

Article

An Empirical Comparison of the Sales Forecasting Performance for Plastic Tray Manufacturing Using Missing Data

Che-Yu Hung ¹, Chien-Chih Wang ^{2,*}, Shi-Woei Lin ¹ and Bernard C. Jiang ¹

¹ Department of Industrial Management, National Taiwan University of Science and Technology, Taipei 10607, Taiwan; d10701003@mail.ntust.edu.tw (C.-Y.H.); shiwoei@mail.ntust.edu.tw (S.-W.L.); bcjiang@mail.ntust.edu.tw (B.C.J.)

² Department of Industrial Engineering and Management, Ming Chi University of Technology, New Taipei 24301, Taiwan

* Correspondence: iecawang@mail.mcut.edu.tw; Tel.: +886-2-2908-9889

Abstract: The problem of missing data is frequently met in time series analysis. If not appropriately addressed, it usually leads to failed modeling and distorted forecasting. To deal with high market uncertainty, companies need a reliable and sustainable forecasting mechanism. In this article, two propositions are presented: (1) a dedicated time series forecasting scheme, which is both accurate and sustainable, and (2) a practical observation of the data background to deal with the problem of missing data and to effectively formulate correction strategies after predictions. In the empirical study, actual tray sales data and a comparison of different models that combine missing data processing methods and forecasters are employed. The results show that a specific product needs to be represented by a dedicated model. For example, regardless of whether the last fiscal year was a growth or recession year, the results suggest that the missing data for products with a high market share should be handled by the zero-filling method, whereas the mean imputation method should be for the average market share products. Finally, the gap between forecast and actual demand is bridged by employing a validation set, and it is further used for formulating correction strategies regarding production volumes.

Keywords: mean imputation; zero-filling; forecaster; sustainable manufacturing; case study



Citation: Hung, C.-Y.; Wang, C.-C.; Lin, S.-W.; Jiang, B.C. An Empirical Comparison of the Sales Forecasting Performance for Plastic Tray Manufacturing Using Missing Data. *Sustainability* **2022**, *14*, 2382. <https://doi.org/10.3390/su14042382>

Academic Editors: Ripon Kumar Chakraborty and Chia-Lin Chang

Received: 16 December 2021

Accepted: 16 February 2022

Published: 19 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

To deal with increasingly fierce market competition, manufacturers have transformed their policies by providing customers with customized products and services, quickly responding to diversified needs, reducing competition uncertainty, and obtaining satisfactory services. Manufacturers expect to maintain or even increase their sales through such a transformation under a potentially increasing inventory pressure. As the green production and the circular economy have gradually formed a consensus between production and sales, manufacturers have tried to address the above challenges and turn them into a positive force to solve market uncertainty and effectively manage their inventory of existing production models.

During the last two decades, significant research work has been reported in the literature. This work has demonstrated that demand forecasting is one of the main tools for evaluating and maintaining the market and has become the cornerstone of companies' decision-making strategies [1–12]. In practice, demand forecasting includes at least two parts, namely, production forecasting and inventory control. Production forecasting relates to actual sales, and the time series analysis has become one of the best solutions for sales forecasting. The use of time series models to assist business decision-making has proven its success in many sectors and industries, such as energy consumption forecasting in the petrochemical industry [13,14], station expansion and capacity growth forecasting in the bus system [15], and economic and financial growth forecasting [16]. The current literature

on inventory control is often associated with supply chain management research [3,7,8,11]. Different production modes, such as make-to-stock (MTS), make-to-order (MTO), assembly-to-order (ATO), and build-to-forecast (BTF), apart from providing downstream customers with different levels of customized services, enable manufacturers to verify their management ability to balance revenue generation and inventory control.

From a practical viewpoint, companies require both accuracy and sustainability from the predictive models. Usually, there is a difference between observed and forecast results. Thus, the prediction results may not necessarily be acceptable (model accuracy) or reasonable (model sustainability). Conducting a verification process (accepted and certificated by companies) before selecting the predictive models is very important to effectively respond to this difference. For example, the sales record of commercial activities will inevitably encounter the no-order situation (for a day or consecutive days), not only during holidays but also during normal working days. However, in sales forecasting, the subjective comments (considering that the no-order situation is occasional and rare) often dominate the objective sale results and are unacceptable for predicting zero sales. They may also affect the preprocessing data methods before modeling. For example, in terms of data storage, records of no orders are usually indicated as blanks. In the following analysis, these blank records are classified as missing data, and the original series is defined as incomplete.

In the currently available literature on time series analysis with missing data, it has generally been assumed that the missing data are randomly missed, i.e., missing at random (MAR) or missing completely at random (MCAR). Moreover, different data imputation methods have been discussed and compared, and predictive modeling for the imputed data series has been performed [17–26]. The zero-forecasting method, which is used to predict a rare event or an intermittent demand [27], has also been reported. Since a rare event is treated as a particular case as part of a business activity and the result is always predicted as zero, this method is not as popular as other methods, which are based on statistical learning (such as the autoregressive integrated moving average (ARIMA)) or machine learning (such as the long short-term memory (LSTM)). In the research conducted on intermittent demand forecasts, the missing data are preprocessed either by combining adjacent time periods [8] or by defining the missing value as noise and then smoothing it out [7,21] or using min-max normalization to revise the data [9]. Then, the transformed data series is processed using a typical time series analysis. Generally, in the studies reported in the literature, the best solution is sought to obtain a better model accuracy. Compared with model accuracy, in model sustainability, the long-term data background is observed and understood both before and after modeling. In actual business activities, companies may accept the additional costs predicted by inaccurate forecasting results. For example, overproduction will increase the inventory cost (high forecast, but low actual demand), and underproduction will increase the labor cost because of the overtime work required (low forecast, but high actual demand). Therefore, companies are concerned about taking appropriate actions in the shortest possible time to correct the discrepancies caused by forecasting. Moreover, the nature of the data is a key issue.

This article aims to determine appropriate methods that are capable of dealing with missing data by establishing an accurate and sustainable forecasting model on the basis of a specific sales data background and to provide a business reference. To this end, a set of real sales data regarding plastic injection tray products is empirically investigated. The products are the outer trays of consumer electronics chips; MTS and MTO are the company's existing production modes. Many blank records in the dataset exist. In this article, zero-filling values in place of the blank records are proposed. Then, time series forecasting is performed on the recovered series. The results are compared with those obtained from the mean imputation method applied to different forecasters, including the Naive forecasting, the ARIMA, and the LSTM. As a conclusion, managerial insights are also proposed.

The rest of this article is organized as follows. The literature review is presented in Section 2. The materials and methods used are described in Section 3, and the numerical results are presented in Section 4. Section 5 provides the discussion and conclusion.

2. Literature Review

2.1. Analysis of Time Series with Missing Data

Nowadays, the time series analysis is widely applied in various sectors, including consumption and businesses [5,9,10,22,28,29], demand forecasting and supply chains [1–3,6,7,11,12], economics [30], industrial applications [4], traffic and automatic system controls [8,21,31], meteorology and the environment [17], epidemiology [19,20,23–25], and others [26,32,33]. In the above studies, the authors have proposed assumptions regarding time series data, the corresponding modeling methods, and evaluation indicators for various contexts.

The problem of missing data is frequently met in time series analysis. This issue, if not appropriately addressed, usually leads to failed modeling and distorted forecasting. The Kalman filter is a tool used to calculate the likelihood of a stationary autoregressive moving average (ARMA) process which describes how the missing data are handled during modeling in both the stationary ARMA and nonstationary ARIMA processes. These studies [14,15] are based on the assumption of normal distribution and estimate the marginal likelihoods.

Furthermore, Kohn and Ansley [16] demonstrated the prediction and interpolation of missing data and used the mean squared error (MSE) metric for performance evaluation. Their work is considered to be a milestone in the research of forecasting time series with missing data. Assisted by the innovation and development of machine learning and deep learning algorithms along with the computing performance of modern computers, further developments in time series forecasting with missing data have been achieved.

2.2. Missing Data Processing

The improper processing of missing data definitely affects the analysis results and the subsequent decision-making process. A general mechanism of missing data based on randomness includes MAR, MCAR, and missing-not-at-random (NMAR) [34–36] data. Here, it is assumed that data are missed because of human involvement during the data collection process. In reality, determining whether the data are entirely missed randomly is a challenging task because of various reasons. For example, in recent years, missing data due to machine maintenance, failed devices, delayed transfer, and other factors have been reported [20,37]. The general approaches used to handle missing data include deletion, mean substitution by valid observations, mean substitution by adjacent observations [38], and maximum likelihood estimation [39,40]. The machine learning approach [41] has also recently been presented. These approaches have become the principal mode of processing missing data in recent decades. The min–max normalization for smoothing noisy data, eliminating outliers, and fixing gaps due to missing data [9,10] has also been reported.

In recent years, the articles related to the keyword of missing value have mainly discussed the imputation and estimation methods. For example, multivariate imputation is recommended for food composition data [42]. Deep learning methods are recommended to deal with the missing value for water quality monitoring due to differences in sensor systems. In articles related to the keywords of both sale forecasting and missing value, the recommended methods are based on whether the data are traceable or not. For example, the missed E-commerce data can be filled by front-end operations [43]. If data are randomly missing and this is not easy to investigate, for example, in the retail business, mean and median imputation methods are mostly recommended [44–48]. For the food and beverage service in restaurants, the missing value would be due to the fact that no customers ordered them (zero order) or the stores were closed (no service). To avoid causing the inferior model fits, such circumstances are suggested to remove missing records before modeling [49].

The original data can be revised by employing any of the above approaches. In practice, it is most important to first recognize the background data. Suppose that the

reason for the missing data is known (no orders collected; company out of power due to snow), and the source of missing data is traceable (customers do not work on holidays or they choose other products; the company cannot serve customers before power recovery) and retrievable (missing data should be recorded as zeros in both cases). In this case, it is not necessary to impute any values instead. In this article, the methods of deletion, mean imputation, and filling with zero are used with different time series forecasters for handling missing data.

2.3. Classic and Modern Time Series Forecaster

The classic time series forecasters are based on statistical learning and include the Naive forecasting [4,7,50], the moving average [2–4,7,22,32,33], the exponential smoothing, the ARMA [14,28,51], and the ARIMA [6,14–18,24,26,28–30,50–52] processes. The modern time series forecasters include machine learning and deep learning algorithms such as the support vector regression [6,10,11], k-nearest neighbor [10,31], artificial neural network [1,7,33], recurrent neural network (RNN) [6,9,10,12], and LSTM [6,9,10,29,30,53] algorithms. Recently, a comparison between the statistical learning and modern approaches in either simulated or real datasets with or without missing data has been reported in the literature [1,10,12,28–30].

2.4. Indicators for Evaluating Time Series Forecasters

In the literature, there are many indicators available for evaluating and comparing the model performance. The most frequently used indicators are the mean squared error MSE [1,4,16,29,33], root mean squared error (RMSE) [4,5,9,16,19,21,22,25,26,28–30,51], mean absolute error (MAE) [5,7,11,12,19,24,25,33], and mean absolute percentage error (MAPE) [1,36,54,55], which are used for calculating accuracy. The MSE and RMSE are calculated using the mean as the center. They are susceptible to missing data and outliers (the variance is increased) and stable while employing the mean imputation (the variance is reduced). The MAE and MAPE only calculate absolute values; the forecasts are supposed to be positive, and the negative predictions are ignored. In specific applications, such as air pollutants [25] and the sale forecasting of consumer products [10], these two indicators are more accepted and more persuasive than the MSE and RMSE. The MAPE is more sensitive than the MAE since it is formatted as a percentage, but its drawback is that it is not compatible with symmetry, and the results cannot be calculated when the original observations have zero values. Additional indicators, such as the symmetric MAPE [9,10,31,56] and the mean absolute scaled error (MASE) [5,6], are specifically used for enhancing and correcting the limitations of MAPE.

3. Materials and Methods

In this article, the data used for the empirical study were obtained from a plastic injection product manufacturing company. Its main business line is to provide downstream firms, which produce electronic chips (such as Sim Cards, ICs, Smart Cards, and Flashes), with packaging boxes (also called trays) of various specifications. In recent years, the theme of green production and the circular economy has been widely discussed and advocated. The recycled used trays are cleaned and reused, or remelted and reinjected to make products with different specifications. These are the practices proposed by manufacturers to deal with market competition. These practices are also environmentally friendly.

Conversely, the above practices have changed the previous supply and demand model. For example, the cycle of customers placing the orders is no longer fixed, the selection of new product specifications continues to increase, but the frequency and amount of a single product demand and a single order form may decrease. These revised business activities also increase the records of zero sale events during a typical working day. To deal with these new cases in their operations, companies must renew their service model and they require more comprehensive scientific management.

In this article, the flowchart presented in Figure 1 was designed to include data acquisition, data preprocessing, modeling, evaluation, and till deployment. Two tasks were arranged in the preprocessing data phase: data transformation and missing data processing. The original data were initially transformed from a daily to a weekly format. Then, both the zero-filling and mean imputation methods were used to fill with values in the transformed series. Subsequently, the models designed by combining the missing data processing methods (zero-filling and mean imputation) and the forecasters in the modeling phase were applied to the filled series. The MAPE and MASE were used as indicators for model evaluation. A set of unused data (extracted from the transformed series) was explicitly used for model deployment to validate that the selected model was sustainable.

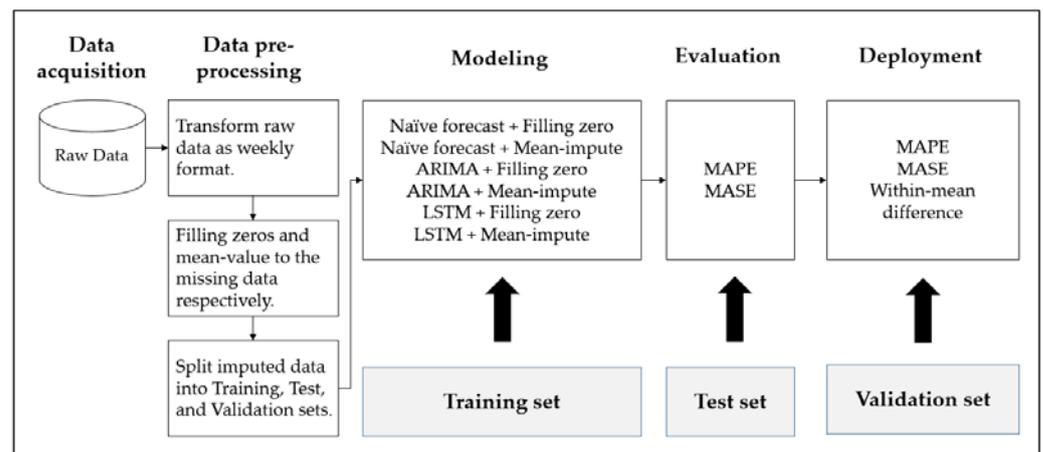


Figure 1. The flowchart of this article.

3.1. Data Preprocessing

Because of commercial confidentiality, customer information and product prices were removed from the data used in this article in advance. The raw data used were the order records of plastic injection products collected from 1 January 2017 to 30 September 2019. To validate the analysis results, the data were filtered, and the top 10 products in 2018 (fiscal year) were selected. The following two assumptions were considered in the analysis: the first one considered the exact unit price, regardless of differences in the specifications; the second one considered that there was no substitution effect between products.

3.1.1. Definitions of and Equations for RMS and MGR

The relative market share (RMS) and the market growth rate (MGR) are significant indicators for creating a Boston Consulting Group (BCG) Matrix (first introduced by Dr. Bruce D. Henderson, the Boston Consulting Group, 1970) [57,58]. The BCG Matrix is also called the Product Portfolio Matrix. A typical 2×2 matrix is used to position a firm's competitiveness or a brand product in the local or global market. From a practical point of view, the term RMS relates to cash generation and cash usage performance.

Relative market share (RMS). The RMS is used to evaluate how far an owned product is from its leading competitor in the market. This indicator represents the competitiveness and completeness of a company's products or brands. A high competitiveness leads to obvious and immediate high profits (the cash) for a company. However, if the company's profit highly depends on a single or a few products, different business problems may arise, once the demand changes. A company with a high market share can gradually expand and boost the growth of other products and establish a complete commercial strategic value and market position. The RMS equation is given as follows:

$$\text{RMS} = \frac{\text{Firm or Brand's Sales this year}}{\text{Leading competitor's Sales this year}} \times 100\% \quad (1)$$

The RMS is always a positive value, and its maximum value is 100% (or 1.00). Furthermore, by adopting the midpoint value (0.50), the market share status (of the product or brand) is indicated. For example, an RMS greater than or equal to 0.50 indicates a high (market) share, whereas an RMS less than 0.50 indicates an average (market) share.

Market growth rate (MGR). The MGR is used to measure the degree to which a firm's or brand's capital gain grows or declines year on year. This indicator represents the degree of change (increase or decrease) in the sales performance or the market share in a specific time (typically a year). The MGR equation is given as follows:

$$\text{MGR} = \frac{\text{Firm or Brand's Sales this year} - \text{Firm or Brand's Sales last year}}{\text{Firm or Brand's Sales this year}} \times 100\% \quad (2)$$

The MGR can reach a very high positive or negative value. If there is no sales record in the previous fiscal year, it cannot be calculated. A 10% annual rate is usually used to assess whether the growth is significant or not. A growth rate of more than 10% indicates a high growth, whereas a growth rate of less than 10% indicates a slow and moderate growth. Accordingly, an annual (negative) growth rate lower than -10% indicates a high decline, whereas a (negative) growth rate higher than -10% indicates a moderate decline.

3.1.2. Missing Data Processing Methods

In this article, the mean imputation and the zero-filling methods were proposed for processing missing data.

Mean imputation. According to its name, in this method, the missing data are replaced by their mean value, which is calculated from other valid data of a variable where the missing data are located. The advantage of this method is that the calculation is simple. Its disadvantage is that both the mean and standard deviation indicators increase after imputation.

Zero-filling. This method is also an imputation-type method, but the missing data are replaced by zeros. From the time series perspective, this method can explain why the data of a specific event (for example, the sale orders) have not been effectively collected at a specific timestamp (for example, during weekends or national holidays). In addition to retaining the nature of the data, the imputed series is also characterized by completeness.

Compared with the mean imputation, the zero-filling method can overcome the disadvantage of the mean value becoming large. Its disadvantages are a more significant variance and not being able to calculate the MAPE.

3.1.3. Data Split into Training, Test, and Validation Sets

The filled series (weekly format) are split into training, test, and validation sets. Table 1 summarizes the definition of each set. The ratio between these sets is 8:2:1.

Table 1. Definition of Training, Test, and Validation sets.

Set	Period	# Of Weeks	The Use
Training	3 January 2017–31 December 2018	104 (72.72%)	Train the models
Test	1 January 2019–30 June 2019	26 (18.18%)	Test if the trained models are appropriate
Validation	1 July 2019–30 September 2019	13 (9.09%)	Validate the performance of trained models deployed on unused data

3.2. Forecaster

In this section, the Naive forecasting, the ARIMA, and the LSTM methods are introduced. These three forecasters were selected on the basis of their specific characteristics. The Naive forecasting method is one of the most frequently used tools by companies. It is a

quick and easy method to use, but its latter forecasts are significantly affected by its former ones, especially when some impacts and uncertainties are not immediately observed (for example, in the case when a former forecast has failed and many latter forecasts become worse). The ARIMA method provides complex but delicate parameter settings. In this model, autoregression and moving average models are integrated, even if the series is stationary or not. The ARIMA is also a data-driven model; it can switch to the ARMA, AR, MA, or even seasonal SARIMA (SARIMA) models, depending on the data characteristics (trend, cycle, seasonality, and more). An effective ARIMA model requires the series data to be complete, but missing data will be frequently encountered in time series analysis. The LSTM is based on RNNs and can address the issue of missing data existence, either by doing nothing (directly ignoring the missing data) or accepting any specific imputation (single or multiple). Overfitting may also occur after executing a large number of iterations.

3.2.1. Naive Forecasting

The Naive forecasting model [50] is the most straightforward time series approach and one of the most frequently used tools by companies. By definition, the last observation of the series is the forecast of the following data point. This is described by the following equation:

$$y_t = \hat{y}_{t+1} \quad (3)$$

where y_t is the observation at time t and \hat{y}_{t+1} is the forecast at time $t + 1$. This approach works remarkably well for many economic and financial time series [50].

3.2.2. ARIMA

The ARIMA [50,52] model is one of the most widely used approaches in time series forecasting. In this approach, the autoregression and moving average models are integrated. The approach also considers series stationarity and the selection of series transformation. The augmented Dickey–Fuller (ADF) test and the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test [59] are valuable tools for detecting series stationarity. Initially, the ADF test is used to check if the series is trend stationary. Then, the KPSS test is used to check if the series is simply difference stationary, even if the series is stationary. Sometimes, modeling may still be complicated because of the presence of white noise and cycle behavior (no trends, no seasonality). To deal with this problem, data transformation, such as differencing and other methods (for example, smoothing and shift), can be used.

Three main parameters are required to configure the ARIMA model; p refers to the AR model, d denotes the integration steps, and q refers to the MA model. For a stationary series, autoregression is modeled using a linear combination of a variable's past value. In other words, the term autoregression means a regression of the variable against itself. An autoregression model of order p is expressed as AR(p), which is formed as follows:

$$y_t = c + \delta_1 y_{t-1} + \delta_2 y_{t-2} + \dots + \delta_p y_{t-p} + \varepsilon_t \quad (4)$$

where ε_t is the white noise and y_t is the forecast value using its lagged value as the predictor.

A moving average model of order q is expressed as MA(q), which is a regression-like model and is formed as follows:

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (5)$$

where ε_t is the white noise and y_t is the weighted moving average of the past few forecast errors (lagged errors). Generally, by combining differencing with an autoregression and a moving average model, a nonseasonal ARIMA model is obtained. This is expressed as ARIMA (p, d, q), which is formed as follows:

$$y'_t = c + \delta_1 y'_{t-1} + \delta_2 y'_{t-2} + \dots + \delta_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (6)$$

where y'_t is a forecast of the transformed series after differencing and the predictors include the lagged values of y_t and the relative lagged errors. The parameters p and q represent the order of the autoregression and the moving average model, respectively, and d denotes the steps of differencing conducted (if necessary) to integrate these two models.

The basic steps to build the ARIMA model involve first conducting the ADF and KPSS tests and then checking if the transformed series (after differencing) is stationary. The next step is to determine the best combination of p and q . The final step is to confirm whether the white noise follows the normal distribution.

3.2.3. LSTM

The LSTM is one of the most popular predictive models used in recent years. It was first introduced by S. Hochreiter and J. Schmidhuber in 1997 [53]. The LSTM prototype comes from an RNN, which is a class of neural network models. By configuring memory feedback during the learning process, an RNN can improve the feedforward learning constraint, which exists in convolutional neural networks (CNNs). It then reduces the bias caused by overlearning. Based on an RNN, the LSTM can solve other complicated problems, which are derived from adding different background factors during the learning process, for example, how to set the feedback position when the occurrence of events is no longer in a fixed order; alternatively, whether the learning memory should be retained or dropped if the time interval is inconsistent, etc. The usual case is the following: in a fixed order of events (for example, first is event A and then followed B, C, D, and E), either C or D could be in a feedback loop, but what if B or E is missing?

The LSTM includes an input gate, an output gate, a forget gate, and a cell. The information enters the cell through the gates and exits as numbers (between 0 and 1). A zero means that all the information has been completely dropped out, whereas a one means that all the information has been completely retained. Figure 2 shows a diagram of an LSTM with a single cell.

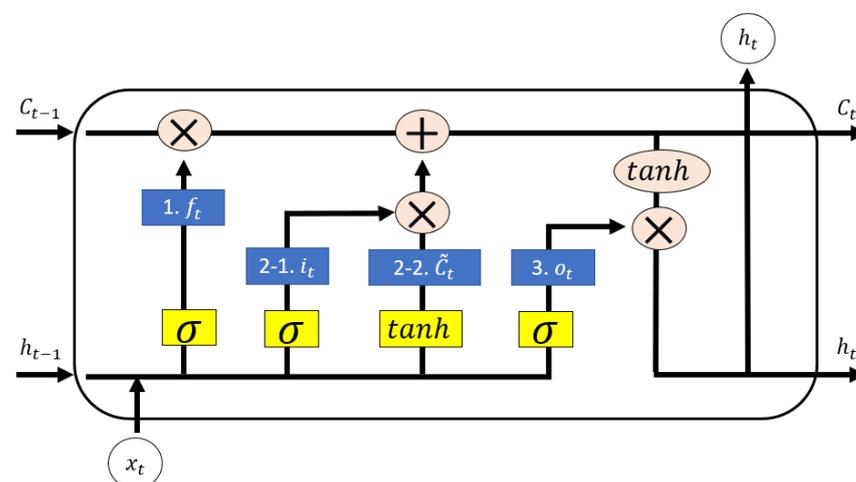


Figure 2. Long short-term memory structure, the single cell.

From the input x_t to the output h_t , four functions are executed in a single cell. These functions are divided into three steps and then are integrated (by applying the addition and multiplication operations) step by step, until the output h_t is ready to be produced.

In Step 1, a function, f_t , is used to determine the information to be dropped out of the cell state. This function is a typical sigmoid function involving the input x_t , the output h_{t-1} of the previous cell, the weighted decay W_f , and the bias b_f . The output is a number between 0 and 1, which is then multiplied with C_{t-1} of a previous cell and moves forward. f_t is described as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \quad 0 \leq f_t \leq 1 \quad (7)$$

Step 2 includes two functions (i_t and \tilde{C}_t), which represent the retained information and a new candidate vector, respectively. This vector is created by the hyperbolic tangent function (\tanh). This step is used to decide the new information to be stored in the cell state. This information is then added to the output generated in Step 1 and moves forward. i_t , \tilde{C}_t , and C_t are given as follows:

$$\begin{aligned}\tilde{C}_t &= \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \\ i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\ C_t &= f_t \times C_{t-1} + i_t \times \tilde{C}_t\end{aligned}\quad (8)$$

Step 3 is used to decide the information to be exited from the cell. The output function o_t is executed by a sigmoid function involving the input x_t , the output of the previous cell h_{t-1} , the weighted decay W_o , and the bias b_o . o_t is described as follows:

$$o_t = \sigma(W_o [h_{t-1}, x_t] + b_o) \quad (9)$$

The last update of the cell state C_t (obtained from Step 2) is first created by applying \tanh (i.e., values between -1 and 1 are produced). It is then multiplied with the output function O_t to generate the final output h_t of the cell as follows:

$$h_t = o_t \times \tanh(C_t) \quad (10)$$

3.3. Model Performance Indicator

By combining the missing data processing methods (Section 3.1.2) and the proposed forecasters (Sections 3.2.1–3.2.3), a total of six combined models are obtained. These include Naive forecasting + zero-filling, Naive forecasting + mean imputation, ARIMA + zero-filling, ARIMA + mean imputation, LSTM + zero-filling, and LSTM + mean imputation. Each of the selected products implements the six combined models and evaluates the model performance based on the produced indicators. Based on the original series, which includes missing data, and the zero-filling method proposed in this article, the MAPE is defined as the first indicator that evaluates and filters which combined model is the most appropriate. The second indicator used for further filtering is the MASE, an indicator that can handle zero counts. These two indicators are used to filter the models and help evaluate if the selected models are reliable.

3.3.1. MAPE

The MAPE is defined as a loss function type by its definition [60]. Similar to other indicators (including the MSE and RMSE), the MAPE is widely used for model accuracy evaluation. This indicator transforms the initial deviation into an absolute value form. In other words, it imposes a heavier penalty on the positive errors usually caused by overestimation. The MAPE equation is given below:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{A_i - F_i}{A_i} \right| \times 100\% \quad (11)$$

where A_i and F_i are the actual value and the forecast value, respectively, of the i -th data point and n is the length (or the number of forecasts) in a given period. A small MAPE value means that, on average, the selected model provides relatively accurate results. The two drawbacks of MAPE are as follows: (1) its inability to handle zero values [61] and (2) the asymmetry problem due to large numbers [55]. The MAPE is unable to calculate if an actual value corresponds to the missing data. Even so, it is relatively effective in using zero values to replace the missing data since the calculated MAPE indicator is always positive. Moreover, it imposes a relatively heavy penalty for positive errors caused, for example, by overestimation [62].

3.3.2. MASE

The MASE was first proposed by Hyndman and Koehler in 2005 [61]. It is a scale-independent indicator for measuring the forecasting accuracy. Because of its scale-independent characteristic, the MASE handles the zero values directly and imposes an equal-weight penalty on both the positive errors (caused, for example, by overestimation) and the negative errors (caused, for example, by underestimation). Generally, the MASE overcomes the significant drawbacks of the MAPE. For time series data, if they are nonseasonal, the MASE is calculated using the following equation:

$$\text{MASE} = \frac{\text{Validation MAE}}{\text{Training MAE of naive forecasts}} = \frac{\frac{1}{v} \sum_{t=n+1}^{n+v} |e_t|}{\frac{1}{n-1} \sum_{t=1}^n |Y_{t-1} - Y_t|} \times 100\% \quad (12)$$

The denominator is the mean absolute error of the one-step Naive forecasting on a training set with n data points. If the series contains seasonal factors, the period t of the training set is redefined. A MASE value of less than one means that the proposed model produces more minor errors than a one-step Naive forecasting [37]. In other words, a MASE value greater than one means that the performance of the proposed model is worse than that of the Naive forecasting.

3.3.3. Within-Mean Difference

To effectively illustrate the achieved performance of the selected models on the validation set, a specific indicator named within-mean difference (WD) is introduced in this article. The WD is the percentage difference (%) between the forecast and the actual values, when it is applied to the validation set. The WD formula is given as follows:

$$\text{WD}_{\text{Validation}} = \frac{\text{Forecast}_{\text{Validation}} - \text{Actual}_{\text{Validation}}}{\text{Forecast}_{\text{Validation}}} \times 100\% \quad (13)$$

A WD value close to zero indicates that the difference between the forecast and the actual values is small, which means that the selected models perform well. A positive WD value indicates that the selected models lead to overestimation, whereas a negative WD value indicates that the trained model lacks fitting (i.e., insufficient estimation). When a WD exceeds 100% or drops below -100% , it is recommended not to further use these models because of their poor performance.

3.4. Research Questions

Through the empirical analysis, this article aims to answer the following two classic research questions:

RQ1: Which combination of the missing data processing methods (deletion, mean imputation, and zero-filling) and the forecasters (ARIMA, LSTM, and Naive forecasting) achieves the best performance (MAPE, MASE) for specific products?

RQ2: Which missing data processing method is mostly recommended for individual forecasters?

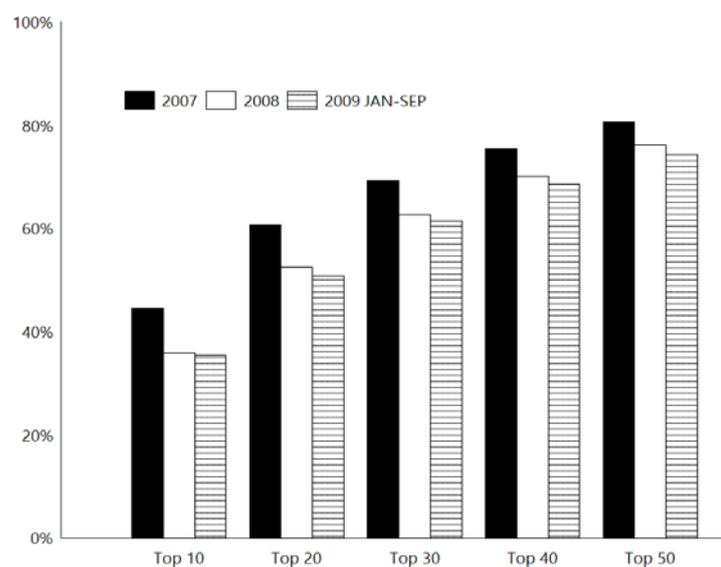
4. Numerical Results

4.1. Used Data Background

In this empirical study, a real dataset of daily sale records of plastic tray products was used. The data were collected from 1 January 2017 to 30 September 2019. Table 2 presents the change in cumulative sales from the best 10 to the best 50 sold products. As shown in Figure 3, a decreasing trend was observed from 2017 to 2018 and continued until September 2019. By combining Table 2 and Figure 3, even if the company provides more than 390 product options each year (the options will not be the same each year), the top 50 sold products will account for more than 70% of the sales in the entire year. Furthermore, the sales of the top 10 sold products will account for a half or more of those of the top 50 sold products.

Table 2. Annual cumulative sales (%), top 50 products sold.

Year	Top 10	Top 20	Top 30	Top 40	Top 50
2017	44.516%	60.753%	69.223%	75.406%	80.593%
2018	35.928%	52.434%	62.699%	70.093%	76.215%
2019 (January–September)	35.480%	50.919%	61.434%	68.694%	74.266%

**Figure 3.** Cumulative sales proportion, September 2017–2019.

Among the 60090 orders collected, sales records existed for 576, of which six products repeatedly (in 2017, 2018, until September 2019) became the top 10 sold products of the year. Table 3 describes the change of the cumulative sales percentage (%) of the top 10 sold and the six recurring products. Two critical observations can be extracted. First, the six recurring products accounted for more than 20% of the annual sales, indicating that downstream customers have a certain degree of dependence on purchasing specific products. Second, the other four products (also in the top 10 sold, but not the same) were combined and accounted for approximately 13–18% (see the Difference row) of the annual sales, indicating the diversity of demands and the complexity of multiproduct management.

Table 3. Accumulated sales proportion (%), Top 10 sold.

	2017	2018	2019 January–September
(1) Top 10 sold	44.516%	35.928%	35.480%
(2) The Recurring 6	26.619%	20.658%	22.011%
Difference = (1)–(2)	17.897%	15.2780%	13.469%

As it was the best-sold product in 2018, the BGA 8 × 13 mm was selected as the benchmark learning object, and its RMS-18 was defined as 1.00. According to Section 3.1.1, the calculated MGR-18 of product BGA 8 × 13 mm was 0.33. The other nine products of the top 10 sold in 2018 were compared with the product BGA 8 × 13 mm. Then, the corresponding RMS-18 indicators were calculated.

Table 4 summarizes the RMS-18 and MGR-18 of the top 10 sold products in 2018. Although six of them were repeated in the top 10, even if some were ranked high, the growth rate exhibited a decline year by year. According to the criteria defined for the RMS and MGR (Section 3.1.1), the top 10 sold products can be categorized into four groups: high share and high growth, high share and high decline, average share and high growth, and average share and high decline. From each group, a specific product was selected for

further time series analysis (BGA 8 × 13 mm, TSOP II 54/86P, TQFP 14 × 14 × 1.4, TSOP II 54/86 135°C, respectively).

Table 4. RMS-18 and MGR-18, top 10 sold of 2018.

Product	Sales Rank		RMS-18	MGR-18	Group
	2017	2018			
BGA 8 × 13 mm	4	1	1.00	0.33	high share and high growth
TSOP I 12 × 20 mm JEDEC TRAY	2	2	0.84	−0.12	high share and high decline
BGA 8 × 12.5	3	3	0.79	−0.14	high share and high decline
TSOP II 54/86P	1	4	0.66	−0.41	high share and high decline
BGA 7.5 × 13mm	60	5	0.64	11.76	high share and high growth
TQFP 7 × 7 × 1.4MM	7	6	0.57	0.08	high share and average growth
QFN 9 × 9	8	7	0.54	0.49	high share and high growth
BGA 11.5 × 13	6	8	0.47	−0.14	average share and high decline
TQFP 14 × 14 × 1.4	9	9	0.45	0.29	average share and high growth
TSOP II 54/86 135°C 400 × 875 mil	5	10	0.41	−0.38	average share and high decline

4.2. Review of the Missing Data Behavior

Table 5 summarizes the amount of missing data presented in daily and weekly formats for training, test, and validation. In the daily format, the amount of missing data was high, but it significantly reduced after transformation into the weekly format. Either using zero-filling or mean imputation, the commonly used statistics, such as the mean and sum were not significantly different when calculated on the basis of the weekly data format. However, using the daily data format, they were quite different.

Table 5. Amount of missing data.

Product	Daily			Weekly		
	Training (728)	Test (181)	Validation (92)	Training (104)	Test (26)	Validation (13)
BGA 8 × 13 mm	312	75	31	1	0	0
TSOP I 12 × 20 mm JEDEC TRAY	280	119	60	2	4	0
BGA 8 × 12.5	267	83	49	0	1	1
TSOP II 54/86P	278	105	50	1	1	0
BGA 7.5 × 13mm	400	62	33	9	1	0
TQFP 7 × 7 × 1.4MM	333	111	55	2	1	0
QFN 9 × 9	299	84	35	0	1	0
BGA 11.5 × 13	361	82	32	1	1	0
TQFP 14 × 14 × 1.4	354	114	43	2	1	0
TSOP II 54/86 135°C 400 × 875 mil	342	89	38	1	1	0

4.3. Model Comparison

For each group identified in Table 4, the representative samples selected for the analysis were the BGA 8 × 13 mm, TSOP II 54/86P, TQFP 14 × 14 × 1.4, and TSOP II 54/86 135°C products. Tables 6–9 present general comparisons. The following rules were adopted for model selection:

- Rule 1: In the test, both the MAPE and MASE are the smallest among all the models. Considering that the zero values affect the MAPE calculation, the smallest MASE is satisfied first;
- Rule 2: If Rule 1 is not satisfied, the minimum MASE model in the validation is selected. If this MASE is more significant than 100%, go to Rule 3;
- Rule 3: If both Rules 1 and 2 are not satisfied, the minimum WD model in the validation is selected.

4.3.1. BGA 8 × 13 mm

For the product BGA 8 × 13 mm, Table 6 summarizes the model evaluation (Test) and model deployment (Validation) performance. The ARIMA + zero-filling model was selected according to the adopted rules (Rules 1 and 2). Furthermore, a negative WD (−16.817%) indicated that the forecast values were underestimated by approximately 17% compared with the actual values. If this model is adopted in decision making, then the forecast production volume may be insufficient to meet the actual demand. In this case, the company should further evaluate and incorporate this expected gap in the decision-making process of the production and inventory management. For example, considering the stipulated safety stock and the first-in-first-out principle, the planned output must be increased by 17%.

A careful observation of Table 6 reveals that regardless of the forecaster it is combined with, the zero-filling method almost achieved the best performance in the model evaluation and deployment. Among the individual forecasters, and the ARIMA achieved the best performance, followed by the LSTM and the Naive forecasting models. Figure 4 presents a trend chart of the BGA 8 × 13 mm deployment, including the actual value, the forecast value, and the forecast value plus the proposed 17% expected gap. After adding the expected gap, the difference between the actual and the forecast values appeared more randomly. The MASE of the Validation was reduced to 64.447%, and the WD was best at −2.975%.

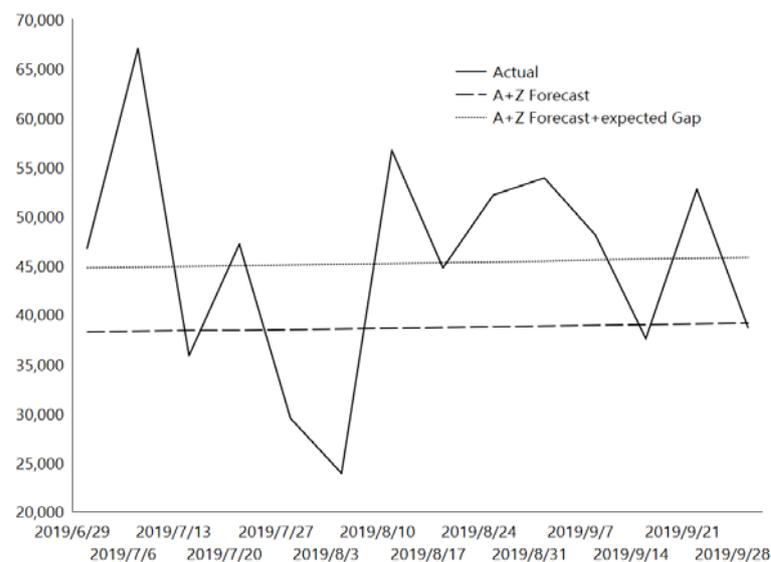


Figure 4. Trend chart, the deployed BGA 8 × 13 mm.

Table 6. Model evaluation and deployment, BGA 8 × 13 mm.

Model	Test (Model Evaluation)		Validation (Deployment)		
	MAPE	MASE	MAPE	MASE	WD
Naïve forecast + zero-filling	61.497%	100.215%	33.202%	95.433%	1.341%
Naïve forecast + mean imputation	61.497%	102.409%	33.202%	97.522%	1.341%
ARIMA + zero-filling	42.955%	82.548%	23.790%	72.481%	−16.817%
ARIMA + mean imputation	42.968%	84.080%	23.793%	74.119%	−16.915%
LSTM + zero-filling	50.246%	82.317%	32.096%	92.897%	−19.655%
LSTM + mean imputation	49.580%	85.797%	31.996%	96.361%	−23.133%

4.3.2. TSOP II 54/86P

Table 7 summarizes both the model evaluation (Test) and the deployment (Validation) for the product TSOP II 54/86P. The selected model was the Naive forecasting + zero-filling,

which exhibited the smallest MAPE and MASE in the Test and the smallest MASE in the Validation. The WD of the Validation was 12.534%, indicating that the forecast values were overestimated by approximately 13% compared with the actual values. In other words, by officially adopting this model, volume overproduction can be expected with a 13% reduction in the planned output.

Table 7 shows that the zero-filling method performed better (smaller MAPE and smaller MASE) than the mean imputation method, regardless of the forecaster used (i.e., the Naive forecasting, the ARIMA, or the LSTM). Among the individual forecasters, Naive forecasting achieved the best performance, followed by the ARIMA and the LSTM models.

Figure 5 presents a trend chart of the TSOP II 54/86P deployment. Because of the Naive forecasting characteristics, the patterns before and after adding the expected gap were almost identical. The MASE of the Validation was reduced from 56.681% to 53.409%, and the WD was best at -0.536% .

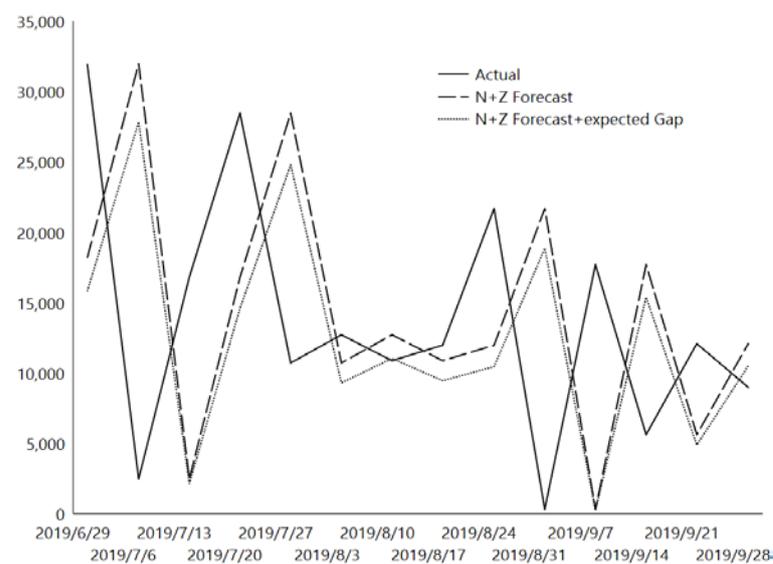


Figure 5. Trend chart, the deployed TSOP II 54/86P.

Table 7. Model evaluation and deployment, TSOP II 54/86P.

Model	Test (Model Evaluation)		Validation (Deployment)		
	MAPE	MASE	MAPE	MASE	WD
Naive forecast + zero-filling	50.738%	38.653%	699.579%	56.681%	12.534%
Naive forecast + mean imputation	882.065%	46.659%	699.579%	56.826%	12.534%
ARIMA + zero-filling	174.608%	56.240%	304.924%	86.779%	340.624%
ARIMA + mean imputation	302.329%	49.508%	173.959%	68.970%	918.302%
LSTM + zero-filling	575.984%	76.111%	1023.013%	81.146%	54.783%
LSTM + mean imputation	1050.151%	78.380%	1051.086%	86.291%	56.368%

4.3.3. TQFP $14 \times 14 \times 1.4$

Table 8 summarizes the model evaluation and deployment of the product TQFP $14 \times 14 \times 1.4$. The results show that only Rule 3 applied. The Naive forecasting + mean imputation model was used because of its relatively low MASE. The WD of the Validation was -3.049% , a relatively low figure; thus, there was no need to immediately add the production volume. Furthermore, regardless of the missing data processing method, the ARIMA performed relatively better than the LSTM and the Naive forecasting method. On the other hand, regardless of the forecaster used, it was not easy to decide whether the zero-filling or the mean imputation performed better.

Figure 6 presents a trend chart of the TQFP $14 \times 14 \times 1.4$ deployment. Because of the characteristics of Naive forecasting, the patterns before and after adding the expected gap were almost identical. Both the MASE and WD of Validation did not change significantly.

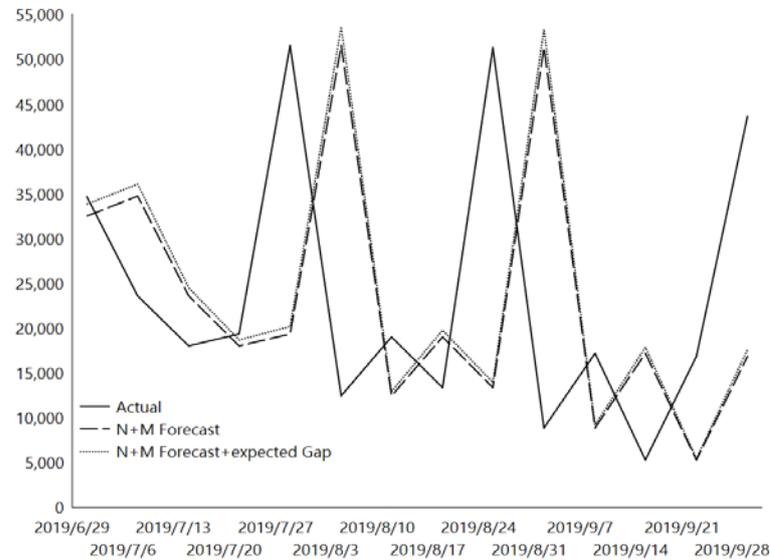


Figure 6. Trend chart, the deployed TQFP $14 \times 14 \times 1.4$.

Table 8. Model evaluation and deployment, TQFP $14 \times 14 \times 1.4$.

Model	Test (Model Evaluation)		Validation (Deployment)		
	MAPE	MASE	MAPE	MASE	WD
Naïve forecast + zero-filling	882.528%	99.793%	114.510%	184.759%	−3.049%
Naïve forecast + mean imputation	891.622%	95.987%	114.510%	183.079%	−3.049%
ARIMA + zero-filling	545.710%	99.839%	47.549%	123.946%	−85.734%
ARIMA + mean imputation	541.693%	96.296%	47.800%	123.896%	−88.193%
LSTM + zero-filling	809.245%	91.828%	71.340%	138.473%	−31.468%
LSTM + mean imputation	853.561%	89.138%	68.755%	131.915%	−29.215%

4.3.4. TSOP II 54/86 135°C

Table 9 summarizes the model evaluation and deployment results of the TSOP II 54/86 135°C product. The LSTM + mean imputation model was used because it exhibited the smallest MASE in both the Test and Validation and the smallest WD in the Validation. A negative WD of −13.230% indicates that the forecast values were underestimated by approximately 14% compared with the actual values. Considering the Validation performance, the LSTM was better than the Naive forecasting and ARIMA models because of its smaller MASE and its relatively low WD.

Table 9. Model evaluation and deployment, TSOP II 54/86 135°C.

Model	Test (Model Evaluation)		Validation (Deployment)		
	MAPE	MASE	MAPE	MASE	WD
Naïve forecast + zero-filling	151.265%	96.453%	93.899%	88.584%	−9.858%
Naïve forecast + mean imputation	150.298%	92.708%	93.899%	89.717%	−9.858%
ARIMA + zero-filling	61.544%	59.725%	78.834%	112.287%	−313.185%
ARIMA + mean imputation	63.320%	60.291%	78.759%	113.512%	−310.384%
LSTM + zero-filling	108.089%	62.501%	90.122%	63.586%	−12.818%
LSTM + mean imputation	106.119%	56.574%	92.424%	62.717%	−13.230%

Figure 7 presents a trend chart of the TSOP II 54/86 135°C deployment. It can be observed that the actual demand fluctuated randomly along the forecast line. The revised WD was also reduced (from -13.230% to 0.675%) as expected. However, some questions, such as whether there was a noticeable difference between forecast and actual demands among single weeks, were not answered. For example, the demand for the week starting at 3 August 2019 exhibited a sharp drop, but the demand for the week starting at 10 August 2019 exhibited a dramatic increase.

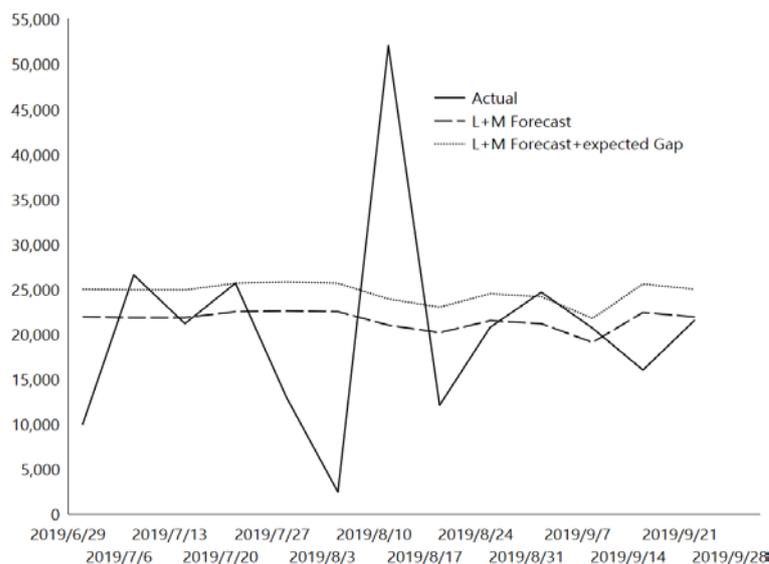


Figure 7. Trend chart, the deployed TSOP II 54/86 135°C.

5. Discussion

The results presented in Section 4.3 have answered the two research questions set in Section 3.4. Table 6 presents the high share and high growth products. It also reveals that regardless of the forecaster it was combined with, the zero-filling method almost achieved the best performance in the model evaluation and deployment. Among the individual forecasters, the ARIMA model achieved the best performance, followed by the LSTM, and the Naive forecasting models. Table 7 presents the high share, but high decline (in the last fiscal year) products. Again, the zero-filling method performed better than the mean imputation method regardless of the forecaster used (i.e., Naive forecasting, ARIMA, or LSTM models). Among the individual forecasters, the Naive forecasting model achieved the best performance, followed by the ARIMA and the LSTM models.

Based on the above results, it is evident that the zero-filling method is the most suitable for high market share products. The ARIMA and Naive forecasting models can also be used, depending on whether the product grew significantly or seriously declined in the previous year. A high market share represents a relatively stable cash flow for companies that provide many diversified and customized products. Regardless of the growth or recession in sales, this is a great challenge and has an impact on cash management. In a BCG matrix, the possible roles are the Star and Cash-cows. For a long-term development, companies must prioritize the changes in demand for products with a high market share.

Table 8 presents the average share and high growth products. The ARIMA model performed better than the LSTM model, followed by the Naive forecasting model, regardless of the missing data processing method. Conversely, regardless of the forecaster used, it was difficult to decide whether the zero-filling or the mean imputation is better. Based on a smaller WD, the Naive forecasting + mean imputation model is suggested. Table 9 presents the average share, but in reference to high decline products. The LSTM + mean imputation model is suggested in this case.

The common features of Tables 8 and 9 are those products that both represent the average market share and use the mean imputation to deal with the missing data. The

selection between the Naïve forecasting and LSTM models depends on the growth or recession of the last fiscal year. Against the previous two products (BGA 8×13 mm and TSOP II 54/86P), TQFP $14 \times 14 \times 1.4$ and TSOP II 54/86 135°C products are significantly affected by environmental factors. For example, high growth indicates potential, but an average share means that the expected potential cannot be fully confirmed and accepted.

Conversely, high decline means the product is possibly out of fashion but still survives because the average-share feature means it can still bring in cash. In a BCG matrix, the possible roles are Problem child and Dogs. Table 10 presents a general summary of the previous results.

Table 10. General summary of the empirical study.

Group	Representative	Suggested Model	Possible Role in BCG Matrix
High share, high growth	BGA 8×13 mm	ARIMA + zero-filling	Star
High share, high decline	TSOP II 54/86P	Naïve forecasting + zero-filling	Cash-cows
Average share, high growth	TQFP $14 \times 14 \times 1.4$	Naïve forecasting + mean imputation	Problem child
Average share, high decline	TSOP II 54/86 135°C	LSTM + mean imputation	Dogs

6. Conclusions

The primary objective of this article was to prove that a dedicated time series model can provide accuracy and sustainability for the sales forecasting of a specific product. Part of the empirical study results achieved this objective. For example, the ARIMA + zero-filling model can predict high share and high growth products. Although there was an underestimation of approximately 17%, this gap could effectively be filled by a correction strategy in real production. The second objective was to prove that a practical observation of the data background helps the appropriate method for processing the missing data to be selected. Four specific products with different backgrounds consistently proved that the zero-filling method achieves the best modeling and deployment performance, regardless of the forecaster it is combined with. By applying the same modeling process, apart from the average share and high growth products (no other products were matched), 6 of the top 10 products sold all led to the same conclusion (see Appendix A). The case company has recognized the case analysis results. Thus, it can be further confirmed that the two propositions of this article can be applied to a company's hot-selling products. They can also be used as a managerial reference for other companies with similar data backgrounds.

In this article, the empirical case of a univariate analysis was presented and the paper successfully dealt with the actual case problems related to the sales forecasting performance for plastic tray manufacturing. By adding other background factors, the contents of the analysis could be enriched to reduce modeling uncertainty and to provide more accurate results. From a practical point of view, when a specific model can be applied to other similar background products and produce effective forecast results, this model can be defined as a guide model. The integration of the BCG matrix and guidance models would enhance the efficiency of multi-item demand forecasting decision making. Moreover, the establishment of the forecasting framework will allow the application of multi-item forecasts. In future research work, the guide models could be figured out and tested for batch processing and multiple product management. The advantages of such an approach include a reduced calculation time, reducing the amount of substitution of other internal products, and even making a direct purchase list for customers. Companies should conscientiously master the production and stock management of specific products, continually provide excellent service regarding fulfillment and shipment, and carefully evaluate the potential of other products.

Author Contributions: C.-Y.H., C.-C.W. and B.C.J. designed the study. C.-Y.H. and C.-C.W. were responsible for methodology design and data analysis. C.-Y.H., C.-C.W. and B.C.J. reviewed relevant literature and interpreted the acquired data. C.-Y.H., C.-C.W., B.C.J. and S.-W.L. drafted the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Ministry of Science and Technology, grant number 110-2221-E-131-027-MY3.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Model Summary of Other Six of the Top 10 Products Sold

Appendix A.1. High Market Share and High Growth Products

Conclusion: Three products included. Following Rule 1 and Rule 2, the zero filling method generally performed better than the mean imputation method in combination with either the Naïve forecast or ARIMA models.

Table A1. Model summary for high market share and high growth products, the Star of the BCG matrix.

Product	Model	Test (Model Evaluation)		Validation (Deployment)		
		MAPE	MASE	MAPE	MASE	WD
BGA 7.5 × 13	Naïve forecast + zero-filling	46.307%	210.839%	25.859%	216.694%	−1.414%
	Naïve forecast + mean imputation	49.205%	167.242%	25.859%	189.118%	−1.414%
	ARIMA + zero-filling	42.644%	182.028%	48.538%	431.875%	−101.771%
	ARIMA + mean imputation	43.612%	161.505%	56.998%	438.494%	−141.008%
	LSTM + zero-filling	45.336%	199.835%	31.478%	284.435%	−23.717
	LSTM + mean imputation	46.435%	156.697%	30.791%	247.197%	−29.617
Product	Model	Test (Model Evaluation)		Validation (Deployment)		
		MAPE	MASE	MAPE	MASE	WD
QFN 9 × 9	Naïve forecast + zero-filling	357.408%	164.455%	257.754%	148.705%	−0.418%
	Naïve forecast + mean imputation	403.848%	168.616%	257.754%	148.705%	−0.418%
	ARIMA + zero-filling	226.165%	90.446%	193.722%	93.259%	1.973%
	ARIMA + mean imputation	226.165%	92.847%	193.722%	93.259%	1.973%
	LSTM + zero-filling	198.610%	102.422%	167.300%	102.350%	−61.201%
	LSTM + mean imputation	185.478%	98.911%	160.700%	101.709%	−69.277%
Product	Model	Test (Model Evaluation)		Validation (Deployment)		
		MAPE	MASE	MAPE	MASE	WD
TQFP 7 × 7 × 1.4MM	Naïve forecast + zero-filling	391.014%	41.522%	136.602%	64.403%	2.243%
	Naïve forecast + mean imputation	391.790%	41.782%	136.602%	63.529%	2.243%
	ARIMA + zero-filling	1569.859%	91.731%	251.179%	102.156%	54.323%
	ARIMA + mean imputation	1681.998%	110.925%	227.725%	91.246%	50.528%
	LSTM + zero-filling	1303.794%	84.954%	155.431%	64.817%	37.895%
	LSTM + mean imputation	1358.349%	83.901%	161.280%	66.864%	39.323%

Appendix A.2. High Market Share and High Decline Products

Conclusion: Two products are included. The MAPE values compared to the Test and the Validation sets are very large, but the MASE values generated by the models with zero-filling methods are relatively average. Following Rule 1 and Rule 2, it is concluded that the zero-filling method combined with either with Naïve forecast or ARIMA models performed better than the mean imputation method combined with the three forecasters.

Table A2. Model summary for high market share and high decline products, the Cash-cows of the BCG matrix.

Product	Model	Test (Model Evaluation)		Validation (Deployment)		
		MAPE	MASE	MAPE	MASE	WD
BGA 8 × 12.5	Naïve forecast + zero-filling	179.464%	65.062%	2452.072%	63.151%	−0.562%
	Naïve forecast + mean imputation	177.407%	52.413%	2452.072%	63.151%	−0.562%
	ARIMA + zero-filling	133.806%	41.726%	1222.734%	39.119%	−2.172%
	ARIMA + mean imputation	135.402%	40.284%	1222.734%	39.119%	−2.172%
	LSTM + zero-filling	276.828%	104.824%	2844.530%	117.981%	48.804%
	LSTM + mean imputation	280.000%	98.208%	2874.039%	119.294%	49.141%
Product	Model	Test (Model Evaluation)		Validation (Deployment)		
		MAPE	MASE	MAPE	MASE	WD
TSOP I 12 × 20	Naïve forecast + zero-filling	117.730%	60.660%	146.703%	70.566%	9.248%
	Naïve forecast + mean imputation	991.467%	96.284%	146.703%	69.797%	9.248%
	ARIMA + zero-filling	375.115%	50.505%	76.511%	73.367%	−578.007%
	ARIMA + mean imputation	1245.648%	116.735%	312.120%	125.780%	59.576%
	LSTM + zero-filling	1167.802%	118.220%	247.478%	95.888%	51.315%
	LSTM + mean imputation	1479.984%	125.727%	262.089%	103.814%	53.622%

Appendix A.3. Average Market Share and High Decline Products

Conclusion: One product is included and the model of LSTM + mean imputation outperformed the others. Further, the mean imputation method generally performed well compared to the zero-filling method whether combined with the Naive forecast, ARIMA, or LSTM models.

Table A3. Model summary for average market share and high decline products, the Problem child of the BCG matrix.

Product	Model	Test (Model Evaluation)		Validation (Deployment)		
		MAPE	MASE	MAPE	MASE	WD
BGA 11.5 × 13	Naïve forecast + zero-filling	79.190%	131.579%	93.991%	126.491%	2.057%
	Naïve forecast + mean imputation	79.035%	120.234%	93.991%	130.903%	2.057%
	ARIMA + zero-filling	147.469%	87.564%	92.565%	105.350%	−16.326%
	ARIMA + mean imputation	148.739%	81.934%	93.381%	108.863%	−15.218%
	LSTM + zero-filling	37.055%	78.889%	98.114%	113.025%	−23.588%
	LSTM + mean imputation	37.128%	73.817%	97.134%	117.246%	−25.090%

References

- Carbonneau, R.; Laframboise, K.; Vahidov, R. Application of machine learning techniques for supply chain demand forecasting. *Eur. J. Oper. Res.* **2008**, *184*, 1140–1154. [\[CrossRef\]](#)
- Ali, Ö.G.; Sayın, S.; Van Woensel, T.; Fransoo, J. SKU demand forecasting in the presence of promotions. *Expert Syst. Appl.* **2009**, *36*, 12340–12348. [\[CrossRef\]](#)
- Babai, M.Z.; Ali, M.M.; Nikolopoulos, K. Impact of temporal aggregation on stock control performance of intermittent demand estimators: Empirical analysis. *Omega* **2012**, *40*, 713–721. [\[CrossRef\]](#)
- Romeijnders, W.; Teunter, R.; Van Jaarsveld, W. A two-step method for forecasting spare parts demand using information on component repairs. *Eur. J. Oper. Res.* **2012**, *220*, 386–393. [\[CrossRef\]](#)
- Kourentzes, N. Intermittent demand forecasts with neural networks. *Int. J. Prod. Econ.* **2013**, *143*, 198–206. [\[CrossRef\]](#)
- Lau, H.C.; Ho, G.T.; Zhao, Y. A demand forecast model using a combination of surrogate data analysis and optimal neural network approach. *Decis. Support Syst.* **2013**, *54*, 1404–1416. [\[CrossRef\]](#)
- Ma, Y.; Wang, N.; Che, A.; Huang, Y.; Xu, J. The bullwhip effect on product orders and inventory: A perspective of demand forecasting techniques. *Int. J. Prod. Res.* **2013**, *51*, 281–302. [\[CrossRef\]](#)
- Li, C.; Lim, A. A greedy aggregation–decomposition method for intermittent demand forecasting in fashion retailing. *Eur. J. Oper. Res.* **2018**, *269*, 860–869. [\[CrossRef\]](#)

9. Abbasimehr, H.; Khodizadeh Nahari, M. Improving demand forecasting with LSTM by taking into account the seasonality of data. *J. Appl. Res. Ind. Eng.* **2020**, *7*, 177–189.
10. Abbasimehr, H.; Shabani, M.; Yousefi, M. An optimized model using LSTM network for demand forecasting. *Comput. Ind. Eng.* **2020**, *143*, 106435. [[CrossRef](#)]
11. Yuan, X.; Zhang, X.; Zhang, D. Analysis of the Impact of Different Forecasting Techniques on the Inventory Bullwhip Effect in Two Parallel Supply Chains with a Competition Effect. *J. Eng.* **2020**, *2020*, 2987218. [[CrossRef](#)]
12. Kiefer, D.; Grimm, F.; Bauer, M.; Van, D. Demand forecasting intermittent and lumpy time series: Comparing statistical, machine learning and deep learning methods. In Proceedings of the 54th Hawaii International Conference on System Sciences, Kauai, HI, USA, 5 January 2021; p. 1425.
13. Deb, C.; Zhang, F.; Yang, J.; Lee, S.E.; Shah, K.W. A review on time series forecasting techniques for building energy consumption. *Renew. Sustain. Energy Rev.* **2017**, *74*, 902–924. [[CrossRef](#)]
14. Borges, C.E.; Kamara-Esteban, O.; Castillo-Calzadilla, T.; Andonegui, C.M.; Alonso-Vicario, A. Enhancing the missing data imputation of primary substation load demand records. *Sustain. Energy Grids Netw.* **2020**, *23*, 100369. [[CrossRef](#)]
15. Chen, C.; Hu, J.; Meng, Q.; Zhang, Y. Short-time traffic flow prediction with ARIMA-GARCH model. In Proceedings of the 2011 IEEE Intelligent Vehicles Symposium (IV), Baden-Baden, Germany, 5–9 June 2011; pp. 607–612.
16. Kohn, R.; Ansley, C.F. Estimation, prediction, and interpolation for ARIMA models with missing data. *J. Am. Stat. Assoc.* **1986**, *81*, 751–761. [[CrossRef](#)]
17. Arumugam, P.; Saranya, R. Outlier detection and missing value in seasonal ARIMA model using rainfall data. *Mater. Today Proc.* **2018**, *5*, 1791–1799. [[CrossRef](#)]
18. Velicer, W.F.; Colby, S.M. A comparison of missing-data procedures for ARIMA time-series analysis. *Educ. Psychol. Indic.* **2005**, *65*, 596–615. [[CrossRef](#)]
19. Junninen, H.; Niska, H.; Tuppurainen, K.; Ruuskanen, J.; Kolehmainen, M. Methods for imputation of missing values in air quality data sets. *Atmos. Environ.* **2004**, *38*, 2895–2907. [[CrossRef](#)]
20. White, I.R.; Carlin, J.B. Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. *Stat. Med.* **2010**, *29*, 2920–2931. [[CrossRef](#)]
21. Musial, J.P.; Verstraete, M.M.; Gobron, N. Comparing the effectiveness of recent algorithms to fill and smooth incomplete and noisy time series. *Atmos. Chem. Phys.* **2011**, *11*, 7905–7923. [[CrossRef](#)]
22. Wongoutong, C. Imputation Methods in Time Series with a Trend and a Consecutive Missing Value Pattern. *Thail. Stat.* **2021**, *19*, 866–879.
23. Gómez-Carracedo, M.P.; Andrade, J.M.; López-Mahía, P.; Muniategui, S.; Prada, D. A practical comparison of single and multiple imputation methods to handle complex missing data in air quality datasets. *Chemom. Intell. Lab. Syst.* **2014**, *134*, 23–33. [[CrossRef](#)]
24. Junger, W.L.; De Leon, A.P. Imputation of missing data in time series for air pollutants. *Atmos. Environ.* **2015**, *102*, 96–104. [[CrossRef](#)]
25. Nur Afiqah, Z.; Norazian, M.N. Imputation methods for filling missing data in urban air pollution data for Malaysia. *Urbanism. Arhitectura. Constr.* **2018**, *9*, 159.
26. Moritz, S.; Sardá, A.; Bartz-Beielstein, T.; Zaefferer, M.; Stork, J. Comparison of different methods for univariate time series imputation in R. *arXiv* **2015**, arXiv:1510.03924.
27. Teunter, R.H.; Duncan, L. Forecasting intermittent demand: A comparative study. *J. Oper. Res. Soc.* **2009**, *60*, 321–329. [[CrossRef](#)]
28. Chujai, P.; Kerdprasop, N.; Kerdprasop, K. Time series analysis of household electric consumption with ARIMA and ARMA models. In Proceedings of the International MultiConference of Engineers and Computer Scientists, Hong Kong, China, 13–15 March 2013; Volume 1, pp. 295–300.
29. Wang, C.C.; Chien, C.H.; Trappey, A.J. On the Application of ARIMA and LSTM to Predict Order Demand Based on Short Lead Time and On-Time Delivery Requirements. *Processes* **2021**, *9*, 1157. [[CrossRef](#)]
30. Siami-Namini, S.; Tavakoli, N.; Namin, A.S. A comparison of ARIMA and LSTM in forecasting time series. In Proceedings of the 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, USA, 17–20 December 2018; pp. 1394–1401.
31. Martínez, F.; Frías, M.P.; Pérez-Godoy, M.D.; Rivera, A.J. Dealing with seasonality by narrowing the training set in time series forecasting with kNN. *Expert Syst. Appl.* **2018**, *103*, 38–48. [[CrossRef](#)]
32. Johnston, F.R.; Boyland, J.E.; Meadows, M.; Shale, E. Some properties of a simple moving average when applied to forecasting a time series. *J. Oper. Res. Soc.* **1999**, *50*, 1267–1271. [[CrossRef](#)]
33. Babu, C.N.; Reddy, B.E. A moving-average filter based hybrid ARIMA–ANN model for forecasting time series data. *Appl. Soft Comput.* **2014**, *23*, 27–38. [[CrossRef](#)]
34. McKnight, P.E.; McKnight, K.M.; Sidani, S.; Figueredo, A.J. *Missing Data: A Gentle Introduction to Missing Data*; The Guilford Press: New York, NY, USA, 2007.
35. Scheffer, J. Dealing with missing data. *Res. Lett. Inf. Math. Sci.* **2002**, *3*, 153–160.
36. Little, R.J.; Rubin, D.B. *Statistical Analysis with Missing Data*; John Wiley & Sons: Hoboken, NJ, USA, 2019; Volume 793.
37. Hung, C.Y.; Jiang, B.C.; Wang, C.C. Evaluating Machine Learning Classification Using Sorted Missing Percentage Technique Based on Missing Data. *Appl. Sci.* **2020**, *10*, 4920. [[CrossRef](#)]

38. Musil, C.M.; Warner, C.B.; Yobas, P.K.; Jones, S.L. A comparison of imputation techniques for handling missing data. *West. J. Nurs. Res.* **2002**, *24*, 815–829. [[CrossRef](#)] [[PubMed](#)]
39. Gardener, G.; Harvey, A.C.; Phillips, G.D.A. An algorithm for exact maximum likelihood estimation of ARMA models by means of the Kalman filter. *Appl. Stat.* **1980**, *29*, 311–322.
40. Jones, R.H. Maximum likelihood fitting of ARMA models to time series with missing observations. *Technometrics* **1980**, *22*, 389–395. [[CrossRef](#)]
41. Emmanuel, T.; Maupong, T.; Mpoeleng, D.; Semong, T.; Banyatsang, M.; Tabona, O. A Survey on Missing Data in Machine Learning. *J. Big Data* **2021**, *8*, 140. [[CrossRef](#)]
42. Tripathi, A.K.; Saini, H.; Rathee, G. Futuristic Prediction of Missing Value Imputation Methods Using Extended ANN. *Int. J. Bus. Anal. (IJBAN)* **2022**, *9*, 1–12. [[CrossRef](#)]
43. Zhang, Y.; Thorburn, P.J. Handling missing data in near real-time environmental monitoring: A system and a review of selected methods. *Future Gener. Comput. Syst.* **2022**, *128*, 63–72. [[CrossRef](#)]
44. Liu, P.; Ming, W.; Hu, B. Sales forecasting in rapid market changes using a minimum description length neural network. *Neural Comput. Appl.* **2021**, *33*, 937–948. [[CrossRef](#)]
45. Zhang, Y. Sales Forecasting of Promotion Activities Based on the Cross-Industry Standard Process for Data Mining of E-commerce Promotional Information and Support Vector Regression. *J. Comput.* **2021**, *32*, 212–225.
46. Tony, A.; Kumar, P.; Rohith Jefferson, S. A Study of Demand and Sales Forecasting Model using Machine Learning Algorithm. *Psychol. Educ. J.* **2021**, *58*, 10182–10194.
47. Sohrabpour, V.; Oghazi, P.; Toorajipour, R.; Nazarpour, A. Export sales forecasting using artificial intelligence. *Technol. Forecast. Soc. Chang.* **2021**, *163*, 120480. [[CrossRef](#)]
48. Gopagani, D.R.; Lakshmi, P.V.; Chaudhary, A. Evaluating Machine Learning Algorithms for Marketing Data Analysis: Predicting Grocery Store Sales. In *Communication Software and Networks*; Springer: Singapore, 2021; pp. 155–163.
49. Posch, K.; Truden, C.; Hungerländer, P.; Pilz, J. A Bayesian approach for predicting food and beverage sales in staff canteens and restaurants. *Int. J. Forecast.* **2022**, *38*, 321–338. [[CrossRef](#)]
50. Hyndman, R.J.; Athanasopoulos, G. *Forecasting: Principles and Practice*; OTexts: Council Bluffs, IA, USA, 2018.
51. Gómez, V.; Maravall, A.; Pena, D. Missing observations in ARIMA models: Skipping approach versus additive outlier approach. *J. Econom.* **1999**, *88*, 341–363. [[CrossRef](#)]
52. Box, G.E.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. *Time Series Analysis: Forecasting and Control*; John Wiley & Sons: Hoboken, NJ, USA, 2015.
53. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)] [[PubMed](#)]
54. Armstrong, J.S.; Collopy, F. Error measures for generalizing about forecasting methods: Empirical comparisons. *Int. J. Forecast.* **1992**, *8*, 69–80. [[CrossRef](#)]
55. Goodwin, P.; Lawton, R. On the asymmetry of the symmetric MAPE. *Int. J. Forecast.* **1999**, *15*, 405–408. [[CrossRef](#)]
56. Tayman, J.; Swanson, D.A. On the validity of MAPE as a measure of population forecast accuracy. *Popul. Res. Policy Rev.* **1999**, *18*, 299–322. [[CrossRef](#)]
57. Farris, P.W.; Bendle, N.; Pfeifer, P.E.; Reibstein, D. *Marketing Metrics: The Definitive Guide to Measuring Marketing Performance*; Pearson Education: London, UK, 2010.
58. Mohajan, H. An analysis on BCG growth sharing matrix. *Noble Int. J. Bus. Manag. Res.* **2017**, *2*, 1–6.
59. Kwiatkowski, D.; Phillips, P.C.; Schmidt, P.; Shin, Y. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *J. Econom.* **1992**, *54*, 159–178. [[CrossRef](#)]
60. Tofallis, C. A better measure of relative prediction accuracy for model selection and model estimation. *J. Oper. Res. Soc.* **2015**, *66*, 1352–1362. [[CrossRef](#)]
61. Hyndman, R.J. Another look at forecast-accuracy metrics for intermittent demand. *Foresight Int. J. Appl. Forecast.* **2006**, *4*, 43–46.
62. Makridakis, S. Accuracy measures: Theoretical and practical concerns. *Int. J. Forecast.* **1993**, *9*, 527–529. [[CrossRef](#)]