“School of Social Sciences”

“Master of Business Administration (MBA)”

Undergraduate Thesis / Postgraduate Dissertation

“AI-Powered Inventory Management: A Machine Learning Framework for Strategic Advantage in Distribution”

“Papathanasiou Dimitrios”

Supervisor: “Bournetas Apostolos”

Patras, Greece, “07 2026”

“AI-Powered Inventory Management: A Machine Learning Framework for Strategic Advantage in Distribution”

“Papathanasiou Dimitrios”

Supervising Committee

Supervisor:	Co-Supervisor:
“Bournetas Apostolos”	“Chalikias Ioannis”

Patras, Greece, “07 2026”

Abstract

Small and medium-sized enterprises (SMEs) generate continuous transactional data through their enterprise resource planning (ERP) systems, yet rarely convert this data into a structured artificial intelligence (AI) framework for demand forecasting and inventory decision-making. This dissertation addresses that gap by developing and empirically validating an integrated AI-driven forecasting and inventory decision-support system, applied to four years (2023–2026) of real ERP transactional data from a participating SME, spanning up to 1,592 distinct stock-keeping units (SKUs).

The proposed framework combines, within a single reproducible pipeline: automated ABC–XYZ inventory segmentation; demand-pattern classification based on the Average Demand Interval and squared coefficient of variation (ADI/CV²); a systematic, statistically validated comparison of classical forecasting methods (moving average, exponential smoothing, Holt-Winters, ARIMA, Croston, SBA, TSB) against machine learning models (Random Forest, XGBoost, Linear Regression, and ensemble learning); and a direct comparison of per-product versus pooled machine learning modelling strategies — translating every forecast into concrete safety-stock and reorder-quantity recommendations.

The findings both confirm and complicate what theory predicts. In line with intermittent-demand theory, Croston-family methods had the lowest out-of-sample error for 60.9% of the 373 Lumpy and Intermittent SKUs evaluated with SBA alone responsible for 47.5% (Wilcoxon signed-rank test, $p < .001$), a real advantage but a moderate one, since simpler methods still held their own in about 39% of cases. Two common assumptions didn't survive testing: per-product machine learning models beat a single pooled model trained on the full assortment by 14.4%, even though the pooled model had far more data to train on and an equally-weighted ensemble performed noticeably worse than its best individual member. Tree-based machine learning improved on a naive benchmark by 19%, while the external indicator tested here, aluminium price, showed no measurable relationship with demand across four lag specifications. A retrospective check of the resulting inventory policy reached a 94.1% service level across 1,954 held-out SKU-months with every ABC–XYZ class within 2.5 percentage points of its target.

By combining AI-based forecasting, automated inventory classification, and decision-oriented validation on real SME data, this dissertation offers both a methodologically rigorous and practically deployable contribution to SME inventory management.

Keywords

Artificial Intelligence, Machine Learning, Demand Forecasting, Inventory Optimization, ABC–XYZ Classification, Intermittent Demand

“Ευφυής Διαχείριση Αποθεμάτων: Ένα Πλαίσιο Μηχανικής Μάθησης για Στρατηγικό Πλεονέκτημα στη Διανομή”

“Παπαθανασίου Δημήτριος”

Περίληψη

Οι μικρομεσαίες επιχειρήσεις (ΜμΕ) παράγουν συνεχή δεδομένα συναλλαγών μέσω των συστημάτων διαχείρισης επιχειρησιακών πόρων (ERP) τους, χωρίς ωστόσο να τα αξιοποιούν συστηματικά μέσω ενός δομημένου πλαισίου τεχνητής νοημοσύνης (AI) για την πρόβλεψη ζήτησης και τη λήψη αποφάσεων αποθεμάτων. Η παρούσα διπλωματική εργασία καλύπτει αυτό το κενό, αναπτύσσοντας και αξιολογώντας εμπειρικά ένα ολοκληρωμένο σύστημα πρόβλεψης και υποστήριξης αποφάσεων αποθεμάτων βασισμένο σε τεχνητή νοημοσύνη, εφαρμοσμένο σε πραγματικά δεδομένα συναλλαγών ERP τεσσάρων ετών (2023–2026) από μια συμμετέχουσα ΜμΕ, που καλύπτουν έως και 1.592 διακριτούς κωδικούς προϊόντων (SKUs).

Το προτεινόμενο πλαίσιο είναι καινοτόμο στο ότι συνδυάζει, σε ένα ενιαίο αναπαραγώγιμο σύστημα: αυτοματοποιημένη ταξινόμηση αποθεμάτων ABC–XYZ· ταξινόμηση προτύπων ζήτησης βάσει Μέσου Διαστήματος Ζήτησης και τετραγωνικού συντελεστή μεταβλητότητας (ADI/CV^2)· συστηματική, στατιστικά επικυρωμένη σύγκριση παραδοσιακών μεθόδων πρόβλεψης (κινητός μέσος όρος, εκθετική εξομάλυνση, Holt-Winters, ARIMA, Croston, SBA, TSB) έναντι μοντέλων μηχανικής μάθησης (Random Forest, XGBoost, Γραμμική Παλινδρόμηση και ensemble μάθηση)· και άμεση σύγκριση στρατηγικών μοντελοποίησης ανά προϊόν έναντι ενοποιημένης μοντελοποίησης — μεταφράζοντας κάθε πρόβλεψη σε συγκεκριμένες συστάσεις αποθέματος ασφαλείας και ποσότητας παραγγελίας.

Τα εμπειρικά ευρήματα τόσο επιβεβαιώνουν όσο και εξειδικεύουν καθιερωμένες θεωρητικές προσδοκίες. Σε συμφωνία με τη θεωρία διαλείπουσας ζήτησης, οι μέθοδοι της οικογένειας Croston πέτυχαν το χαμηλότερο σφάλμα εκτός δείγματος στο 60,9% των 373 προϊόντων με σποραδική και διαλείπουσα ζήτηση που εξετάστηκαν, με τη μέθοδο SBA να αντιστοιχεί μόνη της στο 47,5% (στατιστικός έλεγχος Wilcoxon, $p < 0,001$)· το πλεονέκτημα είναι στατιστικά ισχυρό αλλά μέτριο, καθώς οι απλούστερες μέθοδοι παρέμειναν ανταγωνιστικές σε περίπου το 39% των περιπτώσεων. Δύο ευρέως αποδεκτές παραδοχές δεν επιβεβαιώθηκαν: τα μοντέλα μηχανικής μάθησης ανά προϊόν υπερτέρησαν ενός ενιαίου συγκεντρωτικού μοντέλου εκπαιδευμένου σε όλο το χαρτοφυλάκιο κατά 14,4%, παρά το σημαντικά μεγαλύτερο δείγμα εκπαίδευσης του τελευταίου, ενώ ένα σταθμισμένο σύνολο απέδωσε αισθητά χειρότερα από το ισχυρότερο μεμονωμένο μέλος του. Τα μοντέλα μηχανικής μάθησης ξεπέρασαν ουσιαστικά μια απλή πρόβλεψη αναφοράς κατά 19%, ενώ μια εξεταζόμενη εξωτερική μεταβλητή (τιμή αλουμινίου) δεν εμφάνισε μετρήσιμη σχέση με τη ζήτηση σε τέσσερις διαφορετικές χρονικές υστερήσεις. Η αναδρομική επικύρωση της προκύπτουσας πολιτικής αποθέματος πέτυχε επίπεδο εξυπηρέτησης 94,1% σε 1.954 δοκιμαστικές περιόδους, με κάθε κατηγορία ABC–XYZ να πετυχαίνει τον στόχο της εντός 2,5 ποσοστιαίων μονάδων.

Συνδυάζοντας πρόβλεψη βασισμένη σε τεχνητή νοημοσύνη, αυτοματοποιημένη ταξινόμηση αποθεμάτων και επικύρωση προσανατολισμένη στη λήψη αποφάσεων σε πραγματικά δεδομένα ΜμΕ, η παρούσα διπλωματική προσφέρει μια συνεισφορά τόσο μεθοδολογικά αυστηρή όσο και πρακτικά εφαρμόσιμη στη διαχείριση αποθεμάτων ΜμΕ.

Λέξεις – Κλειδιά

Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Πρόβλεψη Ζήτησης, Βελτιστοποίηση Αποθεμάτων, Ταξινόμηση ABC–XYZ, Διαλείπουσα Ζήτηση.

Table of Contents

List of Figures

List of Tables

List of Abbreviations & Acronyms

ABC	Inventory value classification (A = high-value, B = medium-value, C = low-value items)
ADI	Average Demand Interval
AI	Artificial Intelligence
ARIMA	Autoregressive Integrated Moving Average
CNN	Convolutional Neural Network
CV / CV^2	(Squared) Coefficient of Variation
DL	Deep Learning
ERP	Enterprise Resource Planning
ES	Exponential Smoothing
ETS	Error, Trend, Seasonality (state-space exponential smoothing family)
GRU	Gated Recurrent Unit
KPI	Key Performance Indicator
LR	Linear Regression
LSTM	Long Short-Term Memory
MA	Moving Average
MAE	Mean Absolute Error
ML	Machine Learning
RF	Random Forest
ROP	Reorder Point
SBA	Syntetos–Boylan Approximation
SES	Simple Exponential Smoothing
SKU	Stock-Keeping Unit
SME	Small and Medium-sized Enterprise

TCN	Temporal Convolutional Network
TSB	Teunter–Syntetos–Babai method
XGBoost	Extreme Gradient Boosting
XYZ	Inventory variability classification (X = stable, Y = moderate, Z = erratic demand)

Author's Statement:

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

Chapter 1 — INTRODUCTION AND RESEARCH FRAMEWORK

1.1 Background of the Research

Over the last decade, Artificial Intelligence (AI) has significantly impacted the structure of planning and decision-making processes within the business organizations. The incorporation of intelligent systems within various consumer and business applications has further strengthened the relevance of Industry 4.0, as defined by the European Union and other relevant scholars. Industry 4.0 is characterized by data-oriented and driven systems that have significantly impacted industrial processes (Ivanov, 2022; Douaioui et al., 2024). In this regard, the scope of forecasting and inventory management is witnessing significant changes, with various organizations across different sectors attempting to alter conventional operating models.

At the same time, the reliability of the supply chain has also became more volatile. Therefore, the relevance of reliability in the context of customer relationships also gained more significance. In this context, inventory management emerges as one of the critical aspects of modern business operations (Chopra & Meindl, 2021). In the context of small and medium-sized enterprises (SMEs), which are often more dependent on supply chain operations, the relevance of inventory management is more pronounced due to the high dependency of these organizations on inventory management and the limited scope of analytics that often requires a more judgment-based decision-making process (Waters, 2021; Koh & Saad, 2006).

Current ERP systems have shown improvements in the accessibility of historical data. This helps small and medium-sized enterprises to go beyond judgmental forecasts and use more systematic approaches. The application of integrated analytics in enterprise resource planning (ERP) systems has the potential to improve forecasting accuracy as well as support near real-time decision making processes (Douaioui et al., 2024). This can be achieved if AI-based models are used for ERP time series. However, there are situations when existing ERP data is not optimally utilized by small and medium-sized enterprises (Haddara & Elragal, 2015; Koh & Saad, 2006). The main reasons for this are a lack of data science expertise and difficulties in integrating ERP data with existing systems (Arend & Lévesque, 2010). This gap clearly shows that there is a need for context-aware forecasting that takes into consideration the limitations of small and medium-sized enterprises.

One of the major contributions to this problem area is the research paper by Sagaert and Kourentzes (2025). The authors have shown improvements to hierarchical models of forecasts by considering leading indicators for aggregated groups and propagating the same to individual product-level demand (SKU). The authors have shown that there are improvements in forecast accuracy and better decision-making for inventory management

when compared to the use of leading indicators and hierarchical models individually. However, it is a data-rich problem that is applicable to large organizations and is statistical in nature rather than ML /DL based. The research paper is focused on a global manufacturer which raises questions regarding the applicability to small and medium-sized enterprise scenarios.

In parallel, research literature has increasingly emphasized the advantages of Machine Learning (ML) and Deep Learning (DL) models over other models (e.g. exponential smoothing methods, ARIMA models) when dealing with strong non-linearity, volatility and the presence of exogenous variables (Douaioui et al., 2024; Petropoulos et al., 2022). For instance, Long Short-Term Memory (LSTM), Convolutional Neural Network (CNN) and other models have been found that have better performance when they are faced with complex and intermittent demand. Moreover, enhancements have been found when models are integrated with other variables in the retail sector (Aldahmani et al.,2024; Petropoulos et al., 2022).

However, for small and medium-sized enterprise distributors dealing with heterogeneous demand, there is an additional layer of complexity that arises from the coexistence side by side of different types of demand. The conventional taxonomy of demand types was proposed by Syntetos et al. (2005). The authors classified demand types into smooth, erratic, intermittent and lumpy types based on the values of the Average Demand Interval (ADI) and the squared coefficient of variation (CV^2). However, it is not only methodologically inappropriate to use a single forecasting technique for a divergent demand type but also results in systematic forecast error (Teunter et al., 2011). This further underlines the significance of classifying and modeling demand types according to their type.

In another related research area, ABC-XYZ method provides an approach to prioritize items based on economic importance (ABC) and predictability (XYZ). Research has proven that connecting ABC-XYZ with forecasting can significantly improve replenishment decisions, safety stock optimization and other related inventory outcomes, particularly in scenarios with large SKU assortment (Venkatesh & Sowmiya, 2024; Stojanović & Regodić, 2017).

Another issue is the fact that ERP systems register transactions but not their context. The demand is influenced in the companies through undocumented processes, project-driven spikes in demand and also local seasonality, which are not reflected in the data, so this study points to this limitation.

In conclusion, the literature highlights an important gap: while advanced AI-based forecasting techniques hold considerable promise, their practical application in SMEs remains limited. Although the application of state-of-the-art work, such as leading indicator-based hierarchical forecasting, provides an important theoretical platform, it remains in distance from SMEs. However, the development of machine learning, deep learning

methods and improvements in the quality of ERP data have now provided an important window of opportunity for the development of an integrated, practical and SME-focused forecasting framework. This research precisely targets this gap: advanced, ERP-based AI forecasting that meets the practical constraints of SME distribution

1.2 Research Problem

Nevertheless, it is noted that despite advances in forecasting methodologies and the increasing of AI systems in various supply chains, SMEs face challenges in attaining reliable demand forecasting or efficient inventory results. In other words, although ERP data availability is a given, distribution companies usually apply judgmental or simple statistical approaches due to lack of expertise or economic and support tool limitations (Koh & Saad, 2006; Arend & Lévesque, 2010). Consequently, it is noted that demand forecasting is not able to take into consideration the inherent variety of product portfolios offered by SMEs, which may vary from smooth or seasonal to irregular demand. In other words, it is noted that significant forecasting errors emerge in the form of stockout or excessive inventory levels (Syntetos, Boylan & Croston, 2005; Waters, 2021; Teunter et al., 2011).

The literature has also suggested that ML/DL methodologies can outperform other methodologies in nonlinear, irregular or noisy environments (Douaioui et al., 2024; Petropoulos et al., 2022). However, it is noted that there is a gap between theoretical possibilities and SMEs reality. Moreover, it is noted that most methodologies are based on a supportive infrastructure and pipelines or experts that are not available to SMEs (Arend & Lévesque, 2010). As a conclusion, it is noted that few SMEs have access to reliable and resource-efficient AI frameworks to support their available ERP data.

At the same time, Sagaert & Kourentzes (2025) demonstrate that leading indicators and hierarchical methods can increase the degree of accuracy and enhance the quality of inventory decisions in large corporations. However, this approach assumes the availability of a great amount of macro data, hierarchical structures and analytics capabilities. Furthermore, it is based on advanced statistical methods as opposed to ML/DL approaches. Therefore, there is no off-the-shelf, empirically validated methodology that links ERP time series data, AI/ML models, inventory classification (e.g., ABC-XYZ) and inventory decision rules in a cohesive and end-to-end manner for SMEs.

Finally the core problem statement can be stated as:

SMEs need an integrated, practical and analytically AI-based forecasting and inventory management methodology that can accommodate mixed demand patterns, utilize ERP systems' data and take into consideration external or contextual data in order to enhance the quality of inventory decisions beyond traditional or statistical means.

1.3 Justification of the Research (Including Aims)

The research is justified on three topics:

1. **Academic Relevance:** There is a considerable advance in research on forecasting techniques. However, there is a challenge associated with incorporating AI-based techniques into SMEs due to cost and complexity issues (Koh & Saad, 2006; Arend & Lévesque, 2010). Even if a technique is highly relevant, such as leading indicator hierarchical forecasting (Sagaert & Kourentzes, 2025), it is rarely SME-tailored because of data constraints or evaluated against machine learning alternatives. The proposed framework addresses a significant research gap.
2. **Practical Relevance:** SMEs experience mixed demand patterns, which are characterized by inconsistent and intermittent demand among others. In such situations, a general approach to forecasting is not applicable (Syntetos et al., 2005). The proposed framework is computationally tractable, interpretable and applicable to SMEs with the available technology. The objective is to minimize stockouts, overstocking and service performance under stringent cost constraints (Chopra & Meindl, 2021; Waters, 2021).
3. **Managerial Relevance:** Tacit knowledge based on short, focused interviews may further improve the framework to account for demand drivers that are not captured by ERP data, such as project-based demand or local seasonal demand. This is a potential area for future enhancement of the framework.

The research aims to develop an AI-enabled framework for forecasting and inventory management for SMEs. The framework is to be validated using a case study. The research aims to:

- Exploit ERP data to automatically develop forecasting models.
- Apply ABC-XYZ in order to classify demand and allocate models.
- Evaluate the impact on inventory levels beyond forecasting accuracy.
- Compare machine learning models against classical statistical benchmarks, selecting model families according to the data volume available per product.
- Test whether an external market indicator improves forecast accuracy in this setting.

1.4 Outline of the Dissertation

- **Chapter 1 – Introduction & Research Framework**
  Context, research problem, justification and aims, scope and methodology.
- **Chapter 2 – Literature Review**
  Forecasting theory, AI/ML/DL techniques, inventory principles, ABC–XYZ and demand taxonomy, ERP data and external variables.

- **Chapter 3 – Methodology**
  Research design, ERP extraction and preprocessing, ABC–XYZ segmentation, model development, integration of external variables and evaluation metrics.
- **Chapter 4 – Results & Discussion**
  Empirical results, pattern-specific performance, inventory simulations and validation.
- **Chapter 5 – Conclusions & Recommendations**
  Key findings, managerial implications for SMEs, limitations and future research

1.5 Delimitation of Scope

To keep the research focused and feasible:

1. **Organization:** Focus on a single SME distributor in detail.
2. **Data:** The primary data source is the ERP SKU-level data. Additional data points are added selectively when relevant.
3. **Forecasting:** Examination of classical and ML methods. DL architectures were assessed at the design stage but excluded on data-volume grounds (Section 3.5.3) Reinforcement learning and large cloud stacks are out of scope.
4. **Inventory:** Examination of safety stock, reorder points, stockouts and service levels. Broader supply-chain elements are excluded.
5. **Qualitative:** A small number of interviews are considered during research design but not included in the analysis.
6. **Industry:** Focus on the applicability of distribution-oriented SMEs with heterogeneous portfolios.

1.6 Methodology

The study follows quantitative research structured around the following::

1. **ERP data extraction and screening**: Develop SKU-level time series data, handle missing data and outliers and apply proper aggregation and feature engineering techniques (Koh & Saad, 2006; Haddara & Elragal, 2015).
2. **ABC-XYZ classification:** Classify items based on economic value and demand variability (Venkatesh & Sowmiya, 2024).
3. **Pattern identification:** Apply ADI & CV^2 to guide method selection (Syntetos, Boylan & Croston, 2005).
4. **Model development:**
   - Classic methods: Tools for intermittent demand
   - Machine Learning methods: Tools for non-linear relationships
   - DL methods: Tools for complex and long-dependency series (Douaioui et al., 2024; Petropoulos et al., 2022).
   - Hybrid models if deemed necessary.

5. **External variables:** Consider including seasonality, external commodity indicators or calendar effects if deemed necessary, following the leading indicator theory (Sagaert & Kourentzes, 2025).
6. **Inventory simulation and evaluation:** Evaluate models on inventory-related KPIs

1.7 Summary

Chapter 1 describes the difficulties of a SME with forecasting and inventory in an unstable market. The research problem was clearly stated and the research's academic, practical and managerial significance was justified. The methodology is data-based and quantitatively oriented. The research aims to provide a solution to the problem through the application of ERP-based AI forecasting, ABC-XYZ analysis and the application of experts' qualitative knowledge to provide a solution to the problem in the context of SMEs where it matters the most.

Chapter 2 – Literature Review

2.1 Traditional Forecasting Methods

The traditional forecasting methods provide the intellectual foundation of time series analysis and they are the most important benchmarks in the context of research studies as well as applied fields. Despite the fact that the Machine Learning and Deep Learning methods have been successful in various situations, the traditional statistical methods are relevant especially in small and medium-sized enterprises due to their ease of comprehension and low computational requirements. Recent literature has reinforced the importance of these methods, especially in the context of stable or moderately fluctuating demand conditions. Recent studies indicate that the traditional methods provide an important methodological foundation, even though they are increasingly being replaced in the context of highly nonlinear and externally driven situations (Hyndman & Athanasopoulos, 2021; Makridakis et al., 2018; Petropoulos et al., 2022).

2.1.1 Exponential Smoothing Families (SES, Holt, Holt-Winters / ETS)

In general, the single exponential smoothing still remains one of the first choices for forecasting due to its ease of understanding and surprisingly good performance in capturing level, trend and sometimes seasonality. There are studies that describe Single Exponential Smoothing (SES) as being particularly good for data with no trend and seasonality (Hyndman et al., 2008; Hyndman & Athanasopoulos, 2021), while others like Holt's method, extend to add a trend component or Holt-Winters (also known as ETS) extends to incorporate seasonality (Hyndman et al., 2008). Of course, there are known issues with these types of model. For instance, these models assume that demand will not fluctuate significantly and are not designed to handle sudden changes or interactions. For example, if the actual demand behavior is volatile, intermittent and influenced by environmental factors, these models will clearly follow ML/DL models that incorporate many variables. In other words, exponential smoothing models are efficient when used, however, their performance rapidly deteriorates when complex dependencies exist in the time series data (Makridakis et al., 2018; Petropoulos et al., 2022).

2.1.2 ARIMA and Auto-ARIMA

The Autoregressive Integrated Moving Average (ARIMA) models have emerged as the most prominent toolkit for statistical forecasting (Box et al., 2015; Hyndman & Athanasopoulos, 2021). ARIMA models have been successful in dealing with autocorrelation, finding patterns in historical data and providing accurate forecasts for short-term predictions. Recent comprehensive reviews praise ARIMA as an essential part of statistical time series analysis, particularly for situations where trends and seasonality can be separated and become stationary using differencing (Hyndman & Athanasopoulos, 2021). However, the limitations of ARIMA have also been existed as:

- assumes linearity in the relationship between past and future values,
- requires extensive diagnostic effort,
- struggles with noisy, volatile and/or structurally changing data.
  (Makridakis et al., 2018; Petropoulos et al., 2022)

Recent forecast reviews suggest that the ARIMA model may perform well in structured and relatively stable environments but may not perform well in dealing with dynamic, nonlinear and rapidly changing demand conditions (Petropoulos et al., 2022; Makridakis et al., 2018).

2.1.3 Intermittent-Demand Forecasting (Croston, SBA, TSB)

In smooth or seasonal patterns, traditional ETS and ARIMA models tend to perform satisfactorily. The problem starts when they are faced with intermittent demand patterns, like long sequences of zeros followed by unexpected spikes in demand. In situations like these, these models tend to struggle (Syntetos & Boylan, 2005). This is a regular challenge for SMEs that dealing with large numbers of SKUs, spares or project-based items.

In 2005, a study reminded us that traditional models consistently underperform when they deal to intermittent demand forecasting (Syntetos et al., 2005). Syntetos and his team introduced two useful diagnostic tools to identify when traditional models are not performing as required: Average Demand Interval (ADI) and Squared Coefficient of Variation (CV^2).

More recent reviews simply reinforce that study: Croston's method and its advanced versions such as SBA and TSB, outperform ARIMA/ETS models on intermittent demand because they take into account both demand size and timing (Syntetos & Boylan, 2005; Teunter et al., 2011; Boylan & Syntetos, 2010).

In reality, the advantages of using intermittent demand models include:

- forecasting accuracy for sparse time series,
- safety-stock estimation,
- replenishment planning for long-tail items.

Intermittent demand forecasting is connected closely to inventory management. High demand variability and infrequency influence the safety stock levels and the proper inventory strategy (Gonçalves et al., 2020; Silver et al., 2016).

2.1.4 Strengths and Weaknesses

Traditional methods provide interpretability, efficiency and robustness, especially when dealing with stable time series.

However, these methods may not perform well when the demand is driven by nonlinear effects, external factors or high variability.

The following table summarizes the main characteristics of the classical forecasting methods, according to the insights of recent studies and the intermittent demand research field.

Table 2.1— Summary of Classical Forecasting Methods

Method	Strengths	Limitations	Key Literature
SES / Holt / ETS	Simple, transparent, strong for smooth/seasonal demand	Weak on nonlinearities, shocks, external drivers	Hyndman et al. (2008); Hyndman & Athanasopoulos (2021)
ARIMA / Auto-ARIMA	Statistically grounded, strong for autocorrelation	Requires stationarity, struggles with volatile patterns	Box et al. (2015); Hyndman & Athanasopoulos (2021)
Croston / SBA / TSB	Specifically designed intermittent demand	Not suitable for smooth or seasonal data	Syntetos & Boylan (2005); Teunter et al. (2011)

Furthermore, the figure below shows how traditional methods of forecasting are conceptually grouped based on time series characteristics, smooth, seasonal, stationary and intermittent.

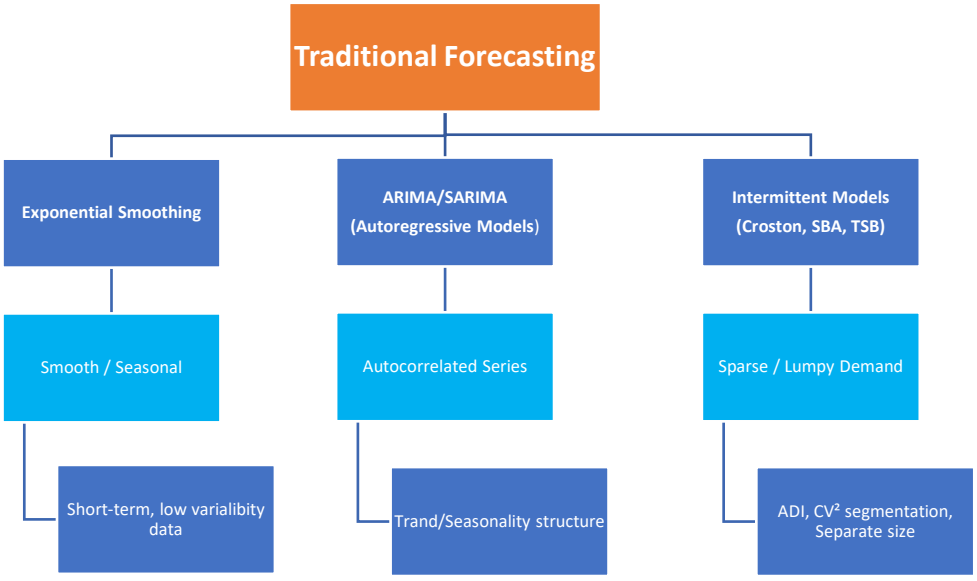

Figure 2.1 Conceptual Taxonomy of Traditional Forecasting Methods (Hyndman & Athanasopoulos, 2021; Syntetos & Boylan, 2005)

As a conclusion, the traditional methods of forecasting play an important role in demand planning, especially for small and medium-sized businesses. However, literature has identified some essential elements:

- ETS and ARIMA methods do not work well in more complicated environments, especially when things get nonlinear or externally driven.

- Modeling intermittent demand is critical for dealing with sparse and diverse product portfolios.

- Classical methods of forecasting are often used as benchmarks for measuring the value added by machine learning and deep learning methods.

In summary, traditional methods of forecasting are still important and far from being outdated. They form part of the demand planning framework, especially when combined with segmentation strategies such as ABC-XYZ and ADI/CV² (Hyndman & Athanasopoulos, 2021; Makridakis et al., 2018).

2.2 Machine Learning Forecasting

During the years, machine learning (ML) has emerged as an essential technology in demand forecasting. Nowadays, markets are becoming increasingly volatile and unpredictable including a growing variety of products. In contrast to traditional methods, Machine Learning models learn patterns based on data and thus are able to identify non-linear relationships and interactions. In recent years, several academic reviews have recognized the importance of ML in demand forecasting and have recognized it as an essential part of modern predictive analytics in supply chain management. Recent research demonstrates the increasing popularity of random forest, gradient boosting and other tree-based algorithms for forecasting purposes in the field of demand management due to their possibilities to determine nonlinear relations between the variables (Douaioui et al., 2024). Furthermore, the tree-based algorithms such as Random Forest and XGBoost have frequently demonstrated strong forecasting performance in forecast accuracy relatively to other AI approaches (Douaioui et al., 2024).

Industry-focused analysis also reinforces this trend. The Oracle overview citing by McKinsey on AI-driven demand forecasting points out that one of the benefits is that machine learning approaches that combine internal and external signals have been reported to cut forecasting errors substantially in data-rich retail settings. Whether that holds in SME distribution, where external signals are harder to come by, is one of the questions this study puts to the test.

In other words, ML approaches may offer significant advantages over traditional approaches that characterized by nonlinearity and multiple explanatory variables. They are scalable with increasing data, handle mixed input signals such as calendar effects, weather, promotions

and macroeconomic factors and can uncover complex relationships that may not be apparent with other approaches. However, there are also challenges with ML approaches that have been reported in the literature. They require significant data preprocessing and model validation, especially with time series data in order to avoid overfitting (Hyndman & Athanasopoulos, 2021; Makridakis et al., 2018).

Table 2.2— Machine Learning Methods in Demand Forecasting

ML Method	Strengths	Suitable Demand Contexts	Key Evidence
Random Forest	Robust to noise, handles nonlinearities	Erratic or intermittent demand	Douaioui et al. (2024); Makridakis et al. (2018)
XGBoost / LightGBM	High accuracy, efficient	Retail, manufacturing and supply chain multi-signal forecasting	Douaioui et al. (2024); Makridakis et al. (2018)

2.3 Deep Learning Forecasting

Deep Learning (DL) models are at the forefront of demand forecasting. They have demonstrated strong performance in complex forecasting environments because they automatically identify and extract features that are temporal and contextual and thus there is a little need for feature engineering as ML require.

According to a comprehensive study in this area, Douaioui and his team found that DL models, such as the LSTM, GRU and CNN/TCN, have strong forecasting performance compared to traditional approaches in cases where there are long-range trends in demand or where demand patterns are changing quickly. They attribute the success of DL models to their ability to quickly adapt in response to changing market conditions and their capacity to handle non-linear relationships between internal and external factors (Douaioui et al., 2024).

Recent literature pays attention to the emergence of modern forecasting architectures based on transformers and attention mechanisms. Nevertheless, their proper applications involve having large datasets, tremendous computing capabilities and long-term historical data, which limits their applicability in small and medium enterprise (SME) forecasting (Douaioui et al., 2024; Aldahmani et al., 2024). In recent years, there have been many developments in the field of predictive analytics in the retail industry. DL models have been found to be

beneficial in cases where there are multiple variables in demand forecasting. These models are also useful in cases where there are external factors such as holidays, promotions and weather that are included in the model and in cases where there is both a need for short-term and long-term scenario planning. However, there are certain issues that must be addressed in DL models. These issues include the need for large amounts of data, the complexity of the model, which may require significant computational resources that may exceed typical SMEs and they require robustness checks such as early stopping and dropout. However, one of the most important issues is the lack of interpretability of DL models. In such cases, tree-based ML models are often used in conjunction with DL models (Douaioui et al., 2024; Makridakis et al., 2018).

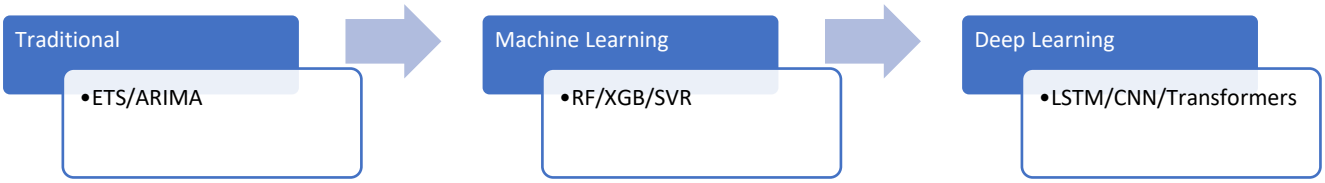

Figure 2.2 — Evolution of Forecasting Approaches (adapted from Hyndman & Athanasopoulos, 2021; Douaioui et al., 2024)

Table 2.3— Deep Learning Models for Demand Forecasting

DL Model	Strengths	Limitations	Literature
LSTM	Long-term dependencies; strong for noisy data	Requires large datasets & compute	Douaioui et al. (2024)
GRU	Faster training; fewer parameters	Less expressive than LSTM	Douaioui et al. (2024)
CNN / TCN	Great for local patterns	Weak on long dependencies	Douaioui et al. (2024)
Transformers	High performance for multi-signal data; parallel processing	High compute; emerging for forecasting	Douaioui et al. (2024); Aldahmani et al. (2024)

2.4 ABC–XYZ Classification and Demand Segmentation

In real-life operations, forecasting depends on more than just the accuracy of the model. The shape of the portfolio also plays an essential role. Therefore, the ABC–XYZ classification system is one of the widely accepted dual-dimensional techniques in inventory management and forecasting (Lagoda & Klumpp, 2024).

The ABC system focuses on the economic importance of the items, which is usually calculated by the annual consumption value. On the other hand, the XYZ system focuses on the predictability of the items using statistical variability as a key driver. Combining the ABC and XYZ systems enables the identification of items such as high value with stable demand (AX), low value with unstable demand (CZ), etc. (Venkatesh & Sowmiya, 2024).

Several recent studies confirm the importance of the ABC–XYZ system in modern analytics. According to the journal article by Venkatesh and Sowmiya (2024), ABC–XYZ segmentation combined with ML-based forecasting using the Random Forest algorithm resulted in more accurate predictions at the SKU level, improving inventory allocation and reducing stockouts as well as excess stock levels. Another article by Lagoda and Klumpp published in 2024 confirms the importance of ABC–XYZ classification in forecasting by choosing the right forecasting technique and establishing the link between segmentation and replenishment strategy. According to the article, the link between segmentation and forecasting improves service levels and reduces inventory costs for different SKU groups (Lagoda & Klumpp, 2024).

This is echoed in professional literature as well. The 2024 guide published by EyeOn, for example, states that ABC is insufficient on its own and XYZ must be used in combination with it in order to capture volatility and tailor service level differentiation. Case studies of actual companies have demonstrated the value of using AX/BX/CX groupings for structured policy development and safety stock optimization (Driessen, M., & van den Bremer, W. (2024).

Recently, the Kubasakova and Kubanova (2024) study uses ABC-XYZ intersections for strategy development in warehouses in a highly unstable demand environment, like the ones affected by the pandemic.

In general, the literature suggests that there is a clear consensus on the value of ABC-XYZ. It is not simply a classification tool, but a framework for sophisticated forecasting approaches, in which pattern-aware models are selected and tailored policies are developed.

Table 2.4 — ABC–XYZ Findings from Recent Literature

Source	Contribution	Application context
Venkatesh & Sowmiya (2024)	ABC–XYZ + ML improves accuracy and reduces stockouts	SKU-level inventory optimization (SMEs)

Lagoda & Klumpp (2024)	Links segmentation to forecasting & inventory methods	Integrated forecasting & inventory control
Kubasakova & Kubanova (2024)	Stock planning with ABC×XYZ under volatility	Warehouse operations (Covid period)

The segmentation guide also tells us where each model is best applied:

- Which patterns each model is best matched with (for instance, which ones are best matched with AX – ETS/ARIMA or with AY – XGBoost or with CZ – Croston and SBA, etc.).
- Where ML can make the greatest contribution: primarily where there is high variation.
- Where simpler models are sufficient: stable segments with low variation (AX and BX).

This strategy avoids unnecessary computation and improves accuracy by using each model on the patterns that it is most likely to recognize.

2.5 External Variables and Leading Indicators in Forecasting

External variables and leading indicators have been widely studied as important drivers of demand variability and forecast accuracy, particularly in volatile and uncertain environments (Makridakis et al., 2018; Hyndman & Athanasopoulos, 2021). An analysis published in 2025 by AI in the Chain found that the Purchasing Managers' Index (PMI) data acts as an early indicator of changes in the way the environment around an organization is behaving, giving an early head start before the actual changes in demand happen.

Furthermore, the 2025 ASCM trends report also found that the integration of macroeconomic indicators, geopolitical risk and variability in lead times into the forecasting model helps in becoming more resilient during uncertain times, as well as improving the scenario planning process.

Recent studies have also identified the following as key influencers on supply chains:

- macroeconomic trends such as industrial production and retail
- seasonal patterns and weather events
- fluctuating commodity costs such as plastics, chemicals and metals
- geopolitics
- marketing events

A recent trends report by AGR (2025) found that incorporating the patterns of the seasons and long lead times can improve service levels and reduce the likelihood of mismatch.

Similarly, it has been found that changes in geopolitics between the US and China or between other nations have a direct impact on supply chains. Finally, the analysis by the International Trade Council in 2024 found that scenario-based forecasting with the inclusion of external variables leads to more accurate inventory planning in long-term.

2.5.1 Integration of External Variables in ML/DL Models

Deep learning architecture really shines when they are allowed to utilize external information, i.e. external factors. In a study by Aldahmani, Alzubi and Iyiola (2024), the effectiveness of BiLSTM and NARX architectures was demonstrated when incorporating exogenous variables, including macroeconomic indicators, into the demand forecasting process. These models demonstrated improved performance compared to several traditional approaches in forecasting accuracy (Aldahmani et al., 2024).

This was not the first study that highlighted the effectiveness of tree-based machine learning architectures, i.e. XGBoost, LightGBM, as these architectures were shown to perform well when external information was utilized, especially when the environment was erratic and volatile (Douaioui et al., 2024; Makridakis et al., 2018).

Table 2.5 — External Variables Used in Forecasting

External Variable	Impact on Forecasting	Source
Macroeconomic indicators (PMI, IP)	Leading indicators; early shift detection	Makridakis et al. (2018)
Weather & seasonality	Higher accuracy for FMCG, apparel, construction	Hyndman & Athanasopoulos (2021)
Commodity prices	Reflect upstream cost and demand shocks	Makridakis et al. (2018)
Promotional events	Predictable peaks	Hyndman & Athanasopoulos (2021)
Exogenous DL inputs	Higher accuracy via nonlinear drivers	Aldahmani et al. (2024); Douaioui et al. (2024)

2.6 Inventory Theory: ROP, Safety Stock and Service Levels

Inventory theory is the link between the forecast results and the actual decision-making process in the supply chain. Forecasting is an attempt to predict future demand while inventory theory is an attempt to translate the forecast results into decision related to when to order, what quantities to order and what service level is appropriate. This is especially

important in small and medium-sized enterprises, as the budget constraints and service level requirements have to be carefully balanced.

2.6.1 Reorder Point (ROP)

The Reorder Point is the actual inventory level at which an order needs to be placed and defined as:

ROP = Demand during lead time + Safety Stock

The 2025 AGR report states that the ROP has to account for seasonal demand as well as variability in lead times, as these could cause stockouts if not properly managed (AGR, 2025).

2.6.2 Safety Stock (SS)

Safety Stock is an additional buffer that is built into the inventory system as a protection against uncertainties in demand as well as uncertainties in the supply chain. The inventory management literature states that safety stock is usually calculated based on statistical patterns between forecast accuracy, lead time and targeted service level (Silver et al., 2016; Gonçalves et al., 2020).

$SS = Z * \sigma LT$

where Z is equal to the target service level and σLT denotes the standard deviation of demand during lead time.

This study shows that the commonly used method of keeping 20% extra is not effective, especially when the variability is high, as opposed to the statistically determined Safety Stock. Also shows that the Safety Stock is the key inventory metric, despite the advances in technology, as it has to balance service level with inventory costs, taking into account demand variability and forecast errors. The 2025 Supply Chain Whitepaper by the International Trade Council shows that the Safety Stock targets have to adjust according to the geopolitical situation, as opposed to keeping them static. Finally, a study done by Gonçalves et al. in 2020 shows that the mathematically determined Safety Stock is the most effective method of managing inventory, because dealing effectively with the uncertainties that occur in the real world (Gonçalves et al., 2020; Silver et al., 2016).

2.6.3 Service Levels (SL)

Service level is the probability of never running out of stock during the refilling process. In reality, companies generally target service levels of 90-98%. These are based on the importance of the particular SKU. New research indicates the relationship between SL and SS as follows:

- increasing service level increases safety stock, thus the need for working capital

- decreasing service level increases the chance of stockout and reduces customer satisfaction

In the inventory policy developed for this dissertation, service level targets are differentiated by ABC-XYZ class rather than kept constant across all products. This approach aligns with the ABC-XYZ approach, where AX may warrant service level above 95%, while CZ may warrant service level of 80-85% (Hyndman & Athanasopoulos, 2021; Lagoda & Klumpp, 2024).

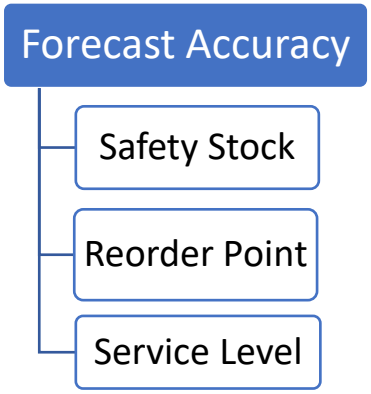

Figure 2.3 — Forecast Accuracy, Inventory and Service Levels (adapted from Hyndman & Athanasopoulos, 2021)

Table 2.6 — Key Inventory Parameters & Literature Insights

Parameter	Role	Insight	Source
Reorder Point (ROP)	Defines order timing	Must adapt to volatility & seasonal shifts	Hyndman & Athanasopoulos (2021); Makridakis et al. (2018)
Safety Stock (SS)	Uncertainty buffer	Should be statistically computed	Gonçalves et al. (2020); Silver et al. (2016)
Service Level (SL)	Stockout probability	Varies by SKU importance	Hyndman & Athanasopoulos (2021); Lagoda & Klumpp (2024)

2.7 Summary

This chapter discusses the various forecasting techniques, classical, machine learning and deep learning, as well as the segmentation of inventory and the inclusion of external variables. The following are the key findings that were consistently noted across the various works discussed in this chapter:

- Classical methods are valid, but they cannot handle nonlinear spikes.
- ML is often effective when dealing with irregular time series data.
- DL may be beneficial when dealing with demand that has many dependencies.
- ABC–XYZ is often valuable when it comes to segmentation, which helps in the optimization of inventory.
- The inclusion of external variables helps with the improvement of the forecast.
- The optimization of inventory, i.e. reorder point, safety stock and service level, is best done by connecting it with the forecast accuracy.

While some progress has been noted in the various fields, the following is still not possible with the current known works: developing a framework, specifically designed for SMEs, that combines the following, with the implementation of the Python language, i.e. data pre-processing using ERP, segmentation of the inventory using the ABC–XYZ method, identification of the demand pattern, classical, machine learning forecasting, evaluating the practical applicability of deep-learning methods, inclusion of external variables, as well as inventory simulation.

Chapter 3 – Methodology

3.1 Research Design and Methodological Approach

The goal of this study is to analyze and evaluate different demand forecasting methods and how affect inventory control using quantitative, experimental and data-driven approaches. The methodology will attempt to fill in the gap between the advanced methods and how can be implemented in small and mid-sized businesses (SMEs), as discussed in Chapter 2.

The research employs an experimental and comparative approach. where multiple forecasting methods, like traditional statistical forecasting models, machine learning (ML) and deep learning (DL)) are implemented through training, validating and testing against actual historical demand data (Hyndman & Athanasopoulos, 2021). Great emphasis is given on practical application and relevance to operational decision-making within the supply chain environment than on advances in methodology alone.

The design of this study is based upon three general objectives:

1. Provide a comparative analysis of the performance of different forecasting demand models under a variety of demand conditions, like stable, seasonal, intermittent and outlier demand patterns.
2. Investigate the influence of segmentation on the forecast accuracy by separating historical demand into ABC–XYZ categories and define the impact of other demand characteristics on the forecast accuracy using ABC–XYZ segmentation method.
3. Evaluate the effects of forecasting accuracy on downstream inventory considerations focusing on reorder point (ROP), safety stock and service levels as opposed to simply forecasting error alone.

Using this multi-objective approach will help to provide consistency with the literature outlined in Chapter 2, which states that forecasting accuracy needs to be examined in conjunction with the operational effects, rather than independently (Makridakis et al., 2018).

3.1.1 Quantitative and Applied Research Orientation

The study is positioned within the quantitative research paradigm utilizing numerical data, algorithmic models and objective performance (Creswell, 2014). The analysis process will be conducted on historical demand time series from operational datasets, therefore the results are empirically grounded and duplicated.

This research is an applied study and therefore, it is different from pure theoretical studies or studies that are based on simulation. This dissertation does not create new algorithms. It rather uses the established forecasting methods systematically and compares each one under realistic conditions where often small and medium enterprises encounter limitations due to limited data availability and demand intermittency, as well as SKU importance

variability (Saunders et al., 2019).This research addresses the research gap identified in Section 2.7 by demonstrating the lack of an integrated end to end framework that combines forecasting, segmentation and inventory decision making and provides a practical implementation context.

3.1.2 Comparative Experimental Framework

The study follows a controlled experimental environment in order to provide methodological rigor and fairness across all evaluated models using the same underlying datasets, preprocessing steps, validation strategy and performance metrics. Thus, this controlled framework provides an opportunity for meaningful comparisons among the forecasting methods and eliminates biases that may otherwise have resulted from inconsistent experimental conditions (Hyndman & Athanasopoulos, 2021).

The comparative framework includes three major model categories:

- Traditional forecasting models, such as exponential smoothing and ARIMA-based approaches, used as interpretable and computationally efficient baselines.
- Machine learning models, including tree-based ensemble methods, chosen for their flexibility and strong performance in nonlinear and irregular demand environments.
- Deep learning models, designed to capture complex temporal dependencies and multi-signal relationships (though, as detailed in Section 3.5.3, excluded from empirical implementation due to data volume constraints).

Through this side-by-side comparison using the same experimental infrastructure, this study will address not only the absolute forecasting accuracy but also the relative forecasting strengths and weaknesses and the suitability of these methods for different demand segments.

3.1.3 Integration of Forecasting and Inventory Decision-Making

The main aspect and the primary advantage of this methodology is that combines the results of forecasting processes with the logic of managing inventory. As shown in Chapter 2, simply achieving accurate forecasts does not create improved operational performance unless they are incorporated into the decision rules (Makridakis et al., 2018).
The design of the research is more than just a means to predict. It includes a simulation of inventory as an additional layer, utilizing the results of the forecasting process directly into the factors calculated to control the inventory. Reorder points, safety stock levels and service level outcomes can all be developed within the different models in such a way that allows the comparison of business-related performance factors such as stockouts, excessive stock risk and overall inventory stability. This inventory assessment is a way to retrospectively validate the company's internal inventory policy by comparing forecasted demand and their safety stock policy against historical, actual demand that was not used for inventory

planning, rather than against synthetic or Monte Carlo generated inventory demand scenarios.

By providing an integrated model, we have addressed one of the major limitations of using only error measures to assess forecasting models, which is the lack of consideration for how forecasting errors will affect the inventory decision policy (Hyndman & Athanasopoulos, 2021).

3.1.4 Time-Series-Aware Methodology

The time-dependent characteristic of demand data results in the application of a methodology that is clearly developed reflecting the time series characteristics and the implications of causality. Training and validating models will occur using only chronological splits of the data so as not to leak future information during the training phase (Hyndman & Koehler, 2006).

Overall, this choice reflects best practices in the forecast literature and avoids overly optimistic performance estimates that occur when random sampling is done inappropriately. The overall design has been developed with the actual deployment conditions as forecast must be generated in a sequential manner using only past information.

3.1.5 Alignment with Research Objectives

The objective of the methodological design is to meet the overall research aim of this dissertation, through:

- Providing an objective evaluation of the methodology through quantitative assessment and standardized experimental conditions.
- Being relevant to others beyond the novelty of the algorithms used by creating experimental conditions based on the operational outcomes.
- Using commonly accepted forecasting methods will ensure that the forecasting methods, in general, used in this project are considered and replicated in similar business environments.
- Emphasizing the computational efficiency of the methods employed will enable to be implemented feasibly with SMEs (small/medium enterprises) through interpretability of the models developed where possible.

Therefore, the study design provides a well-structured and justifiable basis for the following chapter 3 sections, which discuss the data characteristics, data-processing methods, data segmentation strategies, model-building and validation procedures and metrics used to evaluate the model.

3.2 Data Description and Sources

This section describes the datasets used in this study, including structure, scope, time-related characteristics and relevance to the research objectives. The real-world nature of the data and the challenges that are related to operational demand series, including intermittency, volatility and heterogeneity across products are also highlighted (Hyndman & Athanasopoulos, 2021).

3.2.1 Data Sources

The core dataset used in this study is historically based on confirmed orders from the ERP and operational sales systems, which correspond to actual demand, not simulated or synthetic demand (Lapide, 2006). These sets of data reflect actual, as opposed to theoretical, business operations. Therefore, the usage of operational data in a professional context validates the empirical evidence/results of this study and reflects the unique structural irregularities that observed in actual supply chain (Makridakis et al., 2018). These datasets are used to support the focus of the dissertation on meeting the applied aspect of Chapters 2's literature and also demonstrate the need to test forecasting techniques on real-world data.

The time-series demand datasets are introduced at stock keeping unit (SKU) level. Therefore, they can be analyzed both at detailed and aggregate level across multiple products with a variety of demand patterns and economic significance.

3.2.2 Temporal Structure and Granularity

The demand data sets are recorded in discrete time-series with a repetitive temporal occurrence, like daily, weekly or monthly). The technicality of the data collection process is determined by:

- capturing meaningful demand dynamics through seasonality and short-term variability and
- providing sufficient amount of data to adequately train and validate machine learning and deep learning models (Hyndman & Athanasopoulos, 2021).

The observation time period represents multiple years of sufficient historical data in order to identify long-term trends, cyclical seasonal patterns and structural demand fluctuations. The time-series across all products are aligned to one base calendar to enable them to be compared across all SKUs and forecasting models.

3.2.3 Target Variable Definition

The total amount of units requested or delivered per SKU within each time frame, that represents the demand quantity, serves as the primary target variable of this study. Demand is a non-negative quantity and has considerable variability across products.

Observed demand patterns include:

- smooth and stable demand series,
- seasonal and trend-driven demand,
- intermittent demand patterns with many zero demand periods
- highly volatile and erratic demand behavior.

This demand patterns diversity is a key motivation for the segmentation and comparative evaluation strategy adopted in this study, as different forecasting models are expected to perform differently, in terms of forecast accuracy, across a variety of demand structures (Syntetos et al., 2005).

3.2.2 Supporting and Contextual Variables

The dataset has a small number of contextual attributes for preprocessing, segmenting and feature engineering beyond the demand variable.

The contextual attributes include:

- information related to calendars (e.g., time indices, seasonal flags),
- the identification and grouping of products
- basic operational indicators available

In order to reduce the complexity of the baseline model evaluation, the external explanatory variables such as economic or environmental indicators are treated independently (Makridakis et al., 2018).

3.2.3 Data Heterogeneity and Challenges

An important feature of the datasets is that there are differences between the SKUs regarding their demand patterns, such the amount that they are purchased or how erratically they are purchased over time. Some common challenges that observed include:

- gaps in data collection leading to missing observations,
- periods of time when certain SKUs demand is zero because of intermittent demand,
- outlier observations that are due to one-time purchasing occurrence or because of unanticipated events and therefore not representative of typical demand,
- changing patterns in how customers purchase particular SKUs based on changes to their needs, that occur over time.

Rather than viewing these differences as data quality issues that need to be removed from the analysis, the methodology approach in this study rconsiders them as inherent characteristics of the operational demand data and therefore are consistent with the

guidance provided by the forecasting literature (Syntetos et al., 2005; Hyndman & Athanasopoulos, 2021).

3.2.4 Representativeness and Scope

This dataset reflects a common demand scenario of small and medium enterprises (SMEs), that are true to their nature of limited historical depth on certain SKUs, have varied demand patterns and typically constrained resources for analysis. It allows for the conclusions drawn to be generalized beyond just large-scale retail, while also maintaining their relevance to businesses constrained by realistic data and operational conditions (Lapide (2006).

To summarize, the datasets used in this study represent a realistic and challenging environment for testing and evaluating demand forecasting models. The structure, historical depth and variability permit an analysis of how forecasting performance translates into operational inventory outcomes in order a SME to have the maximum benefit.

3.3 Data Preprocessing and Feature Engineering

The data preparation and feature engineering step in a methodology framework are vital to success because the forecast accuracy is heavily impacted by both the quality and the data presentation that used for forecasts. This part of the process is not just to clean the data but involves the raw demand data transforming from day-to-day use into proper and usable, structured and consistent, formats that can be employed in producing forecasts using either traditional statistical forecasting models, machine learning (ML) or deep learning (DL) techniques (Hyndman & Athanasopoulos, 2021).

3.3.1 Data Cleaning and Consistency Checks

After establishing initial datasets cleaning with regards to missing observations, a range of consistency and integrity data checks are applied to ensure that temporal coherence and logical correctness was confirmed. All the time-series are examined for checking duplicate timestamps, issues with calendar alignment and aggregation levels matching checks.
The datasets are checked for duplicate SKUs in time level and decisions are made about removing or aggregating the data based on where the duplicate records originated. The demand values are further validated to ensure their realistic magnitude ranges and non-negative values.
The outliers were treated conservatively. Since extreme demand is often an example of a genuine business event, like one-time customer orders, promotional events or exceptional situations, outliers are kept, unless it is determined to be data entry or system error. This method is followed in order to preserve variability in demand, a necessary step for realistic forecast evaluation (Makridakis et al., 2018).

3.3.2 Feature Scaling and Transformation

Depending on the model requirements, feature transformation is performed selectively based on the model needs. While traditional forecasting models could utilize raw demand

series data without scaling, ML and DL models require the raw data to be scaled before use in order to obtain numerical stability and optimized training. Thus, for deriving features normalization or standardization were performed as appropriate to obtain the correct input for each purpose model. Scaling parameters for each derived feature were calculated solely from the training dataset in order to eliminate any possibility for leakage information into validation or testing datasets (Hyndman & Athanasopoulos, 2021).

3.3.3 Lagged Features and Temporal Windows

Lag-based features are also created for the purpose of capturing temporal dependency in demand for ML and DL models. Specifically, lag-based features allow the model to learn the autoregressive patterns of demand by representing the historical demand value over pre-defined temporal windows.

The lag lengths selection considers both how expressive the model can be and how much data availability exists, particularly with respect to the intermittent demand series. The use of an excessively long lag window is avoided as they reduce effective sample size and therefore increase sparsity of the data leading to degradation of the model performance (Syntetos et al., 2005). Additionally, rolling statistics, such as moving average and rolling standard deviation, are calculated to identify local trends and short-term volatility. This method also captures additional information not available in the raw lag values and help distinguish stable from unstable demand behavior.

3.3.4 Demand Pattern Metrics

To assist with demand segmentation and model assignment, two key demand pattern measures are calculated:

- Average Demand Interval (ADI), the average of the number of time periods between occurrences of non-zero demand.
- Squared Coefficient of Variation (CV^2), a measure of relative demand variability.

These two measures, provide the basis for classifying demand patterns as smooth, intermittent, erratic or lumpy, as established through previous demand classification frameworks (Syntetos et al 2005) and continuing by supporting ABC/XYZ classification approach as described in a later section.

3.3.5 Calendar-Based Features

To establish a basis for identifying systematic temporal effects such as seasonality and long-term trends, both calendar and time index variables are used, consistent with evidence that additive calendar-effect terms can meaningfully improve forecast accuracy in business demand contexts (Hyndman & Athanasopoulos, 2021). These variables allow the machine learning (ML) and deep learning (DL) models to learn recurring patterns in demand without adding any explicit parametric assumptions into the model.

Calendar variables are used consistently across models to ensure fair comparisons, while avoiding complex feature variables introduction that could introduce noise or cause overfitting (Makridakis et al., 2018).

3.3.6 Preparation for Model-Specific Requirements

Considering the diverse inputs of various forecast models, the last preprocessing step assures compatibility among multiple model types:

- Traditional models accept single variable demand serries
- ML models require structured feature matrices.
- DL models received windowed sequences of historical observations

Each of the experimental datasets are produced through a single preprocessing process, thereby guaranteeing consistent test conditions across different types of forecast methods and allowing legitimate comparison.

In conclusion, the preprocessing and the feature engineering process convert the raw operational demand data into a dataset ready to be used by the forecasting models. The approach preserves that demand variance being maintained, unnecessary smoothing is avoided and all forecast techniques having compatible formats so that demand segmentation and model implementation can occur without difficulty in the sections that follow.

3.4 Demand Segmentation Strategy: ABC–XYZ Classification

Demand segmentation is a key component in the way the research methodology was designed. The study acknowledges that SKUs do not form as homogeneous group. The evidence shows that there is considerable variability in demand patterns exhibited by each product as well as their significance to the business. These differences are significant to the forecasting accuracy and inventory management performance according demand classification literature (Syntetos et al., 2005; Silver et al., 2016).

In order to take into account this variability in a structured and operationally applicable way, this study uses a combined ABC-XYZ segmentation before implementing the forecasting model. As a result, using demand segments provides a multiple dimensional view and acts as a guiding mechanism throughout the forecasting and inventory evaluation process.

3.4.1 ABC Classification: Economic Importance

The first dimension of the segmentation is ABC analysis, which divides stock keeping units (SKUs) into groups according their economically contribution. To define how much each SKU costs annually, the number of units sold each year has been taken and multiplied it by the unit price. Then there was a rank based on how much they contributed annually and finally grouped them according to their cumulative contribution:

- A items representing the 70-80% of total consumption value,
- B items for the next 15-20%,
- C items are low-value products of the remaining contribution.

The classifications above are based on inventory management principles normally utilized within the business. Therefore, decisions on forecasting and inventory are directed towards those items that will provide the greatest financial impact (Silver et al., 2016). Thus, the ABC dimension introduces an explicit business value perspective, a business value-based approach, into the research methodology.

3.4.2 XYZ Classification: Demand Variability

The second segmentation dimension shows how often demand occurs and how much it changes over time. It is not related to how much demand is worth and helps to understand the nature of demand so that forecasts can be prepared, like to determine if stock should be built up or not.

Two demand pattern metrics were agreed to be appropriate:

- Average Demand Interval (ADI) that measures the demand regularity in time by computing the average internal between two demands,

$$ADI = \frac{Total\ number\ of\ periods}{Number\ of\ demand\ buckets}$$

- Squared Coefficient of Variation (CV^2) that measures the variation in demand quantities.

$$CV = \frac{Standard\ deviation\ of\ a\ population}{Average\ value\ of\ a\ population}$$

The following three classes of product stock keeping units (SKUs) were identified based on these two metrics:

- X - SKUs with consistently stable and regular demand characteristics,
- Y - SKUs with moderately varying demand characteristics or seasonal patterns,
- Z - SKUs with highly erratic demand characteristics or intermittent demand.

This classification methodology is based on well-established demand pattern classification frameworks and enables a systematic means of distinguishing between predictable and unpredictable demand structures (Syntetos et al. 2005; Boylan & Syntetos 2010).

3.4.3 Integrated ABC–XYZ Segmentation

The last part of demand segmentation combines the ABC and XYZ dimensions, so all nine classes of demand (AX, AY, AZ, BX, BY, BZ, CX, CY, CZ) can be represented. By creating this combined segmentation structure, demand is represented more completely by combining both the business importance and level of uncertainty into a single representation or structure.

The ABC-XYZ framework is an effective decision-support structure rather than a means of requiring a strict one-to-one correspondence between demand segments and forecasting models. The ABC-XYZ framework assists users in providing insights about how well a particular forecasting model is meeting user expectations, providing information about what segments may be at greater risk for forecasting errors and allowing results from various demand categories to be interpreted in different ways (Silver et al. 2016).

3.4.4 Role of ABC–XYZ in the Forecasting Pipeline

In addition to impacting different forecasting methods as noted within the ABC-XYZ framework, the ABC-XYZ classification is used to influence the following forecasting methods:

- Forecasting performance evaluations, where the accuracy of forecasts is evaluated at the aggregate or overall level as well as evaluating specific performance of segments, identifying strengths and weaknesses of each segment,
- Operational interpretation, where the degree of forecast error is weighted more critically for high value and highly variable SKUs as compared to low value SKUs,
- Inventory policy analysis, where the implications for service level targets and safety stock are consistent with the importance of each SKU and the amount of uncertainty demand (Boylan & Syntetos 2010).

This perspective ensures that the forecasting evaluation remains aligned with the real operational priorities rather than purely statistical criteria.

In conclusion, the ABC-XYZ demand segmentation method offers an organized and analytical tool to assist in understanding demand variability. The ABC-XYZ demand segments have been defined within two dimensions, economic significance and demand variance, thus allowing for a fair basis for judging forecasting comparisons, as well as assisting with the transferability of the results of the research into practice.

In the subsequent section, we will continue to apply the ABC-XYZ approach through a description of how to employ forecasting models from a variety of methodologies.

3.5 Forecasting Models Implementation

The forecasting models implementation consists of three groups or categories: the traditional statistical models, machine learning (ML) models and deep learning (DL) models

as referenced in chapter 2. All selected and implemented models were based on the literature reviewed in chapter 2, as well as on the segmentation methodology described in section 3.4 and shows that there is a similarity between theoretical and empirical applications.
All models were implemented in a single experimental framework using the same processed pipeline for pre-processing, training and validation. Therefore, any differences in the comparative results can be attributed to differences in the model capability rather than differences in implementing them or data managing (Hyndman & Athanasopoulos 2021).

3.5.1 Traditional Forecasting Models

The traditional statistical forecasting tools that were used as benchmark baselines as they are highly interpretable, computationally efficient and have been widely used in practice are approaches that applied to univariate demand series and are based on an explicit parametric assumption regarding trend, seasonality and error structure.
The following traditional approaches were used:

- Naïve forecasting method as the minimal baseline reference.
- Exponential smoothing methods (e.g., Holt and Holt Winters) for capturing the trends in stable and seasonal demand.
- ARIMA and seasonal ARIMA (SARIMA), are statistical methods that attempt to model the autoregressive and moving average dynamics of stationary and seasonal time series respectively. The ARIMA method has been applied as a non-seasonal methodology due to the fact that the majority of the monthly SKUs exhibit no historical records outside of the last several months. A SARIMA methodology was not considered. The lack of sufficient Volume/Percentage of Data from which an estimate can be made renders both models ineffective.
- Croston-type methods for intermittent demand including Syntetos–Boylan Approximation (SBA) and Teunter–Syntetos–Babai (TSB) are applied to intermittent demand series.

These models have a long history in the forecasting literature and consistently provide reliable forecasts in smooth and less structured demand environments (Hyndman & Athanasopoulos, 2021; Syntetos et al., 2005) making them suitable to provide valid transparent benchmarking against more complicated forecasting methods.

3.5.2 Machine Learning Models

Machine learning (ML) was used to overcome traditional forecasting techniques' limitations in nonlinear, non-stationary and/or non-feature-rich demand environments. ML models can recognize and learn complex relationships between lagged demand values and other structured inputs such as rolling statistics and calendar-based features.

The primary ML models implemented include:

- Random Forest regression that generates a set of decision trees from aggregated multiple data for reducing variability and improving robustness.
- Gradient Boosting models (e.g., XGBoost or LightGBM), designed to sequentially learn from residual errors and capture nonlinear feature interactions.
- Linear regression serves as a baseline model for machine learning as it is both simple to implement and fully interpretable.
- An ensemble of random forests, XGBoost, and linear regression model by taking a simple average of the three models to see if combining models gives us a more robust result than using just one model.

All of these models have consistently outperformed traditional forecasting techniques in empirical studies (Douaioui et al., 2024; Makridakis et al., 2018), particularly for intermittent and erratic demand patterns where the underlying parametric modelling assumptions of traditional forecasting models are violated.

ML models rely on a clear definition of input data/features through feature engineering and formally designed input matrices. For fair comparison across all models, a common set of input features, produced from preprocessing steps explained in Section 3.3, was used to train all forecasting models.

3.5.3 Deep Learning: Considered but Excluded from Implementation

During the methodological design stage, deep-learning architectures such as LSTM and CNN-LSTM Hybrid Models have been identified as being beneficial for capturing long-range dependencies as well as the ability to capture non-linear demand patterns (Douaioui et al., 2024; Aldahmani et al., 2024). Nevertheless, deep-learning architectures were excluded from empirical implementations for two reasons, both of which were sourced from the data.

The small number of monthly per item forecasted time series (20–40) are typically too small of a sample to produce reliable forecasts using deep learning than simpler statistical or tree-based methods (Douaioui et al., 2024).

Also, tree based ensemble methods (Random Forest and XGBoost) were tested on various per-SKU time series in advance of empirical studies and the results of this testing were recorded in Chapter 4. These results demonstrated that these relatively data efficient ML methods could not maintain as strong levels of predictive accuracy using only these limited amounts of historical data in their training datasets, suggesting that DL models, which require even greater numbers of training variables than those used with tree based ML methods, would similarly have poor predictive accuracy if developed using these same

limited numbers of historical monthly per-SKU time series as input. This decision is revisited in Chapter 5 Limitations and Future Research sections.

3.5.4 Consistency Across Model Implementations

In order to maintain the proper method, each forecast model produced had to have a strict implementation of control and standards.

- For all models, the same data was used in both pre-processing and modelling.
- For all models, the same length of forecasting time was used for evaluation.
- For all models, the model training and validation was done with respect to time series causality.

This standardized way of implementing allows for a performance comparison without bias and an indication of the true strengths and weaknesses rather than experimental artefact (Hyndman & Athanasopoulos, 2021).

3.5.5 Evaluation Protocol

All forecasting comparisons share a common protocol, so no method is favored by how it's evaluated. Each SKU's monthly series is split chronologically: the first 70% of observations train the model, the last 30% is held out for testing. The split is never random, that would let future observations inform predictions about the past, which makes no sense here.

Every method's parameters are tuned only on the training portion. The size of the moving-average window is chosen from 2, 3, 4, 6 or 12, while the smoothing constant α for exponential smoothing and for each Croston-family variant is selected from a grid of 0.05 to 0.50. Holt-Winters and ARIMA are fit with orders chosen by AIC and accuracy is measured only on the held-out period.

Giving one method family per-SKU tuning while leaving the others at default settings would confound method quality with tuning effort. In-sample evaluation has the same problem, since fitted values already "know" the observations they're predicting. The cost of this rigor is a smaller sample: SKUs with fewer than eight monthly observations can't support a meaningful split and are excluded.

3.6 Summary

In conclusion, this section provides a comprehensive and controlled process to apply forecasting methods in an identical experimental manner across three different paradigms (traditional, ML and DL) and connects the choice of method to the demand segmentation. This results in a consistent framework of evaluating the forecasts based on differing demand environments. The following section provides an overview of the training, validation and evaluation of all models. The training, validation and evaluation of models were conducted using statistical and inventory-based performance metrics.

Chapter 4: Results and Discussion

This chapter describes the features and characteristics observed from the application of the previously described methodology (chapter 3) to 4 years (2023-2026) of ERP Data at the SME participating in the study; following the same analytical sequence: segmentation, classification of the demand pattern, comparison between models and translation into an inventory policy.

4.1 Data Overview

The dataset is composed of SKU's from January 2023 until June 2026, with June 2026 being a partial month (6 months). For each year, total distinct SKU counts were identified: 2023 = 942, 2024 = 1,068, 2025 = 1,080 and 2026 = 755 and there was a total of 1,592 unique SKU's crossing all years. For the purpose of completing the forecasting experiments (Sections 4.4 - 4.8), each SKU needed a minimum of a 12-month history to be eligible for consideration for the forecast. After filtering based upon the 12-month history requirement, the eligible SKU count was reduced to 321 total SKU's; all eligible SKU's fell into all ABC-XYZ classes and all demand patterns. This data structure, a large heterogeneous assortment of SKU's with limited historical support for each SKU, is highly consistent with the SME conditions that were identified in section 3.2.4. The participating SME was selected on a convenience basis, reflecting practical access to a complete, multi-year ERP transactional history, a common constraint in SME-focused operations research (Lapide, 2006).

4.2 ABC–XYZ Segmentation Results

The 2025 ABC classification was very similar to the traditional Pareto principle (Silver et al., 2016), where A-items accounted for 12.87% of SKU's but 79.91% of overall value and C items accounted for 64.81% of SKU's, but only 5.02% of overall value. In contrast to the ABC dimension, the distribution of SKU's by XYZ classification was an opposite skew where over half of the SKU's (52.50%) were classified as unpredictable Z items and only 16.39% were classified as well-established X items. This volatility of demand at the SME-level has been documented elsewhere (Lapide, 2006)

Table 4.1 - ABC–XYZ distribution, 2025 (n = 1,080). Source: own elaboration

Class	Count	Value % (of total)	Avg. CV
AX	5	0.9%	0.40
AY	54	28.9%	0.80
AZ	80	50.1%	1.41

Class	Count	Value % (of total)	Avg. CV
BX/BY/BZ	32/86/123	1.8/5.2/8.1%	0.25/0.79/1.45
CX/CY/CZ	140/196/364	1.0/2.1/2.0%	0.24/0.75/1.36

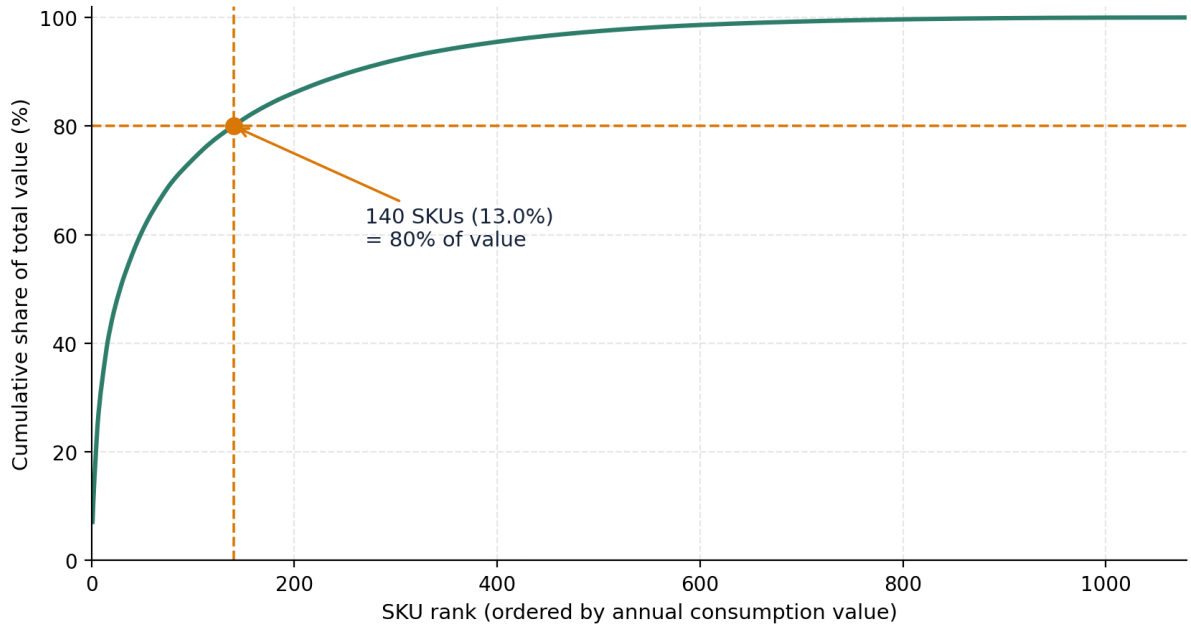

Figure 4.1 - Pareto curve of cumulative consumption value by SKU rank, 2025 (n = 1,080). Source: own elaboration.

An example of the chapter's introductory notable finding (AZ segment) was that only 80 total SKU's accounted for over 50% of total A-item value, and that these SKU's had the highest variability of all A-item SKU segments. Therefore, it follows that high value SKU's and high unpredictability SKU's are co-located in the area where forecasting error is the most expensive (Boylan & Syntetos 2010). Essentially, the application of a standard ABC only inventory system would leave the vast majority (all 139 A-class SKUs) vulnerable to differences in service level definitions and ultimately leads to the company's greatest liability resting with the exact items that provide the most income. The AZ segment will thus require more detailed, frequent monitoring than any ABC classification could warrant, supporting the use of an XYZ dimension in the inventory policy table (see section 3.4.4) instead of the conventional use of ABC classification (Venkatesh & Sowmiya, 2024).

4.3 Demand Pattern Classification (ADI/CV²)

Applying the ADI/CV² classification of Syntetos et al. (2005) identified 563 SKUs (33.3%) as Lumpy, 361 (21.4%) as Intermittent, 164 (9.7%) as Erratic, and 66 (3.9%) as Smooth, out of 1,690 classified series; a further 536 (31.7%) lacked sufficient observations for reliable classification. This total of 1,690 differs from the 1,592 SKUs reported in Section 4.1 because

the demand-pattern classification includes SKUs with recorded quantity movement but zero net sales value (e.g., returns), which the value-based ABC–XYZ segmentation excludes.

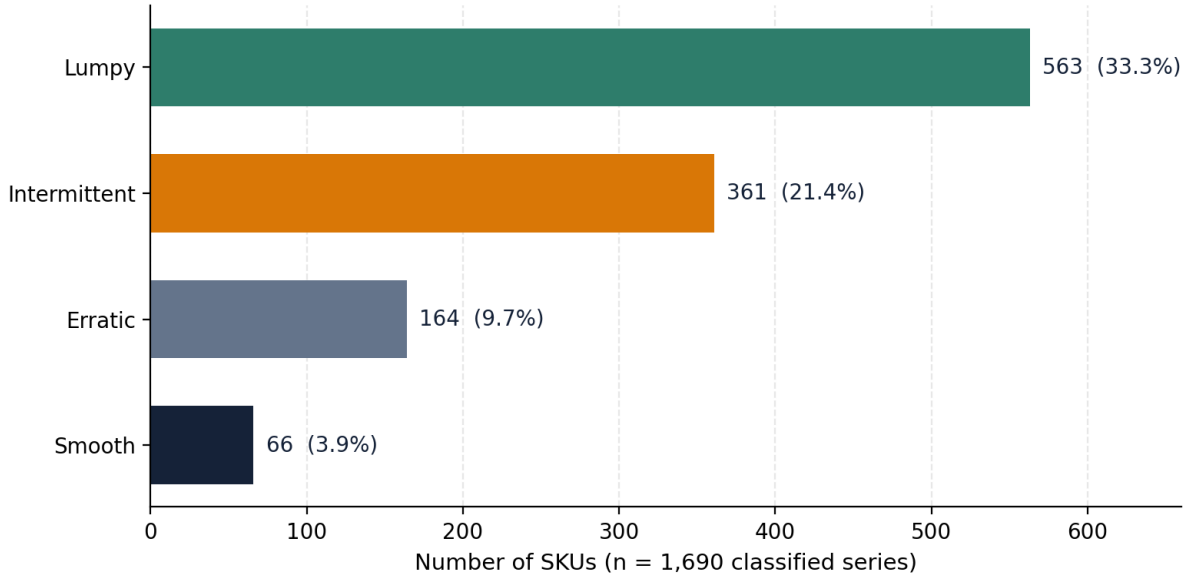

Figure 4.2 - Distribution of demand patterns under the ADI/CV² classification (n = 1,690 classified SKUs). Source: own elaboration.

ADI is computed over each SKU's active range, from its first to its last recorded sale, rather than across the full dataset calendar. This distinction matters in an assortment with meaningful product turnover: applying a common calendar to all SKUs would credit products introduced late or discontinued early with artificial zero-demand months, inflating their ADI and misclassifying them as intermittent.

Two independent measures converge on the same conclusion. The CV-based XYZ classification (Section 4.2) and the ADI/CV² framework are computed from different quantities, yet both indicate that this assortment is dominated by irregular demand, fewer than one SKU in twenty exhibits smooth, regular behaviour.

4.4 Traditional Models: Class-A Products

Class-A SKUs were evaluated under the protocol described in Section 3.5.5. The Syntetos–Boylan approximation achieved the lowest out-of-sample MAE most frequently in all three complete years: 35 of 77 SKUs in 2023, 37 of 87 in 2024, and 43 of 98 in 2025. Exponential smoothing and moving average followed with Croston consistently third and Holt-Winters and ARIMA rarely competitive.

The consistency across three independent years is notable. Even among high-value products with comparatively regular demand, a method designed for intermittent series outperforms conventional smoothing in roughly 45% of cases, an early indication that the demand irregularity documented in Sections 4.2 and 4.3 extends into the Class-A segment rather than being confined to low-value items.

4.5 Traditional Models: Lumpy/Intermittent Products

The central comparison of this study applied the seven-method evaluation specifically to SKUs classified as Lumpy or Intermittent, the demand patterns for which Croston-family methods were developed (Syntetos & Boylan, 2005; Teunter et al., 2011). Of 924 such SKUs, 373 had sufficient history for out-of-sample evaluation.

Croston-family methods achieved the lowest test MAE for 227 of 373 SKUs (60.9%). The Syntetos–Boylan approximation alone accounted for 177 cases (47.5%), Croston for 49 and TSB for one. Moving average won 75 cases (20.1%) and exponential smoothing 71 (19.0%). Median test MAE was 8.96 for the best Croston-family variant per SKU against 11.01 for exponential smoothing, a 19% reduction in typical error, confirmed by a Wilcoxon signed-rank test on paired per-SKU errors ($n = 373$, $p = 1.3 \times 10^{-16}$).

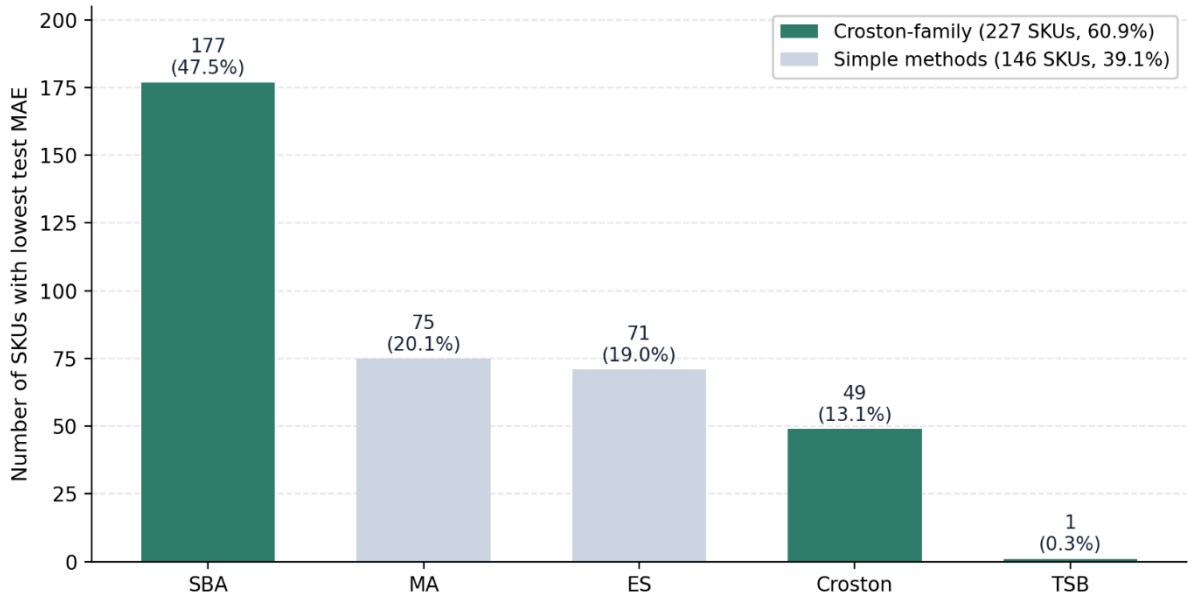

Figure 4.3 - *Best-performing forecasting method by SKU, Lumpy/Intermittent segment, out-of-sample evaluation (n = 373). Source: own elaboration.*

Disaggregating by pattern, SBA's advantage is more pronounced among Lumpy SKUs (147 of 292 wins) than Intermittent ones (30 of 81), consistent with its bias-correction term being most valuable where both demand intervals and demand sizes vary substantially.

Two qualifications temper this result. First, the effect size is moderate rather than decisive: SBA is the single best method for roughly half of intermittent-demand SKUs and simple methods remain competitive in approximately 39% of cases. For an SME weighing implementation effort against accuracy gain, this matters. The theoretically superior method is not universally superior in practice. Second, the advantage is expressed in median terms across a heterogeneous portfolio. Individual SKUs vary considerably, which is precisely why the framework selects a method per product rather than imposing one globally.

4.6 Machine Learning versus Naive Baseline

The Random Forest algorithm has the best average MAE (69.1) across 321 eligible products and won out for 179 products (or 55.8% of overall products) while XGBoost produced 76.7 average MAE and won 95 of the products. Both machine learning-based approaches significantly outperformed the naive lag-1 statistical benchmark (85.3 average MAE), demonstrating that tree-based methods add true value relative to a non-advanced statistical benchmark, which support earlier studies (Douaioui et al., 2024; Makridakis et al., 2018). Moreover, Linear Regression (630.8 average MAE and 14 wins) and the three-model Ensemble technique just have not been close (232.3 average MAE and 33 wins).

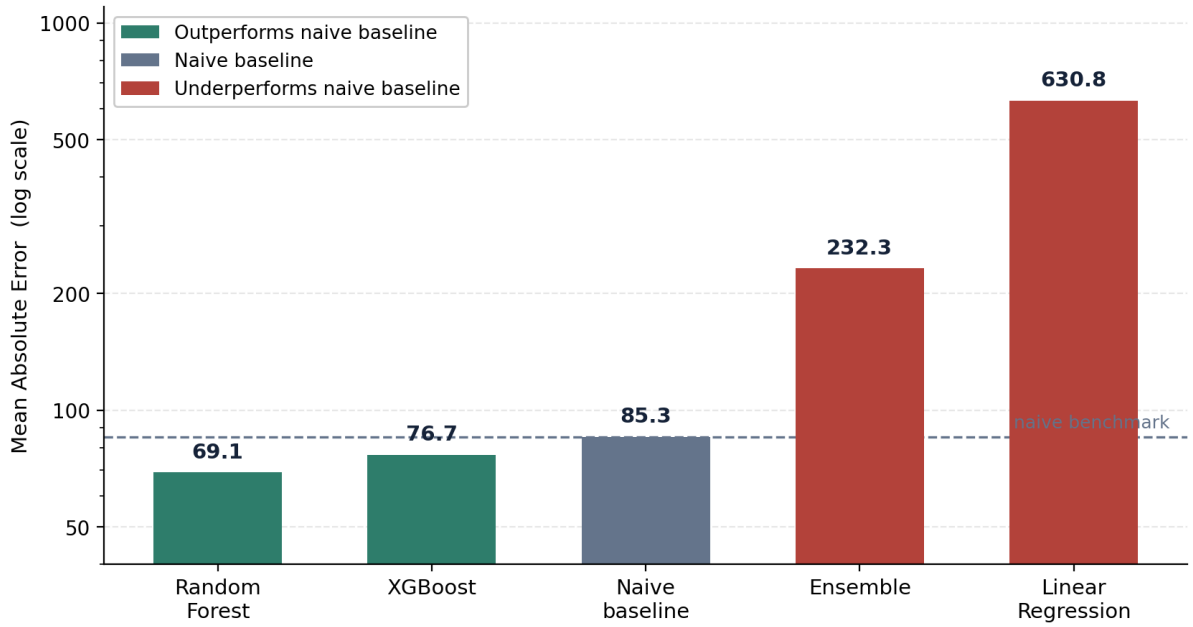

Figure 4.4 -. Mean absolute error by forecasting model across 321 eligible SKUs, logarithmic scale. Lower values indicate better accuracy. Source: own elaboration.

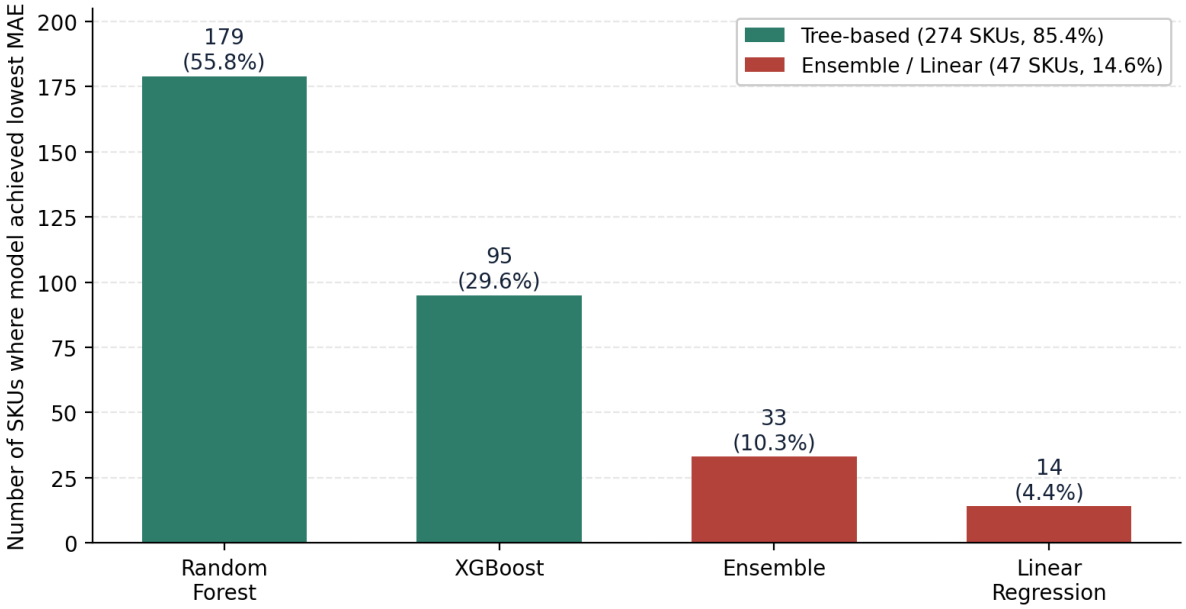

Figure 4.5 - Number of SKUs for which each machine learning model achieved the lowest error (n = 321). Source: own elaboration

Combining one strong model with two substantially weaker models resulted in less predictive performance on average. Therefore, this goes against a belief that ensemble methods always add strength. This means that ensemble methods perform better only when all component models perform similar to one another. The basis for this outcome is straightforward. The linear assumption that Linear Regression makes of a linear relationship between lagged demands and current demands is not appropriate for much of the seasonal and nonlinear nature of our dataset (section 3.3.3). The fact that we used simple average within the Ensemble had significantly diminished the Random Forest's strong contribution to the overall result. A weighted average Ensemble in which the individual models are combined based on their relative back tested performance instead of equally could help address this issue. Sections 5.5 provides the basis for continued research to extend this work.

4.7 Per-SKU versus Pooled (Global) Modelling

The per-SKU strategy outlined above was examined in a second comparative methodology to that of developing a single pooled Random Forest model based on the combined history of all 321 SKUs, with the results normalizing the forecasts based on the historical mean for each SKU (Section 3.5.2). The pooled Random Forest model, while incorporating significantly more data for training, yielded an overall MAE of 80.7, significantly greater than the per-SKU based Random Forest method MAE value of 65.8. The results indicate contrary to the general expectation of improved forecast accuracy through pooling related times series as a result of the pooling learning (Sagaert & Kourentzes, 2025) that in this instance the heterogeneity in the SME's product assortment demand behaviour already identified through the ABC–XYZ and ADI/CV² results (Sections 4.2–4.3) outweighs the sample size

related benefits of pooling. The individual SKU-based forecast showed to more accurately model the unique demand characteristics for each SKU using only 20–40 observations per SKU, than the single pooled Random Forest model developed using an average between diverse SKU's demand behavior The comparison was restricted to 301 of the 321 eligible SKUs, reflecting the intersection of SKUs with valid test-period observations under both evaluation schemes: the per-SKU approach uses a chronological split unique to each SKU's own history, whereas the global pooled model applies a single calendar-date cutoff uniformly; the remaining 20 SKUs did not have observations falling within this shared evaluation window. This difference was confirmed statistically via a paired Wilcoxon signed-rank test across the 301 SKUs with comparable test data ($p < 0.001$). The median MAE was 24.6 for the per-SKU approach versus 28.1 for the global pooled model, a 12% reduction. Because the per-SKU approach selects the best of four candidate models using the same held-out period on which accuracy is reported, a control comparison was conducted using a single pre-specified model. Random Forest alone achieved a mean MAE of 69.11 against 80.69 for the pooled model, confirming that the per-SKU advantage reflects genuine benefit from product-specific modelling rather than an artefact of model selection.

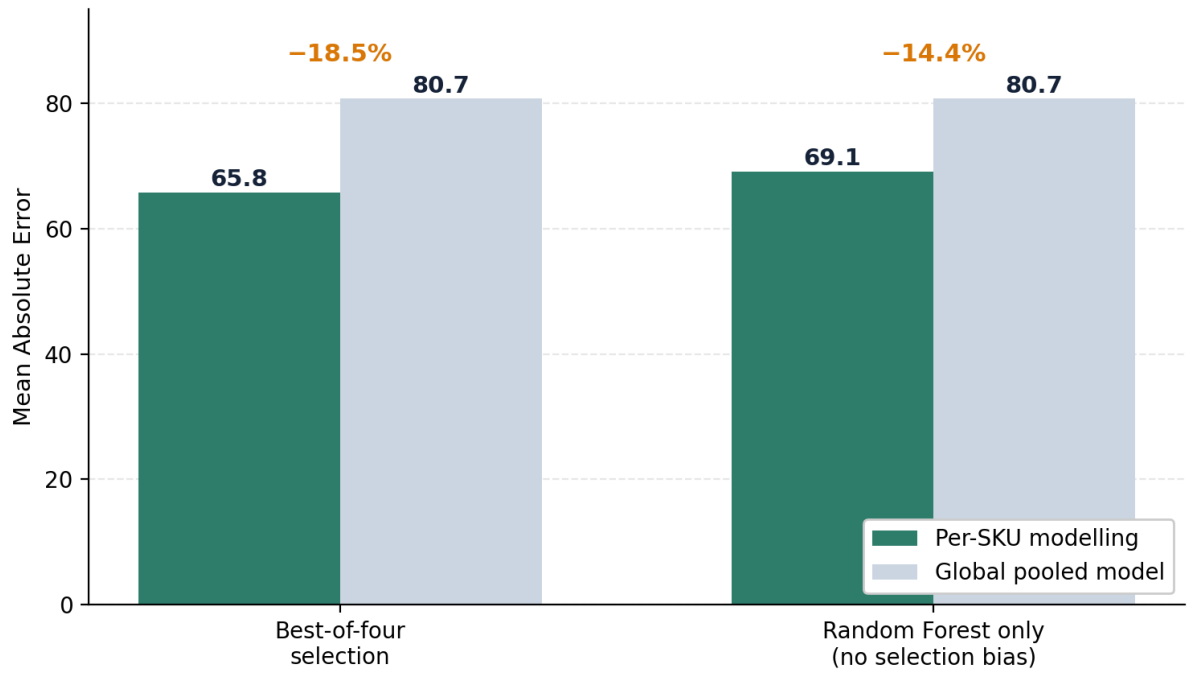

Figure 4.6 - Per-SKU versus global pooled modelling accuracy, with and without model-selection effects (n = 301). Source: own elaboration.

Beyond just the specific SME, this finding suggests to the researcher and practitioner that they should not assume that forecasting strategies based on transfer learning or mean-normalized pooling will outperform per-series modelling simply because there is now more aggregated training data. In situations, such as those typically found in SMEs with a heterogeneous demand behavior across very different product categories in the portfolio,

the benefits of having developed a specific model for a particular series may outweigh any statistical advantage from having a larger pooled sample. Therefore, this should be tested on an individual basis rather than being assumed to be true across the board.

4.8 External Variable Experiment: Aluminium Price

The study described in Section 3.3 added a feature (monthly aluminum futures prices from January 2023 through June 2026) to the per-SKU forecasting pipeline for 321 SKUs, assuming that changes in the cost of materials could affect demand for the material's finished product. Under a controlled comparison — same model, same SKUs, only the feature set changing — per-SKU mean MAE was 68.8 without aluminium price included, versus 74.8 with a one-month lag. Including the feature didn't improve accuracy; if anything it made things worse. This lack of observable cause/effect with aluminum pricing indicates that the purchasing behavior of end customers does not change (detectable results) due to fluctuations in materials pricing at the SME and supplier levels, creating a lag of one month or less in time prior to being reflected in the purchase decision. The inability to demonstrate an observable cause/effect relationship between material pricing fluctuation and end customer purchasing behavior within this dataset is treated as a valid negative result rather than a failure to implement the feature and is discussed further in Section 5.5. As a robustness check on the lag assumption, the experiment was repeated using aluminium price lags of 1, 2, 3, and 6 months. None of the four alternative lags produced an improvement in mean MAE relative to the no-external-variable baseline (68.8), with all configurations performing marginally worse (72.1–74.8). This confirms that the absence of a detectable relationship between aluminium price and end-product demand is not an artefact of the specific one-month lag originally tested.

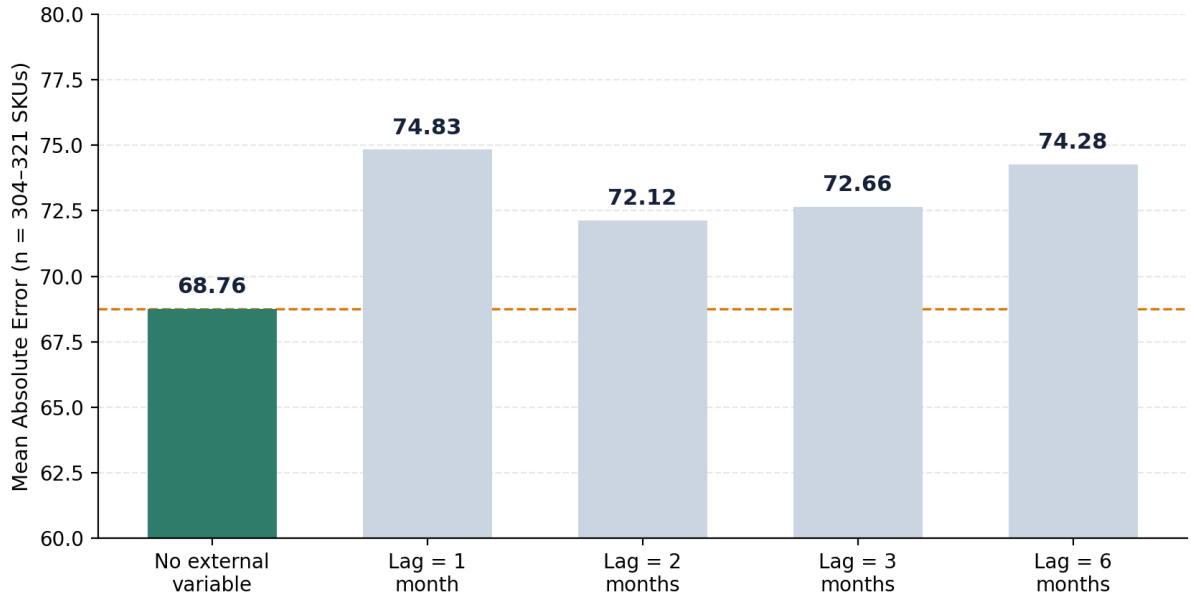

Figure 4.7 - Mean absolute error by aluminium-price lag specification, relative to the no-external-variable baseline (n = 304–321). Source: own elaboration

4.9 Inventory Policy Validation

Using the classification of each stock keeping unit (SKU) into one of the Inventories ABC – XYZ (Section 3.4.4), the safety stock policy has been examined in a retrospective context against the same periods held-out for testing in Sections 4.6 and 4.7, rather than through synthetic Monte Carlo simulation demand situations (Section 3.1). A total of 1,954 SKU-month combinations were evaluated with the policy achieving an overall service level of 94.1%. Safety stock is computed as $z \times 1.2533 \times MAE$, the multiplier converting mean absolute error to an estimate of the error standard deviation: for normally distributed errors, $MAE = \sigma\sqrt{(2/\pi)} \approx 0.798\sigma$ hence $\sigma \approx 1.2533 \times MAE$ (Section 3.4.4, Gonçalves et al., 2020).

Table 4.2 -. Achieved versus target service level by ABC–XYZ class. Source: own elaboration.

Class	N (test periods)	Achieved SL	Target SL	Gap	Avg. shortfall when stockout (units)
AX	42	100.0%	99.0%	+1.0%	-
AY	234	99.1%	98.0%	+1.1%	7.5
AZ	322	96.0%	95.0%	+1.0%	55.1
BX	65	96.9%	97.0%	−0.1%	7.8
BY	269	95.2%	95.0%	+0.2%	162.5
BZ	333	93.7%	92.0%	+1.7%	92.2
CX	199	93.0%	95.0%	-2.0%	27.5
CY	233	91.8%	90.0%	+1.8%	58.6
CZ	257	87.5%	85.0%	+2.5%	38.0
Total	**1,954**	**94.1%**	-	-	-

Note: Gap expressed in percentage points. AX recorded no stockouts across 42 test periods, hence no shortfall magnitude is reported.

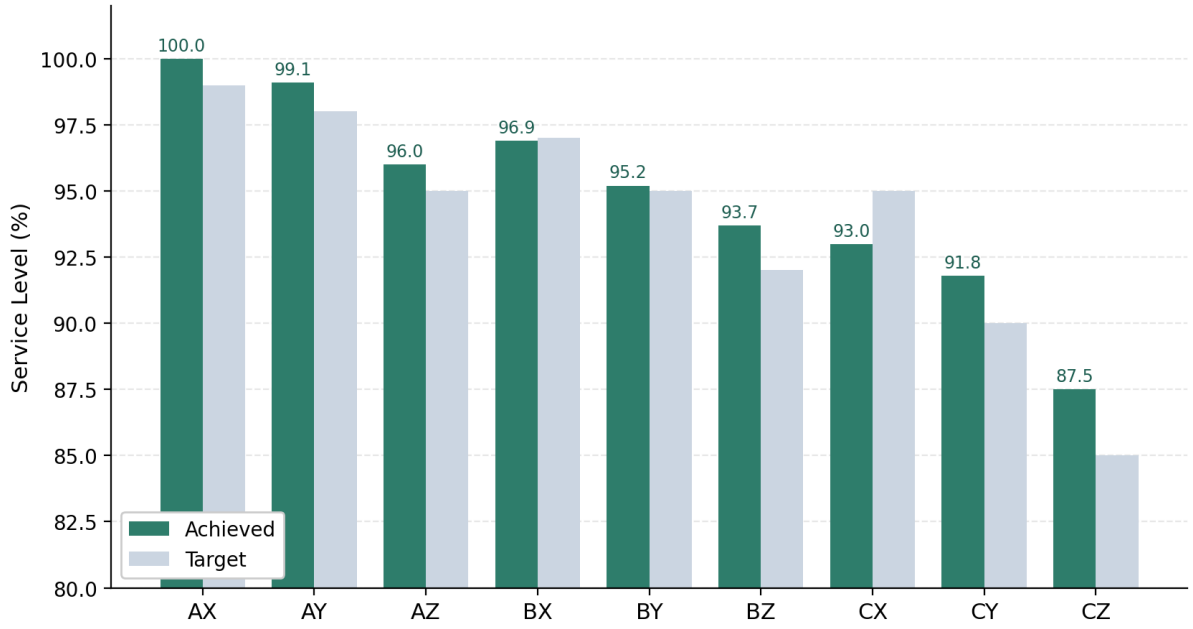

Figure 4.8 - Achieved versus target service level by ABC–XYZ class (n = 1,954 held-out SKU-months). Source: own elaboration.

The pattern across classes is as intended. Attainment declines monotonically from AX to CZ, mirroring the declining service-level targets assigned by the policy table and protection is therefore allocated in proportion to economic importance rather than uniformly. The two classes falling marginally short, BX by 0.1 percentage points and CX by 2.0, are both X-class items whose stable demand yields small absolute buffers, where a single unusual period has disproportionate effect. A separate annualized year-over-year comparison, correcting for the partial 2026 period, showed no systematic distortion, confirming that the methodological adjustment introduced in Section 3.4 was effective. For the SME's management, the policy can be deployed as specified across all classes without additional manual buffers.

Service level and average shortfall measure different things. One tells you how often demand gets met while the other tells you how bad it is when it doesn't. BY is a good example of why the distinction matters: its service level is 95.2%, above target, but its stockouts are the largest in the assortment, 162.5 units on average, simply because the class carries such high absolute demand. BX runs the other way: marginally below target on service level but shortfalls average under eight units when they do happen. A report that only tracks service level would miss this entirely and miss exactly the cases where a stockout actually costs something operationally.

4.10 Synthesis

Sections 4.4 through 4.9 do not add up to a single claim about model complexity. What they support instead is narrower and easier to defend: which forecasting approach works depends on the structure of the demand itself and that has to be established empirically rather than assumed.

The evidence for this comes from a few directions. Methods built specifically for intermittent demand beat general-purpose smoothing on exactly the demand patterns they were designed for. SBA had the lowest out-of-sample error for 47.5% of Lumpy and Intermittent SKUs and Croston-family methods together accounted for 60.9% (Section 4.5). Tree-based machine learning improved on a naive benchmark by 19% but a linear model and an equally-weighted ensemble did considerably worse than doing nothing sophisticated at all (Section 4.6). Product-specific models beat a single pooled model trained on the whole assortment, even though the pooled model had far more training data to work with (Section 4.7).

Put together, this looks less like a complexity gradient and more like a matching problem. Croston-family methods win where demand is sparse and irregular. Tree ensembles win where there is enough per-product history to learn non-linear structure. Simple smoothing still holds its own in roughly 39% of intermittent cases and never gets fully displaced. The failures are just as telling, the ensemble underperformed because averaging a strong model with a weak one just carries the weakness through and the pooled model underperformed because a portfolio this heterogeneous doesn't have a representative "average" product for it to learn from.

Two more findings narrow the framework's scope. The external-variable experiment showed no measurable improvement at any of four lag specifications (Section 4.8). This SME's demand doesn't detectably respond to upstream commodity price movements, which is a negative result but still useful for anyone considering commodity-linked demand sensing. The inventory policy validation reached a 94.1% service level with every class within 2.5 percentage points of its target (Section 4.9), showing that forecast accuracy does translate into operational performance once safety stock is derived correctly from the spread of forecast error.

This also bears on the integration gap flagged in Section 2.7. Most of the literature reviewed there evaluates forecasting accuracy on its own, reporting error metrics without checking whether better accuracy actually produces better inventory outcomes. Here the full chain is traced on one real dataset (segmentation, method selection, safety stock, achieved service level) and the payoff shows up in the results: things that sound reasonable on paper, like pooling data or ensembling models being universally beneficial, didn't hold up under testing in this setting.

Chapter 5 builds its conclusions and managerial recommendations on this basis.

Chapter 5: Conclusions and Recommendations

5.1 Summary of Findings

This study looks at demand forecasting and inventory decisions using four years of ERP transaction data from an SME distributor, covering 1,592 distinct SKUs. The ABC–XYZ and ADI/CV² segmentations show an assortment dominated by irregular demand (fewer than one SKU in twenty behaves smoothly) and flag the AZ segment as a particular concern: 80 products that carry half of all A-tier value at the highest demand variability of any A-class segment.

Three comparative experiments drive the main findings. Methods built specifically for intermittent demand beat general-purpose alternatives on exactly the patterns they were designed for: SBA had the lowest out-of-sample error for 47.5% of Lumpy and Intermittent SKUs and Croston-family methods together accounted for 60.9% (Section 4.5). Tree-based machine learning improved on a naive benchmark by 19% while linear regression and an equally-weighted ensemble did considerably worse than the benchmark itself (Section 4.6). Product-specific models beat a single pooled model trained on the full assortment by 14.4%, even though the pooled model had a much larger training sample to draw on (Section 4.7).

Two more results mark out where the framework's limits are. Aluminium price showed no measurable relationship with demand at any of four lag specifications (Section 4.8). The resulting inventory policy reached a 94.1% service level across 1,954 held-out SKU-months with every ABC–XYZ class within 2.5 percentage points of its target (Section 4.9), confirming that forecast accuracy does translate into operational performance, provided safety stock is derived correctly from the spread of forecast error.

5.2 Managerial Implications

The SME participating in this study must focus their forecasting efforts on demand patterns and categories instead of forecast model sophistication. Method choice should track demand pattern, not be fixed in advance. For the large share of the assortment that's Lumpy or Intermittent, SBA is the strongest single method, and it's cheap to implement (a short recursive calculation, no specialised software needed). Where products have enough regular history, per-product Random Forest models add real value on top. Simple exponential smoothing still holds up for roughly two-fifths of intermittent cases, so it's a reasonable default when implementation effort needs to stay low. What the results rule out is applying one method across the whole assortment (whether that's the simple option or the sophisticated one). The use of machine learning for SKUs requiring reliability enhancements will primarily occur for higher-value and/or less-stable SKUs. In such cases, Random Forest models developed for specific SKUs rather than pooled across the total assortment will yield better quality forecasting results despite smaller sample sizes. The findings from this study regarding the decision-support tools developed in Section 3.4.4 provide managers with

actionable order quantity, safety stock level and reliability indicators. Managers should view the reliability indicator as an actual signal requiring further judgement because of the limited historical data available for those particular SKUs. Additionally, one further practical lesson learned while developing the decision layer was that volume-based prioritisation (which products are largest) and trend-based prioritisation (which products are increasing/decreasing in volume) must be reported together rather than as one single recommendation because a high-volume declining product requires a different managerial response than a high-volume growing product.

5.3 Theoretical and Academic Contribution

This dissertation's main contribution is empirical, not methodological as it tests established assumptions against real SME operational data and sorts out which ones held up.

The sharpest counter-finding is about data pooling. Current practice leans toward global models trained across many series, on the expectation that cross-learning improves accuracy (Sagaert & Kourentzes, 2025). Here the opposite happened: per-product models beat a pooled model by 14.4%, under a control comparison designed to rule out model-selection effects. That doesn't overturn the pooling literature but puts a boundary on it. Pooling helps when series share dynamics and when a portfolio is this heterogeneous, per-series specialisation can win instead. Whether that boundary applies is something to test, not assumption. A second counter-finding concerns ensembling: combining models of unequal quality made accuracy worse, not better, which cuts against the usual presumption that ensembles are robust.

What the evidence confirmed matters too. The theoretical expectation that intermittent-demand methods beat general-purpose smoothing on sparse, irregular series (Syntetos & Boylan, 2005; Teunter et al., 2011) held up, with the caveat that the advantage is moderate, not decisive. Replicating this in a context that hadn't really been tested before (an SME with short per-product histories, where it wasn't obvious the original theory would even apply) has value in its own right.

The study also closes some of the integration gap raised in Section 2.7. Most of the reviewed literature evaluates forecasting accuracy on its own. Here the full chain is traced on one real dataset, from segmentation through method selection to safety stock and achieved service level, which is what Makridakis et al. (2018) have been pushing for (evaluation tied to operational outcomes, not just accuracy metrics).

The methodology itself is worth noting too. Parameters were tuned on training data for every method, with no exceptions. Accuracy was measured strictly out-of-sample, differences were tested with paired significance tests and reported effect sizes rather than just win counts and the null result on aluminium price, tested rigorously rather than

assumed, shows how a well-specified negative finding can be genuinely useful (it steers an SME away from investing in something that wouldn't pay off).

5.4 Limitations

There are many limitations of this research. DL (deep learning) architectures were not implemented as there was insufficient per-SKU (stock-keeping unit) history (section 3.5.3). Although this decision was based upon a data-driven analysis, the results means that there is no direct comparison to those using deep learning models. Every method got the same treatment on parameters: moving average window length, exponential smoothing α and Croston-family α were each tuned by grid search on the training portion only. That rules out tuning asymmetry as an explanation for the differences observed, though the results still depend, more generally, on which parameter grids were searched in the first place. The back testing methodology throughout this study relies on a single chronological train-test split per SKU rather than rolling-origin or k-fold cross-validation; given the limited per-SKU history (12–40 observations), MAE estimates may therefore be somewhat sensitive to which specific months fell within each SKU's test period. Requiring a train-test split cut into the effective sample too: SKUs with fewer than eight monthly observations couldn't be evaluated at all and the partial 2026 period was left out of the annual comparisons entirely. Rolling-origin cross-validation would make better use of the available history while still respecting time order and is worth adopting in future work. The inventory assessment in section 4.9 was based upon historical demand from the retained sample. Therefore, the demand conditions represented only that which actually occurred. One external variable was assessed (aluminium cost) and there was no qualitative assessment conducted via manager interviews, therefore, the results do not take into account tacit operational knowledge of which ERP (enterprise resource planning) information is unable to capture. Finally, this research investigates one SME (small-medium enterprise) in a particular industry, which does not allow for determination as to how far the results would apply in a generic sense to other businesses or industries. The inventory policy validation in Section 4.9 derives safety stock exclusively from the machine learning models (RF/XGBoost/LR/Ensemble) evaluated for the 321 ML-eligible SKUs. It does not incorporate the classical methods (SBA, Croston) shown in Section 4.5 to outperform on Lumpy/Intermittent SKUs specifically. Integrating pattern-specific method selection into the safety-stock pipeline, rather than restricting it to the ML model family, is identified as a priority extension in Section 5.5.

5.5 Future Research

Future research initiatives could take a look at deep learning and pooled-model structures using bigger multi-firm data sets in which per-series history is less limiting. This would allow for a true investigation of whether the pooling disadvantage found here (Section 4.7) is a feature of this dataset’s mixture of elements, or is a larger pattern.

In conclusion, extending the inventory assessment to a complete stochastic simulation as well as testing a broader range of the external variables (such as construction sector activity indices for example relative to the SME's product category would further enhance the practical and theoretical conclusions established in this study. Extend the safety-stock derivation to select forecasts from the full method set (classical and ML) per SKU, rather than from ML models alone, integrating the pattern-specific findings of Section 4.5 directly into the inventory policy

Finally, utilising structured interviews with managers would facilitate triangulating the quantitative results from this project against the managerial operational knowledge that ERP-based analysis does not yield.

5.6 Data Availability and Ethics Statement

The ERP transactional dataset used in this research was received from the SME participant through a confidentiality agreement and is not publicly available. However, the code used to analyse the data is available upon reasonable request. The study did not involve the use of human subjects research, therefore no formal ethics review was needed, and the SME has agreed to allow the use of the anonymised operational data for research purposes after consenting to it through an informed consent process.

References

Academic / Scholarly Sources

Aldahmani, E., Alzubi, A., & Iyiola, K. (2024). Demand forecasting in supply chains using a uni-regression deep approximate forecasting model. Applied Sciences, 14(18), 8110. https://doi.org/10.3390/app14188110

Arend, R. J., & Lévesque, M. (2010). Is the resource-based view a practical organizational theory?, Organization Science, 21(4), 913–930. https://doi.org/10.1287/orsc.1090.0484

Boylan, J. E., & Syntetos, A. A. (2010). Spare parts management: A review of forecasting research and extensions. IMA Journal of Management Mathematics, 21(3), 227-237. https://doi.org/10.1093/imaman/dpp016

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: Forecasting and control (5th ed.). Wiley.

Chopra, S., & Meindl, P. (2021). Supply chain management: Strategy, planning, and operation (7th ed.). Pearson.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage.

Douaioui, K., Oucheikh, R., Benmoussa, O., & Mabrouki, C. (2024). Machine Learning and Deep Learning Models for Demand Forecasting in Supply Chain Management: A Critical Review, Applied System Innovation, Vol 7(5), 93 https://doi.org/10.3390/asi7050093

Gonçalves, J. N. C., Sameiro Carvalho, M., Cortez, P. (2020). Operations research models and methods for safety stock determination: A review, Operations Research Perspectives, 7, 100164 https://doi.org/10.1016/j.orp.2020.100164

Haddara, M., & Elragal, A. (2015). The Readiness of ERP Systems for the Factory of the Future. Procedia Computer Science, 64,721- 728. https://doi.org/10.1016/j.procs.2015.08.598

Hyndman, R. J., & Athanasopoulos, G. (2021). Forecasting: Principles and practice (3rd ed.). OTexts.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting, 22*(4), 679–688.

Hyndman, R. J., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). Forecasting with exponential smoothing: The state space approach. Springer.

Ivanov, D. (2022). Digital supply chains, Industry 4.0, and supply chain resilience. International Journal of Production Research, 60(8), 2391–2404.

Koh, S. C. L., & Saad, S. (2006). Managing uncertainty in ERP-controlled manufacturing environments in SMEs. International Journal of Production Economics, 101(1), 109–127. https://doi.org/10.1016/j.ijpe.2005.05.011

Kubasakova, I., & Kubanova, J. (2024). Utilization of the intersection of ABC and XYZ analysis in stock planning in the warehouse by Covid period. Acta Logistica, 11(3). https://doi.org/10.22306/al.v11i3.532

Lagoda, L., & Klumpp, M. (2024). Efficient warehouse and inventory management: The modified ABC-XYZ analysis as a framework to integrate demand forecasting and inventory control. In M. Freitag et al. (Eds.), Dynamics in Logistics: Proceedings of LDIC 2024 (pp. 390–405). Springer.
https://doi.org/10.1007/978-3-031-56826-8_30

Lapide, L. (2006). Demand forecasting, ERP, and S&OP: What's the connection? *Journal of Business Forecasting*.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and machine learning forecasting methods: Concerns and ways forward. PLoS ONE, 13(3), e0194889.

Petropoulos, F., Apiletti, D., Assimakopoulos, V.,Babai, M. Z., Barrow, D. K., Ben Taieb, S., et al. (2022). Forecasting: Theory and Practice. International Journal of Forecasting, 38(3), 705–871. https://doi.org/10.1016/j.ijforecast.2021.11.001

Sagaert, Y. R., & Kourentzes, N. (2025). Inventory management with leading indicator augmented hierarchical forecasts, Omega,136,103335.
https://www.sciencedirect.com/science/article/pii/S0305048325000510

Saunders, M., Lewis, P., & Thornhill, A. (2019). *Research methods for business students*. Pearson.

Silver, E. A., Pyke, D. F., & Thomas, D. J. (2016). *Inventory and production management in supply chains* (4th ed.). CRC Press.

Stojanović, Đ., & Regodić, D. (2017). The significance of the integrated multicriteria ABC-XYZ method for the inventory management process. Acta Polytechnica Hungarica, 14(5), 29–48

Syntetos, A. A., & Boylan, J. E. (2005). The accuracy of intermittent demand estimates. International Journal of Forecasting, 21(2), 303–314.

Syntetos, A. A., Boylan, J. E., & Croston, J. D. (2005). On the categorization of demand patterns. Journal of the Operational Research Society, 56(5), 495–503.

Teunter, R. H., Syntetos, A. A., & Babai, M. Z. (2011). Intermittent demand: Linking forecasting to inventory obsolescence. European Journal of Operational Research, 214(3), 606–615.

Venkatesh, P., & Sowmiya, P. (2024). ABC-XYZ classification and forecasting for inventory optimization. International Journal of Research Publication and Reviews, 5(12), 5348–5357. https://doi.org/10.55248/gengpi.5.1224.0227

Waters, D. (2021). Inventory control and management (3rd ed.). Wiley.

Industry / Professional Sources (Non-Academic)

AGR. (2025). 2025 supply chain trends report. https://www.agrinventory.com/guides-and-ebooks/2025-supply-chain-trends-report/

AI in the Chain. (2025, June 8). Global PMI signals and supply chain insights for May 2025: A data-driven forecasting guide. https://aiinthechain.com/2025/06/08/global-pmi-signals-and-supply-chain-insights-for-may-2025-a-data-driven-forecasting-guide/

Association for Supply Chain Management. (2025). *Top 10 supply chain trends in 2025*. https://www.ascm.org/making-an-impact/research/top-10-supply-chain-trends-in-2025/

Driessen, M., & van den Bremer, W. (2024). *How to do ABC/XYZ segmentation properly for better inventory optimization*. EyeOn. https://eyeonplanning.com/blog/abc-xyz-segmentation/

International Trade Council. (2024). Scenario exploration: Global state of supply-chain and trade for 2025. https://tradecouncil.org/global-state-of-supply-chain-and-trade-for-2025/

Appendix A: "Supplementary Results Tables"

Table A1 Aluminium price lag-sensitivity

Configuration	**N_SKUs**	**Avg_MAE**
No external variable (baseline)	321	68.76
Aluminum price, lag=1 month(s)	304	74.83
Aluminum price, lag=2 month(s)	304	72.12
Aluminum price, lag=3 month(s)	304	72.66
Aluminum price, lag=6 month(s)	304	74.28

Table A2 Best-performing method by demand pattern, out-of-sample evaluation (n = 373)

Demand Pattern	Best Model	Count	% within pattern
Lumpy (n = 292)	SBA	147	50.3%
	MA	55	18.8%
	ES	54	18.5%
	Croston	36	12.3%
Intermittent (n = 81)	SBA	30	37.0%
	MA	20	24.7%
	ES	17	21.0%
	Croston	13	16.0%
	TSB	1	1.2%

Note: Croston-family methods (Croston, SBA, TSB) account for 183 of 292 Lumpy SKUs (62.7%) and 44 of 81 Intermittent SKUs (54.3%).

Appendix B: "Analytical Pipeline — Code & Data Availability"

The python implementation of this dissertation's results can be broken down into four parts:

1. Extract and transform ERP data into monthly demand series in long format,
2. ABC-XYZ segmentation and ADI/CV^2 classification of demand patterns,
3. Chronological out-of-sample comparison of forecasting methods, with all model parameters optimised on training data only (traditional statistical methods, machine learning and pooled modelling),
4. Translate forecasts into the inventory policy parameters to retrospectively validate results. The complete source code and the underlying ERP database are available upon reasonable request, subject to the confidentiality terms described in Section 5.6.