Evaluating Classical Statistical Models for Short-Term Retail Sales Forecasting

Group 7: Thom van der Ent (13372971), Sten Paffen (13896342), Merlijne Reeuwijk (13706039), Floor Verkade, (14615215)
Applied Forecasting in Complex Systems 2025
University of Amsterdam

ABSTRACT

This study examines whether classical statistical forecasting models improve short-term retail sales forecasts relative to a seasonal naïve benchmark. Using daily sales data for 823 food products from the M5 Walmart dataset, ARIMA, SARIMAX, and a Time Series Linear Model (TSLM) were compared in a pilot study using training and validation data. All structured models outperformed the seasonal naïve baseline. TSLM was selected for full-scale evaluation due to its scalability and stability. TSLM reduced RMSE by 13.8% on the validation set and achieved an RMSE of 2.16 on the unseen evaluation set, compared to 2.65 for the seasonal naïve model. The results demonstrate that incorporating external covariates improves short-term retail sales forecasts beyond seasonal baselines.

1 INTRODUCTION

Accurate sales forecasting is a critical challenge in retail, as it directly affects inventory management, customer satisfaction, and operational efficiency [6, 9]. Supermarkets employ various promotional strategies such as price discounts, special events, and in-store promotions to stimulate demand [8]. While these strategies are widely used, their actual impact on sales can be difficult to quantify due to the complex and dynamic nature of consumer behaviour.

Sales data are influenced not only by historical demand patterns but also by external factors such as promotions, pricing changes, day-of-week effects, and holidays [1]. Traditional forecasting approaches often rely solely on past sales, failing to capture the influence of these explanatory variables [1, 7]. As a result, forecasts may be inaccurate, particularly during promotional periods when demand deviates from historical trends [5].

Various forecasting models exist, ranging from classical statistical methods, such as exponential smoothing and ARIMA/SARIMA [5], to modern machine learning approaches, including tree-based ensembles [3] and neural networks [2]. Each model handles seasonal patterns, promotions, and variability in the data differently. Consequently, it is not immediately clear which model is most suitable for accurately predicting daily sales in a real-world retail environment.

Using a subset of the M5 Forecasting dataset from Walmart, this study focuses on forecasting daily sales for the next 28 days at a single store (TX3). Multiple classical statistical forecasting models are compared using the same input data and evaluation procedure to assess their short-term predictive performance. The objective is to determine which statistical model performs best in a retail forecasting context and to evaluate how the inclusion of promotional and calendar-related explanatory variables affects forecast accuracy.

This leads to the following research question: *"To what extent does the best-performing classical statistical forecasting model improve short-term retail sales forecasts compared to a seasonal naïve benchmark?"* To help answer this question the following sub-question is used:

- How do different classical statistical forecasting models (ARIMA, SARIMAX, and TSLM) perform relative to a seasonal naïve baseline?

Hypotheses:

- H1: Structured classical statistical forecasting models will outperform the seasonal naïve benchmark in short-term retail sales forecasting.

2 DATASET

2.1 Description of dataset

The dataset includes historical daily unit sales, product information (item, department, and category), store details, prices, calendar events, and promotional indicators. The subset contains 3,049 products from three categories (Hobbies, Foods, and Household) and seven departments. For this project, the focus is on the Food3 products sold at the TX3 store.

The dataset consists of six files:

- **calendar_afcs2025.csv**: Contains date information, weekday, month, year, events, and SNAP eligibility.
- **sell_prices_afcs2025.csv**: Contains weekly product prices per store.
- **sales_train_validation_afcs2025.csv**: Historical daily unit sales used for training models.
- **sales_test_validation_afcs2024.csv**: Sales data for the next 28 days used for testing forecasts.
- **sample_submission_afcs2025.csv**: Template for submitting point forecasts for 28 days (F1–F28).

- **sales_test_evaluation_afcs_2025.csv**: Reserved for evaluating model performance; should not be used for training.

The relationships between these files are illustrated in an Entity-Relationship (ER) diagram which is provided in Appendix A.1.

2.2 Exploratory Data analysis

To better understand the sales dynamics at the TX3 Walmart store, an exploratory data analysis was conducted. This analysis examines general trends, seasonal patterns, autocorrelation, and provides summary statistics of daily sales. The obtained insights will be used to guide feature engineering and model selection for sales forecasting.

2.2.1 General trend

Figure 1 shows the daily total sales for all products in TX3, along with 7-day (blue) and 28-day (red) moving averages. The short-term (weekly) and long-term trends are highlighted. Peaks correspond to high-demand days or promotional periods.

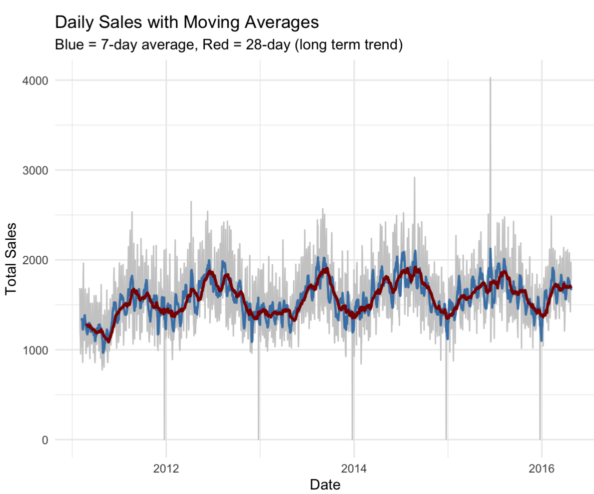

Figure 1: General Trend (weekly and monthly).

2.2.2 Seasonality

Figure 9 shows the distribution of daily sales per weekday. Sales tend to peak on weekends. Figure 10 illustrates monthly seasonality. The seasonal and subseries plots (Figures 2 and 3) further confirm recurring weekly and monthly patterns.

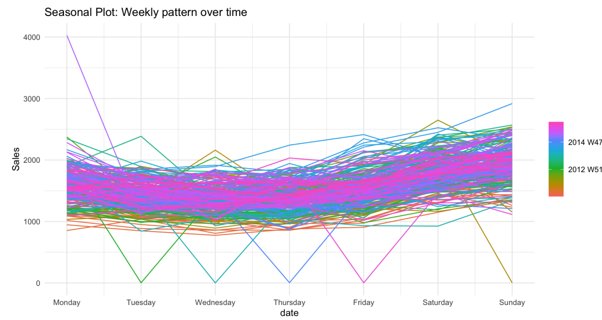

Figure 2: Seasonal subseries plot (weekly).

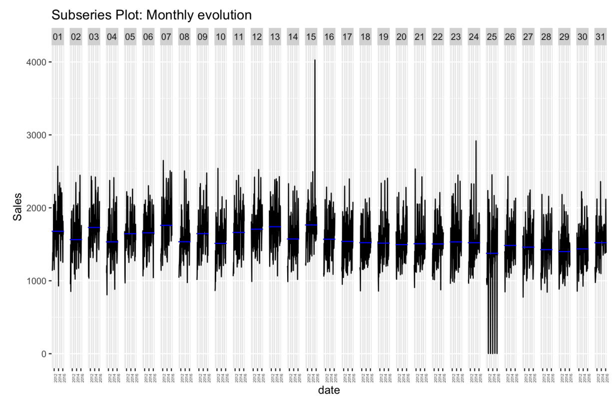

Figure 3: Seasonal subseries plot (monthly).

2.2.3 Autocorrelation

Figure 4 shows the autocorrelation function (ACF) of daily sales. Clear peaks at lags of 7, 14, and 21 days indicate a strong weekly seasonality in sales.

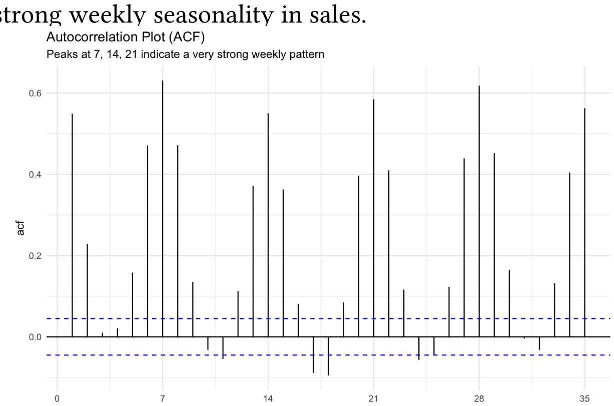

Figure 4: Autocorrelation function of daily sales.

2.2.4 Summary Statistics

Table 1 presents key statistics of daily total sales at the TX3 store.

Statistic	Value
Mean Sales	1567
Median Sales	1529
Minimum Sales	1
Maximum Sales	4027
Standard Deviation	342
Total Days	1913

Table 1: Summary statistics of daily total sales at TX3.

2.2.5 Differences in products

Figure 5 illustrates the ACF for nine randomly selected products. These plots reveal varying seasonality across products, different decay rates across products and different scales of the ACF per product.

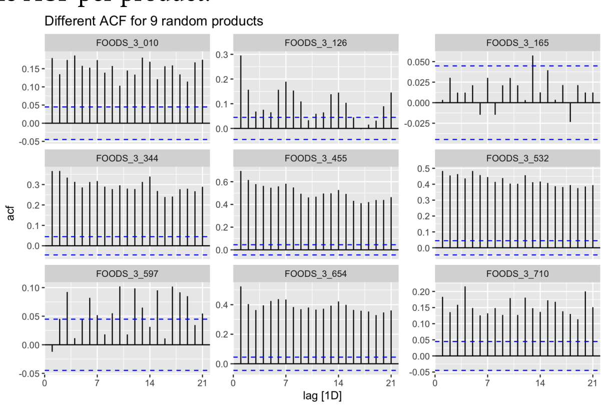

Figure 5: Multiple Autocorrelation functions of 9 random products.

3 METHODS

3.1 Overview of Modelling Strategy

Forecasting daily sales across 823 distinct products requires balancing predictive accuracy, computational feasibility, and interpretability. While more complex time-series models may offer theoretical advantages, their scalability and stability become critical constraints at this scale.

Accordingly, a two-phase modelling strategy was used. Firstly, a pilot study on a randomly selected subset of products compared several candidate forecasting approaches to assess whether increasing the model complexity also led to more meaningful accuracy gains. Secondly, based on the outcomes of this exploratory analysis and practical considerations, a single scalable forecasting framework was selected and applied uniformly across all products, with item-specific parameter estimates.

This separation between exploratory model selection and final implementation ensures that the chosen methodology is both empirically motivated and computationally realisable.

3.2 Model exploration

3.2.1 Baseline Model: Seasonal Naive

To establish a performance benchmark, a Seasonal Naive (SNaive) model was implemented, because the data shows a seasonal pattern in Figure 9. This non-parametric approach assumes that sales on day t will be identical to the sales observed on the same day of the previous week ($t - 7$).

$$\hat{y}_{T+h} = y_{T+h-7} \tag{1}$$

While simple SNaive is an essential baseline in retail forecasting because it effectively captures the 7-day weekly cycle (e.g., Saturday peaks). However, its primary limitation is its inability to account for other factors; it assumes that demand is only a function of the calendar day while ignoring the impact of price changes, SNAP benefit distribution, or special events.

3.2.2 Time Series Linear Model

The Time Series Linear Model (TSLM) is included for three practical reasons. First, it allows the explicit inclusion of external factors that influence sales. By estimating coefficients for price, SNAP benefit days, and event indicators, the model provides an interpretable representation of how these factors are associated with changes in demand, such as the expected change in sales following a price decrease. Second, TSLM conditions forecasts on observed external covariates instead of extrapolating from past sales trends. As noted in the introduction, retail demand is strongly influenced by calendar- and policy-driven factors [1], enabling the model to forecast demand increases associated with future events even in the absence of recent sales momentum. Finally, because TSLM uses standard mathematical formulas it is fast, making it possible to build and train 823 separate models in a short time.

The model used is estimated via Ordinary Least Squares (OLS) with the following specification:

$$\begin{aligned} \log(y_t + 1) = \beta_0 + \beta_1(\text{Weekday}_t) + \beta_2(\text{Price}_t) + \\ \beta_3(\text{SNAP}_t) + \beta_4(\text{Holiday}_t) + \epsilon_t \end{aligned} \tag{2}$$

- **Target Variable ($\log(y_t + 1)$):** A log-plus-one transformation was applied to the raw sales data. This ensures that back-transformed forecasts are strictly non-negative and models the impact of regressors as multiplicative percentage changes (elasticity) rather than additive units.
- **Seasonality (Weekday_t):** A set of dummy variables representing the 7-day weekly cycle was used to capture deterministic recurring patterns.

- **Economic Drivers:** Price_t captures the weekly sell price, while SNAP_t is a binary indicator for benefit distribution days, accounting for the fact that many shoppers wait for benefit days to buy groceries..
- **Calendar Events:** A binary flag indicating major events was derived from the metadata.

Despite the practical advantages regarding interpretability and speed, the reliance on a deterministic linear framework introduces specific limitations. First, the fundamental assumption of Ordinary Least Squares (OLS) is that the error terms are independent, yet retail time series can exhibit "momentum" where high sales persist regardless of external factors. A pure TSLM lacks an autoregressive mechanism to model this. Furthermore, the inherent linearity constraint means the model may fail to capture complex non-linear interactions, unless these are explicitly engineered into the feature set. Finally, the standard TSLM estimates a single set of fixed coefficients for the entire training period, assuming that consumer behaviour remains static. Consequently, the model cannot naturally adapt to structural breaks or evolving trends.

3.2.3 ARIMA

Given the absence of an autoregressive mechanism in TSLM, Autoregressive Integrated Moving Average (ARIMA) models are a natural next step, as they are designed to capture serial dependence and persistence directly from the historical evolution of the series. The model characterizes the data through parameters;autoregressive terms(p), differencing operations(d), and moving-average components(q), enabling it to represent short-term autocorrelation, trends, and when extended to SARIMA: seasonal patterns.

A key characteristic of ARIMA is that it does not incorporate exogenous variables. All variation in the series is explained through its own past observations and error terms. As a result, ARIMA is well suited for settings in which future behaviour is primarily driven by internal dynamics rather than external drivers. Looking at the high autocorrelation of sales in figure 4 and Figure 5, there's evidence that ARIMA could be a valid approach

However, this univariate formulation also has a limitation in applied retail forecasting. Sales volumes can be influenced by known external factors such as calendar effects, price changes, promotions, and policy-driven events (e.g., SNAP) [1]. Since ARIMA has no mechanism to explicitly model such effects, it can limit forecast accuracy during event-driven demand fluctuations and it can reduce interpretability as these external factors are included implicitly.

Another limitation is that different products exhibit different levels of intermittency, seasonality, and persistence, which can be seen in figure 5, requiring distinct ARIMA specifications. As a result, the model parameters (p,d,q) must be selected automatically for each product, typically via information-criterion-based procedures. When applied to a large collection of items, this repeated model identification and estimation process becomes computationally expensive, limiting the scalability of univariate ARIMA approaches in high-dimensional retail settings.

3.2.4 ARIMAX

To overcome the inability of univariate ARIMA models to incorporate external factors while preserving their ability to model serial dependence, the ARIMAX model is considered. By extending ARIMA with the same external covariates used in the proposed TSLM model, ARIMAX combines the strengths of TSLM and ARIMA: modeling of external factors with the temporal dependence captured by ARIMA in a single framework.

$$\begin{aligned} \log(y_t + 1) = \beta_0 + \beta_1(\text{Weekday}_t) + \beta_2(\text{Price}_t) \\ + \beta_3(\text{SNAP}_t) + \beta_4(\text{Holiday}_t) + \eta_t \end{aligned} \tag{3}$$

In this framework, the deterministic component explains variation attributable to known external drivers such as price, calendar effects, and policy-related events, whereas the stochastic component models remaining autocorrelation in sales that arises from internal dynamics, including demand persistence and short-term momentum. By decomposing sales variation into structural and stochastic components, ARIMAX provides a more flexible and theoretically grounded representation of retail demand than either TSLM or univariate ARIMA alone.

Despite its appeal, ARIMAX is also the most computationally demanding of the considered models. Estimating both regression coefficients and ARIMA parameters requires iterative optimization and automatic order selection for each individual product. When applied across a large collection of heterogeneous time series, this substantially increases computation time compared to both TSLM and univariate ARIMA, limiting its practical efficiency at scale.

3.3 Pilot Study for Model Selection

To empirically assess the trade-off between model complexity, predictive performance, and computational feasibility, a pilot study was conducted on a representative subset of products. Rather than evaluating all 823 items, a random sample of 30 distinct products was selected using a fixed random seed (seed = 3) to ensure reproducibility.

The pilot analysis was performed using the fpp3 ecosystem in R, which provide a framework for time-series modeling, forecasting, and evaluation. For each sampled product, identical training and validation splits were used. During the pilot study the final test set is remained untouched.

Four candidate models were fitted to each item: a Seasonal Naive benchmark, TSLM with external regressors as defined in the model exploration, a univariate ARIMA model, and an ARIMAX specification incorporating the same covariates as the TSLM. Model estimation was carried out independently for each product, using automatic order selection for ARIMA and ARIMAX.

Forecasts were generated over the validation window, and predictive accuracy was assessed using Root Mean Squared Error (RMSE). To inform model selection as part of the study design, rather than to report definitive forecasting performance, accuracy was summarized across items by computing the mean RMSE for each model. These aggregated statistics were used solely to assess whether increases in model complexity yielded practically meaningful improvements that justified further investigation.

Table 2 summarises the aggregated RMSE results from this pilot study. ARIMA achieved the lowest mean RMSE (1.76), followed closely by the TSLM (1.82), while the ARIMAX model scored marginally lower. The Seasonal Naive benchmark performed worst. Importantly, these results are not interpreted as comparative performance outcomes; instead, they serve to motivate modeling choices in the subsequent analysis. Differences between ARIMA and TSLM were small relative to the variability across products, as reflected by similar RMSE standard deviations, indicating that added complexity did not clearly translate into robust gains at this exploratory stage.

Table 2: Pilot study results on a random subset of 30 products. RMSE statistics are reported to support model selection only.

Model	Mean RMSE	SD RMSE
ARIMA	1.76	1.73
TSLM	1.82	1.81
ARIMAX	1.94	2.15
SNaive	2.39	2.46

Item level inspection further revealed practical limitations of stochastic models. While ARIMA produced valid forecasts for all sampled products, ARIMAX frequently failed to converge or returned undefined predictions, resulting in missing RMSE values for several items.

Moreover,the ARIMA-based models' automatic order selection led to substantially higher computational costs than regression-based models. Given the marginal accuracy gains over TSLM, combined with increased estimation time and reduced numerical stability, stochastic models were deemed impractical for large-scale application across all 823 products. To speed up the ARIMA models, predefined orders could be set, but differences in individual products as can be seen in the standard deviation make this unfeasible. Consequently, the TSLM was selected as the final forecasting framework, offering a balance between predictive accuracy, interpretability, and computational scalability.

3.4 Final Forecasting Procedure

Following model selection, the Time Series Linear Model was applied to the full dataset comprising 823 distinct products. All data connections provided in the assignment specification were used without modification; an overview of the relational structure is shown in Figure 8. To preserve the temporal structure of the data and avoid information leakage, no random splitting was performed.

3.4.1 Data Partitioning

The dataset was partitioned chronologically into training, validation, and evaluation periods:

- **Training Set:** The first 1,913 days of historical data were used to estimate model parameters for each product.
- **Validation Set:** The subsequent 28 days (d_{1914} to d_{1941}) were reserved as a hold-out set for out-of-sample evaluation and model validation prior to final forecasting.
- **Evaluation Set:** After validation, forecasts were generated for the blind test period (d_{1942} to d_{1969}), which was used for final performance assessment.

3.4.2 Estimation, validation and evaluation

For each product, an individual TSLM was estimated using Ordinary Least Squares with identical model specifications but item-specific parameter estimates. The validation set was used to assess performance and to confirm that the selected model generalised adequately across heterogeneous demand patterns. EDA confirms the need for a flexible yet scalable approach. As can be seen in the ACF plots for nine random products (Figure 5), individual items exhibit widely varying seasonal patterns, decay rates and scale levels. To adequately capture these heterogeneous demand patterns, the chosen model must allow item-specific parameter estimates without compromising computational efficiency.

Next, on the validation set, residual diagnostics were inspected visually for a randomly selected product model (product 42). Given the scale of the analysis, it was not feasible to conduct detailed diagnostic checks for all 823 fitted models. To assess the presence of remaining autocorrelation in the model errors, the Ljung–Box test was applied to the residuals of all products, and averaged to summarise overall residual dependence. To illustrate the estimated effects of the predictors, coefficient estimates were taken from the model for this randomly selected product.

Following validation, the final model's performance was measured on the unseen evaluation set with RMSE (Root Mean Squared Error). This metric was chosen over Mean Absolute Error (MAE) because, in retail inventory management, large errors like missing a demand spike entirely, are much more costly than small errors. As noted by Hyndman and Athanasopoulos [4], RMSE penalizes these large deviations more heavily, making it the safer metric for this problem.

4 RESULTS

4.1 Validation Performance

The Seasonal Naive baseline yielded an RMSE of 2.32 on the validation set. The proposed Time Series Linear Model (TSLM) achieved an RMSE of 2.00 on the same validation period. This corresponds to an absolute reduction of 0.32 units in error and a relative performance improvement of 13.8 % compared to the benchmark.

4.2 Residual Diagnostics

The Ljung-Box test for independence was performed on the residuals of all 823 items to check for remaining patterns. The tests yielded an average p-value of 0.0347, with the vast majority of items falling below the significance threshold ($p < 0.05$). Figure 6 provides a visual breakdown of these residuals. The time plot reveals distinct clusters of errors where the model consistently over- or under-predicted for several consecutive days. Furthermore, the ACF plot shows significant spikes at almost each lag.

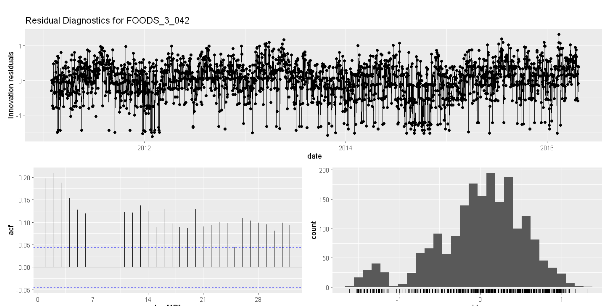

Figure 6: Residual diagnostics for the model of a random food item ('FOODS_3_042').

4.3 Coefficients

Table 3 reports the estimated coefficients of the TSLM for a randomly selected product (product 42). The dependent variable is log-transformed, and coefficient estimates are reported on this scale.

Table 3: Estimated coefficients TSLM for product 42

Predictor	Estimate
Sell price	−1.06
SNAP benefit day	0.04
Event indicator	−0.08

4.4 Test Performance

Upon finalizing the model selection, forecasts were generated for the blind evaluation file, resulting in a final test RMSE of 2.16 for the TSLM and an RMSE of 2.65 for the seasonal naive model. The forecast on randomly selected product 42 created by the the TSLM can be seen in figure 7

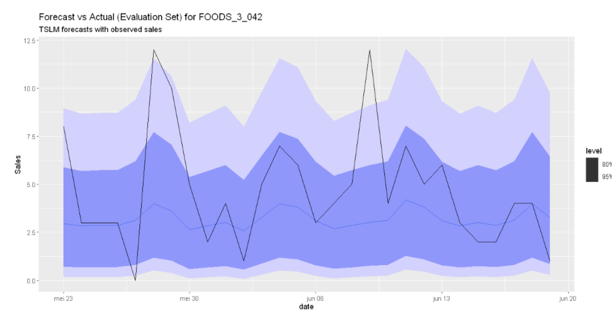

Figure 7: Forecast of 28 days for a random food item ('FOODS_3_042').

5 DISCUSSION

To first address the subquestion on comparative model performance, the classical forecasting models ARIMA, SARIMAX, and TSLM were compared against a seasonal naïve benchmark using a pilot subset of the data, restricted to the training and validation sets. The results indicate that all considered models achieved lower RMSE than the seasonal naïve baseline. However, due to the substantial computational burden associated with fitting item-specific ARIMA-based models, it was not computationally feasible to extend this comparative evaluation to the full dataset. Consequently, these findings should be interpreted as exploratory and informative for model selection rather than as definitive evidence of performance across all products.

Turning to the main research question, the improvement achieved by the best-performing scalable model was assessed using the TSLM. This model was selected due to its computational efficiency and numerical stability. On the validation set, RMSE decreased from 2.32 for the seasonal naïve model to 2 with TSLM, corresponding to a 13.8% reduction in forecast error. On the unseen evaluation set, the TSLM achieved an RMSE of 2.16 compared to 2.65 for the seasonal naïve benchmark, indicating that the observed gains generalize beyond the validation period.

Residual diagnostics revealed that the TSLM residuals were not independent. The average Ljung–Box p-value across items was 0.0347, with most products exhibiting significant autocorrelation. This suggests that the model does not fully capture short-term sales momentum. However, regression-based models without autoregressive errors are expected to leave temporal structure in the residuals when applied to time series data [4]. In the pilot study, ARIMA models achieved lower forecast errors for some products, indicating that autoregressive structure can be informative. Extending the model to an ARIMAX specification, however, did not yield additional gains, potentially due to collinearity between lagged sales terms and external regressors. Addressing this interaction may allow future models to better exploit both autoregressive and exogenous information.

The improvement over the seasonal naïve benchmark indicates that external predictors contribute information beyond seasonal patterns alone. While the seasonal naïve model captures weekly cycles, its higher RMSE suggests that seasonality explains only part of demand variability. This is further illustrated by the general trend plot (Figure 1), which shows substantial daily fluctuations and peaks coinciding with promotional periods and specific events. By incorporating selling price, SNAP benefit timing, and event indicators, the TSLM is able to model these structural demand shifts more effectively.

Several limitations must be acknowledged. First, the deterministic linear structure of the TSLM assumes stable consumer behavior over time. Second, the absence of an autoregressive component limits the model's ability to adapt to short-term momentum or structural breaks. Third, although computationally efficient, residual diagnostics (Figure 6) indicate that temporal information remains unmodeled. Furthermore, coefficient analysis was illustrative rather than systematic across all 823 products, limiting inferential generalization. Finally, the analysis focused exclusively on short-term daily forecasts for the Foods category in store TX3, restricting external validity.

Future research could explore machine learning based forecasting approaches, such as recurrent neural networks (e.g., LSTM models), which are better suited to capturing non-linear relationships and long-range temporal dependencies. Hierarchical modeling frameworks may also help stabilize parameter estimates across products and improve scalability in large retail environments.

CONCLUSIONS

This study examined the extent to which the best-performing classical statistical forecasting model improves short-term retail sales forecasts relative to a seasonal naïve benchmark. Using a scalable Time Series Linear Model (TSLM), consistent reductions in forecast error were observed across both validation and evaluation periods. These results indicate that, within the class of models considered, the selected model provides measurable improvements over a purely seasonal baseline for short-term retail demand forecasting. This demonstrates that external predictors such as price, SNAP benefit timing, and event indicators capture variation beyond seasonal effects alone.

While pilot experiments and residual diagnostics suggested that more complex autoregressive models can yield lower errors for some products, computational constraints limited their applicability at scale. As a result, the final analysis prioritized a model that balances interpretability, robustness, and scalability. Overall, the findings indicate that incorporating external drivers enhances forecasting performance in large-scale retail settings, while also highlighting the need for future work that integrates autoregressive dynamics in a computationally efficient manner.

REFERENCES

[1] Robert Fildes, Shaohui Ma, and Stephan Kolassa. 2022. Retail forecasting: Research and practice. *International Journal of Forecasting* 38, 4 (2022), 1283–1318. https://doi.org/10.1016/j.ijforecast.2019.06.004 Special Issue: M5 competition.

[2] Priyam Ganguly and Isha Mukherjee. 2024. Enhancing Retail Sales Forecasting with Optimized Machine Learning Models. In *2024 4th International Conference on Sustainable Expert Systems (ICSES)*. 884–889. https://doi.org/10.1109/ICSES63445.2024.10762950

[3] Luka Hobor, Mario Brcic, Lidija Polutnik, and Ante Kapetanovic. 2025. Comparative Analysis of Modern Machine Learning Models for Retail Sales Forecasting. arXiv:cs.LG/2506.05941 https://arxiv.org/abs/2506.05941

[4] Rob J Hyndman and George Athanasopoulos. 2021. *Forecasting: principles and practice* (3rd ed.). OTexts. https://otexts.com/fpp3/

[5] Shaohui Ma, Robert Fildes, and Tao Huang. 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 (2016), 245–257. https://doi.org/10.1016/j.ejor.2015.08.029

[6] Patrícia Ramos, Nicolau Santos, and Rui Rebelo. 2015. Performance of State Space and ARIMA Models for Consumer Retail Sales Forecasting. *Robotics and Computer-Integrated Manufacturing* 34 (Aug. 2015), 151–163. https://doi.org/10.1016/j.rcim.2014.12.015

[7] Gylian Verstraete, El-Houssaine Aghezzaf, and Bram Desmet. 2020. A leading macroeconomic indicators' based framework to automatically generate tactical sales forecasts. *Computers Industrial Engineering* 139 (2020), 106169. https://doi.org/10.1016/j.cie.2019.106169

[8] Joselita Pancine Vigna and Emerson Wagner Mainardes. 2019. Sales Promotion and the Purchasing Behavior of Food Consumers. *Revista Brasileira de Marketing* 18, 3 (Dec. 2019), 101–126. https://doi.org/10.5585/remark.v18i3.16368

[9] Xiande Zhao, Jinxing Xie, and R. S. M. Lau. 2001. Improving the Supply Chain Performance: Use of Forecasting Models versus Early Order Commitments. *International Journal of Production Research* 39, 17 (Jan. 2001), 3923–3939. https://doi.org/10.1080/00207540110072236

A APPENDIX

A.1 ER-Diagram Dataset

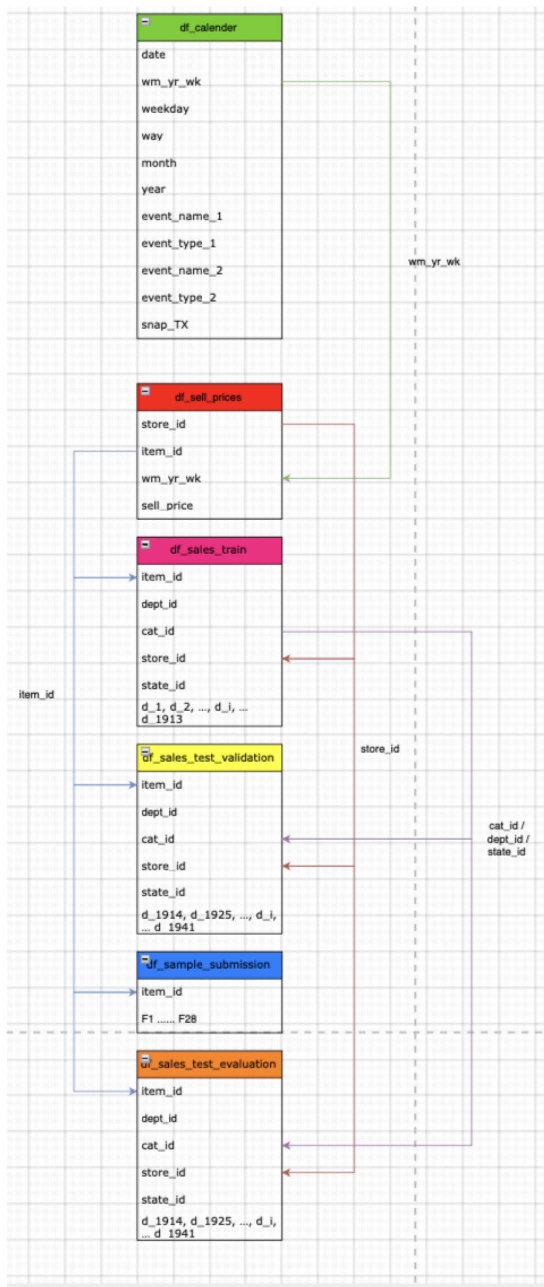

Figure 8: Entity-Relationship diagram of the M5 TX3 dataset subset.

A.2 Exploratory data analysis

A.2.1

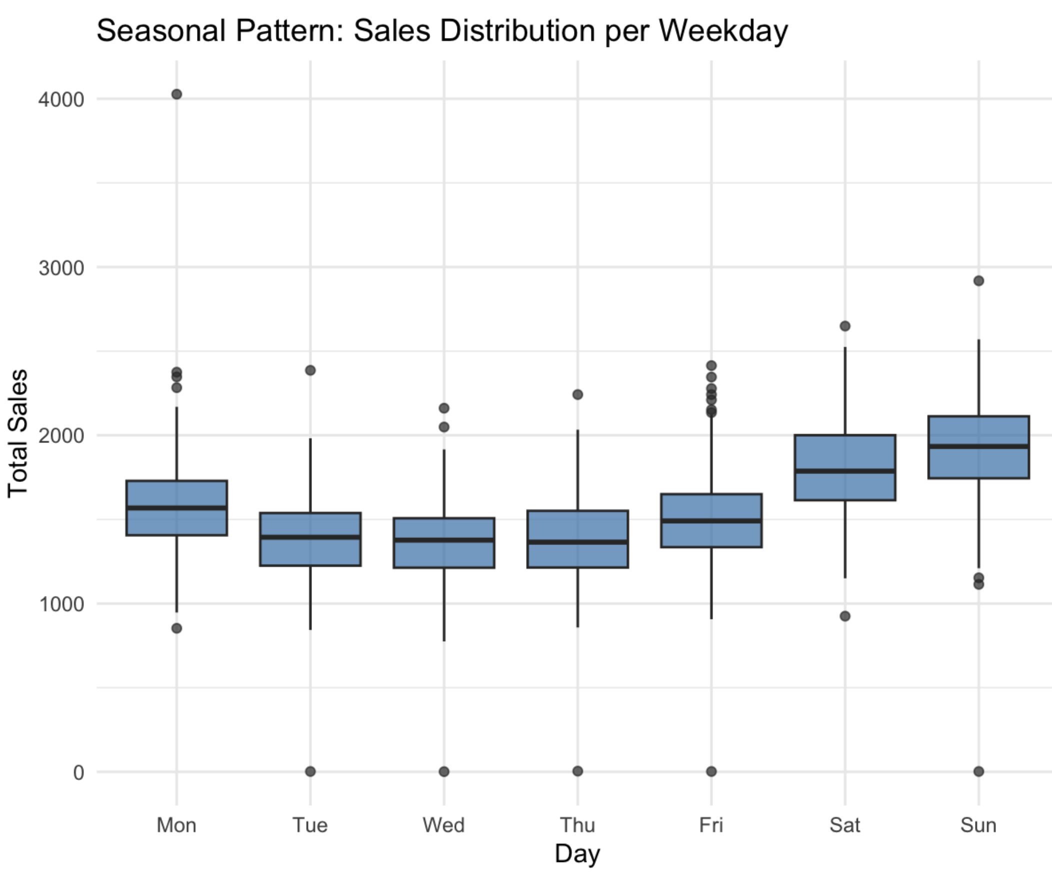

Figure 9: Distribution daily sales per weekday.

A.2.2

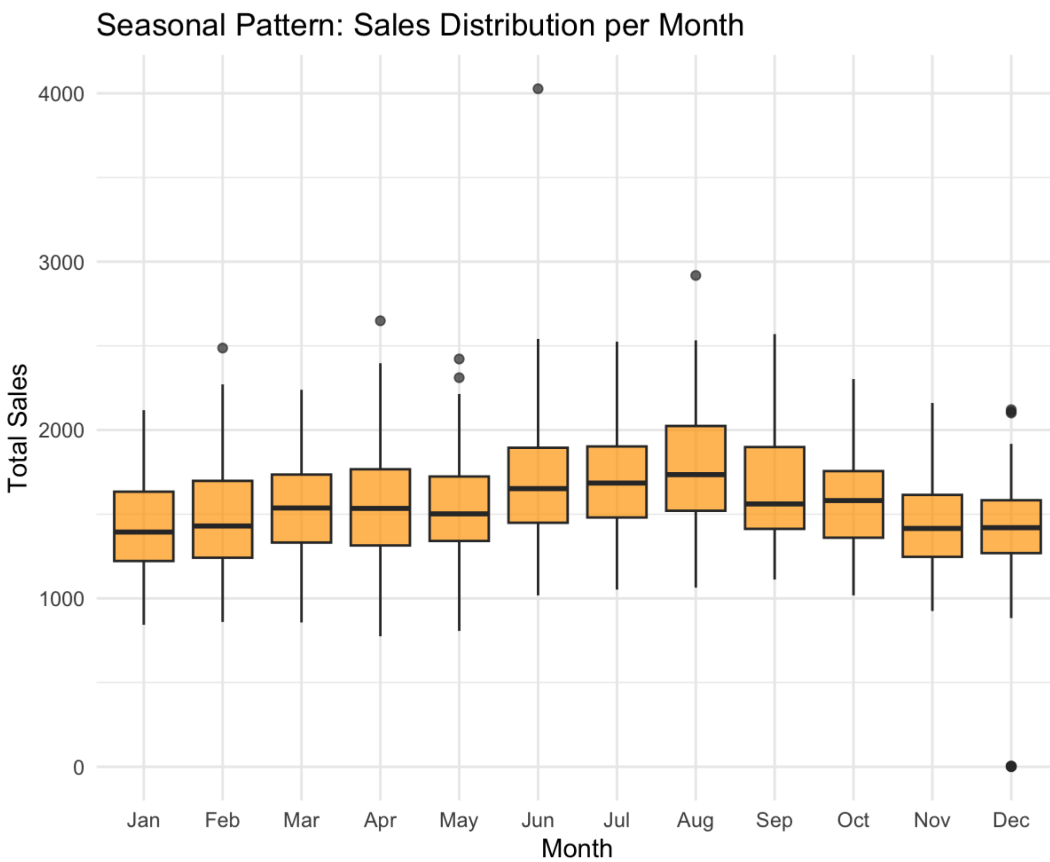

Figure 10: Distribution daily sales per month day.