



“School of Social Sciences”

“Master’s Degree Program in Supply Chain Management”

Postgraduate Dissertation

Analyzing sales patterns in a pharmaceutical consumer healthcare
company in Greece: the case of “Haleon Plc”

“Alexandra Ioanna Anastasiadou”

Supervisor: “Nikolaos Thomaidis”

Patras, Greece, “January” “2024”

Dissertations remain the intellectual property of students (“authors/creators”), but in the context of open access policy they grant to the HOU a non-exclusive license to use the right of reproduction, customization, public lending, presentation to an audience and digital dissemination thereof internationally, in electronic form and by any means for teaching and research purposes, for no fee and throughout the duration of intellectual property rights. Free access to the full text for studying and reading does not in any way mean that the author/creator shall allocate his/her intellectual property rights, nor shall he/she allow the reproduction, republication, copy, storage, sale, commercial use, transmission, distribution, publication, execution, downloading, uploading, translating, modifying in any way, of any part or summary of the dissertation, without the explicit prior written consent of the author/creator. Creators retain all their moral and property rights.



Analyzing sales patterns in a pharmaceutical consumer healthcare company in Greece: the case of “Haleon Plc”

“Alexandra Ioanna Anastasiadou”

Supervising Committee

Supervisor:

Nikolaos Thomaidis

Associate Professor

Department of Financial and Management

Engineering

University of the Aegean

Patras, Greece, “January” “2024”

Firstly, i would like to thank my family and friends who continuously support and motivate me to be better and overcome possible difficulties and obstacles. In addition, I would like to express my deep thanks to my supervisor Dr. Nikolaos Thomaidis for his valuable guidance and encouragement throughout the completion of this dissertation.

“In memory of my beloved father”

Abstract

In today's modern era, the inherent uncertainty that shapes product and goods demand, influenced by a multitude of controllable and uncontrollable factors, holds paramount importance across global industries. Key elements such as production materials, essential human resources, efficient supply chain operations, and organizations' capability to meet demand have made the advancement of effective demand forecasting techniques imperative. Additionally, the precise prediction of customer preferences in today's dynamic environment not only enhances overall supply chain performance but also empowers companies to attain competitive advantages, foster business growth, and enhance decision-making processes at large.

Consumer demand in the pharmaceutical sector is significantly influenced by multifaceted factors, both internal and external. Unlike other sectors, the sensitive nature of the healthcare industry is evident in how individuals seek solutions from doctors and local pharmacies, not just for critical health issues but also for psychological well-being or minor concerns. The COVID-19 pandemic significantly altered people's behavior and demand patterns across various product categories. Heightened concerns about health protection against the virus, coupled with apprehensions about potential shortages, prompted shifts in consumer habits. This shift extended beyond pharmaceutical items, impacting various aspects of everyday life. This alteration in behavior emphasized the critical necessity for substantial improvements in forecasting processes and mechanisms, surpassing the needs observed in previous periods.

The objective of this dissertation is to explore and analyze the sales patterns of products supplied in the Greek market by one of the biggest pharmaceutical companies that operates globally. We intend to examine seven over-the-counter products supplied by ‘Haleon Greece’ to warehouses of pharmaceuticals and retail pharmacies throughout Greece. The aim is to identify which are the statistical features of these products’ time series, their behavior throughout a six-years period from 2017 to 2022, the possible seasonal component and COVID-19 pandemic outbreak effect. The study aims to identify robust prediction models by employing exponential smoothing moving average and linear regression forecasting methods. By evaluating their performance accuracy metrics, the goal is to ascertain the most effective models that align with the actual sales data throughout the entire analysis period.

Keywords

Time series analysis, sales patterns, over-the-counter products, consumer healthcare products, pharmaceutical industry, forecasting methods and techniques, statistical properties, prediction accuracy measures and techniques

Ανάλυση προτύπων πωλήσεων καταναλωτικών προϊόντων υγειονομικής περίθαλψης σε μια φαρμακευτική εταιρεία στην Ελλάδα: η περίπτωση της «Haleon Plc»

Αλεξάνδρα – Ιωάννα Αναστασιάδου

Περίληψη

Στη σημερινή σύγχρονη εποχή, η εγγενής αβεβαιότητα που διαμορφώνει τη ζήτηση προϊόντων και αγαθών, επηρεασμένη από ένα πλήθος ελεγχόμενων και μη ελεγχόμενων παραγόντων, έχει ύψιστη σημασία σε όλες τις παγκόσμιες βιομηχανίες. Βασικά στοιχεία όπως τα υλικά παραγωγής, οι βασικοί ανθρώπινοι πόροι, οι αποτελεσματικές λειτουργίες της εφοδιαστικής αλυσίδας και η ικανότητα των οργανισμών να ανταποκρίνονται στη ζήτηση έχουν καταστήσει επιτακτική την προώθηση αποτελεσματικών τεχνικών πρόβλεψης ζήτησης. Επιπλέον, η ακριβής πρόβλεψη των προτιμήσεων των πελατών στο σημερινό δυναμικό περιβάλλον όχι μόνο ενισχύει τη συνολική απόδοση της εφοδιαστικής αλυσίδας, αλλά επίσης δίνει τη δυνατότητα στις εταιρείες να επιτύχουν ανταγωνιστικά πλεονεκτήματα, να προωθήσουν την επιχειρηματική ανάπτυξη και να βελτιώσουν τις διαδικασίες λήψης αποφάσεων γενικότερα.

Η ζήτηση των καταναλωτών στον φαρμακευτικό τομέα επηρεάζεται σημαντικά από πολύπλευρους παράγοντες, εσωτερικούς και εξωτερικούς. Σε αντίθεση με άλλους τομείς, η ευαίσθητη φύση του κλάδου της υγειονομικής περίθαλψης είναι εμφανής στο πώς τα άτομα αναζητούν λύσεις από γιατρούς και τοπικά φαρμακεία, όχι μόνο για κρίσιμα ζητήματα υγείας αλλά και για ψυχολογική ευεξία ή μικρές ανησυχίες. Η πανδημία COVID-19 άλλαξε σημαντικά τη συμπεριφορά και τα πρότυπα ζήτησης των ανθρώπων σε διάφορες κατηγορίες προϊόντων. Οι αυξημένες ανησυχίες για την προστασία της υγείας από τον ιό, σε συνδυασμό με τις ανησυχίες για πιθανές ελλείψεις, προκάλεσαν αλλαγές στις καταναλωτικές συνήθειες. Αυτή η στροφή επεκτάθηκε πέρα από τα φαρμακευτικά είδη, επηρεάζοντας διάφορες πτυχές της καθημερινής ζωής. Αυτή η αλλαγή συμπεριφοράς τόνισε την κρίσιμη ανάγκη για

ουσιαστικές βελτιώσεις στις διαδικασίες και τους μηχανισμούς πρόβλεψης, ξεπερνώντας τις ανάγκες που παρατηρήθηκαν σε προηγούμενες περιόδους.

Στόχος της παρούσας διπλωματικής εργασίας είναι η διερεύνηση και ανάλυση των μοτίβων πωλήσεων των προϊόντων που προμηθεύει στην ελληνική αγορά μια από τις μεγαλύτερες φαρμακευτικές εταιρείες που δραστηριοποιείται παγκοσμίως. Σκοπεύουμε να εξετάσουμε επτά μη συνταγογραφούμενα προϊόντα που προμηθεύει η «Haleon Greece» σε φαρμακαποθήκες και φαρμακεία λιανικής σε όλη την Ελλάδα. Ο στόχος είναι να εντοπιστούν τα στατιστικά χαρακτηριστικά των χρονοσειρών αυτών των προϊόντων, η συμπεριφορά τους σε μια περίοδο έξι ετών από το 2017 έως το 2022, η πιθανή εποχική συνιστώσα και η επίδραση της πανδημίας COVID-19. Η μελέτη στοχεύει στον εντοπισμό ισχυρών μοντέλων πρόβλεψης χρησιμοποιώντας μεθόδους πρόβλεψης κινητού μέσου όρου εκθετικής εξομάλυνσης και γραμμικής παλινδρόμησης. Αξιολογώντας τις μετρήσεις ακρίβειας απόδοσής τους, ο στόχος είναι να εξακριβωθούν τα πιο αποτελεσματικά μοντέλα που ευθυγραμμίζονται με τα πραγματικά δεδομένα πωλήσεων σε όλη την περίοδο ανάλυσης.

Λέξεις-κλειδιά

Ανάλυση χρονοσειρών, μοτίβα πωλήσεων, προϊόντα χωρίς ιατρική συνταγή, καταναλωτικά προϊόντα υγειονομικής περίθαλψης, φαρμακευτική βιομηχανία, μέθοδοι και τεχνικές πρόβλεψης, στατιστικές ιδιότητες, μέτρα και τεχνικές ακρίβειας πρόβλεψης

Table of Contents

Abstract.....	v
Περίληψη	vii
List of Figures.....	xii
List of Tables	xiv
List of Abbreviations & Acronyms	xvi
1. Introduction	1
1.1 Scope and research questions.....	3
1.2 Research outline.....	4
2. Literature Review in Pharmaceutical Sector	5
2.1 Prior Work	6
2.2 The case study: “Haleon Plc”	8
3. Research Methodology.....	12
3.1 Regression analysis.....	13
3.1.1 Model building framework.....	14
3.1.2 Ordinary least squares	17
3.2 Exponentially Weighted Moving Average	19
3.3 Forecast accuracy measures.....	20
3.3.1 Mean Absolute Error	21
3.3.2 Mean Absolute Percentage Error.....	21
4. Empirical Study.....	22
4.1 Data time series selection.....	22
4.2 Data time series presentation	23
4.3 Descriptive Statistics.....	23
4.3.1 Graphical summary per product	27
4.4 Multiple Linear Regression (MLR) model analysis and construction.....	38
4.5 Sales data transformation into natural logarithmic set.....	40
4.6 Statistical significance of the explanatory variables in MLR model	41
4.6.1 Statistical Properties of product P1	41
4.6.2 Statistical Properties of product P2	43
4.6.3 Statistical Properties of product P3	45
4.6.4 Statistical Properties of product P4	46

4.6.5	Statistical Properties of product P5	48
4.6.6	Statistical Properties of product P6	50
4.6.7	Statistical Properties of product P7	51
4.7	Cross-dependencies between the studied products - MLR model M2	53
4.7.1	Interdependencies of product P1 with P2, P3, P4, P5, P6, P7	54
4.7.2	Interdependencies of product P2 with P1, P3, P4, P5, P6, P7	55
4.7.3	Interdependencies of product P3 with P1, P2, P4, P5, P6, P7	56
4.7.4	Interdependencies of product P4 with P1, P2, P3, P5, P6, P7	57
4.7.5	Interdependencies of product P5 with P1, P2, P3, P4, P6, P7	57
4.7.6	Interdependencies of product P6 with P1, P2, P3, P4, P5, P7	58
4.7.7	Interdependencies of product P7 with P1, P2, P3, P4, P5, P6	59
4.8	The effect of the Covid-19 pandemic period - MLR model M3.....	60
4.8.1	Covid-19 impact on sales of product P1	62
4.8.2	Covid-19 impact on sales of product P2	62
4.8.3	Covid-19 impact on sales of product P3	63
4.8.4	Covid-19 impact on sales of product P4	63
4.8.5	Covid-19 impact on sales of product P5	64
4.8.6	Covid-19 impact on sales of product P6	65
4.8.7	Covid-19 impact on sales of product P7	65
4.9	Sales Forecasting Models Designation and Implementation	66
4.9.1	Forecasting MLR model M1 of product P1	67
4.9.2	Forecasting MLR model M1 of product P2	69
4.9.3	Forecasting MLR model M1 of product P3	70
4.9.4	Forecasting MLR model M1 of product P4	72
4.9.5	Forecasting MLR model M1 of product P5	74
4.9.6	Forecasting MLR model M1 of product P6	76
4.9.7	Forecasting MLR model M1 of product P7	78
4.10	EWMA model implementation.....	79
4.10.1	EWMA for product P1	80
4.10.2	EWMA for product P2	81
4.10.3	EWMA for product P3	83
4.10.4	EWMA for product P4	84
4.10.5	EWMA for product P5	85

4.10.6	EWMA for product P6	87
4.10.7	EWMA for product P7	88
4.11	Forecasting performance and accuracy.....	90
5.	Main Findings	91
6.	Conclusions and future research	95
	References	98
	Appendix A: “EWMA in-sample analysis results products P1 to P4”	101
	Appendix B: “EWMA in-sample analysis results products P5 to P7”.....	102
	Appendix C: “EWMA out-of-sample performance results products P1 to P4”.....	103
	Appendix D: “EWMA out-of-sample performance results products P5 to P7”.....	104
	Appendix E: “MLR M1 performance results products P1 to P4”.....	104
	Appendix F: “MLR M1 performance results products P5 to P7”	105

List of Figures

Figure 1, “Timeline of the history of Haleon and the brands in company’s portfolio”.	
Source: Haleon Plc, https://www.haleon.com/who-we-are/history	8
Figure 2, “Revenues 2022 segmentation by geographic operation area”. Source: Haleon Plc, https://www.haleon.com/investors/annual-report-2022	9
Figure 3, “Revenues 2022 segmentation by market category”. Source: Haleon Plc, https://www.haleon.com/investors/annual-report-2022	9
Figure 4, “Revenues 2022 segmentation by geographic operation area”. Source: Haleon Plc, https://www.haleon.com/investors/annual-report-2022	10
Figure 5, “The global consumer healthcare market”. Source: Haleon Plc, https://www.haleon.com/investors/annual-report-2022	10
Figure 6, “The European supply network supporting Greek portfolio”. Source: Haleon Plc,	11
Figure 7, “The global supply network supporting Greek portfolio”. Source: Haleon Plc, ..	11
Figure 8, “Model Building Sequence”. Source: National Institute of Standards and Technology, U.S Department of Commerce, https://www.itl.nist.gov/div898/handbook/pmd/section4/pmd41.htm	15
Figure 9, “The Regression analysis process”. Source: Mentzer and Moon, (2004)	16
Figure 10, “Time series plot of sales P1 2017-2022”	28
Figure 11, “Histogram of P1 sales”	28
Figure 12, “Monthly Boxplot of product P1”	29
Figure 13, “Time series plot of sales P2 2017-2022”	29
Figure 14, “Histogram of P2 sales”	30
Figure 15, “Monthly Boxplot of product P2”	30
Figure 16, “Time series plot of sales P3 2017-2022”	31
Figure 17, “Histogram of P3 sales”	31
Figure 18, “Monthly Boxplot of product P3”	32
Figure 19, “Time series plot of sales P4 2017-2022”	32
Figure 20, “Histogram of P4 sales”	33
Figure 21, “Monthly Boxplot of product P4”	33
Figure 22, “Time series plot of sales P5 2017-2022”	34
Figure 23, “Histogram of P5 sales”	34
Figure 24, “Monthly Boxplot of product P5”	34
Figure 25, “Time series plot of sales P6 2017-2022”	35
Figure 26, “Histogram of P6 sales”	36
Figure 27, “Monthly Boxplot of product P6”	36
Figure 28, “Time series plot of sales P7 2017-2022”	37
Figure 29, “Histogram of P7 sales”	37
Figure 30, “Monthly Boxplot of product P7”	38
Figure 31, “Time plot of forecasted sales of product P1”	68

Figure 32, “Residual Analysis of product P1”	68
Figure 33, “Time plot of forecasted sales of product P2”	70
Figure 34, “Residual Analysis of product P2”	70
Figure 35, “Time plot of forecasted sales of product P3”	72
Figure 36, “Residual Analysis of product P3”	72
Figure 37, “Time plot of forecasted sales of product P4”	74
Figure 38, “Residual Analysis of product P4”	74
Figure 39, “Time plot of forecasted sales of product P5”	75
Figure 40, “Residual Analysis of product P5”	76
Figure 41, “Time plot of forecasted sales of product P6”	77
Figure 42, “Residual Analysis of product P6”	77
Figure 43, “Time plot of forecasted sales of product P7”	79
Figure 44, “Residual Analysis of product P7”	79
Figure 45, “Graphical presentation of P1 product's forecasting models”	81
Figure 46, “Graphical presentation of P2 product's forecasting models”	82
Figure 47, “Graphical presentation of P3 product's forecasting models”	84
Figure 48, “Graphical presentation of P4 product's forecasting models”	85
Figure 49, “Graphical presentation of P5 product's forecasting models”	87
Figure 50, “Graphical presentation of P6 product's forecasting models”	88
Figure 51, “Graphical presentation of P7 product's forecasting models”	89

List of Tables

Table 1 “Products, Brands, Categories and respective code names”	23
Table 2 “Descriptive Statistics of products P1, P2, P3 and P4”	24
Table 3 “Descriptive Statistics of products P5, P6 and P7”	25
Table 4 “Mean and sample variance of P1, P2, P3 and P4 products during years”	26
Table 5 “Mean and sample variance of P5, P6 and P7 products during years”	26
Table 6 “Regression summary with monthly dummies for product P1”	42
Table 7 “Regression summary with monthly dummies and lag variables for product P1”	42
Table 8 “Regression summary with monthly dummies for product P2”	43
Table 9 “Regression summary with monthly dummies and lag variables for product P2”	44
Table 10 “Regression summary with monthly dummies for product P3”	45
Table 11 “Regression summary with monthly dummies and lag variables for product P3”	46
Table 12 “Regression summary with monthly dummies for product P4”	47
Table 13 “Regression summary with monthly dummies and lag variables for product P4”	47
Table 14 “Regression summary with monthly dummies for product P5”	48
Table 15 “Regression summary with monthly dummies and lag variables for product P5”	49
Table 16 “Regression summary with monthly dummies for product P6”	50
Table 17 “Regression summary with monthly dummies and lag variables for product P6”	51
Table 18 “Regression summary with monthly dummies for product P7”	52
Table 19 “Regression summary with seasonal dummies and lag variables for product P7”	52
Table 20 “Summary output of M2 model investigating cross-dependencies between product P1 and the remaining six products”	55
Table 21 “Summary output of M2 model investigating cross-dependencies between product P2 and the remaining six products”	55
Table 22 “Summary output of M2 model investigating cross-dependencies between product P3 and the remaining six products”	56
Table 23 “Summary output of M2 model investigating cross-dependencies between product P4 and the remaining six products”	57
Table 24 “Summary output of M2 model investigating cross-dependencies between product P5 and the remaining six products”	58
Table 25 “Summary output of M2 model investigating cross-dependencies between product P6 and the remaining six products”	59
Table 26 “Summary output of M2 model investigating cross-dependencies between product P7 and the remaining six products”	60
Table 27 “Summary output of M3 model investigating possible impact of Covid-19 pandemic on product P1”	62
Table 28 “Summary output of M3 model investigating possible impact of Covid-19 pandemic on product P2”	63
Table 29 “Summary output of M3 model investigating possible impact of Covid-19 pandemic on product P3”	63

Table 30 “Summary output of M3 model investigating possible impact of Covid-19 pandemic on product P4”	64
Table 31 “Summary output of M3 model investigating possible impact of Covid-19 pandemic on product P5”	64
Table 32 “Summary output of M3 model investigating possible impact of Covid-19 pandemic on product P6”	65
Table 33 “Summary output of M3 model investigating possible impact of Covid-19 pandemic on product P7”	65
Table 34 “Summary output of M1 model implemented in product P1”	67
Table 35 “Summary output of M1 model implemented in product P2”	69
Table 36 “Summary output of M1 model implemented in product P3”	71
Table 37 “Summary output of M1 model implemented in product P4”	73
Table 38 “Summary output of M1 model implemented in product P5”	75
Table 39 “Summary output of M1 model implemented in product P6”	77
Table 40 “Summary output of M1 model implemented in product P7”	78
Table 41 “MAE and MAPE results for optimal smoothing parameter λ of EWMA model”	90
Table 42 “R-square, MAE and MAPE results for MLR model M1”	90
Table 43 “Comparison matrix of MAPE rate for all products and forecasting methodologies”	94

List of Abbreviations & Acronyms

ANNs	Artificial Neural Networks
ARIMA	Autoregressive Integrated Moving Average
DFM	Demand Forecast Model
DSI	Demand and Supply Integration
ERP	Enterprise Resource Planning
EWMA	Exponential Weighted Moving Average
LSE	London Stock Exchange
MAD	Mean Absolute Deviation
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MLR	Multiple Linear Regression
MSE	Mean Square Error
NYSE	New York Stock Exchange
OLS	Ordinary Least Squares
OTC	Over-the-Counter
RMSE	Root Mean Square Error
S&OP	Sales and Operational Planning
SD	Standard Deviation
SLR	Simple Linear Regression
SSR	Sum of Squared Residuals
TOC	Theory of Constraints
VMS	Vitamins, Minerals and Supplements
VOC	Voice Of Customers
WHO	World Health Organization

1. Introduction

In current era, the organizations that operate in a global market level and not only, are continuously seeking to find the optimal solutions on balancing the consumers' demand needs and the effective supply strategies. As it was stressed in the research of de Albuquerque Dallegrave et al., (2021), in the face of a progressively competitive business landscape, companies strive to enhance their operations to optimize profits and minimize waste, focusing on gaining effective competitive advantages and improving business performance. In addition, the study underlines the fact that efficient management necessitates correct planning and the cooperative and coordinated collaboration of various functions within the supply chain to achieve synergies and objectives. Those functions require a comprehensive understanding of future conditions in which the company will operate and how the elements influencing this outlook interconnect.

Stank et al., (2012) in their research support that companies adopting cost leadership strategies often prioritize operational efficiency but may overlook the importance of understanding the 'voice of the customer' (VOC). Conversely, those pursuing differentiation strategies prioritize customer needs but may compromise operational excellence. Furthermore, they note that strategic position accomplishments involve a consideration of both the VOC, as determined by the demand creation processes within an organization, and the voice of the supply chain as determined by the supply chain processes within and outside the company. The organizations, in order to mitigate these imbalances, should incorporate the supply chain research that the concept of demand and supply integration (DSI) introduces. DSI entails aligning customer-focused activities and processes with the operational and supply-side activities essential for fulfilling demand.

According to Moon (2018), the collaborative approach to supply chain functions within the framework of DSI is quite essential. In its optimal state, DSI represents the business planning process that involves sharing knowledge of market demand, identifying constraints in supply capacity, considering corporate financial targets, and strategically overseeing and deciding on sustainable value chain initiatives. This decision-making process significantly influences the key aspects of supply chain operations, including sales and operational planning (S&OP), delivery and replenishment times, demand planning, order fulfillment, inventory control, and more. Central to these processes is the reliance on future demand estimates, which play a

crucial role in providing essential insights into market dynamics and consumer behaviors. Therefore, sales forecasting emerges as a pivotal element in guiding decision-making procedures.

As many researches in the past have shown, forecasting demand plays a crucial role in the management of business processes. Despite the intricacies and variations in the execution of forecasting procedures across various businesses, the fundamental objective remains consistent. The optimal purpose is to derive a reasonably precise estimate of future demand conditions for a product or service by considering historical data and the current environmental factors (such as political, social, and economic features). This enables businesses to design, organize and shape their operations effectively. Achieving accurate forecasts continues to be a significant challenge, particularly within the pharmaceutical industry, as the supply chain in this sector is characterized by increasing levels of complexity (Merkuryeva, G., Valberga, A., & Smirnov, A., 2019).

The planning of production stands out as a crucial procedure within the pharmaceutical industry, exerting a pivotal influence in a competitive market environment. Given the inherent uncertainty, forecasting tools and techniques are extensively employed as valuable tools for decision-making. These forecasting methods contribute significantly and play a crucial role, in supporting decision makers in the realm of production planning (Goodwin, 2005). Moreover, the demand forecasting capabilities provide companies important competitive advantages and represent a substantial and intricate subprocess within the domain of supply chain management. Effective forecasting methods could be characterized as a systematic process involving the formulation of models and the computation of associated quantities in an on-time way. This process facilitates precise planning across all aspects of the supply chain (İMECE, S., & BEYCA, Ö. F., 2022).

Various forecasting methods exist, and it is widely acknowledged that no single method can be deemed universally superior to others. The prevalent methods documented in the literature are categorized as either quantitative or qualitative/judgmental techniques. This dissertation study focuses on the quantitative analysis of forecasting methods, specifically delving into the examination of historical sales data of a pharmaceutical company. The analysis utilizes time series data on a monthly basis and incorporates causal techniques, employing methods such as exponential smoothing and ordinary least squares (OLS) linear regression models.

Subsequently, a comparison of the Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) indices between the aforementioned methods is undertaken to derive valuable and pertinent conclusions.

1.1 Scope and research questions

The aim of the current dissertation is to investigate the sales time series of seven out-of-order products supplied by ‘Haleon Greece’ to warehouses of pharmaceuticals and retail pharmacies throughout the country, in order to detect the features and analyze the sales patterns of these products. Furthermore, the thesis purpose is to make the appropriate conclusions about the products’ sales time series, to explore the potential impact of the Covid-19 pandemic period and to finalize with the most suitable models in providing useful estimations and forecasting benefits in a continuously evolving and unstable market environment.

By applying an array of time series analysis and regression models spanning a period of six years from January of 2017 to December of 2022, we attempt to conduct a comprehensive analysis of time series data oriented to arriving at accurate and reliable estimations and yielding valuable insights for the company. Such information is very important and necessary for decision-making purposes. The time period that we focused on, is deliberately chosen to include periods before, during and after the pandemic time-span.

Through a thorough study of the time series properties of seven products’ sales in the case of “Haleon Plc.” in Greece in the current thesis, we attempt to answer the following research questions:

1. Which are the main patterns and the statistical properties of the products’ sales time series? Which of these statistical properties are more decisive in shaping the course of sales over time? Are there specific product codes with stronger seasonal signature?
2. Do sales of different product types exhibit interdependent behavior?
3. What is the typical accuracy rate at which monthly sales can be forecasted? Does this accuracy vary across product types?
4. How did the Covid-19 pandemic outbreak affect sales and the totality of the examined products sales patterns?

1.2 Research outline

The present dissertation is structured by six main chapters, each addressing the previously mentioned research questions. The first chapter serves as an introductory overview of the study, where the research scope is examined, and the research questions are clearly presented to the reader. This chapter also outlines the overall structure and development of the thesis.

In the second chapter, a concise review of literature pertaining to forecasting concepts within the pharmaceutical industry is outlined. This section emphasizes the pivotal role of forecasting development and underscores the significance of demand forecasting, particularly the integration of Time Series and Regression Models, within pharmaceutical supply chains. Furthermore, a comprehensive description of the case company, forming the empirical foundation of this study, will be presented.

In the third chapter, the research methodology employed to provide insights and establish theoretical groundwork for the subsequent empirical research analysis in the fourth chapter is presented in details. This chapter extensively covers the forecasting methods, statistical methodologies, and tools that will be utilized in analyzing the case study sales time series. The primary purpose is to assess the dynamic behavior of the sales trajectory and evaluate the forecasting accuracy of the models developed through this analysis.

In chapter four, the empirical study is developed offering a comprehensive analysis of the sales data of company's products that will be examined. This section delves into a detailed investigation of the sales patterns and it investigates the potential interdependencies among various product sales by employing fundamental statistical methods. The chapter also investigates the impact of unexpected global events, such as the outbreak of the recent Covid-19 pandemic, on the sales flow. Concluding this chapter, the validity of the forecasting models will be assessed using common accuracy measures. These measures will form the basis for the statistical evaluation and interpretation of the final outcomes derived from the study.

In the fifth chapter, a conclusive summary is provided, encapsulating the primary inferences derived from the empirical analysis. The focus is on summarizing and discussing the research questions that were outlined in the first chapter.

The final sixth chapter serves as a culmination, consolidating the conclusions that were drawn and the insights derived from the dissertation's outcomes. This section encapsulates the knowledge gained through the results obtained, highlighting key takeaways. Additionally, it proposes potential topics for future research within the domain of forecasting analysis, aiming to expand upon and enhance the existing body of knowledge in this field.

2. Literature Review in Pharmaceutical Sector

The pharmaceutical sector universally stands as one of the most rapidly evolving and consistently advancing industries worldwide. A multitude of multinational corporations thrive within this domain, yielding substantial profits and actively contributing to the advancement of pharmaceutical products, encompassing both prescription and non-prescription items, as well as overall consumer healthcare products.

Forecasting demand stands as a crucial component within business process management. Despite variations in complexity and implementation across diverse industries, the fundamental objective remains consistent: deriving a reasonably precise projection of forthcoming product or service demand by leveraging historical data and current environmental conditions (such as political, social, and economic factors) to effectively structure and manage business operations. Achieving accuracy in forecasting continues to pose a significant challenge within the pharmaceutical sector.

The complexity that characterizes the pharmaceutical industry supply chain consists an obstacle in enhancing the chain's overall performance and effectiveness. Forecasts of demand serve as the cornerstone for all strategic and logistical planning within pharmaceutical supply chain management. Still, forecasting remains a relatively recent undertaking within the pharmaceutical sector, potentially elucidating the prevalent use of rudimentary methodologies primarily executed through Excel spreadsheets (Merkuryeva, G., Valberga, A., & Smirnov, A., 2019).

Rathipriya et al., (2023) claim that demand forecasting represents a systematic and scientific evaluation of the anticipated future demand for critical products. The authors in their research point out that an efficient Demand Forecast Model (DFM) serves as a pivotal tool for pharmaceutical enterprises, empowering their success within the global market. Applying such machine learning tool, companies achieve to effectively match their supply capacity

with the continuously changing demand requirements, and, simultaneously keeping inventories at plausible minimum levels.

As previously emphasized, the pharmaceutical companies' supply chain comprises intricate and interconnected activities, contributing to a multitude of challenges faced by these entities worldwide. Effective management of this complex supply chain is paramount for organizations, especially within the critical service sector, and this rings particularly true for the pharmaceutical industry. The period during the Covid-19 pandemic unveiled various issues and weaknesses within global logistics chains, impacting the efficiency of even the well-established global pharmaceutical sector.

2.1 Prior Work

Numerous studies conducted in the past and in the contemporary era underscore the paramount importance of the pharmaceutical system and its proficient supply chain. Shashi (2023) in his article through a qualitative analysis and research on a multiple-case study, highlights the significance of the effective information digitalization, the beneficial data sharing practices and the importance of the blockchain technologies within the pharmaceutical industry and its supply chain management. The author contends that heightened complexities and unforeseen disruptions in global supply chains necessitate attention from managers and industry leaders in the pharmaceutical sector. Success, as posited by the author, lies in the meticulous recording, exploitation, and monitoring of information, thereby fostering improved cost-effective digital sustainability, heightened profitability, and long-term viability through cost reduction and enhanced performance.

Many scholars employ the Theory of Constraints (TOC) as a framework to identify, leverage, and eliminate constraints within their operations (Trojanowska, Justyna & Dostatni, Ewa., 2017). Constraints may manifest in various facets of manufacturing, spanning the supply chain, logistics, or internal processes. The adoption of sustainable digitalized supply chain management in a pharmaceutical company holds the potential to achieve several benefits, including the reduction of operational costs, enhancement of asset management, augmentation of shareholders' value, responsiveness to customer demands, and the generation of profits (Shashi, 2023).

In their paper, Wang et al., (2013) underscored the significance of embracing hybrid models for time series forecasting. They highlighted the efficacy of simultaneously leveraging the strengths of both Autoregressive Integrated Moving Average (ARIMA) and Artificial Neural Networks (ANNs) models. This hybrid approach was proposed to effectively address the attributes of both linear and nonlinear behaviors encountered within time series datasets.

In a recent study conducted by Siddiqui et al., (2022), the importance of precise forecasting within the pharmaceutical sector was emphasized. The researchers aimed to address demand uncertainty by introducing a hybrid forecasting model named ARWOH. This model combined two statistical methods, the Autoregressive Integrated Moving Average Model (ARIMA) and Holt’s Winter forecasting method. Through this fusion, the authors sought to improve forecasting accuracy, successfully demonstrating enhanced outcomes based on metrics such as Mean Absolute Percentage Error (MAPE), goodness of fit tests (including Mean Absolute Deviation (MAD), Mean Square Error (MSE), and Root Mean Square Error (RMSE)). Their evaluation highlighted the effectiveness of the hybrid model, suggesting the potential implementation of similar techniques for forecasting practices within the pharmaceutical industry.

In another recent study conducted by Rathipriya et al., (2023) the emphasis was placed on the advancement of diverse forecasting methodologies within the pharmaceutical industry. Their focus lay in investigating the ARIMA model performance and validating both shallow and deep neural network methods for demand forecasting in time series dataset. The researchers innovated by applying different forecasting techniques and assessed their accuracy by comparing the RMSE results. This evaluation aimed to recommend tailored sales and marketing strategies based on the trend and seasonal impacts across eight distinct groups of pharmaceutical products, each characterized by unique features.

In this thesis, our aim is to assess the most suitable and precise forecasting tool that can offer the research case company valuable insights into sales course over time. Our goal is to identify interconnections among various product categories and the impact of unexpected events, facilitating informed planning and effective decision-making processes in the volatile contemporary business landscape. In that manner, our objective is to make a meaningful contribution to the existing literature on sales forecasting methodologies.

2.2 The case study: “Haleon Plc”

“Haleon” is a world-leading consumer healthcare company with a clear purpose to deliver better everyday health to its customers with the contribution of science, innovation and the deep human understanding. The company emerged as the rebranded entity of “GSK Consumer Health” subsequent to its separation from the Group, establishing itself as an independent company in mid-2022. Originally, it was a result of the amalgamation between the consumer healthcare businesses of GSK, Novartis and Pfizer, which decided to focus on and prioritize their prescription-based products. On 18 July 2022, “Haleon” listed as an independent company on the London (LSE) and New York (NYSE) stock exchanges, aiming to outperform its competitors and strengthen its position by possessing a world-class portfolio of brands and an attractive geographic footprint. The company's primary aim is to eventually gain competitive advantages through technical and scientific innovations (Haleon Plc, 2024). The establishment of the company and the subsequent stages leading to its complete segregation and independence from the parent corporations are outlined and elucidated in **Figure 1**.

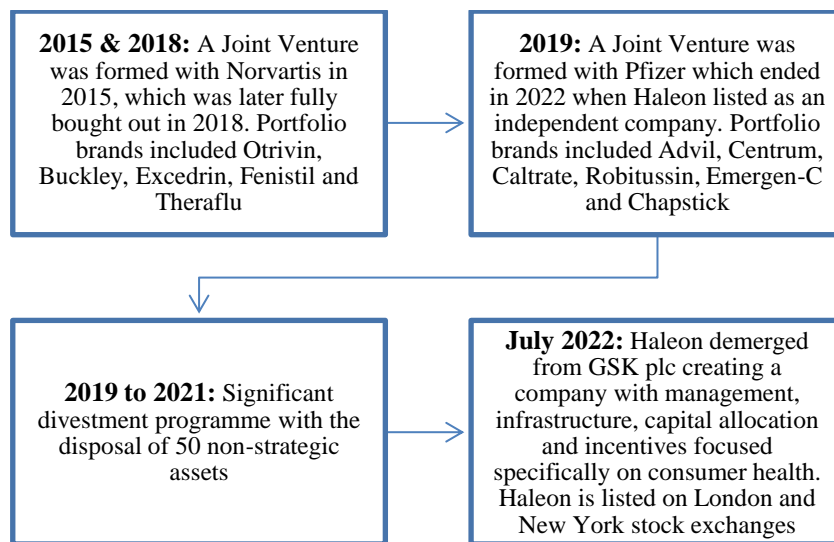


Figure 1, “Timeline of the history of Haleon and the brands in company’s portfolio”. Source: Haleon Plc, <https://www.haleon.com/who-we-are/history>

“Haleon Plc” is a multinational corporation based in Weybridge, England, and operates within five distinct market categories: Oral health, VMS (Vitamins, Minerals, and Supplements), and the Over-the-Counter (OTC) category which encompasses products relevant with Pain relief, Respiratory Relief, Digestive Relief, and others. According to the 2022 annual report, the company's revenues were derived with 27% and 15% from the first Postgraduate Dissertation

two categories respectively, while the remaining 58% originated from the last category of products. The following **Figures 2-4** illustrate information about the how the company’s revenues in 2022 segmented depending on the operating geographic area and the products’ category. In addition, **Figure 5** showcases the reasons why the global consumer healthcare market is defined as one of the largest, most resilient and fastest-growing segments across the consumer staples space, reaching £160bn+ in global value in 2022.

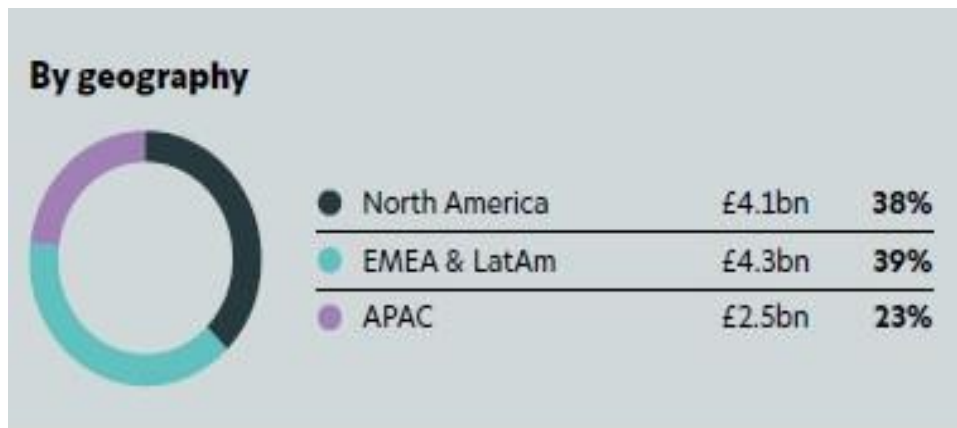


Figure 2, “Revenues 2022 segmentation by geographic operation area”. Source: Haleon Plc, <https://www.haleon.com/investors/annual-report-2022>

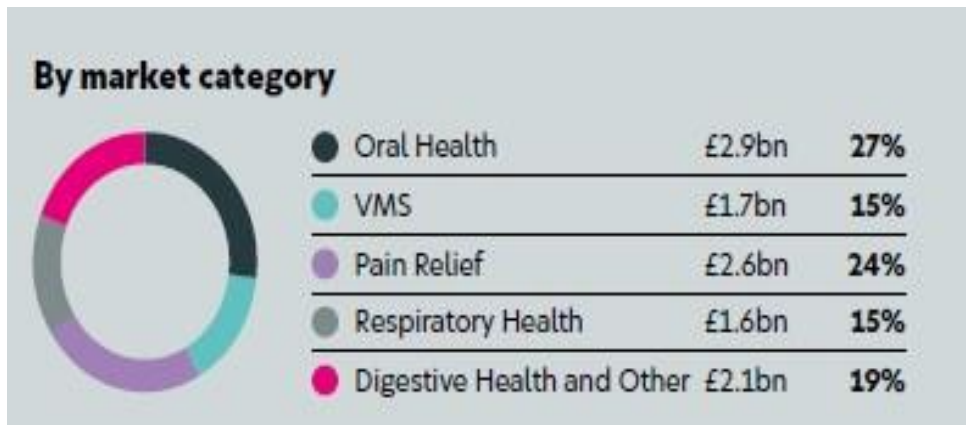


Figure 3, “Revenues 2022 segmentation by market category”. Source: Haleon Plc, <https://www.haleon.com/investors/annual-report-2022>

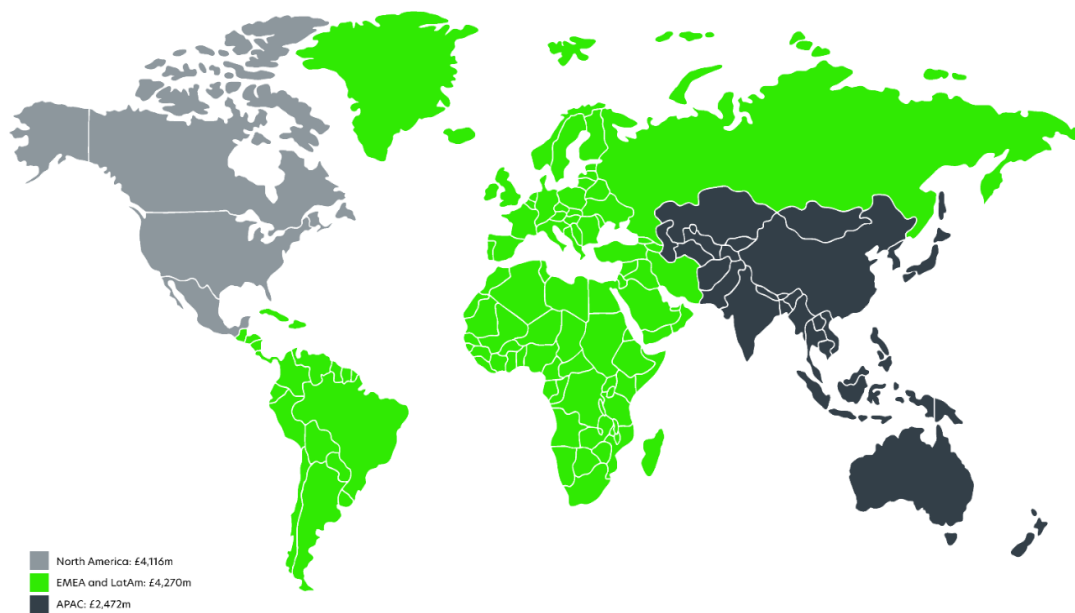


Figure 4, “Revenues 2022 segmentation by geographic operation area”. Source: Haleon Plc, <https://www.haleon.com/investors/annual-report-2022>

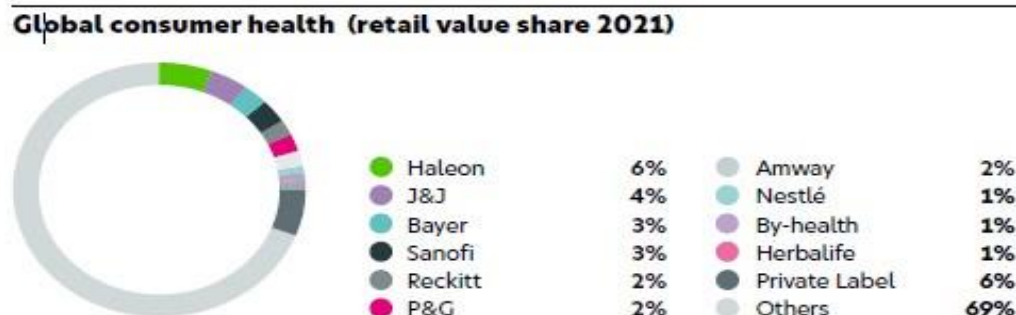
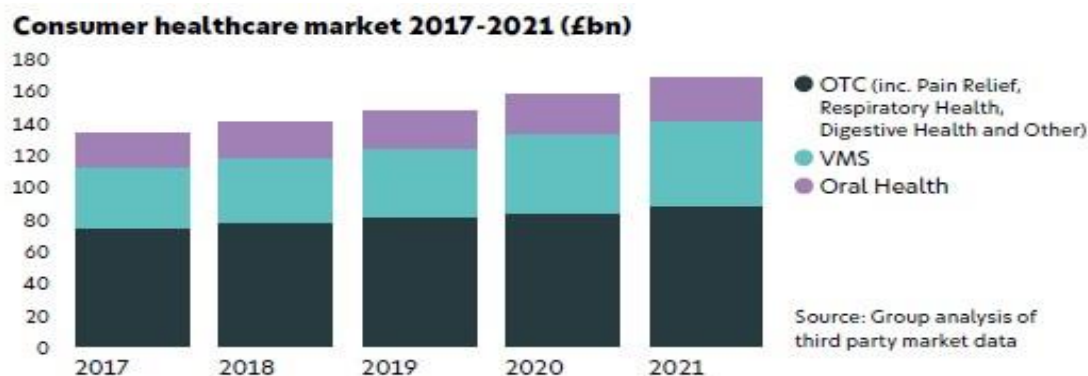


Figure 5, “The global consumer healthcare market”. Source: Haleon Plc, <https://www.haleon.com/investors/annual-report-2022>

Renowned as a global leader in consumer healthcare products, the company operates extensively worldwide, including within the Greek market, which constitutes the focal point

of our interest. Within Greece, “Haleon” services 100 pharmacy warehouses and collaborates with 3,500 retail pharmacies, representing a portion of the total 11,000 operational pharmacies in the country. Sales data, consisting of monthly sales volume measured in thousands of items for seven specific products, have been provided by the company's finance department and form the basis of examination within the current dissertation. The Greek market and the factories supporting Greek portfolio are showcased in the **Figures 6 and 7** below.

Factories supporting Greek portfolio



Supply EU Network

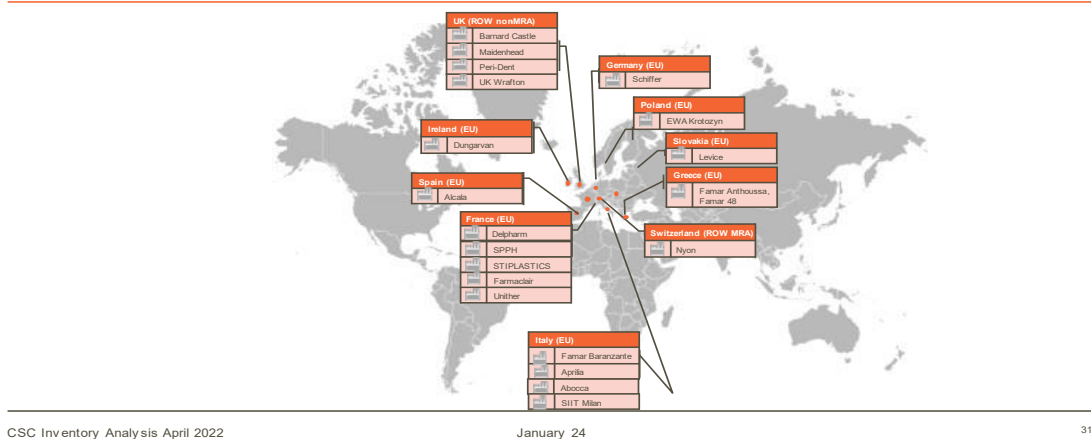


Figure 6, “The European supply network supporting Greek portfolio”. Source: Haleon Plc,

Factories supporting Greek portfolio



Supply Global Network



Figure 7, “The global supply network supporting Greek portfolio”. Source: Haleon Plc,

3. Research Methodology

This chapter is dedicated to scrutinizing the methodology employed to investigate the research questions outlined in the initial chapter of this dissertation. The goal is to offer the reader a comprehensive understanding of the methodology applied in the study, specifically how it was implemented in the context of “Haleon Plc” in Greece. The study primarily involves the utilization of quantitative statistical tools, data analysis techniques, and forecasting methods. These tools are employed to draw conclusions from observed data pertaining to the sales volume of seven over-the-counter consumer healthcare products supplied by the company in the Greek market.

In addition to collecting data and presenting the monthly sales data spanning a six-year period from January 2017 to December 2022 for seven over-the-counter consumer healthcare products ($P_k, k = 1, 2, \dots, 7$), the study initiates by examining their statistical attributes using Excel's descriptive statistics tool. The primary objective is to discern the fundamental statistical properties of the dataset pertaining to each product. Subsequently, insights regarding potential outliers, trends, and cycles will be derived through a graphical summary and boxplot graphs. This analysis aims to facilitate the extraction of conclusions about the data set's characteristics and underlying patterns for each specific product.

Moreover, employing Multiple Linear Regression (MLR) models for each product aims to evaluate their statistical characteristics, ascertain potential seasonality effects, identify possible autocorrelation conditions, unveil interdependencies among the sales datasets, and assess the impact of the Covid-19 pandemic period on these variables. Each facet of this investigation involves the implementation of a respective MLR model. Based on the outcomes, new variables will be introduced or existing ones will be eliminated to address the research questions and determine the most appropriate regression model. The finalized model will comprise solely those variables that demonstrate 5% statistical significance.

Following this, two forecasting models will be formulated and their outcomes compared and assessed using metrics such as Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). A proficient quantitative forecasting tool holds significant value across various market industries, contributing to enhanced business performance, acquisition of advantageous competitive positions, and streamlining decision-making processes, as expounded upon in the introduction chapter. The estimation of these forecasting models aims

to determine the most effective forecasting tool tailored to each product, facilitating informed decision-making and optimizing business outcomes.

3.1 Regression analysis

Regression (correlation) analysis stands as one of the most frequently utilized qualitative tools in market research, primarily employed to scrutinize the cause-and-effect relationship between dependent and independent variables. In marketing contexts, the dependent variable typically represents the primary focus, such as sales, while the independent variables encompass those factors presumed to exert an influence on the dependent variable. Implementing regression analysis yields valuable insights regarding if the dependent variable exhibits significant correlation with the independent variables and it elucidates the extent to which variations in the dependent variable can be elucidated by the independent ones. Furthermore, regression analysis allows for estimations concerning the strength of relationships and effects between the dependent and independent variables. Ultimately, it delivers forecasting outcomes and enables an evaluation of the effectiveness of predictions derived from regression models. (Sarstedt, 2019). The regression models, whether they are simple linear regression (SLR) or multiple regression models (MLR), are commonly represented by the following equation:

For Simple Linear Regression:

$$Y = \beta_0 + \beta_1 X_1 + e \quad (3.1.1)$$

For Multiple Linear Regression:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_j X_j + e \quad (3.1.2)$$

where β_0 is the intercept or sometimes called the constant term in the regression models signifying the mean value of the dependent variable when all other variables are set to zero, β_j denotes the coefficients or slopes associated with each explanatory variable X_j illustrating the effect of these predictors X_j on the response variable Y . Additionally, e represents the “disturbance term” or “error term” within the model, capturing the variability in the dependent variable that remains unpredictable by the model itself.

3.1.1 Model building framework

In general, the fundamental steps involved in constructing a model remain consistent across various modeling methods. While specific details may differ among methods, comprehending the shared steps, coupled with the typical underlying assumptions required for the analysis, establishes a framework enabling the interpretation and comprehension of results derived from nearly any method.

The iterative process of model development involves three key steps. First, in the model selection phase, data plots, process knowledge, and assumptions guide the determination of the appropriate model form that best fits the data. Next, employing the chosen model and potentially additional data insights, a suitable model-fitting technique is applied to estimate the unknown parameters within the model. Once parameter estimates are obtained, a thorough assessment is conducted to verify the plausibility of the assumptions underlying the analysis. Validated models can then be utilized in many sectors and industries. However, if model validation reveals discrepancies or issues, the modeling process is repeated, incorporating insights from the validation step to select or fit an improved model. The **Figure 8** summarizes the aforementioned procedure that is generally applied for all the model-building processes.

The creation of a regression model involves developing a probabilistic representation that accurately captures the relationship between dependent and independent variables. The primary challenges lie in identifying the appropriate relationship form and pinpointing the independent variables with the most substantial impact. This analysis serves as an effective sales forecasting technique, depicted in **Figure 9**, outlining the fundamental steps in this process (Mentzer and Moon, 2004).

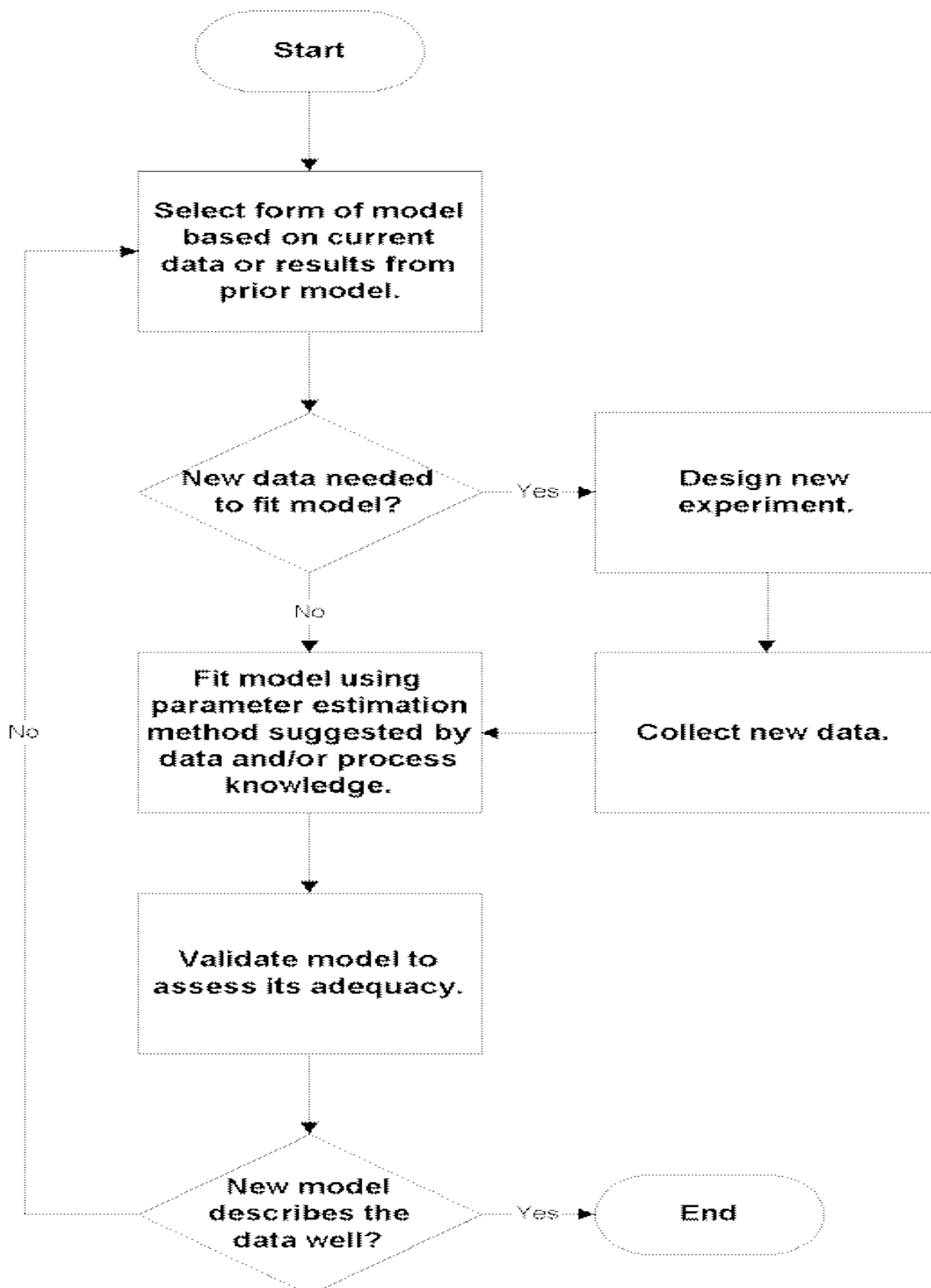


Figure 8, “Model Building Sequence”. Source: National Institute of Standards and Technology, U.S Department of Commerce, <https://www.itl.nist.gov/div898/handbook/pmd/section4/pmd41.htm>

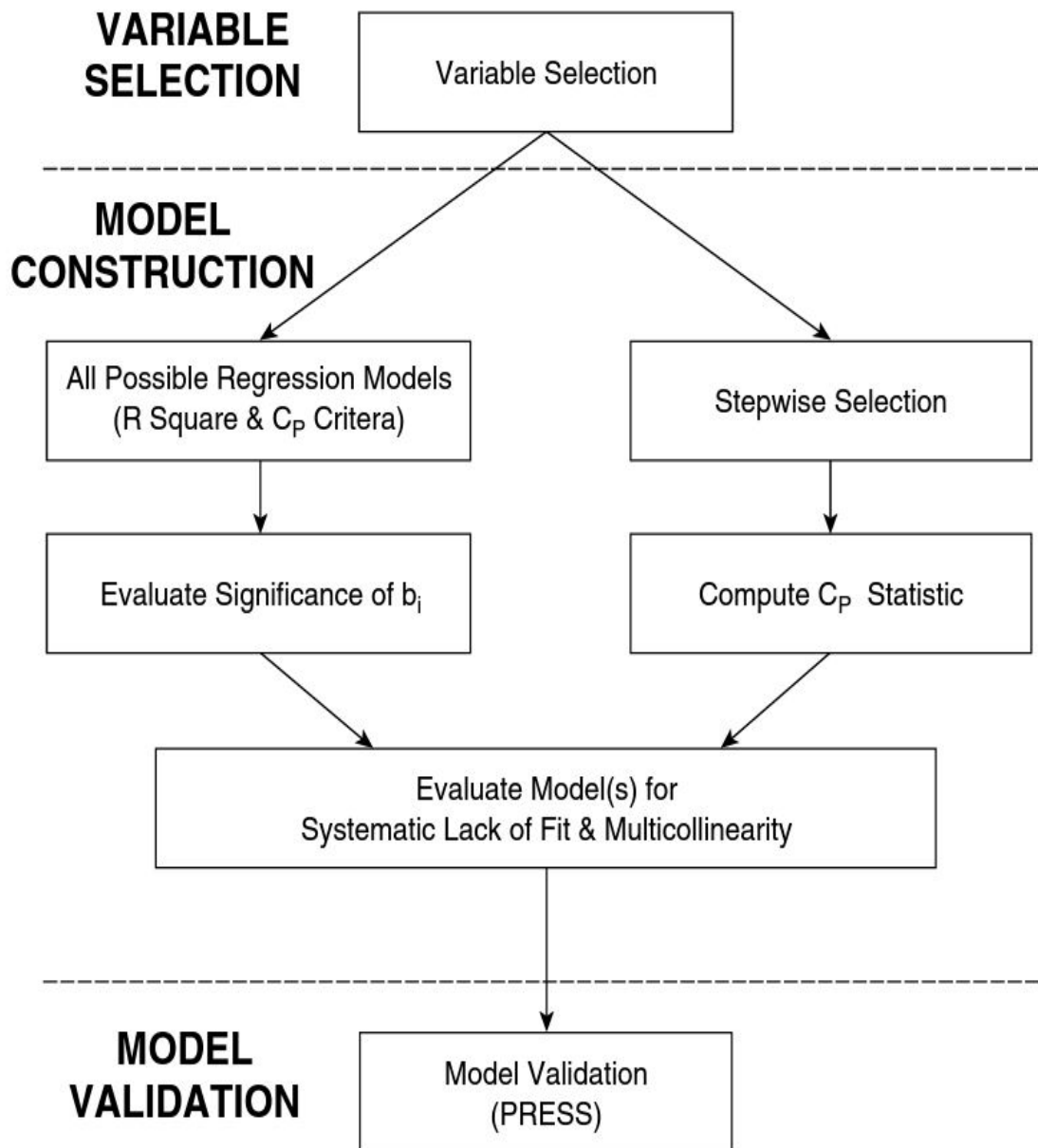


Figure 9, “The Regression analysis process”. Source: Mentzer and Moon, (2004)

The initial step, variable selection, forms the foundation of the entire regression procedure. This phase heavily relies on the analysts' abilities and insights to discern the factors significantly impacting customers' demand. Moving on to the subsequent stage, constructing the model involves employing various statistical measures essential for evaluating the significance of each parameter. The assessment criteria for evaluating the model includes several factors. These factors incorporate the r-square value, which signifies the proportion of the dependent variable (such as demand) that can be accounted for by the independent

ones. Additionally, the C_p statistic (Mallows C) plays a role in determining the model's bias and stability. The statistical significance of the coefficients β_j for the regressors is assessed using appropriate statistical hypothesis techniques like t-tests and others. Finally, it's crucial to evaluate adherence to the fundamental assumptions of the Ordinary Least Squares (OLS) method, including aspects like absence of fit and autocorrelation, which will undergo a thorough examination in subsequent stages of this study.

The ultimate phase of constructing a regression involves assessing the performance of the final model. Essentially, validating the model gauges its capacity to forecast future estimates of the dependent variable, irrespective of the datasets and the timeframe within which the model is being utilized. The simple and linear regression models are applied in sample data sets in order to predict features of population parameters. The following equation depicts, in a mathematical way, the regression model utilized in those sample sets:

$$\hat{Y} = \hat{\beta}_0 + \sum_{j=1}^N \hat{\beta}_j X_j \quad (3.1.1.1)$$

where, $\hat{\beta}_0$ and $\hat{\beta}_j$ are the estimates of the population's parameters.

In linear regression models, the OLS method dominates as the primary and widely adopted approach for estimating population parameters. This method is a cornerstone in statistical analysis, especially when it comes to fitting a linear model to data by minimizing the sum of squared differences between observed and predicted values.

3.1.2 Ordinary least squares

According to Burton (2021), the OLS method stands as the most widely acknowledged and accepted mathematical technique for projecting future estimates of population parameters, utilizing either a simple or multiple linear regression model. The linear regression process employing OLS constructs a line that best represents the data points, aiming for the most accurate depiction of their distribution using a single line. The fundamental principle of least squares asserts that the line generated through the OLS method minimizes the summed squared differences of each data point from the line, resulting in the smallest possible value for these deviations.

The objective of OLS regression is to minimize the sum of squared residuals (SSR) in the sample regression model. SSR, representing "Sum of Squared Residuals" or "Sum of Squared

Errors" denotes the total of squared variances between observed values and those predicted by the regression model in regression analysis. It quantifies the overall discrepancy between observed data and the model's predictions by summing the squared differences between them. In mathematical terms, SSR is computed by squaring the variances between each data point's observed dependent variable values and the values anticipated by the regression equation. After squaring these differences for every data point, they are summed up. SSR serves as a metric that quantifies the total disparity between the observed data and the predictions made by the model. The following equation depicts the SSR calculation:

$$SSR = \sum_{t=1}^T (Y_t - \hat{Y}_t)^2 = \sum_{t=1}^T \hat{e}_t^2 \quad (3.1.2.1)$$

The OLS regression relies on various assumptions that should be verified before continuing with analysis. The most critical of these assumptions are succinctly outlined below:

- Linearity in variable association. This assumption asserts that the independent variables must exhibit a linear relationship with the dependent one, in order the model to be accurately defined.
- Independence or non-autocorrelation assumption. This assumption implies that in a regression model, the residuals should exhibit no correlation with each other, signifying independence between consecutive error terms.
- Normality. The normality assumption asserts that the errors or residuals in a multiple regression model should adhere to a normal distribution around the regression plane.
- Constant error variance is the assumption that indicates that the variability of the dependent variable around the regression plane, known as error variance, remains consistent/constant across all levels of the independent variables.
- Absence of perfect collinearity. When perfect collinearity is present in a regression model, it means that one or more independent variables can be precisely predicted as a linear combination of the other variables. This situation poses problems in regression analysis because it becomes impossible to estimate unique coefficients for each independent variable.

In this thesis, we intend to utilize multiple linear regression extensively across our empirical study. Our aim is to delve into various aspects: analyzing statistical attributes of time series data, exploring interdependencies among product categories, examining the impact of Covid-19 on sales trends, and ultimately establishing a predictive model for future sales.

3.2 Exponentially Weighted Moving Average

Apart from employing regression analysis for the identification of sales time series characteristics and the anticipation of future demand, another extensively utilized method in time series forecasting is the Exponential Weighted Moving Average (EWMA) model. The EWMA model allocates varying weights to successive observations gathered chronologically. This model accords greater significance to recent observations compared to earlier ones, thereby enabling a more accurate representation of temporal patterns within forecasting models. According to Yu et al., (2020), the EWMA model possesses an advantageous recursive structure, enhancing computational efficiency. Additionally, its application circumvents the necessity for presumptions about the distribution of data or possessing prior knowledge concerning the time series.

The EWMA model adeptly captures short-term dynamics and temporal volatility within a time series by integrating prior observations and forecasts to derive subsequent period predictions. However, its applicability diminishes when multiple factors influence the time series. This model assigns weights to historical observations through the smoothing parameter, λ , a variable ranging between 0 and 1. Higher λ values confer greater significance to past observations in forecasting, while lower λ values emphasize the impact of recent observations on estimations (Thomaidis, 2021b). The following equation mathematically expresses the aforementioned introduction and represents the formula for the EWMA model:

$$\hat{Y}_{t+1} \equiv \hat{Y}_{t+1}(\lambda) = (1 - \lambda)Y_t + \lambda\hat{Y}_t \quad (3.2.1)$$

where, \hat{Y}_{t+1} is the forecasted value in the next period $t + 1$, Y_t denotes the actual observed value in the current period t and λ is the smoothing parameter, sometimes referred as the smoothing factor or weight.

The λ parameter takes values from zero to unity ($0 < \lambda < 1$). The meticulous selection of the λ parameter holds paramount significance as it directly influences the EWMA model's incorporation of pertinent information and alignment with the forecasting objectives set forth by the analyst.

3.3 Forecast accuracy measures

Historically, diverse criteria have been employed to evaluate the efficacy of forecasting methodologies. These criteria encompass assessments of forecasting errors, computational efficiency, and the method's capacity to elucidate and interpret forecast outcomes. According to Shcherbakov et al., (2013) the paramount significance of forecast error measures or accuracy in addressing practical issues has been stressed emphasized. Commonly, when dealing with multiple objects, widely accepted forecast error metrics are employed to gauge the effectiveness of forecasting techniques and to select the optimal forecasting approach. Despite inherent limitations, a standard set of usual error measurements prevails across various domains for this purpose.

The forecast error is defined as the difference between the observed value for period t and the estimated value for that corresponding time period. The following equation presents with the mathematically formula that illustrates that discrepancy:

$$\hat{e}_t = Y_t - \hat{Y}_t \quad (3.3.1)$$

where, \hat{e}_t is the forecast error Y_t is the actual observed value for the time period t and \hat{Y}_t is the estimated value for the same time period.

The pivotal performance indicators essential for assessing forecasting errors are consolidated and presented below:

- Root mean squared error (RMSE)
- Mean squared error (MSE)
- Mean absolute error or deviation (MAE or MAD) and
- Mean absolute percentage error (MAPE)

To assess the accuracy of the prognostication methods utilized by analysts for projecting future sales trends, the MAE and MAPE metrics hold significant prominence. Widely adopted, these metrics offer reliable evaluations, aiding in determining the suitability of forecasting techniques based on their performance measurements. In this research, the use of the ultimate pair of forecasting error metrics, namely the MAE and MAPE, is the point of interest. The aim of using these metrics is to assess the efficiency of the employed forecasting methods, allowing for meaningful comparisons between them. Chopra and Meindl (2013) in their book, stressed the importance of accurate evaluation of the forecasting methods.

According to the authors that ability comprises prominent position in companies' vested interest, focusing on predicting the future with optimal accuracy and efficiency. This pursuit contributes to manifold advantages, including improved supply chain performance and planning processes, as well as fostering precision in decision-making procedures.

3.3.1 Mean Absolute Error

The MAE is a metric used to evaluate the accuracy of forecasts or predictions. It measures the average magnitude of errors between predicted and actual values, disregarding the direction of the errors. MAE is calculated by taking the average of the absolute differences between predicted and observed values across all data points. It provides a straightforward understanding of the average magnitude of mistakes in predictions, where lower MAE values indicate better accuracy. This metric is widely used in various fields, including economics, finance, and machine learning, to assess the performance of forecasting models or predictive algorithms. Mathematically, it is represented by the following formula:

$$MAE = \left(\frac{1}{T}\right) \sum_{t=1}^T |Y_t - \hat{Y}_t| = \left(\frac{1}{T}\right) \sum_{t=1}^T |\hat{e}_t| \quad (3.3.1.1)$$

where, $|\hat{e}_t|$ is the absolute value of the forecast error in period t and T is the sample size.

3.3.2 Mean Absolute Percentage Error

The MAPE stands as a widely utilized metric for assessing forecasting model accuracy, similar to the previously discussed MAE metric. It evaluates the average absolute variance between predicted and observed values, expressed in percentage form. To compute MAPE, the absolute difference between predicted and actual values for identical time periods is determined. Subsequently, these differences are divided by the actual values and averaged across all data points. MAPE's methodology ignores error direction, concentrating exclusively on the scale of discrepancies between predicted and observed values.

As highlighted in the discussion about the MAE index, the utilization of the MAPE metric emphasizes the consideration of solely absolute error values, with lower MAPE values signifying enhanced accuracy. Yet, it is crucial the limitations of MAPE to be acknowledged, particularly when handling data points near zero. In these instances, percentage errors might disproportionately inflate, impacting the overall accuracy assessment. Nevertheless, despite these constraints, MAPE finds extensive application across diverse industries such as

finance, supply chain management, and economics. It serves to evaluate forecasting model performance and facilitates comparisons between the accuracies of distinct forecasting methodologies.

Mathematically, MAPE index is represented by the following formula:

$$MAPE = \left(\frac{1}{T}\right) \sum_{t=1}^T \frac{|Y_t - \hat{Y}_t|}{Y_t} = \left(\frac{1}{T}\right) \sum_{t=1}^T \frac{|\hat{e}_t|}{Y_t} \times 100\% \quad (3.3.2.1)$$

where, $|\hat{e}_t|$ is the absolute value of the forecast error in period t , Y_t is the observed actual value for the same period and T is the sample size.

4. Empirical Study

In this chapter, the sales time series for the seven products analyzed in the “Haleon Greece” case study will be presented. The products’ statistical characteristics will be illustrated and the additional insights that derived from the outcomes of both the applied regression analysis and EWMA models will be delved. Additionally, the effectiveness and accuracy of all the forecasting methods used will be assessed, by applying the metrics introduced in the preceding chapter. Furthermore, the potential interconnections among the analyzed time series data will be investigated and the potential impact of the Covid-19 period will be explored. This chapter aims to the thorough examination of the sales time series and to provide responses to the research questions posed in the thesis.

4.1 Data time series selection

In this specific section, an exhaustive exposition of the foundational data underpinning this dissertation will be provided. The dataset for the seven products ($P_k, k = 1, 2, \dots, 7$), was obtained from the finance department of “Haleon Greece” through the company’s enterprise resource planning (ERP) software, SAP S/4HANA Cloud. This dataset encompasses the monthly sales volume over a six-year period, spanning from January 2017 to December 2022. Each product's dataset comprises 72 observations, meticulously compiled without any missing or flawed records that might potentially disrupt the analytical process.

“Haleon Plc” manufactures its goods in Europe, with a significant portion of “Haleon Greece's” inventory primarily sourced from production facilities in England, Ireland, Slovakia, and Switzerland. The company's customer base within the Greek market comprises

pharmaceutical warehouses and retail pharmacies spread across the entirety of Greece. The absence of zero or missing values within the dataset is noteworthy, as these could significantly impact the integrity of the statistical analysis and forecast models. This absence helps mitigate potential limitations and biases that could arise from erroneous or incomplete data, enhancing the accuracy and reliability of the applied forecasting models.

Despite the company's lack of concerns regarding data confidentiality and encryption for the study and analysis, code names will be allocated to the products. This measure aims to facilitate their presentation to readers, ensuring clarity and ease of comprehension while maintaining anonymity and confidentiality.

4.2 Data time series presentation

In **Table 1**, products, their corresponding brand and categories are presented alongside the respective code names, as previously mentioned.

PRODUCT	BRAND	CATEGORY 1	CATEGORY 2	CODE NAME
OTRIVIN PRESERVATIVE FREE SPRAY	Otrivin	Respiratory	Nasal Decongestion	P₁
PANADOL COLD & FLU	Panadol C&F	Respiratory	Cold & Flu Multi-Symptom	P₂
PANADOL EXTRA EFFERVESCENT TABS	Panadol	Pain Relief	Analgesics/Systemic Pain	P₃
VOLTAROL FORTE	Voltarol	Pain Relief	Analgesics/Topical Pain	P₄
FENISTIL GEL 30	Fenistil	Skin irritation	Skin Health	P₅
COREGA SUPER 70	Corega	Oral health	Oral Care	P₆
SENSODYNE REPAIR & PROTECT	Sensodyne	Oral health	Oral Care	P₇

Table 1 “Products, Brands, Categories and respective code names”

4.3 Descriptive Statistics

Utilizing Excel's Descriptive Statistics Tool, our aim is to analyze the time series of product sales, extracting significant insights, and briefly delineate the primary characteristics of the sales data. As Fisher and Marshall (2009) stressed in their research, descriptive statistics refer to the numerical methods and graphical approaches employed to arrange and elucidate the attributes or variables within a specific sample. The authors also noticed that the primary objective of descriptive statistics is to delineate the central point of a distribution of values,

often termed as the mean or the measure of central tendency, as well as to illustrate the extent of variability among the rest values in the dataset. This degree of variability within a dataset is frequently referred to as dispersion or variance.

Tables 2 and 3 encapsulate the primary attributes pertaining to product sales for each of them within the examined period spanning from 2017 to 2022. These tables serve to outline and detail essential attributes associated with the sales data during this specified timeframe. The tables include the sales central point, the dispersion of sales data denoted by their standard deviation, as well as insights into the shape of the dataset distribution through the assessment of skewness and kurtosis values.

STATISTICS	P ₁	P ₂	P ₃	P ₄
MEAN	151045	97917	180729	79976
STANDARD ERROR	8779	9029	9119	3566
MEDIAN	142976	85518	169458	77007
MODE	397576	#N/A	426786	99200
STANDARD DEVIATION	74490	76611	77377	30262
SAMPLE VARIANCE	5548715278	5869290456	5987126935	915792971
KURTOSIS	1.89	0.78	0.90	0.04
SKEWNESS	0.88	0.97	0.71	0.19
RANGE	395590	321273	405152	139485
MINIMUM	1986	2473	21634	13116
MAXIMUM	397576	323746	426786	152601
SUM	10875263	7049992	13012456	5758263
COUNT	72	72	72	72

Table 2 “Descriptive Statistics of products P₁, P₂, P₃ and P₄”

Upon analyzing the characteristics of each product, a notable trend emerges. The product **P₃** appears to hold a dominant position within the company's portfolio compared to the others. However, this prominence is accompanied by a significant level of variability, evident from its notably high standard deviation (SD) value in sales data.

STATISTICS	P ₅	P ₆	P ₇
MEAN	44796	15896	45786
STANDARD ERROR	6920	455	1225
MEDIAN	22676	15943	45451
MODE	#N/A	18186	70307

STANDARD DEVIATION	58716	3864	10396
SAMPLE VARIANCE	3447549356	14930255	108067681
KURTOSIS	3.26	0.08	-0.35
SKEWNESS	1.93	0.36	0.10
RANGE	245684	17646	47438
MINIMUM	1373	8586	22869
MAXIMUM	247057	26232	70307
SUM	3225285	1144483	3296627
COUNT	72	72	72

Table 3 “Descriptive Statistics of products P_5 , P_6 and P_7 ”

Similar patterns manifest in the sales data of products P_1 and P_2 , characterized by substantial volatility across their respective time series, as indicated by their high SD values. Furthermore, all three aforementioned products exhibit skewness values near unity, implying a right-skewed distribution in their datasets. Additionally, the positive kurtosis values for these products suggest distributions with heavier tails and higher peaks compared to a normal distribution. These observations hint at the potential presence of extreme values, often referred to as outliers, within the sales time series of these products.

Extending the analysis to encompass the remaining products, product P_4 demonstrates a comparatively lower standard deviation, suggesting a lack of pronounced volatility within its sales data. Moreover, the skewness and kurtosis values, proximate to zero and marginally positive, indicate an almost symmetrical distribution with a slight rightward skewness. Notably, the kurtosis value approaching zero signifies that the tails in the distribution are similar to those of a normal distribution. Additionally, with the mean nearly equivalent to the median, the sales data exhibit a tendency towards normal distribution.

Upon studying products P_6 and P_7 it becomes evident that product P_6 exhibits comparatively lower consumption among the examined set of products, while both of them demonstrate a lack of notable volatility, as evidenced by their SD values. This observation aligns plausibly with the nature of oral category products, which typically display inelastic demand due to their classification as essential commodities for consumers. Furthermore, it's notable that both products showcase a near equivalence between their mean and median values, while their skewness values approach zero, indicating an almost symmetrical distribution in sales observations with a slight rightward inclination. Additionally, the kurtosis values for P_6 and P_7 are nearly zero and slightly positive and negative, respectively.

This leads to the inference that the tails in their distribution resemble those of a normal distribution, further reinforcing the notion of a distribution pattern akin to normality in their sales data.

The investigation into product **P₅** culminates the descriptive statistics analysis. Specifically, this product demonstrates a notably elevated standard deviation in contrast to its dataset mean. This discrepancy suggests a considerable dispersion and variability among the sales data points and a substantial deviation from the central tendency or average. Such increased volatility typically implies a broader dispersion of data points around their mean, thereby signaling heightened unpredictability or risk within the measured variables and the potential presence of outliers in the dataset. Lastly, the markedly high positive values observed in both kurtosis and skewness for this product indicate a distribution pattern characterized by leptokurtosis and right-skewness.

Tables 4 and 5 provides a concise summary of two pivotal statistical attributes for each product across the years: the mean and the sample variance. This compilation aims to facilitate the forthcoming analysis by offering essential information about these key statistical measures for each product within the dataset.

	P₁		P₂		P₃		P₄	
YEARS	MEAN	VARIANCE	MEAN	VARIANCE	MEAN	VARIANCE	MEAN	VARIANCE
2017	168458	2376827984	156152	9633609691	183054	2771516336	76152	240990371
2018	155741	11098397880	111904	7138266958	177070	5640216240	87698	1390841006
2019	132825	1840957422	109821	3990235423	180857	10477979649	72115	900558392
2020	150601	2584812736	92622	4819875836	201596	10121208684	87280	1238038589
2021	130343	5781624558	44718	3078415997	156995	4641684484	73940	1129699906
2022	168304	10622161833	72283	1320656086	184800	3863446140	82671	756637642

Table 4 “Mean and sample variance of P₁, P₂, P₃ and P₄ products during years”

	P₅		P₆		P₇	
YEARS	MEAN	VARIANCE	MEAN	VARIANCE	MEAN	VARIANCE
2017	48428	5102468666	12375	4207642	40766	132057627
2018	51727	3512143359	15054	2381598	42775	73248144
2019	47576	4700492935	17201	5167214	47820	120539623
2020	42836	4635723609	18432	26432204	48056	163437133
2021	41113	2179343142	17390	7723553	47642	86403857
2022	37094	1963267846	14922	23818657	47661	66734622

Table 5 “Mean and sample variance of P₅, P₆ and P₇ products during years”

4.3.1 Graphical summary per product

In time series analysis, a primary objective is to examine and comprehend the nature of variations present within the datasets under examination. This analysis yields valuable information allowing researchers to thoroughly understand how variables evolve across the specified period, aiding in the evaluation of their forecasting capabilities. According to R Rosca (2011), the graphical representations of these variables afford analysts the opportunity to discern essential issues concerning the presence of systematic or non-systematic fluctuations and patterns within the time series, providing valuable insights into their temporal behaviors over the duration studied.

Within time series analysis, graphical representation serves as a robust and valuable tool. It facilitates the identification of three fundamental recurring components throughout the entirety of the time series. These components encompass the trend, seasonality, and the irregular or random component, often referred to as the noise within the time series data. Through visual data representation, the distinct patterns of these components become discernible, aiding in comprehensive analysis and understanding of the temporal data structure. In subsequent paragraphs, an analytical approach will be undertaken for each product employing time series graphs, boxplots, and histograms for comprehensive examination and interpretation. These visual representations will provide insights into the characteristics and distributions of the respective datasets for detailed analysis.

Product P_1

Figures 10, 11 and 12 showcase the time series plot, histogram, and boxplot chart, offering a visual representation of the data distribution, temporal patterns, and the variability present within the examined datasets. These graphical illustrations serve as instrumental tools in comprehending the dynamics, distributions, and characteristics of the data under analysis.

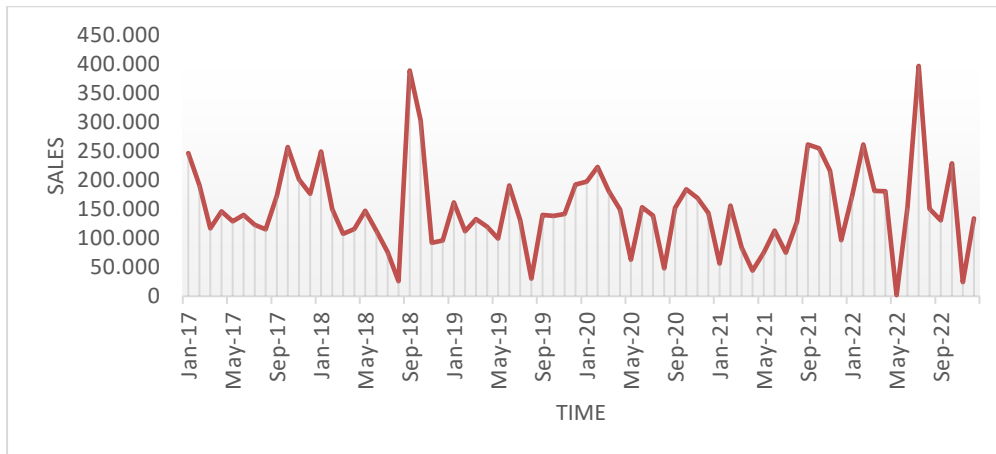


Figure 10, “Time series plot of sales P_1 2017-2022”

The time plot depicting the sales of product P_1 over the total examined period does not display any noticeable pattern or discernible upward or downward trend. Additionally, the sales time series for the product do not demonstrate any significant or apparent fluctuations that could potentially complicate the analysis process. These observations correspond with the data presented in **Table 4**, confirming the consistent mean behavior of the product's demand throughout the six-year period.

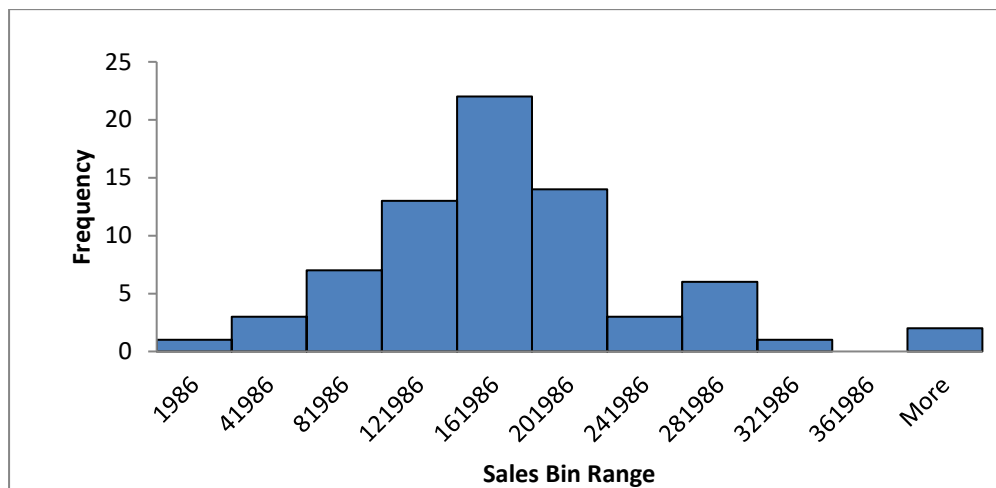


Figure 11, “Histogram of P_1 sales”

The Histogram reveals a distribution slightly skewed to the right, in line with the positive skewness observed (skewness=0.88). Additionally, there appear to be observations that deviate from the clustering of the main data, suggesting the possibility of outliers within the sales data.

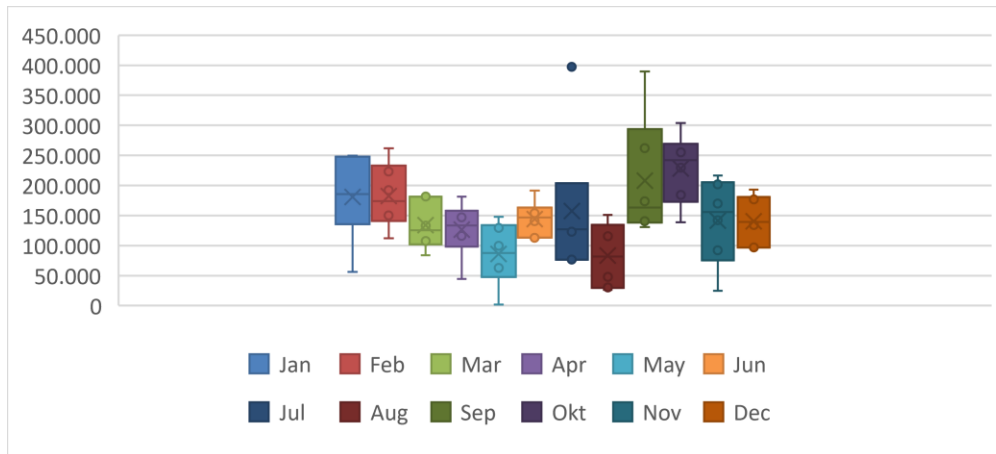


Figure 12, “Monthly Boxplot of product P_1 ”

Interpreting the Boxplot chart reveals some key observations. There is a discernible uptick in product sales during Autumn and Winter months. Additionally, it's noteworthy that an outlier appears exclusively in July among the observed data points.

Product P_2

The visual representation proceeds with product P_2 in **Figures 13-15**. These figures comprise the time series plot, histogram, and boxplot chart respectively. These graphical depictions aim to assist reader in comprehending the characteristics of the product and identifying potential issues within the datasets.

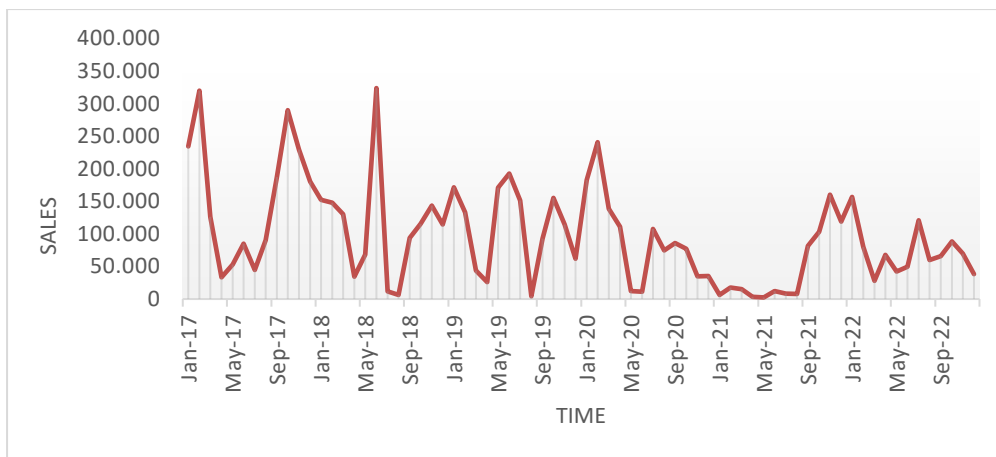


Figure 13, “Time series plot of sales P_2 2017-2022”

The above time series plot illustrating the sales of product P_2 for the total examined period, indicates a descending trend in sales over the years. The time series seem not to reverse

around a specific mean and this observation coincides with the data presented in **Table 4**, in which the continuously decrease of product’s mean over time, is apparent.

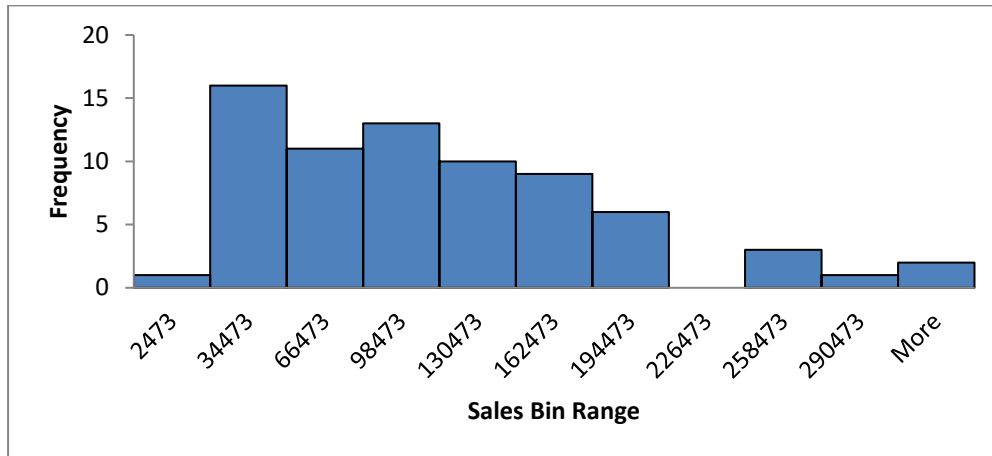


Figure 14, “Histogram of P₂ sales”

The Histogram illustrates a distinctly right-skewed data distribution, also confirmed from the positive skewness nearly approaching one. Additionally, there are observations noticeably deviating from the main cluster, suggesting the possible presence of outliers within the dataset.

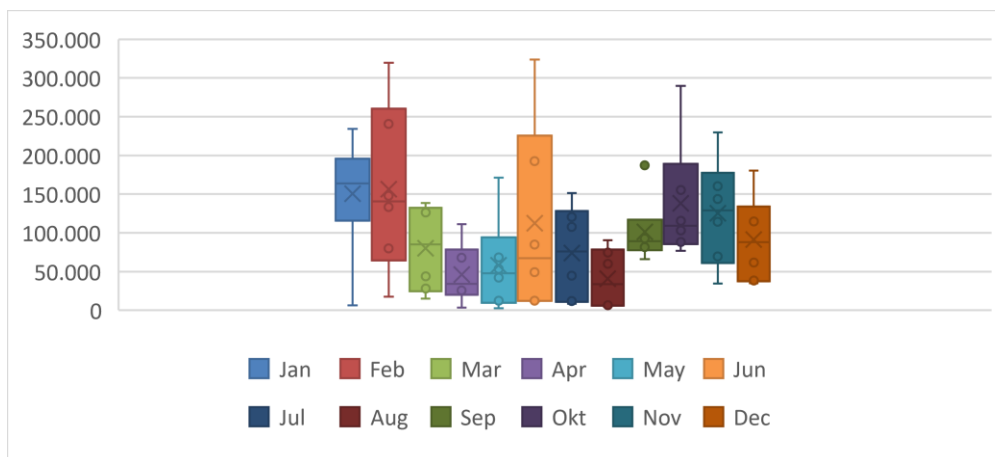


Figure 15, “Monthly Boxplot of product P₂”

The presentation of the Boxplot chart reveals an uptick in product consumption at the onset of each season. Additionally, there’s a distinct outlier exclusively present in September among the observed data points.

Product P_3

Figures 16, 17, and 18 continue the visualization of sales data for product P_3 , maintaining the same approach and explanatory methodology applied to previous products.

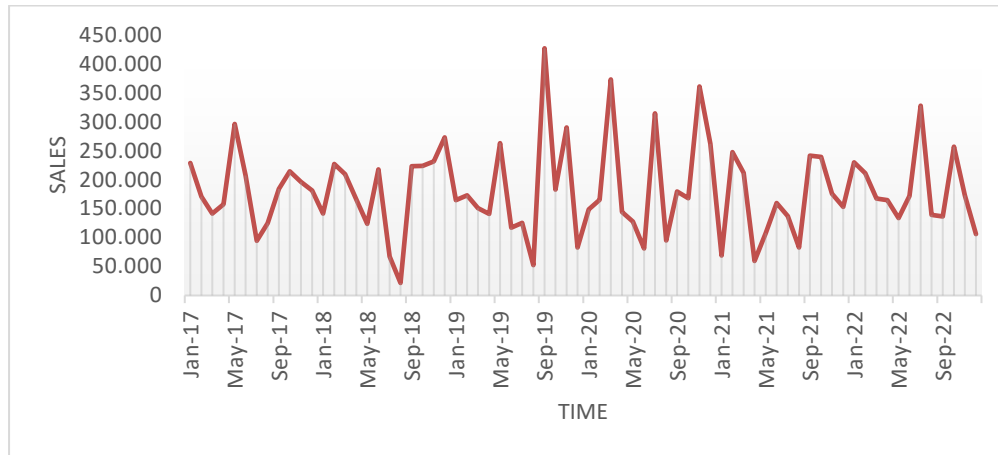


Figure 16, “Time series plot of sales P_3 2017-2022”

The preceding time series plot illustrates the sales pattern of product P_3 , showcasing a lack of trend features within the product's sales observations. There is no discernible upward or downward trend evident from the chart. Moreover, the sales time series for the product do not exhibit any notable or significant fluctuations that could potentially complicate the analysis process. These observations align with the data provided in **Table 4**, reaffirming the consistent mean behavior of the product's demand over the six-year period.

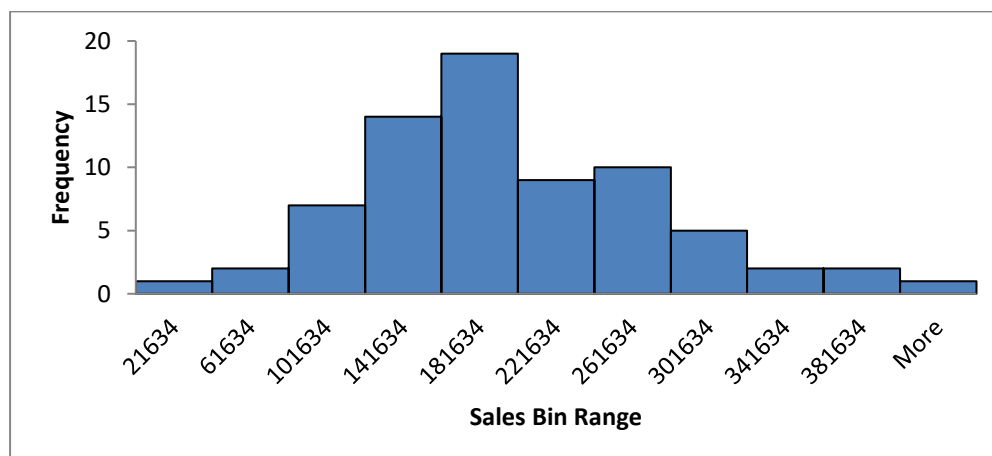


Figure 17, “Histogram of P_3 sales”

The empirical distribution of P_3 sales is clearly right-skewed, supported by the positive skewness coefficient approaching unity. Moreover, observations appear clustered, indicating

the potential absence of outliers within the dataset. The subsequent Boxplot aligns with this assumption. Furthermore, there does not seem to be a clear seasonal trend or noticeable increase in product consumption across specific months.

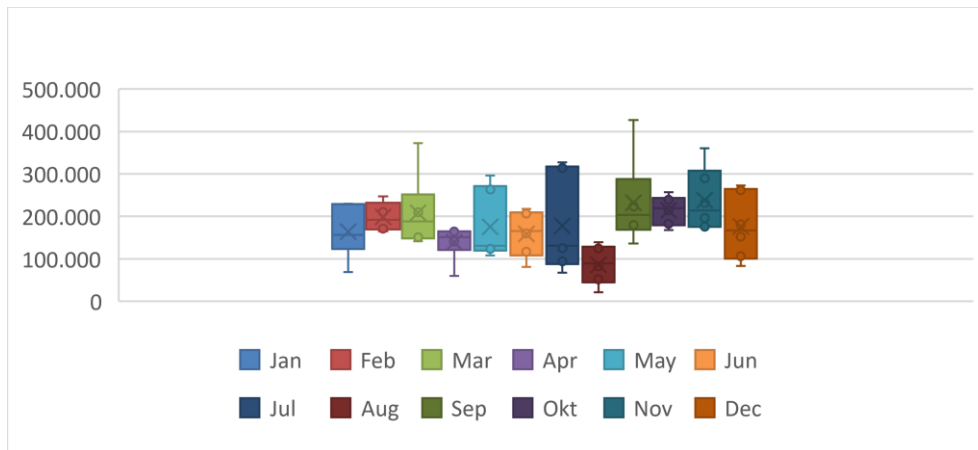


Figure 18, “Monthly Boxplot of product P_3 ”

Product P_4

The products analysis carries on with product P_4 . **Figures 19, 20 and 21** exhibit the respective time plot, histogram and boxplot, continuing the analogous presentation method employed for previous products.

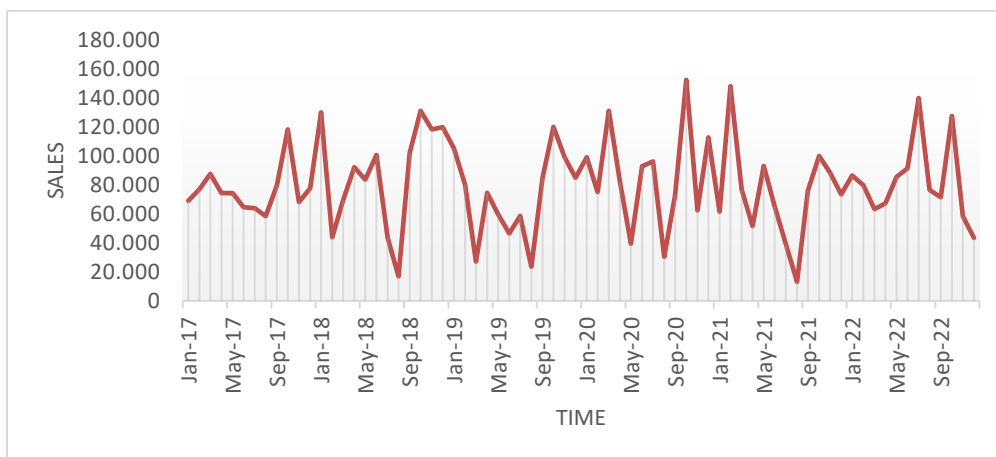


Figure 19, “Time series plot of sales P_4 2017-2022”

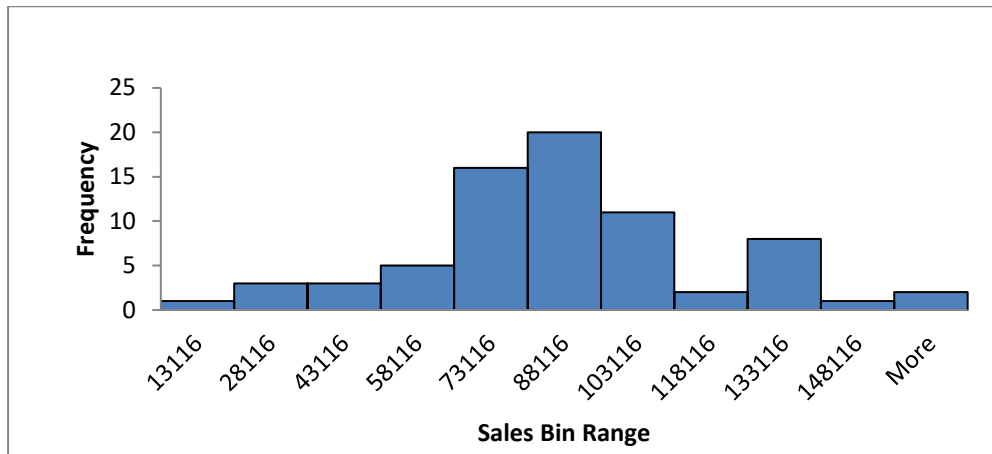


Figure 20, “Histogram of P₄ sales”

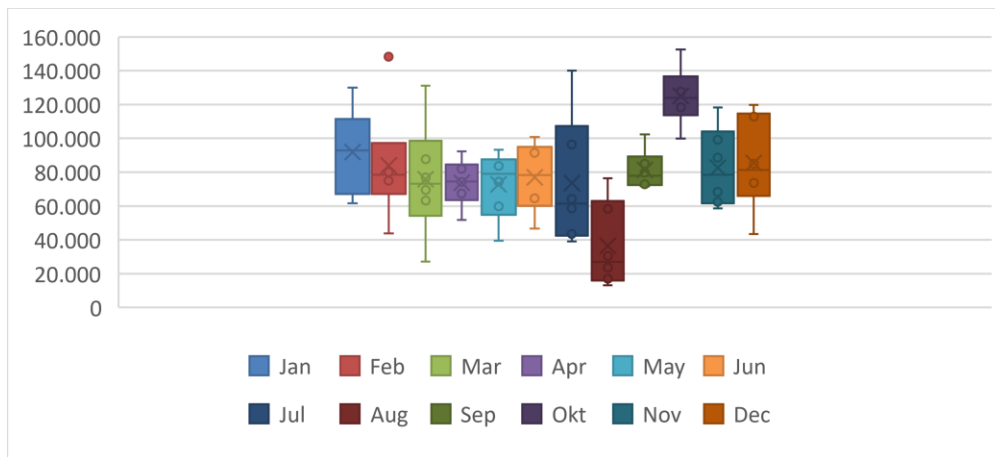


Figure 21, “Monthly Boxplot of product P₄”

The above displayed graphs provide a visualization of product P₄'s sales patterns. The time plot chart illustrates a consistent behavior in the sales of product P₄ throughout the examined time period, despite the observing fluctuations. This observation aligns with the relatively stable mean and variance values presented in **Table 4** across those years, reinforcing the notion of a consistent, non-changing pattern within the dataset. The histogram graph in **Figure 20** reveals an almost symmetrical distribution within our dataset, highlighted by the near equal values of the product's mean and median, with all the datas clustered. The skewness, nearly zero but slightly positive, signifies an almost symmetrical distribution in the sales series, albeit with a slight rightward skew. Finally, the Boxplot graph indicates the presence of a single outlier occurring in February, alongside a noticeable decrease in product sales during August. Conversely, October portrays notably higher sales compared to the other months, suggesting a peak in sales activity during that period.

Product P_5

Continuing the visual analysis process initiated for the previous products, the following **Figures 22, 23** and **24** present the time series plot, the histogram and the boxplot of product P_5 respectively.

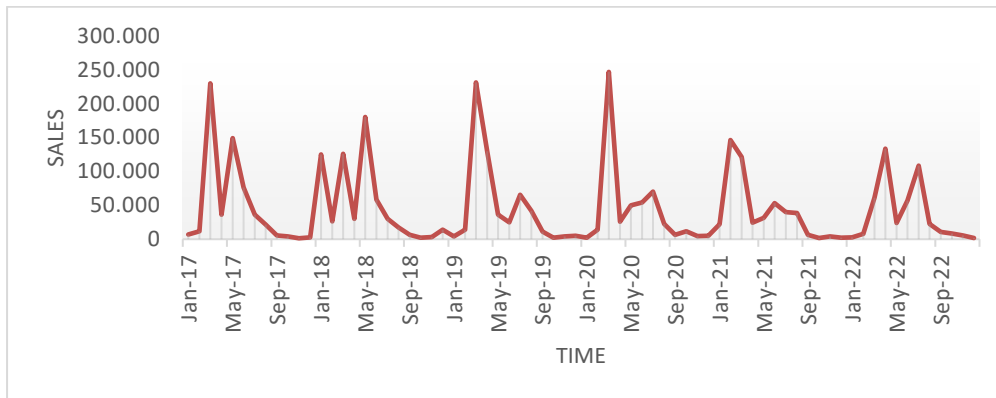


Figure 22, “Time series plot of sales P_5 2017-2022”

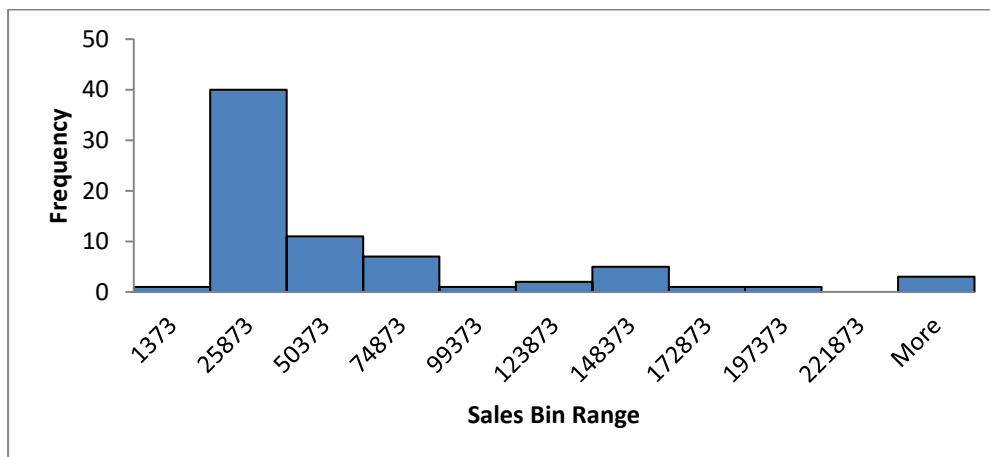


Figure 23, “Histogram of P_5 sales”

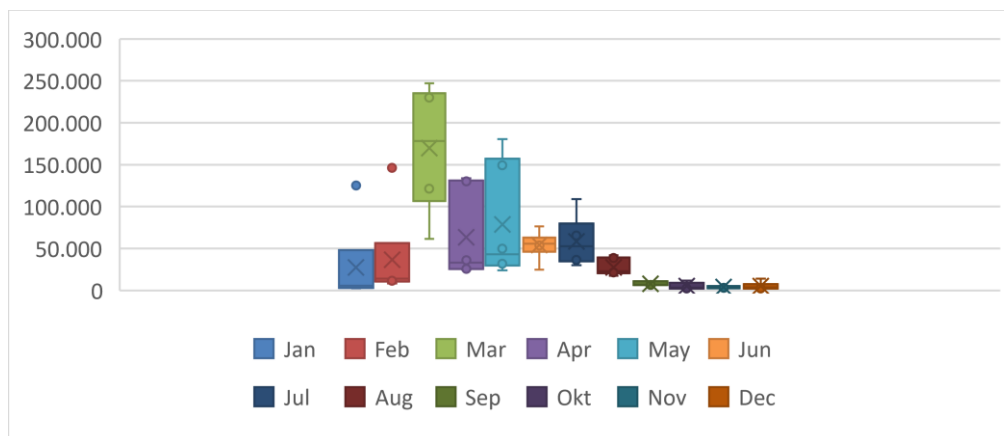


Figure 24, “Monthly Boxplot of product P_5 ”

Analyzing the graphs, it is noticed that the seasonality effect is present and obvious in sales of product **P₅** due to the series time plot. According to the product's time plot chart, it is not observed any ascending or descending trend to the datas and the time series appear to reverse around their mean, as confirmed by the product's almost constant mean and varriance depicted in **Table 5**. Seasonality involves repeating patterns or fluctuations that occur at regular intervals within a time series, such as daily, weekly, monthly, or yearly patterns. The seasonality impact seems to exist, as product's demand increases during Spring and Summer months and afterwards decreases abruptly. The seasonal effect will be thoughroughly explored in the subsequent regression analysis.

The histogram chart indicates that our datas distribution is clearly right-skewed. Tthat observation is supported by a considerably large positive skewness value (nearly equal to two). Additionally, the chart indicates the potential presence of some observations that might be considered as extreme values/ outliers. Similarly, the boxplot reinforces these features observed in the time series and histogram graphs for the product **P₅** sales dataset. Outliers are detected in January and February and the increased product's sales during the Spring and Summer months also noticed in the time series analysis, are obvious in the Boxplot as well.

Product P₆

Following the same process, **Figures 25, 26 and 27** present the time series plot, the histogram and the boxplot of product **P₆** respectively.

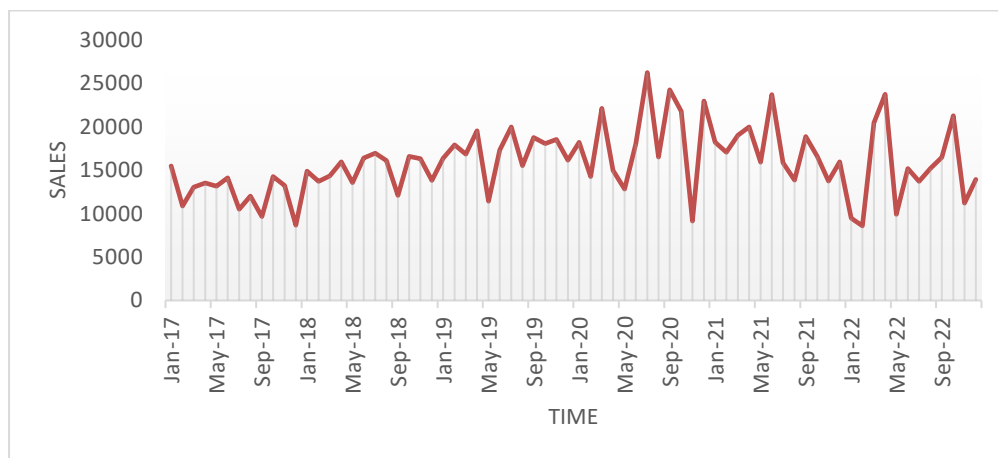


Figure 25, “Time series plot of sales P₆ 2017-2022”

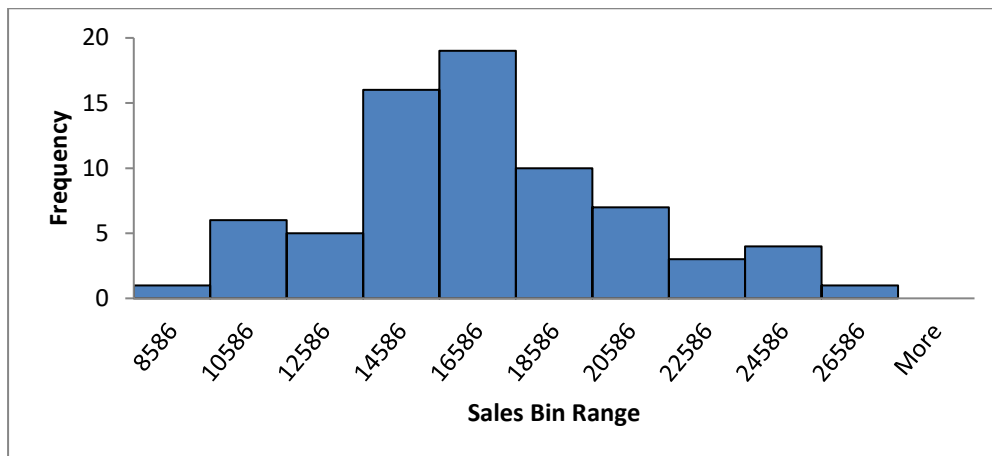


Figure 26, “Histogram of P₆ sales”

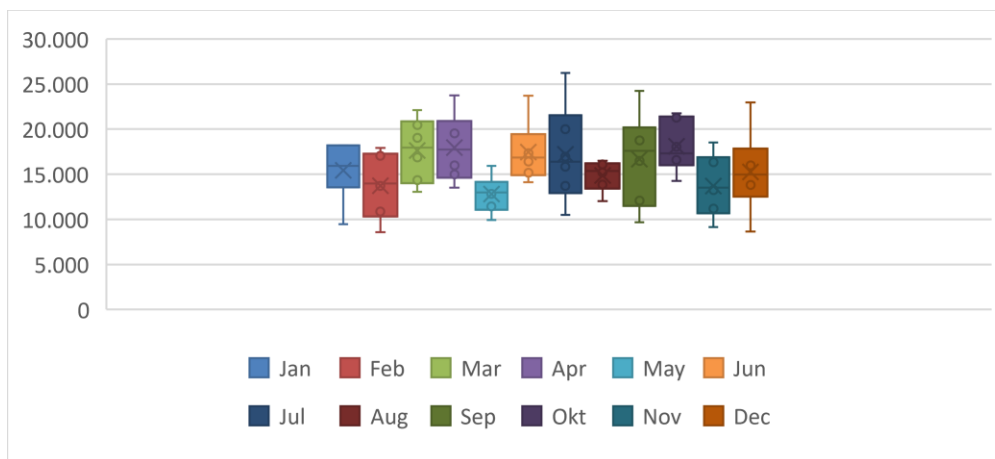


Figure 27, “Monthly Boxplot of product P₆”

The time series chart depicting the product's sales illustrates an upward trend in sales from 2017 to 2019, succeeded by a period showcasing a slight downward trend until the conclusion of the observed period. Furthermore, the presence of a clear seasonality effect is not apparent, which is a typical observation in oral health products. Such products often exhibit stable conditions and inelastic demand owing to their inherent nature.

The Histogram displays an almost symmetrical distribution within our dataset, indicated by the nearly equivalent values of the mean and median. The skewness, close to zero but positive, suggests an almost symmetrical distribution in the sales series, albeit with a slight rightward skew.

The Boxplot aligns with the earlier observations, indicating the absence of a clear and pronounced seasonal trend within the sales dataset. Furthermore, it suggests no significant

increase in product consumption across specific months, reinforcing the notion of stable sales patterns without distinct fluctuations.

Product P₇

The study will adopt a similar approach to explore the attributes of sales for product **P₇**, belonging to the same category as the previously studied **P₆** product. The aim is to evaluate the sales characteristics, particularly focusing on the stability observed in the mean and variance across all recorded years. Subsequently, the analysis will encompass the examination of attributes through visual representations such as the time series plot, histogram, and boxplot to detect any potential outliers within the dataset. **Figures 28, 29** and **30** present the respective charts.

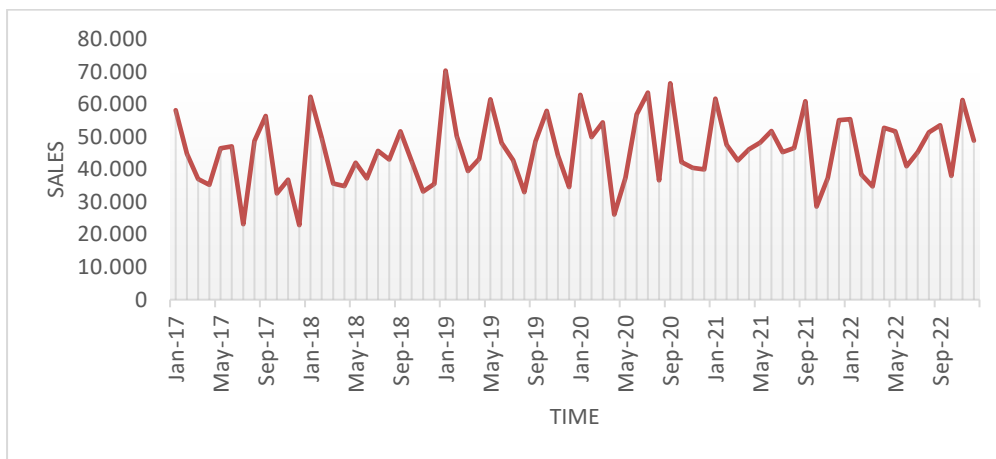


Figure 28, “Time series plot of sales P₇ 2017-2022”

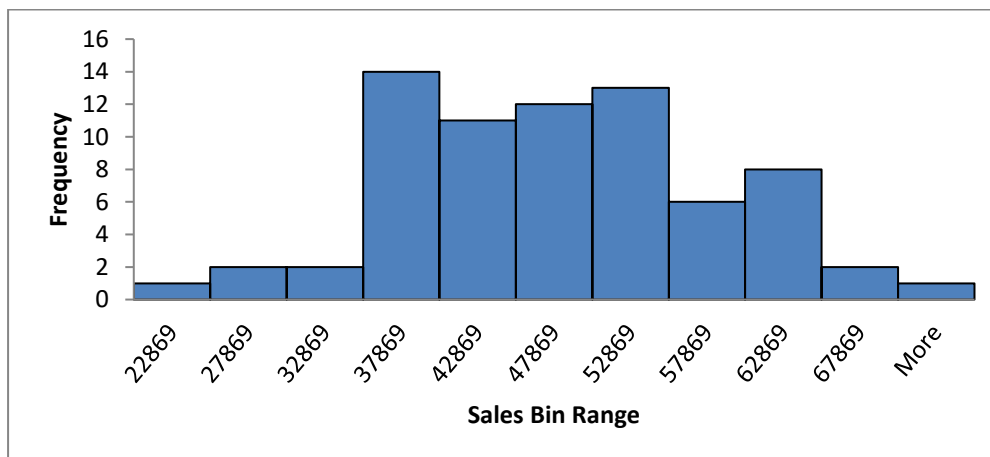


Figure 29, “Histogram of P₇ sales”

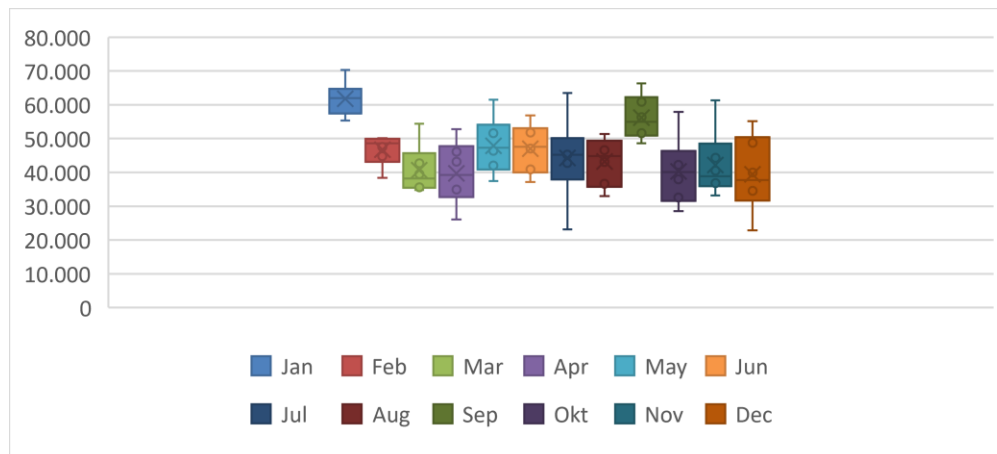


Figure 30, “Monthly Boxplot of product P₇”

The time series plot in **Figure 28** depicts a notably consistent sales evolution over the years for product **P₇**, aligning with the information presented in **Table 5** concerning the product's stable mean and variance values. There is no apparent discernible upward or downward trend observed in sales and the sales data seem to display fluctuations around their mean throughout the examined period. Moreover, there is no apparent seasonality effect, because probably of the stable and inelastic demand observed in oral health product categories.

The Histogram representation, as displayed in **Figure 29**, portrays a distribution within our dataset that is nearly symmetrical, supported by the close resemblance between the mean and median values. With skewness approaching zero, the sales series exhibit an almost impeccable symmetry in their distribution. Additionally, the clustering of all observations intimates the lack of potential outliers within the sales dataset. The monthly boxplot graph illustrated in **Figure 30** validates the absence of any extreme values within the dataset and highlights the increased sales exhibit in January and September compared to the remaining months.

4.4 Multiple Linear Regression (MLR) model analysis and construction

To explore the statistical characteristics the examined set of products over time and ascertain the most suitable regression model, a stepwise regression analysis will be conducted. The regression models will incorporate pertinent variables encompassing potential seasonal influences via monthly dummy variables, trend characteristics (linear or quadratic), autoregressive patterns using lag variables to discern short or long-term sales memory, cross dependencies among the examined products' datasets and the potential impact of the Covid-

19 pandemic employing an appropriate dummy variable. Through this procedure, the study aims to address the research questions outlined in Chapter 1.1.

The regression model that will be applied in order to check the aforementioned features of the sales dataset of each product, can be described by the following equation:

$$M1: S_t = \beta_0 + \beta_t t + \gamma t^2 + \sum_{i=1}^{11} \beta_i M_{t,i} + \sum_{j=1}^2 \delta_j S_{t-j} + e_t \quad (4.4.1)$$

where, β_0 is the intercept term of the time series corresponding to effect of the baseline month January, β_t and γ are the slope coefficients of the linear and quadratic trend respectively, β_i is the coefficient for each monthly dummy variables $M_{t,i}$ where $i = 2, 3, \dots, 12$, δ_j is the coefficient for each lag variables S_{t-j} where $j = 1, 2$ and e_t is the disturbance or error term.

Monthly dummies, a common fixture in statistical analysis, serve to account for seasonal or monthly fluctuations within data. Within this context, “dummies” refer to binary variables representing distinct months of the year. Each dummy variable assumes a value of 1 for observations occurring in that specific month and 0 otherwise. These dummies function to capture and manage any systematic disparities across months. Integration of monthly dummies into a regression model facilitates the estimation of the average impact of each month on the dependent variable while maintaining other factors constant. To circumvent issues related to perfect multicollinearity, which can impair forecasting accuracy and lead to estimation complications, a single month is consistently omitted and utilized as the reference point. The same process will be followed for the entire examined period from 2017 to 2022 for each of the seven products.

Within the MLR analysis, the determination to include or exclude an explanatory variable in the regression equation relies on a hypothesis test aimed at assessing the significance of the respective regression coefficient. A fundamental criterion for this evaluation involves scrutinizing whether the coefficient value β_i equates to zero, denoting the variable's lack of explanatory influence on the dependent variable. Typically, this evaluation is conducted utilizing a two-tailed t -distribution, where the null hypothesis ($H_0: \beta_i = 0$) implies the insignificance of the coefficient, while the alternative hypothesis ($H_1: \beta_i \neq 0$) suggests its significance.

The test statistic's position within its distribution under the null hypothesis is crucial. Extreme positioning suggests potential contradiction with the null hypothesis, while a position closer

Postgraduate Dissertation

to the distribution's center indicates compatibility with it. By calculating the probability value (*p-value*) from the test statistic's distribution, we gauge the likelihood of observing a deviation as extreme as or more extreme than the observed one by chance, assuming the null hypothesis is true. This *p-value*, known as the achieved significance level, helps us determine the significance of the observed deviation. If the *p-value* is less than 0.05, we consider the deviation significant at the 95% confidence level, indicating strong evidence against the null hypothesis (Cox, 1982) (Albright, 2010).

4.5 Sales data transformation into natural logarithmic set

The first thing that is necessary to the study, is to convert the sales time series dataset into natural logarithms applying the Excel's function LN. That choice is many times beneficial and contributes to limit the negative features of a dataset that might significantly affect the results of the analysis. According to Benoit (2011), employing logarithmic transformations on variables within a regression model is a prevalent strategy when dealing with scenarios involving non-linear relationships between independent and dependent variables. Utilizing the logarithm of one or more variables, as opposed to their original forms, introduces a non-linear relationship while retaining the framework of the linear model.

Berenson et al., (2020), in their book supported that logarithmic transformations play a significant role concerning various aspects: they aid in normalizing skewed distributions within a dataset, stabilize variance across the dataset, linearize relationships allowing for the application of linear techniques, identify and manage extreme values, and enhance the interpretability of a model when presented in logarithmic form. Considering these benefits, the sales time series of products were transformed using natural logarithms prior to applying explanatory and forecasting techniques. Thus, the prior model M1 presented in the equation (4.4.1) is transformed in a logarithmic way as follows:

$$M1: \log(S_t) = \beta_0 + \beta_t t + \gamma t^2 + \sum_{i=1}^{11} \beta_i M_{t,i} + \sum_{j=1}^2 \delta_j \log(S_{t-j}) + e_t \quad (4.5.1)$$

where, $\log(S_t)$ is the natural logarithm of the observed sales S_t , β_0 is the intercept term of the time series corresponding to effect of the baseline month January, β_t and γ are the slope coefficients of the linear and quadratic trend respectively, β_i is the coefficient for each monthly dummy variables $M_{t,i}$ where $i = 2, 3, \dots, 12$, δ_j is the coefficient for each logarithmic lag variables S_{t-j} where $j = 1, 2$ and e_t is the disturbance or error term.

4.6 Statistical significance of the explanatory variables in MLR model

By employing the regression model M1 across the sales time series of each product, the objective is to explore the connection between the independent variables and the dependent variable. Utilizing the *p-values* derived from the regression output for each independent estimator will serve as the means to draw pertinent conclusions regarding the statistical significance of the explanatory variables. These *p-values* enable an assessment of how these predictors contribute to explaining the behavior of the variable of interest. As elucidated in the preceding subsection 4.4, the rejection of the Null Hypothesis (H_0) occurs if the *p-value* is less than the critical value α , meaning that the variable's coefficient β_i does not equal zero at a specified confidence level (set at 95% in this study). Conversely, when the *p-value* is greater than or equal to α , the Null Hypothesis cannot be dismissed in favor of the Alternative Hypothesis (H_1).

The identical methodology will be applied to assess the statistical significance of each explanatory variable, including monthly dummies, autoregressive lag variables, cross-dependency variables, and the pandemic dummy variable, that will be employed in the MLR models. Through this systematic approach, the research questions that were set in chapter 1.1 will be investigated and the final answers will be thoroughly explained.

4.6.1 Statistical Properties of product P_1

Beginning with product P_1 , the application of the M1 multilinear regression (MLR) model, as previously outlined, aims to yield valuable insights into the statistical characteristics of the product's sales time series. Additionally, this approach endeavors to assess the statistical significance of the explanatory variables integrated into the regression model. The summary output derived from Excel's regression implementation is presented below in **Tables 6 and 7**.

TERM	COEFFICIENTS	STANDARD ERROR	T STAT	P-VALUE
INTERCEPT	12,2198	0.3492	34.9963	0.0000
LINEAR TREND	-0.0093	0.0154	-0.6009	0.5502
QUADRATIC TREND	0.0001	0.0002	0.2614	0.7947
M2	0.0767	0.3873	0.1980	0.8437

M3	-0.2257	0.3874	-0.5826	0.5625
M4	-0.3253	0.3875	-0.8394	0.4047
M5	-1.1412	0.3877	-2.9439	0.0047
M6	-0.1169	0.3878	-0.3014	0.7642
M7	-0.1864	0.3880	-0.4803	0.6328
M8	-0.8630	0.3883	-2.2227	0.0301
M9	0.2020	0.3886	0.5200	0.6051
M10	0.3502	0.3889	0.9004	0.3716
M11	-0.2946	0.3892	-0.7570	0.4521
M12	-0.1318	0.3896	-0.3382	0.7364

Table 6 “Regression summary with monthly dummies for product P₁”

TERM	COEFFICIENTS	STANDARD		
		ERROR	T STAT	P-VALUE
INTERCEPT	15.4790	2.3781	6.5088	0.0000
LINEAR TREND	-0.0131	0.0174	-0.7523	0.4552
QUADRATIC TREND	0.0001	0.0002	0.4100	0.6834
LAG VARIABLE T-1	-0.0111	0.1320	-0.0843	0.9331
LAG VARIABLE T-2	-0.2584	0.1392	-1.8560	0.0689
M2	0.1040	0.4265	0.2439	0.8082
M3	-0.1807	0.4105	-0.4403	0.6615
M4	-0.2642	0.4092	-0.6457	0.5212
M5	-1.1598	0.4107	-2.8239	0.0066
M6	-0.1708	0.4314	-0.3959	0.6937
M7	-0.4402	0.4390	-1.0029	0.3204
M8	-0.8535	0.4091	-2.0864	0.0417
M9	0.1854	0.4202	0.4414	0.6607
M10	0.1699	0.4294	0.3956	0.6940
M11	-0.1987	0.4154	-0.4783	0.6343
M12	-0.0055	0.4124	-0.0133	0.9894

Table 7 “Regression summary with monthly dummies and lag variables for product P₁”

Based on the summary output, it becomes evident that both the linear and quadratic trend components lack statistical significance, as indicated by their coefficients' *p-values* importantly exceeding the critical value of $\alpha = 0.05$, the designated significance level at 5%. This observation is agreed with the explanation about no-trend behavior that was illustrated in **Figure 10** and the insights detailed in subchapter 4.3.1 and summarized in **Table 4**. Similar insignificance is observed across most of the coefficients associated with the monthly

dummies ($M_{t,i}$), suggesting an absence of seasonality in the dataset, with exceptions noted for the months of May and August.

Furthermore, the analysis reveals a lack of observable memory attributes within the sales time series, highlighted by the lag variables' coefficients possessing *p-values* greater than 0.05. This lack of statistical significance suggests an inability to explain product P_1 's current sales through its previous occurrences.

In summary, the comprehensive assessment indicates that the sales pattern of product P_1 does not exhibit a relationship with prior sales occurrences or a discernible seasonal trend, as most coefficients associated with these factors fail to achieve statistical significance.

4.6.2 Statistical Properties of product P_2

Implementing the same regression model M1 for product P_2 aims to delve into the distinctive features within its time series and scrutinize the statistical significance of the independent variables employed to elucidate the behavior of the dependent variable. This approach facilitates a comparative analysis and offers insights into how the model's performance and variable's significance might differ across distinct products. The summary output derived from Excel's regression implementation is presented below in **Tables 8 and 9**.

TERM	COEFFICIENTS	STANDARD		
		ERROR	T STAT	P-VALUE
INTERCEPT	12.5122	0.5229	23.9266	0.0000
LINEAR TREND	-0.0528	0.0231	-2.2841	0.0260
QUADRATIC TREND	0.0005	0.0003	1.5529	0.1259
M2	0.1278	0.5801	0.2202	0.8265
M3	-0.5031	0.5802	-0.8671	0.3895
M4	-1.1557	0.5803	-1.9915	0.0511
M5	-1.1223	0.5806	-1.9332	0.0581
M6	-0.4661	0.5808	-0.8024	0.4256
M7	-0.6864	0.5811	-1.1812	0.2424
M8	-1.4298	0.5815	-2.4588	0.0169
M9	0.0887	0.5819	0.1524	0.8794
M10	0.3667	0.5824	0.6297	0.5313
M11	0.2287	0.5829	0.3923	0.6962
M12	-0.0861	0.5835	-0.1475	0.8833

Table 8 “Regression summary with monthly dummies for product P_2 ”

TERM	COEFFICIENTS	STANDARD ERROR	T STAT	P-VALUE
INTERCEPT	6.8327	1.8747	3.6447	0.0006
LINEAR TREND	-0.0263	0.0247	-1.0631	0.2925
QUADRATIC TREND	0.0002	0.0003	0.7318	0.4675
LAG VARIABLE T-1	0.5031	0.1362	3.6953	0.0005
LAG VARIABLE T-2	-0.0515	0.1363	-0.3782	0.7068
M2	0.0691	0.5789	0.1193	0.9055
M3	-0.5857	0.5575	-1.0506	0.2981
M4	-0.9145	0.5589	-1.6361	0.1076
M5	-0.5852	0.5736	-1.0203	0.3122
M6	0.0207	0.5845	0.0354	0.9719
M7	-0.5281	0.5803	-0.9100	0.3669
M8	-1.1265	0.5641	-1.9971	0.0509
M9	0.7548	0.5830	1.2947	0.2009
M10	0.2309	0.6040	0.3823	0.7038
M11	0.0316	0.5589	0.0565	0.9551
M12	-0.1990	0.5562	-0.3577	0.7220

Table 9 “Regression summary with monthly dummies and lag variables for product P₂”

Taking into consideration the results extracted from the regression model’s implementation the linear trend coefficient's *p-value* indicate statistical significance, confirming the aforementioned conclusions about trend behavior. The negative value of the coefficient of linear trend variable indicates a descending trend over years in product’s sales time series set. This observation aligns to the earlier visualization of the product's sales time series that was presented in **Figure 13** and detailed in the comprehensive analysis within **Table 4**. However, the significance of the time trend estimator vanishes with the inclusion of lag variables into the regression model.

Furthermore, a similar lack of statistical significance is observed across the majority of coefficients associated with the monthly dummies ($M_{t,i}$), pointing towards an absence of seasonality within the dataset. However, an exception is apparent for the month August, where the coefficient's *p-value* nearly equals 0.05, suggesting potential seasonal effects during this specific period.

The only coefficient displaying a *p-value* clearly below the designated significance level belongs to the lag variable representing the sales from a previous time period [$\log(S_{t-1})$].

This suggests a short-term memory characteristic within the time series. In essence, it signifies that the current behavior of the dependent variable (sales) can be explained by examining the sales observation from the preceding period and that there is a memory momentum effect in product’s sales that should be taken into consideration by the company.

4.6.3 Statistical Properties of product P_3

The study continues with the analysis of product P_3 , leveraging the regression summary output, as was presented in the preceding products. **Tables 10** and **11** encapsulates the results obtained from the application of regression model M1 to the sales time series, encompassing the explanatory monthly dummy and lag variables.

TERM	COEFFICIENTS	STANDARD		
		ERROR	T STAT	P-VALUE
INTERCEPT	12.0105	0.2265	53.0170	0.0000
LINEAR TREND	-0.0062	0.0100	-0.6142	0.5415
QUADRATIC TREND	0.0001	0.0001	0.6238	0.5352
M2	0.2559	0.2513	1.0184	0.3127
M3	0.2601	0.2513	1.0347	0.3051
M4	-0.1438	0.2514	-0.5721	0.5695
M5	0.0608	0.2515	0.2416	0.8099
M6	-0.0082	0.2516	-0.0327	0.9740
M7	-0.0153	0.2518	-0.0609	0.9516
M8	-0.7296	0.2519	-2.8962	0.0053
M9	0.3551	0.2521	1.4088	0.1642
M10	0.3316	0.2523	1.3144	0.1939
M11	0.4093	0.2525	1.6210	0.1104
M12	0.0560	0.2528	0.2215	0.8255

Table 10 “Regression summary with monthly dummies for product P_3 ”

TERM	COEFFICIENTS	STANDARD		
		ERROR	T STAT	P-VALUE
INTERCEPT	15.5295	2.4462	6.3485	0.0000
LINEAR TREND	-0.0080	0.0112	-0.7076	0.4822
QUADRATIC TREND	0.0001	0.0001	0.7680	0.4459
LAG VARIABLE T-1	-0.0553	0.1343	-0.4116	0.6823
LAG VARIABLE T-2	-0.2340	0.1348	-1.7365	0.0882

M2	0.2757	0.2804	0.9833	0.3299
M3	0.2225	0.2726	0.8164	0.4179
M4	-0.1217	0.2663	-0.4568	0.6496
M5	0.0611	0.2691	0.2270	0.8212
M6	-0.0915	0.2780	-0.3293	0.7432
M7	-0.0550	0.2714	-0.2028	0.8401
M8	-0.7863	0.2738	-2.8722	0.0058
M9	0.2568	0.2995	0.8576	0.3949
M10	0.1256	0.3115	0.4032	0.6884
M11	0.4554	0.2659	1.7126	0.0925
M12	0.1003	0.2674	0.3750	0.7091

Table 11 “Regression summary with monthly dummies and lag variables for product P_3 ”

The regression model's summary results reveal resemblances between the sales time series of product P_3 and those of product P_1 . The coefficients demonstrate statistical insignificance, except for the coefficient linked to the monthly dummy variable representing month August. This behavior suggests an absence of seasonality within the sales time series, except for month August. Moreover, the coefficients of the lag variables exceed the critical value $\alpha=0.05$. Thus, unlike the preceding product P_2 , the time series of the current product lack short-memory effect.

Finally, the high *p-values* (exceeding 0.05) of the coefficients linked to the linear and quadratic trend indicate no-trend behavior within the product's time series. This observation gains further reinforcement from the visual depiction of the time series in **Figure 16**. Moreover, it is thoroughly explained in the comprehensive analysis of the product's characteristics within subchapter 4.3.1, gaining support from the detailed insights presented in **Table 4**.

4.6.4 Statistical Properties of product P_4

The analysis procedure employed for prior products has been replicated for product P_4 , aiming to evaluate the outcomes derived from the implementation of the regression model M1. **Table 12** and **13**, encapsulates the output summary resulting from the MLR model, encompassing linear and quadratic trends, alongside monthly dummy and lag variables.

TERM	COEFFICIENTS	STANDARD	T STAT	P-VALUE
		ERROR		
INTERCEPT	11.4252	0.1985	57.5713	0.0000
LINEAR TREND	-0.0021	0.0088	-0.2417	0.8099
QUADRATIC TREND	0.0000	0.0001	0.2353	0.8148
M2	-0.1211	0.2201	-0.5501	0.5844
M3	-0.2589	0.2202	-1.1758	0.2445
M4	-0.2037	0.2202	-0.9248	0.3589
M5	-0.2396	0.2203	-1.0875	0.2813
M6	-0.1799	0.2204	-0.8162	0.4177
M7	-0.2909	0.2205	1.3188	0.1924
M8	-1.0906	0.2207	-4.9420	0.0000
M9	-0.0981	0.2208	-0.4442	0.6586
M10	0.3324	0.2210	1.5041	0.1380
M11	-0.1076	0.2212	-0.4863	0.6286
M12	-0.0921	0.2214	-0.4157	0.6792

Table 12 “Regression summary with monthly dummies for product P₄”

TERM	COEFFICIENTS	STANDARD	T STAT	P-VALUE
		ERROR		
INTERCEPT	10.6772	2.2197	4.8101	0.0000
LINEAR TREND	-0.0046	0.0101	-0.4610	0.6467
QUADRATIC TREND	0.0001	0.0001	0.4236	0.6735
LAG VARIABLE T-1	0.1181	0.1404	0.8412	0.4039
LAG VARIABLE T-2	-0.0440	0.1412	-0.3112	0.7568
M2	-0.1664	0.2467	-0.6743	0.5030
M3	-0.3009	0.2382	-1.2632	0.2120
M4	-0.2341	0.2409	-0.9718	0.3355
M5	-0.2820	0.2411	-1.1694	0.2474
M6	-0.2151	0.2410	-0.8925	0.3721
M7	-0.3342	0.2403	-1.3906	0.1701
M8	-1.1178	0.2420	-4.6184	0.0000
M9	-0.0353	0.2867	-0.1232	0.9024
M10	0.2431	0.2806	0.8664	0.3901
M11	-0.2039	0.2405	-0.8479	0.4002
M12	-0.1173	0.2436	-0.4815	0.6321

Table 13 “Regression summary with monthly dummies and lag variables for product P₄”

The high *p-values* associated with the coefficients related to the linear and quadratic trends, suggest the absence of trend attributes within the sales time series of the product. This observation finds support in **Figure 19** and is further reinforced by the comprehensive analysis of mean and variance values provided in **Table 4**.

Furthermore, the elevated *p-values* linked to the remaining explanatory variables, specifically concerning seasonality (monthly dummy variables) and autoregressive variables (lag variables) within the product's sales data, suggest an absence of a seasonal effect and the lack of short- memory effect within the sales time series. However, an exception is observed in the *p-value* of the coefficient associated with the monthly dummy variable representing August (M_8), which notably differs as it is much lower than the critical value of $\alpha = 0.05$. Consequently, that observation leads to the rejection of the Null Hypothesis (H_0) for that particular month, meaning that evidence exists that seasonality is present solely during month August.

4.6.5 Statistical Properties of product P_5

The study and analysis proceed with product P_5 , and the outcomes derived from implementing the MLR model M1, using Excel's functions and formulas, are encapsulated in the following **Table 14** and **15**.

TERM	COEFFICIENTS	STANDARD		
		ERROR	T STAT	P-VALUE
INTERCEPT	9.1200	0.4034	22.6097	0.0000
LINEAR TREND	0.0013	0.0178	0.0746	0.9408
QUADRATIC TREND	0.0000	0.0002	-0.1549	0.8775
M2	0.8126	0.4474	1.8161	0.0745
M3	2.8237	0.4475	6.3097	0.0000
M4	1.6680	0.4476	3.7261	0.0004
M5	1.8582	0.4478	4.1494	0.0001
M6	1.7401	0.4480	3.8840	0.0003
M7	1.7718	0.4483	3.9525	0.0002
M8	1.0650	0.4485	2.3742	0.0209
M9	-0.1793	0.4489	-0.3995	0.6910
M10	-0.8028	0.4492	-1.7870	0.0792
M11	-0.9203	0.4496	-2.0467	0.0452
M12	-0.8086	0.4501	-1.7964	0.0776

Table 14 “Regression summary with monthly dummies for product P_5 ”

TERM	COEFFICIENTS	STANDARD		
		ERROR	T STAT	P-VALUE
INTERCEPT	10.8767	1.6000	6.7978	0.0000
LINEAR TREND	-0.0065	0.0197	-0.3302	0.7425
QUADRATIC TREND	0.0001	0.0003	0.1973	0.8443
LAG VARIABLE T-1	0.0412	0.1328	0.3100	0.7577
LAG VARIABLE T-2	-0.2356	0.1331	-1.7700	0.0824
M2	0.9323	0.5004	1.8632	0.0679
M3	2.9325	0.5245	5.5907	0.0000
M4	1.8872	0.6947	2.7163	0.0088
M5	2.6004	0.7500	3.4672	0.0010
M6	2.2037	0.6696	3.2910	0.0018
M7	2.2862	0.6757	3.3833	0.0013
M8	1.5514	0.6692	2.3183	0.0242
M9	0.3445	0.6346	0.5429	0.5894
M10	-0.3935	0.5481	-0.7179	0.4759
M11	-0.7779	0.4842	-1.6068	0.1139
M12	-0.8080	0.4730	-1.7081	0.0934

Table 15 “Regression summary with monthly dummies and lag variables for product P₅”

The summary output of the regression model distinctly indicates the presence of seasonality within the sales time series of product **P₅**. Notably, the *p-values* associated with the coefficients of the monthly dummy variables corresponding to the Spring and Summer months are considerably lower than the critical value. This underscores the significance of the seasonality effect during these specific periods. This consistent behavior was previously emphasized in the graphical summary presentation of product **P₅** within subchapter 4.3.1. It was also confirmed by the visualization of the product’s time series plot in **Figure 22** and the examination of the product's Boxplot in **Figure 24**. Both graphs highlighted the observable seasonality during Spring and Summer months, aligning well with the findings obtained from the regression model's summary output.

Continuing the analysis, the elevated *p-values* associated with the coefficients of the linear and quadratic trend validate the no-trend attribute within the sales time series. This aligns with the observations made in subchapter 4.3.1 during the presentation of product **P₅**, where graphical summary and associated notes emphasized this nony behavior. Furthermore, the sales data of the product does not exhibit short-memory attributes, as indicated by the high

p-values of the corresponding coefficients of the lag variables that are depicted in **Table 15**. This implies a lack of statistical significance in the coefficients related to autoregressive variables within the sales time series.

4.6.6 Statistical Properties of product P_6

The study continues examining product P_6 , and **Tables 16** and **17** below showcase the summary output derived from implementing model M1 on the sales time series. This MLR model integrates components such as time trend, autoregression, and monthly dummy variables, similar to the methodology applied in earlier products analyses.

TERM	COEFFICIENTS	STANDARD		
		ERROR	T STAT	P-VALUE
INTERCEPT	9.3068	0.1088	85.5367	0.0000
LINEAR TREND	0.0215	0.0048	4.4780	0.0000
QUADRATIC TREND	-0.0003	0.0001	-4.0068	0.0002
M2	-0.1304	0.1207	-1.0802	0.2845
M3	0.1285	0.1207	1.0645	0.2915
M4	0.1404	0.1207	1.1626	0.2498
M5	-0.1939	0.1208	-1.6049	0.1140
M6	0.1098	0.1208	0.9082	0.3675
M7	0.0651	0.1209	0.5382	0.5925
M8	-0.0476	0.1210	-0.3933	0.6956
M9	0.0264	0.1211	0.2184	0.8279
M10	0.1393	0.1212	1.1494	0.2551
M11	-0.1548	0.1213	-1.2767	0.2068
M12	-0.0614	0.1214	-0.5055	0.6151

Table 16 “Regression summary with monthly dummies for product P_6 ”

TERM	COEFFICIENTS	STANDARD		
		ERROR	T STAT	P-VALUE
INTERCEPT	11.6474	1.7640	6.6028	0.0000
LINEAR TREND	0.0318	0.0070	4.5341	0.0000
QUADRATIC TREND	-0.0004	0.0001	-4.3144	0.0001
LAG VARIABLE T-1	-0.1041	0.1336	-0.7789	0.4394
LAG VARIABLE T-2	-0.1666	0.1300	-1.2820	0.2053
M2	-0.0703	0.1306	-0.5383	0.5926
M3	0.2194	0.1267	1.7320	0.0890

M4	0.2356	0.1283	1.8356	0.0719
M5	-0.0551	0.1338	-0.4121	0.6819
M6	0.2150	0.1309	1.6425	0.1063
M7	0.1455	0.1277	1.1395	0.2595
M8	0.0783	0.1310	0.5975	0.5526
M9	0.1327	0.1281	1.0363	0.3047
M10	0.2341	0.1266	1.8490	0.0699
M11	-0.0362	0.1308	-0.2766	0.7831
M12	0.0454	0.1307	0.3471	0.7299

Table 17 “Regression summary with monthly dummies and lag variables for product P_6 ”

Upon examining the regression results of M1 model and assessing the *p-values* of the variables' coefficients, it becomes evident that the Null Hypothesis (H_0) cannot be rejected for the coefficients of monthly dummies and lag variables. This outcome implies an absence of both the seasonality and the short-memory effect within the product's sales time series. Additionally, the coefficients' *p-values* associated with the linear and quadratic trend suggest statistical significance. The linear trend exhibits a positive coefficient indicating an increase in sales over time. Conversely, the quadratic trend displays a slightly negative coefficient suggesting a minor decrease in this growth rate over time.

The linear trend variable's coefficient signifies the linear change, which, in the case of product P_6 , depicts an ascending trend until year 2019, representing a consistent shift over time. Meanwhile, the quadratic coefficient illustrates the deviation from this linear change, capturing the alteration or evolution of this linear shift over time. It essentially delineates the rate of change within the linear trend itself, providing insight into how the linear trend is altering its trajectory over time. This analysis was also illustrated in subchapter 4.3.1, where the characteristics of the product's time series plot were presented.

4.6.7 Statistical Properties of product P_7

The study concludes its analysis about linear and quadratic trend, seasonal effect and autoregressive sales conditions, with product P_7 . The details of the regression model M1 summary are outlined in the following **Tables 18** and **19**.

TERM	COEFFICIENTS	STANDARD		
		ERROR	T STAT	P-VALUE
INTERCEPT	10.9099	0.1037	105.2399	0.0000
LINEAR TREND	0.0053	0.0046	1.1483	0.2556
QUADRATIC TREND	0.0000	0.0001	-0.5295	0.5985
M2	-0.2842	0.1150	-2.4713	0.0164
M3	-0.4339	0.1150	-3.7722	0.0004
M4	-0.4735	0.1150	-4.1156	0.0001
M5	-0.2772	0.1151	-2.4085	0.0192
M6	-0.2966	0.1151	-2.5756	0.0126
M7	-0.3901	0.1152	-3.3862	0.0013
M8	-0.3882	0.1153	-3.3675	0.0014
M9	-0.1209	0.1154	-1.0481	0.2989
M10	-0.4770	0.1155	-4.1315	0.0001
M11	-0.4281	0.1156	-3.7046	0.0005
M12	-0.5148	0.1157	-4.4501	0.0000

Table 18 “Regression summary with monthly dummies for product P₇”

TERM	COEFFICIENTS	STANDARD		
		ERROR	T STAT	P-VALUE
INTERCEPT	16.9537	1.9740	8.5886	0.0000
LINEAR TREND	0.0115	0.0050	2.3111	0.0247
QUADRATIC TREND	-0.0001	0.0001	-1.4963	0.1404
LAG VARIABLE T-1	-0.1361	0.1246	-1.0921	0.2796
LAG VARIABLE T-2	-0.4506	0.1294	-3.4833	0.0010
M2	-0.2401	0.1350	-1.7785	0.0810
M3	-0.1577	0.1363	-1.1571	0.2523
M4	-0.3459	0.1175	-2.9442	0.0048
M5	-0.2226	0.1135	-1.9615	0.0550
M6	-0.2331	0.1180	-1.9760	0.0533
M7	-0.2408	0.1216	-1.9802	0.0528
M8	-0.2602	0.1181	-2.2036	0.0318
M9	-0.0345	0.1158	-0.2982	0.7667
M10	-0.3530	0.1267	-2.7862	0.0073
M11	-0.2317	0.1240	-1.8683	0.0671
M12	-0.4717	0.1140	-4.1362	0.0001

Table 19 “Regression summary with seasonal dummies and lag variables for product P₇”

The regression results presented in the preceding tables unequivocally indicate the presence of seasonality within the sales time series of product **P₇**. Notably, the *p-values* associated with the coefficients for the majority of the monthly dummy variables either equal or slightly exceed the critical value. An exception to this observation is only evident in the months of March and September. This underscores the fact that respective of the significance level defined in the study, the seasonality effect is conspicuous, if not at the 5% level, certainly at the 10% significance level.

Moreover, the *p-value* of the linear trend coefficient also falls below the critical value, indicating the significance of the linear trend. The positive value of the coefficient ($\beta_t = 0.0115$) signifies a marginal ascending trend within the product's sales data throughout the analyzed period. Lastly, the significantly low *p-value* of the lag variable's ($t-2$) coefficient, suggests the presence of the short-memory effect within the sales time series indicating that the product's demand does not expires at the same time observation, but affects the two forthcoming periods. This observation could be proved valuable information for company's decision-makers and sales managers in terms of effective inventory management.

4.7 Cross-dependencies between the studied products - MLR model M2

It is frequently noticed that when similar or category-related products coexist within the same market, especially when they exhibit minimal differentiating characteristics to consumers, influence each other's demand patterns. Identifying such interconnected demand behaviors becomes crucial, and, by acknowledging and understanding these cross-dependencies, managers can integrate appropriate estimations into their decision-making processes, especially regarding future inventory requirements for the company's products. This proactive approach helps in optimizing inventory levels, production planning, S&OP strategies, efficient supply chain management in total, and to align with the dynamic demands of the market. Overall, recognizing and analyzing cross-dependencies, enables companies to make more informed decisions across various aspects of their business, ultimately contributing to improved performance and competitiveness.

Expanding the prior regression model M1 (as detailed in equation 4.5.1) for each product, the study aims to incorporate the natural logarithm of sales observations from the other products. To mitigate model complexity, considering statistically insignificant explanatory variables as discussed in chapter 4.6 for each product, the revised regression model M2 will

Postgraduate Dissertation

selectively encompass solely the variables demonstrating substantial evidence of explanatory dynamics. The updated MLR model M2 that will be applied for the whole dataset observations (72 in total), can be generally described by the following equation:

$$M2: \log(S_{t,k}) = \beta_{0,k} + \beta_{t,k}t + \gamma_k t^2 + \sum_{i=1}^{11} \beta_{i,k} M_{t,i} + \sum_{j=1}^2 \delta_{j,k} \log(S_{t-j,k}) + \sum_{j \neq k} \alpha_{j,k} \log(S_{t,j}) + e_{t,k} \quad (4.7.1)$$

where $k = 1, 2, \dots, 7$ indicates the examined product, $\beta_{0,k}$ is the intercept term of the time series corresponding to effect of the baseline month January, $\beta_{t,k}$ and γ_k are the slope coefficients of the linear and quadratic trend respectively, $\beta_{i,k}$ is the coefficient for each monthly dummy variable $M_{t,i}$ where $i = 2, 3, \dots, 12$, $\delta_{j,k}$ is the coefficient for each lag variable $S_{t-j,k}$ where $j = 1, 2$, $\alpha_{j,k}$ is the coefficient of the logarithmic sales time series variable $S_{t,j}$ for the different product brands where $j = (1, 2, \dots, k - 1)$, and $e_{t,k}$ is the disturbance or error term.

In line with the prior discussion and evaluation presented in chapter 4.6 regarding the implementation of regression model M1, the significance of coefficients for each product's sales ($\alpha_{j,k}$) will be evaluated using a two-tailed t -distribution. Here, the null hypothesis ($H_0: \alpha_{j,k} = 0$) signifies that the true value of the coefficient is zero, whereas the alternative hypothesis ($H_1: \alpha_{j,k} \neq 0$) indicates its significance. Similar to the previous chapter's analysis for each product, the assessment of ***p-values***, derived from the regression summary output, will be conducted at a 95% confidence level.

4.7.1 Interdependencies of product **P₁** with **P₂, P₃, P₄, P₅, P₆, P₇**

Beginning with the first product **P₁**, the following **Table 20** presents the regression summary output of model M2 implementation in product's sales time series.

TERM	COEFFICIENTS	STANDARD ERROR	T STAT	P- VALUE
INTERCEPT	4.4682	4.5431	0.9835	0.3291
M5	-0.8591	0.2988	-2.8550	0.0055
M8	-0.1760	0.3321	-0.5301	0.5979
log(<i>S</i> ₂)	0.1350	0.0872	1.5478	0.1267
log(<i>S</i> ₃)	0.3707	0.2150	1.7239	0.0896
log(<i>S</i> ₄)	0.1375	0.2277	0.6038	0.5481
log(<i>S</i> ₅)	0.0128	0.0577	0.2211	0.8257

$\log(S_6)$	0.1860	0.3463	0.5370	0.5932
$\log(S_7)$	-0.1892	0.3217	-0.5883	0.5584

Table 20 “Summary output of M2 model investigating cross-dependencies between product P_1 and the remaining six products”

From the outcomes stemming from the application of model M2 to the sales data of product P_1 , coupled with the notably high *p-values* associated with coefficients linked to the other products, indicate the absence of interdependencies between product P_1 and the remaining products at a 95% confidence level. Nonetheless, in case of product P_3 , it should be stressed that if not at 5% significance level, but certainly at 10%, a cross-dependence between the product’s sales and the examined product P_1 exists.

4.7.2 Interdependencies of product P_2 with $P_1, P_3, P_4, P_5, P_6, P_7$

The examination of cross-dependencies within product sales, proceeds focusing on exploring potential correlations involving product P_2 with the other products. The summarized output of implementing regression model M2, utilizing the sales of P_2 as the dependent variable and the sales of the other products as explanatory independent variables, is encapsulated in **Table 21**.

TERM	COEFFICIENTS	STANDARD ERROR	T STAT	P- VALUE
INTERCEPT	-17.3180	5.7780	-2.9972	0.0039
LAG VARIABLE	0.4964	0.0789	6.2904	0.0000
T-1				
M8	0.4376	0.3713	1.1784	0.2431
$\log(S_1)$	0.3110	0.1294	2.4034	0.0192
$\log(S_3)$	0.9496	0.2310	4.1114	0.0001
$\log(S_4)$	0.3361	0.2553	1.3166	0.1928
$\log(S_5)$	-0.0902	0.0620	-1.4566	0.1503
$\log(S_6)$	-0.4149	0.3652	-1.1359	0.2604
$\log(S_7)$	0.8316	0.3760	2.2116	0.0307

Table 21 “Summary output of M2 model investigating cross-dependencies between product P_2 and the remaining six products”

The observed low *p-values* associated with the coefficients of products P_1, P_3 and P_7 signify statistical significance. This indicates robust positive cross-dependencies between the sales of product P_2 and the sales of products P_1, P_3 and P_7 . Understanding these cross-dependencies can provide valuable insights for decision-making. For example, it may inform

product development strategies, pricing decisions, or marketing campaigns by highlighting which products have synergistic effects or how changes in one product might affect the performance of others. Conversely, the notably higher *p-values* linked to the coefficients of the other products (P_4 , P_5 , P_6) lead to the failure of rejecting the Null Hypothesis. This suggests a lack of evidence supporting the possibility of cross-dependence between product P_2 and these specific products.

4.7.3 Interdependencies of product P_3 with P_1 , P_2 , P_4 , P_5 , P_6 , P_7

The study progresses by exploring potential cross-dependencies involving the sales of product P_3 and those of the remaining products. The outcomes derived from implementing regression model M2 are summarized in **Table 22**, showcasing the relationships between the sales of product P_3 and the sales of the other products.

TERM	STANDARD			
	COEFFICIENTS	ERROR	T STAT	P-VALUE
INTERCEPT	4.8463	2.5175	1.9250	0.0587
M8	-0.3073	0.1847	-1.6637	0.1011
$\log(S_1)$	0.1002	0.0666	1.5041	0.1375
$\log(S_2)$	0.1427	0.0469	3.0423	0.0034
$\log(S_4)$	0.3467	0.1228	2.8234	0.0063
$\log(S_5)$	0.0357	0.0316	1.1300	0.2627
$\log(S_6)$	-0.1075	0.1863	-0.5768	0.5661
$\log(S_7)$	0.1148	0.1820	0.6310	0.5303

Table 22 “Summary output of M2 model investigating cross-dependencies between product P_3 and the remaining six products”

The detected low *p-values* associated with the coefficients of products P_2 and P_4 suggest a robust positive cross-dependence between the sales of product P_3 and the sales of these specific products. It is observed that all these three products belong to the respiratory or pain relief category of products, thus, it could be concluded that the company should take such indications into consideration in order to effectively design and develop production strategies and inventory policies. On the contrary, the high *p-values* linked to the coefficients of the other products (P_1 , P_5 , P_6 and P_7) result in the inability to reject the Null Hypothesis. This observation signifies the lack of evidence supporting cross-dependence between product P_3 and these particular products.

4.7.4 Interdependencies of product P_4 with $P_1, P_2, P_3, P_5, P_6, P_7$

The analysis proceeds with the investigation of potential intercorrelations between the sales time series of product P_4 and those of the other products. **Table 23** contains the outcomes derived from implementing model M2, shedding light on the relationships between the sales of product P_4 and the sales of the other products. The results of the regression model implementation indicates that the sales of product P_4 exhibit cross-dependencies specifically with the sales of product P_3 , evidenced by the notably low *p-value* of the coefficient associated with that product. This observation aligns with the earlier findings from **Table 22**, confirming a strong interrelationship between these particular products’ sales. That assessment could be explained by the fact that these two products belong to the same products’ category, providing company the evidence that sales and promotion campaigns should be employed with taking into account that information. On the other hand, the considerably high *p-values* associated with the coefficients of the other products, drive to failing the rejection of the Null Hypothesis. This signifies a lack of evidence in supporting the hypothesis of cross-dependence between product P_4 and the remaining products at a 95% confidence level.

TERM	COEFFICIENTS	STANDARD ERROR	T STAT	P- VALUE
INTERCEPT	4.5051	2.4206	1.8611	0.0673
M8	-0.5858	0.1656	-3.5373	0.0008
log(S_1)	0.0275	0.0649	0.4236	0.6733
log(S_2)	0.0625	0.0475	1.3155	0.1930
log(S_3)	0.3194	0.1131	2.8234	0.0063
log(S_5)	-0.0255	0.0305	-0.8376	0.4054
log(S_6)	0.2817	0.1758	1.6022	0.1140
log(S_7)	-0.0530	0.1751	-0.3030	0.7629

Table 23 “Summary output of M2 model investigating cross-dependencies between product P_4 and the remaining six products”

4.7.5 Interdependencies of product P_5 with $P_1, P_2, P_3, P_4, P_6, P_7$

The same methodology was followed to investigate possible interdependencies between the sales of product P_5 and those of the remaining products. The summarized outcomes obtained from implementing the M2 model are presented in **Table 24**.

TERM	COEFFICIENTS	STANDARD		P-VALUE
		ERROR	T STAT	
INTERCEPT	-6.2768	6.5075	-0.9646	0.3387
M3	3.2968	0.4091	8.0586	0.0000
M4	2.1503	0.4160	5.1687	0.0000
M5	2.3099	0.4392	5.2600	0.0000
M6	2.0884	0.3881	5.3813	0.0000
M7	2.2071	0.3971	5.5781	0.0000
M8	1.8564	0.5007	3.7074	0.0005
log(S_1)	0.2032	0.1763	1.1521	0.2539
log(S_2)	-0.1465	0.1232	-1.1887	0.2393
log(S_3)	0.1784	0.3127	0.5589	0.5783
log(S_4)	0.2625	0.3262	0.8049	0.4241
log(S_6)	-0.0026	0.4929	-0.0054	0.9957
log(S_7)	0.8585	0.4748	1.8084	0.0756

Table 24 “Summary output of M2 model investigating cross-dependencies between product P_5 and the remaining six products”

Assessing the summary output depicted in the previous table, it becomes apparent that the sales of product P_5 do not exhibit cross-dependencies with the sales of the other products. This conclusion is drawn from the notably high *p-values* associated with the products' sales coefficients, indicating a lack of evidence to reject the Null Hypothesis. Such an observation is indeed logical, given that the product category is not correlated with the others under consideration. Nonetheless, in case of product P_7 , it should be stressed that if not at 5% significance level, but certainly at 10%, a cross-dependence between the product's sales and the examined product P_5 exists.

4.7.6 Interdependencies of product P_6 with P_1 , P_2 , P_3 , P_4 , P_5 , P_7

The investigation of possible cross-dependencies among the products' sales continues with product P_6 . The summarized outcomes obtained from implementing the M2 model are presented in **Table 25**.

TERM	COEFFICIENTS	STANDARD		P-
		ERROR	T STAT	VALUE
INTERCEPT	5.8658	1.3442	4.3638	0.0000
LINEAR TREND	0.0202	0.0051	3.9200	0.0002
QUADRATIC TREND	-0.0002	0.0001	-3.6224	0.0006
log(S_1)	0.0747	0.0401	1.8619	0.0673

$\log(S_2)$	-0.0199	0.0330	-0.6012	0.5499
$\log(S_3)$	-0.0850	0.0756	-1.1240	0.2653
$\log(S_4)$	0.1146	0.0705	1.6248	0.1092
$\log(S_5)$	0.0421	0.0197	2.1337	0.0368
$\log(S_7)$	0.1987	0.1089	1.8246	0.0728

Table 25 “Summary output of M2 model investigating cross-dependencies between product P_6 and the remaining six products”

The analysis of the preceding table predominantly indicates the rejection of cross-dependencies between the sales of product P_6 and the sales of the majority of the other products. However, an exception to this observation is detected in the case of product P_5 that stands out. The low *p-value* associated with this specific product's sales coefficient implies a positive influence of P_5 sales on the sales of P_6 . This suggests that, if all else is equal, a potential increase in the sales of product P_5 could correspond to a 4.21% increase in thousands of products in sales for P_6 . While the two products belong to different and unrelated categories, the observed phenomenon suggests that there may still be some interconnectedness between their sales demand. This subtle increase in sales demand resulting from the increase in sales of another product should not be overlooked by the company's managers. It underscores the importance of considering such observations when making operational decisions, as they may have implications for overall business strategy and performance. Furthermore, in case of products P_1 and P_7 , it should be stressed that if not at 5% significance level, but certainly at 10%, cross-dependencies between these products' sales and the examined product P_6 exist.

4.7.7 Interdependencies of product P_7 with $P_1, P_2, P_3, P_4, P_5, P_6$

The last product that is going to be investigated for possible sales cross-dependencies with the other products' sales, is the P_7 . **Table 26** presents the summarized output derived from implementing regression model M2 in the sales time series of product P_7 , incorporating the natural logarithms of the remaining products' sales as independent variables.

TERM	COEFFICIENTS	STANDARD ERROR	T STAT	P- VALUE
INTERCEPT	11.5682	1.6084	7.1922	0.0000
LINEAR TREND	0.0030	0.0013	2.2817	0.0263

LAG VARIABLE	-0.4097	0.1112	-3.6863	0.0005
T-2				
M4	-0.1876	0.0852	-2.2012	0.0319
M8	-0.0384	0.1032	-0.3722	0.7112
M10	-0.3697	0.1031	-3.5858	0.0007
M11	-0.1266	0.1084	-1.1676	0.2479
M12	-0.3610	0.0946	-3.8142	0.0003
$\log(S_1)$	-0.0211	0.0373	-0.5672	0.5729
$\log(S_2)$	0.0116	0.0293	0.3949	0.6944
$\log(S_3)$	0.0514	0.0724	0.7104	0.4804
$\log(S_4)$	0.0079	0.0742	0.1066	0.9155
$\log(S_5)$	-0.0429	0.0247	-1.7345	0.0833
$\log(S_6)$	0.3460	0.1027	3.3700	0.0014

Table 26 “Summary output of M2 model investigating cross-dependencies between product P_7 and the remaining six products”

The *p-values* resulting from the implementation of the M2 regression model in the sales data of product P_7 , suggest that the coefficients associated with all other products lack statistical significance (exceeding the critical value $\alpha=0.05$). However, an exception arises with product P_6 , where its *p-value* is notably lower than the critical value. This signifies evidence that the sales of product P_6 impact the dependent variable (P_7 ' sales) within the model. If all other factors remain constant, a potential increase in the sales of product P_6 corresponds to a 34.6% increase, in thousands of sales products for P_7 . That assessment could be explained by the fact that these two products belong to the same products' category, providing company the evidence that sales and promotion campaigns should be employed with taking into account that information. Nonetheless, in case of product P_5 , it should be highlighted that if not at 5% significance level, but certainly at 10%, a cross-dependence between this product's sales and the examined product P_7 exists.

4.8 The effect of the Covid-19 pandemic period - MLR model M3

The pandemic period had a profound impact on supply chain resilience and its efficient operations worldwide. It underscored the critical importance of maintaining a smooth flow of products within global supply chains and highlighted the necessity of sustainable operations across all parties within these chains. The crisis revealed the substantial consequences arising from supply chain disruptions at various levels, spanning from

upstream to downstream members. These disruptions significantly affected the effective performance of all industries, irrespective of their focus or sector.

The emergence of the coronavirus disease, officially termed the Covid-19 pandemic by the World Health Organization (WHO) in early March of 2020, and its swift escalation in subsequent months, led to a variety of challenges across global economies and industries, impacting also the pharmaceutical sector and disrupting the flow of medicines within the market. The consequential short-term and long-term effects within this sector underscored the critical necessity for meticulous planning and well-founded decision-making processes to navigate the arising challenges effectively (Ayati, N., Saiyarsarai, P., & Nikfar, S., 2020).

In this research, it is endeavored the assessment of the pandemic's potential influence on consumer behavior in the Greek market. In order to evaluate a possible alteration in sales patterns, a new regression model will be developed taking into account as covid period that period from the beginning of the first preventive measures imposed by the Greek government until the time that all of them were withdrawn. The study spans the 72 observations in total, including the covid period from the initial lockdown in March 2020 to the complete removal of restrictions in April 2022 (Wikipedia, 2023). To analyze this impact, a dummy variable labeled CE_t will be employed. The beginning month will be March 2020, meaning that this dummy variable will assume a value of 0 for observations outside the pandemic period and the value 1 within it (from March 2020 to April 2022). Following this, the new regression model (M3) will be formulated, comprising solely as explanatory variables those that exhibited statistical significance for every product in the earlier analyses.

The subsequent equation illustrates the updated MLR model M3, encompassing the newly introduced dummy variable detailed previously. This model presents as an extension of the preceding MLR model M2 depicted in equation (4.7.1).

$$M3: \log(S_{t,k}) = \beta_{0,k} + \beta_{t,k}t + \gamma_k t^2 + \sum_{i=1}^{11} \beta_{i,k} M_{t,i} + \sum_{j=1}^2 \delta_{j,k} \log(S_{t-j,k}) + \sum_{j \neq k} \alpha_{j,k} \log(S_{t,j}) + \beta_C CE_t + e_{t,k} \quad (4.8.1)$$

where, $k = 1, 2, \dots, 7$ indicates the examined product, $\beta_{0,k}$ is the intercept term of the time series corresponding to effect of the baseline month January, $\beta_{t,k}$ and γ_k are the slope coefficients of the linear and quadratic trend respectively, $\beta_{i,k}$ is the coefficient for each monthly dummy variables $M_{t,i}$ where $i = 2, 3, \dots, 12$, $\delta_{j,k}$ is the coefficient for each lag

variable $S_{t-j,k}$ where $j = 1, 2$, $\alpha_{j,k}$ is the coefficient of the logarithmic sales time series variables $S_{t,j}$, of the different product brands where $j = (1, 2, \dots, k - 1)$, β_C is the coefficient of the dummy variable CE_t and $e_{t,k}$ is the disturbance or error term.

The upcoming analysis will follow a methodology similar to previous chapters, focusing on appraising the significance of the pandemic coefficient (β_C) for each product k . The assessment of significance will employ a two-tailed t -distribution. Under this framework, the Null Hypothesis ($H_0: \beta_C = 0$) suggests the coefficient's insignificance, while the alternative hypothesis ($H_1: \beta_C \neq 0$) implies its significance. Similar to analyses in preceding chapters, the evaluation of ***p-values*** extracted from the regression summary output will be undertaken at a 95% confidence level.

4.8.1 Covid-19 impact on sales of product **P₁**

Beginning with the first product **P₁**, the following **Table 27** presents the regression summary output of model M3 implementation in product's sales time series.

TERM	COEFFICIENTS	STANDARD ERROR	T STAT	P- VALUE
INTERCEPT	11.9120	0.1038	114.7539	0.0000
M5	-1.0687	0.2825	-3.7834	0.0003
M8	-0.8067	0.2825	-2.8556	0.0057
CE	-0.0038	0.1619	-0.0238	0.9811

Table 27 “Summary output of M3 model investigating possible impact of Covid-19 pandemic on product **P₁”**

From the outcomes stemming from the application of model M3 to the sales data of product **P₁**, it is observed that the pandemic period did not affect the product's sales at a 95% confidence level, as the variable's coefficient presents a ***p-value*** that is significantly higher than the critical value $\alpha=0.05$.

4.8.2 Covid-19 impact on sales of product **P₂**

The subsequent product under examination is **P₂** and **Table 28** exhibits the outcomes derived from the application of regression model M3 to its sales data.

TERM	COEFFICIENTS	STANDARD ERROR	T STAT	P- VALUE
INTERCEPT	-18.3470	4.3783	-4.1909	0.0001
LAG VARIABLE				
T-1	0.4440	0.0767	5.7868	0.0000
$\log(S_1)$	0.3423	0.1189	2.8778	0.0054
$\log(S_3)$	1.0559	0.1772	5.9593	0.0000
$\log(S_7)$	0.7472	0.3481	2.1464	0.0356
CE	-0.5971	0.1740	-3.4310	0.0010

Table 28 “Summary output of M3 model investigating possible impact of Covid-19 pandemic on product P_2 ”

The regression outcomes reveal statistical significance across all explanatory variables, including the covid effect variable, evidenced from their coefficients' notably lower *p-values* compared to the critical value $\alpha=0.05$. Hence, there is clear evidence at a 95% confidence level that the sales of the product were influenced by the Covid-19 period.

4.8.3 Covid-19 impact on sales of product P_3

The next product under examination for the possible impact of the pandemic period on its sales, is product P_3 . **Table 29** includes the results of applying the M3 regression model.

TERM	COEFFICIENTS	STANDARD ERROR	T STAT	P- VALUE
INTERCEPT	4.8109	0.9834	4.8919	0.0000
$\log(S_2)$	0.2065	0.0443	4.6647	0.0000
$\log(S_4)$	0.4312	0.1003	4.2998	0.0001
CE	0.2276	0.0934	2.4361	0.0175

Table 29 “Summary output of M3 model investigating possible impact of Covid-19 pandemic on product P_3 ”

According to regression summary output, all the applied explanatory variables, including the covid variable, give signs of statistical significance as their coefficients' *p-values* indicate. Thus, there is evidence at a 95% confidence level, that the sales of the product were influenced by the pandemic period.

4.8.4 Covid-19 impact on sales of product P_4

The same analysis continues with product P_4 . **Table 30** illustrates the regression outcomes following the implementation of the M3 model on the sales time series of this product.

TERM	COEFFICIENTS	STANDARD		
		ERROR	T STAT	P-VALUE
INTERCEPT	6.2553	1.1349	5.5117	0.0000
M8	-0.6136	0.1666	-3.6830	0.0005
log(S_3)	0.4179	0.0939	4.4497	0.0000
CE	-0.0524	0.0835	-0.6278	0.5322

Table 30 “Summary output of M3 model investigating possible impact of Covid-19 pandemic on product P_4 ”

The coefficient associated with the covid variable exhibits a high *p-value*, surpassing the critical value. This implies that, at a 95% confidence level, there is insufficient evidence to reject the Null Hypothesis ($H_0: \beta_C = 0$), indicating the statistical insignificance of this coefficient. Hence, it can be concluded that the pandemic period did not influence P_4 's sales.

4.8.5 Covid-19 impact on sales of product P_5

Continuing with the Covid-19 period possible effect on products' sales, the study will proceed examining product P_5 . **Table 31** depicts the regression model M3 results, that was implemented in that product sales.

TERM	COEFFICIENTS	STANDARD		P-
		ERROR	T STAT	VALUE
INTERCEPT	8.7908	0.1619	54.2827	0.0000
M3	3.1489	0.3857	8.1642	0.0000
M4	1.9920	0.3857	5.1646	0.0000
M5	2.1785	0.3840	5.6730	0.0000
M6	2.0593	0.3840	5.3624	0.0000
M7	2.0896	0.3840	5.4413	0.0000
M8	1.3813	0.3840	3.5970	0.0006
CE	-0.0145	0.2155	-0.0672	0.9467

Table 31 “Summary output of M3 model investigating possible impact of Covid-19 pandemic on product P_5 ”

The coefficient of the covid variable exhibits a high *p-value*, significantly greater than the critical value. This indicates that, at a 95% confidence level, the Null Hypothesis ($H_0: \beta_C = 0$) cannot be rejected. Therefore, it can be concluded that the pandemic period had no impact on product's sales.

4.8.6 Covid-19 impact on sales of product P_6

The subsequent product under scrutiny regarding potential sales variations during the pandemic period, is P_6 . The summary output derived from the MLR model M3 that was applied to its sales time series is detailed in **Table 32**.

TERM	COEFFICIENTS	STANDARD ERROR	T STAT	P- VALUE
INTERCEPT	8.9384	0.2119	42.1797	0.0000
LINEAR TREND	0.0208	0.0058	3.6181	0.0006
QUADRATIC TREND	-0.0002	0.0001	-3.4361	0.0010
$\log(S_5)$	0.0379	0.0189	2.0041	0.0491
CE	0.0206	0.0695	0.2965	0.7678

Table 32 “Summary output of M3 model investigating possible impact of Covid-19 pandemic on product P_6 ”

Much like the analysis conducted on previous products, the elevated *p-value* associated with the coefficient of the covid effect dummy variable, suggests that there is no evidence that the pandemic period influenced the product’s sales.

4.8.7 Covid-19 impact on sales of product P_7

The final product under investigation for potential impact of the Covid-19 pandemic on its sales is product P_7 . The outcomes derived from applying the M3 model to the sales data of this product are summarized in **Table 33**.

TERM	COEFFICIENTS	STANDARD ERROR	T STAT	P- VALUE
INTERCEPT	12.6170	1.3376	9.4329	0.0000
LINEAR TREND	0.0040	0.0013	3.0080	0.0038
LAG VARIABLE T-2	-0.4702	0.0968	-4.8584	0.0000
M4	-0.1958	0.0814	-2.4041	0.0192
M10	-0.2599	0.0816	-3.1848	0.0023
M12	-0.2765	0.0812	-3.4075	0.0012
$\log(S_6)$	0.3160	0.0964	3.2786	0.0017
CE	-0.0542	0.0557	-0.9729	0.3344

Table 33 “Summary output of M3 model investigating possible impact of Covid-19 pandemic on product P_7 ”

The summary output of MLR M3 model indicates that the ***p-value*** associated with the covid effect dummy variable's coefficient exceeds the critical value of $\alpha=0.05$. As a result, the Null Hypothesis cannot be rejected. Consequently, there is insufficient evidence suggesting that the pandemic period had a discernible influence on the sales of this product at a 95% confidence level.

4.9 Sales Forecasting Models Designation and Implementation

As outlined in chapter 4.4, the methodology employed for this dissertation entails the utilization of various MLR models (M1, M2, M3) for each product under scrutiny. This approach is undertaken to investigate the statistical properties of the examined set of products and to facilitate the analysis of sales patterns. In subsequent chapters, the initial MLR model M1 will be utilized as forecasting tool, enabling the generation of future demand estimates. Considering the ***p-values*** of the explanatory variables, that will be calculated by the Excel's regression tool, and evaluating their statistical significance as discussed in prior chapters, only those that seem significant and the lag variables will be retained in the forecasting models. Irrelevantly lag variables' statistical significance, they are retained in forecasting models due to their ability to offer comprehensive insights into the relationships between variables, aiding in the interpretation of how past values influence present outcomes.

The upcoming subchapters will introduce the regression models that will be utilized within each product's estimation sample, comprising 48 out of the total 72 observations (spanning from January 2017 to December 2020). The summary output that will be derived from Excel's linear regression tool will be also presented and the final structure of the forecasting MLR M1 model will be designated. Subsequently, the M1 model will be applied to the forecasting sample encompassing 24 observations, spanning from January 2021 to December 2022. This process aims to assess the model's estimation accuracy utilizing indicators like MAE and MAPE.

The second forecasting technique to be employed is the EWMA model, previously detailed in chapter 3.1. Equation (3.2.1) will be applied to the estimation sample, and subsequently, the MAE and MAPE metrics will be computed. Following this, the Solver package within Excel will be utilized to ascertain the optimal value for the λ parameter. The objective is to minimize the MAPE metric within the estimation sample, thereby enhancing the forecasting performance of the model. In order a comprehensive analysis of the outcomes to be derived,

Postgraduate Dissertation

graphical representations depicting the forecasting techniques and models' errors will be generated. These visualizations aim to facilitate a deeper comprehension of the accuracy and adequacy of the models. Finally, a comparative evaluation of the MAE and MAPE indicators' values derived from the forecasting samples for both models, will be conducted to draw important conclusions.

4.9.1 Forecasting MLR model M1 of product P_1

The following **Table 34** presents the summary output derived from implementing the regression model M1 in the sales data of the product's estimation sample. The model's results enable the researcher to determine the form of the forecasting model that will be applied that will produce the sales estimations.

TERM	COEFFICIENTS	STANDARD ERROR	T STAT	P- VALUE
INTERCEPT	14.8356	2.8736	5.1627	0.0000
LINEAR				
TREND	-0.0289	0.0184	-1.5737	0.1260
QUADRATIC				
TREND	0.0005	0.0003	1.3786	0.1782
LAG				
VARIABLE T-1	0.1191	0.1739	0.6852	0.4985
LAG				
VARIABLE T-2	-0.3100	0.1741	-1.7802	0.0852
M2	-0.2605	0.2891	-0.9013	0.3746
M3	-0.3631	0.2759	-1.3162	0.1981
M4	-0.4060	0.2706	-1.5002	0.1440
M5	-0.7012	0.2694	-2.6032	0.0142
M6	-0.3259	0.2760	-1.1808	0.2470
M7	-0.6827	0.2726	-2.5039	0.0180
M8	-1.4588	0.2723	-5.3578	0.0000
M9	0.0254	0.3396	0.0748	0.9409
M10	-0.3476	0.3365	-1.0331	0.3098
M11	-0.2809	0.2790	-1.0068	0.3221
M12	-0.1973	0.2785	-0.7085	0.4841

Table 34 “Summary output of M1 model implemented in product P_1 ”

The equation below is the MLR model that will be implemented in product's forecasting sample.

$$M1: \log(\widehat{S}_{t,1}) = 14.84 + 0.12 \log(S_{t-1,1}) - 0.31 \log(S_{t-2,1}) - 0.70M_5 - 0.68M_7 - 1.46M_8 \quad (4.9.1.1)$$

It should not be forgotten that the sales data was transformed into natural logarithms, thus, in order to apply the aforementioned model and derive conclusions, the built-in function EXP(.) in Excel has to be applied to inverse the function LN(.) that was initially applied.

The diagram in **Figure 31** presents the predicted sales against the actual sales during the performance period, subsequent to the application of the model M1 in the out-of-sample dataset (regression equation 4.9.1.1).

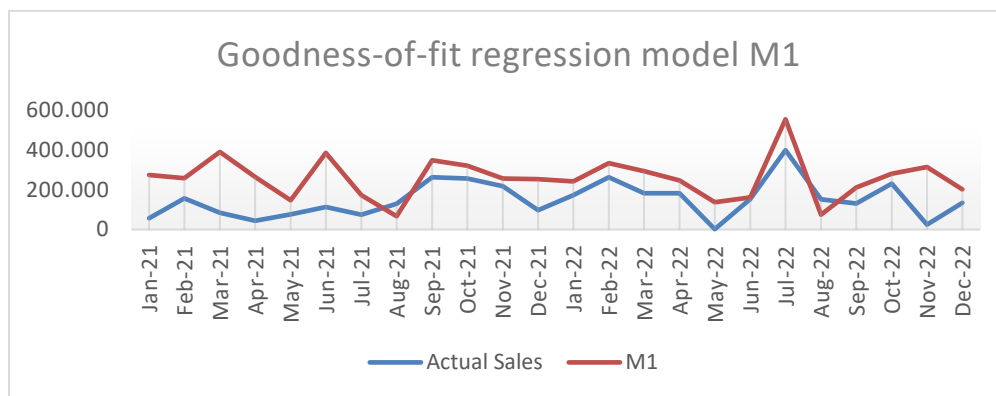


Figure 31, “Time plot of forecasted sales of product P₁”

Additionally, **Figure 32** presents the residuals' time chart of the applied M1 regression model in order to detect any possible errors' patterns.

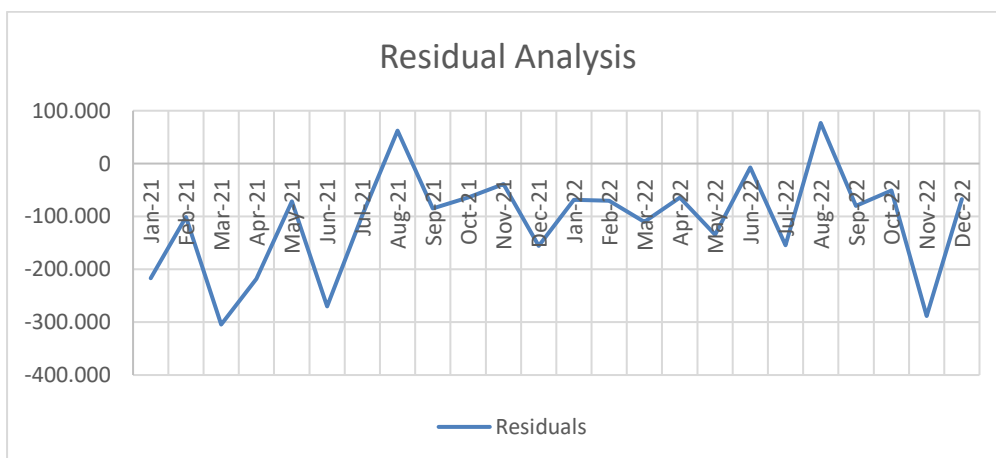


Figure 32, “Residual Analysis of product P₁”

Figure 31 shows that our model is partially a good fit to product's actual sales as it follows the sales fluctuations while it seems to continuously overestimate them. Furthermore, **Figure**

Postgraduate Dissertation

32 confirms this observation, as the residuals of the M1 regression model are mainly negative.

4.9.2 Forecasting MLR model M1 of product P_2

The following **Table 35** presents the summary output derived from implementing the M1 regression model on the sales data of the product's estimation sample.

TERM	COEFFICIENTS	STANDARD ERROR	T STAT	P- VALUE
INTERCEPT	14,5669	2,8674	5,0801	0,0000
LINEAR TREND	-0,0101	0,0438	-0,2308	0,8190
QUADRATIC TREND	-0,0001	0,0008	-0,1478	0,8835
LAG VARIABLE				
T-1	0,2033	0,1700	1,1961	0,2410
LAG VARIABLE				
T-2	-0,3806	0,1742	-2,1846	0,0369
M2	-0,2120	0,7162	-0,2960	0,7693
M3	-0,6346	0,6756	-0,9393	0,3551
M4	-1,3024	0,6692	-1,9462	0,0610
M5	-1,1633	0,6899	-1,6862	0,1021
M6	-0,9987	0,7121	-1,4025	0,1710
M7	-1,4953	0,6939	-2,1550	0,0393
M8	-2,1243	0,6821	-3,1144	0,0040
M9	-0,4622	0,7357	-0,6282	0,5346
M10	-0,8741	0,7535	-1,1600	0,2552
M11	-0,5660	0,6742	-0,8395	0,4078
M12	-0,6599	0,6703	-0,9845	0,3328

Table 35 “Summary output of M1 model implemented in product P_2 ”

The equation below is the MLR model that will be implemented in product's forecasting sample.

$$M1: \log(\widehat{S}_{t,2}) = 14.57 + 0.20 \log(S_{t-1,2}) - 0.38 \log(S_{t-2,2}) - 1.50M_7 - 2.12M_8 \quad (4.9.2.1)$$

After implementing the previous model (regression equation 4.9.2.1) in the product's forecasting sample, estimations about product's sales were derived. **Figure 33** visualizes the predicted sales against the actual sales during the performance period.

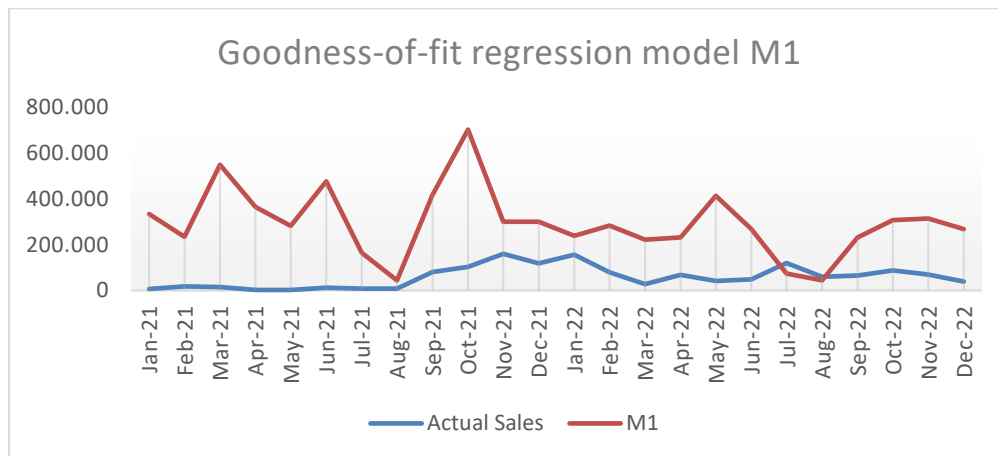


Figure 33, “Time plot of forecasted sales of product P_2 ”

Furthermore, **Figure 34** presents the residuals' time chart of the applied M1 regression model in order to detect any possible errors' patterns.

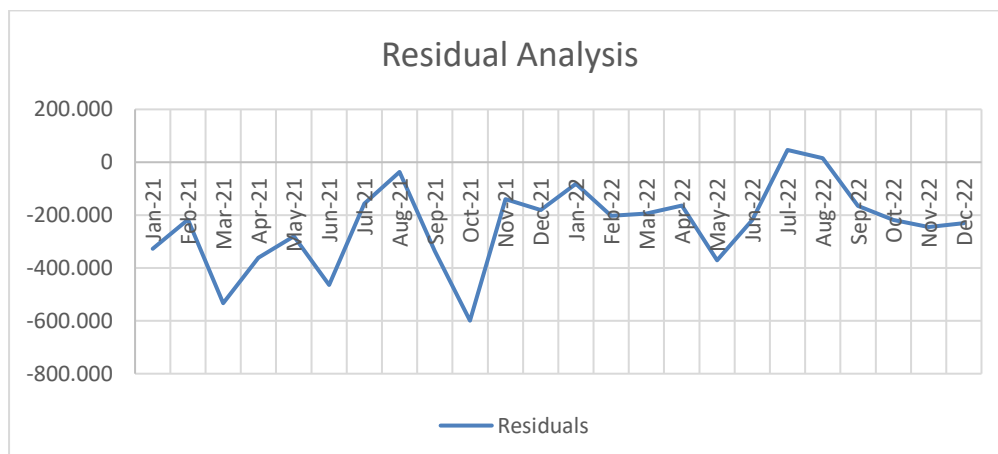


Figure 34, “Residual Analysis of product P_2 ”

Figure 33 shows that our model is not a good fit to product's actual sales as it continuously overestimates them, while in **Figure 34** the consistently negative errors' values of the forecasting model confirms that observation.

4.9.3 Forecasting MLR model M1 of product P_3

The following **Table 36** presents the summary output derived from implementing MLR model M1 in the sales data of the product's estimation sample. Following the same methodology with the previous products, the model's results will be utilized to enable the determination of the forecasting model.

TERM	COEFFICIENTS	STANDARD		
		ERROR	T STAT	P-VALUE
INTERCEPT	16.4923	3.2962	5.0035	0.0000
LINEAR TREND	-0.0283	0.0230	-1.2294	0.2285
QUADRATIC TREND	0.0006	0.0004	1.3347	0.1920
LAG VARIABLE				
T-1	-0.1141	0.1781	-0.6406	0.5266
LAG VARIABLE				
T-2	-0.2355	0.1782	-1.3219	0.1962
M2	0.1094	0.3753	0.2916	0.7726
M3	0.1693	0.3548	0.4773	0.6366
M4	-0.0818	0.3527	-0.2319	0.8182
M5	0.1209	0.3490	0.3463	0.7316
M6	-0.1858	0.3568	-0.5206	0.6065
M7	-0.3012	0.3508	-0.8587	0.3973
M8	-1.1118	0.3634	-3.0598	0.0046
M9	0.1353	0.4126	0.3280	0.7452
M10	-0.0757	0.4271	-0.1772	0.8606
M11	0.5120	0.3490	1.4668	0.1528
M12	0.1228	0.3590	0.3421	0.7347

Table 36 “Summary output of M1 model implemented in product P₃”

The equation below is the MLR model M1 that will be implemented in product’s forecasting sample.

$$M1: \log(\widehat{S_{t,3}}) = 16.49 - 0.11 \log(S_{t-1,3}) - 0.24 \log(S_{t-2,3}) - 1.11M_8 \quad (4.9.3.1)$$

Following the implementation of the preceding model (regression equation 4.9.3.1) on the product's forecasting sample, estimations regarding the product's sales were generated. **Figures 35** and **36** depict the comparison between predicted and actual sales, as well as the time chart of residuals, respectively. This visualization aims to offer the reader a thorough understanding of the model's predictive capacity and the evolution of residuals throughout the performance period, facilitating the assessment of the model's fit and the identification of potential error patterns.

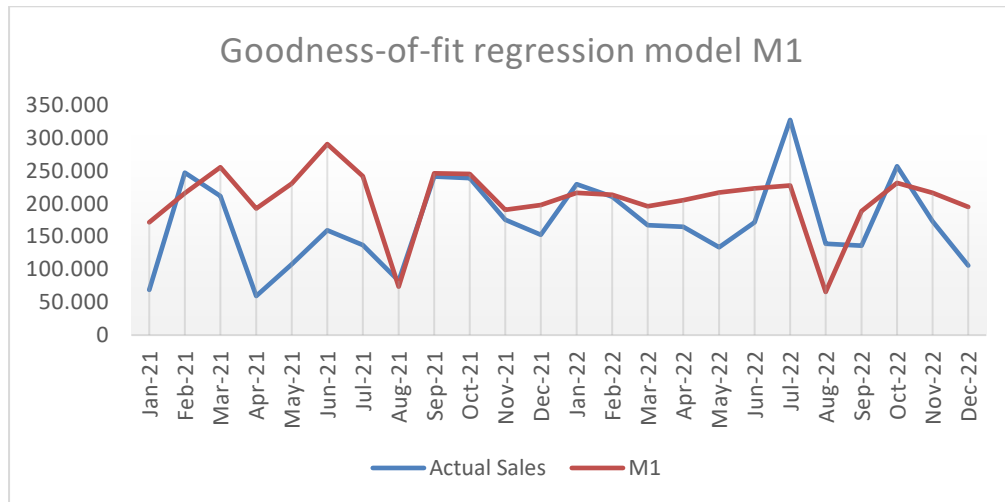


Figure 35, “Time plot of forecasted sales of product P₃”

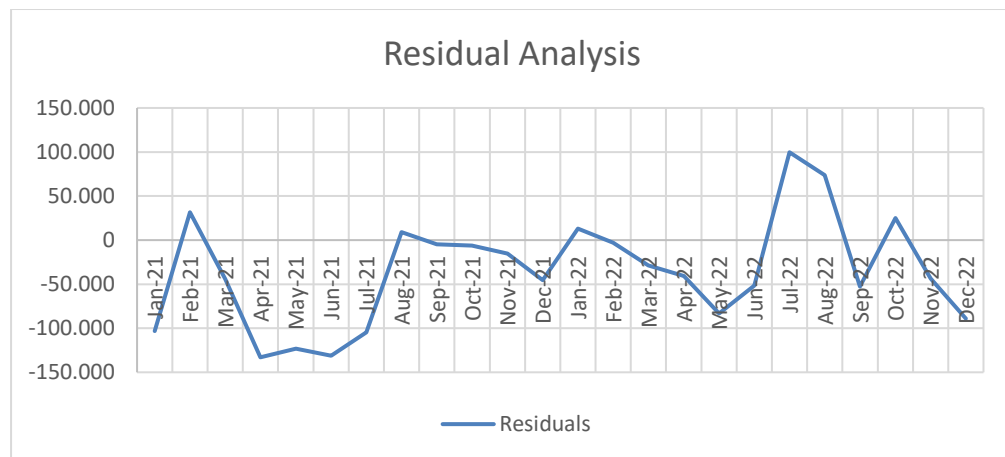


Figure 36, “Residual Analysis of product P₃”

The preceding graphs serve to assess the adequacy of the regression model in accurately forecasting the actual sales of the product. As observed, model M1 is partially a good fit to product’s actual sales, as it follows the sales fluctuations while it seems to, most of the times, overestimate them. Furthermore, **Figure 36** confirms this observation, as the residuals of the M1 regression model are mostly negative.

4.9.4 Forecasting MLR model M1 of product P₄

Table 37 below displays the summary output resulting from applying the regression model M1, to the product’s estimation sample. Following the same methodology with the previous products, the model’s results will be utilized to enable the determination of the forecasting model.

TERM	COEFFICIENTS	STANDARD ERROR	T STAT	P- VALUE
INTERCEPT	14.9154	2.9718	5.0190	0.0000
LINEAR TREND	-0.0224	0.0180	-1.2432	0.2234
QUADRATIC TREND	0.0004	0.0003	1.2554	0.2190
LAG VARIABLE T-1	-0.0455	0.1780	-0.2558	0.7998
LAG VARIABLE T-2	-0.2213	0.1762	-1.2561	0.2188
M2	-0.5390	0.2871	-1.8770	0.0703
M3	-0.5329	0.2771	-1.9231	0.0640
M4	-0.4473	0.2831	-1.5799	0.1246
M5	-0.6950	0.2770	-2.5091	0.0177
M6	-0.5058	0.2810	-1.7999	0.0819
M7	-0.7006	0.2833	-2.4733	0.0193
M8	-1.4461	0.2825	-5.1183	0.0000
M9	-0.4482	0.3507	-1.2781	0.2110
M10	-0.1397	0.3422	-0.4082	0.6860
M11	-0.3210	0.2757	-1.1641	0.2536
M12	-0.1032	0.2757	-0.3741	0.7109

Table 37 “Summary output of M1 model implemented in product P₄”

The equation below is the MLR model M1 that will be implemented to make estimations about product’s sales.

$$M1: \log(\widehat{S}_{t,4}) = 14.92 - 0.05 \log(S_{t-1,4}) - 0.22 \log(S_{t-2,4}) - 0.70M_5 - 0.70M_7 - 1.44M_8 \quad (4.9.4.1)$$

After applying equation 4.9.4.1 to the product's forecasting sample, predictions for the product's sales were derived. **Figures 37** and **38** showcase the comparison between forecasted and actual sales, alongside the time chart illustrating the residuals. As was explained previously, the visual representation aims to provide a comprehensive view of the model's predictive performance and the behavior of its residuals over the forecasting period.

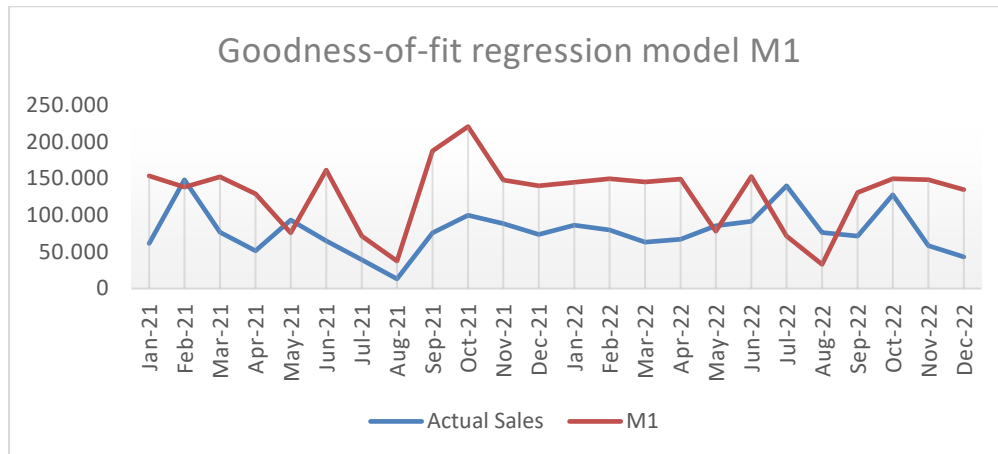


Figure 37, “Time plot of forecasted sales of product P₄”

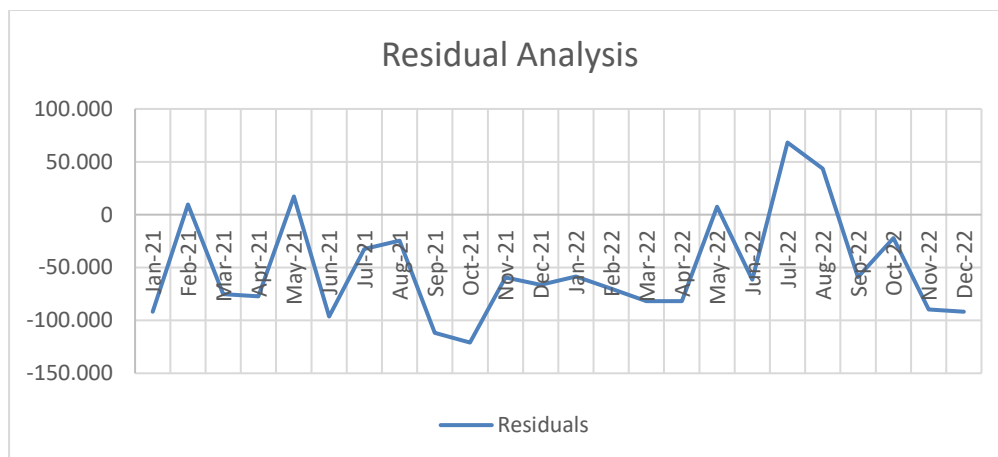


Figure 38, “Residual Analysis of product P₄”

The preceding graphs indicate that the MLR model M1 does not offer a good fit to the actual sales of the product. Moreover, the errors resulting from the model's application do not display any recognizable pattern but mostly are negative.

4.9.5 Forecasting MLR model M1 of product P₅

Table 38 displays the summary output resulting from applying the regression model M1 to the product's estimation sample.

TERM	COEFFICIENTS	STANDARD ERROR	T STAT	P- VALUE
INTERCEPT	13.3696	2.0761	6.4398	0.0000
LINEAR TREND	-0.0302	0.0365	-0.8272	0.4146
QUADRATIC TREND	0.0006	0.0007	0.8090	0.4249

LAG VARIABLE				
T-1	-0.0572	0.1649	-0.3470	0.7310
LAG VARIABLE				
T-2	-0.4132	0.1648	-2.5065	0.0178
M2	0.8426	0.6159	1.3679	0.1815
M3	3.4718	0.6224	5.5780	0.0000
M4	2.2944	0.8656	2.6506	0.0127
M5	3.9172	0.9732	4.0251	0.0004
M6	2.8003	0.8543	3.2778	0.0026
M7	2.9980	0.8787	3.4121	0.0019
M8	2.1176	0.8210	2.5795	0.0150
M9	0.8246	0.7752	1.0637	0.2959
M10	-0.0927	0.6701	-0.1384	0.8908
M11	-0.9161	0.5857	-1.5641	0.1283
M12	-0.5810	0.5748	-1.0107	0.3202

Table 38 “Summary output of M1 model implemented in product P₅”

Considering the regression results, the following equation is the regression model M1 that will be implemented in product’s forecasting sample.

$$M1: \log(\widehat{S}_{t,5}) = 13.37 - 0.06 \log(S_{t-1,5}) - 0.41 \log(S_{t-2,5}) + 3.47M_3 + 2.29M_4 + 3.92M_5 + 2.80M_6 + 2.99M_7 + 2.12M_8 \quad (4.9.5.1)$$

After applying the previous model (regression equation 4.9.5.1) to the product's forecasting sample, the predictions for the product's sales and the model’s errors were computed and are presented in **Figures 39** and **40**.

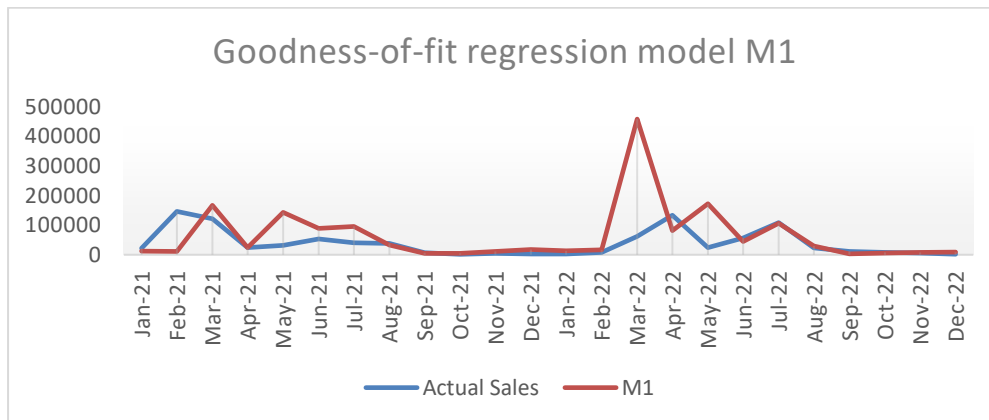


Figure 39, “Time plot of forecasted sales of product P₅”

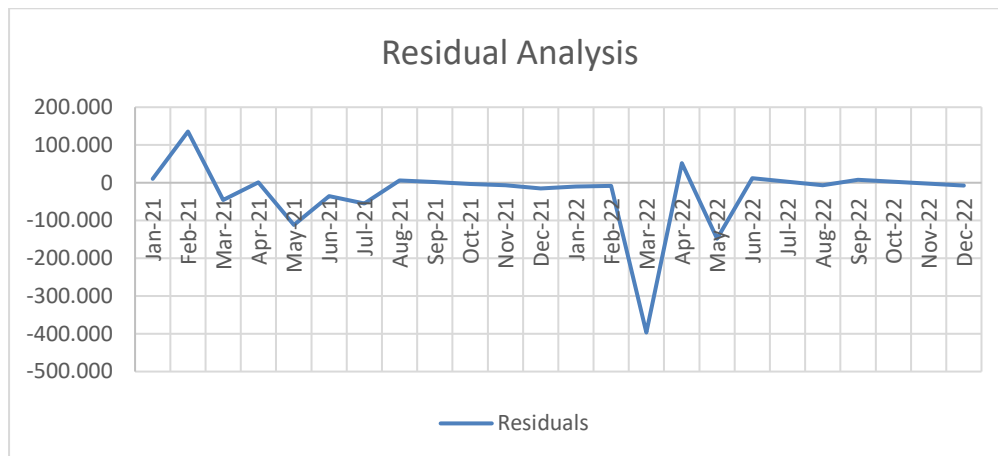


Figure 40, “Residual Analysis of product P₅”

Figure 39 depicts that the forecasting model M1 closely aligns with actual sales, while **Figure 40** illustrates that the residuals primarily approach the zero line across the model's performance duration, apart from some exceptions. The proximity of residuals to zero signifies that, on average, the model's predictions correspond closely to the observed values. This implies that the model captures a significant portion of the data's variance, and the errors (residuals) are relatively minor and random. Furthermore, consistent closeness to zero in residuals suggests that the assumptions of the regression model, such as linearity, homoscedasticity, and normality of residuals, are likely to have been met.

4.9.6 Forecasting MLR model M1 of product P₆

The next product that will be studied is **P₆**. **Table 39** includes the results of implementing the model M1 (equation 4.9.6.1) in product's estimation sample.

TERM	COEFFICIENTS	STANDARD		
		ERROR	T STAT	P-VALUE
INTERCEPT	13.7704	2.8123	4.8966	0.0000
LINEAR TREND	0.0286	0.0115	2.4974	0.0182
QUADRATIC TREND	-0.0003	0.0002	-1.2731	0.2127
LAG VARIABLE				
T-1	-0.3117	0.1859	-1.6772	0.1039
LAG VARIABLE				
T-2	-0.1713	0.2240	-0.7648	0.4503
M2	-0.0496	0.1793	-0.2767	0.7839
M3	0.1173	0.1598	0.7340	0.4687
M4	0.0968	0.1668	0.5802	0.5661

M5	-0.1232	0.1642	-0.7504	0.4589
M6	0.0418	0.1548	0.2697	0.7892
M7	0.1306	0.1700	0.7683	0.4483
M8	0.0205	0.1680	0.1221	0.9037
M9	-0.0161	0.1595	-0.1008	0.9204
M10	0.0871	0.1598	0.5450	0.5898
M11	-0.1135	0.1669	-0.6799	0.5018
M12	-0.1274	0.1564	-0.8150	0.4215

Table 39 “Summary output of M1 model implemented in product P₆”

$$M1: \log(\widehat{S}_{t,6}) = 13.77 + 0.03t - 0.31 \log(S_{t-1,6}) - 0.17 \log(S_{t-2,6}) \quad (4.9.6.1)$$

This equation will be applied to the forecasting sample data to generate estimations and assess the model's fit. **Figures 41** and **42** illustrate the predictions derived from implementing model M1 compared with actual sales data and the corresponding errors, respectively.

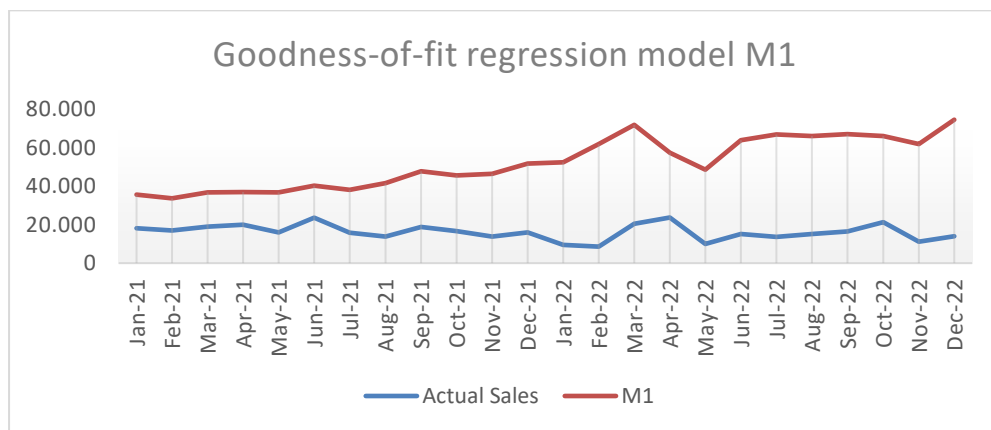


Figure 41, “Time plot of forecasted sales of product P₆”

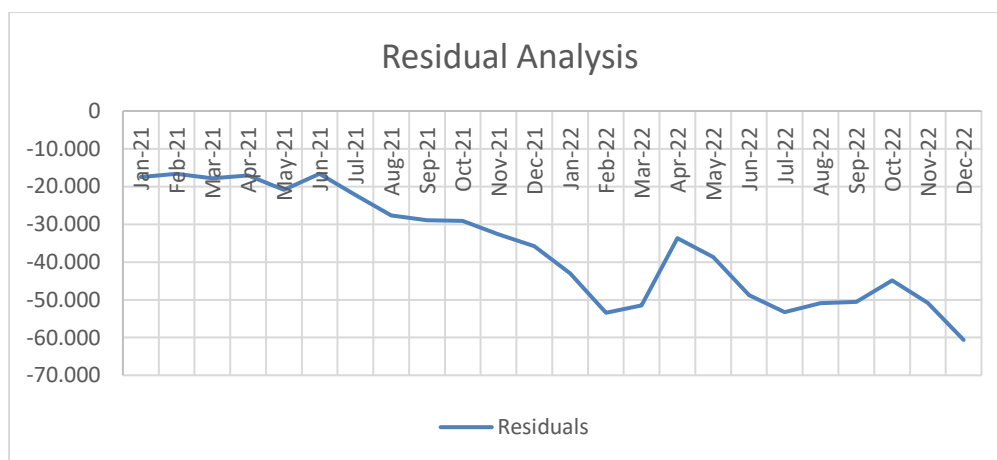


Figure 42, “Residual Analysis of product P₆”

Both graphs illustrate a consistent overestimation of actual sales data by the model. Throughout the model's performance period, the errors consistently maintain negative values with a downward trend, indicating to the analyst that the model is not adequately capturing the underlying structure of the data, and further investigation is warranted to improve the model's fit. This could involve exploring alternative model specifications, identifying and addressing influential data points, or considering additional predictor variables.

4.9.7 Forecasting MLR model M1 of product P_7

The concluding examination centers on product P_7 , with **Table 40** presenting the summarized outcomes of the M1 model. Subsequently, the final form of the model, to be applied in the product's forecasting sample, will be presented.

TERM	COEFFICIENTS	STANDARD ERROR	T STAT	P- VALUE
INTERCEPT	16.4124	2.7150	6.0450	0.0000
LINEAR TREND	0.0102	0.0101	1.0166	0.3175
QUADRATIC TREND	0.0000	0.0002	-0.2545	0.8008
LAG VARIABLE T-1	-0.2012	0.1739	-1.1569	0.2564
LAG VARIABLE T-2	-0.3296	0.1739	-1.8952	0.0677
M2	-0.1919	0.2077	-0.9240	0.3629
M3	-0.1611	0.2041	-0.7892	0.4362
M4	-0.4686	0.1705	-2.7490	0.0100
M5	-0.2748	0.1545	-1.7781	0.0855
M6	-0.2667	0.1681	-1.5868	0.1230
M7	-0.3019	0.1750	-1.7252	0.0948
M8	-0.3619	0.1661	-2.1788	0.0373
M9	-0.0901	0.1595	-0.5650	0.5762
M10	-0.2972	0.1838	-1.6173	0.1163
M11	-0.3571	0.1774	-2.0127	0.0532
M12	-0.6374	0.1585	-4.0228	0.0004

Table 40 “Summary output of M1 model implemented in product P_7 ”

$$M1: \log(\widehat{S}_{t,7}) = 16.41 - 0.20 \log(S_{t-1,7}) - 0.33 \log(S_{t-2,7}) - 0.47M_4 - 0.36M_8 - 0.36M_{11} - 0.64M_{12} \quad (4.9.7.1)$$

After applying the aforementioned equation to the product's forecasting sample, **Figures 43** and **44** showcase the resulting graphs. These demonstrate that the MLR model M1 partially

aligns with the actual sales data, showing no consistent trend of overestimation or underestimation. Additionally, the model's errors fluctuate across the performance period without displaying any discernible pattern.

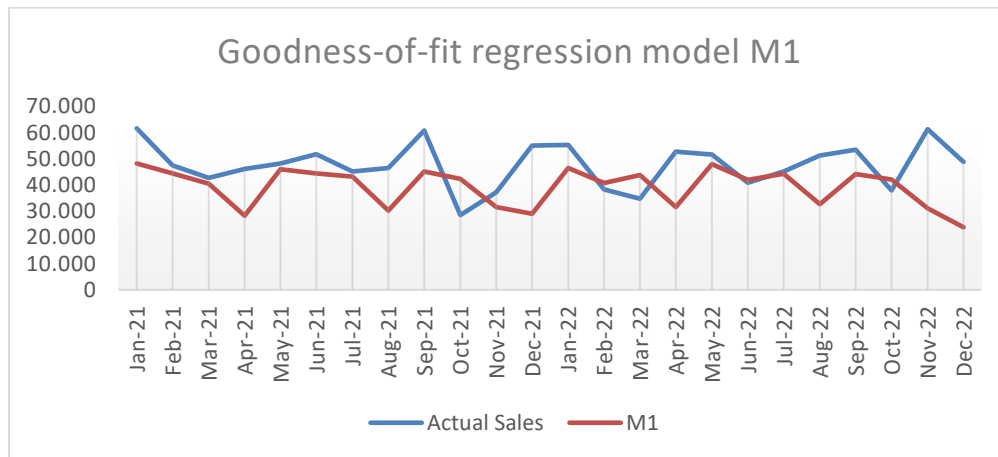


Figure 43, “Time plot of forecasted sales of product P₇”

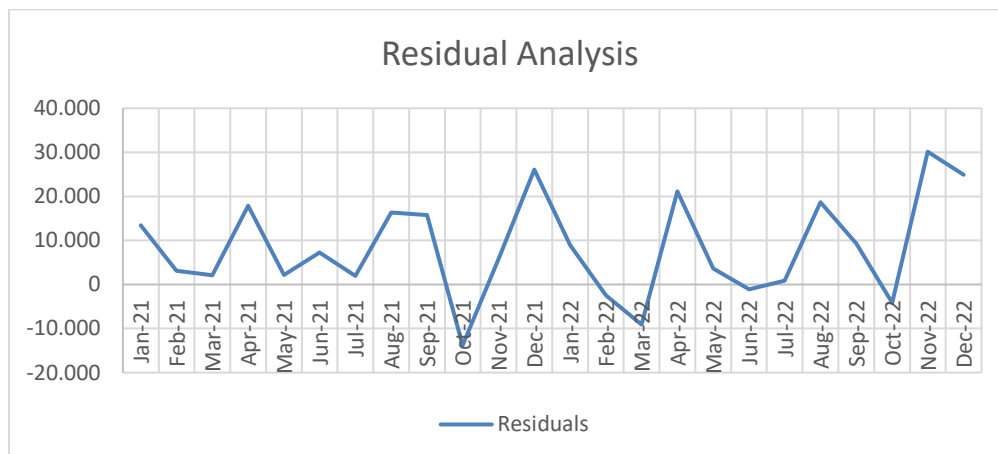


Figure 44, “Residual Analysis of product P₇”

4.10 EWMA model implementation

The subsequent sections will introduce the EWMA model for each product, leveraging Excel's solver software to identify the optimal value for the λ parameter that minimizes the MAPE index. Subsequently, graphs will be provided to showcase the EWMA outcomes alongside the previous MLR M1 models for each product. These comparisons aim to facilitate meaningful assessments, aiding in the determination of the most suitable forecasting technique.

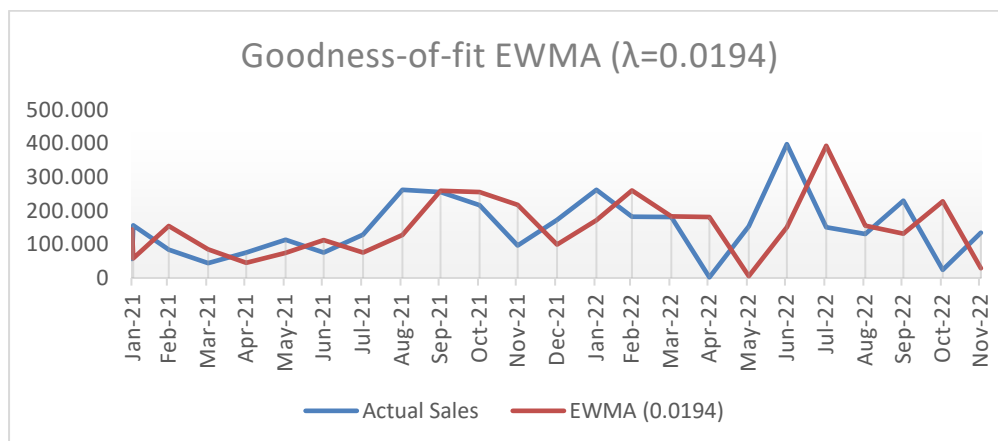
Same methodology will be applied across all products. The sales data will be split into estimation and forecasting samples as previously in chapter 4.9 explained. The EWMA model outlined in equation (3.2.1) will initially use a random smoothing factor λ , and then, utilizing Excel’s Solver Tool, the optimal λ value will be determined with the objective of minimizing MAPE value in the estimation sample. Afterwards, the final EWMA model, created with this optimal λ , will be employed for generating predictions in each product’s forecasting sample. Ultimately, the forecasting outcomes of this method will be compared graphically with those produced with the previous MLR method in order to make comparisons.

4.10.1 EWMA for product P_1

The following equation depicts the EWMA model and the corresponding optimal λ parameter’s value, calculated via Excel’s Solver tool, for product P_1 .

$$EWMA(\lambda) : \hat{Y}_{t+1} = (1 - 0.0194)Y_t + 0.0194\hat{Y}_t$$

The following **Figure 45** illustrates the graphs that represent the goodness-of-fit for the EWMA model, the fitting of both forecasting methods applied in product’s forecasting sample (MLR M1 and EWMA ($\lambda=0.0194$)) and the models’ derived errors. The purpose of such graphical presentation, is to summarize the excluded results of the forecasting techniques’ application, to make the appropriate models’ comparisons.



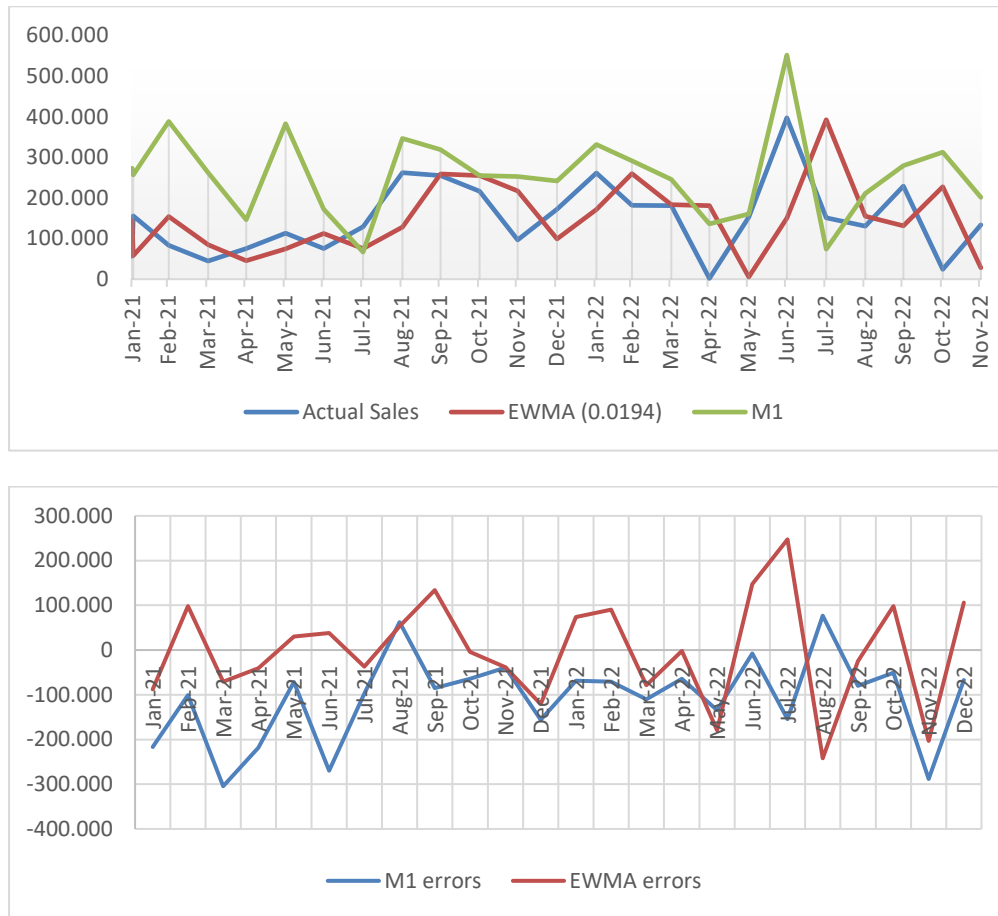


Figure 45, “Graphical presentation of P_1 product's forecasting models”

The previous graphs offer significant insights into the models' performance and residuals. The computed low value of the λ smoothing parameter in the EWMA forecasting method signifies the model's rapid adjustment nature, portraying it as volatile and erratic. This model showcases fast adaptation to abrupt changes within the time series data, as detailed in chapter 3. It prioritizes recent observations over older ones within the product's data series. Moreover, the error graph suggests that both models yield relatively similar forecasting errors, without any notable differences.

4.10.2 EWMA for product P_2

The following equation presents the EWMA model and the corresponding optimal λ parameter's value, calculated via Excel's Solver tool, for product P_2 .

$$EWMA(\lambda) : \hat{Y}_{t+1} = (1 - 0.0105)Y_t + 0.0105\hat{Y}_t$$

As previously with product **P₁**, the **Figure 46** includes the graphs that correspond to the examined forecasting methodologies implemented for product **P₂**.

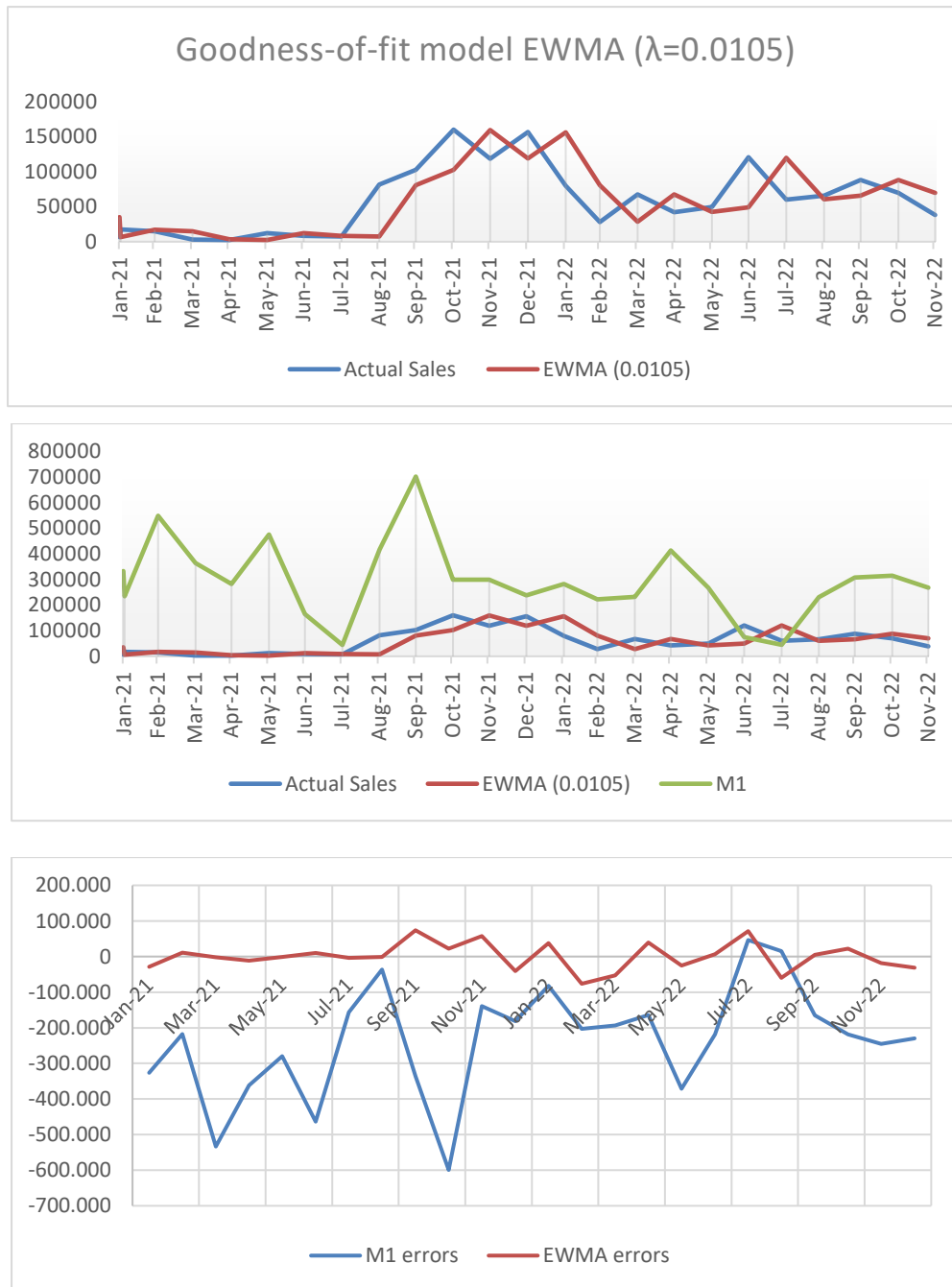


Figure 46, “Graphical presentation of P₂ product's forecasting models”

As evident from the preceding graphs, the calculation of a low optimal value for the smoothing parameter via the Excel's Solver tool, leads to the model swiftly adapting to abrupt fluctuations in sales. Consequently, this results in a shorter memory span and more abrupt estimations, effectively capturing the short-term variations within the sales time series.

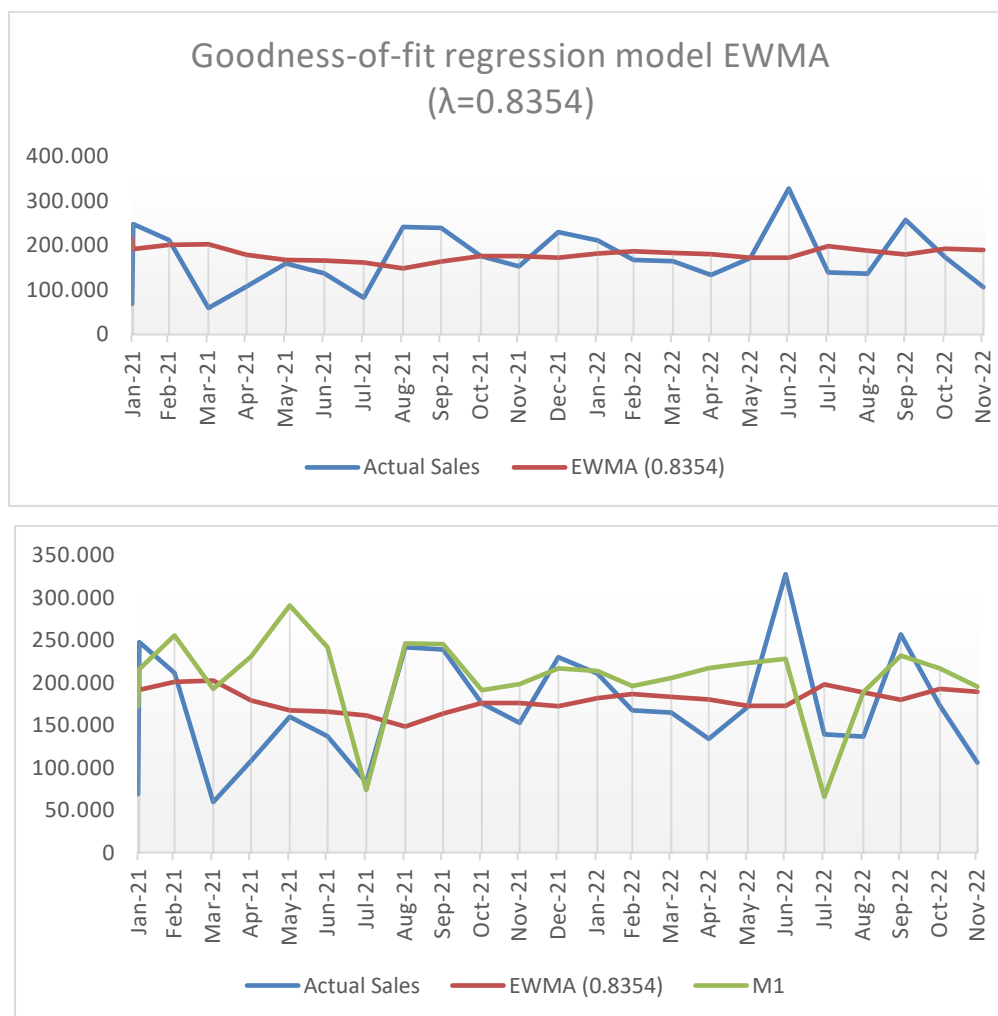
Additionally, the errors generated by the EWMA model are continuously close to the zero line, while the error outcomes of the M1 model remain negative throughout the forecasting period.

4.10.3 EWMA for product **P₃**

The following equation presents the EWMA model and the corresponding optimal λ parameter's value, calculated via Excel's Solver tool, for product **P₃**.

$$EWMA(\lambda) : \hat{Y}_{t+1} = (1 - 0.8354)Y_t + 0.8354\hat{Y}_t$$

The next **Figure 47** includes the graphs that correspond to the examined forecasting methodologies implemented for product **P₃**.



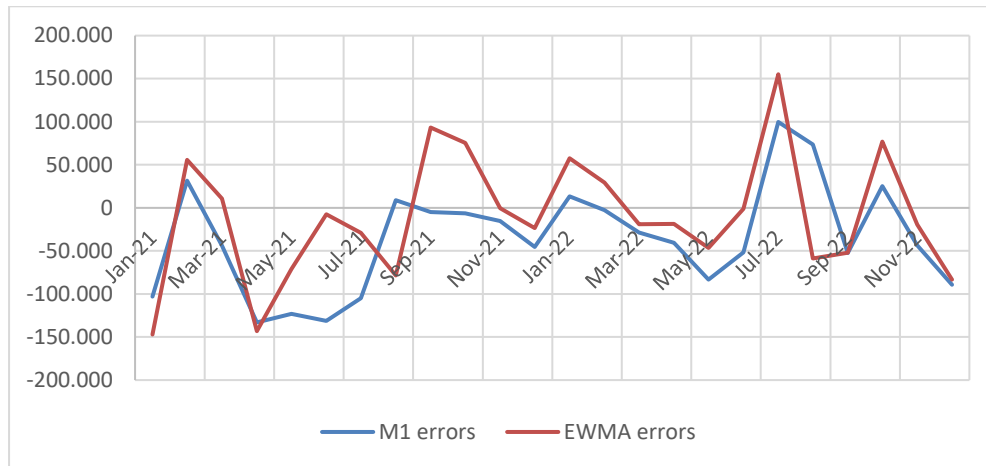


Figure 47, “Graphical presentation of P_3 product's forecasting models”

The graphs illustrate the inability of EWMA model to capture the sales variations, due to the high value of the smoothing parameter calculated by the Solver tool. This parameter leads to smoother sales estimations that overlook short-term sales fluctuations observed in the performance period. Furthermore, the errors in the M1 model present similar results compared to EWMA errors. This suggests that none of the models seem to be superior to the other.

4.10.4 EWMA for product P_4

The following equation presents the EWMA model and the corresponding optimal λ parameter's value, calculated via Excel's Solver tool, for product P_4 .

$$EWMA(\lambda) : \hat{Y}_{t+1} = (1 - 0.9998)Y_t + 0.9998\hat{Y}_t$$

The graphs below depicted in **Figure 48**, correspond to the forecasting methodologies implemented for product P_4 . It is observed that the forecasting results produced by the EWMA model do not capture the sales variations observed in the performance period. Furthermore, the errors in the M1 model remain negative in almost the entire performance period, indicating a continuous overestimation of the actual sales data by the model.

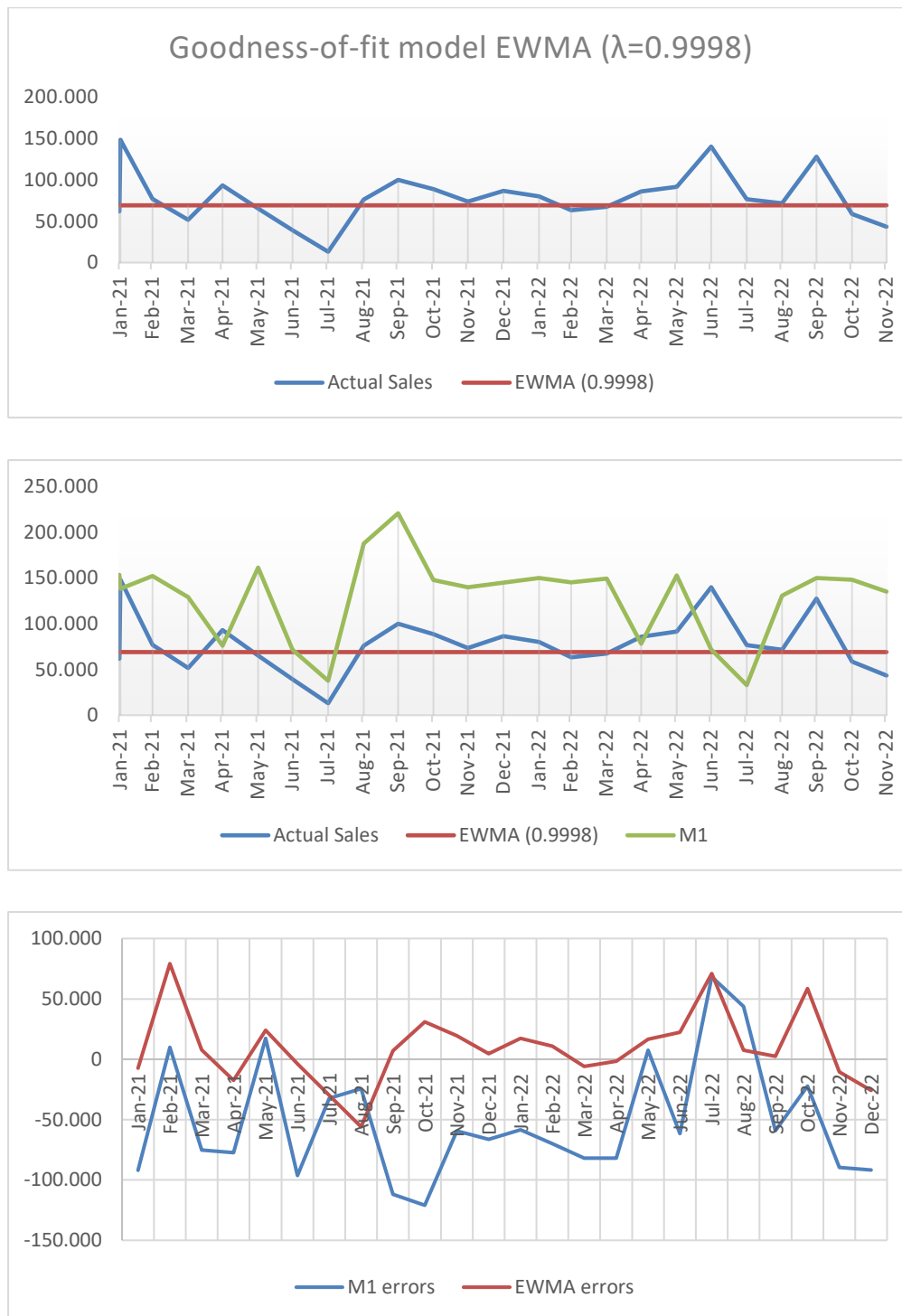


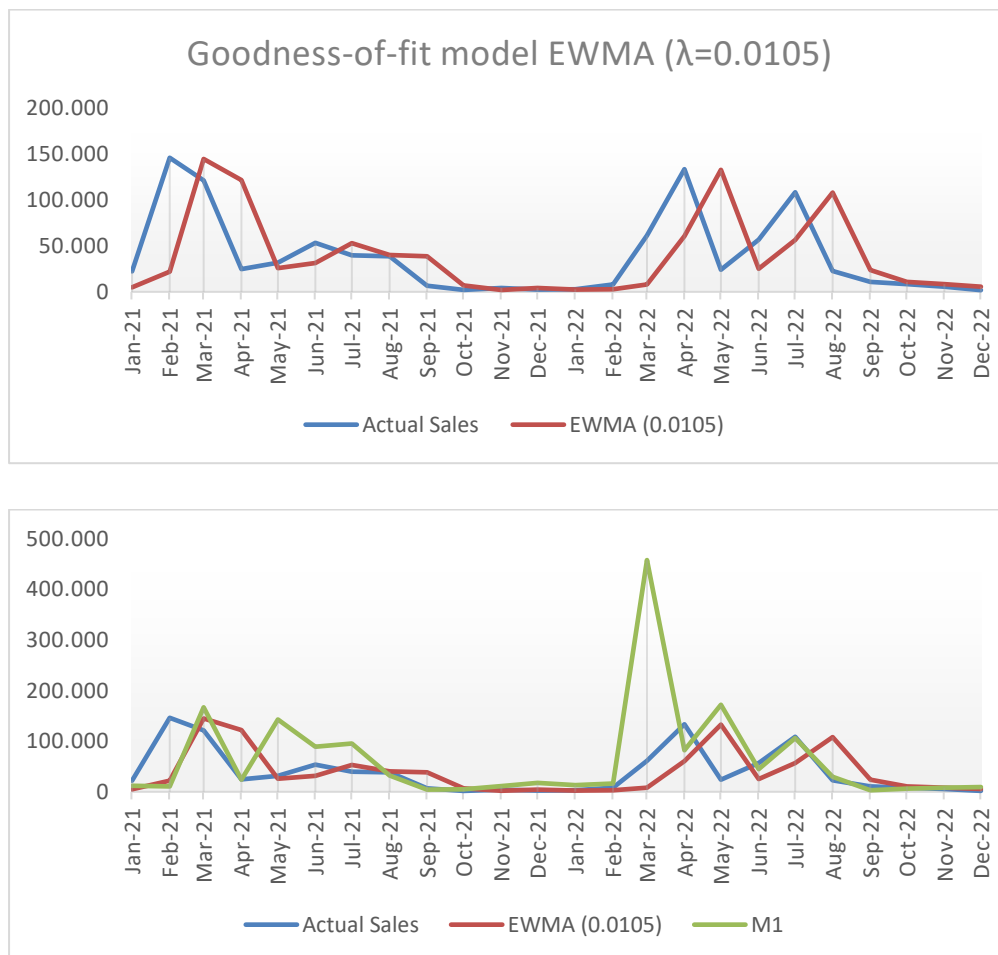
Figure 48, “Graphical presentation of P_4 product's forecasting models”

4.10.5 EWMA for product P_5

The following equation presents the EWMA model and the corresponding optimal λ parameter's value, calculated via Excel's Solver tool, for product P_5 .

$$EWMA(\lambda) : \hat{Y}_{t+1} = (1 - 0.0105)Y_t + 0.01059\hat{Y}_t$$

The graphical representations presented in **Figure 49** pertain to the applied forecasting methodologies utilized for product **P₅**. Notably, the forecasting outcomes generated by the EWMA model successfully encapsulate the fluctuations in sales observed during the specified performance period. As elucidated in chapter 3, the model's utilization of a low parameter value primarily emphasizes recent observations, thereby registering fluctuations occurring within the forecasting timeframe from January 2021 to December 2022. Additionally, the errors produced by the M1 model closely resemble those observed in the EWMA model. This observation leads to the conclusion that both models exhibit analogous behavior, capturing both ascending and descending sales trends.



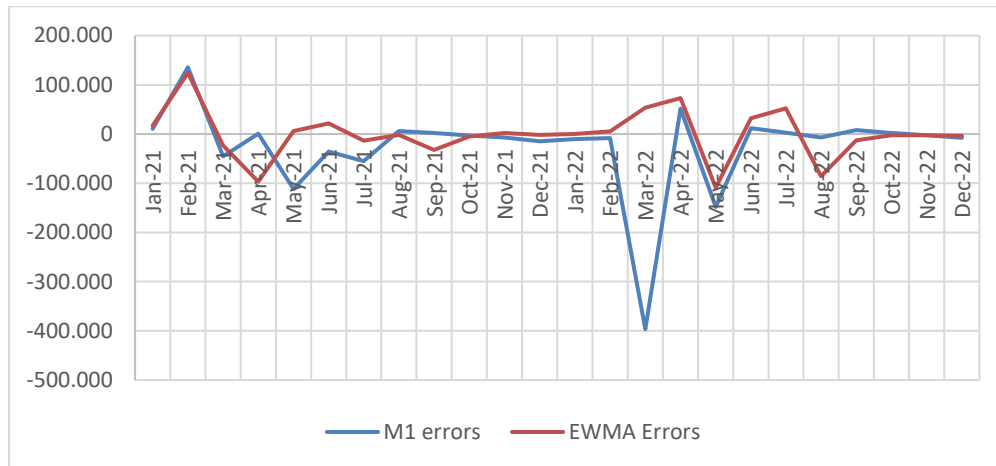


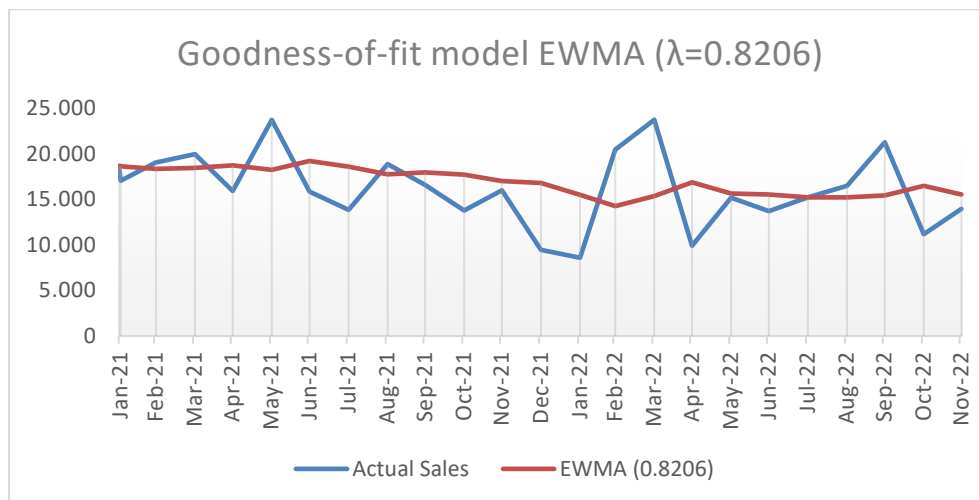
Figure 49, “Graphical presentation of P₅ product's forecasting models”

4.10.6 EWMA for product P₆

The next product that will be examined for the EWMA model performance, is product P₆. The following equation presents the EWMA model and the corresponding optimal λ parameter's value, calculated via Excel's Solver tool.

$$EWMA(\lambda) : \hat{Y}_{t+1} = (1 - 0.8206)Y_t + 0.8206\hat{Y}_t$$

The graphs presented in **Figure 50**, correspond to the forecasting methodologies implemented for product P₆.



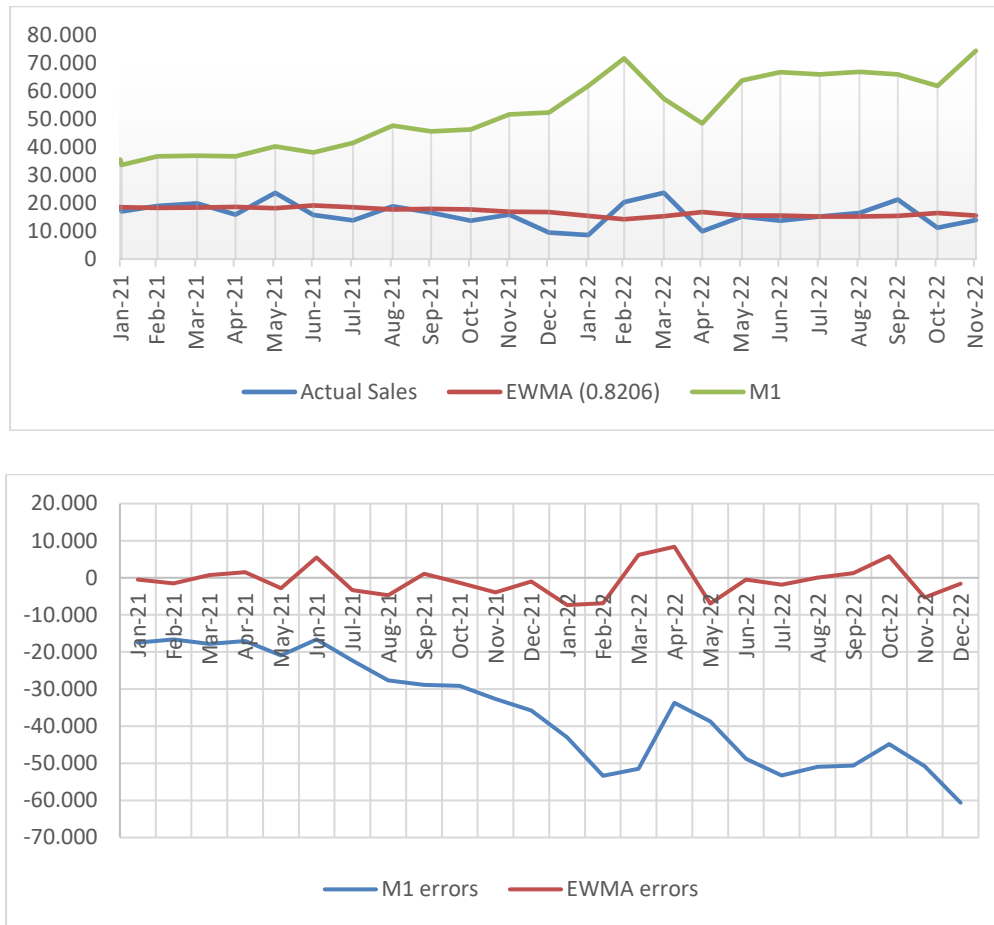


Figure 50, “Graphical presentation of P_6 product's forecasting models”

The previous graphs present some interesting information. Specifically, the EWMA model for product P_6 does not seem to capture sales volatility effectively. The high value of the smoothing parameter contributes to favoring older observations. Nevertheless, the EWMA model's errors remain close to the zero line throughout the performance period. On the other hand, the M1 model's errors consistently maintain negative values with a downward trend, indicating to the analyst that the model is not adequately capturing the underlying structure of the data, and further investigation is warranted to improve the M1 model's fit.

4.10.7 EWMA for product P_7

The last product that will be under examination for the EWMA model's implementation results and performance, is product P_7 . The following equation presents the EWMA model and the corresponding optimal λ parameter's value, calculated via Excel's Solver tool.

$$EWMA(\lambda) : \hat{Y}_{t+1} = (1 - 0.7895)Y_t + 0.7895\hat{Y}_t$$

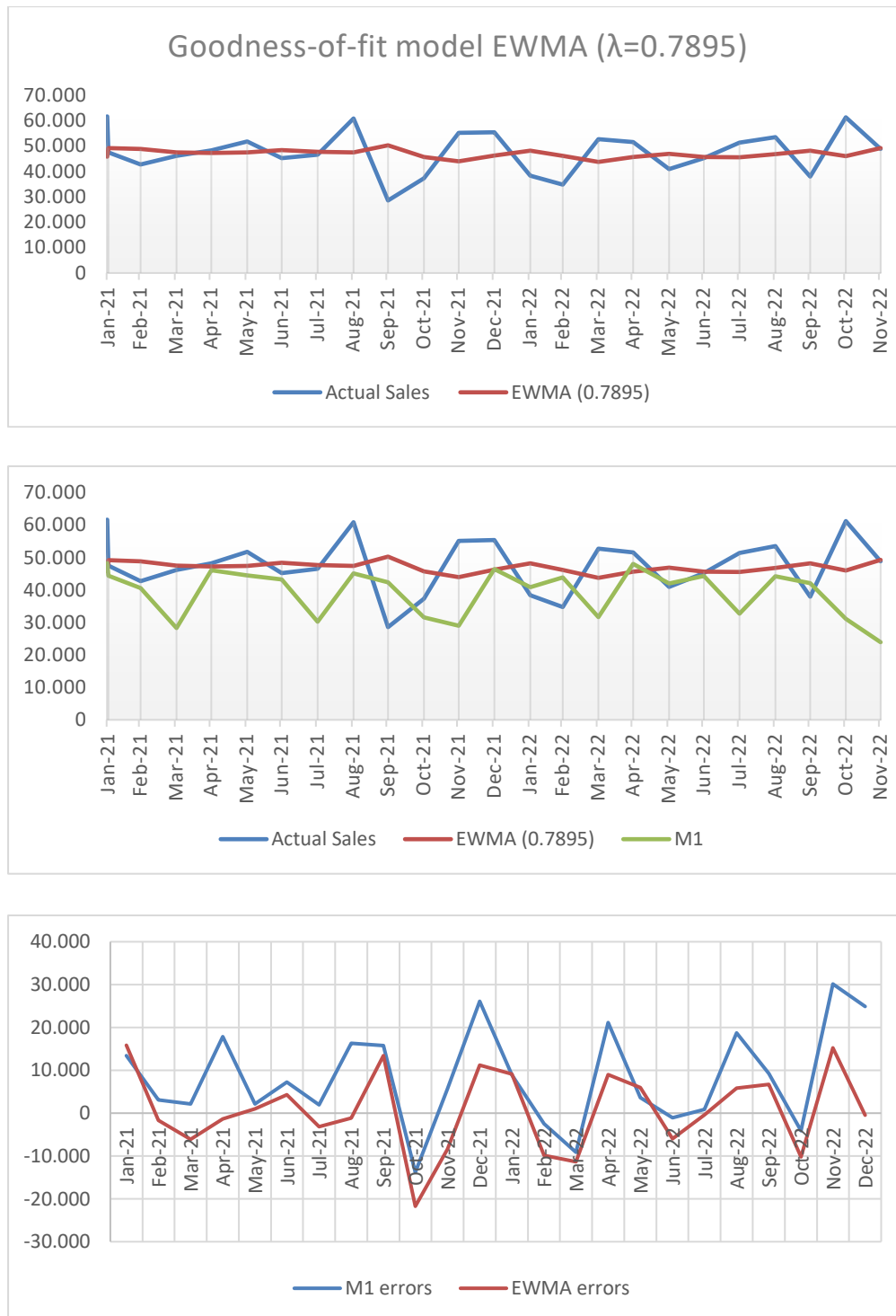


Figure 51, “Graphical presentation of P₇ product's forecasting models”

As evidenced from the preceding illustration of the forecasting models' results, depicted in **Figure 51**, it is apparent that the EWMA model inadequately accounts for the fluctuations in sales, whereas the M1 model demonstrates greater suitability in this regard. However, the forecasting errors generated by the implementation of both models do not exhibit significant

differentiations, suggesting that neither model consistently underestimates nor overestimates the actual sales data.

4.11 Forecasting performance and accuracy

Upon the conclusion of the study analysis and the aforementioned presentation of research findings for each of the seven products, **Tables 41** and **42** summarize the MAE and MAPE outcomes derived from the EWMA and MLR models respectively, that was applied to the forecasting samples. In accordance with Lewis (1982) and Moreno et al., (2013), the MAPE index values undergo classification to assess the forecasting performance, aligning with predetermined thresholds that delineate the forecasting accuracy of the respective methodologies.

- MAPE < 10%, highly accurate forecasting
- 10% < MAPE < 20%, good forecasting
- 20% < MAPE < 50%, reasonable forecasting
- MAPE > 50%, inaccurate forecasting

Products	EWMA (λ)	MAE	MAPE
P ₁	0.0194	93,568	466.18%
P ₂	0.0105	29,584	82.40%
P ₃	0.8354	56,367	45.77%
P ₄	0.9998	22,419	41.21%
P ₅	0.0105	32,392	129.28%
P ₆	0.8206	3,337	23.84%
P ₇	0.7895	7,478	16.79%

Table 41 “MAE and MAPE results for optimal smoothing parameter λ of EWMA model”

Products	R ²	MAE	MAPE
P ₁	67.92%	90,345	439.64%
P ₂	48.89%	241,080	1813.72%
P ₃	53.17%	56,477	46.48%
P ₄	63.50%	63,286	96.51%
P ₅	83.32%	45,161	169.95%
P ₆	53.86%	35,953	250.72%
P ₇	61.57%	10,830	22.16%

Table 42 “R-square, MAE and MAPE results for MLR model M1”

The compiled forecasting accuracy results outlined in the preceding tables offer numerous insights. Particularly, the notably elevated MAPE values evident in products **P₁**, **P₂** and **P₅** suggest that both forecasting models fail to yield precise forecasts. Such exceptionally high MAPE values may signify several underlying issues, including the limited forecasting capability of the models and their inability to effectively capture inherent patterns, trends, and relationships within the data series. Additionally, the potential presence of outliers within the dataset could drive to these inappropriate MAPE outcomes.

Concerning product **P₃**, it can be observed that both forecasting models yield similar MAPE results, generating reasonable forecasts. This suggests that both techniques could be employed to make future estimates, although the EWMA model appears to be the preferred choice. Similarly, for product **P₄**, the EWMA model is favored due to its notably superior MAPE value.

Conversely, for products **P₆** and **P₇**, it is evident that the EWMA model outperforms the MLR model M1, demonstrating its ability to generate good and reasonable forecasts respectively. However, for product **P₇**, both forecasting models produce good MAPE results, suggesting that they could be applied equivalently.

5. Main Findings

Following the completion of the research methodology outlined in the preceding chapter 4, the present chapter is dedicated to synthesizing the outcomes and answering the research questions that were set in subchapter 1.1. The study's conclusions pertaining to the case of “Haleon Plc” in Greece, and, its health care consumer products under examination will be thoroughly presented providing valuable information for variable reasons. Such insights could lead to the appropriate managing of undesired increased inventories which commit a significant portion of the companies' capital, the limitation of unpredicted stock shortages, thereby allowing well-informed planning and decision-making procedures. Each research question is listed subsequently followed by the corresponding response.

Which are the main patterns and the statistical properties of the products' sales time series? Which of these statistical properties are more decisive in shaping the course of sales over time? Are there specific product codes with stronger seasonal signature?

In the initial phase of products presentation, fundamental visual representations of their sales time series were introduced, employing time series graphs, histograms, and monthly

boxplots. Following this, key statistical properties were computed and exhibited using Excel's descriptive statistics tool. Subsequent investigation into trend indications and the probabilities of seasonal effects was conducted through the application of MLR analysis models. Synthesizing the insights garnered from visual observations and regression analysis results allowed for the derivation of more substantiated conclusions. Furthermore, the expansion of MLR models through the incorporation of time lag variables facilitated the identification of short-term memory characteristics within the time series of the products.

Beginning with product P_1 , the sales time series do not exhibit linear or quadratic trends, without important seasonality or short-memory effects. Although some sporadic observations could be considered as extreme values, the sales of the product demonstrate a consistent behavior that tends to fluctuate around its mean. In case of product P_2 , a linear trend pattern in its sales time series exists, characterized by a continuous decline in sales volume over consecutive years, but this trend disappears when including lag variables in the regression model. Furthermore, any impact of seasonality seems minimal, while short-memory effect is evident. Similar to the properties observed in the first product, the third product P_3 , does not also exhibit time trends, important seasonality and short-memory effect in its sales across years. The same attributes are observed in product P_4 , showcasing no substantial differences when compared to the preceding product P_3 . This observation might be pertinent to the fact that both products belong to the same category (pain relief/analgesics) within the company's product range. Notably, for both products, the presence of seasonality impact on sales volume seems noticeable solely in the month August.

Concerning product P_5 , its sales exhibit no time trend behavior over the years without a discernible short-memory effect. However, a notable seasonality is evident within this product's sales time series, particularly during the Spring and Summer months. This seasonal effect is likely attributed to the nature of product (skin irritation/skin health) resulting in heightened demand during these seasons.

Concluding with the last two products P_6 and P_7 , despite the fact that both products belong to the same category within the company's range (oral health/oral care), they demonstrate different time series features. For product P_6 , the sales data series exhibit linear and quadratic trends. Specifically, sales display an ascending trend until the year 2019, followed by a period until 2022 marked by a slight decline in the product's sales volume. Neither seasonality or

short-memory features are significant in product's data set. Finally, the sales time series of product **P₇** showcase linear trend driving to a slight upward trend over years. The short-memory effect is apparent. Seasonality impact is evident on certain months throughout the year, although these months are not consecutive but rather sporadic.

Do sales of different product types exhibit interdependent behavior?

The primary aim of this research question was to delve into and unveil substantial insights that could aid the company in understanding sales interactions and consumption patterns within the same market area. This knowledge is pivotal in preparing for the production planning phase, ensuring alignment with observed sales dynamics and consumer behavior. To ascertain potential interrelationships among the studied products' codes, the MLR model M2 was deployed. The investigation has revealed the existence of cross-dependencies among the products, albeit not uniformly across all of them.

Notably, the sales of product **P₁** do not seem to be affected by the remaining products' sales. However, in the case of **P₂**, cross-dependencies are evident with **P₁**, **P₃**, and **P₇**, signifying associations between their sales demand. Specifically, under the condition of all other factors remaining constant, a potential escalation in **P₁** sales, corresponds to a projected 31.1% upsurge in **P₂** sales, measured in thousands of items. Likewise, a potential increase in **P₃** sales is estimated to yield an almost 95% increment in **P₂** sales, while an increase in **P₇** sales is associated with an 83.2% proportional increase in **P₂** sales. Notably, the significant coefficients' values within the model are represented as percentages, owing to the utilization of natural logarithms in the sales observations during the implementation of the model.

Continuing the research and focusing on the analysis of product **P₃** in relation to other products' sales, it was found that only **P₂** and **P₄** exhibit noteworthy cross-dependencies with this product. Specifically, a potential increase in sales of **P₂** and **P₄** corresponds to an increase of 14.3% and 34.7%, respectively, in the sales of **P₃**, if all else equal. In the case of product **P₄**, it appears that the sole product influencing its sales volume is **P₃**. This implies that a potential increase in **P₃** sales is associated with a corresponding increase of 31.9% in the sales volume of **P₄**. The product **P₅** remains unaffected by the sales of other products. However, **P₆** demonstrates cross-dependencies with product **P₅**, suggesting that a potential increase in **P₅** sales results in a minor 4.2% increase in **P₆** sales. Finally, focusing on product

P_7 , its sales are influenced solely by P_6 . This implies that an increase in P_6 sales, positively impacts P_7 sales, projecting a 34.6% increase in the latter, if all else equal.

What is the typical accuracy rate at which monthly sales can be forecasted? Does this accuracy vary across product types?

To generate future estimates and determine the most appropriate model for each product within the pharmaceutical company “Haleon Plc” in the Greek market, we employed MLR and EWMA forecasting techniques. The evaluation criteria utilized for comparison were the MAPE and MAE indicators. These measures allowed us to assess the accuracy of the derived results and draw conclusions regarding the suitability of each technique for individual products. The summarized outcomes in **Table 43** lead us to the conclusion that it does not exist a universally ideal forecasting model applicable to all products. This finding aligns with the fact that products belong to different category, have distinct attributes and satisfy different consumer needs, thereby justifying unique forecasting approaches.

Forecasting methodology	P_1	P_2	P_3	P_4	P_5	P_6	P_7
EWMA (2)	466.18%	82.40%	45.77%	41.21%	129.28%	23.84%	16.79%
MLR model M1	439.64%	1813.72%	46.48%	96.51%	169.95%	250.72%	22.16%

Table 43 "Comparison matrix of MAPE rate for all products and forecasting methodologies"

Upon analyzing the results presented in the preceding table, it becomes evident that for products P_1 , P_2 and P_5 , none of the employed forecasting methods demonstrate accuracy. The notably elevated MAPE indicators indicate that the forecasting models employed in the research, fail to capture the products’ sales patterns adequately. This inadequacy implies that crucial information and variables pertinent to the evolution of sales in the examined period are overlooked or not adequately incorporated, leading to inaccurate forecasting outcomes.

On the other hand, the remaining products show better forecasting accuracy. Specifically, for products P_3 , P_4 and P_6 the EWMA models produce reasonable forecasts exhibiting median accuracy rates. For products P_3 and P_4 , the EWMA models display better forecasting accuracy superior to the MLR models M1, measuring at 45.77% and 41.21% respectively. Additionally, for product P_6 , EWMA model displays a significantly better accuracy rate compared to the respective MLR model M1. Similarly, in the case of product P_7 , the EWMA model is preferred, as its MAPE value at 16.79% signifies a good forecasting

accuracy rate. However, in this product’s case, the MLR M1 model similarly provide quite accurate forecasts measuring its MAPE at 22.16%.

How did the Covid-19 pandemic outbreak affect sales and the totality of the examined products sales patterns?

Finalizing the thesis research, an investigation was conducted to assess the potential impact of the Covid-19 pandemic on the products produced by the pharmaceutical company “Haleon Plc” and are distributed in the Greek market. The study focused on evaluating the covid effect over the period from March 2020 to April 2022 to ascertain its potential impact on these products sales.

The outcomes from the MLR model M3 highlight that among the examined products, only P_2 and P_3 appear to have been significantly impacted by the outbreak of the pandemic. This conclusion stems from the statistical significance observed in the corresponding *p-values* of the dummy variable's coefficient, denoted as CE_t . During the pandemic period, P_2 experienced a decline in sales, with a decrease of approximately 59.7% in comparison to its pre-covid era sales, holding all other factors constant. Conversely, P_3 recorded an increase in sales during the same period, showing an uptick of around 22.8% compared to its sales in the pre-covid period.

6. Conclusions and future research

In the dissertation it was extensively presented the pharmaceutical sector's status as one of the most progressive and dynamically evolving industries globally. Within this landscape, companies continually augment their investments in innovation to enhance performance and gain competitive advantages, ultimately boosting their profits. The constantly evolution of global market coupled with unforeseen occurrences, as for instance the outbreak of the coronavirus pandemic, frequently impact the pharmaceutical companies' supply chain decisions. Consequently, effective forecasting processes stand as the cornerstone of their strategic position, playing a pivotal role in navigating the emerging challenges.

Within the current research, the sales properties of seven products were captured and the development of effective forecasting methods was attempted. In that way, we were focused on improving the supply chain operations of the “Haleon Plc” case company, in terms of order and production planning. Moreover, within the realm of sales forecasting and market

analysis in the pharmaceutical industry, this dissertation highlights the importance of acknowledging the existing limitations. It underscores the need for additional quantitative factors and qualitative criteria to be incorporated into future studies, as identified through the examination of the case study. Moreover, the dissertation suggests for further research within the literature, aiming the development of more robust and unbiased forecasting methodologies.

A challenging area of investigation for future studies could involve the incorporation of qualitative criteria that may influence product demand within the Greek market, such as gender, age, and other demographic features of customers. These qualitative drivers could be explored and identified through systematic surveys and questionnaires conducted by the Sales & Marketing department of the company. Moreover, future researchers could consider the possibility of simultaneously employing qualitative demand forecasting methods, either judgmental or experimental models, apart from the usually applied quantitative methods, and assess their results as well.

The upcoming studies could improve the forecasting models utilized in the present research by integrating additional quantitative explanatory variables (like price variations, the impact of promotional/advertising/discount strategies, and more) beyond those investigated in the dissertation, aiming to identify more decisive factors that affect the sales time series data. The development of such expanded models has the potential to improve forecasting accuracy, particularly for product codes that exhibited inaccurate results in the current study. Hence, it is recommended to pursue an integrated approach that combines the adoption of judgmental forecasting methods with the incorporation of expanded quantitative techniques. This approach aims to achieve more precise and reliable estimates for long-term predictions.

Another avenue of interest, that warrants consideration for further research in the pharmaceutical sector, is the exploration of additional modeling approaches incorporating econometric and extrapolation techniques. This may involve methods such as Holt’s-Winter’s exponential smoothing with trend and seasonality, ARIMA and S-ARIMA techniques, or contemporary non-linear approaches such as Artificial Neural Networks (ANN). As elucidated in chapter 2, particularly in subchapter 2.1, various studies have previously developed and utilized such techniques and their combinations. Therefore, expanding the forecasting techniques and approaches within the context of the examined

pharmaceutical company could offer more robust insights and managerial perspectives in future research literature. Furthermore, a possible direction of research is the application of machine learning methods (especially architectural neural networks or deep learning) in sales analysis. These models provide adaptable functional forms capable of accommodating nonlinear features within data.

However, as Thomaidis and Dounias (2011) (2012) supported, the inherent flexibility of these models also poses challenges and limitations, as exhaustive attempts to fit the data may lead to overparametrized model architectures, potentially resulting in poor out-of-sample performance (commonly known as the issue of overfitting). To address this challenge, it becomes crucial to develop procedures that can effectively allocate the optimal level of nonlinearity to specific areas within the data feature space. Exploring techniques to mitigate overfitting and enhance the interpretability of complex models in the context of sales analysis could also be a valuable and innovative direction for future research.

References

- Albright, S. C. (2010). *Data Analysis and Decision Making (4th ed.)*. South-Western Cengage Learning.
- Ayati, N., Saiyarsarai, P., & Nikfar, S. (2020). Short and long term impacts of COVID-19 on the pharmaceutical sector. *DARU Journal of Pharmaceutical Sciences*, 28, 799-805.
- Benoit, K. (2011). Linear regression models with logarithmic transformations. *London School of Economics, London*, 22(1), 23-36.
- Berenson M. L., L. D. (2020). *Basic business statistics : concepts and applications (14th Edition)*. Harlow, United Kingdom: Pearson Education Limited.
- Burton, A. L. (2021). OLS (Linear) regression. In J. B. Forde (Ed.), *The Encyclopedia of Research Methods in Criminology and Criminal Justice, Volume II, First Edition* (Vol. 2, pp. 509-514). John Wiley & Sons, Inc.
- Chopra, S., & Meindl, P. (2013). *Supply Chain Management - Strategy, Planning and Operation (5th ed.)*. Pearson Education Inc.
- Cox, D. R. (1982). Statistical significance tests. *British journal of clinical pharmacology*, 14(3), 325-331.
- de Albuquerque Dallegrave, T. L. D., Da Silva, M. V. B., de Souza Neto, D. A., Brasil, P. H., Junior, S., de Albuquerque Filho, J. E., & Santos, W. B. (2021). Sales Forecast Optimization: Ensemble and Time Series Comparison. *Revista de Engenharia e Pesquisa Aplicada*, 6(5), 110-119.
- Fisher, M. J., & Marshall, A. P. (2009). Understanding descriptive statistics. *Australian critical care*, 22(2), 93-97.
- Goodwin. (2005). Providing support for decisions based on time series information under conditions of asymmetric loss. *European Journal of Operational Research*, 163(2), 388-402.
- Haleon Plc. (2023, November 14). Retrieved from Annual Report 2022: <https://www.haleon.com/investors/annual-report-2022>
- Haleon Plc. (2024, January 10). Retrieved from Haleon Plc: <https://www.haleon.com/who-we-are/history>
- İMECE, S., & BEYCA, Ö. F. (2022). Demand Forecasting with Integration of Time Series and Regression Models in Pharmaceutical Industry. *International Journal of Advances in Engineering and Pure Sciences*, 34(3), 415-425.

- Lewis, C. D. (1982). Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting. *Butterworth Scientific*.
- Mentzer and Moon. (2004). *Sales Forecasting Management: A Demand Management Approach (2nd ed.)*. SAGE Publications.
- Merkuryeva, G., Valberga, A., & Smirnov, A. (2019). Demand forecasting in pharmaceutical supply chains: A case study. *Procedia Computer Science*, 149, 3-10.
- Model Building Sequence*. (2023, November 20). Retrieved from Source: National Institute of Standards and Technology, U.S Department of Commerce:
<https://www.itl.nist.gov/div898/handbook/pmd/section4/pmd41.htm>
- Moon, M. A. (2018). *Demand and Supply Integration, The Key to World-Class Demand Forecasting*. De Gruyter, 2nd edition.
- Moreno, J. J. M., Pol, A. P., Abad, A. S., & Blasco, B. C. (2013). Using the R-MAPE index as a resistant measure of forecast accuracy. *Psicothema*, 25(4), 500-506.
- R Rosca, E. (2011). Stationary and non-stationary time series. *The USV Annals of Economics and Public Administration*, 10(1), 177-186.
- Rathipriya, R., Abdul Rahman, A. A., Dhamodharavadhani, S., Meero, A., & Yoganandan, G. (2023). Demand forecasting model for time-series pharmaceutical data using shallow and deep neural network model. *Neural Computing and Applications*, 35(2), 1945-1957.
- Sarstedt, M. M. (2019). *Regression Analysis. In: A Concise Guide to Market Research. Springer Texts in Business and Economics*. Springer, Berlin, Heidelberg.
doi:https://doi.org/10.1007/978-3-662-56707-4_7
- Shashi, M. (2023). Sustainable Digitalization in Pharmaceutical Supply Chains Using Theory of Constraints: A Qualitative Study. *Sustainability*, 15(11), 8775.
- Shcherbakov, M. V., Brebels, A., Shcherbakova, N. L., Tyukov, A. P., Janovsky, T. A., & Kamaev, V. A. E. (2013). A survey of forecast error measures. *World applied sciences journal*, 24(24), 171-176.
- Siddiqui, R., Azmat, M., Ahmed, S., & Kummer, S. (2022, April). A hybrid demand forecasting model for greater forecasting accuracy: the case of the pharmaceutical industry. *In Supply Chain Forum: An International Journal*, 23(2), 124-134.
- Stank, T. P. (2012). Creating relevant value through demand and supply integration. *Journal of Business Logistics*, 33(2), 167-172.
- Thomaidis, N. S. (2021b). Simple Forecasting Techniques. Lecture Notes of Module SCM41: “Forecasting Theory”, of the MSc in Supply Chain Management,. 10-11.

- Thomaidis, N. S., & Dounias, G. D. (2011). On detecting the optimal structure of a neural network under strong statistical features in errors. *Journal of Time Series Analysis*, 32(3), 204-222.
- Thomaidis, N. S., & Dounias, G. D. (2012). A comparison of statistical tests for the adequacy of a neural network regression model. *Quantitative Finance*, 12(3), 437-449.
- Trojanowska, Justyna & Dostatni, Ewa. (2017). Application of the Theory of Constraints for Project Management. *Management and Production Engineering Review*, 8(3), 87-95.
- Wang, L., Zou, H., Su, J., Li, L., & Chaudhry, S. (2013). An ARIMA-ANN hybrid model for time series forecasting. *Systems Research and Behavioral Science*, 30(3), 244-259.
- Wikipedia. (2023). *The COVID-19 pandemic in Greece*. Retrieved from Wikipedia, The Free Encyclopedia: https://el.wikipedia.org/wiki_COVID-19_
- Yu, J., Kim, S. B., Bai, J., & Han, S. W. (2020). Comparative study on exponentially weighted moving average approaches for the self-starting forecasting. *Applied Sciences*, 10(20), 7351.

Appendix A: “EWMA in-sample analysis results products P_1 to P_4 ”

MONTH	Product P_1		Product P_2		Product P_3		Product P_4	
	SALES	EWMA	SALES	EWMA	SALES	EWMA	SALES	EWMA
Jan-17	247,326	247,326	234,366	234,366	228,950	228,950	68,991	68,991
Feb-17	192,035	247,326	319,633	234,366	170,937	228,950	77,113	68,991
Mar-17	117,288	193,105	126,380	318,736	141,289	219,402	87,742	68,992
Apr-17	146,721	118,756	33,578	128,405	157,685	206,546	74,304	68,995
May-17	129,137	146,180	53,227	34,576	296,253	198,505	74,470	68,996
Jun-17	140,442	129,467	84,951	53,031	207,120	214,592	64,619	68,997
Jul-17	123,070	140,230	44,627	84,615	94,166	213,362	64,107	68,996
Aug-17	115,732	123,402	90,293	45,048	124,733	193,745	58,374	68,996
Sep-17	173,748	115,880	187,198	89,817	184,057	182,387	79,662	68,994
Oct-17	257,357	172,628	289,742	186,173	214,268	182,662	118,424	68,996
Nov-17	201,682	255,717	229,527	288,652	196,118	187,863	68,267	69,004
Dec-17	176,956	202,728	180,299	230,149	181,073	189,222	77,750	69,004
Jan-18	249,558	177,455	152,172	180,824	141,353	187,881	130,023	69,005
Feb-18	150,149	248,162	147,989	152,474	226,805	180,223	43,821	69,015
Mar-18	107,786	152,046	130,150	148,036	209,524	187,890	69,455	69,011
Apr-18	116,088	108,643	34,410	130,338	165,736	191,450	92,352	69,011
May-18	147,834	115,944	68,505	35,420	123,614	187,218	83,634	69,015
Jun-18	112,696	147,217	323,746	68,157	217,696	176,750	100,763	69,017
Jul-18	76,449	113,364	11,755	321,056	67,394	183,489	43,592	69,022
Aug-18	26,201	77,164	6,450	15,010	21,634	164,382	16,932	69,018
Sep-18	389,783	27,188	93,875	6,540	223,330	140,888	102,309	69,010
Oct-18	304,115	382,763	115,631	92,956	223,595	154,457	131,318	69,015
Nov-18	92,049	305,638	143,557	115,392	231,375	165,836	118,287	69,025
Dec-18	96,184	96,184	114,613	143,261	272,781	176,622	119,884	69,033
Jan-19	162,163	96,184	171,210	114,915	164,655	192,448	105,353	69,042
Feb-19	112,008	160,886	133,227	170,617	173,094	187,874	80,194	69,048
Mar-19	132,993	112,954	43,714	133,621	151,009	185,441	27,085	69,050
Apr-19	120,048	132,605	25,802	44,660	141,052	179,774	74,671	69,043
May-19	99,461	120,291	171,131	26,000	263,080	173,401	59,857	69,044
Jun-19	191,260	99,864	192,637	169,604	117,169	188,161	46,629	69,042
Jul-19	131,368	189,491	151,471	192,395	125,334	176,477	58,604	69,038
Aug-19	30,456	132,493	4,580	151,902	52,191	168,060	23,643	69,037
Sep-19	140,359	32,431	92,631	6,131	426,786	148,990	85,036	69,029
Oct-19	138,828	138,270	155,459	91,721	182,751	194,710	120,109	69,032
Nov-19	142,086	138,817	114,342	154,788	290,155	192,742	99,200	69,040
Dec-19	192,869	142,023	61,652	114,768	83,008	208,774	85,000	69,045
Jan-20	197,867	191,885	182,811	62,211	148,790	188,075	99,200	69,048
Feb-20	223,348	197,751	240,526	181,542	165,172	181,610	75,000	69,053
Mar-20	181,412	222,852	138,520	239,905	372,774	178,904	131,142	69,054
Apr-20	149,943	182,214	111,150	139,587	144,281	210,812	81,967	69,064
May-20	62,870	150,568	12,277	111,449	127,531	199,862	39,554	69,066
Jun-20	153,867	64,568	11,347	13,321	80,909	187,958	92,967	69,061
Jul-20	139,364	152,138	107,532	11,368	314,200	170,339	96,308	69,065
Aug-20	47,834	139,611	74,437	106,520	95,157	194,016	30,381	69,070
Sep-20	152,580	49,611	86,084	74,775	179,574	177,746	72,811	69,063
Oct-20	184,386	150,587	76,904	85,965	167,978	178,047	152,601	69,064
Nov-20	169,877	183,732	34,724	76,999	360,617	176,390	62,536	69,078
Dec-20	143,866	170,145	35,149	35,169	262,164	206,710	112,890	69,077

Appendix B: “EWMA in-sample analysis results products P₅ to P₇”

MONTH	Product P ₅		Product P ₆		Product P ₇	
	SALES	EWMA	SALES	EWMA	SALES	EWMA
Jan-17	6,979	6,979	15,460	15,460	58,143	58,143
Feb-17	11,493	6,979	10,871	15,460	44,761	58,143
Mar-17	229,887	11,445	13,062	14,637	37,050	55,326
Apr-17	36,150	227,585	13,523	14,354	35,268	51,478
May-17	149,305	38,168	13,139	14,205	46,430	48,066
Jun-17	76,254	148,134	14,128	14,014	47,036	47,722
Jul-17	36,215	77,012	10,500	14,034	23,165	47,577
Aug-17	21,515	36,645	12,030	13,400	48,711	42,438
Sep-17	5,398	21,674	9,666	13,154	56,370	43,759
Oct-17	4,054	5,570	14,244	12,528	32,587	46,414
Nov-17	1,373	4,070	13,230	12,836	36,806	43,503
Dec-17	2,514	1,401	8,652	12,907	22,869	42,093
Jan-18	125,255	2,502	14,898	12,143	62,299	38,046
Feb-18	26,422	123,961	13,695	12,638	49,680	43,152
Mar-18	126,249	27,450	14,340	12,827	35,633	44,526
Apr-18	30,100	125,208	15,962	13,099	34,901	42,654
May-18	180,639	31,102	13,554	13,613	42,046	41,022
Jun-18	58,395	179,063	16,398	13,602	37,172	41,237
Jul-18	30,125	59,667	16,932	14,104	45,674	40,382
Aug-18	17,460	30,436	16,092	14,611	43,037	41,496
Sep-18	6,647	17,597	12,068	14,877	51,643	41,820
Oct-18	2,392	6,762	16,568	14,373	42,510	43,888
Nov-18	2,953	2,438	16,320	14,767	33,173	43,598
Dec-18	14,085	2,948	13,821	15,045	35,526	41,403
Jan-19	4,083	13,968	16,368	14,826	70,307	40,166
Feb-19	14,177	4,187	17,921	15,102	50,157	46,511
Mar-19	231,489	14,072	16,858	15,608	39,440	47,279
Apr-19	130,040	229,197	19,537	15,832	43,218	45,629
May-19	36,306	131,085	11,417	16,497	61,466	45,121
Jun-19	24,776	37,305	17,327	15,586	48,100	48,562
Jul-19	65,593	24,908	19,986	15,898	42,790	48,465
Aug-19	41,796	65,164	15,522	16,632	32,963	47,270
Sep-19	11,012	42,042	18,750	16,432	48,616	44,258
Oct-19	2,286	11,339	18,048	16,848	57,937	45,176
Nov-19	4,311	2,381	18,534	17,064	44,276	47,862
Dec-19	5,042	4,291	16,146	17,327	34,571	47,107
Jan-20	2,405	5,034	18,186	17,115	62,868	44,468
Feb-20	14,038	2,433	14,268	17,307	49,866	48,342
Mar-20	247,057	13,916	22,092	16,762	54,425	48,662
Apr-20	25,670	244,600	14,976	17,718	26,098	49,876
May-20	49,918	27,978	12,834	17,226	37,409	44,870
Jun-20	54,231	49,687	18,012	16,438	56,866	43,299

Jul-20	70,319	54,183	26,232	16,721	63,507	46,155
Aug-20	22,655	70,149	16,494	18,427	36,620	49,808
Sep-20	6,375	23,156	24,246	18,080	66,356	47,032
Oct-20	11,530	6,552	21,762	19,187	42,198	51,100
Nov-20	4,822	11,478	9,126	19,649	40,488	49,226
Dec-20	5,008	4,892	22,950	17,761	39,966	47,386

Appendix C: “EWMA out-of-sample performance results products P_1 to P_4 ”

MONTH	Product P_1		Product P_2		Product P_3		Product P_4	
	ABS ERROR	%	ABS ERROR	%	ABS ERROR	%	ABS ERROR	%
Jan-21	88,023	156.20%	28,711	445.96%	147,068	213.86%	7,455	12.10%
Feb-21	98,131	62.83%	10,716	61.39%	55,885	22.58%	79,163	53.40%
Mar-21	70,530	84.21%	2,225	14.72%	10,778	5.09%	7,805	10.15%
Apr-21	40,695	91.60%	11,850	360.09%	143,093	240.45%	17,407	33.68%
May-21	30,068	39.94%	943	38.12%	71,200	66.02%	24,106	25.86%
Jun-21	38,372	33.94%	9,877	79.91%	7,643	4.79%	3,879	5.95%
Jul-21	37,195	49.50%	3,741	43.93%	29,063	21.21%	29,939	76.46%
Aug-21	53,192	41.22%	882	11.50%	78,542	94.91%	55,976	426.78%
Sep-21	134,046	51.15%	73,773	90.57%	93,148	38.57%	6,968	9.16%
Oct-21	4,228	1.66%	22,283	21.64%	75,419	31.54%	30,848	30.87%
Nov-21	38,640	17.83%	57,472	35.88%	260	0.15%	19,465	21.98%
Dec-21	120,556	124.45%	40,919	34.48%	23,329	15.27%	4,498	6.11%
Jan-22	73,589	42.59%	37,583	23.99%	57,573	25.05%	17,405	20.12%
Feb-22	90,385	34.53%	76,411	95.66%	29,200	13.85%	10,923	13.65%
Mar-22	78,322	43.11%	52,725	188.57%	19,080	11.40%	5,864	9.27%
Apr-22	2,033	1.12%	39,325	57.97%	18,626	11.31%	1,690	2.51%
May-22	179,221	9024.24%	25,091	59.27%	46,369	34.62%	16,623	19.39%
Jun-22	147,641	96.44%	6,804	13.77%	1,077	0.63%	22,354	24.44%
Jul-22	247,337	62.21%	71,337	59.12%	155,094	47.34%	70,890	50.64%
Aug-22	241,933	160.37%	59,782	99.41%	58,720	42.15%	7,384	9.65%
Sep-22	24,768	18.94%	5,187	7.86%	51,972	38.11%	2,472	3.45%
Oct-22	98,076	42.77%	22,473	25.43%	77,014	29.99%	58,611	45.89%
Nov-22	202,938	828.69%	18,392	26.37%	19,441	11.24%	10,573	18.06%
Dec-22	105,719	78.81%	31,513	82.02%	83,205	78.44%	25,768	59.43%
MAE:	93,568		29,584		56,367		22,419	
MAPE:	466.18%		82.40%		45.77%		41.21%	

Appendix D: “EWMA out-of-sample performance results products P_5 to P_7 ”

MONTH	Product P_5		Product P_6		Product P_7	
	ABS ERROR	%	ABS ERROR	%	ABS ERROR	%
Jan-21	17,221	77.48%	506	2.78%	15,842	25.69%
Feb-21	124,212	84.93%	1,555	9.12%	1,649	3.47%
Mar-21	23,478	19.33%	704	3.70%	6,100	14.28%
Apr-21	97,044	393.31%	1,514	7.58%	1,377	2.98%
May-21	5,950	18.80%	2,796	17.56%	993	2.06%
Jun-21	21,818	40.86%	5,482	23.13%	4,324	8.35%
Jul-21	13,263	33.23%	3,368	21.27%	3,132	6.93%
Aug-21	1,449	3.75%	4,744	34.24%	1,143	2.45%
Sep-21	32,046	487.84%	1,106	5.86%	13,427	22.05%
Oct-21	4,941	251.31%	1,373	8.28%	21,729	76.09%
Nov-21	2,198	52.13%	3,940	28.64%	8,409	22.54%
Dec-21	1,777	73.54%	1,025	6.42%	11,200	20.31%
Jan-22	448	15.55%	7,351	77.74%	9,095	16.42%
Feb-22	5,230	64.50%	6,902	80.39%	9,857	25.70%
Mar-22	53,411	86.90%	6,186	30.27%	11,371	32.71%
Apr-22	72,795	54.45%	8,364	35.26%	9,008	17.08%
May-22	108,854	452.14%	6,943	70.00%	5,984	11.59%
Jun-22	31,756	55.73%	441	2.91%	5,985	14.63%
Jul-22	52,083	47.90%	1,826	13.32%	412	0.91%
Aug-22	85,481	376.63%	4	0.03%	5,814	11.32%
Sep-22	12,913	120.86%	1,257	7.63%	6,726	12.57%
Oct-22	2,546	30.77%	5,825	27.40%	10,229	26.94%
Nov-22	2,576	44.99%	5,294	47.34%	15,231	24.86%
Dec-22	3,930	215.71%	1,584	11.36%	445	0.91%
MAE:	32,392		3,337		7,478	
MAPE:	129.28%		23.84%		16.79%	

Appendix E: “MLR M1 performance results products P_1 to P_4 ”

MONTH	Product P_1		Product P_2		Product P_3		Product P_4	
	ABS ERROR	%	ABS ERROR	%	ABS ERROR	%	ABS ERROR	%
Jan-21	216,711	384.57%	326,763	5075.53%	103,161	150.01%	91,914	149.14%
Feb-21	100,935	64.62%	217,401	1245.42%	31,604	12.77%	9,750	6.58%
Mar-21	304,460	363.50%	533,692	3530.18%	44,081	20.83%	75,241	97.84%
Apr-21	218,347	491.47%	361,351	10979.99%	133,005	223.50%	77,393	149.73%
May-21	71,314	94.73%	280,009	11322.63%	123,013	114.06%	17,345	18.61%
Jun-21	270,035	238.81%	463,796	3752.40%	131,156	82.13%	96,338	147.72%
Jul-21	97,405	129.64%	156,552	1838.54%	104,750	76.45%	32,370	82.66%
Aug-21	62,379	48.34%	36,554	476.46%	8,963	10.83%	24,475	186.61%
Sep-21	85,048	32.45%	336,002	412.50%	4,797	1.99%	111,791	146.99%
Oct-21	64,138	25.13%	599,247	582.01%	6,350	2.66%	120,946	121.03%

Nov-21	38,921	17.96%	139,512	87.09%	15,101	8.59%	59,301	66.97%
Dec-21	155,848	160.88%	181,251	152.73%	45,492	29.78%	66,362	90.18%
Jan-22	68,769	39.80%	81,794	52.20%	13,235	5.76%	58,466	67.59%
Feb-22	70,414	26.90%	202,986	254.11%	2,777	1.32%	69,898	87.35%
Mar-22	110,017	60.55%	193,942	693.64%	28,559	17.06%	81,931	129.57%
Apr-22	64,378	35.54%	163,797	241.45%	40,584	24.63%	81,872	121.46%
May-22	134,348	6764.76%	371,266	876.97%	83,266	62.17%	7,460	8.70%
Jun-22	7,592	4.96%	218,788	442.86%	51,646	30.10%	61,468	67.21%
Jul-22	154,465	38.85%	46,434	38.48%	99,725	30.44%	68,242	48.75%
Aug-22	76,838	50.94%	15,384	25.58%	73,619	52.85%	43,583	56.97%
Sep-22	80,208	61.33%	165,393	250.78%	52,638	38.60%	59,173	82.66%
Oct-22	50,774	22.14%	218,899	247.71%	25,057	9.76%	22,196	17.38%
Nov-22	288,560	1178.33%	245,106	351.46%	43,655	25.23%	89,629	153.07%
Dec-22	67,347	50.21%	230,009	598.65%	89,210	84.10%	91,720	211.55%
MAE:	90,345		241,080		56,477		63,286	
MAPE:	439.64%		1,813.72%		46.48%		96.51%	

Appendix F: “MLR M1 performance results products P₅ to P₇”

MONTH	Product P ₅		Product P ₆		Product P ₇	
	ABS ERROR	%	ABS ERROR	%	ABS ERROR	%
Jan-21	10,403	46.80%	17,471	96.07%	13,398	21.73%
Feb-21	135,568	92.69%	16,642	97.63%	3,085	6.49%
Mar-21	45,446	37.41%	17,785	93.48%	2,130	4.99%
Apr-21	812	3.29%	17,051	85.42%	17,874	38.73%
May-21	111,379	351.94%	20,898	131.24%	2,159	4.48%
Jun-21	35,762	66.97%	16,625	70.15%	7,257	14.02%
Jul-21	55,240	138.41%	22,272	140.66%	1,974	4.36%
Aug-21	6,284	16.28%	27,683	199.82%	16,322	35.06%
Sep-21	2,175	33.12%	28,897	153.28%	15,752	25.87%
Oct-21	2,964	150.75%	29,098	175.59%	13,799	48.32%
Nov-21	6,763	160.42%	32,651	237.33%	5,711	15.31%
Dec-21	14,885	616.12%	35,773	224.06%	26,110	47.35%
Jan-22	10,149	352.04%	43,019	454.94%	8,911	16.09%
Feb-22	8,130	100.28%	53,390	621.82%	2,471	6.44%
Mar-22	396,574	645.21%	51,457	251.80%	9,127	26.25%
Apr-22	51,722	38.69%	33,677	141.95%	21,124	40.05%
May-22	147,975	614.64%	38,683	390.03%	3,626	7.02%
Jun-22	11,920	20.92%	48,806	321.64%	1,103	2.70%
Jul-22	2,590	2.38%	53,247	388.38%	883	1.95%
Aug-22	7,012	30.89%	50,916	334.89%	18,694	36.39%
Sep-22	7,690	71.98%	50,573	307.17%	9,254	17.30%
Oct-22	2,303	27.84%	44,854	211.00%	4,121	10.85%
Nov-22	2,546	44.48%	50,788	454.11%	30,141	49.19%
Dec-22	7,567	415.29%	60,626	434.78%	24,883	50.99%
MAE:	45,161		35,953		10,830	
MAPE:	169.95%		250.72%		22.16%	

Author’s Statement:

I hereby expressly declare that, according to the article 8 of Law 1559/1986, this dissertation is solely the product of my personal work, does not infringe any intellectual property, personality and personal data rights of third parties, does not contain works/contributions from third parties for which the permission of the authors/beneficiaries is required, is not the product of partial or total plagiarism, and that the sources used are limited to the literature references alone and meet the rules of scientific citations.