DEMAND FORECASTING IN THE SUPPLY CHAIN: CASE STUDY OF A BANGLADESHI RETAILER

BY

NONTHAPHAN KINGRUNGPHET

AN INDEPENDENT STUDY SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF ENGINEERING (LOGISTICS AND SUPPLY CHAIN SYSTEMS ENGINEERING)
SIRINDHORN INTERNATIONAL INSTITUTE OF TECHNOLOGY
THAMMASAT UNIVERSITY
ACADEMIC YEAR 2025

THAMMASAT UNIVERSITY
SIRINDHORN INTERNATIONAL INSTITUTE OF TECHNOLOGY

INDEPENDENT STUDY

BY

NONTHAPHAN KINGRUNGPHET

ENTITLED

DEMAND FORECASTING IN THE SUPPLY CHAIN: CASE STUDY OF A BANGLADESHI RETAILER

was approved as partial fulfillment of the requirements for
the degree of Master of Engineering (Logistics and Supply Chain Systems Engineering)

on November 20, 2025

Member and Advisor
(Associate Professor Jirachai Buddhakulsomsiri, Ph.D.)

Member
(Associate Professor Pham Duc Tai, Ph.D.)

Director
(Associate Professor Kriengsak Panuwatwanich, Ph.D.)

Ref. code: 25686722041115TZM

Independent Study Title	DEMAND FORECASTING IN THE SUPPLY CHAIN: CASE STUDY OF A BANGLADESHI RETAILER
Author	Nonthaphan Kingrungphet
Degree	Master of Engineering (Logistics and Supply Chain Systems Engineering)
Faculty/University	Sirindhorn International Institute of Technology/ Thammasat University
Advisor	Associate Professor Jirachai Buddhakulsomsiri, Ph.D.
Academic Years	2025

ABSTRACT

This study aims to forecast daily sales using time series analysis through two models: Holt-Winters and ARIMA. A real dataset containing 1,826 consecutive days of sales data was used as the basis for analysis.

The first model, Holt-Winters, was optimized by adjusting parameters for level, trend, and seasonality to achieve the lowest possible MAPE. Subsequently, ARIMA (4,1,1) models were developed using Minitab software. Forecasted values from each model were compared against the actual sales values to compute the Mean Absolute Percentage Error (MAPE), which was used as the main metric for evaluating forecasting accuracy.

The results showed that the Holt-Winters model achieved the lowest MAPE of 22.04, indicating the highest accuracy among the two models. These findings suggest that the Holt-Winters method is a suitable approach for short-term sales forecasting in daily time series data.

This study provides useful insights for improving forecasting in business operations, especially for production planning and inventory management.

Keywords: Time series forecasting, Sales prediction, Holt-winter, Arima, Seasonal pattern, Forecast accuracy, Demand forecasting, Daily sales, Comparative analysis

ACKNOWLEDGEMENTS

This project owes much to the generosity of those who offered time and expertise. I would like to acknowledge Assoc. Prof. Dr. Jirachai Buddhakulsomsiri for clear direction, constructive comments at each milestone, and sustained motivation from the first proposal to the final report. I further acknowledge the Logistics and Supply Chain Systems Engineering program, Sirindhorn International Institute of Technology, Thammasat University, for providing a supportive academic setting and essential resources. My heartfelt thanks also go to my family and friends, whose patience and encouragement made it possible to complete this Independent Study.

Nonthaphan Kingrungphet

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS/ABBREVIATIONS

Symbols/Abbreviations	Terms
t	Time index (day)
Y_t	Actual sales on day t (units)
$\acute{Y}_t$	First difference Y_t - Y_{t-1}
L_t	Level component (Holt–Winters)
T_t	Trend component (Holt–Winters)
S_t	Seasonal component (Holt–Winters)
α, β, γ	Smoothing constants for level, trend, seasonality
s	Seasonal length
$\hat{Y}_t$	Forecast value at time t*
p, d, q	ARIMA orders: AR, differencing,
MA	Moving average
HW	Holt–Winters exponential smoothing
ETS	Error-Trend-Seasonal
AR	Autoregressive
MA	Moving average
ARIMA	Autoregressive Integrated Moving Average
ACF	Autocorrelation function
PACF	Partial autocorrelation function
DOE	Design of experiments (HW parameter grid
ANN	Artificial neural network
W	Weight
b	Bi

CHAPTER 1
INTRODUCTION

In today's highly competitive environment, organizations are under sustained pressure to run lean operations and make decisions grounded in data. A central lever for operational excellence is demanding forecasting anticipating near-term sales with enough accuracy to plan procurement, staffing, inventory, transportation, and service levels. Within logistics and supply chain contexts, forecast accuracy directly affects stockouts, excess inventory, working capital, and the ability to react to market changes.

As sales patterns increasingly reflect a mix of trend, seasonality, calendar effects, and noise, single-rule or naïve approaches often fall short. This has motivated broad use of time-series models such as Holt-Winters exponential smoothing and ARIMA (Autoregressive Integrated Moving Average), alongside modern Artificial Neural Networks (ANN) that can learn nonlinear relationships from engineered features (e.g., lags and calendar indicators). Each method brings different assumptions and strengths; selecting a suitable approach depends on both the data's structure and the forecasting objective.

This Independent Study (IS) compares these approaches on a real retail sales series. Using a common evaluation protocol and focusing on Mean Absolute Percentage Error (MAPE) as the primary criterion, the study assesses which method is most appropriate for daily demand prediction and discusses implications for supply-chain planning.

1.1 Background and significance

Reliable short-horizon forecasts are foundational to supply-chain and logistics decisions. When forecasts are well-calibrated, firms can align purchasing with demand, position inventory efficiently, schedule labor and transport more effectively, and deliver higher service levels at lower cost. Because demand patterns can be complex exhibiting weekly and annual seasonality, holidays, and trend model choice matters. Comparing established statistical models with learning-based methods helps practitioners select tools that balance accuracy, transparency, and ease of deployment.

1.2 Problem statement

Multiple forecasting options exist from exponential smoothing families to ARIMA and data-driven ANN. Their performance, however, is not uniform; it varies with structural features of the data such as trend strength, seasonal amplitude, calendar effects, and noise. Choosing an unsuitable model can inflate forecast errors and create operational inefficiencies. A systematic, side-by-side comparison under a common data split and metric is therefore necessary to identify the most appropriate method for the dataset at hand.

1.3 Research scope and methodology

This study analyzes five years of daily sales (1,826 observations) from a retail context. The series displays clear seasonal behavior and trend, making it a realistic testbed for short-term forecasting. Before modeling, the data were cleaned, transformed where required (e.g., differencing for ARIMA), and augmented with time-based features (e.g., day-of-week, month, holidays, and sales lags) for models that can exploit them (e.g., ANN).

A fixed data split is used for a fair comparison: training, validation, and test subsets. Model selection and hyperparameter tuning are performed on the validation set; final performance is reported on the held-out test set using consistent rounding rules for count forecasts.

1.3.1 Forecasting models considered

This study develops and evaluates three forecasting approaches on the same dataset. First, the Holt–Winters exponential smoothing method (triple smoothing) is applied because it explicitly models level, trend, and seasonality; both additive and multiplicative specifications are considered to accommodate constant versus scale-dependent seasonal amplitudes. Second, the ARIMA framework is used to capture linear dependence in (difference) stationary series, with orders guided by ACF/PACF diagnostics following the Box–Jenkins procedure. Third, an Artificial Neural Network (ANN) is trained as a feed-forward model on engineered features including recent sales lags and calendar or holiday indicators to learn potential nonlinear relationships that classical linear models may not capture. All models share a common protocol: they are trained on the same training period, tuned on the validation period, and then refit as appropriate prior to evaluation on the held-out test set.

1.3.2 Evaluation metric

MAPE is the principal metric, reporting average absolute percentage error relative to actual demand facilitating comparison across time and across methods. To provide a fuller view of error magnitude and dispersion, Mean Absolute Deviation (MAD), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) are also reported. Accuracy metrics are computed on non-differenced actuals, and where applicable, forecast values are rounded to integers to reflect count data before metric calculation for consistency across methods.

1.4 Objective of the study

The overarching aim is to identify a practically deployable method for day-ahead demand forecasting in a retail/supply-chain setting.

1.4.1 Develop and compare multiple models

Build and evaluate Holt–Winters, ARIMA, and ANN on the same daily-sales data under a unified training/validation/test protocol. Holt–Winters addresses level–trend–seasonality explicitly; ARIMA benchmarks linear dependence after differencing; ANN leverages lagged sales and calendar features to capture potential nonlinearities.

1.4.2 Evaluate forecasting accuracy using MAPE

Model performance is benchmarked primarily with Mean Absolute Percentage Error (MAPE), which reports error as a percentage of the observed value, enabling comparisons across different demand levels. To add robustness, we also report Mean Absolute Deviation (MAD), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).

1.4.3 Determine the most effective model

By comparing MAPE values across all models, the study aims to identify the model that yields the lowest forecasting error. This model will be recommended as the most suitable for forecasting daily sales in similar operational environments. The identification of the best-performing model will also consider the complexity and interpretability of each method.

1.4.4 Provide practical implications for real-world business use

The ultimate goal of the study is to generate insights that are directly applicable to real-world business operations. By identifying a forecasting method with the lowest error, organizations can make more informed decisions regarding inventory planning, supply chain scheduling, budgeting, and workforce allocation. Improved forecast accuracy can lead to better service levels, cost savings, and strategic agility in logistics and retail environments.

CHAPTER 2
REVIEW OF LITERATURE

2.1 Introduction to time-series forecasting in retail

Time-series forecasting underpins retail and supply-chain operations by informing replenishment, inventory positioning, labor scheduling, and service-level planning. Reviewing a quarter century of research, **De Gooijer and Hyndman (2006)** conclude that exponential smoothing (including Holt–Winters) and ARIMA remain durable baselines for short-horizon demand when identification, diagnostics, and out-of-sample evaluation are applied systematically. Large-scale benchmarking reinforces this view: in the M3 competition, **Makridakis and Hibon (2000)** show that relatively simple approaches especially exponential-smoothing variants often match or surpass more complex methods across thousands of series, underscoring their practical value in business contexts. Extending the evidence, the M4 competition demonstrates that no single method dominates universally; nevertheless, families such as exponential smoothing and ARIMA remain among the strongest general-purpose performers when tuned carefully, while combinations can further improve accuracy as **Makridakis, Spiliotis and Assimakopoulos (2020)** report. Focusing on daily supermarket sales, **Taylor (2007)** highlights volatility and skewness that motivate interval forecasting, while also noting the widespread operational use of exponential smoothing for point forecasts. Taken together, these findings justify evaluating Holt–Winters and ARIMA as baselines for daily retail demand in this study.

2.2 Exponential smoothing and Holt–Winters

Exponential smoothing evolved from level-only updating to formulations that also track trend and seasonality. Building the trend-corrected structure and its forecasting expressions, **Holt (2004)** laid foundations for modern Holt–Winters variants used widely when trends coexist with recurring seasonal patterns. Synthesizing post-1985 advances, **Gardner (2006)** clarifies practical issues initialization, parameter stability, and when additive versus multiplicative seasonality is appropriate and concludes that Holt–Winters remains highly competitive for short-horizon operational forecasting. Linking the family to a statistical framework, **Hyndman, Koehler, Ord**

and Snyder (2002) cast exponential smoothing in innovations state-space (ETS) form, enabling principled estimation, information-criterion selection, and forecast intervals for automated application. Addressing unstable long-run extrapolation in trending demand, **Taylor (2003)** proposes a damped multiplicative trend specification that tempers growth while preserving level-proportional seasonal amplitude, often improving stability without sacrificing responsiveness.

2.3 ARIMA and the Box–Jenkins framework

The Box–Jenkins methodology frames ARIMA as an iterative cycle identify, estimate, diagnose aimed at parsimonious yet accurate forecasting models. Surveying 25 years of research, **De Gooijer and Hyndman (2006)** position ARIMA as a durable statistical baseline once stationarity is enforced (e.g., via differencing) and orders are guided by ACF/PACF with residual checks, a practice consistent with retail use. Comparative work against machine-learning alternatives emphasizes ARIMA's strength in capturing short-run linear dependence; in repairable-system series, **Ho, Xie and Goh (2002)** find Box–Jenkins ARIMA robust for short horizons relative to neural networks, reinforcing its role as a transparent benchmark. To aid order selection beyond heuristic diagnostics, **Ong, Huang and Tzeng (2005)** apply genetic algorithms, showing that search-based identification navigates complex order spaces and avoids local optima while remaining within the Box–Jenkins workflow. With explanatory inputs, ARIMAX can incorporate promotions and category effects: using SKU-level retail data, **Ma, Fildes and Huang (2016)** demonstrate material accuracy gains from promotional covariates, an important signal for retail planning. Hybridization further suggests complementary strengths, as **Zhang (2003)** shows that combining ARIMA (linear) with neural networks (nonlinear) can outperform either alone across multiple series.

2.4 Artificial Neural Networks (ANN) for time-series demand forecasting

Artificial Neural Networks (ANNs) have been adopted in retail and operations forecasting as flexible learners capable of capturing nonlinear interactions among lagged demand, calendar effects, and special-day influences that linear time-series models may not fully represent. On aggregate retail sales, **Alon, Qi and Sadowski**

(2001) compare ANNs with traditional benchmarks and report competitive accuracy across seasonal horizons, motivating ANN as a practical alternative when calendar structure interacts with demand. Examining sales series with potential interventions and nonlinearities, **Ansuj and Sharma (1996)** show that back-propagation networks can learn patterns not easily handled by linear autoregressive structures, supporting a complementary role alongside Box–Jenkins baselines. At product level for short shelf-life items where volatility and seasonality are pronounced **Doganis, Alexandridis, Patrinos and Sarimveis (2006)** develop nonlinear ANN models that improve short-term sales prediction, particularly when engineered inputs such as recent sales lags and encoded calendar/holiday indicators are included. More recently, for daily horizons with rich calendar structure, **Vallés-Pérez and Martínez-Ballesté (2022)** evaluate recurrent architectures for store-item sales and find gains over conventional methods in several settings, especially when networks exploit day-of-week, month, and special-day effects. While ANNs can deliver accuracy improvements, they require hyperparameter tuning and careful control of data leakage (e.g., fitting any scaling on training data and applying the same transformation to validation/test). They may be less transparent than Holt–Winters or ARIMA and can overfit when training samples are limited relative to model capacity trade-offs that motivate evaluating ANN alongside classical baselines under a common protocol.

2.5 Comparative evidence: Holt–Winters, ARIMA, and ANN

Comparative results across operational settings indicate that relative performance is data-dependent and horizon-specific. When structural breaks or interventions occur, ARIMA with explicit intervention terms can be advantageous; in a call-center application, **Bianchi, Jarrett and Hanumara (1998)** show that modeling disturbances within a Box–Jenkins framework improves accuracy, highlighting ARIMA's strength when shocks or level shifts must be accommodated. Where seasonality is strong and recurring, exponential-smoothing variants remain highly competitive; for short-term electricity demand with pronounced weekly/daily cycles, **Taylor, de Menezes and McSharry (2006)** report that advanced exponential-smoothing specifications designed for multiple seasonalities often match or exceed ARIMA at day-ahead horizons, underscoring Holt–Winters/ETS as a robust baseline when seasonal structure dominates. At broader industrial scale, **Hassani, Heravi and Zhigljavsky (2009)** observe broadly similar short-horizon accuracy for Holt–Winters and ARIMA when seasonality is stable, with differences emerging at longer horizons or under structural change. Relative to these statistical baselines, learning-based methods show complementary strengths. **Ho, Xie and Goh (2002)** find that ARIMA offers reliable short-horizon performance and stands as a credible benchmark against neural networks when linear autocorrelation drives predictability, while **Zhang (2003)** demonstrates that hybrid ARIMA–NN models can surpass either component alone, implying coexistence of linear and nonlinear signals. In retail contexts with rich exogenous information, **Ma, Fildes and Huang (2016)** show that adding promotions and category effects via regression with ARIMA errors (ARIMAX) materially improves accuracy, suggesting that feature-enriched models statistical or ANN benefit from explanatory signals beyond pure autocorrelation. These findings motivate the present study's side-by-side evaluation of Holt–Winters, ARIMA, and ANN under a shared train–validation–test protocol.

2.6 Forecast Accuracy Metrics

Choosing appropriate accuracy metrics is essential for fair model comparison and for aligning forecasts with operational decisions. Reviewing and systematizing measures used across time-series applications, **Hyndman and Koehler (2006)** recommend reporting multiple metrics and introduce the mean absolute scaled error (MASE) to enable comparability across series; they also discuss pitfalls of percentage-based measures such as MAPE and sMAPE, including undefined values at zero and small-value bias issues salient for daily retail sales. Empirical comparisons across 191 economic series show that the choice of error measure can materially affect method rankings; evaluating reliability, sensitivity, and robustness, **Armstrong and Collopy (1992)** argue for relative error measures when generalizing performance across heterogeneous series, reinforcing the need to avoid over-reliance on any single metric. Addressing claims that the symmetric MAPE fixes problems with MAPE, **Goodwin and Lawton (1999)** demonstrate that sMAPE is itself asymmetric for positive versus negative errors, particularly when actual values are small. Consistent with these insights, the present study emphasizes MAPE for headline comparability and reports

MAD, MSE, and RMSE as supporting measures, computed on non-differenced actuals and, where applicable, on integer-rounded forecasts to reflect the count nature of daily sales.

2.7 Research Gap and Practical Importance

Although prior studies establish Holt–Winters/ETS and ARIMA as strong baselines and document promising gains from neural networks, most evidence is scattered across different datasets, horizons, and evaluation protocols. Direct head-to-head comparisons on daily retail demand with long seasonal cycles ($s \approx 365$) are still limited, especially when models are tuned under a consistent train–validation–test split, exposed to the same calendar and holiday information, and judged with identical accuracy rules (e.g., rounding forecasts to integers before computing error to respect count data). In particular, few works examine how multi-step dynamic forecasting affects ARIMA behavior (e.g., the tendency to flatten under $d = 1$) relative to Holt–Winters and ANN on the same series and window, nor do they report a unified set of metrics (MAPE with MAD/MSE/RMSE) to triangulate conclusions.

This study addresses that gap by comparing Holt–Winters, ARIMA, and ANN on a single five-year, daily retail dataset using a shared, transparent protocol: (i) feature design aligned with operational signals (lags and calendar/holiday flags) where applicable; (ii) model selection on a held-out validation segment; and (iii) final assessment on a strict test horizon with consistent rounding and metric computation. The practical importance is direct: selecting a method that is demonstrably best for this data and horizon enables more reliable replenishment, leaner inventories, and better labor and transport planning. Beyond the specific dataset, the protocol itself common inputs, identical splits, and comparable error reporting offers a replicable template for organizations to evaluate forecasting options on their own series and deploy the most effective model with confidence.

CHAPTER 3
METHODOLOGY

3.1 Data

3.1.1 Data collection

The study draws on a publicly released daily-sales series from a Bangladeshi retailer hosted on Mendeley Data. The dataset was accessed for academic use under the repository's terms and serves as a transparent basis for replicating the analysis.

3.1.2 Data description

The file contains 1,826 consecutive calendar days from 2013-01-01 to 2017-12-31. Two fields are provided: a date stamp and the corresponding sales count for that day. This structure enables direct exploration of temporal features trend and recurring seasonal patterns relevant to short-horizon retail forecasting.

3.2 Forecast performance measures

To assess the predictive performance of the forecasting models, several statistical error metrics are employed. These measures provide a quantitative evaluation of forecast accuracy and enable comparison across different methods.

3.2.1 Definition of MAD, MSE, RMSE, and MAPE

Four error criteria were used. MAD summarizes the average absolute miss. MSE magnifies large deviations via squaring, while RMSE returns the result to the original unit by taking the square root. MAPE expresses error as a share of the observed value, which eases comparison across demand levels.

3.2.2 Equation of MAD

$$\text{MAD} = \frac{1}{n}\sum_{t=1}^{n}\left|Y_t - \hat{Y}_t\right| \tag{3.1}$$

3.2.3 Equation of MSE

$$\text{MSE} = \frac{1}{n}\sum_{t=1}^{n}(Y_t - \hat{Y}_t)^2 \qquad (3.2)$$

3.2.4 Equation of RMSE

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(Y_t - \hat{Y}_t)^2} \qquad (3.3)$$

3.2.5 Equation of MAPE

$$\text{MAPE} = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right| \qquad (3.4)$$

Where:

- Y_t = actual value at time t
- $\hat{Y}_t$ = forecasted value at time t
- n = number of observations

Among these measures, MAPE serves as the primary evaluation metric in this study because of its intuitive interpretation as the average percentage error. MAD, MSE, and RMSE are also calculated to provide additional insights into the magnitude and distribution of forecast errors.

3.3 Holt-Winters method

Holt–Winters exponential smoothing is a forecasting method specifically designed for time series data that exhibit both trend and seasonal variations. Unlike simple exponential smoothing, which only accounts for level, the Holt–Winters method incorporates two additional components: trend and seasonality. This makes it particularly suitable for datasets such as daily sales, where seasonal cycles can significantly influence demand patterns. There are two common variations of Holt–Winters: additive and multiplicative. The additive version is typically applied when seasonal fluctuations are relatively constant in magnitude, whereas the multiplicative version is used when the amplitude of seasonal variations changes proportionally with the level of the series.

3.3.1 Data preprocessing for Holt-Winters

To ensure fair model selection and evaluation, the dataset was split chronologically into training (first 1,729 days), validation (next 90 days), and test (final 7 days) with no shuffling. Smoothing parameters (α, β, γ)were tuned using a grid on the validation segment (design of experiments over preset levels, yielding 64 combinations), after fitting on the training data. The final HW configuration was then re-estimated as appropriate before generating forecasts for the test horizon. For comparability across methods and to respect count data, HW forecasts were rounded to the nearest integer prior to computing accuracy metrics (MAPE, MAD, MSE, RMSE) against the non-differenced actuals.

3.3.2 Parameter of Holt-Winters

The model involves three smoothing constants:

- Alpha (α): Controls the smoothing of the level component.
- Beta (β): Controls the smoothing of the trend component.
- Gamma (γ): Controls the smoothing of the seasonal component.

In this study, the optimal values of α, β, and γ were determined using grid search approach, starting with values of 0.01, 0.04, 0.07, and 0.10. By combining these values, 64 different parameter configurations were tested to identify the one that produced the lowest forecasting error.

3.3.3 Equations of Holt-Winters

Level equation

$$L_t = \alpha \left(\frac{Y_t}{S_{t-s}}\right) + (1- \alpha)\,(L_{t-1} + T_{t-1}) \tag{3.5}$$

Trend equation

$$T_t = \beta\,(L_t - L_{t-1}) + (1 - \beta)\,T_{t-1} \tag{3.6}$$

Seasonality equation

$$S_t = \gamma \left(\frac{Y_t}{L_t}\right) + (1 - \gamma)\,S_{t-s} \tag{3.7}$$

Forecast equation

$$\hat{Y}_{t+m} = (L_t + mT_t)\, S_{t-s+m} \tag{3.8}$$

Where:

- L_t = level component at time t
- T_t = trend component at time t
- S_t = seasonal component at time t
- s = season length
- m = forecast horizon

3.4 ARIMA model

ARIMA is a stochastic model for univariate time series in which, after differencing to achieve stationarity, the current value is expressed as a combination of past values and past shocks. Within the Box–Jenkins framework, model orders are written (p, d, q): p is the number of autoregressive lags, d is the degree of non-seasonal differencing, and q is the number of moving-average terms.

3.4.1 Data preprocessing for ARIMA

The daily series was split chronologically into train (1,729 days), validation (90 days), and test (7 days). All identification was done on train only: the series showed trend, so first-order differencing was applied ($d = 1$), and ACF/PACF of $\acute{Y}_t$ guided parsimonious (p, q)candidates (with/without constant). No exogenous variables and no seasonal differencing were used (non-seasonal ARIMA baseline).

Each candidate was fitted on train and validated with 90-day dynamic forecasts; accuracy was computed on the original (non-differenced) scale, with forecasts rounded to integers before calculating MAPE, MAD, MSE, RMSE. The best specification by validation MAPE was then re-estimated on train + validation and used to produce 7-step dynamic forecasts for the held-out test set under the same rounding and metrics.

3.4.2 Parameters of ARIMA

- Autoregressive (AR, p): Refers to the number of lagged observations included in the model.
- Integrated (I, d): Represents the number of differencing operations applied to the raw series to achieve stationarity.
- Moving Average (MA, q): Indicates the number of lagged forecast errors included in the prediction equation.

The selection of AR, I, and MA parameters was carried out using the Box–Jenkins methodology, which involves analyzing the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots. The ACF is useful for identifying potential MA terms, while the PACF is applied to determine possible AR terms. Before examining the ACF and PACF plots, the original sales series was checked for stationarity. Since the series exhibited a clear trend, first-order differencing (d = 1) was applied to stabilize the mean and remove non-stationarity. The differenced series was calculated as:

$$\acute{Y}_t = Y_t - Y_{t-1} \qquad (3.9)$$

Where Y_t represents the sales at time t and Y_{t-1} is the sales at the previous time step.

The resulting differenced series ($\acute{Y}_t$) was then used for ACF and PACF analysis to determine the appropriate AR and MA orders.

3.4.3 Equation of ARIMA

Following the formulation used in Minitab, the general ARIMA (p, d, q) model can be expressed as:

$$\hat{Y}_t = c + \phi_1 \acute{Y}_{t-1} + \phi_2 \acute{Y}_{t-2} + \ldots + \phi_p \acute{Y}_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t \qquad (3.10)$$

Where:

- $\acute{Y}_t$ = the differenced series at time t (after applying d differencing operations)
- c = constant term
- ϕ_i = coefficients of the autoregressive terms

- θ_j = coefficients of the moving average terms
- ε_t = random error term (white noise) at time t

3.5 Artificial Neural Network (ANN)

An Artificial Neural Network (ANN) is a data-driven, nonlinear function approximator that maps input features to a target by composing linear transformations with nonlinear activation functions. For univariate demand forecasting, a feed-forward network learns a direct relationship between recent demand lags and calendar effects and the next-day sales level, without requiring an explicit stochastic structure as in ARIMA or explicit decomposition as in Holt–Winters. In this study, we employ a fully connected feed-forward network with one or two hidden layers (sigmoid activation) and a linear output node suitable for continuous-valued forecasts.

3.5.1 Data preprocessing for ANN

The daily sales series was split chronologically into training (1,729 days), validation (90 days), and test (7 days). The dependent variable is the non-differenced daily sales $\boldsymbol{Y_t}$. To encode short-term dynamics and calendar effects, we constructed the following inputs: seven sales lags ($\boldsymbol{t-1}\ldots\ \boldsymbol{t-7}$), day-of-week dummies (Monday…Sunday), month dummies (January… December), a holiday flag (0/1), and optionally day of month (1–31). The date field was kept only as an identifier and not used as a predictor. Because ANN is trained on the original (non-differenced) scale, each lagged input is defined directly from past observations:

$$sales\_lag_k(t) = Y_{t-k} \tag{3.11}$$

For $t \leq k$, $\text{sales\_lag}_k(t)$ is undefined; after generating all lags ($k = 1 \ldots 7$), the first 7 rows were removed so every remaining record has a complete set of lag features.

3.5.1.1 Normalization and leakage control

When feature scaling was used, the transformer was fit on the training set only and then applied unchanged to validation and test. The target Y_t and binary dummies (day-of-week, month, holiday) were not scaled. For consistency with count data and with other models, ANN predictions were rounded to the nearest integer prior to computing MAPE, MAD, MSE, and RMSE against non-differenced actuals.

3.5.2 Parameters of ANN

The ANN is characterized by architectural and optimization parameters. Architectural choices include the number of hidden layers (1–2), the number of hidden nodes per layer, the hidden activation (sigmoid/logistic), and a linear output unit. Optimization choices include the number of training cycles (epochs), learning rate, momentum, and optional weight decay (L2 regularization). Random initialization and a fixed chronological split were used, no shuffling of time order or sequence windowing was performed. Hyperparameters were selected via grid search on the validation set under the fixed split, using MAPE as the primary selection criterion while also monitoring MAD, MSE, and RMSE for supporting evidence. The final configuration was retrained on the combined train + validation data before evaluation on the held-out test horizon.

3.5.3 Equation of ANN (single hidden layer)

The model used is a feed-forward neural network with one hidden layer (linear output). Its mapping is:

$$\hat{Y}_t = \mathrm{w}_2\, \sigma(\mathrm{W}_1 x_t + b_1) + b_2 \tag{3.12}$$

Where:

- $\mathrm{x}_t \in \mathbb{R}^K$: input feature vector at day t, concatenating $[Y_{t-1}, \dots, Y_{t-7}]$ (sales lags), one-hot day-of-week (Mon–Sun), one-hot month (Jan–Dec), holiday flag (0/1), and (optionally) day-of-month; K = total number of features.
- $\hat{Y}_t \in \mathbb{R}$: predicted sales for day t (real-valued before rounding).
- $\mathrm{W}_1 \in \mathbb{R}^{H\times K}$, $\mathrm{b}_1 \in \mathbb{R}^H$: weights and biases of the hidden layer with H units.
- $\mathrm{w}_2 \in \mathbb{R}^H$, $b_2 \in \mathbb{R}$: weights and bias of the linear output unit.
- σ: element-wise hidden activation (sigmoid/logistic in this study).
- H: number of hidden units; K: number of input features.

Predicted values $\hat{Y}_t$ are rounded to the nearest integer (to reflect count data) before computing MAPE, MAD, MSE, and RMSE against non-differenced actuals.

CHAPTER 4
RESULT

4.1 Holt–Winters result

4.1.1 DOE on validation (model selection)

Holt–Winters (multiplicative seasonality, $s = 365$) was fit on the training set (days 1–1,729). Model selection used grid search over smoothing parameters $(\alpha, \beta, \gamma) \in \{0.01, 0.04, 0.07, 0.10\}^3$ (64 combinations).

For each combination, the model produced 90-day dynamic forecasts for the validation window. Forecasts were rounded to the nearest integer before computing MAPE on non-differenced actuals (with MAD, MSE, RMSE recorded as supporting metrics).

Table 4.1 Validation MAPE for Holt–Winters grid search (64 combinations; multiplicative, $s = 365$)

Level	Trend	Seasonal	MAPE
0.01	0.01	0.01	28.19
0.01	0.01	0.04	28.43
0.01	0.01	0.07	28.42
0.01	0.01	0.1	28.42
0.01	0.04	0.01	28.90
0.01	0.04	0.04	28.99
0.01	0.04	0.07	28.82
0.01	0.04	0.1	28.98
0.01	0.07	0.01	30.18
0.01	0.07	0.04	30.22
0.01	0.07	0.07	30.18
0.01	0.07	0.1	29.61
0.01	0.1	0.01	28.41
0.01	0.1	0.04	28.19
0.01	0.1	0.07	28.05
0.01	0.1	0.1	28.45
0.04	0.01	0.01	28.49
0.04	0.01	0.04	28.51
0.04	0.01	0.07	28.73

0.04	0.01	0.1	28.58
0.04	0.04	0.01	28.44
0.04	0.04	0.04	28.53
0.04	0.04	0.07	28.69
0.04	0.04	0.1	28.71
0.04	0.07	0.01	28.72
0.04	0.07	0.04	28.67
0.04	0.07	0.07	28.74
0.04	0.07	0.1	28.74
0.04	0.1	0.01	28.10
0.04	0.1	0.04	28.23
0.04	0.1	0.07	28.59
0.04	0.1	0.1	28.63
0.07	0.01	0.01	28.45
0.07	0.01	0.04	28.46
0.07	0.01	0.07	28.44
0.07	0.01	0.1	28.45
0.07	0.04	0.01	28.22
0.07	0.04	0.04	28.28
0.07	0.04	0.07	28.34
0.07	0.04	0.1	28.39
0.07	0.07	0.01	28.03
0.07	0.07	0.04	28.09
0.07	0.07	0.07	28.12
0.07	0.07	0.1	28.04
0.07	0.1	0.01	28.65
0.07	0.1	0.04	28.60
0.07	0.1	0.07	28.45
0.07	0.1	0.1	28.31
0.1	0.01	0.01	28.44
0.1	0.01	0.04	28.53
0.1	0.01	0.07	28.51
0.1	0.01	0.1	28.47
0.1	0.04	0.01	27.92
0.1	0.04	0.04	28.00
0.1	0.04	0.07	27.96
0.1	0.04	0.1	28.14
0.1	0.07	0.01	28.17
0.1	0.07	0.04	28.13
0.1	0.07	0.07	28.03
0.1	0.07	0.1	28.03
0.1	0.1	0.01	28.16
0.1	0.1	0.04	28.17

0.1	0.1	0.07	28.09
0.1	0.1	0.1	28.14

Table 4.2 DOE result for the best Holt-Winters model

Level	Trend	Seasonal	MAPE
0.1	0.04	0.01	27.92

The best validation configuration was Level = 0.10, Trend = 0.04 and Seasonal = 0.01 with MAPE = 27.92

4.1.2 Validation fit of the winning model

Figure 4.1 shows Actual vs Forecast on the validation window (90 days) for the winning Holt–Winters configuration. The forecasts follow the short-run level and the recurring seasonal pattern implied by $s = 365$, with visible deviations around holiday-related spikes-typical for daily retail demand.

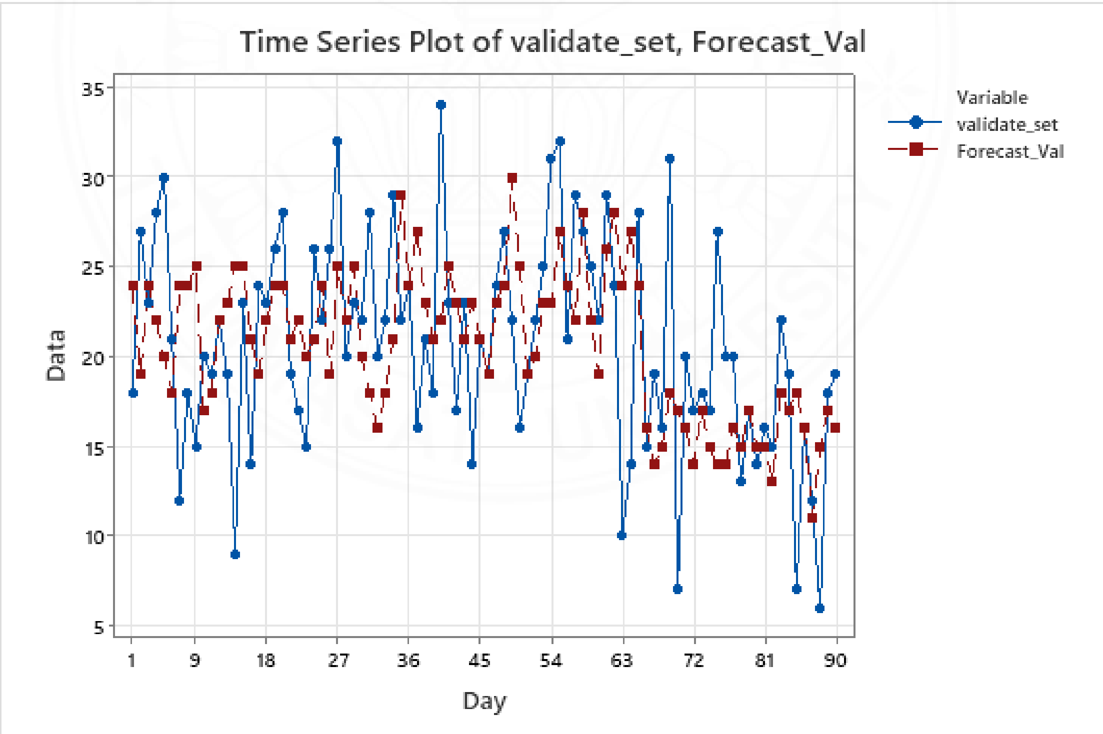

Figure 4.1 Validation (90 days): actual vs forecast (Holt-Winters)

Interpretation. The model reproduces the main seasonal cycle and adapts to local movements sufficiently to attain the lowest validation MAPE among the DOE candidates.

4.1.3 Test-set evaluation (using the selected parameters)

After selection, the Holt–Winters model was re-estimated on train + validation with the winning parameters and used to generate 7-day dynamic forecasts for the test set. As in validation, predictions were rounded to integers prior to scoring against non-differenced actuals. The resulting accuracy was MAPE = 21.12 (with MAD, MSE, RMSE computed under the same rule).

Table 4.3 Holt–Winters accuracy under the split protocol

Data	MAPE	MAD	MSE	RMSE
Validation (90 days)	27.92	4.66	36.37	6.03
Test (7 days)	21.12	4.14	25.00	5.00

Remark. Using a common split and identical rounding/metrics ensures a fair comparison with ARIMA and ANN in subsequent sections.

4.2 ARIMA result

4.2.1 Model identification (ACF/PACF–based orders)

Using the training segment only, the series was first differenced to address non-stationarity, fixing d = 1. Order cues were then taken from the correlograms of the differenced series. The PACF shows clear early significance through lags 1–4 before tapering, so autoregressive orders $p \in \{1,2,3,4\}$ were deemed plausible as in figure 4.2

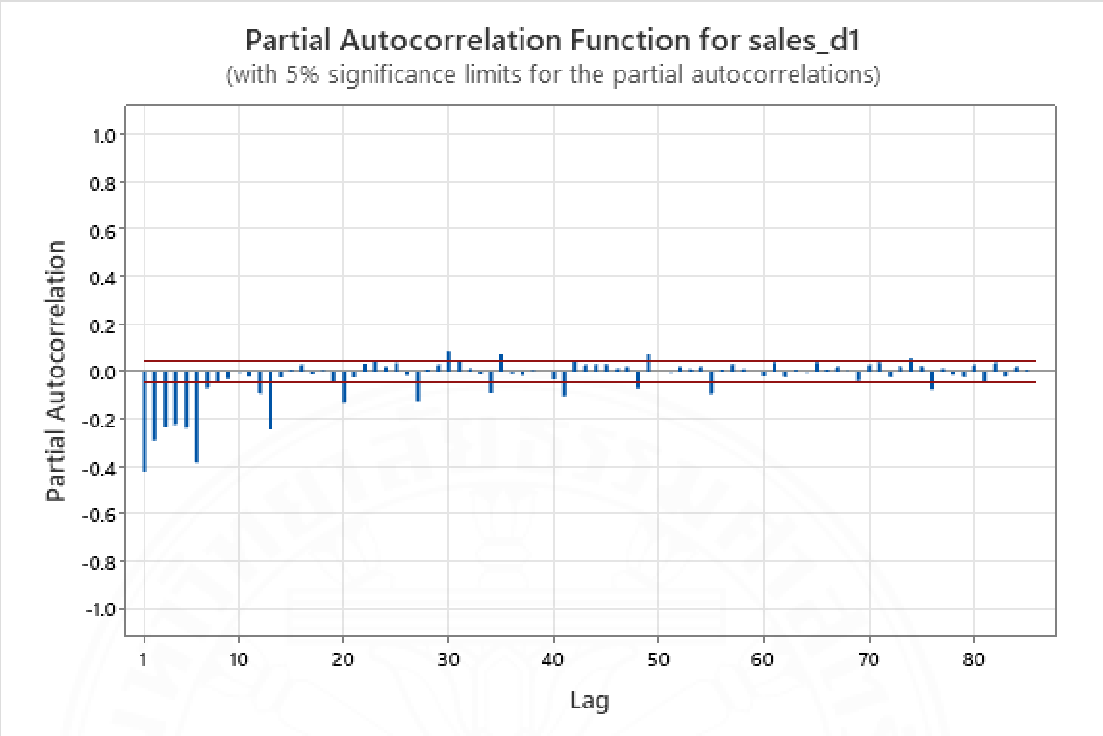

Figure 4.2 Partial Autocorrelation Function (PACF) of differenced data

The ACF displays a pronounced first-lag effect with a gradual decay rather than a sharp multi-lag cut-off, suggesting at most a short moving-average component, i.e., $q \in \{0,1\}$ as in figure 4.3

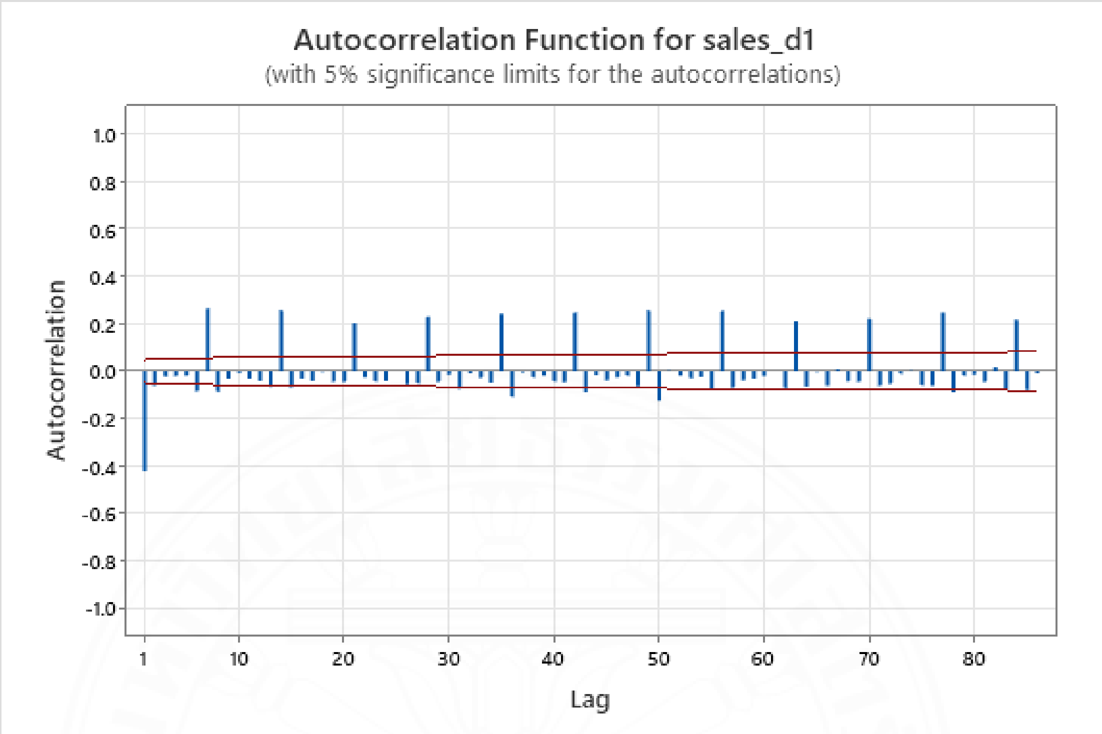

Figure 4.3 Autocorrelation Function (ACF) of differenced data

These readings set the search space to ARIMA($p, 1, q$) with $p \in \{1,2,3,4\}$ and $q \in \{0,1\}$.

4.2.2 Candidate set and validation selection

All candidates ARIMA($p, 1, q$) with $p \in \{1,2,3,4\}$ and $q \in \{0,1\}$ (each with/without a constant) were estimated on the train set and scored on the 90-day validation window under a common rule: forecasts were rounded to the nearest integer and errors computed on the original (non-differenced) sales. Based on lowest validation MAPE.

Table 4.4 Validation accuracy of ARIMA candidates

ARIMA	MAPE
4,1,0	33.98
4,1,1	37.01
3,1,0	31.29
2,1,0	29.58
1,1,1	37.16

1,1,0	28.23

Table 4.5 The best accuracy of ARIMA model

ARIMA	MAPE
1,1,0	28.23

The selected specification was ARIMA (1,1,0) (no constant) with MAPE = 28.23%. Higher p and/or $q = 1$ did not yield lower error.

4.2.3 Validation fit of the selected model

Figure 4.4 plots actuals versus forecasts for the 90-day validation horizon using the selected ARIMA (1,1,0). As expected for a non-seasonal ARIMA with $d = 1$ under multi-step use, the forecast profile flattens toward a local level, capturing short-run persistence but not the pronounced seasonal swings. With integer rounding applied prior to scoring, the validation performance was MAPE = 28.23%.

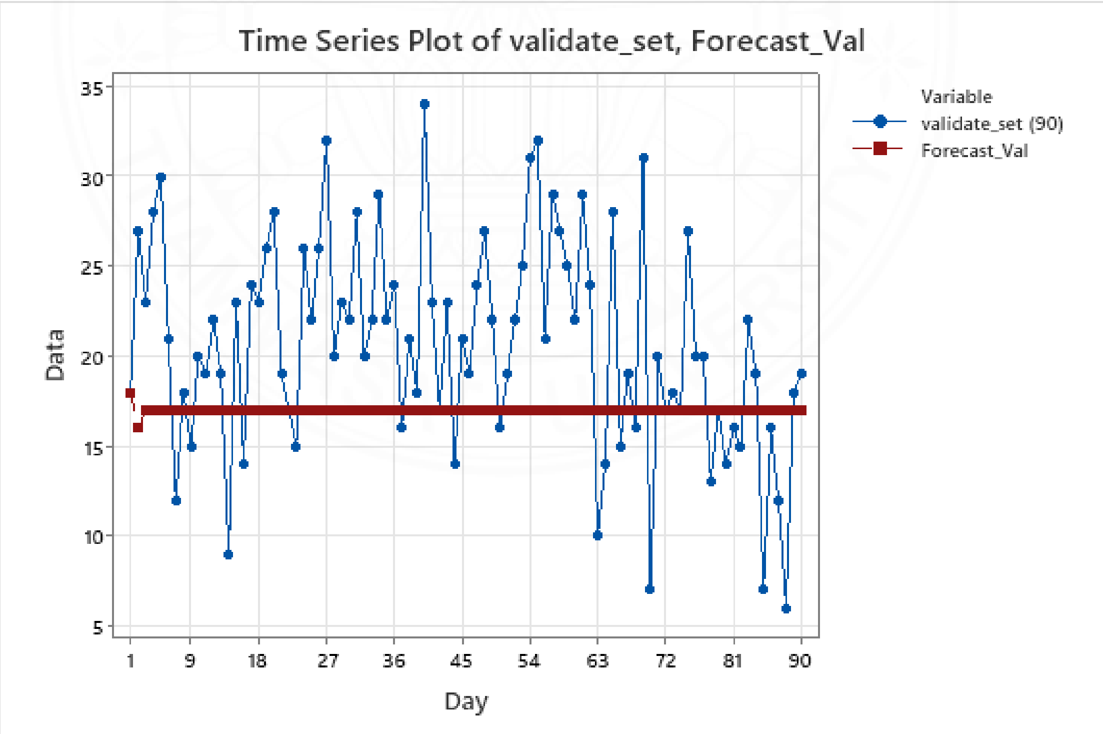

Figure 4.4 Validation (90 days): actual vs forecast (ARIMA (1,1,0))

4.2.4 Test-set evaluation (using the selected parameters)

The winning ARIMA (1,1,0) was re-estimated on train + validation and used to generate 7-day test forecasts, evaluated under the same rounding and metric protocol. The model obtained MAPE = 24.90% on the test set, consistent with validation and confirming ARIMA (1,1,0) as a clear, level-tracking baseline for comparison with Holt–Winters and ANN.

Table 4.6 ARIMA (1,1,0) accuracy under the split protocol

Data	MAPE	MAD	MSE	RMSE
Validation (90 days)	28.23	5.56	48.40	6.96
Test (7 days)	24.90	4.29	23.71	4.87

Remark. Using a common split and identical rounding/metrics ensures a fair comparison with Holt-Winters and ANN in subsequent sections.

4.3 Artificial Neural Network (ANN) results

4.3.1 Model setup and feature design

The ANN is a feed-forward network for single-step regression. Inputs are the engineered time-series features prepared in Chapter 3: seven recent sales lags (sales $\{t$-$1\}$ … sales $\{t$-$7\}$), calendar indicators (day-of-week dummies, month dummies, day-of-month), and a holiday flag. The target is the next-day sales value in levels (no differencing). The training/validation/test protocol follows the same rolling split used for other models: Train = 1,729 days, Validation = 90 days, Test = 7 days. Predictions are rounded to the nearest integer before scoring, and all metrics are computed on the original sales for fair comparison.

4.3.2 Hyper-parameter search (validation)

Architecture and optimizer settings were tuned by grid search on the validation window while fitting on the training segment only. The search covered a single hidden layer (a second layer was trialed but did not improve validation error) with the following optimizer ranges:

- training_cycles: 100 - 1000 (9 linear steps)
- learning_rate: 0.005 - 0.010 (5 linear steps)
- momentum: 0.1 - 0.9 (8 linear steps)
- decay: {true, false}

Table 4.7 Validation accuracy across candidate ANN settings

Architecture (hidden)	MAPE
1 layer, 7 nodes	20.93
1 layer, 14 nodes	20.87
1 layer, 21 nodes	21.77
1 layer, 28 nodes	20.90
2 layer, 14, 7 nodes	21.17
2 layer, 21, 7 nodes	21.50
2 layer, 21, 14 nodes	21.31
2 layer, 28, 14 nodes	21.47

Table 4.8 The best accuracy of ANN model

Architecture (hidden)	MAPE
1 layer, 14 nodes	20.87

Hidden-unit counts from small to moderate sizes were tried; validation error consistently favored a compact single-layer network. The best validation configuration was:

- Hidden layer: 1 layer, 14 nodes
- training_cycles = 800, learning_rate = 0.009, momentum = 0.8, decay = false

This model achieved MAPE = 20.87% on the 90-day validation window under the rounding rule.

4.3.3 Validation fit of the selected ANN

The selected 1×14 network tracks the level and the recurring weekly pattern materially better than linear baselines, while smoothing isolated spikes. The 90-day validation plot (actual vs forecast) should be inserted here.

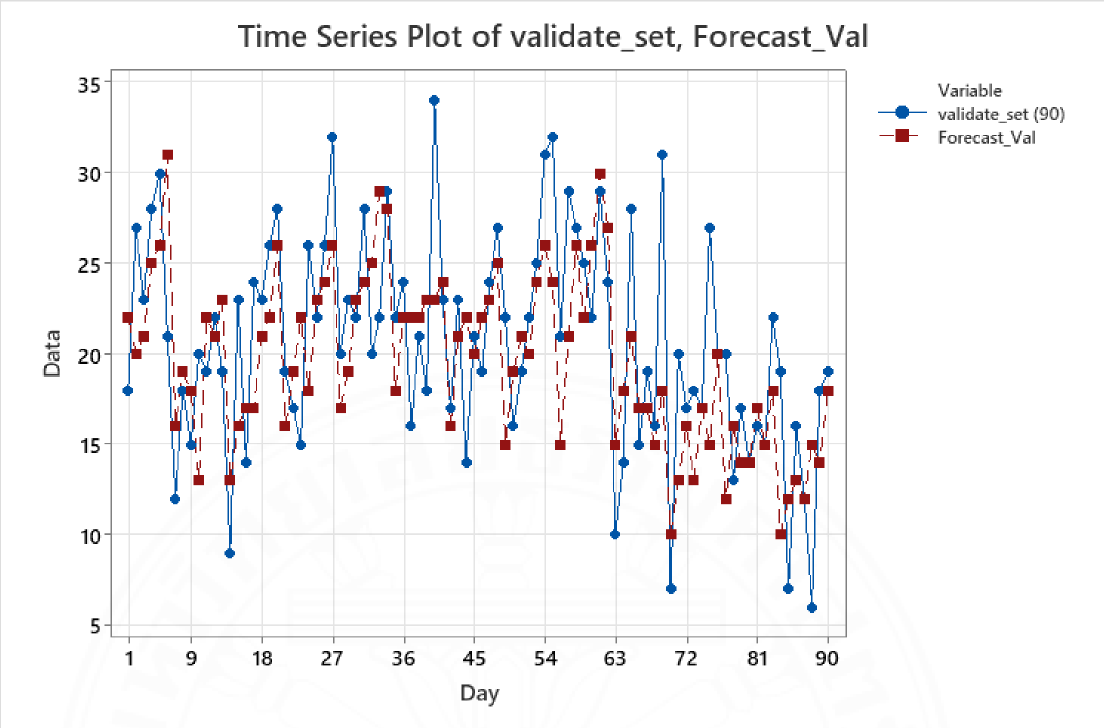

Figure 4.5 Validation (90 days): actual vs forecast (ANN 1×14)

4.3.4 Test-set evaluation (using the selected parameters)

The winning configuration above was retrained on Train + Validation and used to forecast the 7-day test horizon. Under the same integer-rounding and metric protocol.

Table 4.9 ANN (1×14) accuracy under the split protocol

Data	MAPE	MAD	MSE	RMSE
Validation (90 days)	20.87	3.89	23.49	4.85
Test (7 days)	13.47	2.71	15.57	3.95

Remark. Using a common split and identical rounding/metrics ensures a fair comparison with Holt-Winters and ARIMA in subsequent sections.

4.4 Comparative analysis of models

All three approaches were trained, tuned, and tested under the same protocol: the Train (1,729 days) segment was used to fit candidates, Validation (90 days) to select hyper-parameters, and the Test (7 days) horizon to report final generalization. Forecasts from every model were rounded to the nearest integer before scoring, and all metrics were computed on the original (non-differenced) sales to ensure a fair, like-for-like comparison.

Table 4.10 Side-by-side MAPE under the common split/rounding protocol

Model	Validation MAPE	Test MAPE
Holt-Winters	27.92	21.12
ARIMA	28.23	24.90
ANN	20.87	13.47

On the validation window, the ANN reduces MAPE by ~ 7–8 percentage points relative to Holt–Winters and ARIMA, indicating better fit to short-run dynamics given the lag and calendar inputs. On the test horizon, the gap widens: the ANN achieves 13.47%, versus 21.12% for Holt–Winters and 24.90% for ARIMA. The pattern is consistent with the qualitative behavior of the models:

- ARIMA (1,1,0) tends to flatten under multi-step use, tracking local level but not the recurring within-week fluctuations, which raises percentage error.
- Holt–Winters multiplicative handles broad seasonal structure but remains linear in level/trend/season components; without explicit weekday effects it underfits some short-run variation.
- The ANN captures nonlinear interactions between recent lags and calendar/holiday indicators, improving day-to-day tracking without manual seasonal specification.

These results position the ANN (1×14) as the strongest model under the agreed evaluation, with Holt–Winters a competitive classical baseline and ARIMA serving as a simple level-tracking benchmark. The common rounding rule and metric computation on original sales support a transparent.

CHAPTER 5
CONCLUSION

This study compared three time-series approaches for daily sales forecasting under a common and fair protocol. All models were trained on the same training set (1,729 days), tuned on a 90-day validation window, and finally evaluated on a 7-day test horizon. To align with operational use, fitted/forecast values were rounded to the nearest integer, and all errors were computed on the original (non-differenced) sales series.

5.1 Summary of findings

Holt–Winters (multiplicative, $s = 365$). After a grid search, the best setting on the validation window was $\alpha = 0.10$, $\beta = 0.04$, $\gamma = 0.01$. MAPE: Validation 27.92%, Test 21.12%.

ARIMA. Based on ACF/PACF identification with $d = 1$ and a small q, the best candidate on validation was ARIMA (1,1,0) (no constant). MAPE: Validation 28.23%, Test 24.90%.

Artificial Neural Network (ANN). A one-hidden-layer network with 14 nodes trained on lag features ($t-1 \ldots t-7$) and calendar/holiday indicators delivered the strongest results. The tuned training settings were training cycles = 800, learning rate=0.009, momentum=0.8, decay=false. MAPE: Validation 20.87%, Test 13.47%.

Overall, the ANN (1×14) clearly outperformed the classical baselines on both validation and test. Holt–Winters was the best of the two statistical models, while ARIMA served as a transparent, level-tracking benchmark.

5.2 Interpretation and implications

Short-run nonlinear structure matters. Incorporating recent lags and calendar/holiday effects allowed the ANN to capture within-week variation that Holt–Winters (level–trend–season) and ARIMA (1,1,0) did not fully express.

Classical baselines remain useful. Holt–Winters provided competitive accuracy with minimal feature engineering and interpretable components; ARIMA offered a simple check that the series' short-run persistence is being modeled correctly.

Operational takeaway. For short-horizon daily planning, the ANN should be the primary forecasting engine. Holt–Winters is a robust fallback when model transparency or rapid deployment is required. ARIMA may be retained as a sanity-check baseline and for scenarios with structural breaks where intervention modeling is appropriate.

5.3 Practical recommendations

- Deployment. Automate a pipeline that (i) creates lag and calendar/holiday features, (ii) applies the trained ANN, and (iii) rounds forecasts to integers for downstream systems.
- Retraining policy. Refit the ANN on a rolling basis (e.g., monthly) or when monitoring reveals drift (e.g., MAPE deteriorates beyond a threshold).
- Governance. Keep Holt–Winters live as a back-up model; compare nightly to detect anomalies. Log forecasts and realized errors for continuous improvement.

5.4 Limitations

- Single series & short test horizon. Results are for one daily series with a 7-day test; longer and multiple-series evaluations could alter rankings.
- Feature scope. Only lags and calendar/holiday indicators were used; promotions, price, weather, or other exogenous drivers were not included.
- Deterministic point forecasts. The study reports point errors (MAPE/MAD/MSE/RMSE); uncertainty quantification (prediction intervals or quantiles) was not addressed.
- Protocol choices. Integer rounding before error calculation affects MAPE; although applied consistently, different rounding rules might shift absolute values.

5.5 Future work

- Richer features: incorporate promotions, pricing, store events, and weather (ARIMAX/ETSX/ML models).
- Longer & rolling evaluation: adopt rolling-origin cross-validation and extend the test horizon to assess stability.
- Model families: compare with LSTM/Temporal CNN/Transformer baselines; explore gradient-boosted trees and hybrid/ensemble methods (e.g., HW + ANN).
- Probabilistic forecasting: produce quantile forecasts for service-level and inventory decisions; evaluate with pinball loss.
- Multiple seasonality: test ETS models with multiple seasonal cycles and Fourier terms (daily/weekly/annual).
- Automated monitoring: deploy drift detectors and scheduled hyper-parameter re-tuning.

5.6 Concluding statement

Under a unified and transparent evaluation, the ANN with one hidden layer (14 nodes) delivered the lowest error and the most accurate short-horizon forecasts, reducing MAPE materially relative to Holt–Winters and ARIMA. For day-to-day demand planning, adopting the ANN as the primary forecaster backed by Holt–Winters and ARIMA as interpretable safeguards offers a practical, high-accuracy solution that can be incrementally enhanced with additional features, probabilistic outputs, and systematic monitoring.

REFERENCES

Alon, I., Qi, M., & Sadowski, R. J. (2001). Forecasting aggregate retail sales: A comparison of artificial neural networks and traditional methods. *International Journal of Retail & Distribution Management, 29*(10), 477–486. https://doi.org/10.1016/S0969-6989(00)00011-4

Ansuj, A. P., & Sharma, D. (1996). Sales forecasting using time series and neural networks. *Computers & Industrial Engineering, 31*(1–2), 421–424. https://doi.org/10.1016/0360-8352(96)00166-0

Bianchi, L., Jarrett, J. E., & Hanumara, R. C. (1998). Improving forecasting for telemarketing centers by ARIMA modeling with intervention. *International Journal of Forecasting, 14*(4), 497–504. https://doi.org/10.1016/S0169-2070(98)00037-5

De Gooijer, J. G., & Hyndman, R. J. (2006). 25 years of time series forecasting. *International Journal of Forecasting, 22*(3), 443–473. https://doi.org/10.1016/j.ijforecast.2006.01.001

Doganis, P., Alexandridis, A., Patrinos, P., & Sarimveis, H. (2006). Time series sales forecasting for short shelf-life food products based on artificial neural networks and evolutionary computing. *Computers & Operations Research, 33*(11), 3125–3131. https://doi.org/10.1016/j.jfoodeng.2005.03.056

Gardner, E. S. (2006). Exponential smoothing: The state of the art—Part II. *International Journal of Forecasting, 22*(4), 637–666. https://doi.org/10.1016/j.ijforecast.2006.03.005

Goodwin, P., & Lawton, R. (1999). On the asymmetry of the symmetric MAPE. *International Journal of Forecasting, 15*(4), 405–408. https://doi.org/10.1016/S0169-2070(99)00007-2

Hassani, H., Heravi, S., & Zhigljavsky, A. (2009). Forecasting European industrial production with singular spectrum analysis. *International Journal of Forecasting, 25*(1), 103–118. https://doi.org/10.1016/j.ijforecast.2008.09.007

Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting, 20*(1), 5–10. https://doi.org/10.1016/j.ijforecast.2003.09.015

Ho, S. L., Xie, M., & Goh, T. N. (2002). A comparative study of neural network and Box–Jenkins ARIMA modeling in time series prediction. *Computers & Industrial Engineering, 42*(2–4), 371–375. https://doi.org/10.1016/S0360-8352(02)00036-0

Hyndman, R. J., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting, 18*(3), 439–454. https://doi.org/10.1016/S0169-2070(01)00110-8

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting, 22*(4), 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001

Jarrett, J. E. (1989). Forecasting monthly earnings per share—Time series models. *Omega, 17*(1), 37–44. https://doi.org/10.1016/0305-0483(89)90018-2

Ma, S., Fildes, R., & Huang, T. (2016). Demand forecasting with high dimensional data: The case of SKU retail sales forecasting with intra- and inter-category promotional information. *European Journal of Operational Research, 249*(1), 245–257. https://doi.org/10.1016/j.ejor.2015.08.029

Makridakis, S., & Hibon, M. (2000). The M3-Competition: Results, conclusions and implications. *International Journal of Forecasting, 16*(4), 451–476. https://doi.org/10.1016/S0169-2070(00)00057-1

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting, 36*(1), 54–74. https://doi.org/10.1016/j.ijforecast.2019.04.014

Ong, C.-S., Huang, J.-J., & Tzeng, G.-H. (2005). Model identification of ARIMA family using genetic algorithms. *Applied Mathematics and Computation, 164*(3), 885–912. https://doi.org/10.1016/j.amc.2004.06.044

Taylor, J. W. (2003). Exponential smoothing with a damped multiplicative trend. *International Journal of Forecasting, 19*(4), 715–725. https://doi.org/10.1016/S0169-2070(03)00003-7

Taylor, J. W. (2007). Forecasting daily supermarket sales using exponentially weighted quantile regression. *European Journal of Operational Research, 178*(1), 154–167. https://doi.org/10.1016/j.ejor.2006.02.006

Taylor, J. W., de Menezes, L. M., & McSharry, P. E. (2006). A comparison of univariate methods for forecasting electricity demand up to a day ahead. *International Journal of Forecasting, 22*(1), 1–16. https://doi.org/10.1016/j.ijforecast.2005.06.006

Vallés-Pérez, I., & Martínez-Ballesté, A. (2022). Approaching sales forecasting using recurrent neural networks and transformers. *Expert Systems with Applications, 199*, 116962. https://doi.org/10.1016/j.eswa.2022.116993

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing, 50*, 159–175. https://doi.org/10.1016/S0925-2312(01)00702-0

APPENDICES

APPENDIX A
DATASET (SOURCE & DATA DICTIONARY)

A.1 Source and coverage

This study uses a daily retail sales dataset provided via Mendeley Data (Bangladeshi Retailer). The dataset covers 01 January 2013 to 31 December 2017, totaling 1,826 calendar days. The data were downloaded from the public repository and used solely for academic purposes in accordance with the repository's terms.

Dataset DOI: https://data.mendeley.com/datasets/xwmbk7n3c8/1?utm_source

A.2 Data dictionary (brief)

Field	Type	Description
date	Date (YYYY-MM-DD)	Calendar date of the observation
sales_t	Integer (count)	Units sold on date t for the focal item/category

APPENDIX B
PREPROCESSING & EVALUATION SETTINGS

B.1 Preprocessing steps (by model, with split protocol)

Data split. The daily sales series was partitioned chronologically into Train = 1,729 days, Validation = 90 days, and Test = 7 days. Model selection was performed on the validation window; the winning settings were then refit on Train + Validation and evaluated on the Test window.

Holt–Winters (HW)

- Input: original sales series Y_t (no differencing).
- Seasonality: $s = 365$ (annual), multiplicative specification.
- DOE/grid on $\alpha, \beta, \gamma \in \{0.01, 0.04, 0.07, 0.10\}$ using Train; forecasts generated for Validation to select by lowest MAPE.

ARIMA

- Differencing: first difference $Y_t' = Y_t - Y_{t-1}$ on Train to enforce stationarity; ACF/PACF used for order cues.
- Candidate set: ARIMA $(p, 1, q)$ with $p \in \{1,2,3,4\}$, $q \in \{0,1\}$ (with/without constant).
- Models estimated on Train; 90-day Validation forecasts scored to select by MAPE.

Artificial Neural Network (ANN)

- Features: sales lags Y_{t-k} for $k = 1, \dots, 7$; calendar dummies (day-of-week, month), day-of-month, and holiday flag.
- Architecture search on Train with Validation scoring; 1 hidden layer; node count R tuned.
- Training hyperparameters tuned by grid: training cycles, learning rate, momentum, weight decay (on/off).

B.2 Integer rounding rule

Because sales are counts, all validation/test forecasts were rounded to the nearest integer (half-up) prior to error calculation. This preserves interpretability and ensures a fair, common scoring protocol across HW, ARIMA, and ANN.

B.3 Evaluation protocol

- Target for scoring: original (non-differenced) sales Y_t.
- Metrics reported: MAPE (primary), plus MAD, MSE, RMSE.
- Selection: lowest Validation MAPE determines the winning configuration within each model family.
- Test procedure: refit the winning model on Train + Validation and score the 7-day Test under the same rounding/metrics.

BIOGRAPHY

Name	Nonthaphan Kingrungphet
Education	2024: Bachelor of Science (Engineering Management) Sirindhorn International Institute of Technology Thammasat University