DEVELOPING A DEMAND FORECASTING MODEL FOR FICTION BOOKS WITH SALES DATA

Lappeenranta–Lahti University of Technology LUT

Industrial Engineering and Management, Master's Thesis

2025

Aapo Aarvala

Examiner(s): Docent, Post-doctoral researcher Lasse Metso

Post-doctoral researcher Minnu Joseph

ABSTRACT

Lappeenranta–Lahti University of Technology LUT

LUT School of Engineering Sciences

Industrial Engineering and Management

Aapo Aarvala

Developing a demand forecasting model for fiction books with sales data

Master's thesis

2025

75 pages, 9 figures, 21 tables and 3 appendices

Examiners: Docent, Post-doctoral researcher Lasse Metso and Post-doctoral researcher Minnu Joseph

Keywords: Demand forecasting, fiction books, publishing industry, supply chain management, quantitative analysis, seasonal demand

This thesis investigates demand forecasting in the context of fiction books and develops a demand forecasting model for a Finnish publishing company using sales data. The research objectives were to identify forecasting methods that capture the seasonal and volatile patterns of fiction demand and to create a forecasting method that supports data-driven decision making across the company's supply chain.

The study was conducted as a case study consisting of theoretical and empirical parts. The theoretical section covered the literature review of demand forecasting from the fiction book context and introduced common forecasting methods and error metrics. The empirical case study used quantitative data analysis to daily sales data of fiction titles released between 2022 and 2024. The forecasting was conducted on a monthly aggregation level using naïve, simple moving average, simple exponential smoothing and a newly developed coefficient forecasting methods. Forecasting performance was evaluated using mean absolute error, mean error and mean squared error. Additionally, a total demand projection model was created to estimate each book's total first-year sales based on cumulative sales patterns.

The results indicated that the coefficient model provided the most accurate monthly forecasts across all error metrics, showcasing the least errors, bias and volatility. The model effectively reduced over-forecasting and adapted to the seasonal and downward trends typical of fiction books. The projection model demonstrated better accuracy as more sales data becomes available. The developed models enable the case company to improve print run planning, inventory management and supply chain coordination.

TIIVISTELMÄ

Lappeenrannan–Lahden teknillinen yliopisto LUT

LUTin insinööritieteiden tiedekunta

Tuotantotalous

Aapo Aarvala

Kaunokirjojen kysynnän ennustemallin kehittäminen myyntidatan avulla

Tuotantotalouden diplomityö

2025

75 sivua, 9 kuvaa, 21 taulukkoa ja 3 liitettä

Tarkastajat: Tutkijatohtori Lasse Metso ja Tutkijatohtori Minnu Joseph

Avainsanat: Kysynnän ennustaminen, kaunokirjat, kustannusala, toimitusketjun hallinta, kvantitatiivinen analyysi, kausivaihtelu

Tässä diplomityössä tutkitaan kaunokirjojen kysynnän ennustamista ja kehitetään suomalaiselle kustantamolle myyntidataan perustuva ennustemalli. Tutkimuksen päätavoitteena oli tunnistaa ennustemetodit, jotka pystyvät parhaiten mukautumaan kausiluonteiseen ja vaihtelevaan kaunokirjojen kysyntään sekä luoda ennustemalli, joka tukee datapohjaista päätöksentekoa kohdeyrityksen toimitusketjussa.

Tutkimus toteutettiin tapaustutkimuksena, joka koostui teoreettisesta ja empiirisestä osasta. Teoriaosiossa tehtiin kirjallisuuskatsaus kysynnän ennustamisen periaatteista kaunokirjallisuuden kontekstissa sekä esiteltiin keskeisiä ennustemenetelmiä ja virhemittareita. Empiirisessä osassa hyödynnettiin kvantitatiivista data-analyysia vuosina 2022–2024 julkaistujen kaunokirjojen päivittäiseen myyntidataan. Ennustaminen suoritettiin kuukausitasolla käyttäen naiivia menetelmää, yksinkertaista liukuvaa keskiarvoa, yksinkertaista eksponentiaalista tasoitusta ja kehitettyä kerroinmallia. Tarkkuutta arvioitiin käyttämällä keskimäärästä absoluuttista virhettä, keskimääräistä virhettä ja keskimääräistä neliövirhettä. Lisäksi laadittiin kokonaiskysynnän ennustemalli, jolla arvioitiin kunkin kirjan ensimmäisen vuoden kokonaismyyntiä kumulatiivisten myyntimallien perusteella.

Tulokset osoittivat, että kerroinmalli tuotti tarkimmat kuukausiennusteet kaikilla virhemittareilla mitattuna. Malli vähensi ennustevirheitä, vinoumaa ja vaihtelua. Kerroinmalli vähensi yliennustamisen ongelmaa ja mukautui tehokkaasti kaunokirjoille tyypillisiin kausiluonteisiin ja laskeviin trendeihin. Kokonaiskysynnän ennustemallin tarkkuus parani, kun myyntidataa kertyi lisää. Kehitetyt ennustemallit auttavat kohdeyritystä parantamaan painosmäärien suunnittelua, varastonhallintaa ja toimitusketjun koordinointia.

ACKNOWLEDGEMENTS

The completion of this thesis marks the culmination of my studies at LUT University. I am extremely grateful for the people I met and the friends I made along the way. I will always cherish the moments we have shared and look forward to what the future will bring.

I would like to sincerely thank the case company for providing me the opportunity to explore new and engaging topics related to demand forecasting in the publishing business. Special thanks to my supervisor at the case company, Timo, who gave me valuable tips and insights along the way.

Finally, I would like to express my gratitude to my family and friends who have supported during the completion of this thesis and throughout my studies.

November 20th, 2025

Aapo Aarvala

ABBREVIATIONS

AHP	Analytic Hierarchy Process
AI	Artificial Intelligence
ARIMA	Autoregressive Integrated Moving Average
CV2	Squared Coefficient of Variation
ES	Exponential Smoothing
IDI	Inter-demand Interval
MA	Moving Average
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
ME	Mean Error
ML	Machine Learning
MSE	Mean Squared Error
POS	Point of Sales
RMSE	Root Mean Squared Error
SES	Simple Exponential Smoothing
SKU	Stock Keeping Unit
SMA	Simple Moving Average
SMAPE	Symmetric Mean Absolute Percentage Error

Table of contents

Appendices

Appendix 1. Fiction releases by month for each year

Appendix 2. Multipliers of the coefficient model

Appendix 3. The average cumulative sales at the end of each month for monthly releases

Figures

Tables

DECLARATIONS

Turnitin

The originality of this thesis has been reviewed with the Turnitin similarity checking service.

AI usage

The author of the thesis, Aapo Aarvala, used the following AI-tools during the preparation of the thesis:

1. Generative AI
    a. Purpose of use: Idea generation, structuring, text flow
    b. Explanation of the use of the tool: Generative AI was used to identify more topics related to the thesis, to help with thesis structure and to improve readability as well as to reduce repetition

Responsibility

The author, Aapo Aarvala, takes full responsibility for the content of this thesis and has reviewed and edited the content generated by the possible use of AI tools.

1 Introduction

The first chapter discusses the background, objectives and scope as well as research questions for the thesis. The section also highlights the research methods and data utilized in the study. Moreover, the structure of thesis is introduced at the end.

1.1 Background

Forecasting is an important activity in businesses for all decision-making tasks, everywhere form inventory management and scheduling to planning and strategic management. (Petropoulos et al., 2014) In the book publishing industry, forecasting plays an equally critical role in helping companies adapt to market changes and technological disruptions.

Global digitalization has rapidly transformed the modern lifestyles, leading to the swift displacement of traditional industries. The rise of electronic books has contributed to the decline of the paper book market, putting significant pressure on the broader publishing industry. Initially, there were concerns that advancements in e-book technology and software would pose a serious threat to conventional publishing. However, while the magazine sector has largely shifted to digital formats, the demand for physical books has remained relatively stable across different genres from fiction, nonfiction, practical to educational materials. (Park, Lee & Doo, 2020) This suggests that book demand forecasting remains necessary for physical copies and requires a strong understanding of traditional consumption patterns.

In today's dynamic and competitive landscape achieving long-term sustainability requires companies to acknowledge the critical role of demand forecasting. Companies must allocate sufficient resources to improve the accuracy and efficiency of forecasting. Robust forecasting systems should be capable of combining changing and emerging information while offering proficient analytical tools to interpret relevant data. The ultimate goal of demand forecasting should extend beyond generating prediction to deriving strategic implications. Therefore, decision makers and data analysts within organizations must be equipped to extract meaningful insights and propose strategies that align forecasting outcomes with long-term business objectives. (Park, Lee & Doo, 2020) In the context of

book publishing, analytical forecasting capabilities are essential for predicting market trends, managing printing volumes and optimizing logistics and inventory.

Inaccurate forecasting of book demand and improper estimation of print quantities can lead to challenges in inventory management due to overproduction. Moreover, if bookstores fail to stock the book titles customers are looking for, the risk of losing customers to competitor retailers increases. Therefore, accurate demand forecasting of books is crucial for making sound business decisions, minimizing operational inefficiencies and to ensuring the long-term competitiveness of publishing companies.

1.2 Objectives, scope and research questions

This master's thesis explores effective forecasting methods for fiction book demand. As highlighted in the background, there is a need for accurate demand forecasting in the evolving book market. Precise demand forecasting is especially pivotal in the fiction book market where product life cycles are short, seasonal fluctuations are evident, and reader preferences can be volatile and influenced by unpredictable trends and events. Inaccurate forecasting can result in excess inventory or lost sales opportunities. Given these challenges, this thesis seeks to address several key questions. First, how can sales data be utilized to forecast fiction book demand in a way that captures the relevant patterns and trends. Second, which forecasting models produce the most accurate forecasts for fiction titles. Third, how can the resulting forecasts be applied to inform more data-driven decisions in the supply chain from book printing volumes to inventory management. By exploring these questions, this thesis aims to contribute practical insights and help improve operational efficiency in fiction book supply chains. The objective is to identify the most accurate method for fiction book demand forecasting based on sales data of previous releases. The research attempts to provide the case company suggestions about the most efficient forecasting techniques and ways to utilize the demand forecasts to improve decision-making in the supply chain. The research is limited to fiction books sold at the physical retail locations. The study seeks to answer the following research questions:

- *How can sales data be utilized to forecast fiction book demand?*
- *Which forecasting methods produce the most accurate forecasts for fiction books?*
- *How can the demand forecasts be used to support decision-making in the case company?*

1.3 Research methods and data

The thesis is conducted as a case study that examines fiction book demand forecasting using real sales data. The analysis utilizes three years of fiction book sales data from 2022, 2023 and 2024. Each dataset contains daily sales figures for each individual fiction titles. The titles included in each dataset are unique, as the data only covers books that were considered newly released in the corresponding year. Therefore, forecasting activities are carried out separately for each dataset.

Demand forecasting is performed on a monthly basis for each individual book. The demand forecasting is conducted by using quantitative forecasting methods: naïve forecasting, simple moving average forecasting, simple exponential smoothing and coefficient model forecasting. Forecasting accuracy is measured by using mean absolute error, mean error and mean squared error. These error metrics provide a comprehensive view of forecasting performance since they measure monthly forecasting accuracy, forecasting bias and the variance of errors.

The forecasting methods are compared both on a yearly basis and according to the release month of each title. Additionally, a projection model based on average cumulative sales is employed to estimate the total demand for each fiction title. This approach allows the identification of demand forecasting techniques that provide the most reliable and applicable results for fiction book demand forecasting.

1.4 Thesis structure

The first section provides an overview of the thesis, highlighting the background, main objectives, scope, limitations, research questions and the research methods and data used in the thesis. The theoretical foundation of the thesis is divided between chapters two and three.

The second chapter reviews the literature on the objectives and challenges of book demand forecasting, the demand forecasting process and the factors affecting demand forecasting accuracy in the context of fiction books. The third chapter introduces quantitative and qualitative forecasting methods, as well as combined approaches and error metrics used to evaluate forecasting performance. The fourth section presents the empirical case study, including an introduction to the case company, the data collection process, the selection of forecasting methods and the performance evaluation of the chosen models. The fifth chapter discusses the main findings, addresses the research questions and highlights the key challenges and limitations identified in the study. Finally, the sixth chapter… The structure of the thesis is summarized in the table below.

Table 1. Structure of the thesis

Section	**Content**
1. Introduction	Presents the background, objectives, scope, research questions and research methods of the study
2. Demand forecasting	Reviews literature on the objectives, challenges, processes and factors affecting forecasting accuracy in the context of fiction books
3. Demand forecasting methods and techniques	Describes quantitative and qualitative forecasting methods, combined approaches and accuracy metrics
4. Empirical study	Introduces the case company, outlines data collection and method selection and presents the performance evaluation of the forecasting models
5. Results	Discusses the key findings of the empirical analysis, answers the research questions and identifies challenges
6. Conclusions	Highlights main conclusions, provides recommendations for future research and discusses limitations

2 Demand forecasting

This section reviews the existing literature on book demand forecasting with a focus on fiction titles. It first examines the objectives and challenges associated with forecasting book demand, highlighting the unique characteristics of the publishing industry that influence forecast accuracy. The section then discusses the general demand forecasting process and factors affecting forecasting performance.

2.1 Objective of demand forecasting

Demand forecasting involves analysing historical data to build models that identify patterns and predict future demand (Mas Machuca, Sainz & Martínez Costa, 2014). It is widely applied across various business sectors and organizations of all sizes. The relevance of demand forecasting spans multiple industries, including agriculture, construction, manufacturing and service sectors, since their activities rely on accurate demand forecasting. (Park, Lee & Doo, 2020)

Demand forecasting is a key component in the operations of contemporary organizations since most supply chain management decisions depend on forecasts for effective support (Babai, Boylan & Rostami-Tabar, 2022). It supports a wide range of business decisions across operational, tactical and strategic levels, such as capacity planning, resource allocation, marketing and promotional planning, demand planning, competitive analysis and production scheduling. It is essential for companies to clearly define the forecasting goal in terms of forecasting horizon and time granularity to support the decision-making context. (Kourentzes, Rostami-Tabar & Barrow, 2017)

Babai et al. (2022) identify two main dimensions of granularity in supply chain decision-making: time and product. Product-related decisions range from those focused on individual stock keeping units (SKU), such as inventory control, up to broader considerations involving all SKUs, such as aggregate capacity planning. In terms of time, operational decisions typically occur on a daily or weekly basis, while tactical and strategic decisions are made over longer periods, such as monthly or annually. There are advantages to co-ordinating plans across these levels with forecasting to ensure coherence between strategic, tactical and

operational plans and decisions over different time horizons. The required level of detail required in forecasting, both cross-sectionally and temporally, should be guided by specific planning and decision-making needs. (Syntetos et al., 2016) Demand forecasting can also be performed across multiple hierarchical levels simultaneously without being limited to data from a specific level. Forecasting methods can incorporate information from different levels, allowing for greater flexibility through bottom-up and top-down approaches. (Babai et al., 2022) These relationships are illustrated in Figure 1 below.

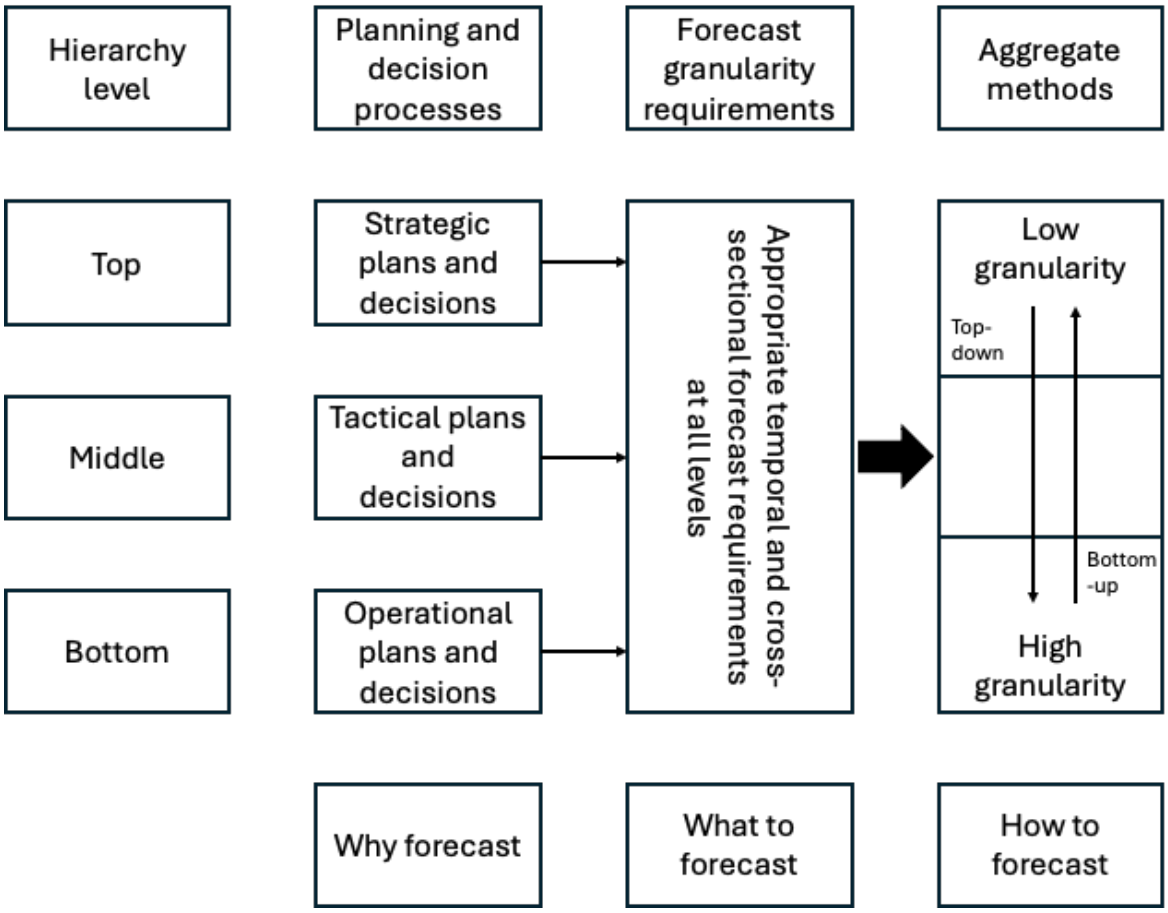


Figure 1. Demand forecasting framework for organizational hierarchies and granularity levels (modified from Babai et al., 2022)

Syntetos et al. (2016) discuss that forecasting demand in retail naturally occurs in three key dimensions: location in the supply chain hierarchy (store, distribution center, supply chain level), product hierarchy (SKU, category, total) and time granularity (day, week, month, year). Typically, lower levels of the supply chain require finer time resolutions, for example retail stores may need daily forecasts, whereas distribution centers rely on weekly forecasts

(Petropoulos et al., 2022). Hierarchical forecasting has emerged as an effective tool for producing coherent multi-level forecasts. The historical time series data used in demand forecasting can be aggregated or disaggregated to align with specific forecasting requirements. Both single and multiple aggregation levels have been found useful in demand forecasting. Single aggregation level forecasting often requires extensive contextual knowledge, while applying multi-level forecasting tends to be more robust to model uncertainty. (Kourentzes, Rostami-Tabar & Barrow, 2017)

Demand forecasting enables businesses to predict customer purchases and plan the production or delivery of goods and services (Syntetos, Boylan & Disney, 2009). To address the complex task of demand forecasting, a wide range of methods are being used and developed, including statistical and qualitative techniques, artificial intelligence-based methods, and hybrid models. (Kolková, Ključnikov, 2021)

In the publishing industry, the objective of book demand forecasting is to estimate the potential purchase level of a book. The goal is to properly allocate resources for a particular book which decreases the derived expenses related to production processes, distribution and storage. (Hildick-Smith, 2006) Accurate demand forecasting is important, since most of these costs are associated with number of copies printed. Reliable demand predictions not only reduce direct production costs but also enhance the entire operational efficiency. Forecasts that accurately estimate future demand allow publishers to optimize production schedules, increase profitability and minimize losses. (Castillo et al., 2017) In essence, accurate forecasting of individual book's market demand is crucial to nearly all major business decisions in publishing, from new title selection to retail purchasing (Hildick-Smith, 2006).

2.2 Challenges of demand forecasting

There have been developed a wide range of methods for demand forecasting. However, these methods often face similar challenges related to their complexity, long computing time and need for high performance IT infrastructure. (Kolková, Ključnikov, 2021) One of the most persistent challenges that forecasters keep facing is the selection of the most appropriate method for a specific dataset or individual time series. (Petropoulos et al., 2014)

The primary challenges in demand forecasting differ across industries and contexts but some common issues are related to intermittent or irregular demand patterns, a lack of historical data for new product forecasting, short product life cycles and hierarchical data structures (Tang & Dong, 2021). The emergence of the big data era has also introduced new challenges: the need to forecast large volumes of related time series and managing the growing number of external variables that considerably impact the future demand. (Petropoulos et al., 2022)

Swaminathan & Venkitasubramony (2024) highlighted that demand forecasting in the fashion industry presents significant challenges due to the frequent introduction of new products, short product life cycles, returns and seasonality. Similarly, Ren, Chan & Siqin (2020) noted that forecasting demand for new items or products is difficult due to the absence of historical data in the fashion industry. These challenges are also relevant to fiction book forecasting due to similar product attributes and forecasting environment. Swaminathan & Venkitasubramony (2024) brought up several demand forecasting obstacles for fashionable products that can also be applied to fiction book forecasting:

- **Limited dataset availability:** New releases, especially from debut authors, often lack historical sales data, which hinders the accuracy of the forecasting model. Additionally, pre-release interest may not convert into actual sales. Although many studies emphasize the need for larger datasets, restricted datasets can be enriched by utilizing comparable titles and external sources.

- **Demand censoring:** Demand censoring effect occurs when the actual demand exceeds the availability of the product. This means that stock-outs can cause lost sales which leads to underestimating the actual demand. Lost sales due to stock-outs can be estimated by combining point-of-sales (POS) data along with inventory records and monitoring waitlist or backorder data.

- **Uncertainty in exogenous variables:** External factors, such as market trends, awards, media coverage and economic indicators significantly impact fiction book demand, but they are difficult to identify, quantify and incorporate into forecasting models. External factors can be accounted for by using more sophisticated machine learning (ML) models with exogenous inputs or expert judgement.

- **Intermittent demand:** Fiction books exhibit short sales cycles with sharp peaks, for example at launch or during Christmas. These periods of high demand are followed by periods of low or zero demand, meaning there are demand fluctuations. This creates struggles for traditional forecasting methods and could be tackled by applying intermittent demand models or more complex artificial intelligence (AI) techniques.

- **Overfitting and underfitting:** ML models risk overfitting to noisy data or underfitting due to oversimplification. Using regularization techniques and ensemble models to balance complexity and generalizability or expert judgement to adjust model outputs could mitigate these issues.

- **Computation time:** AI-based and other more complex models improve forecasting accuracy, but they often require considerable computing time which makes real-time forecasting unattainable. Optimizing or simplifying model architecture, for example by clustering similar books, can help to address these issues.

Demand forecasting is especially challenging for book publishers since they release new products at notably higher frequency than other industries, requiring frequent and accurate demand predictions. The biggest challenge is determining the optimal print quantity. Overestimating demand leads to financial losses and increased inventory costs, whereas underestimating causes missed sales opportunities and potential reputational harm. (Castillo et al., 2017) Hildick-Smith (2006) highlighted that forecasts and production plans tend to be consistently overly optimistic in the publishing industry, resulting in approximately 40% of books remaining unsold at retail locations and being returned to the publisher at considerable cost. Castillo et al (2017) also noted that overprinting is the most prevalent issue in practice, while they underline that underestimating demand is less problematic since reprints can be conducted before the stock is completely depleted. Over production and estimation tends to occur because consequences of a book being unavailable are more severe due to the short lifecycle of books (Hildick-Smith 2006). Underestimating and reprinting reintroduces the same forecasting challenge in deciding the print quantity (Castillo et al., 2017). Limited data availability is also a significant challenge in book forecasting. The short shelf life and relatively modest revenue generated by a single book combined with the large number of new titles released annually, have made it technically and financially challenging to implement data-driven forecasting approaches. (Hildick-Smith, 2006) Lack of historical

data, short lifetime of products, the high number of new titles released, fluctuating demand and influence of external factors are highlighted as the main challenges in book demand forecasting.

2.3 Demand forecasting process

Research has shown that adhering to a well-defined forecasting process leads to more accurate and meaningful results (Fildes & Goodwin, 2021). There often can be a tendency to rush into forecasting by heavily relying on automated tools or software which leads to bypassing foundational steps. While the foundational steps may seem tedious or boring, they are grounded in decades of proven practice and neglecting them can compromise quality of forecasts. Therefore, it is essential to follow a structured process to achieve accurate forecasting results. Structured forecasting process forms a fundamental basis for all forecasting activities whether that be predicting the weather or estimating sales volumes. (Sanders 2017, 17)

There are few core principles of forecasting that apply regardless of the specific application. Firstly, forecasts are rarely perfect since forecasting always involves some degree of error due to the uncertainty of future events. Aiming for perfect forecasting accuracy should not be intended and the focus should be on minimizing large errors and improving average forecasting performance to better prepare for the future. Secondly, forecasting for aggregated groups tend to be more accurate than for individual items since the variation between different items tend to offset each other. This means that the expectation for forecasting accuracy needs to be adjusted according to the aggregation level. (Sanders 2017, 18-19) The higher the aggregation level, the higher the expectation for forecast performance. One common approach is the use of top-down forecasting strategies where high-level forecasts are broken down into finer details. Another approach is a bottom-up strategy where forecasts of individual items are aggregated to higher level projections. (Widiarta, Viswanathan & Piplani, 2009) Thirdly, short-term forecasts are more reliable than long-term ones since shorter time horizons involve less uncertainty and change, as longer-term forecasts are more prone to shifts in trend patterns (Sanders 2017, 18-19). Regularly updating forecasts helps to reduce considerable short-term errors and clarify the longer-term projections (Madhavan

et al., 2023). Together, these principles set realistic expectations and guide efforts to achieve the most accurate and applicable forecasts possible (Sanders 2017, 18-19).

To produce reliable and credible forecasts, a structured forecasting process should be followed regardless of the type of forecast or model being used (Sanders 2017, 20). Sanders (2017, 20-25) has refined the demand forecasting process to six key steps:

1. **Decide What to Forecast:** Identify, define and address the real problem, and decide what needs to be forecasted. Consider the time period and collect data at the most detailed level possible. Try to use short intervals between forecasts and appropriate units of measure of the variable.

2. **Clean the Data:** Forecast accuracy is highly dependent on the data quality, meaning high-quality data produces better forecasts. Data quality can be enhanced by cleaning the data by replacing missing values with moving averages, adjusting working days and other calendar effects, accounting for outliers and using deflating monetary values to adjust for inflation.

3. **Identify Data Patterns:** Historical data needs to be analysed to identify patterns in the data. Common data patterns are level or horizontal (data fluctuates around a stable average), trend (values increasing/decreasing over time), seasonality (regular cycles e.g. based on holidays or seasons) and cycles (created by economic fluctuations or product life cycles, and as a result are difficult to predict). On top of these patterns, it must be recognized that the data contains random variations that cannot be predicted.

4. **Select an Appropriate Model:** After identifying the data patterns, a model should be chosen which matches the data patterns and the forecasting requirements. When selecting between different forecasting models, factors such as amount and availability of data, desired accuracy, forecast time horizon and model suitability for observed patterns should be taken into consideration. Forecasting methods and techniques are discussed in more detail in section 3.

5. **Generate the Forecast:** Forecasts are generated by inputting the cleaned data into the selected model to produce forecasted numerical values. This is typically the

easiest part of forecasting, but it is important to understand what goes to the computation, such as the data and the model being used.

6. **Monitor Forecast Accuracy:** The produced forecasts should be compared to the actual outcomes to measure accuracy. Forecast error information should be used to refine models and improve future forecasting. Forecasting is a continuous improvement process since there is always new information and data becoming available.

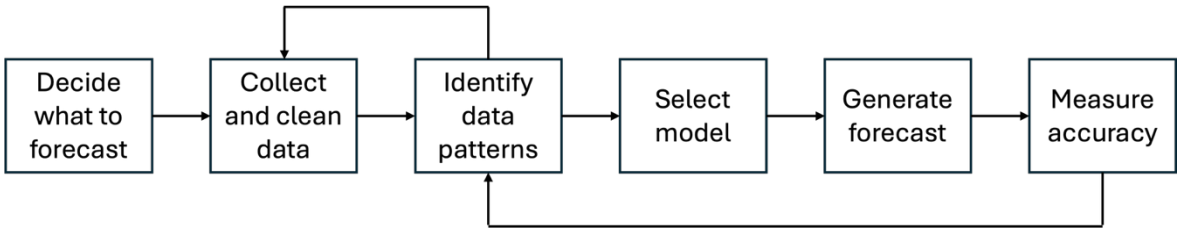


Figure 2. The forecasting process (modified from Sanders, 2017)

There is an evident need for improving forecasting systems and processes in the book publishing industry since its heavy reliance on experience-based decision-making. This is causing inefficiencies since these methods are often not adequate for predicting the potential sales of individual books. Improved forecasting approaches for book demand are required if companies in the industry desire to grow sales and profitability, making them more equipped to face the increasing competition from alternate sources of entertainment. Efficient forecasting techniques and processes help publishers to prioritize, distribute and project sales of books. This also allows book retailers to operate at a higher productivity level. Therefore, adopting structured forecasting methods and processes strengthen the business performance of companies within the publishing industry and the overall book market. (Hildick-Smith, 2006)

2.4 Factors affecting forecasting accuracy

Demand forecasting accuracy is influenced by several key variables such as market volatility, economic conditions, seasonality, data quality, and the forecasting techniques utilized. It is important for companies to understand these factors to improve their forecasting capabilities and operational efficiency. (Chandran & Khan, 2024)

Fluctuations in market have a significant impact on demand forecast accuracy. Hence, forecasting techniques ought to be adaptive to market changes to generate more reliable results. Companies that have demonstrated a greater capability of adapting forecasts to volatile market conditions achieved stronger operational performance and resilience in managing uncertainty. (Chandran & Khan, 2024) Academic literature has shown that macro-economic variables influence consumer behaviour. Key indicators such as consumer price index, index of consumer sentiment and unemployment rates have a central role in explaining retail demand. These relationships are generally well understood and could help to clarify the discrepancies in forecasted and actual demand. However, integration of such economic factors into demand forecasting models remains limited in practice. (Haque, Amin & Miah, 2023)

Petropoulos et al. (2014) defined seven dominant factors affecting forecasting accuracy of time series data extrapolation methods: seasonality, trend, cycle, randomness, the number of observations, inter-demand interval (IDI), squared coefficient of variation (CV2) and forecasting horizon as a strategic decision. They found that for regular data where demand occurs in each and every period, cycle and randomness have the largest negative effect on forecasting accuracy. Increasing the forecasting horizon decreases the forecasting accuracy for this type of data. For intermittent data, higher IDI has the most significant effect on reducing forecasting accuracy. Also, the greater the CV2 of the intermittent time series data, the less precise are the forecasts generated. For all kind of time series data, it was found that increasing the number of observations or the length of the time series has a slight positive effect on forecast accuracy. (Petropoulos et al., 2014)

Data quality related factors also affect forecast accuracy. Inaccurate, missing or noisy data reduces forecasting precision, meaning that low quality data inputs produce low quality forecasting outcomes regardless of the type of model being utilized. (Green and Armstrong, 2012) Gelper, Fried & Croux (2010) emphasize that data pre-cleaning is rarely performed in

practice, especially to non-stationary time series, even though it has the potential to significantly enhance forecasting accuracy. Data availability poses risks for precise forecasts since limited data can lead to incorrect assumptions or improper model usage. Demand censoring must be accounted for since stockouts or store closures can lead to underestimating the actual demand which hinders the performance of the forecasting model.

Data granularity influences the ability to detect patterns in the data. Patterns may be difficult to identify on very fine-grained data, such as on a daily or individual product level, but patterns may emerge when aggregating the data to a higher hierarchy, such as monthly or product category level. This means that granularity must be considered in demand forecasting because it can be difficult to produce precise forecasts on the low aggregation levels. It has been discovered that product family and aggregation forecasts yield more accurate results than individual item forecasting. (Ren, Chan & Siqin, 2020) Aggregation has been found to beneficial in demand forecasting, since higher-level forecasts generally outperform lower-level ones. Accuracy gains are evident when forecasting is conducted on aggregated data, rather than aggregating forecast from disaggregated data. Therefore, it is important that the level of required forecasting matches the level of available data collected. (Kourentzes, Rostami-Tabar & Barrow, 2017)

The forecasting method selected greatly impacts the accuracy of forecasts. It is important to choose a suitable forecasting method for the dataset to achieve accurate forecasting results. (Green and Armstrong, 2012) For instance, quantitative methods are most applicable when sufficient historical data is present. If the data is multidimensional or complex, more advanced computational methods such ML and AI based techniques can perform better. (Lei et al., 2024) For new products or situations with limited data, judgemental or other qualitative demand forecasting methods can be preferred. One approach that has generated accurate demand forecasts in numerous contexts is the combination of multiple forecasting methods. Combining forecasting methods often leads to more accurate forecasts due to reducing the variance of forecasting errors and uncertainty in predictions. Combining approaches utilize more sophisticated weightings and information criteria to integrate results from multiple models. (Petropoulos et al., 2014) However, complex methods should not be used if notable accuracy gains are not achieved because they are more difficult interpret which causes overlooking of errors and improper method application (Green and Armstrong, 2012).

3 Demand forecasting methods and techniques

This section explores different methods for forecasting demand. It begins introducing quantitative methods and then moving to qualitative methods. Also, techniques that combine aspects from multiple methods are discussed in this chapter. Lastly, the section highlights key factors regarding evaluation and selection of forecasting models.

3.1 Quantitative methods

Quantitative or statistical forecasting methods primarily rely on analysing and modelling time series trends and patterns derived from historical sales data. These methods are widely applied in demand forecasting for three key reasons. Firstly, they are relatively easy to implement and operate. Secondly, they often offer fast computation and quick generation of forecasts. Thirdly, they provide clear and interpretable results which makes them simple to integrate with company decision-making processes and operations, such as inventory management or production planning. This makes quantitative approaches suitable for analysing how time-series trends, including seasonality, affects future demand for products with stable demand. (Ren, Chan & Siqin, 2020; Zhang et al., 2024)

Time is often regarded as the central factor in demand forecasting, as customer demand typically evolves in temporal patterns. Time series techniques that rely on the statistical characteristics of historical data, such as trends, seasonality and cycles, are frequently used to estimate the future demand of products. (Petropoulos et al., 2014) Common quantitative methods include moving averages (MA), regression analysis, exponential smoothing (ES), the Box & Jenkins model and the Holt-Winters model. These approaches have been demonstrated to be useful in various contexts, but their effectiveness depends on the specific application and the forecasting horizon. (Swaminathan & Venkitasubramony, 2024)

3.1.1 Simple moving average

Simple moving average (SMA) relies on the fixed number of most recent observations, updating the forecast each period by discarding the earliest data point and incorporating the latest observation (Tliche, Taghipour & Canel-Depitre, 2019). The simplest form of SMA, where the number of observations is one, is called naïve forecasting. The naïve forecast equals to the last observed value, meaning it predicts zero change in demand from the previous period. (Beck, Dovern & Vogl, 2025) The naïve forecasting can be defined as:

$$F_t = D_{t-1} \tag{1}$$

where D_{t-1} is the actual demand at previous period and F_t is the forecasted demand for the current period. The naïve forecasting method is commonly just used as a simple benchmark due to its oversimplified essence. Although, when time series can be approximated by random walks, naïve forecasts can be difficult to outperform. (Goodwin, Petropoulos & Hyndman, 2017)

The SMA formula, when there is more than one observation included, can be expressed at time t + 1 as:

$$F_{t+1} = \frac{1}{N}\sum_{k=0}^{N-1}(D_{t-k}) = \frac{D_t + D_{t-1} + \cdots + D_{t-k}}{N} \tag{2}$$

where F_{t+1} is the forecast at time t + 1, N is the order of the SMA and D_t is the demand at time t. Simply, the SMA is defined as the average of the previous k observations. It is calculated as a simple average, meaning it attaches equal weight to all observations. SMA smooths short-term fluctuations and showcases longer-term trends by continuously updating the average as new data becomes available, making it useful for identifying general demand patterns. (Ali et al., 2017)

SMA is a simple yet widely utilized forecasting technique. It is often competing with exponential smoothing in companies. The appeal of the method lies in its simplicity which makes intuitive and easy for practioners to apply. However, the simplicity comes with notable limitations on its performance and applicability. Firstly, the number of periods N in the MA must be predetermined which is typically based on the forecaster's judgement rather than formal optimization. Secondly, SMA lacks a formal statistical foundation which makes it challenging to construct reliable prediction intervals, posing issues in practical business applications. Also, SMA has received limited attention in academic research because it does not seem interesting due to its straightforward nature. (Svetunkov & Petropoulos, 2018)

3.1.2 Simple Exponential smoothing

Exponential smoothening remains a foundational method in business forecasting since it serves as a strong benchmark despite continuous advancements in the field of forecasting. The core concept of exponential smoothing involves calculating weighted averages of past values. The weights of the past observations decrease exponentially when moving further away from the present observation, meaning the most recent values outweigh the earlier ones. The choice of appropriate exponential smoothing technique depends on the components in the time series. Simple exponential smoothening (SES) is adequate when there is no obvious trend or seasonal pattern. (Petropoulos et al., 2022) This is the simplest form of exponential smoothing and can be defined as:

$$F_{t+1} = \alpha D_t + (1 - \alpha) F_t \tag{3}$$

where α is the exponential smoothing parameter, D_t is the actual demand at time t, F_t is the forecast at time t and F_{t+1} is the forecast at time t + 1. The smoothing constant α is provided a value between 0 and 1. This smoothing parameter determines the extent to which new information influences the model's forecasts. Lower smoothing parameter values reduce the effects of new data on the forecasts which makes the model less reactive. Conversely, higher smoothing constant values make the model more responsive to recent fluctuations. (Pritularga, Svetunkov & Kourentzes, 2023)

SES provides a simple and effective methods for stationary time series. However, real-world time series often include additional components such as trends, seasonal patterns, cycles, and irregular variations. Holt-Winters model has been developed to extend the basic model to account for trend and seasonal components. This model is explored in the next subsection.

3.1.3 Holt-Winters

The Holt-Winters method, also known as triple exponential smoothing, is an extended version of single exponential smoothing tailored for time series data with trends and seasonality. It is a widely utilized technique for forecasting business data that exhibit patterns, evolving trends, and seasonal correlations. (Gelper, Fried & Croux, 2010) The Holt-Winters decomposes a time series into three elements which are level, trend and seasonality. There are two variants when modelling the seasonality factor: either additive or multiplicative form. The additive method is favoured when the seasonal variations are roughly constant through the series. The multiplicative method is utilized when the seasonal variations are changing proportionally to the level of the time series. (Hansun et al., 2019)

The additive version of the Holt-Winters method can be expressed with the following equations:

$$L_t = \alpha(D_t - S_{t-s}) + (1-\alpha)(L_t + b_{t-1}) \quad (4)$$

$$b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1} \quad (5)$$

$$S_t = \gamma(D_t - L_t) + (1-\gamma)S_{t-s} \quad (6)$$

$$F_{t+m} = L_t + b_t m + S_{t-s+m} \quad (7)$$

The multiplicative Holt-Winters equations can be expressed as:

$$L_t = \alpha\left(\frac{D_t}{S_{t-s}}\right) + (1-\alpha)(L_{t-1} + b_{t-1}) \quad (8)$$

$$b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1} \quad (9)$$

$$S_t = \gamma\left(\frac{D_t}{L_t}\right) + (1-\gamma)S_{t-s} \quad (10)$$

$$F_{t+m} = L_t + b_t m + S_{t-s+m} \quad (11)$$

where D_t is demand at time t, L_t is the level component at time t, b_t is the trend at time t, S_t is the seasonality at time t, m is the number of forecast periods ahead, s is the length of seasonality (for instance the number of days/weeks/months in a season), and F_{t+m} is the forecast for m periods ahead. The smoothing parameters α, β, and γ are typically given a value between 0 and 1 and control the degree of responsiveness of each component. The key difference in the equations is that some components are addressed with addition and subtraction in the additive method, rather than multiplication and division in the multiplicative method. (Ferbar Tratar, Strmčnik, 2016)

The Holt-Winters, as well as single exponential smoothing, is particularly sensitive to anomalies and outliers in the data. The outliers influence forecasting in two main ways. Firstly, they distort the estimated level, trend and seasonal components since both current and historical data are incorporated to the smoothed values. Secondly, outliers influence the selection of smoothing parameters which are optimized to decrease the sum of squared forecast errors. The presence of outliers in the data leads to non-robust parameter optimization, resulting in less accurate forecasts. To achieve greater forecasting accuracy, robust versions of the Holt-Winters method have been created. (Gelper, Fried & Croux, 2010) Hansun et al. (2019) further highlight challenges such as estimating the unknowns, choosing the prediction intervals and establishing correct conditions in successful implementation of the Holt-Winters model.

3.1.4 Box-Jenkins

Box-Jenkins, also known as the Autoregressive Integrated Moving Average (ARIMA), has been one of the most popular forecasting models since the 1970s (Zhu et al., 2019). ARIMA models have been considered the dominant benchmark in empirical forecasting evaluations and are extensively utilized in variety of applications including hospitality, production, healthcare and climate forecasting. (Petropoulos et al. 2014) The central assumption in the ARIMA model is that the future values of a time series data can be represented as linear combinations of past observations and associated errors (Zhu et al., 2019). ARIMA model consists of two integrated linear regressions: Autoregressive (AR) and Moving Average (MA). The AR component can be expressed as:

$$D_t = c + a_1 D_{t-1} + \cdots + a_p D_{t-p} + u_t \tag{12}$$

where D_t is the demand at time t, c is a constant, a_1,…, a_p are the AR parameters, p is the order of the AR component, and u_t is the error term. Similarly, the MA component can be expressed as:

$$D_t = \mu + u_t + m_1 u_{t-1} + \cdots + m_q u_{t-q} \tag{13}$$

Where μ is the expected value of D_t, m_1,…, m_q are the MA parameters, and q is the order of MA component. The AR and MA components can be integrated where the ARIMA(p,q) can be expressed as:

$$D_t = c + a_1 D_{t-1} + \cdots + a_p D_{t-p} + u_t + m_1 u_{t-1} + \cdots + m_q u_{t-q} \tag{14}$$

Where p and q are the autoregressive and moving average terms. The fundamental premise of the ARIMA(p,q) model is that the time series data is statistically stationary, meaning that

key statistical properties such as mean and variance remain constant. However, practical data often displays non-stationary behaviour which requires the data to be differenced to achieve stationarity. This is reflected in the ARIMA(p,d,q) model where d represents the degree of differencing applied to the data. (Al-Musaylh et al., 2018) ARIMA is a data-driven linear approach, meaning its performance can be limited when the underlying data is non-linear. Hybrid ARIMA and advanced machine learning techniques have been developed to deal with non-linear data. (Zhu et al., 2019)

3.2 Qualitative methods

Qualitative forecasting methods are especially useful when historical data is scarce or unavailable, meaning that applying quantitative methods can be challenging. Historically, demand forecasting for products with limited data such as fashion products or fiction books has heavily relied on the insights and intuition of retailers and industry experts. (Caniato, Kalchschmidt & Ronchi, 2011) Common qualitative approaches include judgement-based, survey-based and Delphi methods. There has been limited research focused solely on qualitative forecasting. However, these methods are often combined with quantitative approaches in practice to enhance accuracy of the forecasts. (Swaminathan & Venkitasubramony, 2024)

Judgemental forecasting is a commonly used decision-making approach when there is little or no data available, such as during new product launches or novel manufacturing processes. Judgement-based forecasting is typically applied in three ways:

1. Limited or no data available, meaning statistical models are unfeasible and judgemental forecasting is the only viable option

2. Statistical models and results are refined and adjusted through expert opinion

3. Judgemental and statistical forecasting are utilized separately and the outcomes from both are combined into the final forecast

The performance of judgemental forecasts is dependent of several key factors, such as accuracy of acquired and calculated data, the presence of constraints, and influence of cognitive biases. Additionally, strong domain expertise and real-time information is required

to produce high quality judgemental analyses. Although, the inherently subjective nature of judgemental forecasting poses a notable limitation for this method. (Mishra et al., 2022) Marmier & Cheikhrouhou (2010) concluded that structured approach should be applied when using judgemental forecasting. Structured approach allows the experts to identify and explain the contextual factors affecting the demand forecast adjustments. This also makes comparing different perspectives viable when experts have alternating perceptions on how a certain factor affects the forecast. (Caniato, Kalchschmidt & Ronchi, 2011) Structuring judgement was found to have a significant improvement on the accuracy of the demand forecast (Marmier & Cheikhrouhou, 2010).

The Delphi method is a qualitative forecasting technique designed to achieve a reliable consensus from a panel of experts through multiple rounds of structured questionnaires (Swaminathan & Venkitasubramony, 2024). The Delphi method provides a structured approach when dealing with limited quantitative data. The key characteristics of the method are anonymity, iteration, controlled feedback, and aggregation of group responses. Anonymity is maintained by providing a questionnaire outlining the forecasting problem to the experts and ensuring that the individual responses are confidential from other responses. This prevents social pressure and group influence altering individual participant's judgement. The individual responses are statistically analysed and summarised with measures such as mean and variance. Then these are given to the experts, and they are allowed to revise their initial estimates. If there are responses that significantly deviate from the group consensus, the participants are asked to justify their reasoning. This process is repeated for several rounds to gradually form a final consensus. Linear opinion pool is often utilized when combining individual judgements. Studies have shown that the Delphi method produces forecasts with greater accuracy and lower variance than individual human forecasts. Reasons for the increased performance are requirement to provide detailed justifications of judgements and feedback loop for reducing bias. However, the Delphi method has received criticism regarding anonymity since initial group consensus is generated after the first round. (Zellner et al., 2021)

There are also other group-based judgemental forecast methods such as face-to-face discussions with focus groups or nominal group technique (Zellner et al., 2021). Analytic Hierarchy Process (AHP) is a multiple criteria decision-making method which uses experts' relative assessment and ranking of different factors affecting demand forecast (Swaminathan

& Venkitasubramony, 2024). Survey-based approaches are not only limited to being utilized with experts. It is possible to survey the intentions and expectations of potential customers to gather insights about market size and customers' interest towards specific products. This can provide valuable data for demand forecasting. Another way to introduce customer input into qualitative demand forecasting is through simulated interaction. Engagement with target customers occurs through focus groups or beta testing which can yield qualitative insights to demand forecasts. (Hildick-Smith, 2006)

In the context of fiction book demand forecasting, expert judgement can be utilized to assess factors that influence fiction book sales. These factors include author popularity, market trends, target audience and seasonality. These factors can either be directly incorporated into fiction book demand forecasting models or utilized for book categorization. Division of fiction books into categories allows for tailored forecasting models where each category employs specific variables derived from historical data to predict fiction book demand more accurately. Target audience can also be surveyed or analysed to gain insights about their preferences and interests in upcoming titles and the overall fiction book market. Bookstore managers, sales analysts and marketing professionals can engage with readers in focus groups or beta reading to extend their knowledge about fiction books and their potential demand. (Hildick-Smith, 2006) The Delphi method may be used for structured and consensus-driven fiction book forecasting where industry professionals are surveyed to estimate the sales of specific titles and to justify reasons for their predictions. This has the potential to reveal underlying factors that affect or are perceived to affect fiction book sales. Similarly, AHP is also a suitable tool for understanding the most significant features influencing fiction demand. Qualitative methods have potential in fiction book demand forecasting because historical data is scarce due to short product lifecycles and continuous release of new titles. These approaches provide additional benefits since fiction demand can be influenced by unpredictable elements such as media attention or social trends. However, the notable limitations of these methods are high time and resource consumption, as well as susceptibility to bias which undermine their reliability and practical applicability. The qualitative methods are summarized in Table 2 below.

Table 2. Qualitative forecasting methods for fiction book demand

Method	**Principle**	**Use case**
Expert judgement-based	Expert's intuition and contextual knowledge is utilized to form forecast.	Useful when dealing with unpredictable demand, scarce data or new titles. Elements assessed include author popularity, seasonality, and trends to form predictions. Aids in categorization and recognition of affecting factors.
Delphi	Structured consensus through multiple rounds of anonymous questionnaires and feedback.	Ideal for understanding underlying factors influencing demand.
Focus groups	Direct engagement through group discussion or beta reading. Can be done only among industry professionals or with target audience.	Helps in collecting qualitative insights on fiction book appeal.
AHP	Decision-making method using expert ranking of multiple influencing factors.	Identification and weighting of key variables influencing demand, assisting in structured forecasting and categorization.
Customer survey	Gathering data through structured surveys from potential customers.	Used to recognize reader interest in titles, assess market size and gauge demand directly from target customers.
Nominal group technique	Structured group discussion where ideas are ranked anonymously.	Effective for consolidating diverse perspectives and demand-impacting factors while minimizing bias.

3.3 Combining methods

Combining methods for demand forecasting involves integrating multiple forecasting techniques to enhance accuracy and reliability. This approach leverages the strengths of individual models while mitigating their weaknesses, leading to more robust predictions. (Caniato, Kalchschmidt & Ronchi, 2011) Therefore, it may be effective to attempt to combine forecasted generated by different methods in some cases, rather than trying to select a most appropriate single method (Zhou, 2017).

Mathematical and statistical models can be effective in most contexts, but they often fall short when forecasting situations of high uncertainty, for example specific or one-time events. Therefore, adding contextual knowledge, such as insider insights about strikes, promotions or new store openings, plays a critical role in improving forecasting accuracy. (Petropoulos et al., 2022) There is an ongoing need to integrate both quantitative models and expert knowledge across organizations and supply chains to improve the quality of the forecast. Combining expert judgement and mathematical forecasting leverages the stakeholders ability to apply their domain expertise while also utilizing the analytical strengths and consistency of statistical models. Moreover, studies have shown that collaborative forecasting typically yields more accurate results than relying on input from a single source. (Cheikhrouhou et al., 2011)

Combining forecasts can also occur between quantitative methods. The M4 competition revealed that nearly all of the best performing approaches were combinations of statistical models. This finding implies that a single method cannot capture the patterns in the data sufficiently, whereas combining methods achieve more accurate results due to offsetting individual model errors through averaging. This holds true even when comparing simple combining approaches against the most suitable individual methods. Furthermore, the M4 competition highlighted that hybrid models that utilize both statistical and ML features achieve the highest accuracy. (Makridakis, Spiliotis & Assimakopoulos 2020)

There is a consensus that combining individual forecasts tend to improve forecasting performance. However, the degree of error reduction achieved through combining forecasts depends heavily on the extent to which the separate forecasts contain unique or independent information. In practice, obtaining truly independent forecasts can be difficult since many predictions are derived from similar statistical models and rely on overlapping data.

Measuring the coherence between forecasts can help to identify the degree of independence of forecasts. (Thomson et al., 2019) Another challenge of combining forecast is the increased complexity. The effort and expertise required to develop and maintain a more complex combined demand forecast may not justify the performance improvements gained.

3.4 Evaluation and selection of forecasting method

Performance measures are essential in demand forecasting because they help to evaluate the effectiveness of different forecasting methods. The most used performance assessment measures found from the literature are Mean Squared Error (MSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). (Ren, Chan & Siqin, 2020) Root Mean Squared Error (RMSE) and Symmetric Mean Absolute Percentage Error (SMAPE) are also notable demand forecasting evaluation metrics found in the literature (Chicco, Warrens & Jurman, 2021). Forecast accuracy is the primary focus in academic studies when it comes to forecasting performance. In real-world applications, there are also other performance factors to consider, such as computational speed, data requirements, stability, ease of use and industry specific constraints. (Ren, Chan & Siqin, 2020)

MAE is a widely used absolute error measure for forecast evaluation. It measures the average absolute difference between the actual and forecasted demand. (Karunasingha, 2022) MAE can mathematically be expressed as:

$$MAE = \frac{1}{N}\sum_{t=1}^{N}|D_t - F_t| \tag{15}$$

where D_t is the actual demand at time t, F_t is the forecasted demand at time t, and N is the number of observations. (Zhu et al., 2019) MAE quantifies how far the predictions are from the actual demand on average. It produces non-negative values in the same unit as the original target variable with smaller values indicating higher accuracy. While MAE reflects the magnitude of forecast errors, it does not signal whether the model systematically underestimates or overestimates demand. (Karunasingha, 2022) Punia & Shankar (2022)

noted that Mean Error (ME) can be utilized to assess forecasting bias. ME measures the average difference between the actual and forecasted values. It produces positive and negative values, where positive ME values indicate that forecasts tend to underestimate demand and negative values suggest overestimation. ME is useful for detecting bias in forecasting models and is often paired with other error metrics to provide a more complete evaluation of performance. However, positive and negative errors cancel each other out, meaning ME is not a reliable measure for overall accuracy.

MSE is another common metric used to evaluate the accuracy of forecasting methods. It especially utilized in regression and time series forecasting. MSE calculates the mean of squared deviations between the actual and predicted values. The MSE formula can be defined as:

$$MSE = \frac{1}{N}\sum_{t=1}^{N}(D_t - F_t)^2 \tag{16}$$

where D_t is the actual demand at time t, F_t is the forecasted demand at time t, and N is the number of samples. (Ferbar Tratar, Strmčnik, 2016) MSE always provides non-zero values, since the errors are squared. Squaring also penalizes larger errors more than smaller ones which makes MSE sensitive to outliers. Although lower MSE values indicate better model performance, the squared units can make interpretation less intuitive.

RMSE is a mathematically similar forecasting accuracy metric to MSE. In essence, RMSE is the square root of MSE. RMSE can be defined as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(D_t - F_t)^2} \tag{17}$$

where D_t is the actual demand at time t, F_t is the forecasted demand at time t, and N is the number of samples. RMSE has similar prosperities as MSE since the results are always non-

zero and lower values showcase better performance. It also penalizes larger errors but less severely than MSE due to the square root. RMSE yields the results in the same unit as the target unit, making the results more interpretable. (Hodson, 2022)

MAPE is a relative statistical forecasting accuracy method. It is a scale independent measure used to assess the accuracy of a forecasting model by expressing the result as a percentage. MAPE calculates the average absolute error between the predicted and actual values and compares the absolute error to the absolute actual value. The mathematical formula can be expressed as:

$$MAPE = 100\% \frac{1}{N} \sum_{t=1}^{N} \left| \frac{D_t - F_t}{D_t} \right| \tag{18}$$

where D_t is the actual demand at time t, F_t is the forecasted value at time t, and N is the number of samples. Scale independent properties of the MAPE measure makes it an applicable comparison method across different datasets and units of measure. The lower percentages demonstrate more accurate forecasts. The results are easy to interpret and useful when focus is on the relative size of errors. (Ferbar Tratar, Strmčnik, 2016) However, MAPE also has some limitations especially when the values approach zero. In the case of zero demand, it produces undefined values due to division by zero. MAPE can become skewed if the values are close to zero because small absolute errors appear disproportionately large. MAPE also penalizes absolute errors based on the actual demand at a specific time. (Chicco, Warrens & Jurman, 2021) For example, if the absolute error is exactly the same in two different cases, MAPE is producing a larger error when the corresponding actual demand is lower, and a smaller error when the corresponding actual demand is higher (Ferbar Tratar, Strmčnik, 2016).

SMAPE is also a scale independent metric used to evaluate demand forecasting performance. It is a symmetric performance measure, meaning it treats over-predictions and under-predictions equally. (Wallström & Segerstedt, 2010) SMAPE compares actual and predicted demand, and the error is declared as a percentage. The formula can be defined as:

$$SMAPE = 100\% \frac{1}{N} \sum_{t=1}^{N} \frac{|D_t - F_t|}{\frac{|D_t| + |F_t|}{2}} \quad (19)$$

where D_t is the actual demand at time t, F_t is the forecasted value at time t, and N is the number of observations. SMAPE shares similar features to MAPE since the error is represented as a percentage and lower values indicate a better performing model, making the interpretation of the results relatively simple. One key difference is the error range which is capped at 200% in SMAPE compared to the unlimited maximum value in MAPE. SMAPE was developed to address the disadvantages of MAPE. SMAPE is less sensitive to lower values, although if both the actual demand and forecasted demand are near zero, it can become unstable. SMAPE yields balanced penalization for both over and under predictions while MAPE penalizes under forecasting more. (Chicco, Warrens & Jurman, 2021)

These error metrics attempt to summarize the characteristics of errors into a single value, facilitating model comparison. Generally, the forecasting method that demonstrates the lowest error should be selected. (Karunasingha, 2022) However, selecting an appropriate forecasting method can yield to less effective outcomes if the decision is based on only one error measure. Therefore, using multiple error measures is suggested, although these metrics can produce conflicting results. A model that performs well based on one error metric can be outperformed by another method when using a different error measure, making the selection less straightforward. (Badulescu, Hameri & Cheikhrouhou 2021) In practice, organizations should not only focus on statistical accuracy to gain the full benefits of forecasting techniques. Strategic aspects, such as adaptability to market changes, implementation and maintenance costs, ease of integration into existing systems and scalability, are equally important to consider when selecting a forecasting method. (Chandran & Khan, 2024)

4 Empirical case study

This section presents the empirical case study which evaluates different forecasting methods for fiction book demand. It begins by introducing the case company and describing the data as well as forecasting methods used in the analysis. Finally, the section showcases and discusses the results of the applied forecasting methods.

4.1 Introduction of the case company

The case company is a multidisciplinary media group in Finland with a broad influence on Finnish society. Their business areas include Books, Learning Services and Media. The Books business area consists of the nationwide bookstores which serve customers in 56 retail locations around Finland, and they are also available online. Under the Books business area is the publishing company which publishes fiction, nonfiction and educational materials, as well as the book printing company responsible for book manufacturing to customer demand. Learning services produces digital and print content as well as services ranging from pre-school to adult education. The Media division publishes popular magazines and online services for individual customers and provides advertising, event and content solutions for business customers. The corporate structure is visualised in Figure 3 below.

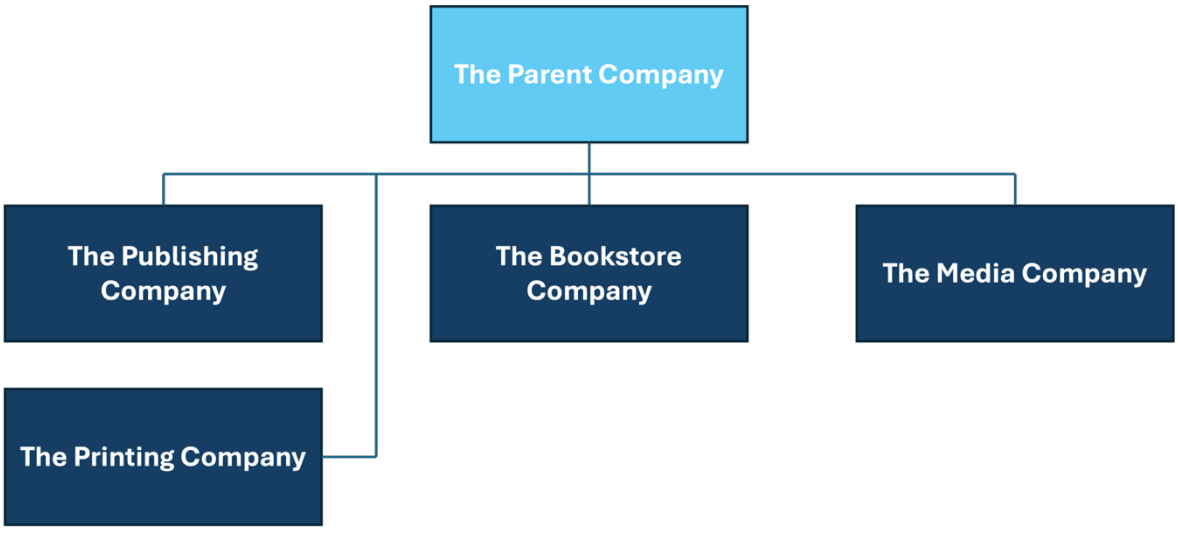


Figure 3. The case company corporate structure

The case company is in a unique position regarding its supply chain since they own and have control over the whole supply chain from publishing decisions to book retail. The publishing company negotiates publishing deals with authors and other parties producing literature material. After that the publishing company makes the decision on which material is being published and pushed into manufacturing. When the publishing decisions are made, the printing company manufactures the physical books. The printing company is in charge of the initial printing quantity which is based on presales figures and past experience with similar books. Presales begin approximately six months before the public release date. During this period, retail stores within the bookstore chain and external vendors place orders. This means that the bookstores in the supply chain and outside vendors purchase the material up to six months in advance, although the books are delivered to the physical locations close to the actual release date of the book. This allows enough time for the printing company to schedule the book manufacturing and plan other supply chain operations. After release, subsequent printing decisions are based on the orders coming from the retail locations and outside vendors. The average lead time for post-release orders is two weeks. When the books have been manufactured, they are distributed to the physical retail locations where individual customers can purchase them. The books are also purchasable online but online or eBook sales channels are out of scope for this thesis. The retail stores in the bookstore company track the daily sales and inventory levels for each individual title.

The company's end-to-end control over the entire supply chain from publishing decisions to retail provides opportunities for utilizing data across multiple functions. This serves as the main motivation for commissioning the thesis. The company is interested in utilizing the point of sales data from the retail locations to reduce the bullwhip effect, enhance manufacturing efficiency and improve predictive capabilities. This case study focuses on forecasting the demand for individual fiction titles sold in the 56 physical retail locations. The case study is limited to only fiction books in the retail locations. The study aims to find an effective method for forecasting individual fiction book demand and explore how demand forecasts can inform more predictive and data-driven decision-making. Currently, the company does not have an established process for forecasting individual fiction titles, meaning no prior benchmark exists. The company expects improvements especially in optimizing subsequent print quantities, inventory management and logistics planning.

4.2 Data collection and method selection

The data selected for the empirical case study consists of fiction book sales data from the past three years. The dataset is divided into three individual subsets: fiction book sales in 2022, 2023, and 2024. This separation is based on the short life cycle of fiction titles, as majority of fiction books remain relevant for only a year at most. Given this one-year lifecycle, each dataset contains unique, newly published titles. Each fiction books has an attribute referred to as its year of novelty, which indicates the year in which the title is considered new. Each fiction book receives a year of novelty based on its release date. Titles released in November or December are assigned the following year as their year of novelty. Consequently, the 2023 dataset includes sales data for fiction books released between November 2022 and October 2023.

Each dataset contains POS data from physical bookstores, covering daily sales since release. The POS data is aggregated from all retail locations automatically by the system at the individual title level. Therefore, each row in the dataset represents daily sales across all retail stores for a specific fiction book. An example of the data rows is showcased in Table 3 below.

Table 3. Example of the sales data rows

Date	**The year of novelty**	**Release date**	**Title**	**Author**	**Sales**
1.5.2023	2023	20.4.2023	Title A	Writer A	10
1.5.2023	2023	25.3.2023	Title B	Writer B	7
1.5.2023	2023	12.12.2022	Title C	Writer C	3
2.5.2023	2023	20.4.2023	Title A	Writer A	18
2.5.2023	2023	25.3.2023	Title B	Writer B	5
2.5.2023	2023	12.12.2022	Title C	Writer C	1

Exploratory data analysis was conducted to gain insights into the datasets, identify trends and patterns, detect anomalies and develop initial hypothesis. This assisted in finding outliers which were primarily related to titles with exceptionally high single-day sales, generally resulting from external factors such as awards or “Book of the month” selections. Outliers were identified by capping the daily sales at 200 units. Titles that exhibited higher daily sales were removed from the datasets. These books were mainly “Book of the month” selections which demonstrated abnormally inflated daily sales and returns that did not represent the demand patterns of other releases. Therefore, these books had to be removed to ensure proper forecasting model evaluation and selection. It was also found that the sales data from the first quarter of 2022 was missing, since all titles exhibited zero sales during this period. Visualising the data showcased clear seasonal trends, including the increase of sales during Christmas period, Mother’s Day and Father’s Day. Another key finding was the importance of the first three to four months after a book’s release, during which most of the total sales occur. The sales graphs also highlight the short lifecycle of fiction titles, as sales figures are depleted in the subsequent years. These findings are illustrated in the figures and discussed in more detail below.

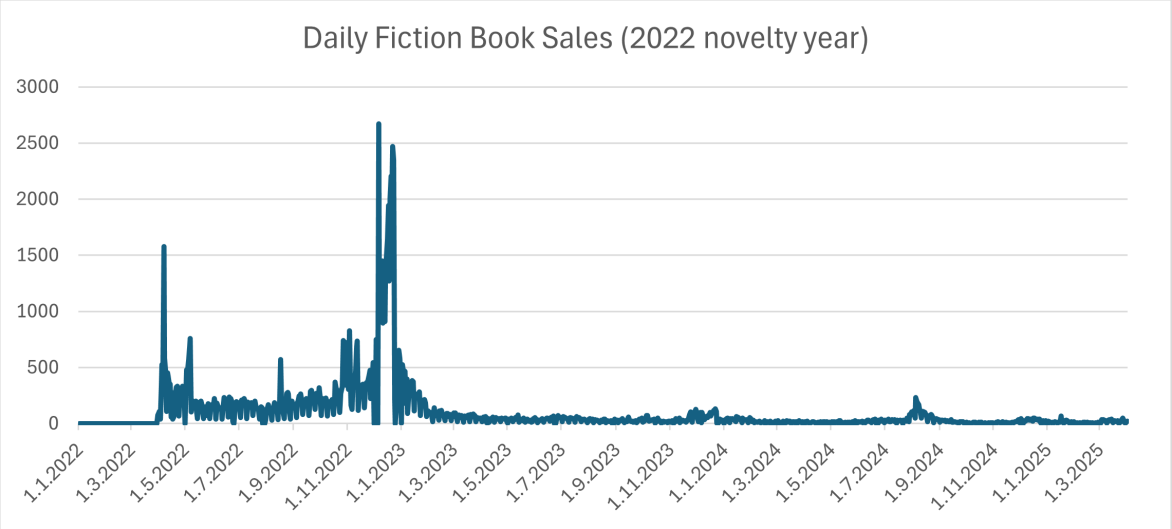


Figure 4. Daily sales for 2022 novelty year fiction books

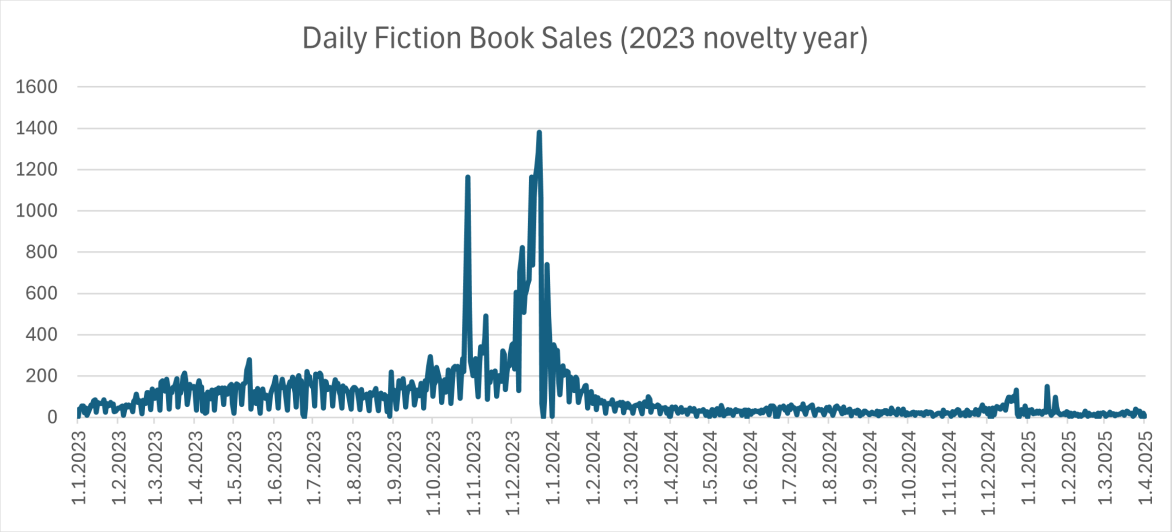


Figure 5. Daily sales for 2023 novelty year fiction books

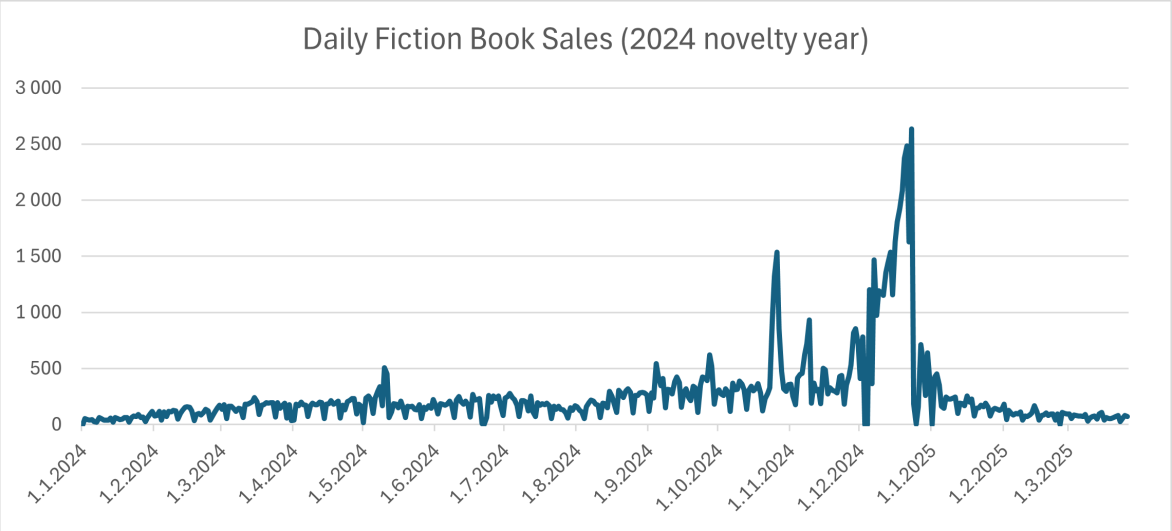


Figure 6. Daily sales for 2024 novelty year fiction books

Figures 4, 5 and 6 above reflect the daily fiction book sales at retail locations for each novelty year. All graphs illustrate a consistent pattern across years. The baseline sales remain relatively low for the first three quarters of the calendar year apart from occasional spikes. The spike in the beginning of May can be explained by gift purchases due to Mother's Day. During the last quarter of the year, the sales begin rapidly increasing towards end of the year. The most significant spike at the end of the year is driven by Christmas sales. The smaller increase in sales at the beginning of November aligns with Father's Day. After the end of

year holiday period, there is a remarkable drop in sales. The sales drop sharply and remain at a steady and low level through the following periods. The main difference in these charts is the peak quantity in the 2023 dataset. The 2023 dataset has a peak quantity around 1500 where the other years have a peak quantity around 2500. More outliers and cleanable data were found in the 2023 dataset, which explains the noticeably lower quantity. The 2022 dataset lacks sales data for the first quarter, as can be seen form Figure 4, which may account for the irregular initial spike at the start of the second quarter.

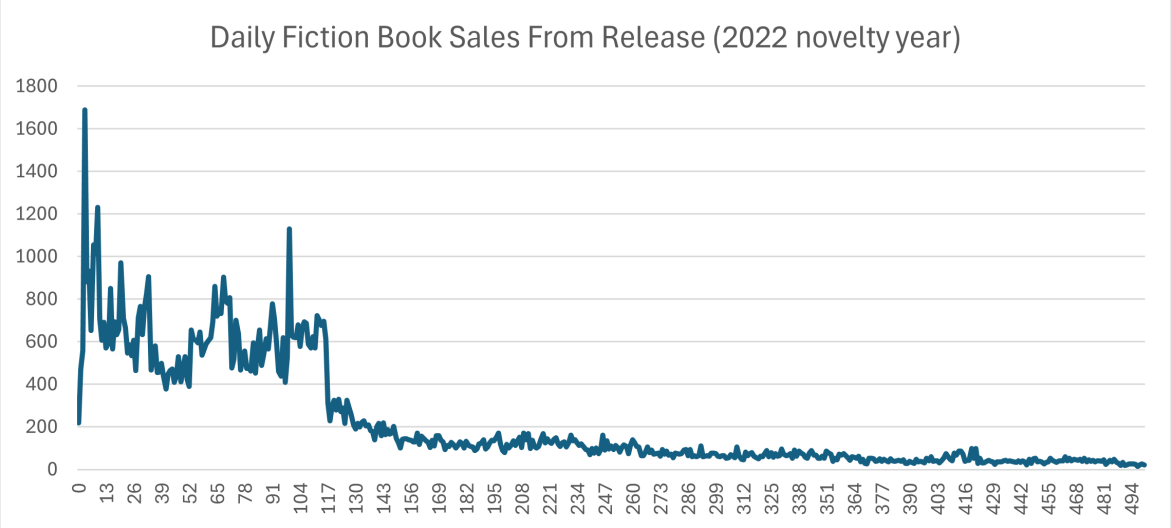


Figure 7. Daily sales for 2022 novelty year fiction books from release

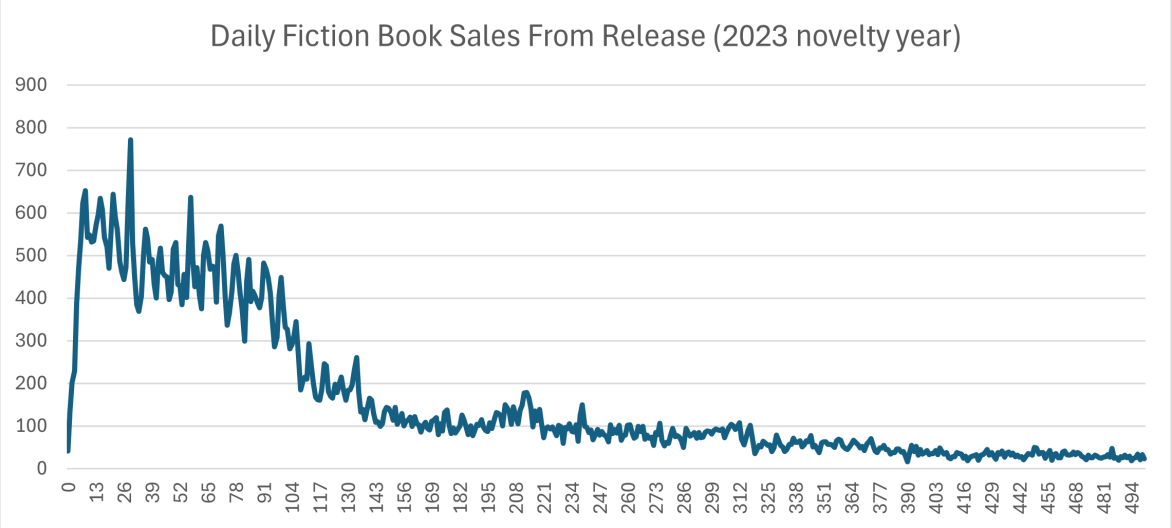


Figure 8. Daily sales for 2023 novelty year fiction books from release

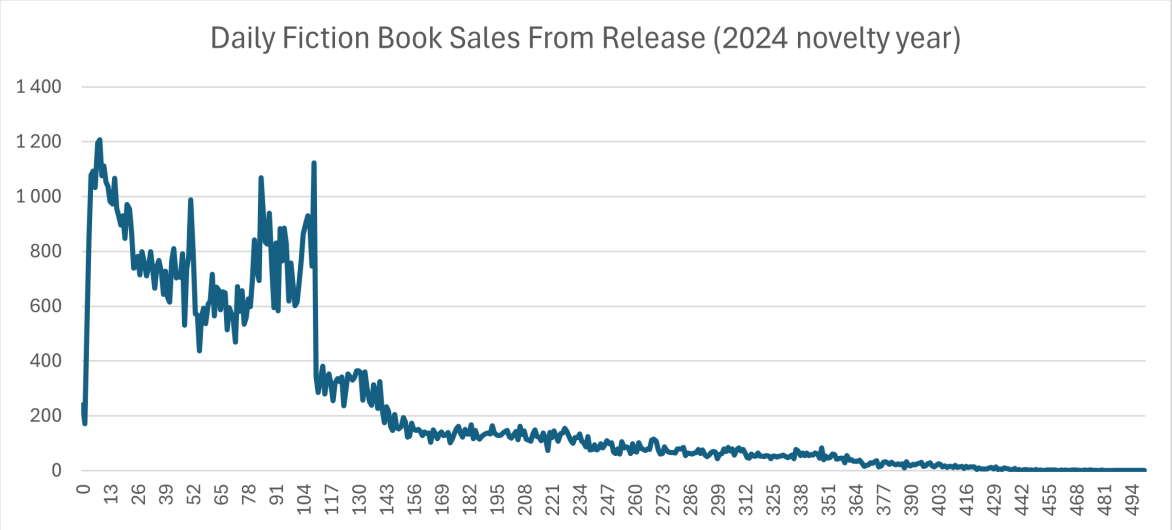


Figure 9. Daily sales for 2024 novelty year fiction books from release

Figures 7, 8 and 9 show the daily sales for fiction books from their release date. The x-axis displays days since release, allowing comparison of sales patterns independent of the calendar date. All figures demonstrate a similar pattern where the high initial demand following release declines over time. The elevated sales volumes persist approximately for the first four months where most of the sales occur. A sharp drop in sales is noticed around 120 days post-release in the 2022 and 2024 datasets. After this period, the sales volume is considerably lower and decreasing slightly and steadily over time. The declining is prominently smoother in the 2023 graph since the dataset underwent more extensive data cleaning. The lower initial sales immediately following release, suggest potential challenges in book deliveries to retail locations or in marketing communication reaching customers.

The forecasting in this thesis focuses on next-period monthly forecasting, attempting to find the best method for predicting next-month demand using data from previous months. A monthly forecasting horizon was selected because the printing quantities of books are planned to cover the demand over multiple months. Therefore, shorter forecast intervals would be too granular for production planning and longer time horizons would not be practical due to the short lifecycle of fiction books. Shorter forecast intervals may be useful in other contexts, but reliable forecasts are difficult to generate due to the high short term demand volatility as seen from the previous figures. Additionally, projections of each title's total demand were generated to support in predictive book printing decisions, help to quantify the total sales and reduce stockouts.

The forecasting is conducted at the SKU level which means that forecasts are generated for every single book. This is necessary since book printing decisions are made separately for individual titles. The daily sales of individual fiction books were aggregated to a monthly level to represent monthly demand since Kourentzes, Rostami-Tabar & Barrow (2017) indicated that the level of data utilized should match the level of generated forecasts. The aggregation of monthly demand was conducted in two different ways for comparison. The first aggregation method combined the sales based on calendar month. The demand for the individual books was calculated for all calendar months from January to December. The second method aggregated the sales based on the release date of the book. From the release date of the book, the sales data was aggregated in 30-day periods to indicate monthly demand. The months were numbered from 0 to 11, where month 0 consisted of sales from days 0 to 29, month 1 consisted of sales from days 30 to 59, etc. The hypothesis was that the release date based aggregation generates more accurate forecasts for initial months after release, while the calendar based aggregation performs better in forecasting Christmas demand. Initial forecasting and extensive analysis highlighted the varying demand patterns of the individual books based on their release month. It was found that books released in the spring, had significantly different demand patterns than books released in the autumn. Consequently, books within the datasets were grouped by release month. The forecasting results are reported according to this monthly categorization for each year.

Due to the missing data in the 2022 dataset and the absence of November releases in the 2023 dataset, no assessment was conducted for November releases. No normalization steps were applied in attempt to compensate for these gaps. Regenerating the sales figures for November titles in the 2022 dataset would have required assumptions which could lead to inaccurate results. Additionally, the 2024 November releases did not have a full year of sales at the time of the analysis, meaning the November results would rely on two incomplete years of data. These data gaps have a minimal impact on cross-year comparison since the main purpose of the analysis was to compare the performance of various forecasting methods rather than individual months across years. Therefore, the exclusion of November releases does not undermine the validity of the findings.

There is only twelve months of available data for individual titles. Due to the scarcity of demand data, more complex or data-intensive forecasting methods could not be applied. Such models require longer historical time series than the fiction book context provides.

Therefore, the study employs naïve forecasting, SMA, SES and a self-developed coefficient-based method for monthly forecasting. The performance is measured using ME, MAE and MSE to gain a comprehensive view of forecasting bias, error magnitude and error variance. Percentage-based performance measurements were excluded since monthly sales volumes for individual books can be quite low, leading to disproportionately large percentual errors and less interpretable results. Finally, a projection model is introduced and used to estimate the total demand for each title. Percentage-based error metric was applied for this method as total demand values are larger and thus less sensitive to extreme percentage deviations.

4.3 Performance measurement of selected forecasting methods

Forecasting performance of naïve, SMA, SES and coefficient methods are evaluated on a yearly and release month basis. The error metrics used for the evaluation are MAE, ME and MSE which measure the accuracy, bias and volatility in forecasting. Lastly, the projection model is used to estimate the total demand for the first year for each individual title.

4.3.1 Naïve forecasting performance

Naïve forecasts were generated for all books in the 2022, 2023 and 2024 datasets. The forecasts were conducted on a monthly level where the next month's demand forecast was the sales of the previous month. The books have been grouped by their release month, and the forecasts were generated for both calendar month and release date aggregation methods. The performance of the naïve forecasting method is reported both on a yearly level and for the monthly release groups. The performance of the naïve forecasting method can be found on the tables below.

Table 4. Naive forecasting yearly performance, release date based aggregation

	MAE	**ME**	**MSE**
2022	30	-9	8735
2023	23	-7	3097
2024	38	-16	7690

Table 5. Naive forecasting yearly performance, calendar based aggregation

	MAE	**ME**	**MSE**
2022	35	-4	10913
2023	30	-5	5899
2024	45	-9	10533

Tables 4 and 5 showcase the performance of the naïve forecasting method for each of the three years. The average MAE, ME and MSE have been calculated for each year to evaluate forecasting accuracy. The release date based aggregation method produces more accurate results for each year, when comparing MAE and MSE. On the other hand, ME indicates better performance for the calendar based aggregation. Therefore, forecasting errors balance each other more when data is aggregated on a calendar month basis. Across all datasets, ME only has negative values, suggesting consistent over forecasting. This result was expected as fiction sales typically decline over time, meaning naïve forecasting should systematically produce higher forecasts than the actual value.

Table 6. Naive forecasting performance for the monthly releases, release date based aggregation

	MAE	**ME**	**MSE**
January	14	-3	661
February	17	-6	721
March	18	-8	1345
April	29	-14	2147
May	23	-6	1808
June	23	-8	1711
July	16	-3	1055
August	49	-18	23381
September	64	-16	21666
October	59	-25	14025
November	-	-	-
December	15	-4	847

Table 7. Naive forecasting performance for the monthly releases, calendar based aggregation

	MAE	**ME**	**MSE**
January	15	0	868
February	24	-1	1255
March	20	-3	2013
April	34	-7	2869
May	26	-2	2589
June	25	-5	2207
July	19	-2	1176
August	54	-11	24644
September	80	-8	33171
October	80	-16	25404
November	-	-	-
December	17	-12	948

In tables 6 and 7, the books have been grouped by their release month. The average MAE, ME and MSE have been calculated for each monthly release group. The results are similar to the yearly comparison where the release date based aggregation consistently showcases better performance according to MAE and MSE. Also, the forecasting errors tend to have a greater balancing effect when using the calendar based aggregation, as the ME values are closer to zero. The ME also reveals that naïve forecasting generates higher predictions than the actual value for nearly every monthly category. The monthly grouping reveals higher predictions errors for books released in August, September and October. This is explained by the increased quantity of books released in these months, which includes a bigger proportion of best-selling titles. Larger sales volumes of these titles amplify the error metric values for the aforementioned release months.

4.3.2 Simple moving average forecasting performance

SMA forecasting was conducted for all fiction books in the 2022, 2023 and 2024 datasets at the SKU level. Forecasts were generated for each month of a book's first year after release. Averaging windows of two and three months were used for comparison, referred to as SMA2 and SMA3. Longer averaging periods haven't been applied due to the general downward trend in fiction book sales after release, since it would increase in over forecasting. SMA2 forecasts are generated by calculating the average of the previous two months' sales, whereas SMA3 uses the preceding three months' average. The first month's forecast is a naïve forecast because there is only one month of available data at that point. The second month's forecast in SMA3 equals SMA2 since only two months of historical data exists. The performance of the SMA2 and SMA3 forecasts have been presented on a yearly basis and for the monthly release groups for both release date based and calendar based aggregation methods. The results of SMA forecasts can be found on the tables below.

Table 8. SMA forecasting yearly performance, release date based aggregation

	SMA2			**SMA3**		
	MAE	**ME**	**MSE**	**MAE**	**ME**	**MSE**
2022	35	-12	10350	39	-15	10494
2023	27	-11	4340	31	-15	4768
2024	45	-23	8460	51	-30	9089

Table 9. SMA forecasting yearly performance, calendar based aggregation

	SMA2			**SMA3**		
	MAE	**ME**	**MSE**	**MAE**	**ME**	**MSE**
2022	39	-6	11900	42	-9	12031
2023	31	-6	6626	34	-9	6622
2024	49	-13	10917	54	-17	11038

Tables 8 and 9 present the annual results of the SMA2 and SMA3 forecasting methods. Across all years in both tables, the results indicate that SMA2 generally produces lower forecasting errors than SMA3. The results suggest that SMA2 is more suitable for fiction book demand forecasting across all the compared error metrics. This implies that a shorter averaging window is better at capturing the demand patterns of fiction books which was expected, since the demand is volatile and fluctuates seasonally. Shorter windows react more quickly to rapid changes in demand, whereas longer averaging windows smooth demand changes too heavily and lag behind actual sales patterns. The negative ME values indicate that both SMA methods overestimate demand. The bias is particularly evident in 2024 but MAE and MSE values are also elevated. SMA forecasts perform best for books released in 2023 across all error metrics. The release date based aggregation produces lower MAE and MSE values. The calendar based aggregation results in better ME values, implying less bias but slightly worse overall accuracy.

Table 10. SMA forecasting performance for the monthly releases, release date based aggregation

	SMA2			**SMA3**		
	MAE	**ME**	**MSE**	**MAE**	**ME**	**MSE**
January	16	-4	824	18	-6	953
February	19	-9	2462	22	-10	3147
March	21	-11	1568	23	-13	1888
April	36	-20	4615	42	-26	5453
May	26	-9	2723	29	-12	2816
June	28	-13	2054	31	-16	2261
July	17	-5	1180	18	-6	1171
August	57	-27	23980	63	-34	24180
September	72	-26	25004	80	-35	25233
October	71	-38	17047	83	-50	18348
November	-	-	-	-	-	-
December	18	-2	998	21	-4	1198

Table 11. SMA forecasting performance for the monthly releases, calendar based aggregation

	SMA2			SMA3		
	MAE	**ME**	**MSE**	**MAE**	**ME**	**MSE**
January	17	-2	960	19	-3	1049
February	23	-1	5272	25	-2	5520
March	23	-4	2120	23	-4	2129
April	39	-10	5895	43	-14	6480
May	29	-3	3602	31	-5	3499
June	28	-7	2170	31	-10	2202
July	19	-3	1275	19	-4	1232
August	60	-16	23586	63	-21	23477
September	85	-13	32694	94	-18	33375
October	92	-23	26366	100	-32	25752
November	-	-	-	-	-	-
December	15	-8	804	18	-10	884

Tables 10 and 11 display the SMA forecasting performance for the monthly groups. The results are consistent with the yearly comparison. The SMA2 method commonly outperforms the SMA3 method among all error metrics. The release date based aggregation tends to provide more accurate forecasting, except in December. All ME values are negative, confirming systematic over-forecasting of demand. The calendar based aggregation causes less forecasting bias which can be noted when comparing the ME values. Certain months have a particularly poor performance for both forecasting and aggregation methods. Books released in August, September, and October showcase notably larger errors than books released in other months. This is caused by higher sales volumes due increased amount of releases during these months.

4.3.3 Simple exponential smoothing forecasting performance

SES forecasts were generated for all fiction titles in the datasets. SES forecasting was conducted on a monthly level to predict next month's demand. The smoothing constant α values of 0,4; 0,6 and 0,8 were used for comparison. Lower α values would generate excessive over-forecasting and higher α values would lead to resembling naïve forecasting. Solving the optimal value for α was not feasible since forecasting was performed separately for hundreds of titles across multiple datasets. Automated optimization methods could theoretically be applied, but they were not suitable in this context. Automated α optimization across hundreds of titles would increase the computational complexity which was not in the interest of the case company. Also, optimizing α for each title would reduce the comparability of the results. Therefore, the SES forecasting was produced using several smoothing constant values to evaluate model responsiveness. The first demand forecast generated is the first month's actual sales since there is only one month of available data at that point. The performance of SES forecasting can be found on the tables below.

Table 12. SES forecasting yearly performance, release date based aggregation

	α = 0,4			**α = 0,6**			**α = 0,8**		
	MAE	**ME**	**MSE**	**MAE**	**ME**	**MSE**	**MAE**	**ME**	**MSE**
2022	40	-19	9830	37	-13	9273	32	-10	9118
2023	32	-18	4652	28	-12	4065	25	-9	3979
2024	52	-33	9083	45	-25	7998	41	-20	7863

Table 13. SES forecasting yearly performance, calendar based aggregation

	α = 0,4			**α = 0,6**			**α = 0,8**		
	MAE	**ME**	**MSE**	**MAE**	**ME**	**MSE**	**MAE**	**ME**	**MSE**
2022	42	-10	10986	41	-7	10870	36	-5	11429
2023	34	-9	6090	31	-7	6193	30	-5	6673
2024	55	-18	11229	52	-18	11217	50	-18	12611

Tables 12 and 13 summarize the performance of the SES models with smoothing constant α values of 0,4; 0,6 and 0,8. The results in both tables demonstrate that higher α values generally improve forecasting accuracy. When considering MAE and ME, the highest α value of 0,8 produces lowest errors in all instances. The MSE values tend to decrease as the value of α is increased when release date based aggregation is used. However, when calendar based aggregation is used, the α value of 0,8 generates the highest MSE values, suggesting larger individual forecasting errors even though the average monthly forecast improves. The ME values demonstrate similar findings with previous methods since the SES models have a bias towards over-forecasting and calendar based aggregation produces less bias. Overall, the results indicate that a higher weighting of recent observations generates more accurate forecasts. Higher values of α allow the model to adapt more effectively to the short term fluctuations that are imminent to fiction books. Thus, more reactive models generate more precise forecasts and reduce systematic over-forecasting.

Table 14. SES forecasting performance for the monthly releases, release date based aggregation

	α = 0,4			**α = 0,6**			**α = 0,8**		
	MAE	**ME**	**MSE**	**MAE**	**ME**	**MSE**	**MAE**	**ME**	**MSE**
January	19	-8	917	16	-5	751	15	-3	682
February	24	-14	3581	20	-9	2444	18	-7	1937
March	25	-17	2054	21	-11	1528	19	-9	1350
April	43	-31	5645	36	-22	4341	32	-17	3808
May	28	-14	2514	26	-10	2443	24	-8	2521
June	32	-19	2228	27	-14	1865	25	-11	1731
July	18	-7	1040	17	-5	1001	16	-4	1015
August	65	-40	23148	65	-29	21615	51	-22	21906
September	81	-35	23757	73	-28	23071	67	-22	25055
October	90	-57	17671	74	-41	15905	65	-32	15823
November	-	-	-	-	-	-	-	-	-
December	22	-6	1170	18	-3	932	16	-2	845

Table 15. SES forecasting performance for the monthly releases, calendar based aggregation

	α = 0,4			**α = 0,6**			**α = 0,8**		
	MAE	**ME**	**MSE**	**MAE**	**ME**	**MSE**	**MAE**	**ME**	**MSE**
January	19	-4	944	17	-2	857	15	-1	839
February	24	-2	4883	23	-1	5024	23	-1	5436
March	23	-5	1953	22	-3	1862	21	-3	1900
April	42	-14	6037	37	-11	5394	35	-8	5232
May	29	-4	3162	28	-3	3237	27	-3	3442
June	29	-10	2000	27	-8	1953	26	-6	2031
July	19	-4	1101	18	-3	1079	18	-2	1113
August	65	-22	22336	68	-17	21610	56	-13	22474
September	91	-11	30084	86	-11	31415	82	-10	34281
October	113	-44	27324	103	-42	27278	96	-44	31317
November	-	-	-	-	-	-	-	-	-
December	16	-10	586	15	-9	603	15	-8	718

The SES forecasting performance for the monthly release groups can be found on tables 14 and 15. The results demonstrate that a higher α values typically yield lower error values across most monthly groups when comparing MAE and ME values. However, the values of MSE are less consistent and interpretable. MSE values were occasionally increasing with higher α values. For instance, the highest MSE values were generated for the fiction titles released in September when α is given the value of 0,8. This indicates that highly reactive models may over-adjust at times. The forecasting bias is consistent with the yearly performance as all the ME values are negative in both tables and calendar based aggregation creates less biased forecasting. Also, larger forecasting errors can be found on the books released in August, September and October due to higher sales volumes linked to titles released during these months. Although the results for the monthly release groups are less interpretable, the findings indicate that the α value of 0,8 still provides the best overall SES forecasting performance.

4.3.4 Coefficient forecasting performance

Demand forecasts were first generated using the naïve, SMA and SES methods. It was found that the naïve forecasting method provided the best overall performance of the compared methods. This finding is further supported by SMA and SES evaluations, as SMA2 outperformed SMA3 and SES forecasts with higher α values produced less errors. Although the naïve method generated the most accurate demand forecasts, forecasting accuracy can be improved because the model does not react to seasonal volatility and showcased consistent over-forecasting. To overcome these issues, a coefficient-based demand forecasting model was developed.

The coefficient model is based on naïve forecasting but incorporates monthly adjustment multipliers assigned that reflect typical changes in fiction demand. The multipliers for each month were calculated by dividing the current month's demand by the preceding month's demand. The ratio captures how much the demand generally increases or decreases from month to month. This ratio was calculated for every book across all monthly release groups. After this, the average of the corresponding multipliers within the monthly release groups were calculated to obtain the final multipliers used in the forecasting. As a result, each month within each monthly release group is assigned a multiplier that represents the typical relative demand change from previous month. The coefficient model forecasts next month's demand by multiplying the most recent month's sales by the multiplier associated with the upcoming month. This approach should respond better to the empirically observed trend and seasonality changes in fiction book demand. Forecasts were generated at a monthly level for both aggregation methods. The results are presented in a similar format than previously on the tables below.

Table 16. Coefficient forecasting yearly performance, release date based aggregation

	MAE	**ME**	**MSE**
2022	24	-1	6320
2023	16	0	2063
2024	29	-6	4092

Table 17. Coefficient forecasting yearly performance, calendar date based aggregation

	MAE	**ME**	**MSE**
2022	20	1	3954
2023	16	0	1662
2024	28	-6	2936

The yearly performance of the coefficient demand forecasting model can be found on tables 16 and 17. The overall performance of the coefficient method exceeds the previously applied methods. All the error metrics showcase that the coefficient method generates more accurate forecasts. Lower MAE values indicate better monthly forecasting, lower MSE values demonstrate reduced variance and ME values close to zero suggest less forecasting bias. The issue of over-forecasting was addressed since the ME values are near zero for books released in 2022 and 2023. However, this issue remains for the books released in 2024 as the ME value in both tables is -6. When comparing the aggregation methods, calendar based aggregation clearly provides more accurate forecasts. This is especially evident when comparing the MSE, indicating less prediction variance and higher degree of stability.

Table 18. Coefficient model performance for the monthly releases, release date based aggregation

	MAE	**ME**	**MSE**
January	11	-1	489
February	10	1	637
March	12	0	956
April	17	0	1420
May	18	-3	2190
June	17	-1	1483
July	15	-4	1463
August	48	-11	18703
September	54	-8	11838
October	29	-1	5368
November	-	-	-
December	10	0	318

Table 19. Coefficient forecasting performance for the monthly releases, calendar based aggregation

	MAE	**ME**	**MSE**
January	11	1	645
February	17	0	1076
March	14	2	1362
April	27	-3	3441
May	18	0	1982
June	15	-1	911
July	12	-1	642
August	31	-3	8808
September	43	-6	6563
October	31	-2	5041
November	-	-	-
December	9	1	124

Tables 18 and 19 exhibit the forecasting accuracy for the coefficient method for the monthly release groups. These tables also demonstrate greater performance of the coefficient method when compared to the other methods. The coefficient forecasting has corrected the over-forecasting bias as the ME values for some monthly groups are zero or even positive. Consistent with previous findings, forecasting errors are the smallest for the books released in the beginning of the year and the greatest for the books released in August, September and October. The forecasting results between the aggregation methods are mixed. The release date based aggregation generates more accurate forecasts for titles released between January and April, but the calendar based aggregation generates better forecasts for the remaining releases. Overall, the coefficient model with calendar based aggregation provides the most accurate results across both yearly and monthly analyses.

4.3.5 Projection model performance

The monthly forecasts were generated, and their accuracy was evaluated in the previous subsections. On top of the monthly forecasts, total demand projections were produced to estimate each book's cumulative first year sales. To generate the total first year demand for each book, the average cumulative sales percentage was calculated at the end of every month for the monthly release groups. The cumulative percentage represents the proportion of total sales achieved at the end of each month. The projected total demand for each book was calculated by dividing the current cumulative sales by the average cumulative sales percentage corresponding to the current month. The total demand for each book was projected at the end of each month, and the projection error was calculated as a percentage deviation between the projected demand and the actual total sales. The results of the total demand forecasts can be found on the tables below.

Table 20. Total demand forecasting performance at the end of each month

Months from release	**Projection Error**
0	35 %
1	27 %
2	23 %
3	19 %
4	15 %
5	14 %
6	11 %
7	9 %
8	7 %
9	4 %
10	2 %
11	0 %

Table 20 represent the total demand forecasting errors at the end of each month. The projection error indicates the percentage deviation between the forecasted and the actual

annual demand. The data showcases a clear and consistent downward trend in projection error, reflecting that the accuracy of demand estimates improve as additional sales data is available. In the first months, projection errors remain relatively high, but the percentage errors decrease noticeably in the middle of the year. Therefore, early forecasts provide directional insights, while latter stage projections can be used to support accurate decision-making regarding inventory planning and book printing.

Table 21. Total demand forecasting performance at the end of each month for the monthly releases

		The Forecasting Error at the End of Month											
		Jan	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Jul**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**
The Release Month of the Book	**January**	35%	26%	19%	15%	13%	12%	10%	9%	8%	6%	5%	0%
	February	0%	61%	26%	23%	22%	19%	18%	16%	15%	14%	11%	5%
	March	1%	0%	58%	29%	21%	19%	16%	14%	13%	12%	11%	6%
	April	3%	1%	0%	65%	29%	23%	21%	18%	17%	14%	13%	10%
	May	4%	3%	1%	0%	62%	26%	24%	21%	21%	18%	16%	12%
	June	7%	5%	3%	2%	0%	54%	38%	34%	30%	25%	21%	14%
	July	20%	6%	5%	4%	2%	0%	49%	29%	21%	18%	16%	15%
	August	7%	5%	4%	3%	3%	2%	0%	74%	41%	33%	22%	15%
	September	8%	7%	5%	4%	3%	2%	1%	0%	58%	34%	23%	15%
	October	5%	4%	3%	2%	2%	1%	1%	0%	0%	35%	25%	11%
	November	-	-	-	-	-	-	-	-	-	-	-	-
	December	18%	14%	13%	13%	10%	9%	8%	8%	6%	3%	0%	56%

The total demand projection error at the end of each month for the monthly release groups is presented in Table 21. Each row represents books released in a given month and the columns showcase the forecasting error at the end of each forecasting month. As in the previous analysis, the projection errors decrease over time as more sales data accumulates. The initially high projections errors are largely related to the fact that the first month's demand data may only contain a partial sales period. This is particularly significant when it comes to mid-month and late-month releases. Although slight variation exists across release months, there are no particular months that clearly exhibit higher errors. Contrary to the monthly forecasting results, the releases of August, September and October don't generate larger forecasting errors. Even though these months include titles with higher absolute demand, the relative errors are similar to other release months. The findings are consistent

with those in Table 20, reinforcing that the early total demand estimates are most useful for understanding general range of demand and later projections provide a reliable basis for more precise decision-making. The forecasts generated by the projection are the most important when it comes to determining the optimal printing quantity since it covers the book's full lifecycle.

5 Results and interpretation

The theoretical part of the thesis covered the topic of demand forecasting for fiction books. It highlighted the objectives and challenges of demand forecasting, the process of generating demand forecasts and factors affecting forecasting accuracy. The section also introduced quantitative and qualitative forecasting methods, their combinations and variety of error metrics used to measure forecasting accuracy. The theoretical concepts were discussed in the context of the book publishing industry and related industries such as the fashion industry, where similar short product life cycles and seasonal demand patterns exist. The key takeaways from the theoretical section were the demand forecasting process from data collection and cleaning to measuring accuracy, the quantitative forecasting methods and the forecasting accuracy evaluation metrics since these were utilized in the empirical part of the thesis.

The case study aimed for developing a fiction book demand forecasting method for the case company. The case company is in a self-sufficient position when it comes to supply chain management since it controls the entire supply chain from publishing and printing to retail sales. This offers a unique opportunity to utilize the demand forecasts across supply chain operations. The company did not have pre-existing demand forecasting methods for fiction books on the individual level, meaning this study established the first fiction book demand forecasting method for the company which can be utilized as a benchmark in the future.

The data used in the thesis consisted of daily book sales of fiction books released in 2022 to 2024. The data was split into three datasets based on the release year. The lifecycle of a fiction book is a year at most, meaning each of the datasets consisted of different books. The datasets were cleaned from outliers such as books of the month and award winners because their demand volume exceeded and demand pattern did not correspond the typical fiction books. The demand patterns were visualized and analysed to assist in developing the most suitable model for demand forecasting. The key points found from the demand patterns were the highly elevated demand during Christmas, Father’s and Mother’s Day and the first months after release, relatively low demand during other times of the year, general downward trend in sales volume after release and the end of demand after the novelty year. During the demand pattern visualization and analysis, it was found that the books had

different demand patterns based on their release month. Therefore, the books were divided into groups based on their release month. Before generating the monthly forecasts, the daily sales were aggregated to a monthly level since the aggregation level of the data should match the generated forecasting level. The aggregation was conducted as release date based aggregation and calendar based aggregation for comparison. The hypothesis was that the release date based aggregation would produce more accurate forecasts in the early months and the calendar based method would perform better for Christmas demand, but no significant difference was found.

The monthly forecasting was conducted for both aggregation methods using naïve, SMA and SES forecasting methods. Naïve forecasting produces the most accurate results across all error metrics when comparing both the yearly datasets and monthly groups. The release date based aggregation produced generally less errors when using these three methods. Naïve, SMA and SES methods were quite poor at quickly capturing trends in the data which resulted in consistent over-forecasting, indicated by negative ME values. It was also found that large forecasting errors were evident showcased by the high MSE values. Therefore, there was room for improvement in terms of forecasting accuracy, and the coefficient method was developed based on naïve forecasting but enhanced with month-specific multipliers derived from historical sales patterns. These multipliers, as the Holt-Winters trend and seasonal components, accounted for trend and seasonal patterns and were tailored to every monthly release group. The coefficient method outperformed the previous methods. The coefficient model produced better monthly forecasts indicated by the lower MAE value, fixed the issue of over forecasting showcased by the close to zero ME values and reduced error variance demonstrated by the reduced MSE values. The coefficient method generated more accurate demand forecasts when using calendar based aggregation. Additionally, total demand projections were produced for each book based on the average cumulative sales of previous books. The projection model estimated a title's first-year demand at the end of each month, providing more accurate estimations as more sales data became available. The early projections could be utilized for rough estimates while the later projections are more suitable for timely decision-making in the supply chain.

How can sales data be utilized to forecast fiction book demand?

The theoretical framework established that historical sales data is essential in time series forecasting and highlighted the importance of matching aggregation levels between data and forecasts. The empirical findings further confirmed that the sales data of previous fiction books can be used to tailor a demand forecasting model for fiction books. The key is to understand the typical demand patterns of previous fiction releases. These include elevated demand during Christmas, first months after release and Father's and Mother's Day, generally low demand during other times of year and downwards trend in sales volume from release. On top of the volatile demand patterns, it is essential to understand that the usual lifecycle of a fiction title is one year at most and the release month has a significant effect on the demand patterns. Therefore, long-term forecasting for individual books is not viable and forecasting models must be adjusted based on the release month of the book. Understanding the typical demand patterns, it is possible to develop a demand forecasting model for fiction books that can react to the highly volatile demand predictively. When structured correctly, the sales data can be used to predict individual fiction book demand or overall category demand. The forecasting model can be built to forecast short term such as weekly or monthly demand. Additionally, previous cumulative sales can assist in projecting the total demand for fiction titles, meaning sales data can support both monthly forecasting and lifecycle demand projections.

Which forecasting methods produce the most accurate forecasts for fiction books?

The empirical analysis showed the developed coefficient method produced the most accurate monthly forecasts for fiction books across all error metrics. The model generated less errors due to reacting more efficiently to the demand patterns such as the downwards sales trend and high demand during Christmas. Therefore, it tackled the main issues of other models, such as over-forecasting and large variance of errors. The second most accurate model was the naïve forecasting method. This finding also aligns with theory which suggested that naïve forecasts are often difficult to outperform for products with short life cycles and high volatility since they naturally respond to immediate changes in demand. This is further supported by the hindered performance of the SMA3 and SES with the α value of 0,4 since

these smoothen the forecasted demand more and react slowly to rapid changes. The theory likewise stated that longer averaging windows and middling smoothing constant values tend to lag behind actual demand when volatility is high. The general finding is that reactive models perform best when forecasting fiction book demand due to the big changes in monthly demand. Simple and tailored methods must be utilized to forecast fiction book demand if forecasting is based on sales data. Since the lifecycle is only a year at most, the amount of sales data is limited, meaning utilization of traditional time series forecasting methods is less viable. Potentially, ML and AI based forecasting methods could improve forecasting accuracy in an attribute-based forecasting model which incorporates variables such as author popularity or price.

How can the demand forecasts be used to support decision-making in the case company?

The demand forecasts can be used to guide decision-making across the supply chain activities. The monthly forecasts generated by the coefficient model for individual fiction titles can be used in short-term planning and operations. The total demand projections combined with the data from other sales channels help in managing the books throughout their life cycle. One of the main objectives of the thesis was to aid determining the optimal second print quantity for individual fiction titles. The monthly forecasts assist in the correct timing of the second print while the total projections help with the correct printing quantity. These timely and precise print runs have the potential to reduce inventory costs, decrease the number of print runs and costs associated with those. The monthly forecasts can also increase the efficiency of logistics and inventory management in the supply chain. For instance, the monthly forecasts enable predictive stock replenishments in physical stores and pulling out excessive stock and reallocating it to other locations. Overall, the demand forecasts help the case company in timely and analytical decision-making and coordination in the supply chain when it comes to fiction titles. Previously, the company did not use demand forecasting or sales data exhaustively, so this thesis helps to guide the company towards predictive data-driven decision-making across the supply chain.

There were some challenges when conducting this thesis on fiction book demand forecasting. Firstly, the previously carried out research and literature on book publishing and book demand forecasting was relatively limited. There was no theoretical literature solely focusing on fiction books. Therefore, findings from related industries, such as the fashion

industry, were utilized. The short lifecycle of fiction titles caused some challenges when determining useful demand forecasting methods. The one-year lifetime of an individual fiction book meant that traditional time series models were not viable. Another obstacle in forecasting was the absence of reference points to previous year's sales on an individual book level. Generating demand forecasts for all fiction titles as a group was straightforward and the results were accurate, but the demand forecasts did not generate accurate results at the SKU level. It was challenging to develop a universal model that produces accurate results for each book since there are differences in demand volatility and unpredictability between fiction titles. The solution to this issue was the division of books into categories based on the release month and tailoring the coefficient model to each monthly group. The result is a general forecasting method that uses different multipliers for each release month. The same forecasting method for all the release months is better compared to developing a different method for each month because this solution is easier to understand, manage and maintain. Lastly, there was a practical problem when developing the demand forecasting model for the books released in November. The first quarter's sales data was missing in the 2022 dataset, meaning almost half of the sales data was not available since the November books were released in 2021. There was no November books released in the 2023 dataset. Therefore, the only dataset containing reliable data about November fiction titles is the novelty year 2024 dataset. It is not reliable to conduct demand forecasting only based on single year's releases so November releases were not examined further. In the future, when more data becomes available, the company can calculate the multipliers for November releases or use October's or December's multipliers.

This study demonstrates that sales data of previous releases can be used to build a demand forecast model for fiction books. The thesis suggests that the case company can use the coefficient method for monthly forecasting and the projection model to estimate the total demand of fiction titles. These forecasts help the case company in short term decision-making and longer-term planning across the supply chain. Utilizing the sales data and these forecasting methods supports the case company's path towards predictive and data-driven supply chain management.

6 Conclusions

The primary objective of the thesis was to develop a demand forecasting model for fiction books at the SKU level using the sales data of previous releases. The case company commissioned the thesis to improve its decision-making by incorporating demand forecasting. The study successfully achieved these goals by evaluating different quantitative forecasting methods and further developing a tailored coefficient forecasting model that generates reliable demand forecasts.

The research began by reviewing theoretical literature on demand forecasting. The previous studies highlighted the main objectives, challenges and processes of demand forecasting. The literature review also introduced quantitative and qualitative forecasting methods and common error metrics that can be applied in the context of fiction books. Due to the lack of existing research on book demand forecasting, the study established a link to industries such fashion where products also experience high volatility, short life cycles and seasonal patterns.

Empirical analysis of sales data from 2022-2024 informed that fiction book demand follows a pattern where the sales are elevated after release, major seasonal peaks are evident, and the baseline of sales is relatively low during other periods. The empirical results indicate that reactive models are better suited for fiction title demand forecasting. It was found that the coefficient model's simplicity and accuracy make it most practical and effective for the case company. Additionally, a total demand projection model presented used to estimate each title's annual sales based on cumulative monthly percentages.

The thesis offers significant substantial value to the case company as the developed forecasting methods enable the company to align printing volumes, inventory and logistics activities more accurately. The monthly forecasts improve short-term planning, while the total projections guide longer-term decisions. The results support the company's shift towards predictive and data-driven supply chain management, reducing excess printing, minimizing waste and improving operational efficiency.

Previous research has rarely focused on forecasting in the publishing industry, meaning the findings from this thesis help to address the notable research gap. This study exhibits how

time-series models can be adapted to fit to volatile and short life cycle environments. The results align with existent studies from similar industries, such as fashion, where adaptive methods provide the most accurate forecasting in volatile conditions. However, the study extends the findings to the publishing sector and demonstrates that adjusted simple models can achieve great accuracy.

There are some limitations that should be acknowledged. Firstly, the study relied only on sales data from three years, which restricted the historical analysis and the usage of more data-intensive forecasting. Since the study focused on sales in the retail locations, the COVID-19 pandemic limited the usage of extensive historical data. Secondly, fiction book demand is impacted by qualitative factors, such as author popularity, marketing investments and literary awards, that were not taken into consideration in the quantitative methods. Thirdly, missing data of November releases constrained the analysis for this month. Lastly, the thesis focused solely on sales at physical retail locations, excluding online and eBook channels that would provide a more comprehensive view of the total market.

Future studies should expand the data by including additional years. The utilized data could be enriched with title-level attributes, such as author popularity, pricing and marketing efforts, which could be incorporated to forecasting models. This would allow the explorations of attribute-based or ML models to improve accuracy and usability. Additionally, the forecasting framework could be extended to other product categories, such as non-fiction or children's books, to test the applicability of the coefficient model. Integrating external data, such as social media trends, search engine results or pre-order activity, may provide insights on the impact of consumer sentiment to fiction demand in future research.

In conclusion, this thesis successfully achieved its objectives by developing a data-driven forecasting model for fiction book demand. The research generated theoretical and practical insights into forecasting demand in the context of short product life cycles and volatile demand patterns. The findings indicate that grasping the demand patterns and employing properly reactive models can produce accurate and reliable forecasts in the context of fiction books. The case company can utilize the forecasting models to make predictive, efficient and data-driven decisions in their supply chain management. In a broader context, the study contributes on how traditional industries can utilize analytics and forecasting to transition into more intelligent decision-making in a data-centric world.

References

Ali, M.M., Babai, M.Z., Boylan, J.E. & Syntetos, A.A. 2017. Supply chain forecasting when information is not shared. European Journal of Operational Research. vol. 260, no. 3, pp. 984–994.

Al-Musaylh, M., Deo, R.C., Adamowski, J.F. & Li, Y. 2018. Short-term electricity demand forecasting with MARS, SVR and ARIMA models using aggregated demand data in Queensland, Australia. Advanced engineering informatics. vol. 35, pp. 1–16.

Babai, M.Z., Boylan, J.E. & Rostami-Tabar, B. 2022. Demand forecasting in supply chains: a review of aggregation and hierarchical approaches. International Journal of Production Research. vol. 60, no. 1, pp. 324–348.

Badulescu, Y., Hameri, A. & Cheikhrouhou, N. 2021. Evaluating demand forecasting models using multi-criteria decision-making approach. Journal of advances in management research. vol. 18, no. 5, pp. 661–683.

Beck, N., Dovern, J. & Vogl, S. 2025. Mind the naive forecast! a rigorous evaluation of forecasting models for time series with low predictability: Mind the naive forecast! a rigorous evaluation of. Applied intelligence (Dordrecht, Netherlands). vol. 55, no. 6.

Caniato, F., Kalchschmidt, M. & Ronchi, S. 2011. Integrating quantitative and qualitative forecasting approaches: organizational learning in an action research case. The Journal of the Operational Research Society. vol. 62, no. 3, pp. 413–424.

Castillo, P.A., Mora, A.M., Faris, H., Merelo, J.J., García-Sánchez, P., Fernández-Ares, A.J., De las Cuevas, P. & García-Arenas, M.I. 2017. Applying computational intelligence methods for predicting the sales of newly published books in a real editorial business management environment. Knowledge-Based Systems. vol. 115, pp. 133–151.

Chandran, J.M. and Khan, M.R.B. 2024. A Strategic Demand Forecasting: Assessing Methodologies, Market Volatility, and Operational Efficiency. Malaysian Journal of Business Economics and Management. vol. 2, pp. 150–167.

Cheikhrouhou, N., Marmier, F., Ayadi, O. & Wieser, P. 2011. A collaborative demand forecasting process with event-based fuzzy judgements. Computers & Industrial Engineering. vol. 61, no. 2, pp. 409–421.

Chicco, D., Warrens, M.J. & Jurman, G. 2021. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ.Computer science. vol. 7, pp. 623.

Ferbar Tratar, L. & Strmčnik, E. 2016. The comparison of Holt–Winters method and Multiple regression method: A case study. Energy (Oxford). vol. 109, pp. 266–276.

Fildes, R. & Goodwin, P. 2021. Stability in the inefficient use of forecasting systems: A case study in a supply chain company. International Journal of Forecasting. vol. 37, no. 2, pp. 1031–1046.

Gelper, S., Fried, R. & Croux, C. 2010. Robust forecasting with exponential and Holt-Winters smoothing. Journal of Forecasting. vol. 29, no. 3, pp. 285–300.

Goodwin, P., Petropoulos, F. & Hyndman, R.J. 2017. A note on upper bounds for forecast-value-added relative to naïve forecasts. The Journal of the Operational Research Society. vol. 68, no. 9, pp. 1082–1084.

Green, K.C. and Armstrong, J.S. 2012. Demand Forecasting: Evidence-Based Methods. SSRN Electronic Journal.

Hansun, S., Charles, V., Indrati, C.R. & Subanar 2019. Revisiting the Holt-Winters' Additive Method for Better Forecasting. International journal of enterprise information systems. vol. 15, no. 2, pp. 43–57.

Hildick-Smith, P. 2006. Systems and methods for forecasting book demand. U.S. Patent Application 11/030,725.

Hodson, T.O. 2022. Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. Geoscientific Model Development. vol. 15, no. 14, pp. 5481–5487.

Karunasingha, D.S.K. 2022. Root mean square error or mean absolute error? Use their ratio as well. Information Sciences. vol. 585, pp. 609–629.

Khairina, D.M., Daniel, Y. & Widagdo, P.P. 2021. Comparison of double exponential smoothing and triple exponential smoothing methods in predicting income of local water company. Journal of physics.Conference series. vol. 1943, no. 1, pp. 12102.

Kolková, A. & Ključnikov, A. 2021. Demand forecasting: an alternative approach based on technical indicator Pbands. Oeconomia Copernicana. vol. 12, no. 4, pp. 1063–1094.

Kourentzes, N., Rostami-Tabar, B. & Barrow, D.K. 2017. Demand forecasting by temporal aggregation: Using optimal or multiple aggregation levels?. Journal of business research. vol. 78, pp. 1–9.

Lei, C., Zhang, H., Wang, Z. & Miao, Q. 2024. Multi-Model Fusion Demand Forecasting Framework Based on Attention Mechanism. Processes. vol. 12, no. 11, pp. 2612.

Madhavan, M., Ali Sharafuddin, M., Piboonrungroj, P. & Yang, C. 2023. Short-term Forecasting for Airline Industry: The Case of Indian Air Passenger and Air Cargo. Global business review, vol. 24, no. 6, pp. 1145–1179.

Makridakis, S., Spiliotis, E. & Assimakopoulos, V. 2020. The M4 Competition: 100,000 time series and 61 forecasting methods. International Journal of Forecasting. vol. 36, no. 1, pp. 54–74.

Marmier, F. & Cheikhrouhou, N. 2010. Structuring and integrating human knowledge in demand forecasting: a judgemental adjustment approach. Production planning & control. vol. 21, no. 4, pp. 399–412.

Mas Machuca, M., Sainz, M. & Martínez Costa, C. 2014. A review of forecasting models for new products. Intangible capital. vol. 10, no. 1, pp. 1–25.

Mishra, R.S., Kumar, R., Dhingra, S., Sengupta, S., Sharma, T. & Gautam, G.D. 2022. Adaptive grey model (AGM) approach for judgemental forecasting in short-term manufacturing demand. Materials today : proceedings. vol. 56, pp. 3740–3746.

Park, M.H., Lee, J.S. & Doo, I.C. 2020. A Study of the Demand Forecasting Model for Publishing Business using Business Analysis. International Journal of Computing and Digital Systems. vol. 9, no. 5, pp. 801–812.

Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M.Z., Barrow, D.K., Ben Taieb, S., Bergmeir, C., Bessa, R.J., Bijak, J., Boylan, J.E., Browell, J., Carnevale, C., Castle, J.L., Cirillo, P., Clements, M.P., Cordeiro, C., Cyrino Oliveira, F.L., De Baets, S., Dokumentov, A., Ellison, J., Fiszeder, P., Franses, P.H., Frazier, D.T., Gilliland, M., Gönül, M.S., Goodwin, P., Grossi, L., Grushka-Cockayne, Y., Guidolin, M., Guidolin, M., Gunter, U., Guo, X., Guseo, R., Harvey, N., Hendry, D.F., Hollyman, R., Januschowski, T., Jeon, J., Jose, V.R.R., Kang, Y., Koehler, A.B., Kolassa, S., Kourentzes, N., Leva, S., Li, F., Litsiou, K., Makridakis, S., Martin, G.M., Martinez, A.B., Meeran, S., Modis, T., Nikolopoulos, K., Önkal, D., Paccagnini, A., Panagiotelis, A., Panapakidis, I., Pavía, J.M., Pedio, M., Pedregal, D.J., Pinson, P., Ramos, P., Rapach, D.E., Reade, J.J., Rostami-Tabar, B., Rubaszek, M., Sermpinis, G., Shang, H.L., Spiliotis, E., Syntetos, A.A., Talagala, P.D., Talagala, T.S., Tashman, L., Thomakos, D., Thorarinsdottir, T., Todini, E., Trapero Arenas, J.R., Wang, X., Winkler, R.L., Yusupova, A. & Ziel, F. 2022. Forecasting: theory and practice. International Journal of Forecasting. vol. 38, no. 3, pp. 705–871.

Petropoulos, F., Makridakis, S., Assimakopoulos, V. & Nikolopoulos, K. 2014. ‘Horses for Courses’ in demand forecasting. European Journal of Operational Research. vol. 237, no. 1, pp. 152–163.

Pritularga, K.F., Svetunkov, I. & Kourentzes, N. 2023. Shrinkage estimator for exponential smoothing models. International Journal of Forecasting. vol. 39, no. 3, pp. 1351–1365.

Punia, S. & Shankar, S. 2022. Predictive analytics for demand forecasting: A deep learning-based decision support system. Knowledge-Based Systems. vol. 258.

Ren, S., Chan, H. & Siqin, T. 2020. Demand forecasting in retail operations for fashionable products: methods, practices, and real case study. Annals of operations research. vol. 291, no. 1-2, pp. 761–777.

Sanders, N.R. 2017. Forecasting fundamentals. New York, Ny: Business Expert Press.

Svetunkov, I. & Petropoulos, F. 2018. Old dog, new tricks: a modelling view of simple moving averages. International Journal of Production Research. vol. 56, no. 18, pp. 6034–6047.

Swaminathan, K. & Venkitasubramony, R. 2024. Demand forecasting for fashion products: A systematic review. International Journal of Forecasting. vol. 40, no. 1, pp. 247–267.

Syntetos, A.A., Boylan, J.E. & Disney, S.M. 2009. Forecasting for Inventory Planning: A 50-Year Review. The Journal of the Operational Research Society. vol. 60, pp. 149–160.

Syntetos, A.A., Babai, Z., Boylan, J.E., Kolassa, S. & Nikolopoulos, K. 2016. Supply chain forecasting: Theory, practice, their gap and the future. European Journal of Operational Research. vol. 252, no. 1, pp. 1–26.

Tang, Z. & Dong, S. 2021. A total sales forecasting method for a new short life-cycle product in the pre-market period based on an improved evidence theory: application to the film industry. International Journal of Production Research, vol. 59, no. 22, pp. 6776–6790.

Thomson, M.E., Pollock, A.C., Önkal, D. & Gönül, M.S. 2019. Combining forecasts: Performance and coherence. International Journal of Forecasting. vol. 35, no. 2, pp. 474–484.

Tliche, Y., Taghipour, A. & Canel-Depitre, B. 2019. Downstream Demand Inference in decentralized supply chains. European Journal of Operational Research. vol. 274, no. 1, pp. 65–77.

Wallström, P. & Segerstedt, A. 2010. Evaluation of forecasting error measurements and techniques for intermittent demand. International Journal of Production Economics. vol. 128, no. 2, pp. 625–636.

Widiarta, H., Viswanathan, S. & Piplani, R. 2009. Forecasting aggregate demand: An analytical evaluation of top-down versus bottom-up forecasting in a production planning framework. International Journal of Production Economics, vol. 118, no. 1, pp. 87–94.

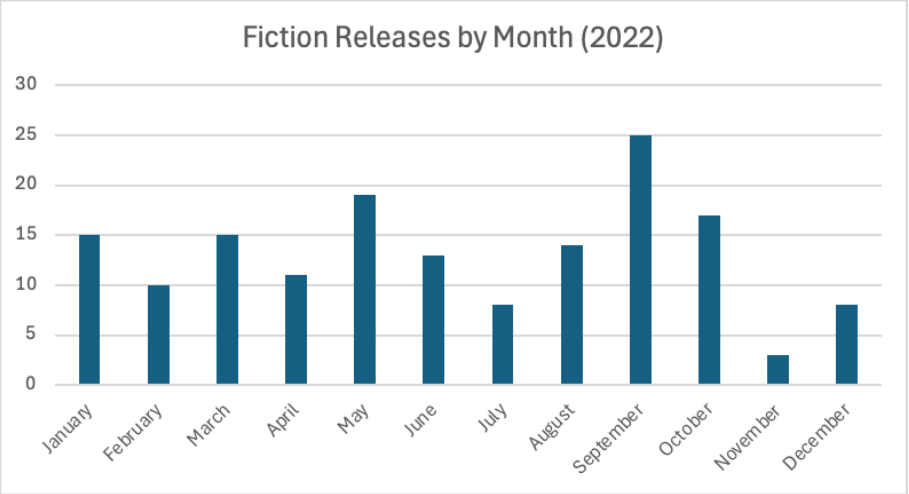
Zellner, M., Abbas, A.E., Budescu, D.V. & Galstyan, A. 2021. A survey of human judgement and quantitative forecasting methods. Royal Society Open Science. vol. 8, no. 2.

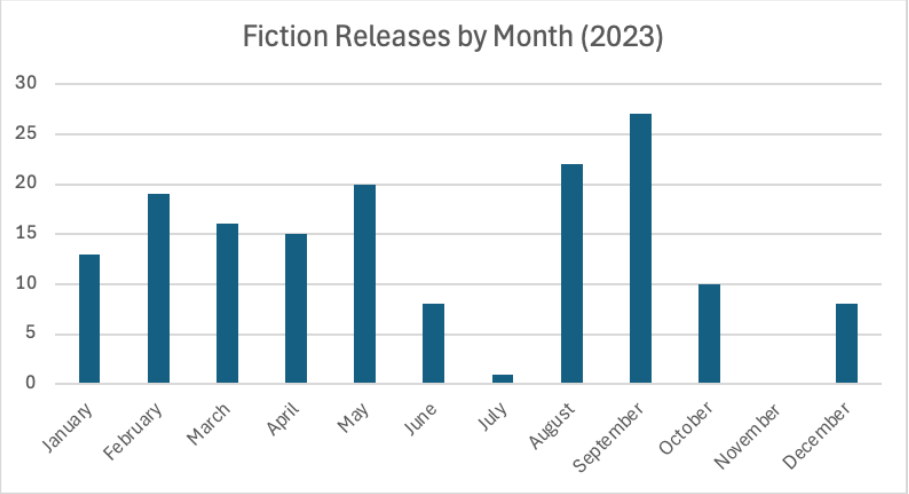
Zhang, X., Li, P., Han, X., Yang, Y. & Cui, Y. 2024. Enhancing Time Series Product Demand Forecasting With Hybrid Attention-Based Deep Learning Models. IEEE access. vol. 12, pp. 190079–190091.

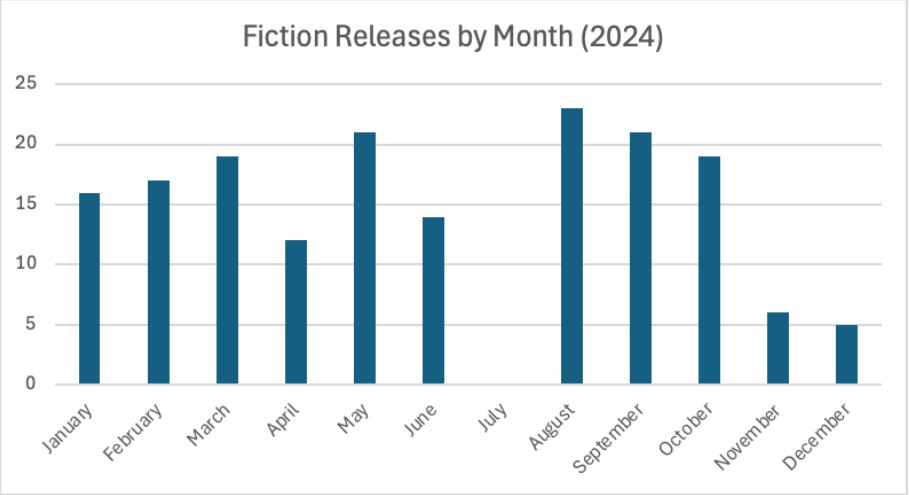
Zhou, L. 2017. Prediction of a service demand using combined forecasting approach. Journal of physics.Conference series. vol. 887, no. 1, pp. 12075.

Zhu, Y., Zhao, Y., Zhang, J., Geng, N. & Huang, D. 2019. Spring onion seed demand forecasting using a hybrid Holt-Winters and support vector machine model. PloS one. vol. 14, no. 7.

Appendix 1. Fiction releases by month for each year







Appendix 2. Multipliers of the coefficient model

		Multipliers Assigned to Each Month											
		Jan	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Jul**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**
The Release Month of the Book	**January**	1.00	1.27	0.73	0.60	1.05	0.57	0.92	0.67	0.68	1.57	0.75	4.43
	February	0.33	1.00	2.81	0.43	0.79	0.61	0.94	0.81	0.69	1.20	1.00	3.80
	March	0.45	0.30	1.00	1.69	0.68	0.51	0.83	0.71	0.72	1.16	0.90	3.71
	April	0.36	0.34	0.74	1.00	2.01	0.68	0.83	0.56	0.78	1.03	1.38	2.35
	May	0.42	0.31	0.77	0.79	1.00	2.39	0.86	0.65	0.67	0.99	0.90	3.50
	June	0.49	0.37	0.86	0.79	1.47	1.00	1.44	0.63	0.62	0.83	1.01	3.40
	July	0.52	0.15	0.74	0.73	0.82	1.39	1.00	1.73	0.61	0.58	1.28	4.04
	August	0.20	0.39	0.75	0.56	0.80	1.36	1.35	1.00	1.30	0.81	0.95	2.76
	September	0.14	0.42	0.73	0.57	0.83	1.17	1.13	0.67	1.00	1.73	0.92	2.89
	October	0.16	0.36	0.71	0.62	0.82	1.05	1.06	0.73	0.93	1.00	1.34	2.21
	November	-	-	-	-	-	-	-	-	-	-	-	-
	December	2.18	0.49	0.69	0.67	1.02	0.57	0.98	0.71	0.98	1.77	1.13	1.00

Appendix 3. The average cumulative sales at the end of each month for monthly releases

		The Average Cumulative Sales at the End of Month											
		Jan	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Jul**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**
The Release Month of the Book	**January**	16%	36%	50%	59%	68%	73%	78%	82%	84%	87%	90%	100%
	February	100%	10%	38%	50%	59%	64%	69%	73%	75%	79%	83%	96%
	March	98%	100%	14%	36%	52%	60%	67%	71%	74%	78%	81%	93%
	April	97%	99 %	100%	13%	35%	48%	59%	65%	69%	73%	78%	91%
	May	95%	97%	99%	100%	8%	25%	40%	49%	56%	62%	67%	86%
	June	90%	93%	96%	97%	100%	12%	29%	40%	47%	53%	59%	79%
	July	80%	90%	93%	95%	98%	100%	8%	22%	32%	38%	45%	70%
	August	92%	95%	96%	97%	98%	99%	100%	10%	23%	35%	48%	86%
	September	90%	92%	94%	96%	97%	98%	99%	100%	10%	27%	41%	82%
	October	93%	95%	96%	97%	97%	98%	99%	99%	100%	15%	36%	85%
	November	-	-	-	-	-	-	-	-	-	-	-	-
	December	39%	53%	63%	70%	77%	81%	85%	88%	90%	96%	100%	10%