $\#F$  can be updated accordingly and thus enabling to anticipate future increases or declines in the manufacturer's demand for  $c$ . This indicator is strictly related to the concept of *component commonality* [38]. The second is the total amount of units of  $c$  required for producing each finished product  $p_i \in F$  ( $U$ ), which strongly reflects the magnitude of the manufacturer's demand: high levels of  $U$  tend to lead to an increased demand for  $c$ , whenever there exists demand for  $p_i \in F$ . This indicator contains product structure information and, for the same reasons as those given in the justification of  $\#F$ , it is also an early leading indicator of component's demand. The third is the field intelligence information regarding production orders, planned in the current period

for the future, of finished products  $p_i \in F$  ( $P$ ). Logistics planning experts operate in close liaison with sales department and end-customers to develop dynamic weeks-ahead production plans for finished products that may account for internal capacity constraints, expected events across the planning horizon and adjustments to possible misaligned end-customer demand signals. This makes it possible to use the resulting planned orders to generate a more refined estimation of future finished product production volume, thereby anticipating manufacturer’s demand for  $c$ . This indicator reveals fundamental insights about the future behavior of end-customer(s) demand, and it is less prone to sharp under and overestimations as it happens in the case of initial end-customer forecasts. It is worth mentioning that previous studies have already shown that expert information may lead to improved forecasting performance [e.g., 29]. To account for different leading effects of each indicator, we consider lagged versions of  $\#F, U$  and  $P$ . In addition, as  $\#F$  can vary significantly over time  $t$ , especially in highly volatile SC environments with frequent ECR at product level, we consider the last two above-mentioned input features in an aggregate form for all the finished products  $p_i \in F$ , rather than separately (as single regressors) for each one of them.

The central problem in this paper can thus be formulated in a typical regression fashion as follows. Given an output sequence of past demands for  $c$  lagged by  $k$  time periods up to time  $T$  ( $[D_t]_{t=(T-k):T}$ ), and an input vector  $\mathbf{x}_t$  that integrates the time lags for the leading indicators  $\#F$  (cardinality of finished products),  $U$  (total number of units of  $c$  necessary for producing each  $p_i \in F$ ) and  $P$  (planned finished product production volume) in the same time intervals, our purpose consists in predicting future demands  $\widehat{D}_{T+1}$  to  $\widehat{D}_{T+H}$ , where  $H > 1$  denotes the maximum forecast horizon. More formally, this multivariate framework can be formulated as:

$$\left[ \widehat{D}_{T+h} \right]_{h=1:H} = f([D_t]_{t=(T-k):T}, [\mathbf{x}_t]_{t=(T-k):T}; \theta) + \omega_{T+h}, \quad k \in \{1, \dots, d\}, \quad (1)$$

where  $d$  denotes the maximum order of time lags,  $\omega_{T+h}$  is the error term, and  $\theta$  is a vector of parameters of the regression function  $f(\cdot)$ . The objective is to find the function  $f(\cdot)$  for which the forecast error is minimal. For each horizon  $h$ , we impose that  $k \geq h$  to ensure that only observed values are used as inputs for forecasts (see Section 4.2 for details).

## 4. Methodology

### 4.1. Proposed forecasting framework

We propose a two-stage forecasting framework for manufacturer’s demand forecasting, which is presented in Fig. 2 and detailed throughout subsequent Sections. In such a setting, we firstly build a multivariate dataset of time series input sequences for each component, using the leading indicators described previously. Then, we adopt a sliding time window [39] to create a set of training instances defined by a pre-specified number of time lags related with each input feature. At this point, since future values of the leading indicators are typically unknown [29], and only information that is available at the time when the forecast is generated can be used to predict future manufacturer’s demand, we formulate an unconditional forecasting setup (see Section 4.2). Further, we design a rolling origin forecasting scheme (see Section 5.2) to collect multiple forecasting errors using different forecast origins and test sets, which allows to increase the reliability of the results. This scheme serves as basis for the modeling and evaluation processes, which include the selection of the suitable time lags, the model training and validation, and the assessment of the generalization capability of the model to unseen data.

In a second stage, the optimized model is used for modeling the relationship between the selected inputs and the target variable, to ultimately forecast manufacturer’s demand and help SC managers in making decisions for improving inventory management and order plans to suppliers. We believe that the proposed forecasting framework is flexible enough to allow its implementation in general supply chain networks involving assembly operations, as is the case with the automotive, manufacturing, semiconductor and computer industries.

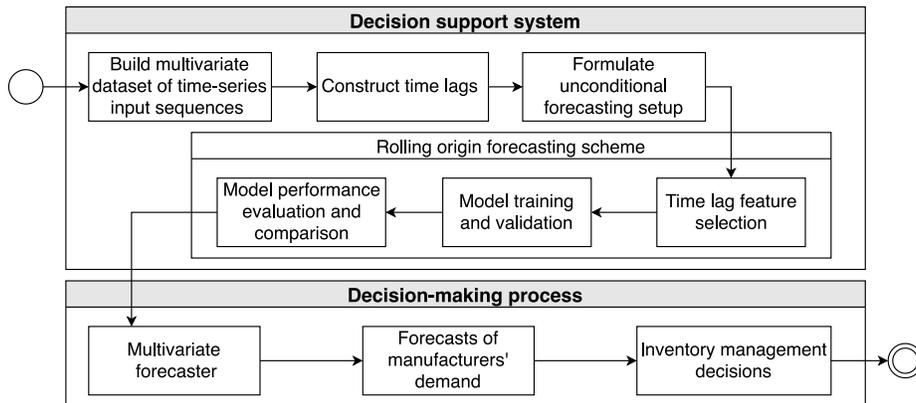


Figure 2: Proposed multivariate forecasting framework.

#### 4.2. Unconditional forecasting

In contrast with conditional (or *ex post*) forecasting, in which later information on the regressors is assumed to be known, unconditional (or *ex ante*) forecasting uses only the information that is available up until the forecast origin at period  $T$  [40]. Concretely, since future values of the leading indicators are generally unknown, lower order lags cannot be used when forecasting higher order horizons, since they will have no data available. Hence, we adjusted the regression models for each horizon in such a way that only time lags greater or equal to the current forecast horizon are included as regressors. For instance, when forecasting horizon is  $h = 1$  all the lagged indicators are considered. Yet, when forecasting horizon is  $h = 2$  only variables lagged by 2 or more time periods can be used, as the values of lower order lags are unknown. In the limit, when forecasting a maximum forecast horizon of  $h = H$  only indicators lagged by  $k \geq H$  periods are used in the process.

Following this strategy, only observed values are considered as inputs for forecasts. Obviously, this setup implies the design and training of  $H$  different forecasting models. Despite this apparent modeling complexity, a natural and important advantage of unconditional forecasting is that it avoids the need to predict all regressors separately, and therefore the addition of noise to the forecasting process. Besides, such a setting allows considering hard-to-predict turning points captured by the leading indicators [29] that would not be included in the model if, for instance, a vector autoregressive (VAR) approach was employed. Note that the unconditional setup setting is commonly used in other contexts, such as in tactical sales forecasting [29] and tactical capacity planning [30].

#### 4.3. Models

In this Section, we elaborate on the choice of the regression models. We investigated a multivariate expansion of the Auto-Regressive Integrated Moving Average (ARIMA) model, commonly known as ARIMAX, that allows the inclusion of exogenous inputs apart from the autoregressive and moving average parameters. In addition, we considered three flexible and popular supervised ML regression models: multilayer perceptron (MLP), support vector regression (SVR) and random forest (RF). Here, these models are considered as implemented in the **rminer** package [41] of the **R** statistical computing language [42]. The use of these non-parametric ML-based models may represent an advantage with respect to the conventional statistical-based forecast techniques, as they can cope with complex non-linear mappings with increased tolerance to noise. In effect, these ML regression models

were previously applied to multi-step forecasting of univariate time series [39], obtaining competitive results when tested with several small sized series (from 108 to 192 observations), and compared with the statistical ARIMA and Holt-Winters forecasting methods.

#### 4.3.1. ARIMAX model

The ARIMAX model can be mathematically expressed in the form  $\text{ARIMAX}(p, d, q, r)$ , which combines the AutoRegressive (AR) model with order  $p$ , the Integrated (I) with degree of differencing  $d$ , the Moving Average (MA) with order  $q$  and the eXogenous (X( $r$ )), where  $r$  denotes the maximum number of exogenous variables included in the model, i.e.,

$$y_t = \rho + \sum_{i=1}^p \beta_i y_{t-i} + \sum_{j=1}^r \omega_j x_j + \sum_{j=1}^q \theta_j z_{t-j} + z_t, \quad (2)$$

where  $y_t$  is a dependent variable at time  $t$ ;  $\rho$  is a constant;  $y_{t-i}$  is a dependent variable lagged by  $i$  periods and  $\beta_i$  are the respective coefficients;  $x_j$  denotes the exogenous variables and  $\omega_j$  are the respective coefficients;  $z_{t-j}$  is the error at time  $t - j$  with coefficient  $\theta_j$  and  $z_t \sim \mathcal{N}(0, \sigma^2)$  is a white noise process. We followed a variation of the Hyndman-Khandakar algorithm [67] for an automatic ARIMAX modeling.

#### 4.3.2. Multilayer perceptrons (MLP)

The MLP is a feedforward artificial NN that consists of an input layer, followed by one or more hidden layers comprised by nonlinearly-activating neurons (nodes), and an output layer. In this work, we adopt a MLP architecture composed by  $m$  inputs, a single hidden layer with  $\mathcal{H}$  neurons and a single neuron in the output layer. The choice of this configuration relies strongly upon its ability to perform, under certain conditions, universal approximations of continuous functions on compact subsets of  $\mathbb{R}^n$  [43]. We have considered the identity function as the activation function for the output node and the sigmoid activation function  $\phi(v) = 1/(1 + e^{-v})$  for the hidden layer nodes. During the MLP learning process, the synaptic weights are constantly adapted based on the Broyden-Fletcher-Goldfarb-Shanno (BFGS) quasi-Newton algorithm [44]. The final MLP solution is, however, strictly dependent on an initial configuration of synaptic weights. To overcome this problem, we considered an ensemble of  $N_r$  distinct trained MLPs, in which the final output is given by an unweighted average of the individual predictions given by each MLP. Since smaller synaptic weights may result in better generalization for the trained networks, we further optimize the weight decay regularization parameter [45].

#### 4.3.3. Support vector regression (SVR)

The support vector regression (SVR) [46] is a powerful learning method resulting from an adaptation of support vector machines [47] to cope with nonlinear regression problems. Suppose we are given  $n$  training observations in a set  $G = \{(\mathbf{x}_i, y_i)\}_{i=1}^n \subset \mathbb{R}^m \times \mathbb{R}$ , where  $\mathbf{x}_i$  is a  $m$ -dimensional vector of input features and  $y_i$  is the corresponding target output. Following common practice [39], we consider the  $\varepsilon$ -SVR model, in which the ultimate purpose is to find a suitable and flat function  $f(\mathbf{x})$  that deviates at most by  $\varepsilon$  from the target outputs  $y_i$  for all training data (see [46] for details). For the SVR training, the **rminer** R-package adopts the Sequential Minimal Optimization (SMO) algorithm [48]. The SVR formulation is also typically dependent on a kernel function  $\kappa(\mathbf{x}, \mathbf{x}')$ . Here, we tried the polynomial kernel  $\kappa(\mathbf{x}, \mathbf{x}') = P(\langle \mathbf{x}, \mathbf{x}' \rangle)$ , where  $P(\cdot)$  is an arbitrary polynomial with positive coefficients, and the common Gaussian kernel (or the Radial Basis Function (RBF))  $\kappa(\mathbf{x}, \mathbf{x}') = \exp(-\|\mathbf{x} - \mathbf{x}'\|^2 / 2\sigma^2)$  since it allows to generate an infinite dimensional feature space only depending on  $\sigma > 0$ , which denotes the width of the kernel. In the case of the popular Gaussian kernel, particular attention should be given to the selection of the hyperparameters  $\sigma, \varepsilon$  and  $C$ , which specifies the trade-off between model fit and the flatness of the mapping. Yet, special emphasis is placed in the first hyperparameter,  $\sigma$ , as it is the one with the greatest potential impact on the performance of  $\varepsilon$ -SVR [49]. For a better generalization, such a selection should be conducted in a reduced input space by choosing suitable time lags to feed the  $\varepsilon$ -SVR model [39].

Overall, due to the existence of a single global minimum and its ability to build flexible and nonlinear regression estimating functions, SVR has proven to perform well in a wide variety of regression datasets [50], including in time series forecasting [51].

#### 4.3.4. Random forest (RF)

Developed by Breiman [52], a random forest is a non-parametric model consisting of an ensemble of randomly generated decision trees. The resulting forest predictions are based on a given aggregation method, which depends on the nature of the problem, i.e., classification or regression. In random forest regression, the forest can be algebraically defined as an ensemble of  $R$  trees  $\{T_1(\mathbf{x}), T_2(\mathbf{x}), \dots, T_R(\mathbf{x})\}$  that produces  $R$  individual predictions  $\{\hat{Y}_1 = T_1(\mathbf{x}), \hat{Y}_2 = T_2(\mathbf{x}), \dots, \hat{Y}_R = T_R(\mathbf{x})\}$ , where  $\hat{Y}_i : i = 1, \dots, R$ , is the prediction derived by the  $i$ th decision tree. During the training process, in which the ultimate goal is to generate multiple de-correlated trees, each decision tree of the ensemble is grown using distinct bootstrap samples with replacement from the original training set,

under the random subspace method [53]. For each bootstrap sample, each tree is grown by choosing, at each node, the best split among a given subset of random features. The resulting forest prediction is then formed by averaging the individual predictions derived from the different  $R$  regression trees using bootstrap aggregation (or bagging, in [54]), which minimizes the variance related to prediction. This is particularly important in order to reduce the propensity of overfitting, which is a serious handicap of decision trees. Bootstrap aggregation also generates de-correlated trees from different training samples, giving to the RF an increased tolerance to noise.

By definition, RF has several hyperparameters. Yet, two of them are deemed relevant [55] to achieve satisfactory results, namely the number of trees ( $R$ ) and the number of random features ( $m_{try}$ ) considered at each split in the forest.

## 5. Empirical evaluation

This Section presents and describes the research methodology carried out in order to (i) empirically evaluate the merits of including multivariate information in the manufacturer’s demand forecasting process and (ii) examine the performance of the derived forecasts, from both a statistical accuracy and supply chain perspective. For this purpose, we adopted a case study approach as it is well-suited to research topics for which relatively little work has been conducted [56]. In what follows, we rely on [57] to outline the research design as well as to describe the underlying research question and methodologies considered in the evaluation part of this study.

### 5.1. Case study design and unit of analysis

Our research setting is the logistics department of Bosch Automotive Electronics Portugal (AE/P). Although fairly profitable when compared with global competitors, Bosch AE/P tends to keep high inventories and safety stocks of components upholding the highest standards of service level. While it is desirable to maintain a high customer service level, it is likewise vital to minimize inventory holding costs. Currently, manufacturer’s demand for components is determined using the classical MRP concept. Complementarily, even though less frequently, univariate forecasting techniques can also be used for this purpose. In the former context, end-customer demand forecasts form the basis for the determination of gross requirements which, in turn, are used to determine the net requirements for components over time. The supplied components are further assembled by the manufacturer

in order to produce finished products. However, the company constantly faces increases, reductions, cancelations and forward-backward movements of finished product orders by the end-customers. This, together with the fact that the maintenance of MRP system parameters is not always efficient, makes it difficult to find suitable supply orders for components. After several face-to-face meetings with several senior experts from the logistics department of Bosch AE/P, we joined forces to study a strategy to improve manufacturer's demand forecasting.

The investigated case examines the dynamics of our multivariate demand forecasting approach throughout the component's life-cycle. At this point, we consider the evolution of the forecasting and supply chain performance derived from our approach as the core unit of analysis of this study.

#### *5.1.1. Research question and hypothesis*

Our experimental study addresses the following research question: To what extent can the introduction of multivariate information improve upon the performance of univariate forecasting, inventory and operational management?

Here, we hypothesized that our multivariate forecasting approach can help us to accurately predict manufacturer's demand throughout the different component's life-cycle stages, which are likely to impact on demand forecast accuracy. If this hypothesis is accepted, the manufacturer can set more adequate empirical safety stocks across the component's life-cycle, leading to a better management of demand variability. In addition, we expect that the smaller forecast deviations from true manufacturer's demand resulting from the application of our approach will provide an opportunity to improve customer service levels and inventory management by better matching manufacturer's demand and supply. In practice, this will enable to foster the buyer-supplier relationship by reducing the bullwhip effect, which govern how good the manufacturer can cope with future downstream needs.

#### *5.1.2. Data collection and analysis*

We considered empirical data, covering the period from the year 2008 up until 2016, related with three procurement components identified by the company experts as key items with a high turnover rate. The data were collected from an enterprise resource planning (ERP) system. Following the forecasting framework described in Section 3, the collected data for every component include weekly information related with each input feature (4

time series per component). As the manufacturer’s demand in the week 52 is relatively small in all components, we aggregated the data of week 52 into week 51, resulting in time series with 459 samples (51 weeks  $\times$  9 years).

For illustration, Fig. 3 depicts the weekly manufacturer’s demand ( $D$ ) for each component together with the leading indicator  $P$ . Each point in the  $P$  curve was collected 10-weeks ahead of the realization of the manufacturer’s demand for components ( $D$ ). From the zoomed parts of this figure, one may argue that weeks-ahead estimations for future finished product production volume ( $P$ ), resulting from field intelligence information, can serve as an interesting leading indicator for actual demand  $D$ . The time series are found to be seasonal with period 51, as verified by the Friedman’s test [58] ( $p < 0.05$ ), and non-stationary by the Augmented Dickey-Fuller (ADF) test [59]. Note that, for the *component 3*, the time series  $D$  and  $P$  are quite related in terms of magnitude because it is only used once in the product structure of all finished products  $p_i \in F$ . In contrast, for the remaining components, we see a difference in magnitude between the curves  $D$  and  $P$  inasmuch as one component can be used more than once in the different  $p_i \in F$ . The unimodal curve associated with each of the time series can be explained by the three component’s life-cycle stages. First is the launch stage, typically characterized by a growing demand for components in alignment of the increasing sales of finished products. After reaching its peak, the demand becomes relatively stable for a finite time period called the maturity stage. Entering into the end-of-life (EOL) stage, the demand for components gradually falls back towards zero as more finished products reach their lifetime.

Although there is a relatively small sample size of components in this study, the respective demand time series exhibit several demand spikes that, without causal information, are challenging to predict. This makes such time series interesting for forecasting purposes.

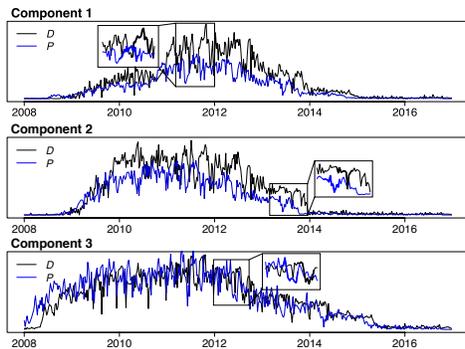


Figure 3: Planned production orders from field intelligence information ( $P$ ) as a leading indicator of weekly manufacturer’s demand ( $D$ ) for the different components, from 2008 up until 2016.

## 5.2. Forecasting design and evaluation

To measure the performance of the forecasting models, we devised a realistic rolling origin forecasting scheme [40] in such a way that the initial training set gradually increases to generate forecasts for  $h = 1$  to  $h = H$  periods ahead from several forecasting origins, which roll forward in time over the component's life-cycle. This setup goes as follows. In the first iteration of the scheme, the oldest  $W$  data records are used to fit the prediction model constructed in an *ex ante* fashion. The trained model then produces the forecasts for the next  $H$  periods, starting from  $t = W + 1$ . In the second iteration, the training set  $W$  is increased by  $K$  periods and the forecasting model is retrained to generate new  $H$  forecasts, starting from  $t = W + K + 1$ . Following this process, the training window  $W$  grows up to  $t = S - H$ , where  $S$  is the available sample size. This rolling origin setup is robust since it allows to compare several forecasts over different training and test sets.

In such a setting, we now discuss the evaluation measures used to assess the performance of the forecasts. We have considered forecasting and supply chain performance measures, computed separately for each horizon  $h \in \{1, \dots, H\}$ , and aggregated over  $n$  forecast origins.

### 5.2.1. Measuring forecasting performance

We used the Normalized Mean Absolute Error (NMAE) as a measure to statistically evaluate and compare the predictions derived from the application of the models:

$$\begin{aligned}
 NMAE_h &= \frac{MAE_h}{Y_{max} - Y_{min}} = \frac{1}{n} \sum_{i=1}^n \frac{|Y_{T+h}^{(i)} - \hat{Y}_{T+h}^{(i)}|}{Y_{max} - Y_{min}} \\
 MAE_h &= \frac{1}{n} \sum_{i=1}^n |Y_{T+h}^{(i)} - \hat{Y}_{T+h}^{(i)}|,
 \end{aligned} \tag{3}$$

where  $T$  is the last known time period for the rolling iteration  $i$ ,  $Y_{T+h}^{(i)}$  is the actual demand at time  $T+h$  for  $i$ ,  $\hat{Y}_{T+h}^{(i)}$  is the target demand, whereas  $Y_{max}$  and  $Y_{min}$  represent the maximum and the minimum target values of the test set. The NMAE is easy to interpret, as it expresses the error as a percentage of the range of the target values, and it has desirable statistical properties. First, in sharp contrast with the MAE, it is scale independent, which allows to compare forecasting errors for data with different ranges of values. Secondly, contrary to other absolute error measures (e.g., Mean Absolute Percentage Error), NMAE is able to cope with zero target values or with aggregation of single errors that could be zero.

### 5.2.2. Measuring supply chain performance

Recalling that, depending on the component's nature, different magnitudes of forecast errors translate into different implications for the organization (e.g., in terms of production planning and service level), it is also important to consider the potential impacts of the forecast errors in the SCM process. Focusing on inventory management, practitioners often concern about the standard deviation of the forecast errors during the lead time  $L^2$  when establishing appropriate safety stock levels to hedge against stock-outs and demand uncertainty. Therefore, we studied the implications of forecast deviations on the dimensioning of safety stocks as a function of the target service level  $\alpha^*$ . For that, we adopted the well-known formulation for the classical safety stock dimensioning problem [60]:

$$SS = \Phi^{-1}(\alpha^*)\sigma_L, \quad (4)$$

where  $\Phi(\cdot)$  is the standard Gaussian cumulative distribution function and  $\sigma_L$  is the standard deviation of forecast errors over a given lead time  $L > 0$ . Following common practice [61], we considered an empirical estimation where  $\sigma_L$  can be directly estimated from the lead time forecast errors according to the following parametric approach:

$$\sigma_L = \sqrt{\frac{\sum_{j=1}^N [\epsilon_j(L) - \bar{\epsilon}(L)]^2}{N}}, \quad (5)$$

where  $\epsilon_j(L) = \sum_{h=1}^L [Y_{T+h}^{(j)} - \widehat{Y}_{T+h}^{(j)}]$  is the lead time forecast error,  $\bar{\epsilon}(L) = \frac{1}{N} \sum_{j=1}^N \epsilon_j(L)$  is the average error for the  $L$  under consideration, and  $N$  is the number of forecasts of length  $L$  considered. Instead of assuming that forecast errors are independent and identically distributed (i.i.d.), as in the theoretical approach, the above-mentioned empirical estimation relaxes the independent variance assumption by allowing  $\sigma_L$  to vary over time.

While the results of most research do not include the financial impacts of forecasting errors in the SC and, when they do so, the case study approach is typically not considered [1], we adopted a cost function to measure the expected inventory-related costs ( $TC$ ), in terms of holding and backlog inventory, induced by a given forecasting deviation at  $h$ :

$$TC_h = \frac{1}{n} \sum_{i=1}^n \left[ c_k \max(\widehat{Y}_{T+h}^{(i)} - Y_{T+h}^{(i)}, 0) + c_b \max(Y_{T+h}^{(i)} - \widehat{Y}_{T+h}^{(i)}, 0) \right] \quad (6)$$

---

<sup>2</sup>We define *lead time* as the moment between a component order is placed by the manufacturer to a supplier and the moment at which the components actually arrive to the plant.

where  $c_k$  and  $c_b$  represent the fixed weight factors for holding and backlog costs, respectively, of a given component. The backlog costs cover all the premium freight costs incurred to face stock-out risks. Note that Eq. (6) allows to reflect the potential effects caused by forecast deviations in terms of inventory holding costs, when  $\widehat{Y}_{T+h}^{(i)} > Y_{T+h}^{(i)}$ , and in terms of backlog costs when  $\widehat{Y}_{T+h}^{(i)} < Y_{T+h}^{(i)}$ . As in [62], we have assumed that holding and backlog costs evolve linearly as a function of the inventory-on-hand and backlog levels, respectively. Of interest, we also examined both the loss rates ( $LR$ ) and fill rates ( $FR$ ) generated by the different forecasting models as follows:

$$LR_h = \frac{1}{n} \sum_{i=1}^n \frac{\max(\widehat{Y}_{T+h}^{(i)} - Y_{T+h}^{(i)}, 0)}{Y_{T+h}^{(i)}} \quad (7)$$

$$FR_h = 1 - \frac{1}{n} \sum_{i=1}^n \frac{\max(Y_{T+h}^{(i)} - \widehat{Y}_{T+h}^{(i)}, 0)}{Y_{T+h}^{(i)}} \quad (8)$$

where  $LR$  represents the fraction of components that need to be discarded over the actual manufacturer's demand due to an overestimated forecast, and  $FR$  traduces the fraction of demand that is fulfilled by the forecast [34].

### 5.3. Baseline & benchmark models

We start by comparing the forecast results with those of three traditional univariate benchmark models, including the Naïve (random walk), Theta and ARIMA, commonly adopted in researches on demand forecasting based on the M3 competition data [62]. The Naïve method is a simple yet fundamental benchmark for forecasting, in such a way that any other forecasting model should outperform it to ensure its appropriate use. In this method, the forecast for the next period equals the last observed value. The Theta method [63] is a decomposition approach that modifies the local curvature of a time series, leading to the creation of two or more Theta-lines that are extrapolated and combined to capture the short and long-term dynamics of the original data. For additional benchmark purposes, we have also considered the well-established Elman Recurrent Neural Network (RNN) model [64] (ERNN) and a recently Automated ML (AutoML [65]) model that is implemented in the **rminer** R-package. The ERNN model presents a similar structure to the MLP but the former contains local feedback recurrent connections allowing storage and the use of past output information to forecast future values. The AutoML is particularly appealing for non expert ML users, since it automatically performs a ML model selection

and its hyperparameter optimization. In particular, the adopted AutoML model automatically trains and tunes a Generalized Linear Model (GLM), an eXtreme Gradient Boosting (XGBoost) model, and a Stacked Ensemble (SE) method, in addition to the SVM, MLP and RF. The SE method (or stacked regression, in [66]) combines the predictions of the above-mentioned base learners by using a second-level ML algorithm. The AutoML model then selects the best performing method on the validation set to be used on unseen data. It is noteworthy that, so far, the literature is quite scarce when it comes to applications of AutoML to time series forecasting.

Since the company considered in this study takes advantage of the ARIMA model as one of its forecasting methodologies, we used it as a separate baseline to compare the performance of our multivariate regression models.

#### 5.4. Modeling setup

Based on business intuitive knowledge, we have restricted the time lags to be of an order less or equal to 15 weeks ( $k \leq 15$ ), a period considered to be sufficiently long to grasp demand patterns by the company experts. The section managers of the case company are interested in short to mid-term forecasts over the entire component’s life-cycle. Thus, we distinguished the life-cycle stages and respective duration for the different components. We considered the final 30 weeks of each component’s life-cycle stage as test set to execute three rolling origin forecasting iterations with  $H = K = 10$  as described in Section 5.2. This yields to three forecasting origins for each life-cycle stage, and therefore nine model updates for each component. In such a setting, we considered a number of different training window sizes ( $W$ ) for each component (Table 2), since the length of the life-cycle stages varies from one component to another. For each iteration of the rolling origin scheme, the training set is further split into training and validation subsets in a timely ordered fashion [40]. Here, we consider the last 10 values of each training set as the validation subset. Each training set is used to identify which of the time lags for the different features have a considerable influence on the manufacturer’s demand. For that, we adopted a forward- and backward-stepwise regression based feature selection procedure [45]. On the other hand, each validation subset was used for tuning the hyperparameters of the regression models, in which the lowest MAE hyperparameter configuration is selected.

We set the number of MLP architectures used in the NN ensemble to  $N_r = 7$ , which is also adopted in other multi-step forecasting experiments [69]. Each MLP in the ensemble follows the architecture described in Section 4.3.2 and was trained with 100 epochs of the

BFGS algorithm. A grid search was performed to determine the weight decay and the number of nodes in the hidden layer. Building on [70], we sequenced over all combinations of  $decay \in \{10^{-5}, 10^{-4}, \dots, 10^{-1}\}$  and  $\mathcal{H} \in \{1, 2, \dots, 10\}$ . In the case of SVR, the hyperparameters were searched over the hyperrectangle  $[\sigma_{min}, \sigma_{max}] \times [C_{min}, C_{max}] \times [\varepsilon_{min}, \varepsilon_{max}]$ , where  $\sigma_{min} = 2^{-8}, \sigma_{max} = 2^0, C_{min} = 2^{-1}, C_{max} = 2^5, \varepsilon_{min} = 2^{-9}, \varepsilon_{max} = 2^{-1}$ . Regarding the RF hyperparameters, we set the number of trees,  $R$ , for each model using a grid search from 10 to 500 in increments of 25. The number of random features ( $m_{try}$ ) considered at each split in the forest was left at the default value of  $m_{try} = m/3$  for regression, where  $m$  is the number of input features. Finally, the ERNN model was constructed with the **R-RSNNs** package [71] using the same MLP inputs and output node, and a hidden layer with 4 nodes (other values were tested, such as 6 or 8 hidden nodes, leading to similar results), trained with 1000 epochs of a backpropagation with momentum algorithm. Once the best hyperparameters were obtained for the different algorithms, the final model performance was derived by retraining the model with all training data and applying them to the out-of-sample evaluation set, which contains examples that remained unseen throughout the whole process of model validation.

Prior to starting the fitting of the forecasting models, the data were also carefully normalized by mapping them to the range  $[0, 1]$  using maximum and minimum values computed over the training data only. This scaling is necessary to ensure an efficient training as well as to smoothen the numerical convergence [45]. The resulting model predictions are then post-processed with the inverse of the standardized function.

Table 2: Experimental settings of the rolling origin forecasting setup.

Dataset	Training window sizes ( $W$ )
<i>Component 1</i>	$W \in \{108, 118, 128, 313, 323, 333, 414, 424, 434\}$
<i>Component 2</i>	$W \in \{75, 85, 95, 253, 263, 273, 414, 424, 434\}$
<i>Component 3</i>	$W \in \{79, 89, 99, 301, 311, 321, 414, 424, 434\}$

## 6. Results and discussion

### 6.1. Performance over the life-cycle stages

We start by distinguishing the forecasting performance evaluation across the entire component’s life-cycle. Table 3 presents the results for the different forecasting models, with respect to %NMAE, at the different life-cycle stages. The forecasting errors are averaged over all time series, forecast origins and horizons.

Table 3: Forecasting performance evaluation (expressed in %NMAE) throughout the life-cycle stages (best values are highlighted in boldface).

Life-cycle stage	Multivariate information						Univariate information		
	MLP	RF	SVR	AutoML	ERNN	ARIMAX	Naïve	ARIMA	Theta
Launch	15.36	19.24	15.49	17.17	15.69	<b>13.02</b>	31.99	20.05	19.55
Maturity	<b>11.32</b>	14.25	12.30	14.14	11.76	13.50	26.76	22.13	22.70
EOL	<b>11.82</b>	13.90	13.83	13.11	18.22	15.57	36.62	22.94	24.47

The results show that, whatever the life-cycle stage, models with multivariate information generate the lowest forecasting errors and outperform the univariate benchmark models in terms of forecast accuracy, including the company benchmark ARIMA. Interestingly, we found that ARIMAX outperforms ML-based models at the launch stage. Yet, its forecasting performance tends to worsen throughout the component’s life-cycle. We argue that the underperformance of ML-based forecasts at early life-cycle stages may be due to the lack of sufficient training data that hinders a proper generalization capability of the models to unseen data. A generalized deterioration of forecasting performance at later life-cycle stages is also observed for the benchmark pure statistical models, in sharp contrast with the results of ML-based models. Of note, multivariate forecasting methods generally exhibit a good overall forecast accuracy at the EOL stage. This finding is particularly meaningful if one consider that such stage depends on an increased forecast accuracy to minimize the risk of overstock and the resulting obsolescence of discontinued components when their demand decreases faster towards the end of the life-cycle.

To probe deeper into the importance of the leading indicators across the life-cycle stages, we further conducted a comprehensive ablation analysis (Table 4). We shall first note that the forecasting results generally degrade from the full model, which includes all the leading indicators, to the simple univariate framework. Recall that the full ARIMAX model performs better than ML-based forecasters at the launch stage (Table 3). Here, the ablation results provide further evidence that the ARIMAX forecasting performance improves when using only the inputs  $D$  and  $P$ . This holds for the remaining life-cycle periods. As such, when forecasting at early life-cycle stages (with necessarily less training data), we favor the use of a more simple statistical-based approach, rather than nonlinear complex ML-methods. Nevertheless, as we move from the launch to the maturity and EOL stages, the ML-based models trained with all indicators start to outperform ARIMAX.

Table 4: Ablation analysis of influence of feature interactions on the forecasting performance for the different multivariate models across life-cycle stages. The last column shows the average (Avg) %NMAE values (best values in boldface) for each interaction and the ranking within the column.

Launch stage, for 1–10 weeks ahead							
Feature interactions	MLP	RF	SVR	AutoML	ERNN	ARIMAX	Avg (rank)
$\#F + U + D + P$	<b>15.36</b>	<b>19.24</b>	<b>15.49</b>	<b>17.17</b>	<b>15.69</b>	13.02	<b>16.00</b> (1)
$\#F + U + D$	34.69	24.59	25.73	30.22	24.06	23.69	27.16 (3)
$D + P$	16.29	20.94	16.52	18.28	17.02	<b>11.50</b>	16.76 (2)
$D$	27.51	26.47	28.19	26.71	31.42	25.27	27.59 (4)
Maturity stage, for 1–10 weeks ahead							
Feature interactions	MLP	RF	SVR	AutoML	ERNN	ARIMAX	Avg (rank)
$\#F + U + D + P$	<b>11.32</b>	<b>14.25</b>	<b>12.30</b>	14.14	<b>11.76</b>	13.50	<b>12.88</b> (1)
$\#F + U + D$	26.45	25.72	23.31	27.19	23.29	26.70	25.44 (3)
$D + P$	12.35	16.32	12.87	<b>13.44</b>	14.92	<b>13.11</b>	13.84 (2)
$D$	29.06	30.13	28.00	26.73	26.81	25.72	27.74 (4)
EOL stage, for 1–10 weeks ahead							
Feature interactions	MLP	RF	SVR	AutoML	ERNN	ARIMAX	Avg (rank)
$\#F + U + D + P$	<b>11.82</b>	<b>13.90</b>	<b>13.83</b>	<b>13.11</b>	<b>18.22</b>	15.57	<b>14.41</b> (1)
$\#F + U + D$	25.98	22.68	20.37	21.13	22.93	30.41	23.91 (3)
$D + P$	14.74	14.73	14.53	14.44	20.59	<b>14.93</b>	15.66 (2)
$D$	29.49	31.52	25.75	24.81	21.00	24.28	26.14 (4)

A comparison between the results obtained with and without the indicator  $P$  clearly reveals the sizable impact of that indicator on the general improvement of forecasting performance over the component’s life-cycle. At the same time, while multivariate models based merely on the indicators  $\#F$  and  $U$  tend to perform close to (or even worse than) univariate benchmarks, it is interesting to observe that the combination of  $\#F$  and  $U$  with the indicator  $P$  yields a notable additive forecasting performance effect in all models, whatever the underlying life-cycle stage. When compared to pure univariate benchmarks, the underperformance of multivariate forecasting models using only  $\#F, U$  and  $D$  is essentially motivated by two reasons: (i) the leading indicators  $\#F, U$  contain only product structure information, which may be useful in anticipating the magnitude of manufacturer’s demand but not its variation; (ii) in contrast, despite  $D$  might contain

important information on demand’s variation, it is a lagging (not leading) indicator of manufacturer’s demand behavior. In short, this is a special case of feature engineering [73], in which the indicators  $\#F$  and  $U$  appear to be irrelevant (or weakly relevant) in isolation but relevant in combination with  $P$ . Overall, these results justify the use of the full model in the subsequent analyses.

Let us now discuss the implications of forecasting errors on inventory management, in particular on the dimensioning of safety stocks over the life-cycle. For the sake of brevity, we focused our attention on *Component 1*. Safety stocks were directly calculated from Eq. (4) with  $L = 10$  weeks, since longer lead times amplify the differences in performance among the tested forecasting models, enabling improved comparisons. We highlight that 10-week lead times are reasonable in real-life SCs, especially those operating with transportation services by sea worldwide, as is the case of the case study company. Figure 4 shows the required safety stocks over the different life-cycle stages as the target service level  $\alpha^*$  varies for the multivariate forecasting methods and the company benchmark ARIMA. Naturally, the required levels of safety stock increase with the target service level. Comparison between the results obtained on each life-cycle stage suggest that more safety stocks are required at the initial phase to cope with the forecast error variance. In contrast, the required size of safety stocks tends to decrease towards the end of the life-cycle. Such considerations are strictly related with the magnitude of the forecast errors observed in each life-cycle stage. Overall, the good performance of the multivariate forecasting models potentially translates into lower safety stock levels and, in the limit, lower holding costs when compared to current forecasting strategy adopted by the case study company. The above-mentioned results hold for the remaining components.

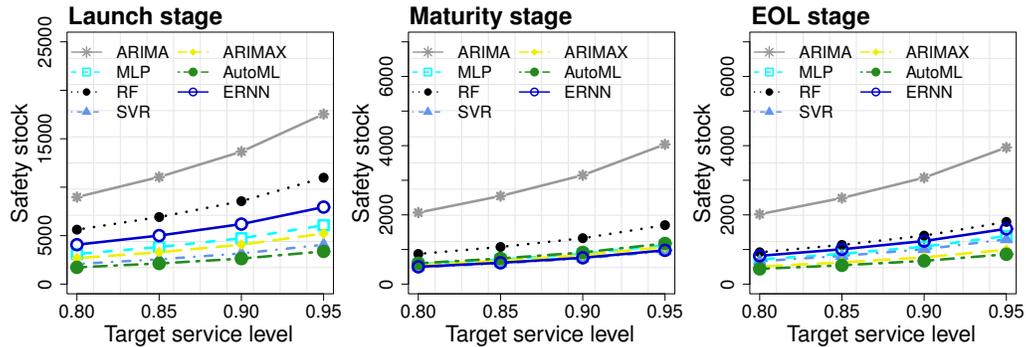


Figure 4: Required safety stocks as a function of target service level for *Component 1* and  $L = 10$ .

## 6.2. Overall forecasting performance

To examine the overall forecasting performance of the tested models as a function of  $h$ , the nine individual %NMAE values obtained per horizon across the different life-cycle stages of each component are now aggregated vertically to derive an estimated median error with corresponding 95% confidence intervals via the non-parametric Wilcoxon signed-rank test [58]. Table 5 presents the forecasting performance of each model across the forecast horizons, together with its ranking in terms of average %NMAE.

Table 5: Forecast accuracy (expressed in %NMAE) across forecast horizons (best values are highlighted in boldface; Avg – denotes the average). The numbers in round brackets represent the model rank in terms of average forecasting performance.

%NMAE for $h$ :	1	2	3	4	5	6	7	8	9	10	Avg
MLP (1)	15.54	<b>9.48</b>	<b>11.77</b>	9.46	<b>9.50</b>	<b>8.51</b>	<b>9.79</b>	12.66	13.78	20.37	<b>12.09*</b>
RF (6)	17.13	13.27	11.88	11.92	14.48	13.65	14.46	12.88	14.58	27.03	15.13
SVR (2)	17.69	10.31	12.34	<b>8.93</b>	12.93	13.04	13.53	12.33	<b>11.28</b>	19.68	13.21
AutoML (4)	16.52	10.52	12.65	9.83	15.57	10.38	11.86	13.96	14.45	23.65	13.94
ERNN (5)	<b>14.60</b>	12.64	16.19	10.31	14.59	12.96	13.97	<b>11.99</b>	15.08	24.17	14.65
ARIMAX (3)	19.62	13.13	12.42	9.63	10.68	11.07	11.98	13.83	12.71	<b>17.45</b>	13.25
Naïve (9)	27.76	30.24	25.82	23.73	31.71	30.99	41.12	31.34	32.94	31.71	30.74
ARIMA (7)	18.53	21.79	19.30	15.96	18.45	19.12	26.11	21.49	18.93	32.94	21.26
Theta (8)	19.45	22.23	16.40	17.86	20.74	22.03	29.37	20.42	20.12	29.92	21.85

\*Statistically significant when compared with all univariate benchmark models at the 95% significance level.

Comparing the different forecasting models, MLP is ranked first in terms of forecasting performance, followed by SVR, ARIMAX and AutoML. The ERNN is among the less accurate ML-based methods. This is not surprising, as a recent study of Makridakis et al. [25] using a large subset of time series used in the M3 Competition has already shown that more advanced ML-based methods, such as RNN, do not necessarily guarantee enhanced forecasting performance. For all horizons, our results demonstrated that demand forecasting is enhanced whenever forecasting models are built on multivariate information rather than univariate information (Fig. 5). In terms of overall forecast accuracy, we found that multivariate models outperform (with statistically significant differences) all the univariate benchmark models. A possible reason for the superiority of multivariate models over univariate techniques might be related to the existence of several demand spikes in the different time series that are challenging to predict with just univariate information. Yet, we found no statistical evidence that the overall differences between the forecasts generated by the different multivariate models are significant over horizons.

In light of the above considerations, the statistically significant difference between the forecasting performance in the multivariate and univariate frameworks supports our hypothesis about the role played by our multivariate approach in improving the manufacturer’s demand forecasting process across the component’s life-cycle.

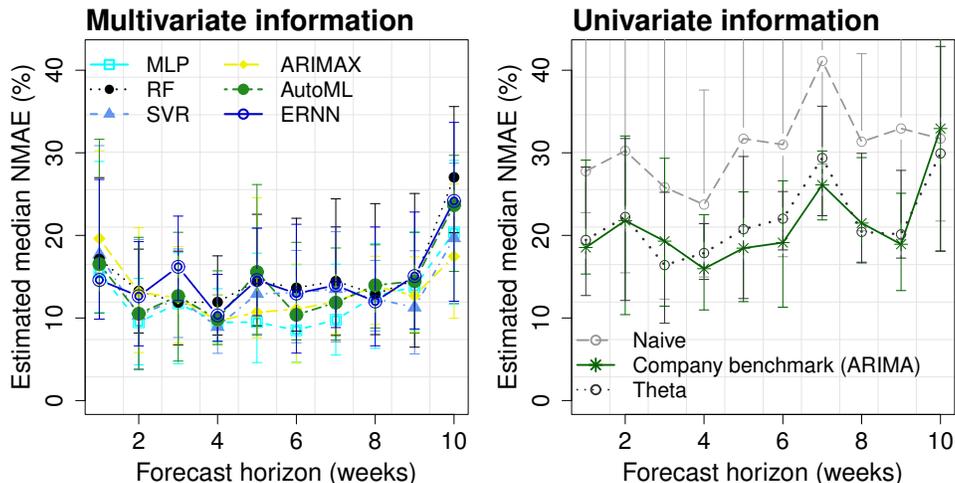


Figure 5: Forecast accuracy obtained using multivariate (left panel) and univariate information (right panel). The whiskers represent 95% confidence intervals for each Wilcoxon median value.

### 6.3. Supply chain performance

Focusing on the forecasting errors only, multivariate models seem to outperform univariate benchmark models. Nevertheless, when choosing a suitable forecasting method, the impacts of overestimated and underestimated forecasts on the SC performance should not be overlooked either. To quantify these impacts, we have considered three evaluation measures: inventory-related costs ( $TC$ ), loss rate ( $LR$ ) and fill rate ( $FR$ ). Likewise, the calculation of each measure follows the same aggregation strategy across forecasts used for examine the overall forecasting performance of the different models. Table 6 presents the  $TC$  (in €), formulated as in Eq. 6, produced by each forecasting method over all horizons. For the sake of business confidentiality, the values of the factors  $c_k$  and  $c_b$  considered for each component are omitted.

The results show that regardless the forecasting model, the inventory-related costs generally increase whenever the forecast horizon is increased, and the models with multivariate information consistently perform best in the sense of minimizing inventory-related costs over all horizons. Interestingly, ERNN is among the bottom three methods in terms

Table 6: Inventory-related costs across horizons. The numbers in round brackets represent the model rank in terms of average (Avg) total cost performance (best values are highlighted in boldface).

$TC$ for $h$ :	1	2	3	4	5	6	7	8	9	10	Avg
MLP (3)	64.83	112.22	100.52	188.68	112.03	185.36	124.37	136.53	171.40	354.27	155.02
RF (4)	78.22	157.71	<b>38.83</b>	95.39	<b>80.40</b>	209.73	<b>51.74</b>	64.45	225.37	587.36	158.92
SVR (2)	97.25	116.11	148.16	<b>61.99</b>	131.87	<b>152.04</b>	137.55	128.11	<b>78.54</b>	<b>268.11</b>	131.97
AutoML (6)	62.46	213.24	216.30	277.95	262.94	225.11	218.74	126.91	207.54	418.31	222.95
ERNN (1)	<b>61.80</b>	<b>102.89</b>	131.25	87.10	105.08	181.24	91.32	<b>60.73</b>	98.25	304.71	<b>122.44</b>
ARIMAX (5)	204.31	150.07	178.45	156.81	258.40	276.90	209.31	181.74	159.13	388.04	216.32
Naïve (7)	187.47	201.39	194.00	164.23	247.47	248.95	256.70	198.30	219.51	583.68	250.17
ARIMA (9)	79.32	328.79	262.02	272.58	308.60	489.05	378.49	279.99	300.96	778.34	347.81
Theta (8)	109.76	270.05	179.63	181.92	208.88	318.01	273.82	252.66	257.03	804.52	285.63

of %NMAE, but it is ranked first in terms of averaged  $TC$ . This is particularly due to *component 2*, for which ERNN forecasts tend to overestimate demand over several rolling origins and  $c_k$  is residual. As, by definition,  $TC$  depend heavily on the factors  $c_k$  and  $c_b$ , the potential disagreement between forecasting performance and inventory-related costs can be easily explained by the magnitude of these factors for the different components. By way of example, for a sufficiently small factor  $c_k$  of a particular component, if  $c_k \ll c_b$  and the forecasts tend to overestimate demand over several rolling origins, then the  $TC$  derived therefrom tend to be low. From the reported results, one may also realize that whenever the company benchmark ARIMA is adopted the  $TC$  induced by forecast deviations from the actual demand values are substantially higher than those obtained by using multivariate approaches, and, to a lesser extent, than those derived by Theta and Naïve methods. Notably, a substantial averaged cost reduction is observed whenever the multivariate models are employed. For instance, the adoption of the AutoML model, which is the worst multivariate performing model in terms of  $TC$ , results in 35.9% averaged cost savings over the standard forecasting strategy employed by the case study company.

For added confidence and validation of our results, we studied the SC performance of all models in terms of loss and fill rates (Table 7). The results are clear. The average fraction of demand that is fulfilled by the forecasts tends to be slightly higher whenever multivariate information is considered during the forecasting process. Moreover, compared with all multivariate models, the forecasted demands generated by the company benchmark ARIMA are more often overestimated than underestimated, potentially leading to increased loss rates and holding costs. To a greater extent this is also true for the Naïve

and Theta methods. Overall, MLP is ranked first in terms of averaged loss and fill rates. Besides, it presents the lowest %NMAE and is among the top three methods in terms of  $TC$ , thus supporting the conclusion that it is the best performing model for our data.

Table 7: Averaged loss and fill rates generated by multivariate and univariate forecasting models (Avg: Average; SD: Standard Deviation; best Avg values are highlighted in boldface).

Forecasting approach	Models	Loss rate		Fill rate	
		Avg	SD	Avg	SD
Multivariate	MLP	<b>0.09</b>	0.06	<b>0.90</b>	0.04
	RF	0.13	0.15	0.89	0.02
	SVR	0.11	0.10	<b>0.90</b>	0.02
	AutoML	<b>0.09</b>	0.04	0.89	0.05
	ERNN	0.16	0.09	<b>0.90</b>	0.02
	ARIMAX	0.12	0.08	<b>0.90</b>	0.04
Univariate	Naïve	0.41	0.23	0.77	0.03
	ARIMA	0.24	0.15	0.89	0.03
	Theta	0.27	0.13	0.89	0.03

#### 6.4. Practical & managerial implications

Our study shows that the proposed multivariate ML approach might provide relevant insights to enhance upstream demand forecasting, helping company managers in the sense of improving their complex operational logistic decisions and defining suitable procurement strategies in a data-driven fashion. Of note, according to a Gartner report<sup>3</sup>, decision intelligence is pointed out as a major trend shaping the evolution of digital businesses. For decision-makers of Bosch AE/P, we highlight several managerial implications.

First, our findings have supported the idea that the generation of accurate manufacturer’s demand forecasts at the launch stage of the life-cycle can be a challenging task. Typically, the company is forced to increase the safety stocks at this stage in order to cope with end-customer demand fluctuations. When compared to the traditional univariate benchmark methods, the usage of the proposed forecasting strategy is able to address these issues by providing improved demand forecasts throughout the entire component’s life-cycle. In particular, when less training data are available, we argue in favor of the

<sup>3</sup>Gartner (2019). Five Major Trends Shaping the Evolution of Analytics and Business Intelligence. URL:<https://www.gartner.com/en/newsroom/press-releases/2019-10-02-gartner-reveals-five-major-trends-shaping-the-evoluti>, last accessed on September 7, 2020.

adoption of ARIMAX with the indicator  $P$ , as it increases our ability to produce better forecasts than those derived from more complex nonlinear models. As a result, accurate forecasts lead to smoother safety stocks and decreased inventory-related costs. The reasonably good performance of the multivariate models at the EOL stage also confirms the opportunity to minimize the holding costs associated with the overstock of discontinued components. In any case, we strongly suggest an increase of collaboration efforts between downstream and upstream SC players to minimize major demand signal distortions.

Second, of note, a potential advantage of the proposed approach over traditional methodologies is that sharp demand turning points captured by the leading indicators are included in the forecasting models. Therefore, we believe that the smaller forecast deviations from true manufacturer's demand resulting from the application of our approach would provide an opportunity to improve inventory service levels, by better matching manufacturer's demand and supply. This reasoning is especially relevant in the automotive industry, in which holding costs are typically high. Moreover, in Vendor-Managed Inventory (VMI) settings, enhanced manufacturer's demand derived from the adoption of our approach could be shared with suppliers, leading to improved collaboration among upstream supply chain players. This is not the case with indicators  $\#F$  and  $U$ , which given the confidential nature of product structure information are not, in general, shared with traditional suppliers with standard contracts with the manufacturer. Yet, in the presence of information exchange or, ideally, synchronized SC collaboration schemes [74] with strategic suppliers, there may be opportunities to share product information towards the improvement of forecasting performance [75]. Such sharing would allow, for instance, improved anticipation of demand drops towards the ramp down stage of finished products that make use of a given component.

However, it is worth pointing out that since ML-based algorithms have not been fully explored in the context of SCM [10], company managers need to understand the complexities inherent to their application in real-life environments with multiple components, as well as to identify the resources with the necessary skills to successfully implement them in a productive system. Bridging this gap is critical for the overall success of the proposed forecasting initiative. Finally, it is also noteworthy that company experts should play a fundamental role in the definition and maintenance of a potential decision support system. Their business sensitivity to the market dynamics ought to be considered as an important factor to re-training the models with new relevant information for the forecasting process.

Future research may develop in (i) finding different strategies to acquire other relevant inputs from domain experts able to support the demand forecasting process, (ii) evaluating the application of transfer learning models for predicting new component’s demand, (iii) trying different ways to leverage different types of time series features (e.g., via automated feature engineering [72]), and (iv) extending this work to include a broader range of components, which allow us to obtain more representative results and a more comprehensive comparison between statistical and ML-based forecasting methods. We also intend to assess the applicability of the proposed approach in other assembly industries.

## 7. Conclusions

Multi-step demand forecasting is a complex problem with serious repercussions at economic, tactical and operational level in real-life supply chains. Focusing on the upstream-end side, distorted demand signals induced by erratic market information can seriously hamper the proper assignment of component order quantities to suppliers for further production of finished products. We have derived a flexible multivariate approach for enhancing multi-step demand forecasting at the upstream-end side of general supply chains with assembly operations. Rather than only using univariate information, we take advantage of several leading indicators of demand shifts that serve as model inputs to forecast future manufacturer’s demand. Our approach resorts to the statistical ARIMAX model as well as to Machine Learning (ML) models adapted for time series forecasting. Numerical data collected from a major automotive electronics manufacturer (Bosch Automotive Electronics, Portugal) provided context for the proposed forecasting methodology. All the forecasting methods were compared to univariate benchmark models, including the one currently used by the case study company, under a realistic rolling origin forecasting procedure. When forecasting manufacturer’s demand, our results demonstrated that the inclusion of multivariate information provides additional explanatory power above that provided by traditional univariate forecasting techniques. In particular, we found that the proposed approach provides more accurate forecasts than univariate benchmark models across all life-cycle stages (launch, maturity and end-of-life), in addition to generating lower inventory-related costs and loss rates, resulting from smaller forecast deviations. We have also discussed the practical implications of forecast deviations on dimensioning safety stocks across the component’s life-cycle. Our results provided evidence on the usefulness of our approach in improving forecasts at early life-cycle stages, where accurate demand forecasts are more difficult to obtain. Particularly, in the multivariate context, we found

that ARIMAX provided the best forecasts for the launch stage of component’s life-cycle, while the ML-based models produced the most accurate predictions in the remaining two life-cycle stages.

From a practical standpoint, this work may provide a suitable benchmark for logistic decision-makers of general supply chains with assembly operations, especially in contexts where demand is subject to high levels of uncertainty.

## Acknowledgments

This work has been supported by FCT – Fundação para a Ciência e Tecnologia within the R&D Units Project Scope: UIDB/00319/2020. We would like to thank to the three anonymous reviewers for their helpful suggestions. The authors wish to thank Sofia Pereira and Rui Sousa (Bosch Automotive Electronics, Portugal) for useful discussions and technical ideas about the methods proposed in this paper.

## References

- [1] A. Kerkkänen, J. Korpela, J. Huiskonen, Demand forecasting errors in industrial context: Measurement and impacts, *International Journal of Production Economics* 118 (1) (2009) 43–48.
- [2] F. Chen, Z. Drezner, J. K. Ryan, D. Simchi-Levi, Quantifying the bullwhip effect in a simple supply chain: The impact of forecasting, lead times, and information, *Management Science* 46 (3) (2000) 436–443.
- [3] C.-J. Lu, T.-S. Lee, C.-C. Chiu, Financial time series forecasting using independent component analysis and support vector regression, *Decision Support Systems* 47 (2) (2009) 115–125.
- [4] Z. Guo, W. K. Wong, M. Li, A multivariate intelligent decision-making model for retail sales forecasting, *Decision Support Systems* 55 (1) (2013) 247–255.
- [5] C. Li, A. Lim, A greedy aggregation–decomposition method for intermittent demand forecasting in fashion retailing, *European Journal of Operational Research* 269 (3) (2018) 860–869.
- [6] M. A. Villegas, D. J. Pedregal, Supply chain decision support systems based on a novel hierarchical forecasting approach, *Decision Support Systems* 114 (2018) 29–36.
- [7] D. Y. Suh, M. S. Ryerson, Forecast to grow: Aviation demand forecasting in an era of demand uncertainty and optimism bias, *Transportation Research Part E: Logistics and Transportation Review* 128 (2019) 400–416.
- [8] E. Hofmann, E. Rutschmann, Big data analytics and demand forecasting in supply chains: a conceptual analysis, *The International Journal of Logistics Management* 29 (2) (2018) 739–766.
- [9] R. N. Boute, S. M. Disney, M. R. Lambrecht, B. Van Houdt, An integrated production and inventory model to dampen upstream demand variability in the supply chain, *European Journal of Operational Research* 178 (1) (2007) 121–142.
- [10] M. Abolghasemi, E. Beh, G. Tarr, R. Gerlach, Demand forecasting in supply chain: The impact of demand volatility in the presence of promotion, *Computers & Industrial Engineering* 142 (2020) 106380.
- [11] T. R. Willemain, C. N. Smart, H. F. Schwarz, A new approach to forecasting intermittent demand for service parts inventories, *International Journal of Forecasting* 20 (3) (2004) 375–387.
- [12] A. A. Syntetos, J. E. Boylan, The accuracy of intermittent demand estimates, *International Journal of Forecasting* 21 (2) (2005) 303–314.

- [13] F. Petropoulos, K. Nikolopoulos, G. P. Spithourakis, V. Assimakopoulos, Empirical heuristics for improving intermittent demand forecasting, *Industrial Management & Data Systems* 113 (5) (2013) 683–696.
- [14] N. Kourentzes, Intermittent demand forecasts with neural networks, *International Journal of Production Economics* 143 (1) (2013) 198–206.
- [15] W. Fu, C.-F. Chien, Unison data-driven intermittent demand forecast framework to empower supply chain resilience and an empirical study in electronics distribution, *Computers & Industrial Engineering* 135 (2019) 940–949.
- [16] K. Afrin, B. Nepal, L. Monplaisir, A data-driven framework to new product demand prediction: Integrating product differentiation and transfer learning approach, *Expert Systems with Applications* 108 (2018) 246–257.
- [17] J. R. Rego, M. A. Mesquita, Demand forecasting and inventory control: A simulation study on automotive spare parts, *International Journal of Production Economics* 161 (2015) 1–16.
- [18] N. Kourentzes, D. K. Barrow, S. F. Crone, Neural network ensemble operators for time series forecasting, *Expert Systems with Applications* 41 (9) (2014) 4235–4244.
- [19] R. S. Gutierrez, A. O. Solis, S. Mukhopadhyay, Lumpy demand forecasting using neural networks, *International Journal of Production Economics* 111 (2) (2008) 409–420.
- [20] T. Efindigil, S. Önüt, C. Kahraman, A decision support system for demand forecasting with artificial neural networks and neuro-fuzzy models: A comparative analysis, *Expert Systems with Applications* 36 (3) (2009) 6697–6707.
- [21] H. C. Lau, G. T. Ho, Y. Zhao, A demand forecast model using a combination of surrogate data analysis and optimal neural network approach, *Decision Support Systems* 54 (3) (2013) 1404–1416.
- [22] L. Aburto, R. Weber, Improved supply chain management based on hybrid demand forecasts, *Applied Soft Computing* 7 (1) (2007) 136–144.
- [23] S. Jaipuria, S. Mahapatra, An improved demand forecasting method to reduce bullwhip effect in supply chains, *Expert Systems with Applications* 41 (5) (2014) 2395–2408.
- [24] K. I. Nikolopoulos, M. Z. Babai, K. Bozos, Forecasting supply chain sporadic demand with nearest neighbor approaches, *International Journal of Production Economics* 177 (2016) 139–148.
- [25] S. Makridakis, E. Spiliotis, V. Assimakopoulos, Statistical and machine learning forecasting methods: Concerns and ways forward, *PloS ONE* 13 (3) (2018) e0194889.
- [26] S. Makridakis, E. Spiliotis, V. Assimakopoulos, The M4 competition: Results, findings, conclusion and way forward, *International Journal of Forecasting* 34 (4) (2018) 802–808.
- [27] R. Carboneau, K. Laframboise, R. Vahidov, Application of machine learning techniques for supply chain demand forecasting, *European Journal of Operational Research* 184 (3) (2008) 1140–1154.
- [28] A. Mukattash, M. Samhouri, Supply planning improvement: a causal forecasting approach, *Journal of Applied Sciences* 11 (12) (2011) 2207–2214.
- [29] Y. R. Sagaert, E.-H. Aghezzaf, N. Kourentzes, B. Desmet, Tactical sales forecasting using a very large set of macroeconomic indicators, *European Journal of Operational Research* 264 (2) (2018) 558–569.
- [30] Y. R. Sagaert, N. Kourentzes, S. De Vuyst, E.-H. Aghezzaf, B. Desmet, Incorporating macroeconomic leading indicators in tactical capacity planning, *International Journal of Production Economics* 209 (2019) 12–19.
- [31] L. Ferbar, D. Čreslovník, B. Mojskerc, M. Rajgelj, Demand forecasting methods in a supply chain: Smoothing and denoising, *International Journal of Production Economics* 118 (1) (2009) 49–54.
- [32] H. E. Sayed, H. A. Gabbar, S. Miyazaki, A hybrid statistical genetic-based demand forecasting expert system, *Expert Systems with Applications* 36 (9) (2009) 11662–11670.
- [33] P. M. Yelland, Bayesian forecasting of parts demand, *International Journal of Forecasting* 26 (2) (2010) 374–396.
- [34] J. Huber, A. Gossmann, H. Stuckenschmidt, Cluster-based hierarchical demand forecasting for perishable goods, *Expert Systems with Applications* 76 (2017) 140–151.

- [35] P. W. Murray, B. Agard, M. A. Barajas, Forecast of individual customer’s demand from a large and noisy dataset, *Computers & Industrial Engineering* 118 (2018) 33–43.
- [36] T. Vollmann, W. Berry, D. Whybark, *Manufacturing planning and control systems*, 3rd. Edition. Richard D. Irwin, Inc., 1992.
- [37] B. G. Kingsman, *Raw materials purchasing: an operational research approach*, Pergamon Press, Oxford, UK, 1985.
- [38] K. R. Baker, Safety stocks and component commonality, *Journal of Operations Management* 6 (1) (1985) 13–22.
- [39] M. Štěpnička, P. Cortez, J. P. Donate, L. Štěpničková, Forecasting seasonal time series with computational intelligence: On recent methods and the potential of their combinations, *Expert Systems with Applications* 40 (6) (2013) 1981–1992.
- [40] R. J. Hyndman, G. Athanasopoulos, *Forecasting: Principles and Practice*, OTexts, 2013.
- [41] P. Cortez, *Data mining classification and regression methods*. R package version 1.4.6, URL <https://cran.r-project.org/package=rminer>.
- [42] T. R. Core, *R: A language and environment for statistical computing*. R foundation for statistical computing, Vienna, Austria, URL <http://www.R-project.org>.
- [43] R. Hecht-Nielsen, Kolmogorov’s mapping neural network existence theorem, in: *Proceedings of the international conference on Neural Networks*, Vol. 3, IEEE Press New York, 1987, pp. 11–14.
- [44] M. F. Møller, A scaled conjugate gradient algorithm for fast supervised learning, *Neural Networks* 6 (4) (1993) 525–533.
- [45] T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction Inference, and Prediction*, 2nd Edition, Springer, Heidelberg, 2017.
- [46] A. J. Smola, B. Schölkopf, A tutorial on support vector regression, *Statistics and Computing* 14 (3) (2004) 199–222.
- [47] C. Cortes, V. Vapnik, Support-vector networks, *Machine Learning* 20 (3) (1995) 273–297.
- [48] J. Platt, Sequential minimal optimization: A fast algorithm for training support vector machines, *Tech. Rep. MSR-TR-98-14*, Microsoft Research (1998).
- [49] B. Scholkopf, A. J. Smola, *Learning with kernels: support vector machines, regularization, optimization, and beyond*, MIT press, 2001.
- [50] M. Fernandez-Delgado, M. Sirsat, E. Cernadas, S. Alawadi, S. Barro, M. Febrero-Bande, An extensive experimental survey of regression methods, *Neural Networks* 111 (2018) 11–34.
- [51] N. I. Sapankevych, R. Sankar, Time series prediction using support vector machines: a survey, *IEEE Computational Intelligence Magazine* 4 (2) (2009) 24–38.
- [52] L. Breiman, Random forests, *Machine Learning* 45 (1) (2001) 5–32.
- [53] I. Barandiaran, The random subspace method for constructing decision forests, *IEEE Trans. Pattern Anal. Mach. Intell* 20 (8) (1998) 1–22.
- [54] L. Breiman, Bagging predictors, *Machine Learning* 24 (2) (1996) 123–140.
- [55] A. Liaw, M. Wiener, Classification and Regression by RandomForest, *R news* 2 (3) (2002) 18–22.
- [56] I. Benbasat, D. Goldstein, M. Mead, The case research strategy in studies of information systems, *MIS Quarterly* 11 (3) (1987) 369–386.
- [57] P. Runeson, M. Höst, Guidelines for conducting and reporting case study research in software engineering, *Empirical Software Engineering* 14 (2) (2009) 131–164.
- [58] M. Hollander, D. A. Wolfe, *Nonparametric statistical methods*, John Wiley & Sons, 1999.
- [59] G. Elliott, T. J. Rothenberg, J. H. Stock, Efficient tests for an autoregressive unit root, *Econometrica* 64 (4) (1996) 813–836.

- [60] J. N. C. Gonçalves, M. S. Carvalho, P. Cortez, Operations research models and methods for safety stock determination: A review, *Operations Research Perspectives*, 7 (2020) 100164.
- [61] J. R. Trapero, M. Cardos, N. Kourentzes, Empirical safety stock estimation based on kernel and garch models, *Omega* 84 (2019) 199–211.
- [62] F. Petropoulos, X. Wang, S. M. Disney, The inventory performance of forecasting methods: Evidence from the M3 competition data, *International Journal of Forecasting* 35 (1) (2019) 251–265.
- [63] V. Assimakopoulos, K. Nikolopoulos, The theta model: a decomposition approach to forecasting, *International Journal of Forecasting* 16 (4) (2000) 521–530.
- [64] J. L. Elman, Finding structure in time, *Cognitive Science* 14 (2) (1990) 179–211.
- [65] I. Guyon, L. Sun-Hosoya, M. Boullé, H. J. Escalante, S. Escalera, Z. Liu, D. Jajetic, B. Ray, M. Saeed, M. Sebag, A. Statnikov, W.-W. Tu, E. Viegas, *Analysis of the AutoML Challenge Series 2015–2018*, Springer International Publishing, Cham, 2019, pp. 177–219.
- [66] L. Breiman, Stacked regressions, *Machine Learning* 24 (1) (1996) 49–64.
- [67] R. Hyndman, Y. Khandakar, Automatic time series forecasting: The forecast package for R, *Journal of Statistical Software* 27 (1–22).
- [68] R. J. Hyndman, G. Athanasopoulos, C. Bergmeir, G. Caceres, L. Chhay, M. O’Hara-Wild, F. Petropoulos, S. Razbash, E. Wang, F. Yasmeen, *forecast: Forecasting functions for time series and linear models*. R package version 8.11, URL <https://cran.r-project.org/package=forecast>.
- [69] P. Cortez, P. J. Pereira, R. Mendes, Multi-step time series prediction intervals using neuroevolution, *Neural Computing and Applications* (2019) 1–15.
- [70] B. D. Ripley, *Pattern Recognition and Neural Networks*, Cambridge University Press, 1996.
- [71] C. N. Bergmeir, B. Sánchez, Neural networks in R using the Stuttgart neural network simulator: RSNNS, *Journal of Statistical Software* 46 (7) (2012) 1–26.
- [72] M. Christ, N. Braun, J. Neuffer, A. W. Kempa-Liehr, Time series feature extraction on basis of scalable hypothesis tests (tsfresh—a python package), *Neurocomputing* 307 (2018) 72–77.
- [73] P. Domingos, A few useful things to know about machine learning, *Communications of the ACM* 55 (10) (2012) 78–87.
- [74] E. Frutos, J. Trapero, F. Ramos, A literature review on operational decisions applied to collaborative supply chains, *PloS ONE* 15 (3) (2020) e0230152.
- [75] M. Attaran, S. Attaran, Collaborative supply chain management, *Business Process Management Journal*, 13 (3) (2007) 390-404.