OPERATIONS AND SUPPLY CHAIN MANAGEMENT
Vol. 17, No. 4, 2024, pp. 267 - 282
ISSN 1979-3561 | EISSN 2579-9363

A Novel Prescriptive Supply Chain Analytics Model for Monitoring the Relationship Between Influential Variables Across the Supply Chain Network

Mohammad Khajehzadeh
Department of Industrial Engineering,
Iran University of Science and Technology, Tehran, Iran
E-mail: khajezadeh.sina@gmail.com

Farhad Pazhuheian
Department of Industrial Engineering,
Iran University of Science and Technology, Tehran, Iran
E-mail: farhad.pzh@gmail.com

Farima Seifi
Management Department, Anderson School of Management,
University of California, Los Angeles, USA
E-mail: farima.seifi.2025@anderson.ucla.edu

Alireza Ghorbani
Department of Statistics,
Shahid Beheshti University, Tehran, Iran
E-mail: ghorbanialireza@outlook.com

Golshan Madraki
College of Business
University of New England, Biddeford, USA
E-mail: gmadraki@une.edu (*Corresponding Author*)

ABSTRACT

The goal of supply chain monitoring is to provide an efficient tracking system for ensuring a secure flow of goods and services throughout the supply chain. Supply chain monitoring helps identify and address unexpected events early. There are five main components in supply chain networks including manufacturing, warehousing, procurement, logistic / transportation and demand. Numerous factors in each of the five components of the supply chain have direct impacts on sales and production. This paper presents a comprehensive method to monitor and analyze the impacts of these factors on both sales and production, ultimately aiming to identify areas for cost reduction and improvement. To achieve this goal, the sales and production are modeled and evaluated. Then, products with out-of-control behavior are simultaneously identified. Finally, to optimize out-of-control products, we considered the most influential factors affecting sales and production. The optimal values for out-of-control products are selected, which minimize operating costs while simultaneously maximizing operating profits within the supply chain. A case study in the personal care industry shows that the method increases the operation profit rate for out-of-control products.

Keywords: *Generalized Estimating Equation (GEE), Hotelling T^2 Control Chart, Joint Optimization Plot, Supply Shain Analytics*

1. INTRODUCTION

A supply chain network consists of five main components: Procurement and suppliers, manufacturing units, warehouses, logistic/transportation such as distribution centers and retailers, and demand/customers. These units are responsible for creating value for customers by optimizing the flow of acquiring raw materials, producing goods and services, and delivering them to end-users (Aamer *et al.*, 2020). Extensive research on supply chain networks highlights the pivotal role of a streamlined network in determining a business's overall economic success (Wang *et al.*, 2020). Accordingly, supply chain operations come across various types of risks, such as delays, poor quality from suppliers, procurement failures, imprecise forecasts, uncertain consumer demands, and potential supply chain disruption like natural disasters (Hudnurkar *et al.*, 2017). Clearly, in the absence of a well-organized supply chain strategy, these risks and vulnerabilities might result in financial losses or even a complete collapse of the supply chain network (Wang *et al.*, 2020). Moreover, growing complexity of supply chains (Gómez and Lee, 2023) as we extend downstream to connect producers and consumers underscores the critical need for a measurement system (Kim and Oh, 2005).

This system should facilitate coordinated decision-making across all supply chain components, aligning with the goals of independent elements while enhancing overall network performance (Wang, 2010). Effective monitoring, achieved through timely detection of abnormal operations, is crucial for a supply chain's economic viability and successful functioning. Supply Chain Monitoring (SCMo) fulfills this role by promptly identifying and providing early warnings of network issues, enabling effective management (Wang *et al.*, 2023). SCMo has become an integral part of supply chain management, with its primary objective being improved decision-making. This is achieved by characterizing normal operating conditions, revealing discrepancies between planning and execution, issuing warnings for abnormal situations, identifying potential root causes, and providing recommendations for mitigation (Wang *et al.*, 2023).

During the last decades, the mainstream focus of supply chain management has been on cost efficiencies by using just-in-time methods and avoiding holding excess inventory (Kovács and Falagara Sigala, 2021). Most studies in literature often analyze only one out of the five components in supply chain, e.g., solely focused on optimizing manufacturing or distribution. Consequently, they use methods that might be sufficient for limited number of variables such as statistical methods based on cross-sectional data, which can only provide one measurement for each response at a specific time point. The primary contribution of this paper is to address these shortcomings through the follows unique features:

1. Our proposed analysis focuses on all five components of supply chain functions over time to reduce operating costs and increase supply chain efficiency.
2. Our method is comprehensive and novel that combines established techniques and can create an efficient approach for handling numerous variables and correlated characteristics in the data when monitoring the supply chain network across its five functions over time. This method is rooted in a robust foundation grounded in principles of multivariate analysis, focusing on longitudinal analysis and optimization techniques, aiming to optimize supply chain networks.
3. Our proposed method can be applied and extended to larger and more complex supply chain networks, where multiple elements exist within each of the five components of the supply chain network.
4. Finally, a novel case study on the supply chain of a personal care company in the Middle East is presented, and the application and performance of the proposed method is assessed over this real-world case study.

2. LITERATURE REVIEW AND BACKGROUND

Supply chain management now relies more than ever on data to capture cost and performance trends, monitor inventory, support process control and improvement, as well as optimize production. To gain insights and make informed decisions about all components of supply chain management, it is important to understand the value of data analytics and its effective use in supply chain management (Sukha and Prabhu, 2023). This application of advanced data analytics techniques to supply chain management is called Supply Chain Analytics (SCA) (Gilvan, 2014). The SCA techniques have a significant role in SCMo and can be classified into three main types:

1. Descriptive analytics extracts valuable insights from the network data to describe “what is happening”. For instance, real-time information about the location and quantities of goods in the supply chain network equips managers with the necessary tools to adjust regarding delivery schedules, replenishment orders, emergency orders, and transportation modes, etc.
2. Predictive analytics derives demand forecasts from historical data and predicts "what will be happening" in the future.
3. Prescriptive analytics generates decision recommendations by combining descriptive and predictive analytics models with mathematical optimization techniques. It addresses the question of "what should be happening" and guides decision-makers towards optimized strategies and solutions (Gilvan, 2014). Notably, prescriptive analytics receives significant attention in academic research, software development, and practical application within the domain of SCA.

2.1 Relevant Studies

This section discusses the details of the most recent and relevant studies existing in the literature in the field of Supply Chain Monitoring.

Nguyen *et al*, (2018) proposed a visual framework **Figure 1**, which helps the understanding of the relationships between different supply chain layers/components and the role of data analytics studies in assessing their performance. The first layer represents the five main key functions of Supply Chain Management, while supply chain data analytics is associated with the second and third layers in this taxonomy, referring to the aforementioned types of SCA studies.

As for procurement, Jain *et al.* (2014) conducted a study employing a data mining approach to uncover the hidden relationships between data used for suppliers' selection and their overall rating based on prior performance. This approach significantly aids in optimizing the supplier selection process. Similarly, Choi *et al.*, (2016) introduced a novel data analytics approach using Fuzzy Cognitive Maps to enhance decision-making in IT service procurement for the public sector. This unique method combines data analytics with intuitive qualitative techniques to create decision models, and its efficacy is validated through a case study, demonstrating its value in facilitating robust public decision-making. Mori *et al.*, (2012) utilized Support Vector Machine and Logistic Regression to build a prediction model for customer-supplier relationships, which can help to identify potential business partners.

As for manufacturing, Zhang *et al.*, (2017) proposed an overall architecture called data-based analytics for product lifecycle. This architecture leverages data analytics and service-driven patterns and through a practical application scenario, it demonstrates impactful benefits for customers, manufacturers, the environment, and all stages of product lifecycle management.

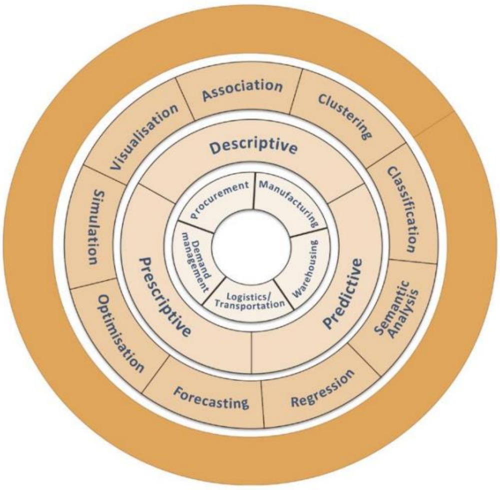


Figure 1 Studies classification framework

In studies on inventory and warehousing, Chiang *et al.*, (2011) introduced an association index and proposed a data mining-based storage assignment approach that enhances the efficiency of order picking. Khurana and Kumar, (2017) conducted a practical usage of data analytics in inventory management and implemented linear discriminant analysis on a large data set to find the dependencies. Chen, (2021) addressed an inventory control problem with active exploration in the inventory through lost sales in a shifting demand environment through historical data analysis. Suwignjo *et al.*, (2023) applied gradient boosting model for solving the inventory status (overstock, understock) by considering demand forecast in an FMCG company.

In terms of logistics/transportation, Zhao *et al.*, (2017) used the upper and lower limits of uncertain parameters from historical data to redesign a green supply chain network. Li *et al.*, (2015) employed Lasso Granger causality models to select the most relevant data to build a traffic prediction model. Dash and Mohanty, (2018) proposed a deterministic linear programming model to address a transportation problem where the unit cost of supplies, transportation and demands are uncertain. They minimized the expected value of an uncertain objective function with respect to some constraints under certain confidence level. Amellal *et al.*, (2023) addressed the lack of accurate lead time for meeting customer demand by developing a hybrid model combining Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) architectures.

As for demand management, Salehan and Kim, (2016) used a sentiment mining approach to study predictors of online consumer review performance. Chong *et al.*, (2016) employed a neural network to examine the impact of different variables such as online reviews, promotion strategies, and sentiments on product sales. Additionally, Mohan *et al.*, (2021) proposed a demand forecasting and route optimization approach for delivering products on time and meeting customer’s growing expectations. Nguyen, (2023) reviewed the artificial intelligence models such as recurrent neural networks for demand forecasting in supply chain over various industries.

Wang *et al.*, (2020) and Kapil *et al.*, (2021) described that the data-driven optimization techniques are playing a significant role in enhancing SCM in uncertain environments. By integrating machine learning, data analytics, and robust optimization, the planning of supply chain network can be more efficient and accurate. These data-driven techniques display their potential in improving SCM under uncertainty. Nitin *et al.*, (2023) employed bibliometric statistical analysis on supply chain analytics to provide a systematic analysis of this area for identifying key research themes and sub-themes for enhancing performance of supply chain management and business value.

A major gap in literature is that most of the current studies have focused solely on one specific function out of the five in the supply chain network i.e., Manufacturing, procurement, demand, warehousing, transportation. To the best of our knowledge, there are no comprehensive studies integrating the analysis of all five main functions of supply chain network through data analytics approaches. Accordingly, a comprehensive analysis through a data-driven statistical model across all five supply chain functions for reducing operating costs will be a compelling research opportunity in this field. Moreover, to monitor the supply chain network across its five functions over time, a statistical method is required to be used due to numerous variables to be considered and the correlated characteristics in the data. Longitudinal data is particularly suitable for handling such data, unlike cross-sectional data, which can only provide one measurement for each response at a specific time point. One of the main advantages of longitudinal data is that the correlation among observations within an experimental unit leads to more specific power level in longitudinal data compared to cross-sectional data. Consequently, a smaller number of experimental units is needed in the sample to achieve a specific power level. Achieving a specific power level can guarantee a high likelihood of detecting meaningful effects or relationships between characteristics (Ten Have, 1995). In this regard, a Generalized Estimating Equation model is employed to create a control chart for detecting inefficient items in the problem and then the paper proposes a joint optimization approach to enhance these items and boost supply chain efficiency.

2.2 Background of the Tools and Methods

The key terms and concepts that will be used in our method are reviewed in this section to aid in the understanding of the proposed solutions and their effectiveness in addressing the challenges presented in the case study.

2.2.1 Longitudinal Data

The primary goal of a longitudinal study is to characterize changes in responses over time and determine the factors that influence these changes. Thus, the main characteristic of longitudinal data is the repeated measurement of subjects over time. One important feature of longitudinal clustered data is that each cluster consists of observations from a single experimental unit at different time points. (Ten Have, 1995 and Fitzmaurice *et al.*, 2011). Finding a useful set requires understanding the sources of random changes in longitudinal data. These sources can be classified into the following three general categories (Fitzmaurice *et al.*, 2011):

- *Random effects*: In a situation where a population is randomly sampled, various aspects of the sample members' behavior represent random variations between experimental units. Random effects are the variables that differ among subjects.
- *Sequential correlation*: At least a portion of each measured unit exhibits a time-dependent response within that unit. These random changes are caused by the correlation between the measured pairs within the same unit, and this correlation depends on the time difference between the measured pairs. Usually, the correlation decreases with increasing time interval.
- *Measurement error*: The measurement process within the experimental unit may lead to changes in the data.

In longitudinal data, y_{ij} represents the response variable of the i-th subject in the j-th measurement, and x_{ij} is a p-dimensional vector of explanatory variables at time t_{ij}, where $j = 1,2,\dots,n$ and $i = 1,2,\dots,m$, where n is the number of subjects and m is the number of measurements. Most longitudinal analyses are based on a regression model, such as the following linear model:

$$y_{ij} = \beta_1 x_{ij1} + \beta_2 x_{ij2} + \cdots + \beta_p x_{ijp} + \epsilon_{ij} \quad (1)$$

This model can be expressed in matrix form as follows:

$$y = X_i \beta + \epsilon_i \quad (2)$$

where X_i is an $n_i \times p$ matrix of explanatory variables, β is the vector of unknown regression coefficients of dimension p, $\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{in_i})$ is the random vector of errors, and $y_i = (y_{i1}, \dots, y_{in_i})$ represents the repeated responses for the i-th subject (Ten Have, 1995 and Fitzmaurice *et al.*, 2011).

2.2.2 Marginal Models

Marginal models are one of the common methods for longitudinal data modeling that will be used in this study. In marginal models, the response variable is modeled on covariate variables apart from the within-subject correlation structure (Fitzmaurice *et al*, 2011). In this model, the marginal expectation of the response variable is expressed as a function of the explanatory variables. The term "*marginal expectation*" refers to the average response in a subpopulation with common values of X. A marginal model is characterized by the following three key features.

The marginal expectation of the response, $\mu_{ij} = E(y_{ij}|X_{ij})$, depends on the covariates with a certain link function Eq. (3):

$$g(\mu_{ij}) = \eta = X'_{ij}\beta \quad (3)$$

The marginal variance of the responses is related to the marginal mean as Eq. (4):

$$Var(y_{ij}) = \phi . v(\mu_{ij}) \quad (4)$$

where, $v(\mu_{ij})$ is a specified variance function; and, ϕ is a scale parameter that may need to be estimated. The correlation between within-subject observations is a function of the marginal mean and additional parameters α.
within-subject communication of the repeated responses vector is modeled as (5) by considering the correlation pattern of the first-order autoregressive model, where $0 \leq \alpha \leq 1$ (Ten Have, 1995):

$$Corr(y_{ij}, y_{ik}|X_{ij}, X_{ik}) = \alpha^{|k-j|} \quad (5)$$

Therefore, in the marginal models, the correlation in the longitudinal data is considered through the variance-covariance matrix. In a marginal model, the identical relationship between the response variable and the matrix of covariate variables is assumed to apply to all subjects in the sample. The key feature of marginal models is that they model the mean response and within-subject relationships separately. Consequently, the regression coefficients in this model are interpreted as population averages, meaning that changes in the mean response relative to the predictor variables are examined in the sub-population defined by these predictors (Carrière *et al.*, 2002 and Fitzmaurice *et al.*, 2011). As a result, when studying time-independent predictor variables, i.e., variables that do not change for each individual during the follow-up period, population-average interpretations are typically preferred (Wu *et al.*, 2012). In these models, the method of *Generalized Estimating Equations* (*GEE*) is employed to estimate the parameters (Fitzmaurice *et al*, 2011). GEE allows for the estimation of parameters while considering the correlation structure between the response variables, without requiring knowledge of their specific distribution. The correlation matrix derived from this structure is assumed to be identical for all subjects in the sample. The data consists of repeated measures of the response variable and covariate variables within a group of subjects. With this method, a suitable model is constructed for the mean of the response variable, incorporating separate observations and correlated variables (Fitzmaurice *et al*, 2011). In most cases, according to the type of response variable and specific design conditions, a generalized linear model such as (6) can be used to model grouped structures. In (6), y_i represents the value of the response variable for subject i, X_i is the correlated variable or covariate, and β is a vector of model parameters or independent coefficients of X_i. ε_i represents the random terms, and g is the link function, which maps the set of possible values of the response variable to a linear function of the X variable.

$$y_i = \mu_i + \varepsilon_i \quad , g(\mu_i) = f(x) = X_i\beta \quad (6)$$

To estimate the parameters of the "generalized linear model" and perform inference, it is typically assumed that the error terms (ε) have the same distributions and are independent. However, this assumption often does not hold in practice. As an alternative, Generalized Estimating Equations (GEE) offer a non-parametric approach that does not rely on the normal distribution assumption for the error term.

2.2.3 Generalized Estimating Equations (GEE)

In GEE, instead of assuming a specific distribution for the data, the best estimate for β is generated using iterative calculation techniques and trial and error. This approach aims to be the most descriptive for capturing the relationship between the response and dependent variables (Ziegler *et al.*, 1998).

The GEE estimator for β in the marginal model is obtained by minimizing the objective function (7):

$$\sum_{i=1}^{N}\{y_i-\mu_i(\beta)\}^{\prime}V_i^{-1}\{y_i-\mu_i(\beta)\} \tag{7}$$

where V is not dependent on β and μ_i, and μ_i is a vector of the average response with the following components:

$$\mu_{ij}=\mu_{ij}(\beta)=g^{-1}\left(X_{ij}^{\prime}\beta\right) \tag{8}$$

Using differential and integral calculus, we can demonstrate that the existence of the minimum function in Eq. (7) requires the solution of the following generalized estimating equations:

$$\sum_{i=1}^{N}D_i^{\prime}V_i^{-1}(y_i-\mu_i)=0 \tag{9}$$

in which V_i is called "working" covariance matrix and $D_i=\partial\mu_i/\partial\beta$ represents the gradient or derivative matrix. Since GEE relies on both β and α, the following iterative two-step estimation procedure is necessary:

1. According to the current estimates of α and ϕ, V_i is estimated and the updated estimate of β is obtained as generalized estimating equations resulting from Eq. (9).
2. According to the current estimate of β, the updated estimates of α and ϕ are obtained based on the standardized residuals Eq. (10):

$$e_{ij}=\frac{y_{ij}-\widehat{\mu_{ij}}}{\sqrt{v\left(\hat{\mu}_{ij}\right)}} \tag{10}$$

Finally, in this two-step estimation method, the process is typically iterated between steps 1 and 2 to ensure convergence (Fitzmaurice *et al.*, 2011).

*2.2.4 The Hotelling Multivariate Control Chart (**T^2** Control Chart)*

In many instances, monitoring multiple related quality characteristics simultaneously is crucial. It helps control these traits effectively and evaluate their potentially deceptive nature. To tackle such scenarios, specialized tools must be employed to detect, identify, and analyze the significant sources of variability in a given process. Among the various techniques, Multivariate control charts stand out since they can simultaneously monitor and control multiple characteristics that define the quality of a single production process. The Hotelling T^2 control chart holds significant recognition in the literature and comes highly recommended for processes involving multiple qualitative characteristics. Since these features are interconnected, monitoring them collectively is crucial. The T^2 test statistic is derived from the Eq. (11) (Montgomery, 2019).

$$T^2=n(\bar{X}-\bar{\bar{X}})^{\prime}+S^{-1}(\bar{X}-\bar{\bar{X}}) \tag{11}$$

In (11), $\bar{X}$ is the mean vector, and S represents the covariance matrix of the process. The application of multivariate control Hotelling T^2 chart is performed in two phases:

- *Phase I*: the chart's upper control bound is calculated using Eq. (12)

$$UCL=\frac{p(m-1)(n-1)}{mn-m-p+1}F_{\alpha,P,mn-m-p+1} \tag{12}$$

In (12), p is the number of variables, m is the number of samples, n is the sample size, and α is the parameter of the F distribution degree (Bersimis *et al.*, 2007 and Tracy *et al.*, 1992).

- *Phase II*: the chart's upper control bound is expressed by Eq. (13)

$$UCL=\frac{p(m+1)(n-1)}{mn-m-p+1}F_{\alpha,P,mn-m-p+1} \tag{13}$$

The lower control limit for both phases is equal to zero in the control chart (Bersimis *et al.*, 2007 and Tracy *et al.*, 1992).

2.2.5 Joint Optimization Plot

In an industrial experiment or decision-making system, there are several control factors (independent variables) denoted as $x_1,\dots,x_k$, multiple control responses (dependent variables) represented by $y_1,\dots,y_N$, and various target values $\tau_1,\dots,\tau_N$. When aiming to optimize such a system, conflicts may arise in the results while attempting to optimize the control factors individually. Consequently, a relative combination of the factors is necessary to bring the multiple responses as close as possible to the specified target values. The application of the Joint Optimization method enables us to achieve this objective. Joint optimization refers to the process of finding the optimal values for multiple variables or parameters simultaneously. It involves considering the trade-offs and compromises between different objectives or constraints. The strategy for simultaneously optimizing multiple responses is presented as follows (Kuhnt and Rudak, 2013 and Pignatiello Jr., 1993):

Consider an experiment with control factors $x_1,\dots,x_k$ and N responses $y_1,\dots,y_N$ with target values $\tau_1,\dots,\tau_N$. The optimal settings for the control factors should be determined to ensure that the means of the responses are on target with minimal variances. This can be achieved by minimizing the expected loss of y with respect to x, which is referred to as the risk function and is defined as follows:

$$\begin{aligned}R(x)=E\left(loss(y|x)\right)&=E((y-\tau)^TC(y-\tau)|x)\\&=trace\left(C\Sigma(x)\right)\\&\quad+(\mu(x)-\tau)^TC(\mu(x)-\tau)\end{aligned} \tag{14}$$

where $(y-\tau)^TC(y-\tau)$ represents the loss function, and C is the cost matrix, $\mu(x)=E(y|x)$ denoting the expected value of y given x, and $\Sigma(\mathrm{x})$ represents the covariance matrix of y given x. In the case of independent responses $y_1,\dots,y_N$, both the covariance matrix $\Sigma(x)$ and C become diagonal matrices, so Eq (14) turns to Eq (15) where $\mathrm{c_i}$ represents the i th element of cost matrix.

$$R(x) = \sum_{i=1}^{N} c_i.(\sigma_i^2(x) + (\mu_i(x) - \tau_i)^2) \quad (15)$$

Minimizing the risk function, as described in Eq. (16), means adjusting the average value (mean) towards the desired goal while keeping the variability (variance) as low as possible (Pignatiello Jr., 1993). In situations involving an unknown matrix C, this cost matrix is decomposed to Eq. (16) in which A is diagonal standardization matrix and W is diagonal weight matrix.

$$C = A^T WA \quad (16)$$

Diagonal elements of weight matrix W are specified through a slop vector $d\epsilon R^N$ for N responses and a stretch value log(a) in the following form

$$log\ \omega = d.log\ a \quad (17)$$

where ω is diagonal of weight matrix W and $\{a_t\}_{t=1}^N$ is an increasing equidistance vector within the interval $[\log a_{\text{low}}, \log a_{\text{high}}]$. Standardization matrix A for k control factors is defined as

$$A_y = diag([\frac{1}{k}\sum_{k=1}^{K} \widehat{var}(y_i|x_k)]_{i=1,\dots,N}^{-1/2}) \quad (18)$$

Therefore, the estimated risk function in Eq. (14) is given by (19) where b_i denotes the inverse of i th element of standardization matrix A.

$$\hat{R}(x) = \sum_{i=1}^{N} \omega_i.\frac{(\widehat{var}(y_i|x) + (\hat{E}(y_i|x) - \tau_i)^2)}{b_i^2} \quad (19)$$

The sequence of weight matrices ensures an optimal solution, and a joint optimization plot displays the optimal parameter setting for every cost matrix $C_t = A^T W_t A$ in one plot and its corresponding predicted response in other plot (Pignatiello Jr., 1993).

3. PROPOSED METHOD

We propose a comprehensive method to monitor and optimize variables across all five main components of a supply chain network. The goal is to enhance the overall performance of any general supply chain network. The method requires data from the five main components of the supply chain over a twelve-month interval and organizing it as a matrix. Due to the high correlation and longitudinal structure of the data, it is necessary to employ a method that does not require assuming normality for the error distribution in the regression model. This leads to the application of the GEE method for modeling the problem. Notably, such a comprehensive statistical method has not been employed in previous studies analyzing supply chain networks. This study explored changes in production & sales over time (using GEE analysis) and subsequently used a Hotelling T^2 multivariate control chart to monitor product performance and identify any supply chain issues. Recognizing the need for optimization, a joint optimization method considering interconnected variables is employed to simultaneously optimize costs and profits. This approach, rarely used in supply chain monitoring, offers a unique and effective solution for improving overall performance. These steps are outlined as follows:

Step 1: The variables resulting from the problem are reshaped into a longitudinal form to make a set of explanatory and response variables for sales and production. The supply chain has essentially been decomposed for further analysis.

Step 2: In this longitudinal supply chain study, the time variable is introduced as a fixed effect. Additionally, variables associated with all five functions of the supply chain are considered as covariates to assess their influence on the response variables. The (GEE) method is utilized to analyze the data, considering the within-*Stock Keeping Unit* (*SKU*) effect as a latent variable. The modeling process involves utilizing the "*xtgee command*" in *Stata software*. Two separate analyses are carried out: one for sales and another for production.

Step 3: Using the Hotelling T^2 control chart, we use this control chart to monitor two variables derived from the fitted values in Step 2 of the Phase I control chart. In this step (3), we detect and optimize out-of-control products using the *Joint Optimization* (*JOP*) method. We performed this crucial step with the *MSQC* package in R-4.3.2 software.

Step 4: The variables related to products beyond the control chart should be optimized. We utilized the JOP model, considering the cost of goods sold and finance costs of the products as the response variables. The decision variables for the modeling process are internal factors within the business. This phase is executed using the JOP package in R-4.3.2 software.

In the following, the efficiency of the proposed method is demonstrated on a real case study in the supply chain of a personal care company.

4. A REAL-WORLD CASE STUDY

The two main goals of any supply chain system are to maximize profit and minimize total cost through efficient coordination of all supply chain facilities in the network. Today, the need for this coordination is more critical than ever due to geopolitical challenges, the repercussions of the global pandemic, economic disruptions, and other factors that have affected both demand and supply. Supply chain networks are complex systems with interrelated functions and various sections across procurement, production, warehousing, logistics, and retail, all of which require effective coordination (Soosay, 2023). This section represents our proposed model through a real-world case study of a company (and its supply chain) in the personal care industry in the Middle East. This case study demonstrates the importance of coordination across all sections of the company's supply chain under uncertain circumstances and economic disruptions in the Middle East. Due to the confidentiality agreement, the company's name and the specifications of products cannot be disclosed.

The study focuses on analyzing a Multi-Echelon Supply Chain network that involved 51 Stock Keeping Unit (SKU) products over a 12-month financial year. Data is comprehensively collected across all five main components of the supply chain network: procurement, manufacturing, warehousing, logistics/transportation, and demand management. Each SKU is evaluated monthly based on 22 variables derived from the five main components of the supply chain network. The dataset recorded the values of these variables, resulting in a matrix with 612 rows (51 SKUs × 12 months) and 22 columns (variables). The response variables are the production quantity and sales of each SKU per month.

Two major costs in SCM are the Cost of Goods Sold (COGS) and operational costs (Brandenburg and Seuring, 2011). Identifying these major costs and defining appropriate variables associated with them is crucial for establishing an optimal business strategy, leading to cost efficiency and positive financial outcomes (profits) in each fiscal year. Therefore, we defined the variables across the following five dimensions, considering criteria that have a direct effect on COGS and operational costs:

- Manufacturing Variables (X_1, X_2, X_3): The consumption of raw materials (material usage) is a direct cost driver in manufacturing, as it directly correlates with production output and their costs (Favi *et al.*, 2021). It is important to note that other factors in the manufacturing processes can also contribute to the COGS, such as facility utilization, setup cost, defect rate, etc. However, this study focuses on the amount of raw materials usage as a major factor. According to the products Bill of Materials, three primary raw materials are used in this manufacturing process, accounting for approximately 68% of the total materials used in each product. These materials are coded as I, II, and III in the study. All these materials are sourced overseas, making their impact on the total cost more critical. These materials are measured in kilograms, and their usage is calculated per SKU per month based on the Bill of Materials for each SKU.
- Warehousing Variables (X_4, X_5, X_6): The amounts of raw materials that need to be stored have a direct impact on costs associated with warehousing space, inventory management, and the interest on capital tied up in the raw material inventory (holding cost). The storage amounts for these three primary materials are calculated based on their weights (kilograms) for each SKU per month.
- Procurement Variables (X_7 to X_{15}): The prices and quantities of the purchased raw materials are the two critical factors in COGS. Additionally, government subsidies on these materials can help reduce costs and improve the overall financial performance of the system. In this case study, the procurement of the three primary raw materials (I, II, III) constitutes a major portion of total purchases. Since all three materials are sourced overseas, they qualify for governmental subsidies. Therefore, our procurement variables include: 1) purchasing prices of the three primary raw materials in US dollars (X_7, X_8, X_9), 2) purchasing quantities of these materials in kilograms (X_{13}, X_{14}, X_{15}), and, 3) government subsidies for these materials in US dollars (X_{10}, X_{11}, X_{12}). These variables are tracked monthly for each SKU.
- Logistics/Transportation Variables (X_{16} to X_{18}): Freight and transportation costs depend on various factors such as distance, cargo volume, transportation methods, and delivery charges between different types of facilities within the supply chain network (Mamonov and Poluektov, 2021). In this case study, three transportation-related variables are defined, which have a direct impact on COGS: the freight cost of raw materials from suppliers to manufacturers in US dollars (X_{16}), transportation costs of delivering finished products (SKUs) to the distribution centers in US dollar (X_{17}), and the order size (quantity) of retailers and/or customers (delivered from distributors to retailers) (X_{18}). These variables are all recorded monthly per SKU.
- Demand Variables (X_{19} to X_{22}): Shelf price, currency exchange rates, consumer price index, and inflation rate are critical external factors that influence consumer demand and purchasing power (Khajehzadeh *et al.*, 2022). These factors should be considered when determining an optimal pricing strategy to ensure competitiveness and profitability. It is important to note that the company's market is in the Middle East, where external factors can be very unstable due to geopolitical circumstances. Therefore, these variables are critical and can substantially affect shelf prices. These variables pertain to the customer side of the supply chain network. Data for each SKU is collected quarterly to ensure comprehensive data collection.

As it is observed, the model covers the entire supply chain network, starting from the factory to the distribution centers (depots), and finally, the presence of products on store shelves. The response variables and independent variables corresponding to each supply chain component are integrated into the dataset for analysis.

Figure 2 illustrates a conceptual model representing the mutual influence of the five major components on each other. The set of variables defined for each of these components impacts both its correlated component and the entire supply chain network. The response variables (Y_1 and Y_2) is related to the overall performance of the supply chain network, integrating the financial effects of these 22 variables. By monitoring these effects, the model provides a comprehensive understanding of how changes in one section can influence the entire network.

This case study provides valuable insights into the complexities of the supply chain network of our targeted personal care company. **Table 1** summarizes the variables:

4.1 Results

A random effects model with the identity link function is fitted to the data for each of the two variables of sales and production. The mixed-effects model for each response is as Eq.20 where β_0 is the intercept, u_i is the random effect of the i-th, β_{ij} is the effect of variable x_{ij}, and ε_{ij} is the random error.

$$y_{ij} = \beta_0 + u_i + \sum_{i=1}^{52} \sum_{j=1}^{12} \beta_{ij} x_{ij} + \varepsilon_{ij} \qquad (20)$$

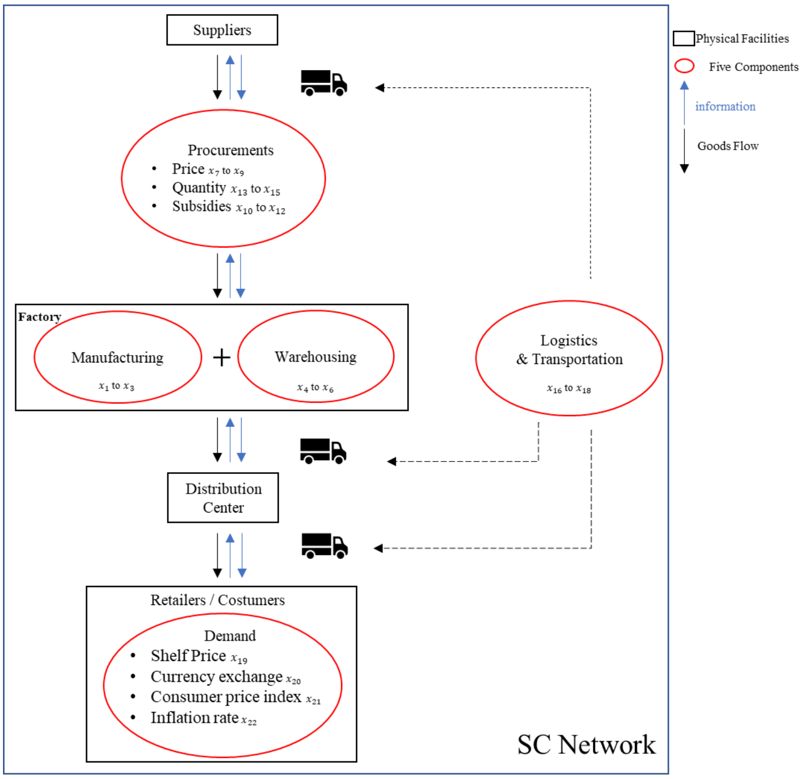


Figure 2 Study conceptual model

Table 1 Variables of the supply chain network

Layers (Component)	Indexes	Variables
Manufacturing	x_1	Consumption of raw material (I) (per unit of kg)
	x_2	Consumption of raw material (II) (per unit of kg)
	x_3	Consumption of raw material (III) (per unit of kg)
Warehousing	x_4	Storage of raw material (I)
	x_5	Storage of raw material (II)
	x_6	Storage of raw material (III)
Procurement	x_7	Raw material purchasing price (I)
	x_8	Raw material purchasing price (II)
	x_9	Raw material purchasing price (III)
	x_{10}	Raw material inputs subsidy (I)
	x_{11}	Raw material inputs subsidy (II)
	x_{12}	Raw material inputs subsidy (III)
	x_{13}	Raw material Purchasing Quantity (I)
	x_{14}	Raw material Purchasing Quantity (II)
	x_{15}	Raw material Purchasing Quantity (III)
Logistic / Transportation	x_{16}	Freight Charges (suppliers to Manufacturer)
	x_{17}	Transportation Cost (manufacturer to distributors)
	x_{18}	Order Size (distributors to retailers)
Demand	x_{19}	Shelf Price
	x_{20}	Currency exchange
	x_{21}	Consumer price index
	x_{22}	Inflation Rate

The results of fitting the GEE model on production and sales are shown in **Table 2** and **Table 3**, respectively. Based on the results of **Table 2**, variables x_4, x_5, x_6, and x_{19} are statistically significant as their p-values are less than 0.05. This means they have a significant impact on production. With each unit increase in x_4, production increases by 0.22 units, and with each unit increase in "storage of raw material (II)" (x_5), production increases by 0.14 units. Furthermore, according to the results, each unit increase in "storage of raw material (III)" (x_6) leads to a significant increase in production by 1.89 units. However, variables "third raw material purchasing quantity" (x_{15}) has a negative coefficient and is statistically significant, suggesting that an increase in x6 leads to a decrease in production, each unit increase will result in a decrease in production by 0.012 and 0.002 units.

Table 2 Results for production

Production	Coef.	Std. Err.	z	$P > \|z\|$	[95% Conf. Interval]	
x_1 |	97593.59	222695	0.44	0.661	-338881	534067 .7
x_2 |	-237101.3	524465.7	-0.45	0.651	-1265035	790832.6
x_3 |	-1534.438	3919.006	-0.39	0.695	-9215.55	6146.673
x_4 |	0.221107	0.012325	17.9	0	0.196951	0.245263
x_5 |	0.14155	0.01231	11.5	0	0.117423	0.165678
x_6 |	1.896452	0.137458	13.8	0	1.62704	2.165865
x_7 |	0.014487	0.033645	0.43	0.667	-0.05146	0.08043
x_8 |	0.006707	0.015286	0.44	0.661	-0.02325	0.036667
x_9 |	0.050431	0.114288	0.44	0.659	-0.17357	0.274432
x_{10} |	2.629571	5.923298	0.44	0.657	-8.97988	14.23902
x_{11} |	-2.462099	5.551389	-0.44	0.657	-13.3426	8.418423
x_{12} |	-0.265546	0.613631	-0.43	0.665	-1.46824	0.937148
x_{13} |	0.001009	0.001019	0.99	0.322	-0.00099	0.003006
x_{14} |	0.000255	0.000204	1.25	0.211	-0.00015	0.000655
x_{15} |	-0.012652	0.005155	-2.45	0.014	-0.02276	-0.00255
x_{16} |	-239563.6	545887	-0.44	0.661	-1309482	830355.3
x_{17} |	400218.4	897608.2	0.45	0.656	-1359061	2159498
x_{18} |	0	(omitted)				
x_{19} |	-0.00213	0.000782	-2.72	0.006	-0.00366	-0.0006
x_{20} |	0	(omitted)				
x_{21} |	0	(omitted)				
x_{22} |	0	(omitted)				
cons |	-33881.01	103491	-0.33	0.743	-236720	168957.7

Table 3 Results for sales

Sale	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
x_1 |	252562.3	225174.2	-1.12	0.262	-693896	188770.9
x_2 |	-605025	530300.7	1.14	0.254	-434345	1644396
x_3 |	-4668.31	3962.623	1.18	0.239	-3098.28	12434.91
x_4 |	0.192825	0.012238	-15.8	0	-0.21681	-0.16884
x_5 |	0.098772	0.012161	-8.12	0	-0.12261	-0.07494
x_6 |	1.73389	0.134552	-12.9	0	-1.99761	-1.47017
x_7 |	0.037457	0.03402	-1.1	0.271	-0.10414	0.029221
x_8 |	0.017453	0.015456	-1.13	0.259	-0.04775	0.012841
x_9 |	0.131243	0.115561	-1.14	0.256	-0.35774	0.095253
x_{10} |	6.742258	5.989259	-1.13	0.26	-18.481	4.996474
x_{11} |	-6.31508	5.613214	1.13	0.261	-4.68662	17.31677
x_{12} |	-0.69629	0.620463	1.12	0.262	-0.51979	1.912379
x_{13} |	-0.00048	0.001027	0.47	0.638	-0.00153	0.002494
x_{14} |	0	(omitted)				
x_{15} |	0	(omitted)				
x_{16} |	-610217	551967.5	1.11	0.269	-471620	1692053
x_{17} |	1028013	907600.1	-1.13	0.257	-2806877	750850.1
x_{18} |	0.000448	0.000206	-2.18	0.03	-0.00085	-4.4E-05
x_{19} |	-0.00143	0.000738	1.94	0.043	-1.7E-05	0.002876
x_{20} |	0	(omitted)				
x_{21} |	-0.02321	0.005192	4.47	0	0.013034	0.033385
x_{22} |	0	(omitted)				
cons |	-99128.2	104646.4	0.95	0.344	-105975	304231.3

Based on the results of **Table 3**, variables $x_4, x_5, x_6, x_{18}, x_{19}$ and x_{21} are statistically significant as their p-values are less than 0.05. This means they have a significant impact on sales; accordingly, with each unit increase in "storage of raw material (I)" (x_4), sales increase by 0.19 units, and with each unit increase in "storage of raw material (II)" (x_5), sales decrease by 0.098 units. Additionally, with each unit increase in "shelf Price" x_{19} and "consumer price index" x_{21}, sales decrease by 0.001 and 0.02 units, respectively. On the other hand, each unit increase in x_{18} leads to a sales increase of 0.023 units. (p-value < 0.05). The result of monitoring the fitted values obtained by the Hotelling T^2 control chart is shown in **Figure 3**.

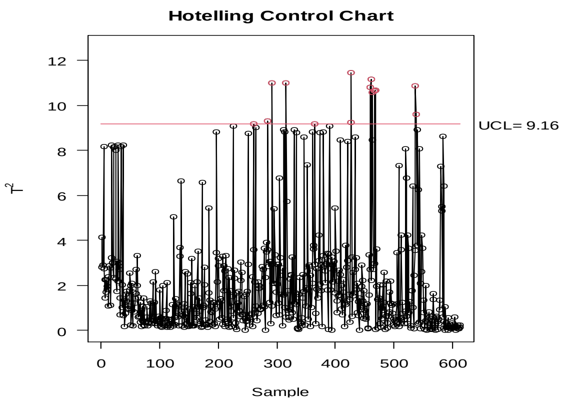


Figure 3 The Hotelling control chart

As shown in the chart, there are 14 samples in the company's supply chain network that are out of control (Sample 260, 285, 292, 316, 365, 426, 427, 460, 461, 464, 467, 468, 537, 538). By analyzing the data, we observed that three products are repeatedly identified as inefficient in the outputs. Therefore, these items are eliminated for further examination. The performance of the remaining eleven sample products should be improved.

The correlation chart of each of the two response variables versus the significant variables is shown in **Figure 4**. In this figure, there is a correlation between the explanatory variables and sale and production. Therefore, the existence of a relationship between the explanatory variables and Y_1 and Y_2 is intuitively confirmed. In the next step, we used this interpretation for applying Joint Optimization plot in rendering optimal values.

4.2 *Applying Joint Optimization to Render Optimal Values*

The analysis of "Hotelling T^2 Control Chart" reveals that 11 products within the supply chain network are experiencing inefficiencies and are out-of-control. To gain a deeper understanding of these problematic products, we conducted an investigation using the *Profit and Loss (P&L)* statements. The P&L statements indicate that each of the 11 out-of-control products incurred losses in at least one month during this 12-month interval. Consequently, transforming these losses into profitable outcomes is crucial for improving the overall performance of the supply chain network.

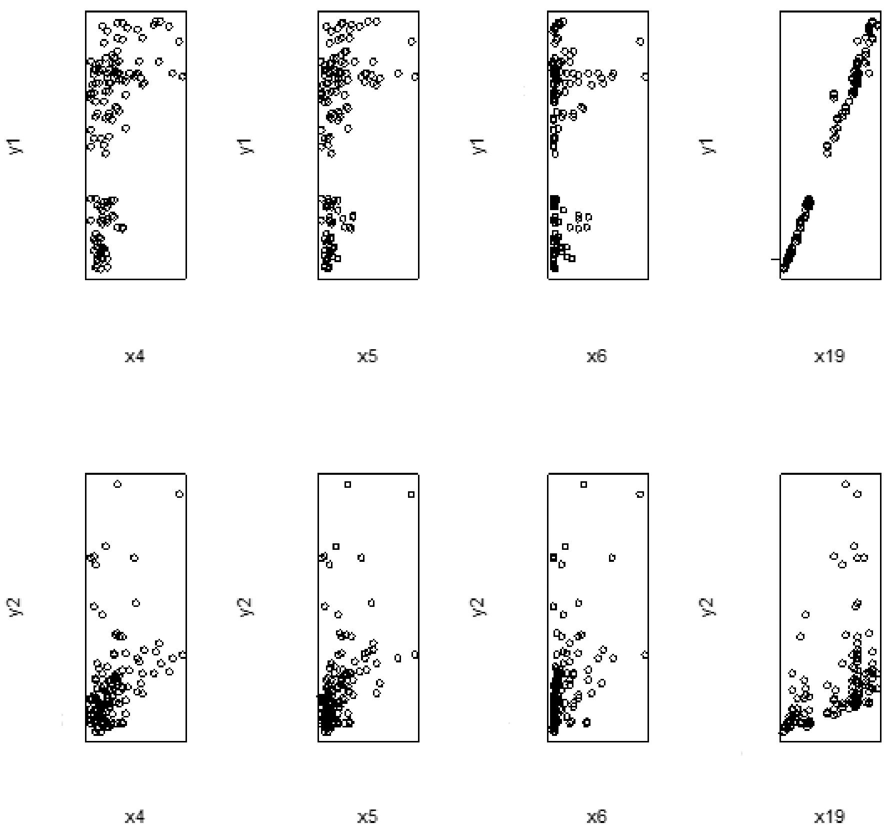


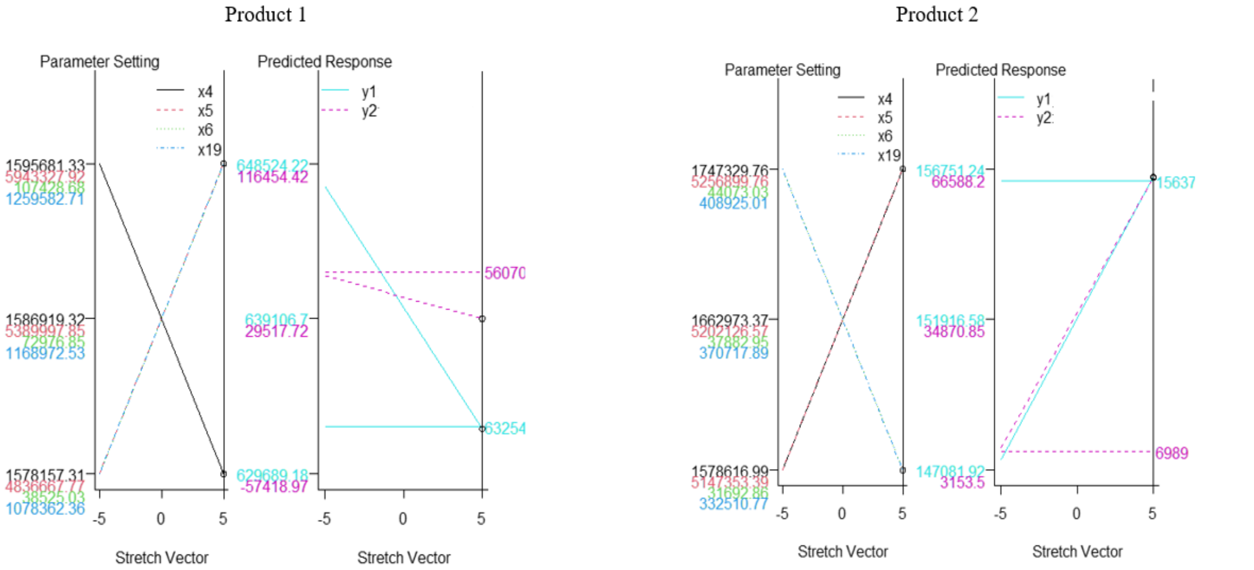
Figure 4 Correlation chart of significant variables versus response variables

Moreover, after a meticulous analysis of the data and the influence of variables on the supply chain during preceding stages, a clear revelation emerged: the storage of raw materials (I, II, and III) and the shelf price are identified as critical variables, bearing a substantial impact on the company's operations, performance, and overall success.

Given that these factors can be controlled by the company, there is an opportunity for their enhancement. To do so, the Joint Optimization modeling is employed to derive new decision variables that boost the company's performance in generating profits.

The simultaneous optimization plot illustrates the results of optimizing multiple responses graphically. In this plot, the optimization of control factors (variables) is shown in one graph, and the corresponding estimated responses aligned with the desired optimum are displayed in another graph (Kuhnt, 2004). As an example, consider Product 1, which incurred losses in the fifth month of sales. By applying the Joint Optimization model, careful adjustments and fine-tuning are made to the values of critical variables throughout each twelve-month interval. These optimized decision variables are precisely customized to ensure profitability rather than loss.

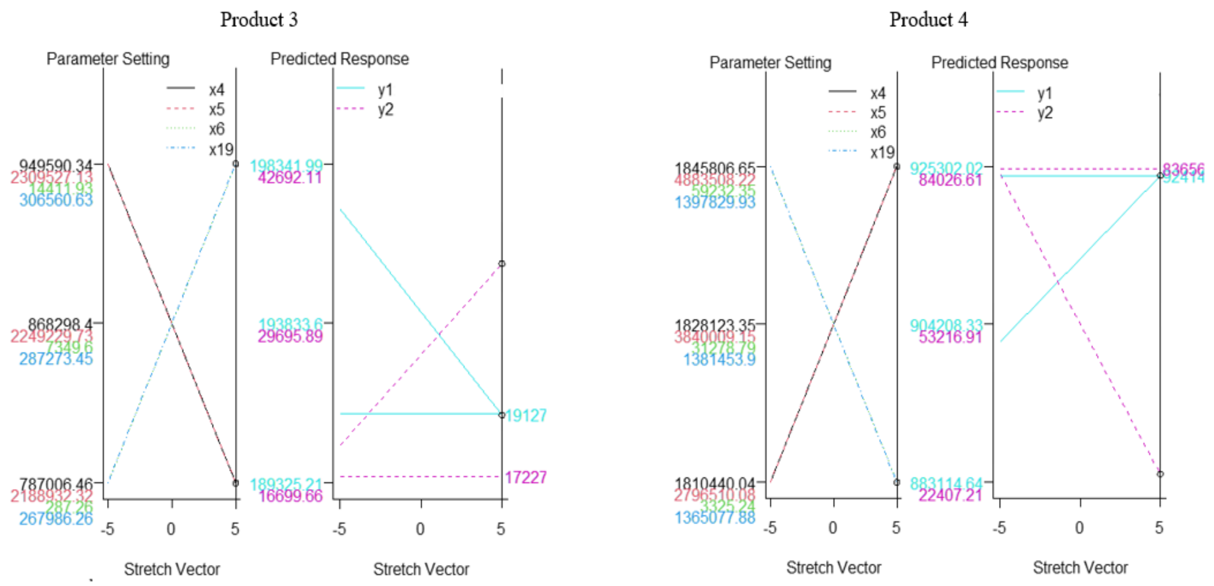
As a result, effectively managing the storage of raw materials and setting appropriate shelf prices can lead to greater cost efficiency, improved revenue streams, and overall success in the competitive market landscape. **Figures 5 to 10** show results for each product.



Optimal Values of Variables				Optimal Values of Responses	
X_4	X_5	X_6	X_{19}	Y_1	Y_2
1578157.3	5943327.9	107428.7	1259582.7	632451.62	29517.72

Optimal Values of Variables				Optimal Values of Responses	
X_4	X_5	X_6	X_{19}	Y_1	Y_2
1747329.76	5256899.76	31692.86	332510.77	156518.69	64864.55

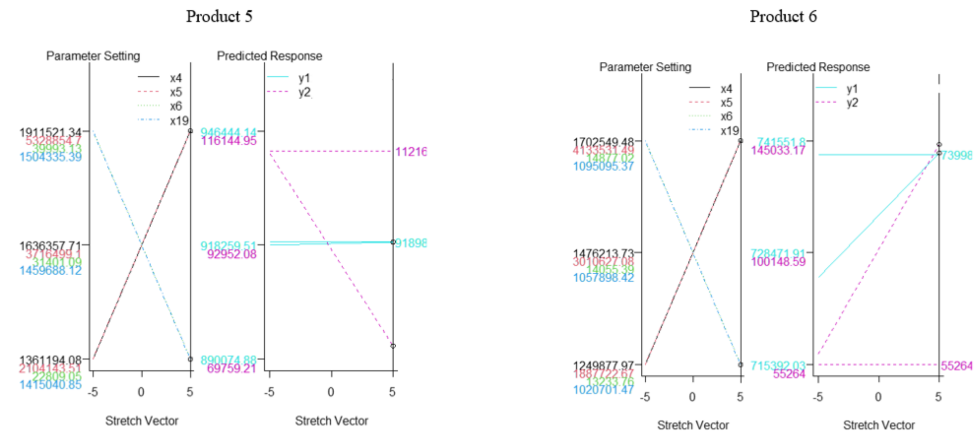
Figure 5 Joint optimization plots for product 1 and 2



Optimal Values of Variables				Optimal Values of Responses	
X_4	X_5	X_6	X_{19}	Y_1	Y_2
787006.46	2188932.32	14411.93	306560.63	191246.33	34566.37

Optimal Values of Variables				Optimal Values of Responses	
X_4	X_5	X_6	X_{19}	Y_1	Y_2
1845806.651	4883508.222	3325.239	1365077.877	924085.20	24034.37

Figure 6 Joint optimization plots for product 3 and 4



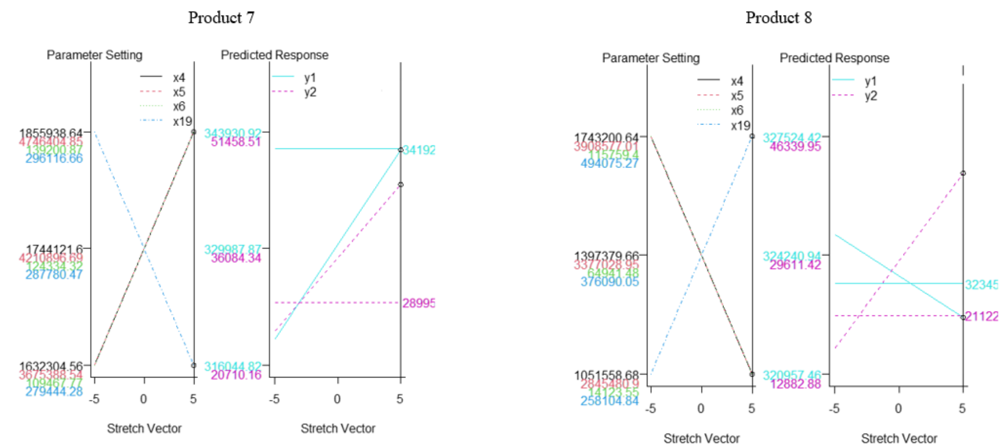
Product 5

Optimal Values of Variables				Optimal Values of Responses	
X_4	X_5	X_6	X_{19}	Y_1	Y_2
1911521.34	5328854.70	22809.05	1415040.85	919001.22	72457.67

Product 6

Optimal Values of Variables				Optimal Values of Responses	
X_4	X_5	X_6	X_{19}	Y_1	Y_2
1702549.48	4133531.49	13233.76	1020701.47	740124.7	143583.7

Figure 7 Joint optimization plots for product 5 and 6



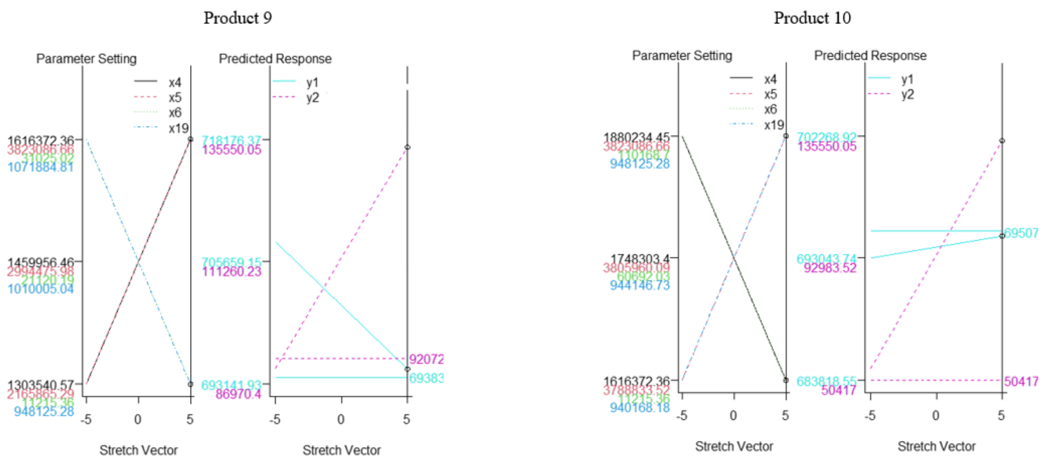
Product 7

Optimal Values of Variables				Optimal Values of Responses	
X_4	X_5	X_6	X_{19}	Y_1	Y_2
1855938.6	4746404.8	139200.9	279444.3	341750.14	44521.19

Product 8

Optimal Values of Variables				Optimal Values of Responses	
X_4	X_5	X_6	X_{19}	Y_1	Y_2
1051558.68	2845480.90	14123.55	494075.27	494075.27	41172.41

Figure 8 Joint optimization plots for product 7 and 8



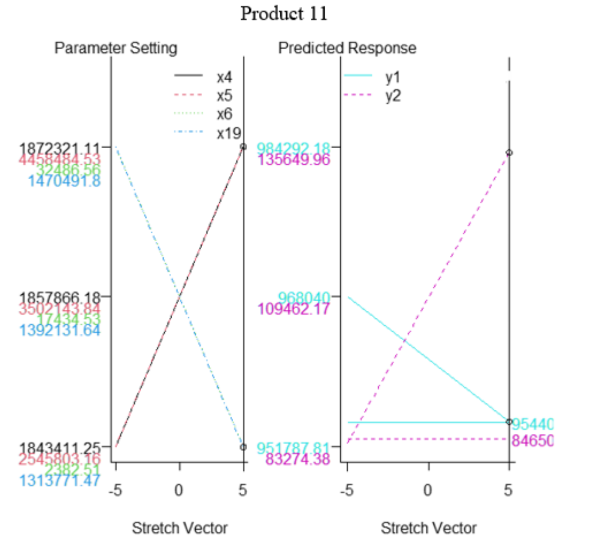
Product 9

Optimal Values of Variables				Optimal Values of Responses	
X_4	X_5	X_6	X_{19}	Y_1	Y_2
1616372.36	3823086.66	11215.36	948125.28	694692.1	134070.2

Product 10

Optimal Values of Variables				Optimal Values of Responses	
X_4	X_5	X_6	X_{19}	Y_1	Y_2
1616372.36	3823086.66	11215.36	948125.28	694692.1	134070.2

Figure 9 Joint optimization plots for product 9 and 10



Optimal Values of Variables				Optimal Values of Responses	
X_4	X_5	X_6	X_{19}	Y_1	Y_2
1872321.108	4458484.526	2382.511	1313771.473	954517.5	134657.9

Figure 10 Joint optimization plots for product 11

Optimal results for x_4, x_5, x_6, and x_{19} are obtained, and the operational profit is recalculated to assess improvements since the last performance evaluation of these 11 products in the supply chain network. The **Table 4** summarizes the performance of these products based on the new optimal values for the determined variables, comparing their previous operational profit to the optimized one:

Table 4 Operational profit for inefficient products

Product Number	Previous Operation Profit	Optimized Operation Profit	Deviation
1	(373,381)	25,619	399,000
2	(31,877)	45,723	77,600
3	(70,080)	2,640	72,720
4	(127,859)	182,141	310,000
5	(458,744)	6,256	465,000
6	(15,948)	216,552	232,500
7	(178,771)	13,529	192,300
8	(134,055)	28,245	162,300
9	(217,158)	118,856	336,014
10	(130,906)	32,994	163,900
11	(96,140)	88,000	184,140

Regarding the outputs, we can conclude that effectively managing the storage of raw materials and setting appropriate shelf prices can lead to cost efficiency, improved revenue streams, and overall success in the competitive market landscape. Emphasizing this approach as part of the company's supply chain strategy is crucial for sustainable growth and continued success.

5. DISCUSSION

This study presents a novel prescriptive model designed to improve supply chain networks by monitoring and analyzing the relationships between influential variables across the entire network. Our comprehensive method integrates data from all five main components of the supply chain—manufacturing, warehousing, procurement, logistics/transportation, and demand management—over a 12-month period. This integration allows for a thorough examination of the factors impacting sales and production, leading to the identification and optimization of out-of-control products affecting the profitability of company and leading to loss in financial year statement.

The findings of this study align with and extend the current literature on supply chain monitoring and analytics. For instance, Singh. (2024) highlighted the notable relationship between different layers of supply chain and the importance of flexibility in multiple aspects of supply chain to build resilience against a range of uncertainties. Our study validates this fact by demonstrating how integrated data from various supply chain components can provide a holistic view that is crucial for identifying inefficiencies and opportunities for optimization.

Taleizadeh *et al.*'s (2023) findings on the 'Demand' component also support our results. They employed developed mathematical models and data analytics to perform sensitivity analysis to understand the impact of shelf price and production costs on the supply chain network. Their findings revealed that manufacturers and retailers can adopt mixed strategies through making counterbalance between influential factors to reach their target profit. Similarly, our analysis of price and procurement variables underscored their significant impact on overall supply chain performance. It is important to note that in our study, the usage of longitudinal data and advanced statistical methods like the Generalized Estimating Equations (GEE) provided deeper insights into these relationships. Our proposed method can go beyond the capabilities of traditional cross-sectional studies and create a path to make a joint optimization model for reaching profit with changing different influential factors affecting the entire supply chain network.

Moreover, we utilized the Hotelling T^2 as a multivariate control chart in our study to effectively monitor and control product performance. This critical toll has been scarcely used in the literature on supply chain analytics and is essential for maintaining the stability and efficiency of the supply chain network. Finally, our Joint Optimization Plot (JOP) optimize critical variables across multiple responses, we demonstrated that the JOP can be an efficient tool for enhancing supply chain performance, particularly for inefficient products in the supply chain network, which directly affect the company's profitability.

The optimization of influential variables leads to substantial cost efficiencies and improved revenue streams. For instance, the recalibration of raw material storage levels and the adjustment of shelf prices have been shown to directly enhance operational profits, transforming previously unprofitable products into profitable ones. By focusing on key cost drivers and aligning them with optimal operational strategies, businesses can achieve higher financial returns and maintain a competitive edge in the market. Our approach also supports environmental sustainability by minimizing waste. Effective inventory management reduces overstocking and understocking, thereby decreasing the likelihood of excess materials becoming obsolete.

Additionally, optimized transportation logistics lower the carbon footprint.

The main advantage of our approach is that it helps companies to proactively manage resources while saving money, reducing the adverse environmental impacts, and meeting their consumers' needs. Unlike reactive strategies that address issues only after they occur, our preemptive approach allows for early detection and correction of potential problems. By continuously monitoring supply chain performance using advanced statistical methods such as the Generalized Estimating Equations (GEE) and Hotelling T^2 control charts, managers can identify out-of-control products and variables before they lead to significant disruptions or losses within the supply chain network. This proactive approach not only ensures smoother operations but also builds resilience against uncertainties and risks inherent in supply chain management.

This study makes a significant contribution to the literature by integrating advanced statistical methods and comprehensive data analysis across all five main components of the supply chain network. Our model results in a win-win strategy that benefits all stakeholders, including managers, decision-makers, and investors. This holistic approach not only identifies key areas for improvement but also provides actionable insights for optimizing supply chain operations, enhancing both efficiency and profitability, promoting sustainability, and mitigating risks.

6. CONCLUSION

The supply chain analytics can enhance supply visibility and improve forecasting, lead to lower inventory levels and cost savings, and increase overall efficiency. The primary contribution of this paper is to enhance multiple facets of the supply chain, encompassing storage levels and optimal shelf prices, to maximize profits. This is achieved through the utilization of statistical methodologies that have not been extensively applied before in the literature. Unlike previous studies that focused on variables from only one or two components of supply chain networks, this study implements a comprehensive analysis across all five components of the supply chain network. By examining variables across all components, the optimal values for the most significant variables can be determined, thereby, enhancing the overall performance of the supply chain network. The method involves: 1) extracting data from the five main components of the supply chain over a twelve-month interval, 2) utilizing the GEE method along with Hotelling T^2 control chart to monitor product performance, 3) detecting any unusual or out-of-control behavior in the supply chain, 4) applying Joint Optimization modeling to the products that exhibited out-of-control behavior during the supply chain monitoring, 5) optimizing relevant variables derived from previous stages to find the optimal cost values, including the cost of goods and operational expenses.

This study examined a real-world Multi-Echelon Supply Chain Network with 51 personal care products to provide prescriptive supply chain analytics for enhancing its performance. Data is collected across all five supply chain components for a 12-month period based on 22 variables from the entire chain, from production to distribution centers and store shelves. As a result, the products exhibiting inefficiencies within the supply chain network undergo substantial improvement through the optimization of prices and the selection of optimal storage levels for raw materials during manufacturing which ensures profitability and enhances the company's supply chain performance. This strategic method led to a profit in their financial statement within a period that had previously incurred losses.

REFERENCES

Aamer, A., Eka Yani, L., & Alan Priyatna, I. (2020). Data analytics in supply chain management: Review of machine learning applications in demand forecasting. *Operations and Supply Chain Management: An International Journal*, 14 (1), 1-13. https://doi.org/10.31387/oscm0440281

Amellal, A., Amellal, I., Seghiouer, H., & Ech-Charrat, M. (2023). Improving lead time forecasting and anomaly detection for automotive spare parts with a combined CNN-LSTM approach. *Operations and Supply Chain Management: An International Journal*, 16 (2), 265-278. https://doi.org/10.31387/oscm0530388

Bersimis, S., Psarakis, S., & Panaretos, J. (2007). Multivariate statistical process control charts: An overview. *Quality and Reliability Engineering International*, 23, 517-543. https://doi.org/10.1002/qre.829

Brandenburg, M., & Seuring, S. (2011). Impacts of supply chain management on company value: Benchmarking companies from the fast-moving consumer goods industry. *Logistics Research*, 3, 233-248. https://doi.org/10.1007/s12159-011-0056-7

Carrière, I., & Bouyer, J. (2002). Choosing marginal or random-effects models for longitudinal binary responses: Application to self-reported disability among older persons. *BMC Medical Research Methodology*, 2 (1), 15. https://doi.org/10.1186/1471-2288-2-15

Chen, B. (2021). Data-driven inventory control with shifting demand. *Production and Operations Management*, 30, 1365-1385. https://doi.org/10.1111/poms.13326

Chiang, D. M.-H., Lin, C.-P., & Chen, M.-C. (2011). The adaptive approach for storage assignment by mining data of warehouse management system for distribution centers. *Enterprise Information Systems*, 5, 219-234. https://doi.org/10.1080/17517575.2010.537784

Choi, Y., Lee, H., & Irani, Z. (2016). Big data-driven fuzzy cognitive map for prioritizing IT service procurement in the public sector. *Annals of Operations Research*, 243(1-2), 1-30. https://doi.org/10.1007/s10479-016-2281-6

Chong, A. Y. L., Li, B., Ngai, E. W. T., Ch'ng, E., & Lee, F. (2016). Predicting online product sales via online reviews, sentiments, and promotion strategies. *International Journal of Operations & Production Management*, 36, 358-383. https://doi.org/10.1108/IJOPM-03-2015-0151

Dash, S. P., & Mohanty, S. (2018). Uncertain transportation model with rough unit cost, demand and supply. *Operations Research*, 55 (1), 1-13. https://doi.org/10.1007/s12597-017-0317-6

Favi, C., Campi, F., Mandolini, M., Martinelli, I., & Germani, M. (2021). Key features and novel trends for developing cost engineering methods for forged components: A systematic literature review. *International Journal of Advanced Manufacturing Technology*, 117, 2601-2625. https://doi.org/10.1007/s00170-021-07611-4

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2011). Applied longitudinal analysis. Wiley Series in Probability and Statistics. *John Wiley & Sons, Inc*. https://doi.org/10.1002/9781119513469

Gilvan, C. Souza., G. C. (2014). Supply chain analytics. *Business Horizons*, 57(5), 595-605. https://doi.org/10.1016/j.bushor.2014.06.004

Gómez, M. I., & Lee, D. (2023). Transforming food supply chains for sustainability. *Journal of Supply Chain Management*, 59(4), 79-92. https://doi.org/10.1111/jscm.12310

Hudnurkar, M., Deshpande, S., Rathod, U., & Jakhar, S. (2017). Supply chain risk classification schemes: A literature review. *Operations and Supply Chain Management: An International Journal*, 10(4), 182-199. https://doi.org/10.31387/oscm0290190

Jain, R., Singh, A. R., Yadav, H. C., & Mishra, P. K. (2014). Using data mining synergies for evaluating criteria at the pre-qualification stage of supplier selection. *Journal of Intelligent Manufacturing*, 25, 165-175. https://doi.org/10.1007/s10845-012-0684-z

Kapil, M., Gumte, P., Pantula, D., Soumitri, M., Miriyala, S., & Mitra, K. (2021). Data-driven robust optimization for handling uncertainty in supply chain planning models. *Chemical Engineering Science*, 246, 116889. https://doi.org/10.1016/j.ces.2021.116889

Khajehzadeh, M., Pazhuheian, F., Seifi, F., Noorossana, R., Ali, A., & Saeedi, N. (2022). Analysis of factors affecting product sales with an outlook toward sale forecasting in the cosmetic industry using statistical methods. *International Review of Management and Marketing*, 12(6), 55-63. https://doi.org/10.32479/irmm.13337

Khurana, M., & Kumar, D. (2017). The study of data analytics in inventory management. *International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions)*, Dubai, United Arab Emirates, 2017, 140-144. https://doi.org/10.1109/ICTUS.2017.8285993

Kim, B., & Oh, H. (2005). The impact of decision-making sharing between supplier and manufacturer on their collaboration performance. *Supply Chain Management*, Vol. 10 No. 3, pp. 223-236. https://doi.org/10.1108/13598540510606287

Kovács, G., & Falagara Sigala, I. (2021). Lessons learned from humanitarian logistics to manage supply chain disruptions. *Journal of Supply Chain Management*, 57, 41-49. https://doi.org/10.1111/jscm.12253

Kuhnt, S., & Rudak, N. (2013). Simultaneous optimization of multiple responses with the R package JOP. *Journal of Statistical Software*, 54(9), 1-23. https://doi.org/10.18637/jss.v054.i09

Kuhnt, S., & Erdbruegge, M. (2004). A strategy of robust parameter design for multiple responses. *Statistical Modelling*, 4 (4), 249-264. https://doi.org/10.1191/1471082X04st078oa

Li, L., Su, X., Wang, Y., Lin, Y., Li, Z., & Li, Y. (2015). Robust causal dependence mining in big data networks and its application to traffic flow predictions. *Transportation Research Part C: Emerging Technologies*, 58, 292-307. https://doi.org/10.1016/j.trc.2015.03.003

Mamonov, V., & Poluektov, V. (2021). The impact of transport costs on sales in supply chains. In Murgul, V., & Pukhkal, V. (Eds.), *International Scientific Conference Energy Management of Municipal Facilities and Sustainable Energy Technologies* 2019 (pp. 565-577). Springer, Cham. https://doi.org/10.1007/978-3-030-57450-5_69

Mohan, B. A., Harshavardhan, B., Karan, S., Mohammed, J., Shariff, P., & Pranav, M. G. (2021). Demand forecasting and route optimization in the supply chain industry using data analytics. *Asian Conference on Innovation in Technology*, Pune, India, 2021, 1-7. https://doi.org/10.1109/ASIANCON51346.2021.9544942

Montgomery, D. C. (2019). Introduction to statistical quality control (8th ed.). *John Wiley & Sons*.

Mori, J., Kajikawa, Y., Kashima, H., & Sakata, I. (2012). Machine learning approach for finding business partners and building reciprocal relationships. *Expert Systems with Applications*, 39, 10402-10407. https://doi.org/10.1016/j.eswa.2012.01.202

Nguyen, T. (2023). Applications of artificial intelligence for demand forecasting. *Operations and Supply Chain Management: An International Journal*, 16 (4), 424-434. https://doi.org/10.31387/oscm0550401

Nguyen, T., Zhou, L., Spiegler, V., Ieromonachou, P., & Lin, Y. (2018). Big data analytics in supply chain management: A state-of-the-art literature review. *Computers & Operations Research*, 98, 254-264. https://doi.org/10.1016/j.cor.2017.07.004

Nitin, K., Singh, K.-h., Lai, K.-h., & Zhang, J. (2023). Intellectual core in supply chain analytics: Bibliometric analysis and research agenda. *International Journal of Information Technology and Decision Making*, 1-29. https://doi.org/10.1142/S0219622023300021

Pignatiello Jr., J. (1993). Strategies for robust multiresponse quality engineering. *IIE Transactions*, 25 (3), 5-15. https://doi.org/10.1080/07408179308964286

Salehan, M., & Kim, D. J. (2016). Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics. *Decision Support Systems*, 81, 30-40. https://doi.org/10.1016/j.dss.2015.10.006

Singh, R. K. (2024). Strengthening resilience in supply chains: The role of multi-layer flexibility, supply chain risks and environmental dynamism. *The International Journal of Logistics Management*. Advance online publication. https://doi.org/10.1108/IJLM-08-2023-0337

Soosay, C. (2023). Coordination in supply chains. In Sarkis, J. (Ed.), *The Palgrave Handbook of Supply Chain*

Management. Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-030-89822-9_57-1

Sukha, C. S., & Prabhu, S. (2023). Framework in supply chain using inventory management, big data and inventory analytics. *IEEE 12th International Conference on Communication Systems and Network Technologies* Bhopal, India, 2023, 549-553. https://doi.org/10.1109/CSNT57126.2023.10134627

Suwignjo, P., Panjaitan, L., Baihaqy, A., & Rusdiansyah, A. (2023). Predictive analytics to improve inventory performance: A case study of an FMCG company. *Operations and Supply Chain Management: An International Journal*, 16 (2), 293-310. https://doi.org/10.31387/oscm0530390

Taleizadeh, A. A., Aboutorabian, F., Moradi, R., Moon, I., & Mohamadi, N. (2023). Pricing, warranty, and shelf space decisions for the supply chain with non-symmetric market and warranty-period dependent demand. *Operations Research International Journal*, 23, 61. https://doi.org/10.1007/s12351-023-00803-8

Ten Have, Thomas R. (1995). [Review of Analysis of Longitudinal Data, by P. J. Diggle, K.-Y. Liang, & S. L. Zeger]. *Journal of the American Statistical Association*, *90*(431), 1123–1124. https://doi.org/10.2307/2291352

Tracy, N. D., Young, J. C., & Mason, R. L. (1992). Multivariate control charts for individual observations. *Journal of Quality Technology*, 24 (2), 88-95. https://doi.org/10.1080/00224065.1992.12015232

Wang, J., Swartz, C. L. E., & Huang, K. (2023). Data-driven supply chain monitoring using canonical variate analysis. *Computers & Chemical Engineering*, 174, 108228. https://doi.org/10.1016/j.compchemeng.2023.108228

Wang, J., Swartz, C. L. E., Corbett, B., & Huang, K. (2020). Supply chain monitoring using principal component analysis. *Industrial & Engineering Chemistry Research*, 59 (27), 12487-12503. https://doi.org/10.1021/acs.iecr.0c01038

Wang, Y. (2010). Imprecise probabilities based on generalized intervals for system reliability assessment. *International Journal of Reliability & Safety*, 4 (4), 319-342. https://doi.org/10.1504/IJRS.2010.035572

Wu, L., Liu, W., Yi, G. Y., & Huang, Y. (2012). Analysis of longitudinal and survival data: Joint modeling, inference methods, and issues. *Journal of Probability and Statistics*, 640153. https://doi.org/10.1155/2012/640153

Zhang, Y., Ren, S., Liu, Y., & Si, S. (2017). A big data analytics architecture for cleaner manufacturing and maintenance processes of complex products. *Journal of Cleaner Production*, 142 (2), 626-641. https://doi.org/10.1016/j.jclepro.2016.07.123

Zhao, R., Liu, Y., Zhang, N., & Huang, T. (2017). An optimization model for green supply chain management by using a big data analytic approach. *Journal of Cleaner Production*, 142, 1085-1097. https://doi.org/10.1016/j.jclepro.2016.03.006

Ziegler, A., Kastner, C., & Blettner, M. (1998). The generalized estimating equations: An annotated bibliography. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 40 (2), 115-139. https://doi.org/10.1002/(SICI)1521-4036(199806)40:2<115::AID-BIMJ115>3.0.CO;2-6

Mohammad Khajehzadeh got his master’s degree in industrial engineering from Iran University of Science and Technology, Iran. Currently, he is pursuing a second master’s degree in Epidemiology at the Ludwig Maximilian University (LMU) of Munich, Germany. His research interests include applying optimization and statistical learning models in practical fields, especially in business analysis to enhance performance.

Farhad Pazhuheian is currently a senior expert in systems and methods. He received his bachelor’s degree in statistics from Shiraz University, Iran, and his master’s degree in industrial engineering from Iran University of Science and Technology, Iran. His primary research interests include statistical process control, process optimization and statistical analysis

Farima Seifi got her bachelor’s degree in Renewable Energy Engineering from the Islamic Azad University, Science and Research Branch, Tehran, Iran, with a keen interest in solar panels and sustainable technologies. Currently, she is an MBA student at The University of California, Los Angeles, USA with research interests in business development and business analytics.

Alireza Ghorbani got his master's degree in Statistics from the Shahid Beheshti University, Tehran, Iran. Currently, he is pursuing a master's degree in Epidemiology at the Ludwig Maximilian University (LMU) of Munich, Germany. His research interests include applying Bayesian statistics and statistical learning models in high-dimensional environments.

Golshan Madraki, Ph.D., is currently a tenured Associate Professor of Supply Chain Management in the College of Business at the University of New England (UNE), Maine, USA. Before joining UNE, she was an Associate Professor of Engineering and Management in the Reh School of Business and the Associate Director of the Honors Program at Clarkson University. She received her Ph.D. in Industrial and Systems Engineering from Ohio University in 2017, and her B.S. and M.S. degrees in Operations Management from Allameh Tabataba’i University, Tehran, Iran, in 2010 and 2013, respectively. Her research focuses on various applications of algorithms, data analytics, graph theory, operations research, and mathematical modeling in manufacturing systems, supply chain networks, and social media networks.