IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 13 October 2024, accepted 25 November 2024, date of publication 2 December 2024, date of current version 10 December 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3510175

RESEARCH ARTICLE

Analyzing One-Step and Multi-Step Forecasting to Mitigate the Bullwhip Effect and Improve Supply Chain Performance

ALPER SARICIOĞLU[1], MÜJDE EROL GENEVOIS[2], AND MICHELE CEDOLIN[3]
[1]Graduate School of Science and Engineering, Galatasaray University, 34349 Istanbul, Türkiye
[2]Faculty of Engineering and Technology, Galatasaray University, 34349 Istanbul, Türkiye
[3]College of Engineering and Technology, American University of the Middle East, Egaila 54200, Kuwait

Corresponding author: Alper Sarıcıoğlu (alper.saricioglu@gmail.com)

This work was supported by Galatasaray University under Project FOA-2022-1128.

ABSTRACT This study presents the results of an investigation into effectiveness of one-step and multi-step (h-step-ahead) forecasting methods in mitigating the Bullwhip Effect and improving supply chain performance within an order-up-to-level inventory control system. the Bullwhip Effect, a phenomenon in which small variations in consumer demand cause increasingly larger fluctuations upstream in the supply chain, presents significant challenges for inventory management and cost control. Traditional forecasting methods, such as Moving Average and Exponential Smoothing, have been extensively studied for their impact on supply chain performance. This study is among the first to introduce machine learning forecasting models, specifically Long Short-Term Memory and LightGBM, within this research domain, comparing their performance under various demand conditions, including autoregressive processes with and without seasonality, as well as the well-known M5 forecasting competition dataset. The results reveal that the multi-step-ahead forecasting capability of Long Short-Term Memory and LightGBM significantly outperforms one-step-ahead forecasting in reducing demand amplification, leading to improvements in key supply chain metrics such as order fulfillment rate, variance of inventory, and average end inventory across all autoregressive and M5 series. The findings demonstrate the superior accuracy and stability of machine-learning methods, particularly in scenarios with high demand autocorrelation, seasonality, and variability. These results provide new insights into the potential of advanced forecasting techniques to better manage supply chain variability and reduce the Bullwhip Effect, thereby offering valuable guidance for optimizing inventory control strategies.

INDEX TERMS Bullwhip effect (BWE), supply chain performance, machine learning forecasting, long short-term memory (LSTM), LightGBM, multi-step forecasting, order-up-to-level (OUT) inventory control, demand amplification, autoregressive processes, M5 competition dataset.

I. INTRODUCTION

The Bullwhip Effect (BWE), also known as demand-variance amplification, remains a significant challenge in contemporary supply chain management. The phenomenon, first conceptualized by Forrester in his system dynamics simulation analysis, occurs when a relatively modest increase in customer demand leads to a disproportionate escalation in production rates. This cascading effect leads to issues, such as excessive backorders, inflated inventory levels, and underutilized capacity at higher stages of the supply chain [1]. While the widely recognized supply chain role-playing game has shed light on the behavioral aspects that drive BWE [2], further research suggests that it is not influenced solely by behavioral biases. In fact, BWE often stems from rational inventory and production decisions that, when combined, magnify demand variance inadvertently and elevate overall supply chain costs [3].

The associate editor coordinating the review of this manuscript and approving it for publication was Asadullah Shaikh.

© 2024 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License.
For more information, see https://creativecommons.org/licenses/by/4.0/

The persistence of BWE across supply chains highlights the need for effective strategies to minimize its impact. Traditionally, businesses have relied on improving communication, reducing lead times (LT), and implementing demand-smoothing techniques to address this problem. However, these supply chain initiatives and practices continue to evolve and improve. A promising avenue lies in leveraging accurate demand-forecasting methods. Recent advances in machine learning (ML) techniques offer new opportunities to enhance demand forecasting accuracy, and consequently, supply chain performance.

In this study, we evaluated the ability of various forecasting methods to predict demand patterns and mitigate BWE. These methods range from Moving Average (MA) and Exponential Smoothing (ES) to special neural network structures designed for time-series forecasting, Long Short-Term Memory (LSTM), and the gradient boosting approach based on decision trees (LightGBM).

The rest of the paper is organized as follows: Section II provides a detailed review of the related literature, focusing on BWE, traditional forecasting techniques, and the emergence of ML methods in supply chain forecasting. Section III presents the methodology used in this study, including the generation of an autoregressive demand series and the simulation of inventory control systems. Section IV discusses the results and compares the performance of different forecasting methods. Finally, Section V concludes the paper and offers insights into the implications of the findings and potential directions for future research.

II. RELATED LITERATURE

Based on seminal work by Lee et al. [3] to the present day The order-up-to-level (OUT) inventory control framework, often coupled with autoregressive demand processes across various supply chain stages, has served as a critical testing ground for research. Dejonckheere et al. [4] demonstrated that the OUT policy induces BWE regardless of the forecasting method used, whether MA or ES. Chen et al. [5] illustrated how increased LT exacerbates BWE by increasing reliance on demand data in a two-stage supply chain using an MA. In a subsequent study, Chen et al. [6] examined both ES and MA to assess their performance in relation to BWE. Their findings indicated that negatively autocorrelated demand tends to amplify volatility, with MA offering a smoother amplification ratio than ES. Chandra and Grabis [7] compared the OUT approach with a material requirement planning (MRP)-based approach, revealing that h-step-ahead forecasting and the MRP approach exert a smoothing effect on BWE. Hosoda and Disney [8] expanded the scope by incorporating inventory variance into their study of a three-stage supply chain forecast using the minimum mean square error (MMSE). They concluded that end-to-end LT exacerbates BWE, not the supply chain level. Recently, Babai, et al. [9] evaluated various strategies for estimating the variance of lead time demand under stochastic conditions. Michna et al. [10] found that forecasting stochastic LTs not only fails to mitigate the BWE but also contributes to it. Hussain et al. [11] explored the ES and MMSE with different parameters, demonstrating that inventory variance is more influenced by LT than by order variance. Ma et al. [12] introduced a price sensitivity coefficient into OUT and ranked MMSE, ES, and MA based on their responsiveness to price changes.

As ML forecasting methods have demonstrated improvements over traditional methods [13], [14], [15], there has been growing interest in applying these advanced techniques in the supply chain context [16], [17]. Within the OUT system, Kourentzes [18] introduced a neural network-based methodology for forecasting intermittent demand. The study revealed that although neural networks may not consistently outperform traditional methods in terms of forecasting accuracy, they significantly enhance inventory metrics such as service levels without necessitating an increase in stockholding volumes. Trapero et al. [19] applied kernel density estimation and GARCH models to refine safety stock calculations by forecasting lead-time demand variance more accurately. Huber et al. [20] demonstrated that integrating ML techniques, such as Artificial Neural Networks (ANNs) and LightGBM, significantly enhances cost reduction and inventory performance, particularly when applied to large datasets. Similarly, Cao and Shen [21] addressed the newsvendor problem and its multi-period extension, showing that their ML approach leads to more accurate inventory decisions under nonstationary demand conditions than traditional methods. Oroojlooy et al. [22] employed deep learning techniques to address the newsvendor problem by integrating forecasting and inventory optimization without depending on the explicit demand distributions. In a comparative study by Babai et al. [23] evaluated neural network-based methods against traditional forecasting techniques for intermittent demand. The results indicate that these neural network methods not only outperform traditional methods but also improve forecast accuracy and inventory efficiency. Spiliotis et al. [24] investigated various probabilistic forecasting methods using a categorized dataset from the M5 competition organized according to demand patterns. Their study emphasized that no single forecasting method consistently outperformed the others across all demand pattern categories. More recently, Seyedan et al. [25] presented ensemble deep learning models that combine Multilayer Perceptron (MLP), LSTM, and 1D Convolutional Neural Network (1D-CNN) to capture diverse patterns in the data, thereby enhancing the forecasting accuracy. They further integrated a modified Economic Order Quantity (EOQ) model into the inventory optimization framework to determine the optimal order quantity that minimizes total inventory costs.

Despite the extensive literature assessing the performance of ML forecasting methods, studies specifically examining these methods in the context of BWE remain limited. Jaipuria and Mahapatra [26] integrated ML methods with statistical approaches within a base stock policy framework to estimate their impact on BWE and inventory variance. Their findings indicate that, whereas ML techniques enhance forecast

accuracy and reduce both BWE and inventory variance, they do not necessarily lead to a reduction in net inventory levels. Prakash et al. [27] demonstrated that artificial neural networks (ANN) significantly improve forecast accuracy over traditional methods, effectively reducing the BWE. Zhang et al. [28] developed a method to reduce the BWE by combining Variational Mode Decomposition (VMD) and Support Vector Machine (SVM) techniques. Their findings demonstrated that this VMD-SVM model effectively filtered out demand noise, thereby improving the accuracy of demand forecasts and reducing BWE.

The contributions of this study are summarized below:

- The literature on the interplay between BWE and other supply chain cost implications was extended by comparing traditional statistical methods, MA and ES, with advanced ML techniques, LSTM, and LightGBM. Autoregressive AR(1) series and AR(1) series with an additive seasonality component were generated to evaluate the forecast accuracy of these methods.
- A subset of the M5 competition series is incorporated to test the methods using real-world data. Inventory control was simulated using the OUT method, and various supply chain metrics were compared based on the forecasting results.
- The impact of seasonality and the autocorrelation coefficient on AR(1), volatility, and LT on both the AR(1) and M5 series were examined.
- Crucially, this study is the first to compare the impact of one-step-ahead versus h-step-ahead forecasts on BWE in the context of ML methods. The findings demonstrate that h-step-ahead forecasting significantly reduces BWE compared to one-step forecasts, offering new insights into the potential of ML to better manage supply chain variability.

III. RESEARCH METHODOLOGY

This study followed the methodology illustrated in Fig. 1. It consists of four key stages: Demand Generation, Forecasting, Simulation, and Post-hoc Analysis.

In the first stage, demand data were generated using an autoregressive process of order 1, AR(1), to simulate different demand conditions. The autocorrelation (ϕ), seasonality amplitude (A), and variance (σ) were adjusted to explore a range of possible scenarios. In addition, a subset of the M5 competition dataset introduced a real-world dimension into the analysis.

In the second stage, the generated and real-world demand data are forecast using both the traditional and ML methods. The traditional methods MA and ES are employed for one-step-ahead forecasts, which predict the demand for one period in the future. Two ML models, LSTM and LightGBM, were trained and tested for both one-step-ahead and h-step-ahead forecasts. For h-step-ahead forecasting, multiple lead times are considered, enabling a more comprehensive assessment of the forecast horizon impact on supply chain performance.

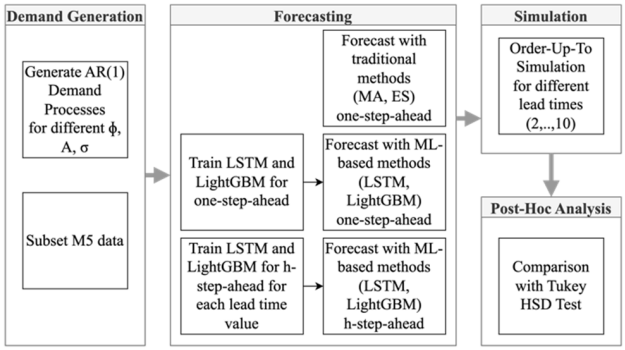


FIGURE 1. Methodology flow of the study.

In the third stage, the OUT method is simulated over lead times of two–ten periods to replicate the inventory control dynamics. The simulated inventory control system uses the forecasts generated in the previous step to replenish the stock based on the predicted future demand. The simulation aimed to replicate real-world supply chain dynamics and assess the performance of the forecasting models in terms of BWE, inventory level, inventory volatility, and service level. By adjusting the lead time, the simulation measures the effect of forecast accuracy over varying time horizons on the overall supply chain efficiency.

Finally, a post-hoc analysis was conducted using the Tukey Honest Significant Difference (HSD) test to statistically compare the performance of the different forecasting methods under each scenario. This analysis ensures that the differences observed in the supply chain metrics are statistically significant, providing a basis for evaluating the difference between one-step-ahead and h-step-ahead forecasting.

A. DEMAND SERIES

Autoregressive demand series were generated with varying parameters to facilitate the evaluation of forecasting methods. The AR(1) demand process is as follows:

$$D_t = \mu + \phi\,(D_{t-1} - \mu) + \epsilon_t \tag{1}$$

where D_t represents the value of the process at time t, μ denotes the mean of the process, and ϕ is the AR(1) coefficient, which captures the influence of the previous value on the current value. The term, ϵ_t represents white noise at time t, which is normally distributed with a mean of 0 and a variance of σ.

Different demand characteristics, including variations in autocorrelation coefficients and variances, were tested to evaluate the methods.

Fig. 2 illustrates a section of the time series for the two AR(1) processes. The series with $\phi = 0.0$ and $\sigma = 20$ and $\phi = 0.9$ and $\sigma = 10$ demonstrate how the different levels of autocorrelation impact the behavior of the time series under varying levels of noise σ.

An additive seasonality component is introduced to test how the forecasting methods handle different demand

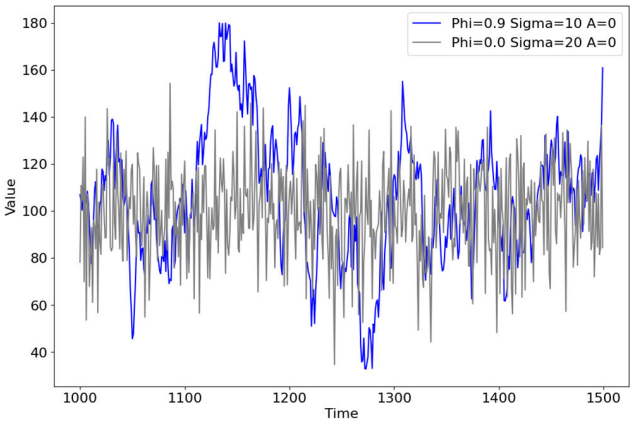


FIGURE 2. Selected time series of AR(1) processes for time steps between 1000 and 1500.

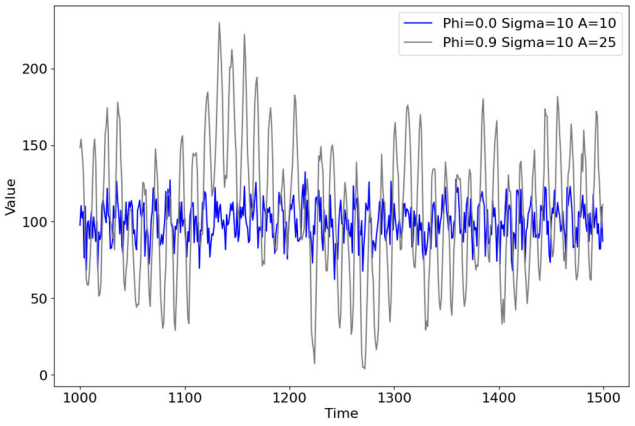


FIGURE 3. Selected time series of AR(1) processes with seasonality for time steps between 1000 and 1500.

patterns. As a standard approach for introducing deterministic seasonality [29], [30] trigonometric functions are employed as follows:

$$D_t = \mu + A \cdot sin\left(\frac{2\pi t}{P}\right) + \phi\left(D_{t-1} - \mu\right) + \epsilon_t \qquad (2)$$

The term $A \cdot \sin\left(\frac{2\pi t}{P}\right)$ introduces a sinusoidal fluctuation to the mean, where A represents the amplitude of the seasonality component and P denotes the period of the seasonality, corresponding to the number of time periods required for one complete cycle of the sine wave. This addition resulted in a recurring pattern every P period.

Fig. 3 illustrates the behavior of a section of AR(1) processes with different seasonality parameters over time. The AR(1) process with $\phi = 0.0$ and seasonality amplitude $A = 10$, showing a more stable and less oscillatory pattern contrast, the AR(1) process with $\phi = 0.9$ and amplitude $A = 25$, displaying more pronounced fluctuations owing to the higher amplitude and positive autocorrelation parameter.

The M5 series, derived from Walmart, one of the largest retail chains globally, captures actual sales data from Walmart

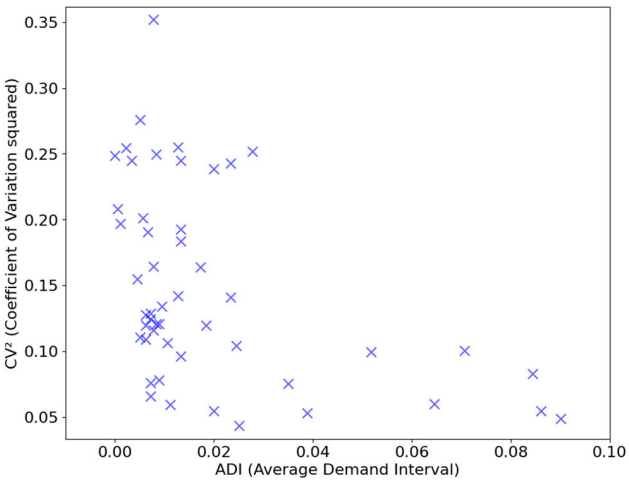


FIGURE 4. Scatter plot of ADI vs CV^2 for M5 series.

stores across the United States, and is used in this study as real-world data. The M5 series was categorized into four distinct classes: intermittent, lumpy, erratic, and smooth [31], [32]. Fig. 4 illustrates the relationship between the Average Demand Interval (ADI), which represents intermittency, and the Coefficient of Variation squared (CV^2), which represents erratic patterns for the time series selected for the study. Given that traditional forecasting methods such as MA and ES are not well suited for intermittent demand [18], and to accurately assess aggregated inventory performance within the dyadic nature of the simulation, 50 product series with non-intermittent demand patterns at the state level were selected.

B. FORECASTING TECHNIQUES

Different forecasting methods were selected based on their complementary strengths, allowing for a comprehensive evaluation of their effectiveness in predicting demand and managing supply chain variability.

1) MOVING AVERAGE (MA)

MA is a widely used time-series forecasting method that smooths fluctuations in data by averaging the values of a fixed number of consecutive time periods. The MA of order n at time t, denoted as MA_t, was calculated as follows:

$$MA_t = \frac{1}{n}\sum_{i=0}^{n-1} D_{t-i} \qquad (3)$$

where D_{t-i} represents the observed demand at time $t - i$, i and n is the number of periods considered in the MA. This method is particularly effective for data with minimal trends and seasonal components.

2) EXPONENTIAL SMOOTHING (ES)

The ES is another common time-series forecasting technique that assigns exponentially decreasing weights to past observations. Unlike MA, ES assigns more weight to recent data points, allowing the model to adapt more quickly to changes

in the underlying data. The forecast at time $t+1$, denoted as F_{t+1}, is given by

$$F_{t+1} = \alpha D_t + (1-\alpha) F_t \tag{4}$$

where α is the smoothing constant, $(0 < \alpha < 1)$, D_t is the observed demand at time t, and F_t is the forecast at time t. The choice of α determines the level of responsiveness to changes in the data, with higher values placing greater emphasis on recent observations.

3) LONG SHORT-TERM MEMORY (LSTM)

LSTM is a type of recurrent neural network (RNN) specifically designed to model sequential data with long-range dependencies. LSTM networks are particularly well suited for time-series forecasting because of their ability to capture complex temporal patterns. The core component of an LSTM network is the memory cell, which maintains a cell state c_t and is governed by three types of gates: the input gate i_t, forget gate f_t, and output gate o_t. These gates control the flow of information into and out of the memory cell, allowing the network to retain or discard information selectively over time.

The key equations for an LSTM cell at time (t) are as follows:

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_i\right) \tag{5}$$
$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{6}$$
$$o_t = \sigma\left(W_o \cdot [h_{t-1}, x_t] + b_o\right) \tag{7}$$
$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tanh\left(W_c \cdot [h_{t-1}, x_t] + b_c\right) \tag{8}$$
$$h_t = o_t \cdot \tanh\left(c_t\right) \tag{9}$$

where x_t is the input at time t, h_{t-1} is the hidden state from the previous time step; W_i, W_f, W_o, W_c are the weight matrices; b_i, b_f, b_o, b_c are the bias terms; and σ and *tanh* are the sigmoid and hyperbolic tangent activation functions, respectively.

4) LightGBM

LightGBM, or a Light Gradient Boosting Machine, is a powerful ML algorithm based on decision tree ensembles, specifically, gradient boosting. It is designed for high performance and efficiency, particularly for large datasets and complex feature spaces. LightGBM builds an ensemble of weak learners (decision trees) by sequentially fitting each new tree to the residual errors of the previous trees, thereby minimizing a specified loss function.

The objective of LightGBM is to minimize the following loss function over the dataset:

$$\text{Loss} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i\right) + \sum_{k=1}^{K} \Omega\left(f_k\right) \tag{10}$$

where $l\left(y_i, \hat{y}_i\right)$ is the loss between the observed value y_i and predicted value $\hat{y}_i$, $\Omega\left(f_k\right)$ is the regularization term to prevent overfitting, and K is the total number of trees in the model. The LightGBM is known for its ability to handle large datasets, support categorical features directly, and provide fast training times while maintaining high accuracy.

C. ORDER-UP-TO-LEVEL INVENTORY CONTROL

A widely used inventory management approach, the OUT method, determines the optimal inventory level that minimizes costs while satisfying the demand [33]. This method relies on forecasting future demand, determining the safety stock, and accounting for LT. This study implements a simulation of inventory control using the OUT under various LTs and different demand forecasts.

1) LEAD TIME DEMAND CALCULATION

The Lead Time Demand, D_{Lt}, is the total demand expected during lead time, LT. Depending on the approach used, this can be calculated by multiplying the forecasted demand F_t by L or summing the forecasted demand F_i, over the LT periods as follows:

$$D_{Lt} = \mathrm{F}_t * LT \tag{11}$$
$$D_{Lt} = \sum_{i=t}^{t+L-1} \mathrm{F}_i \tag{12}$$

The multiplication approach can be used for all forecasting models, as it requires one-step-ahead forecasting, where the model predicts the value for the immediate next period. The forecast was based on past data up to the current period, and the prediction was made only for the next time point.

The summation approach of LT has been shown to outperform the multiplication approach in smoothing order variance and reducing inventory costs [7], [9], [34]. However, because these studies utilize MMSE and ARIMA forecasting with autoregressive processes to generate h-step-ahead forecasts, they are limited in addressing situations where demand parameters are unknown. ML-based methods, however, warrant attention owing to their flexibility in overcoming these barriers. By adapting their output shape, ML models can forecast multiple steps ahead without requiring prior knowledge of demand parameters or using recursive forecasting. Therefore, our study focuses on evaluating these ML-based methods to address such challenges, which are discussed in Section IV-B.

2) SAFETY STOCK CALCULATION

Safety stock, SS, is an additional inventory buffer that accounts for demand variability during the LT. It is typically calculated based on service-level targets as follows:

$$SS = z \times \sigma_D \times \sqrt{\mathrm{LT}} \tag{13}$$

Equation (13) shows a typical SS level, where z is the safety factor (e.g., 1.96 for a 95% service level), σ_D is the standard deviation of demand, LT is the lead time. In this study, the safety stock was set to zero to directly measure the impact of LT [35].

3) ORDER-UP-TO LEVEL

The order-up-to level, S, determines the inventory position to which the inventory should be replenished and is calculated

as follows:

$$S_t = D_{L,t} + SS \tag{14}$$

With zero safety stock, this simplifies to LT Demand.

4) INVENTORY POSITION

The inventory position $\mathrm{IP_t}$, is the inventory balance of the system that sums inventory level I_{t-1} at the end of the previous period, inventory currently in transit IT_t, and inventory AI_t that has arrived during period t. The equation is as follows:

$$\mathrm{IP}_t = I_{t-1} + \mathrm{IT}_t - \mathrm{D}_t + \mathrm{A}_t \tag{15}$$

5) ORDER QUANTITY

The order quantity Q_t, at time t is the amount of inventory required to raise the inventory position $\mathrm{IP_t}$, to the OUT, S:

$$Q_t = \max\left(0, S_t - \mathrm{IP}_t\right) \tag{16}$$

Equation (16) shows that Q_t is a non-negative variable that does not return excess inventory without a cost.

D. SUPPLY CHAIN METRICS

1) AMPLIFICATION RATIO

The Amplification Ratio (AR) measures the degree to which the variability of orders placed by a retailer or distributor is amplified compared with the variability of actual demand. The typical method for quantifying BWE that involves determining the ratio of demand variances between the upstream and downstream levels [36], [37], [38] is as follows:

$$AR = \frac{\mathrm{Var}\left(Q_t\right)}{\mathrm{Var}\left(D_t\right)} \tag{17}$$

Equation (17) shows the AR, which quantifies the degree to which the variability in orders (Q_t) placed by a retailer or distributor exceeds the variability in the actual demand (D_t). A higher AR indicates a more pronounced BWE, reflecting greater amplification of demand variability across the supply chain.

2) VARIANCE OF INVENTORY

The variance of inventory, V_{I} quantifies the variability in inventory levels over time and can be represented by Equation (18) as follows:

$$V_{\mathrm{I}} = \frac{\mathrm{Var}\left(I_t\right)}{\mathrm{Var}\left(D_t\right)} \tag{18}$$

The variance of inventory is a key metric for evaluating the stability of inventory levels relative to demand fluctuations. A higher variance indicates a less stable system, leading to increased holding costs for more safety-stock needs [11].

3) ORDER FULFILLMENT RATE

The Order Fulfillment Rate (OFR) is the proportion of customer demand that is met without delay or backorder and can be expressed as follows:

$$B_t = max\left(0, D_t - I_t\right) \tag{19}$$

$$\mathrm{OFR_t} = \frac{D_t - B_t}{D_t} \tag{20}$$

If the backorder B_t is zero (19), then $\mathrm{OFR_t} = 1$, meaning that 100% of the demand is fulfilled. If backorders exist, $\mathrm{OFR_t}$ is less than 1 (20).

4) AVERAGE END INVENTORY

The average end inventory (22) refers to the average amount of inventory left at the end of each period, where $\mathrm{EI_t}$ is the inventory remaining at the end of the period and T is the total number of periods (21). The equations are expressed as follows:

$$\mathrm{EI_t} = max\left(0, \mathrm{EI_{t-1}} + \mathrm{AI_t} - D_t\right) \tag{21}$$

$$\text{Average EI} = \frac{\sum_{t=1}^{T} \text{End Inventory}_t}{T} \tag{22}$$

The average end inventory serves as an indicator of inventory efficiency, where higher values suggest overstocking and increased holding costs, whereas lower values may indicate a risk of stockouts and unmet demand. Maintaining an optimal average end inventory is crucial for balancing service levels with cost minimization in supply chain management.

IV. RESULTS

A. FORECAST ACCURACY

Forecast accuracy was measured using the Root Mean Squared Error (RMSE) and Symmetric Mean Absolute Percentage Error (sMAPE). The RMSE is preferred for ranking forecast performance, for the AR(1) series, while sMAPE is preferred for M5 series, as it allows for comparison across different scales, ensuring consistent evaluation of diverse data.

1) RESULTS FOR AR(1) SERIES WITHOUT SEASONALITY

The forecasting methods ranked by RMSE show that LSTM and LightGBM consistently outperform ES and MA across different values of ϕ and σ. When analyzing sensitivity to ϕ and σ, it is evident that, as ϕ increases, the RMSE generally tends to increase for MA and ES, reflecting their difficulty in handling positively autocorrelated processes. Conversely, LSTM and LightGBM exhibited more resilience to changes in ϕ, with only slight increases in RMSE, maintaining lower errors even when ϕ approached 0.9.

Fig. 5 shows that at $\phi = 0.0$ and $\sigma = 10$, the RMSE for LSTM is 10.26, which increases slightly to 10.83 at $\phi = 0.9$. LightGBM exhibits a similar trend, with RMSE values ranging from 10.44 at $\phi = 0.0$ to 10.21 at $\phi = 0.9$. In contrast, the RMSE for MA increased more significantly from 11.30 to 13.33 over the same range of ϕ, demonstrating its reduced effectiveness in handling higher autocorrelation.

As σ increased, the RMSE naturally grew rapidly for all methods, reflecting the increased variability in the demand process. For example, at $\phi = 0.0$ and $\sigma = 20$, the RMSE for LSTM increases to 20.82, whereas for LightGBM, it climbs to 20.88. Notably, ES and MA also showed sub-

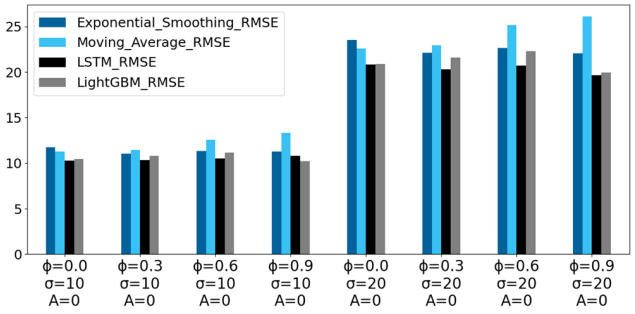


FIGURE 5. RMSE comparison of AR(1) series with no seasonality.

stantial increases, with RMSE values doubling from their $\sigma = 10$ counterparts.

Overall, LSTM and LightGBM demonstrated superior performance, particularly in scenarios with higher autocorrelation and variability, making them preferable choices for forecasting in such contexts.

2) RESULTS FOR AR(1) SERIES WITH SEASONALITY

The RMSE values in Fig. 6 demonstrate that while ES and MA struggle with increasing seasonality amplitude and autocorrelation, LSTM and LightGBM maintain relatively lower error rates. As ϕ increases from 0.0 to 0.9, a notable increase in RMSE is observed for both ES and MA, particularly at higher amplitudes.

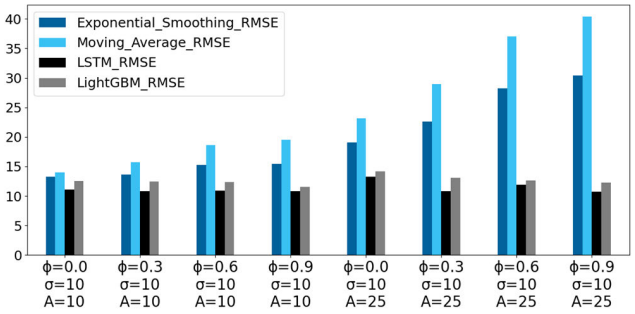


FIGURE 6. RMSE comparison of AR(1) series with seasonality.

At $A = 25$ and $\phi = 0.9$, the RMSE for ES increased to 30.42, and for MA, it reached 40.36, reflecting their limitations in capturing complex seasonal patterns combined with positive autocorrelation. In contrast, LSTM and LightGBM display more stable RMSE values under the same conditions, with LSTM showing only a slight increase from 13.25 at $\phi = 0.0$ and $A = 25$ to 10.71 at $\phi = 0.9$ and $A = 25$.

The sensitivity analysis further underscores the impact of increasing the seasonal amplitude A. At $\phi = 0.0$, the RMSE for LSTM increases from 11.13 at $A = 10$ to 13.25 at $A = 25$, whereas the RMSE rises from 12.53 to 14.22 under the same conditions. This trend continued as ϕ increased, with all methods experiencing higher RMSE values at larger amplitudes. However, the performance of LSTM and LightGBM remained more consistent, indicating their robustness in dealing with seasonal demand patterns.

Overall, LSTM and LightGBM were better suited for forecasting in environments characterized by seasonality autocorrelation, as evidenced by their lower RMSE values and greater resilience to changes in ϕ and A.

3) FORECAST RESULTS FOR M5 COMPETITION SERIES

Fig. 7 illustrates the distribution of sMAPE values for each forecast method across different coefficients of variation (CV) groups for the M5 series. LightGBM consistently exhibited the lowest sMAPE values across all CV groups, with an overall sMAPE of 20.07, reflecting its robustness in accurately capturing demand patterns, even in highly variable datasets. Its performance is particularly strong in the lower CV group ($0.2 < CV \leq 0.3$), where the sMAPE remains below 19, indicating superior forecasting accuracy in more stable environments.

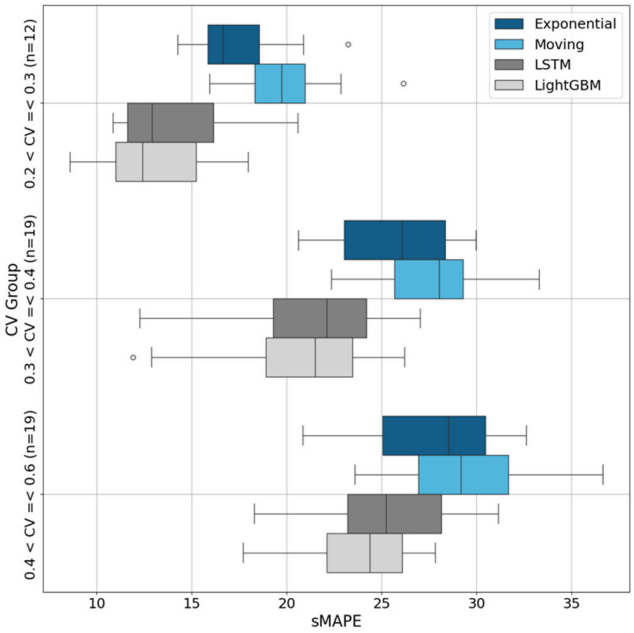


FIGURE 7. sMAPE comparison of M5 competition series.

LSTM also performed well, with an overall sMAPE of 21.07. While slightly higher than LightGBM, LSTM shows strong adaptability, especially in the mid-range CV group ($0.3 < CV \leq 0.4$), where it manages to maintain a low sMAPE around 20.5, demonstrating its effectiveness in handling moderate variability.

On the other hand, ES and MA showed significantly higher sMAPEs, 24.49 and 26.60, respectively. These methods struggle more with datasets that exhibit higher variability, particularly in ($0.4 < CV \leq 0.6$) CV group (27. The pronounced gap in sMAPEs between MA, ES, and ML-based approaches is still evident in the higher CV groups.

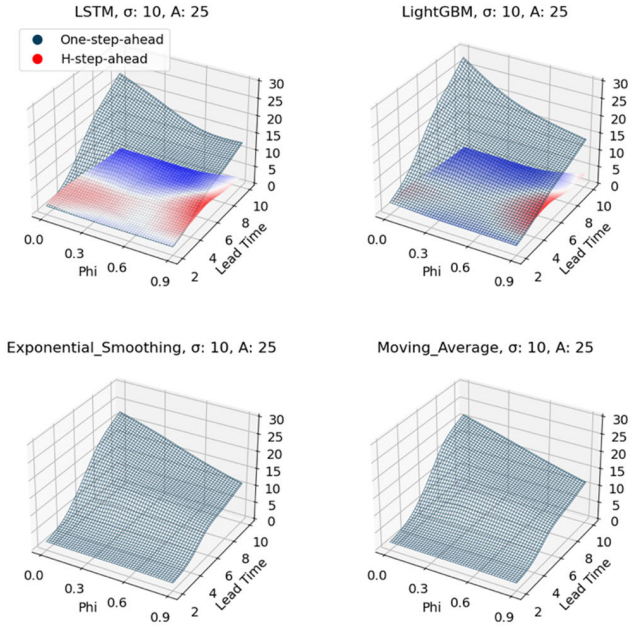


FIGURE 8. Amplification ratio versus ϕ and LT for AR(1) series.

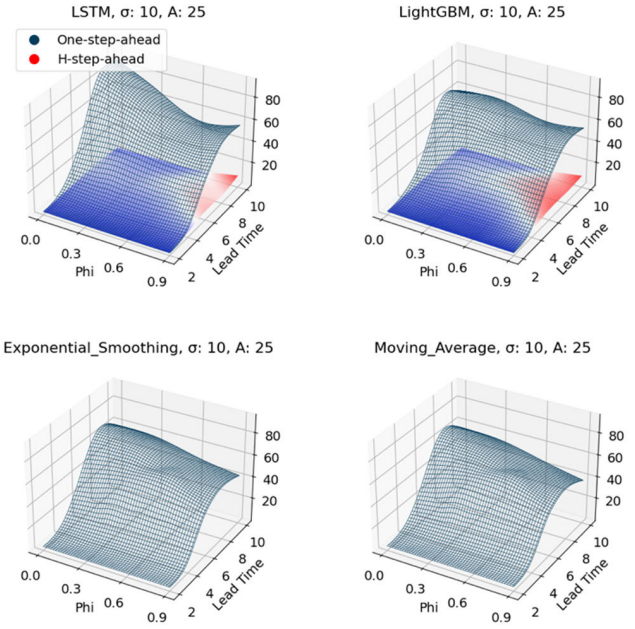


FIGURE 10. Variance of inventory versus ϕ and LT for AR(1) series.

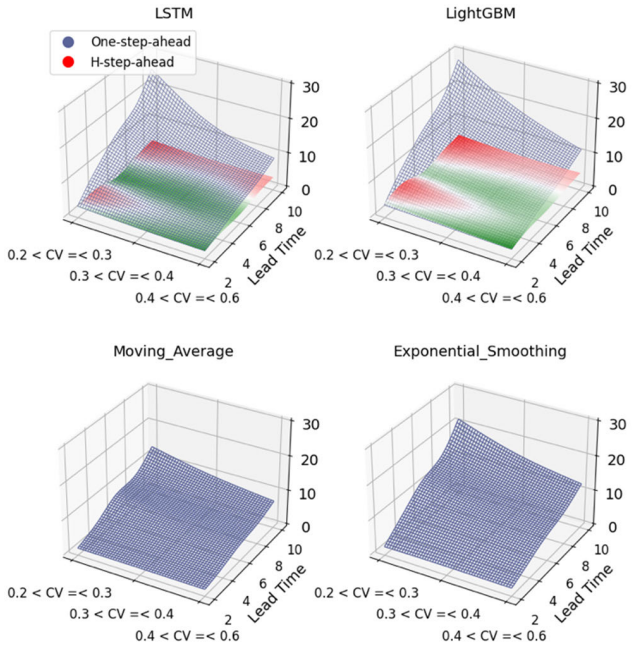


FIGURE 9. Amplification ratio versus CV and LT for M5 series.

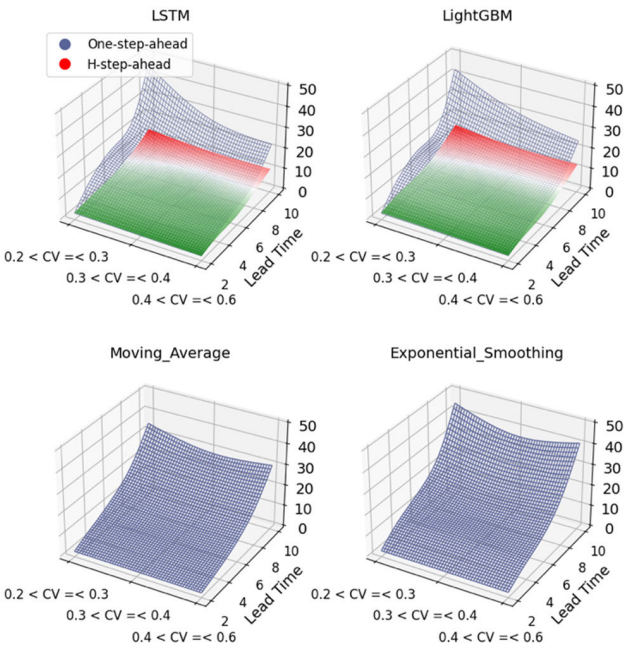


FIGURE 11. Variance of inventory versus CV and LT for M5 series.

B. COMPARISON OF ONE-STEP AND H-STEP AHEAD FORECASTING

This section presents a comprehensive comparison of forecasting methods, forecasting horizons, various demand conditions, and their effects on critical supply chain metrics across both AR(1) demand processes and the M5 dataset. The performance of each method for the corresponding metric is illustrated with plots and supported by statistical significance.

Tables 1 and 3 show the simulation results for both the AR(1) series and M5 dataset, respectively, under varying conditions of autocorrelation (ϕ), seasonality (A), demand variance (σ), and coefficient of variation (CV). They compared the inventory management performance of the corresponding forecast methods at lead times 2 and 10. Tables 2 and 4 present the results of the Tukey Honest Significant Difference (HSD) test for the AR(1) and M5 datasets, providing

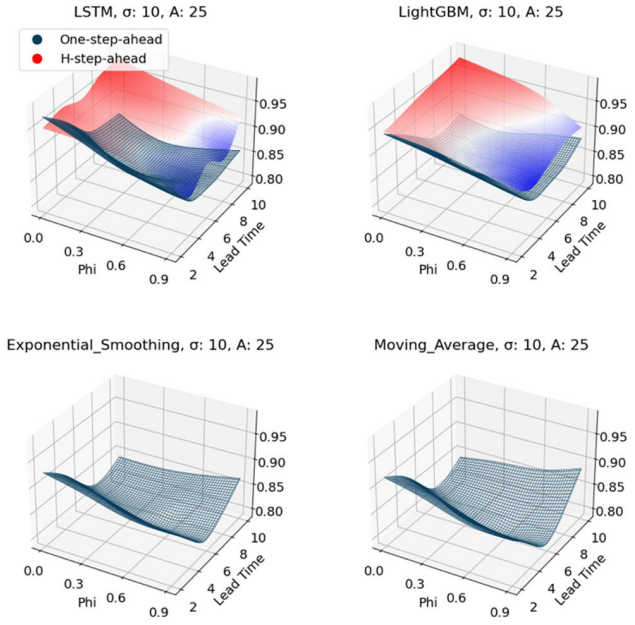


FIGURE 12. Average OFR versus ϕ and LT for AR(1) series.

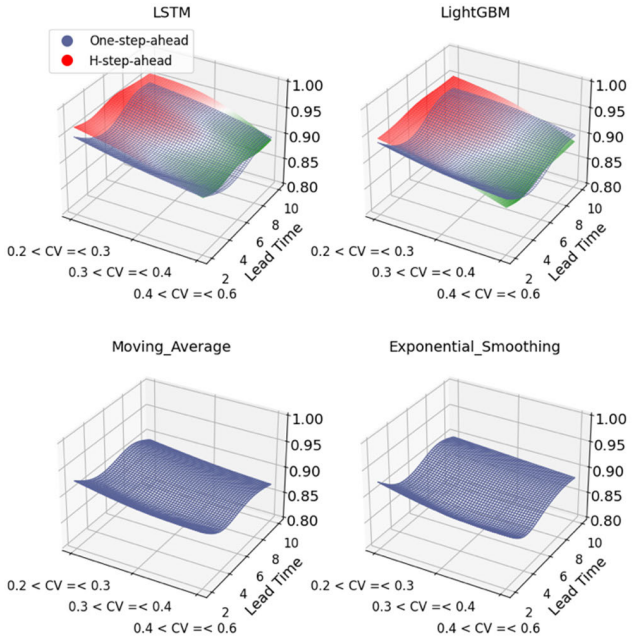


FIGURE 13. Average OFR versus CV and LT for M5 series.

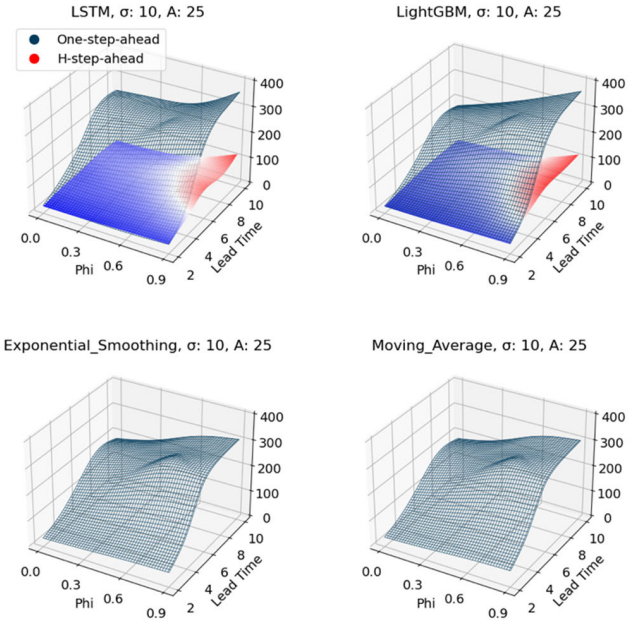


FIGURE 14. Average end inventory versus ϕ and LT for AR(1) series.

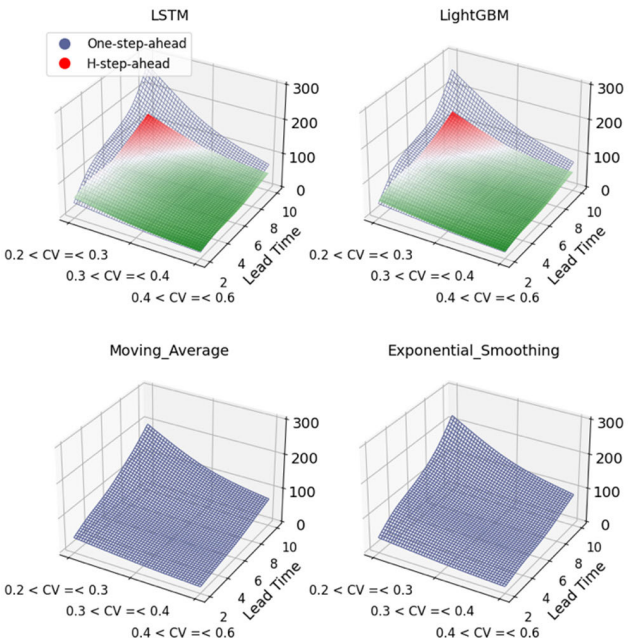


FIGURE 15. Average end inventory versus CV and LT for M5 series.

statistical validation of the performance differences between the forecasting methods.

1) AMPLIFICATION RATIO

The results indicate that, particularly at shorter lead times, the amplification ratios for LSTM and LightGBM were not significantly different from those of ES and MA.

As lead times increased, the differences between the ML-based methods and traditional methods became more apparent, although the improvements in the amplification ratio remained moderate. Specifically, both LSTM and LightGBM showed lower amplification ratios as lead times were extended, but the reduction was not as significant as might be expected given their superior forecast accuracy.

This suggests that while ML models provide better demand predictions, the traditional one-step-ahead structure limits their ability to fully mitigate BWE, which arises from forecast errors and demand variability amplification.

TABLE 1. Simulation results of AR(1) demand process for $\sigma = 10$; LT = 2 / LT = 10.

A	Method	Horizon	ϕ	AR	Avg. OFR	Var. of Inv.	Avg. Backorders	Avg. End Inv.
0.0	ES	One-step	0.0	2.21 / 33.04	0.97 / 0.94	1.75 / 36.16	3.66 / 6.40	11.14 / 70.23
			0.3	2.25 / 29.12	0.96 / 0.92	1.84 / 64.52	4.22 / 8.03	12.50 / 89.73
			0.6	1.76 / 23.53	0.96 / 0.91	1.44 / 71.55	4.68 / 9.94	13.71 / 108.91
			0.9	1.36 / 9.52	0.96 / 0.90	0.73 / 43.28	4.64 / 11.62	14.50 / 139.59
	LightGBM	H-step	0.0	2.02 / 2.70	0.97 / 0.98	1.32 / 6.77	3.34 / 2.30	10.29 / 27.34
			0.3	2.30 / 3.86	0.96 / 0.97	1.65 / 10.96	3.95 / 3.49	11.52 / 37.97
			0.6	2.51 / 6.30	0.95 / 0.94	1.45 / 18.90	5.49 / 6.71	13.44 / 54.07
			0.9	2.09 / 12.24	0.95 / 0.89	0.81 / 30.74	6.11 / 12.68	16.54 / 123.20
		One-step	0.0	2.24 / 20.42	0.97 / 0.96	1.45 / 13.74	3.48 / 4.11	10.80 / 47.07
			0.3	2.60 / 33.22	0.96 / 0.94	1.78 / 33.81	4.08 / 6.43	11.84 / 72.16
			0.6	2.72 / 28.69	0.95 / 0.91	1.56 / 43.85	5.51 / 9.43	14.08 / 97.19
			0.9	2.03 / 17.50	0.95 / 0.90	0.79 / 41.88	5.81 / 11.98	16.15 / 148.70
	LSTM	H-step	0.0	1.25 / 1.35	0.98 / 0.98	1.22 / 5.21	2.41 / 1.94	10.21 / 22.62
			0.3	1.53 / 1.94	0.97 / 0.97	1.30 / 8.45	3.76 / 3.62	10.02 / 28.00
			0.6	1.69 / 3.90	0.96 / 0.94	1.16 / 13.05	4.56 / 6.40	12.35 / 41.18
			0.9	1.84 / 8.98	0.96 / 0.91	0.83 / 25.08	4.61 / 11.29	17.66 / 107.52
		One-step	0.0	1.26 / 1.70	0.98 / 0.99	1.24 / 6.40	2.34 / 1.11	10.45 / 28.86
			0.3	1.67 / 12.12	0.96 / 0.94	1.21 / 18.85	4.53 / 6.24	9.10 / 46.72
			0.6	1.92 / 21.66	0.96 / 0.92	1.19 / 38.48	5.07 / 8.92	12.06 / 84.29
			0.9	1.65 / 13.13	0.94 / 0.88	0.55 / 36.63	7.07 / 14.35	12.04 / 132.02
	MA	One-step	0.0	1.39 / 13.90	0.97 / 0.95	1.33 / 24.75	3.27 / 4.91	9.96 / 53.36
			0.3	1.59 / 16.94	0.96 / 0.93	1.60 / 58.91	3.94 / 7.51	11.67 / 82.91
			0.6	1.37 / 16.20	0.96 / 0.91	1.48 / 68.62	4.53 / 9.53	13.24 / 102.41
			0.9	1.29 / 7.94	0.96 / 0.90	0.82 / 45.30	4.88 / 11.64	15.24 / 136.24
10.0	ES	One-step	0.0	2.08 / 30.86	0.96 / 0.90	1.78 / 58.26	4.60 / 9.49	14.17 / 108.81
			0.3	1.96 / 28.45	0.95 / 0.88	1.78 / 83.52	5.76 / 11.02	17.34 / 136.60
			0.6	1.86 / 24.33	0.94 / 0.86	1.91 / 94.46	7.00 / 13.68	20.99 / 181.15
			0.9	1.53 / 12.15	0.94 / 0.87	1.17 / 54.85	7.06 / 13.15	21.94 / 188.75
	LightGBM	H-step	0.0	2.30 / 1.98	0.96 / 0.98	1.61 / 4.76	3.99 / 2.30	13.40 / 29.18
			0.3	2.83 / 2.23	0.95 / 0.97	1.66 / 6.51	5.52 / 3.28	16.12 / 35.73
			0.6	3.19 / 3.10	0.94 / 0.95	1.55 / 7.99	6.76 / 6.03	18.04 / 54.17
			0.9	2.53 / 9.91	0.93 / 0.88	0.96 / 24.25	8.04 / 13.79	20.04 / 119.39
		One-step	0.0	2.69 / 30.17	0.96 / 0.93	1.70 / 31.44	4.13 / 6.63	13.74 / 81.39
			0.3	3.16 / 34.96	0.95 / 0.89	1.76 / 59.43	5.81 / 9.91	16.90 / 124.30
			0.6	3.28 / 34.08	0.94 / 0.87	1.66 / 76.04	7.00 / 13.24	19.18 / 178.35
			0.9	2.43 / 21.64	0.93 / 0.87	0.99 / 65.53	7.68 / 13.61	20.78 / 213.74
	LSTM	H-step	0.0	1.54 / 1.21	0.97 / 0.98	1.25 / 4.50	3.68 / 2.63	11.81 / 27.63
			0.3	2.10 / 1.13	0.96 / 0.96	1.31 / 4.90	5.04 / 4.00	14.36 / 26.46
			0.6	2.62 / 1.84	0.95 / 0.94	1.36 / 6.89	6.16 / 6.66	17.76 / 46.21
			0.9	2.34 / 6.61	0.94 / 0.92	0.97 / 22.35	7.04 / 9.70	21.55 / 105.98
		One-step	0.0	1.31 / 15.11	0.98 / 0.93	1.13 / 44.67	2.69 / 6.96	12.80 / 98.03
			0.3	1.76 / 23.14	0.96 / 0.87	1.06 / 63.74	4.75 / 12.02	13.08 / 137.46
			0.6	2.17 / 26.53	0.96 / 0.87	1.33 / 79.23	5.01 / 12.94	18.29 / 187.68
			0.9	1.93 / 17.66	0.93 / 0.86	0.76 / 56.44	7.39 / 14.84	17.31 / 201.30
	MA	One-step	0.0	1.58 / 22.69	0.96 / 0.90	1.84 / 55.68	4.82 / 9.05	14.80 / 101.20
			0.3	1.65 / 23.64	0.95 / 0.88	2.10 / 78.16	5.92 / 10.38	17.88 / 128.04
			0.6	1.69 / 22.24	0.94 / 0.86	2.30 / 89.86	7.53 / 13.31	22.55 / 171.13
			0.9	1.49 / 11.80	0.93 / 0.87	1.49 / 53.57	7.83 / 13.38	24.23 / 189.89
25.0	ES	One-step	0.0	1.93 / 22.48	0.93 / 0.84	1.95 / 58.27	8.42 / 12.39	25.89 / 182.21
			0.3	1.85 / 18.82	0.91 / 0.84	1.95 / 57.24	10.96 / 11.89	33.48 / 218.60
			0.6	1.86 / 14.58	0.89 / 0.85	2.09 / 48.73	14.15 / 10.96	42.95 / 277.26
			0.9	1.70 / 11.17	0.88 / 0.87	1.83 / 42.70	15.17 / 10.92	48.71 / 308.38
	LightGBM	H-step	0.0	2.90 / 0.92	0.94 / 0.98	1.31 / 2.51	6.30 / 2.86	19.11 / 37.43
			0.3	2.24 / 0.79	0.94 / 0.96	0.88 / 3.05	6.91 / 4.59	20.61 / 47.22
			0.6	2.17 / 0.86	0.92 / 0.94	0.80 / 3.94	8.44 / 7.06	24.97 / 62.63
			0.9	2.14 / 3.39	0.91 / 0.90	0.70 / 9.14	10.55 / 12.62	29.23 / 118.50
		One-step	0.0	3.01 / 28.01	0.94 / 0.86	1.44 / 55.49	6.83 / 11.41	20.84 / 179.97
			0.3	2.47 / 20.37	0.93 / 0.85	1.09 / 59.17	8.02 / 11.58	23.60 / 233.11
			0.6	2.34 / 16.09	0.91 / 0.87	1.07 / 50.64	9.80 / 10.15	30.00 / 295.32
			0.9	2.19 / 13.62	0.90 / 0.88	0.89 / 53.65	11.36 / 9.86	34.34 / 364.93
	LSTM	H-step	0.0	2.74 / 0.54	0.95 / 0.97	1.19 / 2.26	5.68 / 2.96	19.67 / 35.72
			0.3	2.10 / 0.49	0.95 / 0.95	0.84 / 2.74	6.11 / 5.96	21.24 / 40.52
			0.6	2.13 / 0.48	0.94 / 0.94	0.76 / 3.50	7.34 / 8.22	25.65 / 62.24
			0.9	2.07 / 2.40	0.91 / 0.92	0.64 / 9.33	10.81 / 10.48	28.41 / 120.00
		One-step	0.0	1.96 / 23.31	0.97 / 0.86	1.20 / 86.08	3.09 / 11.28	24.30 / 242.06
			0.3	1.79 / 18.19	0.94 / 0.83	0.82 / 68.36	6.34 / 13.17	21.94 / 252.52
			0.6	1.79 / 13.23	0.91 / 0.84	0.63 / 50.37	9.29 / 12.84	21.47 / 275.09
			0.9	1.84 / 12.65	0.91 / 0.86	0.69 / 56.20	9.98 / 12.35	28.41 / 362.27
	MA	One-step	0.0	1.82 / 21.60	0.92 / 0.84	2.46 / 55.88	9.65 / 12.66	29.58 / 181.79
			0.3	1.79 / 18.20	0.90 / 0.85	2.55 / 54.47	12.40 / 11.36	37.80 / 216.77
			0.6	1.81 / 14.64	0.88 / 0.86	2.70 / 46.19	16.22 / 10.60	49.15 / 277.36
			0.9	1.67 / 11.50	0.87 / 0.89	2.38 / 38.24	17.33 / 10.22	55.40 / 304.36

TABLE 2. Tukey HSD test results of AR(1) for LT = 2 / LT = 10.

		Amplification Ratio		Avg. End Inventory		Avg. OFR		Variance of Inventory	
Group 1	Group 2	p-adj[a]	Mean diff.	p-adj	Mean diff.	p-adj	Mean diff.	p-adj	Mean diff.
ES - One-step	LightGBM - H-step	0.00 / 0.00	0.50 / -15.84	0.80 / 0.00	-4.05 / -99.82	0.99 / 0.00	0.00 / 0.06	0.08 / 0.00	-0.38 / -44.60
	LightGBM - One-step	0.00 / 0.85	0.67 / 2.51	0.96 / 1.00	-2.66 / -3.90	1.00 / 0.89	0.00 / 0.01	0.42 / 0.28	-0.26 / -12.33
	LSTM - H-step	1.00 / 0.00	0.01 / -17.31	0.74 / 0.00	-4.34 / -109.37	0.59 / 0.00	0.01 / 0.06	0.00 / 0.00	-0.54 / -46.72
	LSTM - One-step	0.88 / 0.15	-0.13 / -5.32	0.64 / 1.00	-4.89 / -7.48	0.75 / 0.92	0.01 / 0.01	0.00 / 0.33	-0.58 / -11.83
	MA - One-step	0.08 / 0.23	-0.32 / -4.85	1.00 / 1.00	1.31 / -6.87	1.00 / 1.00	0.00 / 0.00	0.90 / 0.99	0.14 / -3.09
LightGBM - H-step	LightGBM - One-step	0.71 / 0.00	0.17 / 18.36	1.00 / 0.01	1.39 / 95.92	1.00 / 0.02	0.00 / -0.04	0.96 / 0.00	0.12 / 32.27
	MA - One-step	0.00 / 0.00	-0.82 / 11.00	0.54 / 0.01	5.36 / 92.94	0.97 / 0.00	-0.01 / -0.05	0.00 / 0.00	0.52 / 41.51
LightGBM - One-step	MA - One-step	0.00 / 0.01	-0.99 / -7.36	0.81 / 1.00	3.97 / -2.98	1.00 / 0.98	0.00 / -0.01	0.05 / 0.60	0.40 / 9.24
LSTM - H-step	LightGBM - H-step	0.00 / 0.98	0.49 / 1.46	1.00 / 1.00	0.30 / 9.55	0.89 / 1.00	-0.01 / 0.00	0.84 / 1.00	0.16 / 2.12
	LightGBM - One-step	0.00 / 0.00	0.66 / 19.82	0.99 / 0.00	1.68 / 105.47	0.73 / 0.01	-0.01 / -0.05	0.33 / 0.00	0.28 / 34.39
	LSTM - One-step	0.83 / 0.00	-0.14 / 11.98	1.00 / 0.00	-0.55 / 101.89	1.00 / 0.01	0.00 / -0.05	1.00 / 0.00	-0.04 / 34.89
	MA - One-step	0.06 / 0.00	-0.34 / 12.46	0.48 / 0.00	5.66 / 102.49	0.42 / 0.00	-0.02 / -0.06	0.00 / 0.00	0.68 / 43.63
LSTM - One-step	LightGBM - H-step	0.00 / 0.00	0.63 / -10.52	1.00 / 0.01	0.84 / -92.34	0.96 / 0.01	-0.01 / 0.04	0.68 / 0.00	0.20 / -32.77
	LightGBM - One-step	0.00 / 0.01	0.80 / 7.83	0.98 / 1.00	2.23 / 3.58	0.86 / 1.00	-0.01 / 0.00	0.19 / 1.00	0.32 / -0.50
	MA - One-step	0.59 / 1.00	-0.19 / 0.48	0.37 / 1.00	6.20 / 0.60	0.59 / 0.99	-0.01 / -0.01	0.00 / 0.66	0.72 / 8.74

TABLE 3. Simulation results of M5 for LT = 2 / LT = 10.

Method	Horizon	CV Group	AR	Avg. OFR	Var. of Inv.	Avg. Backorders	Avg. End Inv.
ES	One-step	0.2 < CV =< 0.3	1.92 / 21.55	0.93 / 0.90	1.90 / 44.49	12.61 / 15.35	38.80 / 222.04
		0.3 < CV =< 0.4	1.94 / 15.70	0.90 / 0.89	1.78 / 36.48	7.70 / 7.35	23.68 / 131.22
		0.4 < CV =< 0.6	1.74 / 12.37	0.90 / 0.88	1.55 / 40.92	4.55 / 4.82	13.86 / 86.38
LightGBM	H-step	0.2 < CV =< 0.3	3.58 / 5.48	0.94 / 0.95	2.39 / 15.31	9.62 / 7.96	44.69 / 130.13
		0.3 < CV =< 0.4	2.80 / 4.50	0.92 / 0.93	2.01 / 12.63	6.66 / 5.40	24.81 / 68.80
		0.4 < CV =< 0.6	2.11 / 4.48	0.89 / 0.89	1.42 / 12.78	5.09 / 4.67	13.99 / 49.27
	One-step	0.2 < CV =< 0.3	3.12 / 27.88	0.94 / 0.93	2.05 / 42.83	10.74 / 12.09	36.46 / 254.57
		0.3 < CV =< 0.4	2.57 / 17.86	0.91 / 0.92	1.74 / 30.13	6.66 / 5.68	21.73 / 140.67
		0.4 < CV =< 0.6	2.08 / 11.42	0.90 / 0.90	1.35 / 24.28	4.40 / 4.48	13.00 / 79.19
LSTM	H-step	0.2 < CV =< 0.3	2.56 / 3.82	0.96 / 0.96	2.37 / 13.08	6.43 / 7.08	53.72 / 123.28
		0.3 < CV =< 0.4	1.92 / 3.12	0.94 / 0.94	1.72 / 10.81	5.29 / 5.00	26.41 / 66.65
		0.4 < CV =< 0.6	1.58 / 3.18	0.90 / 0.89	1.24 / 10.47	4.34 / 4.56	14.11 / 46.12
	One-step	0.2 < CV =< 0.3	2.77 / 25.23	0.95 / 0.94	2.13 / 45.06	9.30 / 10.68	39.11 / 270.73
		0.3 < CV =< 0.4	2.06 / 13.59	0.92 / 0.92	1.57 / 26.93	6.42 / 5.62	21.73 / 138.21
		0.4 < CV =< 0.6	1.66 / 8.85	0.91 / 0.90	1.24 / 22.61	4.28 / 4.56	12.44 / 72.20
MA	One-step	0.2 < CV =< 0.3	1.49 / 13.34	0.93 / 0.89	1.57 / 34.60	12.65 / 15.66	38.84 / 196.21
		0.3 < CV =< 0.4	1.47 / 9.62	0.91 / 0.88	1.48 / 28.46	7.64 / 8.01	23.39 / 111.52
		0.4 < CV =< 0.6	1.38 / 7.26	0.90 / 0.87	1.30 / 30.34	4.44 / 5.19	13.36 / 73.18

TABLE 4. Tukey HSD test results of M5 for LT = 2 / LT = 10.

		Amplification Ratio		Avg. End Inventory		Avg. OFR		Variance of Inventory	
Group 1	Group 2	p-adj[a]	Mean diff.	p-adj	Mean diff.	p-adj	Mean diff.	p-adj	Mean diff.
ES - One-step	LightGBM - H-step	0.00 / 0.00	-0.86 / 11.12	1.00 / 0.25	-1.89 / -59.88	0.78 / 0.00	-0.01 / 0.03	0.70 / 0.00	-0.16 / -26.76
	LightGBM - One-step	0.00 / 0.41	-0.66 / -1.98	1.00 / 1.00	-1.63 / 8.67	0.53 / 0.00	0.01 / 0.02	1.00 / 0.00	-0.06 / -9.14
	LSTM - H-step	0.98 / 0.00	-0.08 / 12.53	0.96 / 0.19	4.72 / -63.54	0.00 / 0.00	0.03 / 0.04	1.00 / 0.00	-0.03 / -28.87
	LSTM - One-step	0.43 / 0.83	-0.22 / 1.26	1.00 / 1.00	-1.20 / 8.96	0.02 / 0.00	0.02 / 0.03	0.78 / 0.00	-0.14 / -10.45
	MA - One-step	0.01 / 0.00	0.42 / 6.23	1.00 / 0.98	-0.29 / -18.70	0.94 / 0.35	0.00 / -0.01	0.10 / 0.00	-0.29 / -9.44
LightGBM - H-step	LightGBM - One-step	0.49 / 0.00	0.21 / -13.09	0.99 / 0.13	-3.52 / 68.54	1.00 / 0.95	0.00 / 0.00	0.36 / 0.00	-0.22 / 17.62
	MA - One-step	0.00 / 0.00	1.29 / -4.89	1.00 / 0.66	-2.18 / 41.18	1.00 / 0.00	0.00 / -0.04	0.00 / 0.00	-0.44 / 17.32
LightGBM - One-step	MA - One-step	0.00 / 0.00	1.08 / 8.20	1.00 / 0.92	1.34 / -27.37	0.97 / 0.00	0.00 / -0.04	0.29 / 1.00	-0.23 / -0.30
LSTM - H-step	LightGBM - H-step	0.00 / 0.75	-0.78 / -1.41	1.00 / 1.00	-2.82 / 3.66	0.00 / 0.80	-0.02 / -0.01	0.55 / 0.91	0.18 / 2.11
	LightGBM - One-step	0.00 / 0.00	-0.57 / -14.50	0.86 / 0.09	-6.34 / 72.20	0.01 / 0.25	-0.02 / -0.01	1.00 / 0.00	-0.03 / 19.73
	LSTM - One-step	0.87 / 0.00	-0.13 / -11.27	0.89 / 0.09	-5.92 / 72.49	0.33 / 0.43	-0.01 / -0.01	0.89 / 0.00	-0.12 / 18.42
	MA - One-step	0.00 / 0.00	0.51 / -6.30	0.94 / 0.58	-5.01 / 44.84	0.00 / 0.00	-0.02 / -0.05	0.16 / 0.00	-0.26 / 19.43
LSTM - One-step	LightGBM - H-step	0.00 / 0.00	-0.64 / 9.86	0.99 / 0.12	3.09 / -68.83	0.39 / 0.99	-0.01 / 0.00	0.07 / 0.00	0.30 / -16.31
	LightGBM - One-step	0.00 / 0.03	-0.44 / -3.23	1.00 / 1.00	-0.43 / -0.29	0.65 / 1.00	-0.01 / 0.00	0.97 / 0.99	0.09 / 1.31
	MA - One-step	0.00 / 0.00	0.64 / 4.97	1.00 / 0.91	0.91 / -27.66	0.20 / 0.00	-0.01 / -0.04	0.78 / 1.00	-0.14 / 1.01

[a] p-adj values below 0.05 denotes statistically significant difference.

The introduction of h-step-ahead forecasting for ML-based methods, specifically LSTM and LightGBM, led to substantial improvements in controlling the amplification ratio compared with both traditional methods and one-step-ahead ML models. Fig. 8 and Fig. 9 show that, as lead times increased, the h-step-ahead forecasting methods consistently outperformed both the one-step-ahead approaches and traditional methods for both the AR(1) and M5 series.

The Tukey test results in Tables 2 and 4 further validate these findings, showing statistically significant improvements in the amplification ratio for h-step-ahead forecasts, particularly when compared to one-step-ahead forecasts from

both the ML and traditional models. These differences are especially pronounced as lead times increase, where the benefit of multi-step forecasting becomes more evident in mitigating BWE.

2) VARIANCE OF INVENTORY

For one-step-ahead forecasts, the ML-based methods did not consistently exhibit a significantly lower variance of inventory compared to traditional methods. Although LSTM and LightGBM provided more accurate demand forecasts, this accuracy did not always translate into more stable inventory levels. This finding is more evident when lead times are short, where the variance of inventory remains comparable across both the traditional and ML methods.

As lead times increased, some differences began to emerge, with LSTM and LightGBM showing slightly lower inventory variance than ES and MA. However, these differences were not always statistically significant according to Tukey's test results. The results suggest that the forecast accuracy gains from ML methods in one-step-ahead forecasting are not fully realized in the context of inventory variance. The one-step-ahead nature of these forecasts limits their ability to anticipate long-term fluctuations in demand, which contributes to continued variability in inventory levels.

Fig. 10 and Fig. 11 show that the use of h-step-ahead forecasting with ML-based methods significantly reduced the variance of inventory compared to both traditional methods and one-step-ahead ML approaches.

The Tukey test results support these findings, showing statistically significant reductions in inventory variance when using h-step-ahead LSTM and LightGBM compared with both one-step-ahead forecasting methods and traditional approaches.

3) ORDER FULFILLMENT RATE

The results show that despite the superior forecast accuracy of LSTM and LightGBM in predicting demand, the OFRs did not significantly differ between ML-based methods and traditional methods. The Tukey test results indicate that while there are some statistically significant differences between methods for certain lead times, these differences are not as pronounced as expected, considering the forecast accuracy advantage of ML models.

In concurrence with other metrics, Fig. 12. And Fig. 13. illustrate how h-step-ahead forecasting with LSTM and LightGBM offers a substantial improvement in OFR over one-step-ahead forecasting and traditional methods for both demand series.

The Tukey test results indicate statistically significant differences in OFRs when comparing h-step-ahead ML methods with both one-step-ahead and traditional methods for the AR(1) series. These differences become more significant as the lead times increase. However, while there is a positive gap, the Tukey analysis for the M5 series does not confirm the statistical significance of the observed mean differences.

4) AVERAGE END INVENTORY

For the one-step-ahead forecasting comparison, Fig. 14 and Fig. 15 highlight that the one-step-ahead ML methods did not demonstrate a substantial improvement over traditional methods. As the lead-time increased, both categories of methods tended to converge in their performance, indicating that the superior forecast accuracy of LSTM and LightGBM did not translate into significant gains in inventory management.

The introduction of h-step-ahead forecasting methods for ML-based models has yielded notable improvements in the Average End Inventory. This enhancement is particularly evident when compared with the one-step-ahead forecasts of traditional methods and ML-based approaches.

Tukey analysis shows a significant reduction in the Average End Inventory when utilizing h-step-ahead forecasting with LSTM and LightGBM, especially as lead times increase for the AR(1) series. However, the Tukey results for the M5 series do not support the significance of the mean differences, as is evident in Fig. 15.

V. CONCLUSION

This study offers novel insights into the application of ML models, LSTM, and LightGBM to mitigate BWE and enhance supply chain performance. This study is among the first to systematically compare the impact of one-step-ahead and h-step-ahead forecasting in the context of advanced ML methods, focusing on their ability to handle varying demand conditions. Notably, our findings demonstrate that h-step forecasting, particularly with ML models, significantly reduces demand amplification, leading to improvements in key supply chain metrics including order fulfillment rates, variance of inventory, and average end inventory. The results underscore the superiority of LSTM and LightGBM in scenarios characterized by high demand autocorrelation, variability, and seasonality. While traditional methods, such as MA and ES, continue to struggle with more complex demand patterns, LSTM and LightGBM showed resilience in maintaining lower forecast errors across different stochastic environments. This accuracy translated into reduced inventory variance and improved order fulfillment, particularly when h-step-ahead forecasting was applied. The findings also reveal that h-step-ahead forecasting allows supply chains to better anticipate longer-term demand fluctuations, resulting in smoother order adjustments and a notable reduction in BWE. Despite these significant improvements, the research highlights that the benefits of ML models are not fully realized when constrained to one-step-ahead forecasting.

Building on the findings of this study, future research could focus on combining ML models with traditional methods using hybrid forecasting approaches to provide more robust solutions for managing supply chain uncertainty. Integrating these forecasts with advanced inventory optimization algorithms provides an opportunity to further enhance replenishment decisions. Testing these models in multistage

supply chain environments would offer valuable insights into their practical applications and adaptability. In addition, future studies should investigate alternative inventory control mechanisms beyond the OUT policy by applying ML-based forecasts to other inventory systems for a more comprehensive evaluation of supply chain performance.

REFERENCES

[1] J. W. Forrester, *Industrial Dynamics*. Cambridge, MA, USA: MIT Press, 1961.

[2] J. D. Sterman, "Modeling managerial behavior: Misperceptions of feedback in a dynamic decision making experiment," *Manage. Sci.*, vol. 35, no. 3, pp. 321–339, Mar. 1989, doi: 10.1287/mnsc.35.3.321.

[3] H. L. Lee, V. Padmanabhan, and S. Whang, "Information distortion in a supply chain: The bullwhip effect," *Manage. Sci.*, vol. 43, no. 4, pp. 546–558, Apr. 1997, doi: 10.1287/mnsc.43.4.546.

[4] J. Dejonckheere, S. M. Disney, M. R. Lambrecht, and D. R. Towill, "Measuring and avoiding the bullwhip effect: A control theoretic approach," *Eur. J. Oper. Res.*, vol. 147, no. 3, pp. 567–590, Jun. 2003, doi: 10.1016/S0377-2217(02)00369-7.

[5] F. Chen, Z. Drezner, J. K. Ryan, and D. Simchi-Levi, "Quantifying the bullwhip effect in a simple supply chain: The impact of forecasting, lead times, and information," *Manage. Sci.*, vol. 46, no. 3, pp. 436–443, Mar. 2000, doi: 10.1287/mnsc.46.3.436.12069.

[6] F. Chen, J. K. Ryan, and D. Simchi-Levi, "The impact of exponential smoothing forecasts on the bullwhip effect," *Nav. Res. Logistics*, vol. 47, no. 4, Apr. 2000, doi: 10.1002/(SICI)1520-6750(200006)47:4<269::AID-NAV1>3.0.CO;2-Q.

[7] C. Chandra and J. Grabis, "Application of multi-steps forecasting for restraining the bullwhip effect and improving inventory performance under autoregressive demand," *Eur. J. Oper. Res.*, vol. 166, no. 2, pp. 337–350, Oct. 2005, doi: 10.1016/j.ejor.2004.02.012.

[8] T. Hosoda and S. M. Disney, "On variance amplification in a three-echelon supply chain with minimum mean square error forecasting," *Omega*, vol. 34, no. 4, pp. 344–358, Aug. 2006, doi: 10.1016/j.omega.2004.11.005.

[9] M. Z. Babai, Y. Dai, Q. Li, A. Syntetos, and X. Wang, "Forecasting of lead-time demand variance: Implications for safety stock calculations," *Eur. J. Oper. Res.*, vol. 296, no. 3, pp. 846–861, Feb. 2022, doi: 10.1016/j.ejor.2021.04.017.

[10] Z. Michna, S. M. Disney, and P. Nielsen, "The impact of stochastic lead times on the bullwhip effect under correlated demand and moving average forecasts," *Omega*, vol. 93, Jun. 2020, Art. no. 102033, doi: 10.1016/j.omega.2019.02.002.

[11] M. Hussain, A. Shome, and D. M. Lee, "Impact of forecasting methods on variance ratio in order-up-to level policy," *Int. J. Adv. Manuf. Technol.*, vol. 59, nos. 1–4, pp. 413–420, Mar. 2012, doi: 10.1007/s00170-011-3485-1.

[12] Y. Ma, N. Wang, A. Che, Y. Huang, and J. Xu, "The bullwhip effect on product orders and inventory: A perspective of demand forecasting techniques," *Int. J. Prod. Res.*, vol. 51, no. 1, pp. 281–302, Jan. 2013, doi: 10.1080/00207543.2012.676682.

[13] G. P. Zhang and M. Qi, "Neural network forecasting for seasonal and trend time series," *Eur. J. Oper. Res.*, vol. 160, no. 2, pp. 501–514, Jan. 2005, doi: 10.1016/j.ejor.2003.08.037.

[14] R. Carbonneau, K. Laframboise, and R. Vahidov, "Application of machine learning techniques for supply chain demand forecasting," *Eur. J. Oper. Res.*, vol. 184, no. 3, pp. 1140–1154, Feb. 2008, doi: 10.1016/j.ejor.2006.12.004.

[15] S. F. Crone, M. Hibon, and K. Nikolopoulos, "Advances in forecasting with neural networks? Empirical evidence from the NN3 competition on time series prediction," *Int. J. Forecasting*, vol. 27, no. 3, pp. 635–660, Jul. 2011, doi: 10.1016/j.ijforecast.2011.04.001.

[16] M. Pournader, H. Ghaderi, A. Hassanzadegan, and B. Fahimnia, "Artificial intelligence applications in supply chain management," *Int. J. Prod. Econ.*, vol. 241, Nov. 2021, Art. no. 108250, doi: 10.1016/j.ijpe.2021.108250.

[17] E. Weisz, D. M. Herold, and S. Kummer, "Revisiting the bullwhip effect: How can AI smoothen the bullwhip phenomenon?" *Int. J. Logistics Manage.*, vol. 34, no. 7, pp. 98–120, Dec. 2023, doi: 10.1108/ijlm-02-2022-0078.

[18] N. Kourentzes, "Intermittent demand forecasts with neural networks," *Int. J. Prod. Econ.*, vol. 143, no. 1, pp. 198–206, May 2013, doi: 10.1016/j.ijpe.2013.01.009.

[19] J. R. Trapero, M. Cardós, and N. Kourentzes, "Empirical safety stock estimation based on kernel and GARCH models," *Omega*, vol. 84, pp. 199–211, Apr. 2019, doi: 10.1016/j.omega.2018.05.004.

[20] J. Huber, S. Müller, M. Fleischmann, and H. Stuckenschmidt, "A data-driven newsvendor problem: From data to decision," *Eur. J. Oper. Res.*, vol. 278, no. 3, pp. 904–915, Nov. 2019, doi: 10.1016/j.ejor.2019.04.043.

[21] Y. Cao and Z.-J.-M. Shen, "Quantile forecasting and data-driven inventory management under nonstationary demand," *Oper. Res. Lett.*, vol. 47, no. 6, pp. 465–472, Nov. 2019, doi: 10.1016/j.orl.2019.08.008.

[22] A. Oroojlooyjadid, L. V. Snyder, and M. Takáč, "Applying deep learning to the newsvendor problem," *IISE Trans.*, vol. 52, no. 4, pp. 444–463, Apr. 2020, doi: 10.1080/24725854.2019.1632502.

[23] M. Z. Babai, A. Tsadiras, and C. Papadopoulos, "On the empirical performance of some new neural network methods for forecasting intermittent demand," *IMA J. Manage. Math.*, vol. 31, no. 3, pp. 281–305, Dec. 2019, doi: 10.1093/imaman/dpaa003.

[24] E. Spiliotis, S. Makridakis, A. Kaltsounis, and V. Assimakopoulos, "Product sales probabilistic forecasting: An empirical evaluation using the M5 competition data," *Int. J. Prod. Econ.*, vol. 240, Oct. 2021, Art. no. 108237, doi: 10.1016/j.ijpe.2021.108237.

[25] M. Seyedan, F. Mafakheri, and C. Wang, "Order-up-to-level inventory optimization model using time-series demand forecasting with ensemble deep learning," *Supply Chain Anal.*, vol. 3, Sep. 2023, Art. no. 100024, doi: 10.1016/j.sca.2023.100024.

[26] S. Jaipuria and S. S. Mahapatra, "An improved demand forecasting method to reduce bullwhip effect in supply chains," *Expert Syst. Appl.*, vol. 41, no. 5, pp. 2395–2408, Apr. 2014, doi: 10.1016/j.eswa.2013.09.038.

[27] O. Prakash and V. Pandey, "Reducing the bullwhip effect in a supply chain using artificial intelligence technique," *J. Prod. Res. Manag.*, vol. 4, no. 2, Apr. 2014.

[28] M. Zhang, L. Shi, X. Zhuo, and Y. Liu, "A bullwhip effect weakening approach based on VMD-SVM algorithm under the background of intelligent manufacturing," *Processes*, vol. 9, no. 11, pp. 1–13, 2021, doi: 10.3390/pr9111957.

[29] S. F. Crone and N. Kourentzes, "Forecasting seasonal time series with multilayer perceptrons–an empirical evaluation of input vector specifications for deterministic seasonality," in *Proc. DMIN*, 2009, pp. 232–238.

[30] M. Udenio, E. Vatamidou, and J. C. Fransoo, "Exponential smoothing forecasts: Taming the bullwhip effect when demand is seasonal," *Int. J. Prod. Res.*, vol. 61, no. 6, pp. 1796–1813, Mar. 2023, doi: 10.1080/00207543.2022.2048114.

[31] A. A. Syntetos, J. E. Boylan, and J. D. Croston, "On the categorization of demand patterns," *J. Oper. Res. Soc.*, vol. 56, no. 5, pp. 495–503, May 2005, doi: 10.1057/palgrave.jors.2601841.

[32] E. Spiliotis, S. Makridakis, A.-A. Semenoglou, and V. Assimakopoulos, "Comparison of statistical and machine learning methods for daily SKU demand forecasting," *Oper. Res.*, vol. 22, no. 3, pp. 3037–3061, Jul. 2022, doi: 10.1007/s12351-020-00605-2.

[33] G. D. Johnson and H. E. Thompson, "Optimality of myopic inventory policies for certain dependent demand processes," *Manage. Sci.*, vol. 21, no. 11, pp. 1303–1307, Jul. 1975, doi: 10.1287/mnsc.21.11.1303.

[34] X. Zhang, "The impact of forecasting methods on the bullwhip effect," *Int. J. Prod. Econ.*, vol. 88, no. 1, pp. 15–27, Mar. 2004, doi: 10.1016/S0925-5273(03)00128-2.

[35] J. Heydari, R. B. Kazemzadeh, and S. K. Chaharsooghi, "A study of lead time variation impact on supply chain performance," *Int. J. Adv. Manuf. Technol.*, vol. 40, nos. 11–12, pp. 1206–1215, Feb. 2009, doi: 10.1007/s00170-008-1428-2.

[36] S. M. Disney and D. R. Towill, "On the bullwhip and inventory variance produced by an ordering policy," *Omega*, vol. 31, no. 3, pp. 157–167, Jun. 2003, doi: 10.1016/S0305-0483(03)00028-8.

[37] I. Kim and M. Springer, "Measuring endogenous supply chain volatility: Beyond the bullwhip effect," *Eur. J. Oper. Res.*, vol. 189, no. 1, pp. 172–193, Aug. 2008, doi: 10.1016/j.ejor.2007.05.012.

[38] S. Cannella, A. P. Barbosa-Póvoa, J. M. Framinan, and S. Relvas, "Metrics for bullwhip effect analysis," *J. Oper. Res. Soc.*, vol. 64, no. 1, pp. 1–16, Jan. 2013, doi: 10.1057/jors.2011.139.



ALPER SARICIOĞLU received the B.S. degree in industrial engineering from Eskişehir Osmangazi University, Türkiye, in 2010, and the M.Sc. degree in industrial engineering from Bahçeşehir University, Türkiye, in 2014. He is currently pursuing the Ph.D. degree in industrial engineering with Galatasaray University, Türkiye. Since September 2012, he has been a Partner and a Senior Consultant with Poseidon Consulting, Istanbul. In this role, he has contributed to the design of operations and the development of IT strategies for various local and international companies.



MÜJDE EROL GENEVOIS received the B.Sc. degree in business engineering from Istanbul Technical University and the M.Sc. and Ph.D. degrees in industrial engineering from the Institut National Polytechnique de Grenoble. She is currently a Professor in industrial engineering with Galatasaray University. Her research interests include supply chain flexibility, reverse logistics, and sustainable design. Over the years, she has led numerous projects, including studies on ATM cash management and green design for wasteful electrical equipment. She has published extensively in leading journals, with notable contributions in areas, such as ATM demand forecasting and facility-location strategies. In addition to her research, she teaches courses, such as production planning, supply chain management, and engineering systems analysis while supervising graduate theses. She also held administrative roles, including the Vice Dean and General Secretary with Galatasaray University.



MICHELE CEDOLIN received the B.S. and M.S. degrees in industrial engineering and the Ph.D. degree from Galatasaray University, Istanbul, Türkiye, in 2014, 2016, and 2022, respectively. He is currently an Assistant Professor in industrial engineering. His research interests include data analysis, process design, forecasting, supply chain optimization, and the applications of machine learning and artificial intelligence in industrial contexts.