Demand Forecasting Method Proposal for a Packaging Industry at the Manaus Industrial Hub

Dan Ferreira Machado, Marcelo Albuquerque de Oliveira, and Gabriela de Mattos Veroneze

ABSTRACT

Every year, many companies hold meetings with their experts to prospect sales for the following year. Projections are often based on the team's experience and are subject to bias and errors. This article seeks to propose a statistical method of forecasting demand for a packaging company located in the Industrial Pole of Manaus, comparing performance metrics from different analyses. The results suggested that in general, the predictions had a good performance, but in certain cases, the need for a more human vision was also felt in order to ponder certain points. Therefore, the combination of quantitative and qualitative demand forecasting methods is essential.

Keywords: Demand forecast, Packaging industry, Time series.

I. INTRODUCTION

Over the years, the business competition has become more intense each time. Each company has searched for developing ways of standing out on the dispute for market domain and so perpetuate its existence. Within this context, it comes to the demand’s prediction, focusing on supplying to the managers the future sales estimating and helping them on planning the production.

According to Wang and Chang (2010), demand forecasting is fundamental for its planning and, consequently, for the production chain. It is necessary to minimize errors as much as possible to meet the demand of consumers and customers, which, as a result, leads to the achievement of loyalty.

Therefore, the adequate method choice of demand prediction, also called forecasting, enables the company targeting the elaboration of a production plan, storage, and purchases, besides the complex and critical strategies definition to the corporation's success (Cecatto and Belfiore, 2015).

The elaboration of forecasting demand methods can be in a quantitative, qualitative way or the combination of both. The quantitative methods are based on temporal series (demand data over the time), and the qualitative methods are based upon a specialist opinion (Pellegrini and Fogliatto, 2001).

Jain and Malehorn (2006) aim that there isn’t forecasting “magic formula” that assists all the situations. There are at least 70 forecasting methods (Kerkkänen et al., 2009).

Thereby, the most adequate method choice must be taken cautiously and previously, focusing on reducing the maximum of mistakes.

Some conditions are necessary so that forecasting must be able: (i) historical information availability; (ii) possibility of transforming these data into numerical information and (iii) repetition of the supposition from old patterns in the future (Makridakis et al., 1998).

Relating to temporal series, some concepts are important for the better understanding of the subject. According to (Pellegrini and Fogliatto, 2001), period, is the basic time unity utilized on forecasting, generally described within months, weeks, or years, depending on its application. Furthermore, according to Pellegrini and Fogliatto (2001), outlook is the number future period fulfilled by forecasting, determined by the same time unity. They also define seasonality as a variation pattern that occurs during time intervals reasonably increasing or decreasing of the data series over a long period.

II. LITERATURE REVIEW

A. Data Treatment

1) ABC Curve

The ABC curve is based on the 80/20 principle of Vilfredo Paretto. It aims that around 80% of a company’s billing is stemming from 20% from its customers; 15% of the billing
comes from 30% of these and other 5% are derivatives from remnant customers (Dias, 1993).

The focus of this analysis is to recognize the items of bigger demand and so make a better management, especially because these items represent high billing to the organization and planning can entails on big cost reductions (Gonçalves, 2020).

B. Forecasting Methods

1) Exponential Softening

The Exponential Softening (ES) technique is based on the application weighted average on the temporal data series, where bigger weights are assigned to more recent observations and the lower weights to the former ones (Ferro et al., 2019).

The ES has low-cost application and practicality, besides delivering fast results [10]. They are divided in 3: (i) Simple Exponential Softening (SES); (ii) Holt’s Linear Exponential Softening (HES) and (iii) Holt-Winters’s Exponential Softening (HWES) (Makridakis et al., 1998).

The SES is applicable when the series keeps constant during the average, once it is calculated forecasting by the previous weighted sum by the alpha coefficient (\(\alpha\)) (Ferro et al., 2019). The \(\alpha\) varies between zero and one, where the values next to zero assign more weight to recent forecasting, the same way that \(\alpha\) next to one gives the emphasis to recent demands (Morettin and Toloi, 2004). On the other hand, HES also considers the series linear tendency, this is also softened by the new beta (\(\beta\)) coefficient.

The HWES presents itself as an evolution of the previous models, once it takes into account not only the tendency but also the seasonality. It is divided into sub models of additive Holt-Winters Exponential Softening (aHWES) and the Multiplicative Holt-Winters Exponential Softening (mHWES), which the aHWES has constant seasonal amplitude variation, and the mHWES display this amplitude varying over the time (Makridakis et al., 1998).

2) Autoregressive Integrated Moving Average

The model ARIMA (Autoregressive Integrated Moving Average) is an aggregation of three established coefficients like (p), (d) and (q). The (p) represents the autoregressive order to be used on the series, which the autoregressive is defined the relation degree between the current value and the ones from the former series. The coefficient (d) is the number of necessary differentiations to transform the series into motionless and (q) is the maximum order of the mobile averages (Box et al., 2016).

Besides that, there is also the model SARIMA (Seasonal Autoregressive Integrated Moving Average), which includes seasonality on the calculations through the new parameters (P), (D) and (Q), the prediction becomes more precise (Box et al., 2016).

3) Classic Decomposition Model

According to Makridakis et al. (1998), the decomposition model has been proposed in the 20’s decade, and it shows that it is relatively more precise to certain situations and still very used nowadays. Such a model consists in separating the temporal series compounds into three (tendency, cycle, and seasonality), studying them individually and recombining them (Samohyl et al., 2008).

This model is classified as additive and multiplicative. The additive model is required when the oscillation of values around the average keeps constantly, whilst the multiplicative model modifies proportionally the amplitude of data as far as the average also varies (Samohyl et al., 2008).

The tendency is calculated by the half of the difference between the averages of the first and second temporal series half (Makridakis et al., 1998).

The seasonality, when it’s about the additive model, is calculated through the subtraction between the observed value and the tendency, while to the multiplicative model it is given by the reason between them (Samohyl et al., 2008).

The cycle, also called the adjust discrepancy, is the small value remnant from the calculation aforesaid its average must be around zero (Samohyl et al., 2008).

4) Neural Artificial Networks Model

Neural artificial networks (NAN) are systems based on structures of storage and processing of information from a nervous biological system. They are capable of storing and providing experimental knowledge (Machado and Fonseca Júnior, 2013).

A network is built through the entrances and connected through synaptic weights to the processing unity, the neurons. The neuron’s exit is calculated by the activation function, which is given through the product between the entrance, and its correspondent weight added to the constant non-null bias.

The weights and constants are adjusted for each new data entry through a correction vector, in order to minimize the error (Machado and Fonseca Júnior, 2013).

There are still the parameters (p), (P) e (k), which (p) to the autoregressive order of the layer entrance, (P) to the seasonality and (k) as the number of intermediary knots. These parameters are utilized on the software RStudio, which is going to be approached later.

Even though this method is interesting, and it has relative forecasting on its predictions, it has the disadvantage of giving only one-off forecasting, distinctly from the other methods that also generate confidence intervals.

C. Accuracy Measures

1) Akaike’s Information Criterion – AIC

The AIC, the abbreviation in English to Critério de Informação de Akaike is one of the best metrics of accuracy to evaluate forecasting of ES and ARIMA (Walter et al., 2013).

Bueno (2011) asserts that this technique makes the comparison between forecasting’s according to calculations of natural logarithms from the generated residues, which the smallest the value of the AIC index, the most accurate is the forecasting.

2) Root Mean Square Error – RMSE

Another way of calculating forecast is through the RMSE.

This calculation consists in removing the square root of the mistakes found in the forecast. This metric has a huge utility because it is not related to a scale, and it is not limited only to the forecast family ES and ARIMA (Mancuso, 2013).
III. METHODOLOGY

A case study has been done in a package company from Manaus Industrial Hub – MIH. The company has a solid presence on the market delivering products of high quality, which ensures their customers' confidence.

Although it is well structured, the forecast criteria to the demand are done in a qualitative way based on the experience and expectation of the sale team.

The aim of this study is to propose a statistical method to the demand forecast of the factory. The classification can be seen hereinafter according to Table I.

After the classification, it has been realized that the customers C1, C2 and C3 have been responsible for 83.53% of the company’s billing in the quinquennium 2015-2019, so the forecasting techniques are going to be directed to these three. The other customers have been disregarded from the focus of this study.

The software used was RStudio Desktop version 1.3.1093. The choice of this software is justified because it is a free of charge platform, simple and effective, very used in the statistical scope.

The next step was to carry out a statistical demand forecasting method to each product line of the selected customers. It is worth mentioning that the customer C1 has four product lines: L1, L2, L3 and L4. Nonetheless, the customers C2 and C3 have only the lines L5 and L6, respectively.

To each product line did forecasts of types ARIMA, Exponential Softening, Neural Artificial Networks and Decomposing, which each of these models had been applied to the periods of 2015-2019, 2016-2019 and 2017-2019, obtaining 72 forecasts to 2020.

The aim of these different intervals is to verify the influence of former periods on the results of current forecasts. The demands are disposed monthly and in the ton unity.

Lastly, the forecasts have been tested by the Accuracy method RMSE for the choice of better methodologies. The forecast data have been tabled and plotted graphically as the respective chosen models.

IV. RESULTS AND DISCUSSIONS

After the production data import to the software RStudio, it has been done 72 forecasts to 2020, according to what was mentioned in the methodology. It is important to mention that it gives automatically the best model to each demand forecast family: indexes (p), (d) and (q) to the family ARIMA; models SES, HES or HWES to family ES; the additive models or the multiplicative ones in the Decomposition’s case and (p), (P) and (k) to the Neural Networks Forecast.

The RMSE index interpretation must be done by the following way: the smaller the index, more precise is going to be the forecast. The board above suggests that there does not exist a more powerful model of demand forecast.

Below lies, the overview of the best models found according to the accuracy calculation.

Henceforth, each forecast accuracy has been tested with the real value used to the same year. The data are on the following Table II.

A. Best Model to L1: Neural Networks (2,1,2) – 2015 to 2019

The best model found to the product line L1 has been the Neural Networks (2,1,2) between 2015 and 2019. As mentioned before, these coefficients found by the software RStudio represents that the autoregressive has been done between one data and its second predecessor, the seasonality autoregressive has been done within the previous period and the intermediary knots number created has been 2.

It is noticed above that besides being the best model found, the forecasts were distant from what has been done over the same period. There was not demanded for concrete patterns in the latter years that had entailed difficulties with the model on capitating these variations. It is also notable the lack of a confidence interval on the forecasts with NAN's is a great disadvantage to the planning.

The demand forecast from L1 to 2020 is presented in Fig. 1 and Table III.
TABLE II: RMSE INDEXES TO FORECASTING MODELS HELD BETWEEN 2015 AND 2019

<table>
<thead>
<tr>
<th>Period</th>
<th>Model</th>
<th>L1 - RMSE</th>
<th>L2 - RMSE</th>
<th>L3 - RMSE</th>
<th>L4 - RMSE</th>
<th>L5 - RMSE</th>
<th>L6 - RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ES</td>
<td>40,0728</td>
<td>83.0604</td>
<td>12,2511</td>
<td>4,9241</td>
<td>9.7053</td>
<td>9.8310</td>
</tr>
<tr>
<td></td>
<td>NAN</td>
<td>47,8422</td>
<td>119.0594</td>
<td>6,5373</td>
<td>6,1490</td>
<td>12,1058</td>
<td>10,3747</td>
</tr>
<tr>
<td></td>
<td>Decomposition</td>
<td>42,3076</td>
<td>78.8242</td>
<td>12,1113</td>
<td>5,9712</td>
<td>11,3992</td>
<td>9,8412</td>
</tr>
<tr>
<td>2016-2019</td>
<td>ARIMA</td>
<td>54,3042</td>
<td>94.8561</td>
<td>11,5057</td>
<td>4,9474</td>
<td>13,3223</td>
<td>8,0905</td>
</tr>
<tr>
<td></td>
<td>ES</td>
<td>36,4103</td>
<td>84.7292</td>
<td>12,8572</td>
<td>4,9253</td>
<td>11,3040</td>
<td>7,9231</td>
</tr>
<tr>
<td></td>
<td>NAN</td>
<td>39,3799</td>
<td>109.2574</td>
<td>11,4250</td>
<td>10,2032</td>
<td>11,9792</td>
<td>8,6753</td>
</tr>
<tr>
<td></td>
<td>Decomposition</td>
<td>42,3882</td>
<td>90.6455</td>
<td>11,4680</td>
<td>6,0437</td>
<td>12,2544</td>
<td>10,1011</td>
</tr>
<tr>
<td>2017-2019</td>
<td>ARIMA</td>
<td>40,8203</td>
<td>89.2430</td>
<td>11,5350</td>
<td>5,9130</td>
<td>12,0055</td>
<td>11,4157</td>
</tr>
<tr>
<td></td>
<td>ES</td>
<td>36,9434</td>
<td>82.4098</td>
<td>11,5350</td>
<td>4,9734</td>
<td>16,0420</td>
<td>8,2977</td>
</tr>
<tr>
<td></td>
<td>NAN</td>
<td>36,1908</td>
<td>101.2167</td>
<td>27,3350</td>
<td>3,4672</td>
<td>41,4679</td>
<td>7,6303</td>
</tr>
<tr>
<td></td>
<td>Decomposition</td>
<td>45,1779</td>
<td>103.6959</td>
<td>12,9210</td>
<td>7,1387</td>
<td>13,1793</td>
<td>9,4251</td>
</tr>
</tbody>
</table>

![Fig. 1. L1 Forecast Evolution to 2020.](image)

It is noticed above that besides being the best model found the forecasts were distant from what has been done over the same period. There was not demanded for concrete patterns in the latter years that had entailed difficulties with the model on capitating these variations. It is also notable the lack of a confidence interval on the forecasts with NAN’s is a great disadvantage to the planning.

B. Best Model to L2: Decomposition by Addition – 2015 to 2019

The additive decomposition on the quinquennium 2015-2019 has been the model of best accuracy found to L2.

The demand forecast from L1 to 2020 is presented in Fig. 2 and Table IV.

In the decomposition model, differently from the NAN's model, there is the insertion of the confidence interval. It is notable that the forecast was not very precise all of them are hedged by the interval.

C. Best Model to L3: Neural Networks (6,1,4) – 2017 to 2019

The L3 line had as the best model the Neural Networks (6,1,4) to the data from 2017 to 2019. It can be checked on the following results. The demand forecast from L3 to 2020 is presented in Fig. 3 and Table V.

![Fig. 1. L1 Forecast Evolution to 2020.](image)
The L3 line has presented the best accuracy between the six ones studied. Besides not having a confidence interval, the model has shown to be close. It is also observed that it was important removing from the forecast the years 2015 and 2016, once there was not a production during this period.

D. Best Model to L4: Neural Networks (1,1,2) – 2017 to 2019

The NAN's (1,1,2) have been chosen as the best model to L4's line. The triennium from 2017 to 2019 has been the more suitable, according to what can be observed hereinafter. The demand forecast from L4 to 2020 is presented in Fig. 4 and Table VI.

The forecast to this product line has achieved a reasonable precision and possibly, it would be hedged by a confidence interval in case the NAN's provided, except the months of April, September, and October.

<table>
<thead>
<tr>
<th>TABLE VI: L4 Numeric Forecast to 2020</th>
</tr>
</thead>
<tbody>
<tr>
<td>Month</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>jan/20</td>
</tr>
<tr>
<td>feb/20</td>
</tr>
<tr>
<td>mar/20</td>
</tr>
<tr>
<td>apr/20</td>
</tr>
<tr>
<td>may/20</td>
</tr>
<tr>
<td>jun/20</td>
</tr>
<tr>
<td>jul/20</td>
</tr>
<tr>
<td>aug/20</td>
</tr>
<tr>
<td>sep/20</td>
</tr>
<tr>
<td>oct/20</td>
</tr>
<tr>
<td>nov/20</td>
</tr>
<tr>
<td>dec/20</td>
</tr>
</tbody>
</table>

E. Best Model to L5: Exponential Softening (mHWES) – 2015 to 2019

To the line L5 the ideal forecast model found has been the multiplicative Holt-Winters Exponential Softening. Hereinafter, it has presented the forecast to 2020 graphically plotted and disposed on a board. The demand forecast from L5 to 2020 is presented in Fig. 5 and Table VII.

The L5 forecast has been done during the period of 2015 to 2019, capturing the best tendencies and seasonality of demand. It is also notable that the 95% confidence interval has adjusted very well to the results, assuring that all of them must have been contained.
The line L6 has obtained the Neural Artificial Network as the best model, applied to the triennium of 2017 to 2019. According to RStudio, the coefficients (p), (P) and (k) more appropriate have been (1,1,2), the same way that the line L4. The demand forecast from L6 to 2020 is presented in Fig. 6 and Table VIII.

The forecast to L6 has presented good adherence to the results done during 2020, besides the lack of confidence intervals of this forecast family. The model has captured very well the seasonality of the temporal series, with periods of fall in May, November, and December.

**Table VII: L5 Numeric Forecast to 2020**

<table>
<thead>
<tr>
<th>Month</th>
<th>Accomplished</th>
<th>Foreseen</th>
<th>Min (95% of conf.)</th>
<th>Max (95% of conf.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>jan/20</td>
<td>57.25</td>
<td>43.87</td>
<td>26.23</td>
<td>61.52</td>
</tr>
<tr>
<td>feb/20</td>
<td>36.06</td>
<td>34.43</td>
<td>20.35</td>
<td>48.51</td>
</tr>
<tr>
<td>mar/20</td>
<td>45.58</td>
<td>49.72</td>
<td>29.06</td>
<td>70.38</td>
</tr>
<tr>
<td>apr/20</td>
<td>34.92</td>
<td>31.74</td>
<td>18.34</td>
<td>45.14</td>
</tr>
<tr>
<td>may/20</td>
<td>25.72</td>
<td>26.97</td>
<td>15.41</td>
<td>38.53</td>
</tr>
<tr>
<td>jun/20</td>
<td>11.59</td>
<td>48.97</td>
<td>7.67</td>
<td>70.27</td>
</tr>
<tr>
<td>jul/20</td>
<td>40.55</td>
<td>44.16</td>
<td>24.68</td>
<td>63.64</td>
</tr>
<tr>
<td>aug/20</td>
<td>52.36</td>
<td>45.55</td>
<td>25.18</td>
<td>65.93</td>
</tr>
<tr>
<td>sep/20</td>
<td>38.37</td>
<td>36.21</td>
<td>30.72</td>
<td>81.70</td>
</tr>
<tr>
<td>oct/20</td>
<td>59.04</td>
<td>48.06</td>
<td>25.98</td>
<td>70.14</td>
</tr>
<tr>
<td>nov/20</td>
<td>51.25</td>
<td>53.62</td>
<td>28.66</td>
<td>78.57</td>
</tr>
<tr>
<td>dec/20</td>
<td>66.23</td>
<td>57.75</td>
<td>30.53</td>
<td>84.97</td>
</tr>
</tbody>
</table>

**Table VIII: L6 Numeric Forecast to 2020**

<table>
<thead>
<tr>
<th>Month</th>
<th>Accomplished</th>
<th>Foreseen</th>
<th>Month</th>
<th>Accomplished</th>
<th>Foreseen</th>
</tr>
</thead>
<tbody>
<tr>
<td>jan/20</td>
<td>43.48</td>
<td>48.33</td>
<td>jul/20</td>
<td>54.95</td>
<td>47.98</td>
</tr>
<tr>
<td>feb/20</td>
<td>46.16</td>
<td>48.24</td>
<td>aug/20</td>
<td>53.85</td>
<td>48.38</td>
</tr>
<tr>
<td>mar/20</td>
<td>51.46</td>
<td>47.84</td>
<td>sep/20</td>
<td>51.21</td>
<td>48.30</td>
</tr>
<tr>
<td>apr/20</td>
<td>64.46</td>
<td>48.31</td>
<td>oct/20</td>
<td>38.98</td>
<td>48.22</td>
</tr>
<tr>
<td>may/20</td>
<td>39.66</td>
<td>42.63</td>
<td>nov/20</td>
<td>38.87</td>
<td>38.37</td>
</tr>
<tr>
<td>jun/20</td>
<td>42.21</td>
<td>48.85</td>
<td>dec/20</td>
<td>47.46</td>
<td>48.66</td>
</tr>
</tbody>
</table>
V. CONCLUSIONS

The application of demand forecast technique is essential to a good planning of a company and consequently, its spotlight in front of other competitors. Besides the existence of dozens of techniques, many organisations keep on basing on the expertise of the sales team, which many times might not bring the best performance.

Due to this worry, it has been done a case study in a package company of Manaus, with the purpose to propose a statistical method to assist the manager in decision-making.

The six product lines with the highest economic impacts to the corporation have been subjected to quantitative tests of demand forecast helped by software. The performance of each forecast has been variable. The lines L1 and L2 have obtained the worst results because the temporal series variation on the previous years has not been well captured by the chosen models. Nonetheless, whilst L1 has been damaged by the lack of confidence interval, L2 has been fully restrained within the given strip.

The lines L3, L5 and L6 have obtained a good performance on its forecasts to 2020, capturing well the tendency and seasonality of the series. Whereas L4 has obtained a medium result on its forecast, demonstrating also the need of a confidence interval.

The analysis also points out that there is not a forecast family that assists better to all the cases. The Neural Network model has been chosen 4 of the 6 times, while the Exponential Softening and Decomposition families have been chosen once each of them. Whereas ARIMA method has not achieved the best forecast in any cases, which can be explained by the need for much data on the temporal series so it can obtain a better result. The selected periods also vary, which half of the forecasts have adhered more than to the triennium 2017-2019, while the other half has adapted more to quinquennium of 2015-2019.

It is important to mention that the year of 2020 has been an atypical year, and that has suffered a huge impact by the world health crisis caused by the new coronavirus.

The company of the study in question is attached to the chain of food packages, contributing to and expressive growth of its production and its billing, contributing to fluctuations in the result carried.

Effectively, it is notable that the demand forecast study is essential to the planning of a huge company. The choice of the best method must be analysed with very severity, although it does not have to be taken as the only parameter, because it is bound to many economic market variations.

The sales team experience has shown to be primordial to aggregating these forecasts, prospecting these fluctuations, and doing the necessary adjusts on the productive planning.

ACKNOWLEDGMENT

The authors thank the Production Engineering Department (DEP) of the Federal University of Amazonas (UFAM) for their support and structure.

Special acknowledgement to the institutions and people involved to the final consolidation of this work that has been authorized through the agreement 001/2020 from the Cooperação Técnico-Científica celebrated between Fundação Universidade do Amazonas (FUA), Fundação de Apoio ao Ensino, Pesquisa, Extensão e Interiores do IFAM (FAEPI) and the company Samsung Eletrônica da Amazônia LTDA (SEDA-M), within the project’s scope of capacitation and research, which has been done by the academic unity FT-UFAM.

REFERENCES


---

**Dan Machado** is a student of Production Engineering at the Federal University of Amazonas, and researcher under SUPER project.

**Marcelo Oliveira** is an Adjunct Professor of Department of Production Engineering and member of Master Program in Production Engineering, Federal University of Amazonas. He has published several Research Papers related to Engineering area.

**Gabriela Veroneze** is an Adjunct Professor, Department of Production Engineering and member of Master Program in Production Engineering, Federal University of Amazonas. She has published several Research Papers related to Engineering area.