

Optimizing Sales Forecasting, Inventory, Pricing and Sourcing Decisions

Présentée le 31 mars 2023

Collège du management de la technologie
Chaire de technologie et gestion opérationnelle
Programme doctoral en management de la technologie

pour l'obtention du grade de Docteur ès Sciences

par

Yara KAYYALI EL ALEM

Acceptée sur proposition du jury

Prof. G. J. A. de Rassenfosse, président du jury
Prof. R. Seifert, directeur de thèse
Prof. A. Huchzermeier, rapporteur
Prof. C. Tucci, rapporteur
Prof. Ph. Wieser, rapporteur

To my family, for always being my support system and for believing in me.

And to my partner, Gontran Sangouard, for his extraordinary love and support in every step of
this process.

Acknowledgements

I would like to express my deepest gratitude and appreciation to my advisor, Prof. Ralf W. Seifert. He has given me great opportunity to learn and grow. He has believed in me since I joined the program and has provided me with guidance, support, feedback, and trust throughout the process. I am truly grateful for having him as my supervisor.

I would like to thank my coauthors, Prof. Sebastian Maier and Prof. Işık Biçer, who were very helpful during my PhD. Sebastian provided valuable advice towards improving my academic paper writing skills. And Işık took the time to involve me in two very interesting projects which constitute two chapters of my thesis. He is a great cosupervisor and is passionate in explaining new fields of research.

I would like to thank my dissertation committee: Prof. Gaétan de Rassenfosse, Prof. Arnd Huchzermeier, Prof. Christopher L. Tucci, and Prof. Philippe Wieser for accepting to be part of my jury.

I would like to thank the administrative staff of the College of Management (CDM) at EPFL. In particular, I would like to express my sincere gratitude to Ilona Ball for all her kindness and help throughout my PhD. I would like to thank Amandine Weissbrodt for her friendship and support. And I would like to thank the IT team at CDM for their constant support.

I am extremely thankful to have met my office mates, or rather my second family. For my first two years, in particular, Paul for always being there for me and for bringing the happiest vibes to the office; Charles, from being my TA to my colleague and friend, for his constant enthusiasm and advice on research that helped me throughout my journey; Andrey for being there from day one, and supporting me in the darkest times of convex optimization; Rachel for her advice and sharing the same interests as me in discovering new places in Lausanne; Ling for her kindness and travel advices; Yu for her thoughtfulness and best recommendations in Switzerland.

When almost everyone graduated, I was blessed with a new family that has joined the office. I would like to thank Philipp for his positivity, for making sure everyone is included, and his support throughout the PhD process; Andrea for always being helpful whether it is in preparing me for job interviews or helping me brainstorm my meal plan for the week; Onur for our lovely chats, for constantly asking about me, and for always proposing to help; Atin for her constant support and consideration, and for all the great times as gym partners; Yanan for her support and kindness, making sure I am feeling great at all times; and Buse for her amazing personality and her friendship.

I am also grateful for people in the other offices with whom I have become friends with. Roy

Acknowledgements

for being a great listener and for helping me on several occasions; and Wolf for giving me feedback and advice on several projects and decisions. I would also like to thank Anna, Réne, Max, George, Raphael, Dirk, Yves, Micheal, Matthias, Gabriele, Sina, Fatemeh, Enrico, Farnaz, Valeria, Sadegh, and Mohammadsaeed for our coffee breaks and chats over the years.

I would like to express how lucky I feel to have three amazing people by my side through my PhD and throughout my life. Bahar for devoting so much time to helping me on the academic side and on the personal side, and for sharing a special bond with me since the first day we met; Lynn for being my number one cheerleader through my entire PhD journey and for believing so much in me when I am faced with the biggest obstacles; and Karam for his constant understanding, wisdom, and help in countless situations and for always being by my side.

My friends outside the PhD program have also been of great support. A special thank you to Yaman and Fayez for helping me onboard into the coding and machine learning world. A thank you to Liza, Nour, George, Ray, Joanna, Farzane, Dorsa, Rashad, Edouard, Khalid, Malak, Raed, Rima, and Rama.

Finally, and most importantly, a thank you from the bottom of my heart to my family. My parents, Faten and Fayad, for dedicating their lives to making sure I achieved all of what I am capable of, for believing in me, for supporting me, and for being my safe place. I would also like to thank my siblings, Tamara, Iyad, Samir, and Sarah, for always being attentive to my worries, sharing and celebrating all my achievements, and having my back throughout my life. A thank you for my siblings-in-law, Ahmed, Catherine, and Amro and my niece Layan and nephews Rayan, Adam, and Malek. For Gontran, my boyfriend, all the gratitude for being by my side during the PhD years, spending so much time listening to my presentations and reading my papers, making sure I have nothing to worry about except my studies, and easing all the stress I have been through in the process. I couldn't have done it without you!

Lausanne, January 25, 2023

Yara Kayyali Elalem.

Abstract

In this thesis we address various factors that contribute both theoretically and practically to mitigating supply demand mismatches. The thesis is composed of three chapters, where each chapter is an independent scientific paper. In the first paper, we develop a framework that enables forecasters to use quantitative methods to forecast sales of newly launched products with limited historical data. The forecasts are provided over the new product's life cycle by leveraging data from similar older products. Our framework allows us to test autoregressive integrated moving average with exogenous variables (ARIMAX) and machine learning (ML) methods. It does so by exploiting the historical data of similar products through time-series clustering and by performing data augmentation to generate sufficient sales data. We perform a comparative analysis between ARIMAX and three deep neural networks (DNNs). Based on the results of using two data sets, we show that ARIMAX outperforms the more advanced DNNs. However, when adding Gaussian white noise to test the robustness of the methods, DNNs show better performance as the ARIMAX performance deteriorates with an increase in noise level. We provide insights for practitioners on when to use advanced forecasting methods and when to use traditional methods. The right choice of forecasting method leads to an increase in forecasting accuracy which results in better matches between supply and demand.

In the second paper, we study a price-setting newsvendor problem where the decision maker sells multiple products with uncertain demand and is faced with a capacity constraint on the total order quantity. To solve the problem, we first develop analytical expressions that make it possible to derive the optimal order quantities and optimal prices given a capacity limit. We show that the optimal allocation policy for the capacity is a nested-allocation policy. We show that the more limited the capacity, the more convergent the pricing policy. Optimal pricing policies change depending on the demand uncertainty and price elasticity. We therefore develop a decision typology that can be adopted by practitioners to choose the optimal pricing policy. It is important to study pricing and inventory simultaneously. Prices affect demand and demand affects order quantities. Therefore, optimizing both order quantities and prices has a positive impact on inventory management and profit.

Finally, in the third paper, we analyze the benefits of operational flexibility on both profitability and environmental sustainability. We measure the impact on the environment by the reduction in excess inventory at the end of the season when a certain operational-flexibility strategy is employed. We consider three different operational flexibility strategies: (1) lead-time reduction, (2) quantity-flexibility contracts, and (3) multiple sourcing. We find that the

lead-time reduction strategy has the maximum capability to reduce waste in the sourcing process, followed by the quantity flexibility and multiple-sourcing strategies, respectively. Our results show that a firm can improve on both environmental sustainability and profitability by using an operational-flexibility strategy that localizes production and therefore provides better matching supply to demand.

Key words: Forecasting, Machine Learning, Product Life Cycle, Analytics, Inventory Management, Pricing Policy, Product Portfolio Management, Sustainability, Sourcing, Operational Flexibility

Résumé

Dans cette thèse, nous abordons divers sujets qui contribuent à la fois théoriquement et pratiquement à atténuer les inadéquations entre l'offre et la demande. La thèse est composée de trois chapitres, où chaque chapitre est un article scientifique indépendant. Dans le premier article, nous développons un modèle conceptuel qui permet aux prévisionnistes d'utiliser des méthodes quantitatives pour prévoir les ventes de produits nouvellement lancés avec des données historiques limitées. Les prévisions de cycle de vie du nouveau produit sont basées sur les données de produits similaires plus anciens. Notre modèle nous permet de tester la moyenne mobile intégrée autorégressive avec des variables exogènes (ARIMAX) et des méthodes de machine learning (ML). Cela fonctionne par le regroupement des données historiques de produits similaires suivi par une augmentation de données sur le cycle de vie des produits. Nous effectuons une analyse comparative entre ARIMAX et trois deep neural networks (DNN). Sur la base des résultats de l'utilisation de deux ensembles de données, nous montrons qu'ARIMAX surpasse les DNN les plus avancés. Cependant, lors de l'ajout de bruit blanc Gaussien pour tester la robustesse des méthodes, les DNN affichent de meilleures performances que ARIMAX qui se détériorent avec une augmentation du niveau de bruit. Nous fournissons des informations aux managers concernant le meilleur moment pour utiliser des méthodes de prévision avancées ou des méthodes traditionnelles. Le bon choix de la méthode de prévision conduit à une augmentation de la précision des prévisions, ce qui se traduit par une meilleure concordance entre l'offre et la demande.

Dans le deuxième article, nous étudions un problème type basé sur la détermination des prix chez un marchand de journaux. Celui-ci vend plusieurs produits dont la demande est incertaine et a une limite de capacité sur la quantité totale de produits qu'il peut commander. Pour résoudre ce problème, nous développons d'abord des équations qui permettent de dériver les quantités de commande optimales et les prix optimaux compte tenu d'une limite de capacité. Nous montrons que la politique d'allocation optimale de la capacité est une politique d'attribution imbriquée. Nous montrons que plus la capacité est limitée, plus la politique tarifaire est convergente. Les politiques tarifaires optimales changent en fonction de l'incertitude de la demande et de l'élasticité des prix. Nous élaborons donc un système décisionnel pouvant être adopté par les managers pour choisir la politique tarifaire optimale. Il est important d'étudier simultanément les prix et les stocks. Les prix affectent la demande et la demande affecte les quantités de commande. Par conséquent, l'optimisation des quantités et des prix des commandes a un impact positif sur la gestion des stocks et les bénéfices.

Enfin, dans le troisième article, nous analysons les bénéfices de la flexibilité opérationnelle à

la fois sur les profits et sur l'environnement. Nous mesurons l'impact sur l'environnement de la réduction des stocks excédentaires à la fin de la saison lorsqu'une certaine stratégie de flexibilité opérationnelle est utilisée. Nous considérons trois stratégies de flexibilité opérationnelle différentes : (1) la réduction des délais, (2) les contrats de quantité variable et (3) l'approvisionnement multiple. Nous constatons que la stratégie de réduction des délais a la capacité maximale de réduire les déchets dans le processus d'approvisionnement, suivi par les contrats à quantité variable et les stratégies d'approvisionnement multiple, respectivement. Nos résultats montrent qu'une entreprise peut améliorer à la fois la durabilité environnementale et la rentabilité en utilisant une stratégie de flexibilité opérationnelle qui localise la production pour fournir une meilleure adéquation entre l'offre et la demande.

Mots clefs : Prévission, Machine Learning, Cycle de Vie du Produit, Analyse, Gestion des Stocks, Politique Tarifaire, Gestion du Portefeuille de Produits, Durabilité, Approvisionnement, Flexibilité Opérationnel

Contents

Abstract (English/Français)	iv
List of Figures	xi
List of Tables	xiii
Introduction	1
1 A machine learning-based framework for forecasting sales of new products with short life cycles using deep neural networks	7
1.1 Introduction	7
1.2 Literature review	10
1.2.1 Product life cycle	10
1.2.2 Quantitative models for forecasting sales of new products	11
1.2.3 Machine learning for forecasting sales of new products	12
1.3 Data	13
1.3.1 Data sets	13
1.3.2 Data preparation	15
1.4 Clustering	16
1.4.1 Overall approach	17
1.4.2 Assigning products to clusters	19
1.4.3 Comparison between integration and DTW	19
1.5 Forecasting	21
1.5.1 Data augmentation	21
1.5.2 Traditional statistical model	23
1.5.3 Machine learning models	25
1.5.4 PLC shape-based methods	26
1.6 Forecasting results	28
1.6.1 Dell data set	29
1.6.2 Retailer X data set	32
1.7 Robustness of the quantitative methods	34
1.7.1 Impact of adding Gaussian white noise	34

1.7.2 Effect of incorrect cluster assignment	35
1.8 Conclusions	37
2 Optimal policies for a price-setting newsvendor with multiple products and a joint capacity constraint	41
	41
2.1 Introduction	41
2.2 Literature review	43
2.3 Model preliminaries	44
2.3.1 Multi-product model	44
2.3.2 Properties of the optimal pricing solution	48
2.4 Numerical illustration for the multi-product case	50
2.4.1 Convergence of solutions	50
2.4.2 Identical cost structures	51
2.4.3 Variations in cost structures	54
2.4.4 Sensitivity analysis	56
2.5 Conclusion	60
3 Why do companies need operational flexibility to reduce waste at source?	63
	63
3.1 Introduction	63
3.2 Literature review	65
3.3 Model preliminaries	68
3.4 Analysis of the impact of operational flexibility on excess inventory	71
3.4.1 Lead-time reduction	71
3.4.2 Quantity-flexibility contracts	73
3.4.3 Multiple sourcing	75
3.5 Discussion	78
3.6 Conclusion	79
Conclusion	81
A Supplementary material for Chapter 1	83
A.1 Clustering results	83
A.2 Exploratory data analysis	88
A.3 Integration	90
A.4 Dynamic time warping	92
A.5 Comparison of DTW and Integration using MSE	94
A.6 Results of using DTW in clustering and cluster assignment	95
A.7 Estimation and implementation details	98
A.8 Error measurements of PLC shape-based methods	100
A.9 Evaluating out-of-sample forecast accuracy	102

Contents

A.10 Error distributions	103
A.11 Additional results for incorrect cluster assignment analysis	105
B Supplementary material for Chapter 2	107
B.1 Proof of Theorem 1	107
B.2 Proof of Concavity	111
B.3 Proof of Theorem 2	112
B.4 Proof of Theorem 3	115
B.5 Proof of Theorem 4	116
B.6 Detailed explanation of possible price increase with capacity	117
Bibliography	126
Curriculum vitae	127

List of Figures

1	Matching supply and demand by optimizing various supply chain decisions (Desmet, 2016).	2
1.1	Main steps of proposed framework for forecasting sales of new products with short life cycles	8
1.2	Comparison between ARIMAX with and without noise (SD of 0.01) and GRUs with noise after 20 input weeks (Dell data)	10
1.3	Time-series of the two data sets used in this study, with the red, blue and green lines representing weekly normalized sales (before smoothing) for three randomly selected SKUs of each data set.	14
1.4	PLC curve fit by 3rd degree polynomial (SKU150, Dell data)	16
1.5	Clustering results with individual curves both normalized and smoothed (Dell data)	18
1.6	Percentage of correct cluster assignment achieved by DTW and integration for various weeks after product launch (Dell data)	20
1.7	Example of data augmentation results on Dell data set (first weeks of SKU 105)	23
1.8	Example of histogram of residuals without and with data augmentation (SKU 10, Dell data)	24
1.9	Reference to how the Bass-II model works	27
1.10	Comparison of forecasting results of ARIMAX and GRUs (SKU 112 from the Dell data set, with 10 weeks of introduction)	31
1.11	Comparison of forecasting results of ARIMAX and Bass-II (SKU 24 from the Dell data set, with 10 weeks of introduction)	31
1.12	Forecast errors (MASE) after adding Gaussian white noise to Dell data (with a log-scale vertical axis)	35
2.1	Convergence of quantities and prices over 30 iterations of the two sub-problems with a capacity constraint of 300 units.	51
2.2	Impact of capacity on optimal order quantities and optimal prices.	52
2.3	Impact of capacity on capacity utilization.	52
2.4	Impact of capacity on expected leftovers and expected sales.	53
2.5	Impact of capacity on expected profit.	53
2.6	Impact of changes in products' salvage values on optimal prices with a capacity constraint of 1100 units.	55

2.7	Changes in optimal prices with changes in the order cost of product 1 for a capacity constraint of 300 units.	55
2.8	Impact of changes in products' ordering costs on the critical capacity K^C	56
2.9	Sensitivity analysis of price elasticity on optimal prices of products with high demand uncertainty.	57
2.10	Sensitivity analysis of price elasticity on optimal prices of products with low demand uncertainty.	58
2.11	Pricing policy topology based on demand uncertainty and price elasticity. . . .	58
3.1	Closing the loop in fashion industry	66
3.2	Evolution of demand forecasts for the multiplicative process with $D_0 = 1$, $v = 0$, and $\zeta = 1$	69
3.3	Lead-time reduction	72
3.4	Quantity-flexibility contracts	75
3.5	Multiple Sourcing	77
A.1	Result of dendrogram analysis using the Dell data	83
A.2	Scree plot of the Dell data	84
A.3	Result of dendrogram analysis using the single method on Retailer X data	84
A.4	Result of dendrogram analysis using the complete method on Retailer X data . .	85
A.5	Result of dendrogram analysis using the average method on Retailer X data . .	85
A.6	Result of dendrogram analysis using the ward method on Retailer X data	85
A.7	Scree plot of the Retailer X data	86
A.8	Clustering results with individual curves both normalized and smoothed (Retailer X data)	86
A.9	Correlation heat maps for Dell data set clusters	88
A.10	Correlation heat maps for Retailer X data set clusters (15% of SKUs)	89
A.11	Assigning products to clusters by integration (Dell data)	90
A.12	DTW performance when comparing SKU107 and SKU115 (Dell data)	93
A.13	Results of DTW and integration in percentage of correct assignment using MSE instead of MAE (Dell data)	94
A.14	Comparison of cluster assignment performance of DTW and Integration for various weeks after product launch when clustering on the Dell data set is done using either the DTW (-D) or the Euclidean distance (-E) measure	95
A.15	Dendrogram result from using DTW as a distance measure on the Dell data set	96
A.16	Clustering result from using DTW as a distance measure on the Dell data set . .	97
A.17	Forecast errors (MASE) of the quantitative methods using the Dell data set with 20 weeks of input	103
A.18	Forecast errors (MASE) of the PLC shape-based methods using the Dell data set with 20 weeks of input	104
B.1	Explanation of price increase with capacity for some parameters.	118

List of Tables

1.1	Best-fit values for Gaussian Distribution (SKU 10, Dell data)	24
1.2	Summary of PLC shape-based methods	27
1.3	Forecast errors for new products for different weeks after introduction (Dell data)	30
1.4	Forecast errors for new products for different weeks after introduction (Retailer X data)	32
1.5	Average MASE results for (in)correct cluster assignment using ARIMAX on Dell data	36
1.6	Average MASE results for (in)correct cluster assignment using LSTM on Dell data	36
1.7	Average MASE results for (in)correct cluster assignment using GRUs on Dell data	37
1.8	Average MASE results for (in)correct cluster assignment using CNNs on Dell data	37
3.1	Summary of results of the operational-flexibility strategies	78
A.1	Grid search for hyperparameter tuning	98
A.2	Python packages used for implementation	99
A.3	Forecast errors of PLC shape-based methods for new products for different weeks after introduction (Dell data)	100
A.4	Forecast errors of PLC shape-based methods for new products for different weeks after introduction (Retailer X data)	101
A.5	Average MASE results when forecasting using all products in the Dell data set (i.e. without cluster assignment)	105

Introduction

Motivation

Supply and demand mismatches persist even when firms continue to invest in more advanced technology and improve information sharing across the supply chain (Colback, 2022). Mismatches are caused by various factors such as an increase in demand uncertainty or the availability of limited supply. The problem worsens with the change in consumer buying behavior as consumers become more price-sensitive. Reasons such as pandemics, geopolitical conflicts, trade wars, and extreme weather events further increase supply and demand mismatches (Niranjan et al., 2022).

Mismatches between demand and supply affect inventory levels. Whether inventory is in excess at the end of the season or the firm faces shortages both have negative impacts on a firm's profit. The demand uncertainty in 2022 has resulted in inventory glut in several firms, resulting in companies such as Target Corp. to face a 90% decrease in net earnings (Nassauer and Terlep, 2022). Companies such as Samsung expected demand for semi-conductor products to increase after the shortages in supply in 2021 (Seifert and Markoff, 2021). By anticipating an increase in demand for memory chips (semi-conductors), Samsung increased their production and inventory. This resulted in a profit drop of 69% from the previous year due to an unexpected decrease in sales (Wong, 2023).

Managing inventory levels by considering factors such as demand forecasts, price-elasticity, capacity limits, and supplier lead times is therefore crucial for improving a firm's performance. Figure 1 highlights other factors that contribute to excess inventory and stockouts. This thesis focuses on increasing the forecasting accuracy of newly launched products, jointly deciding on order quantities and optimal prices for multiple products, and choosing a supplier based on costs and lead times to diminish differences in supply and demand (factors addressed in this thesis are highlighted in orange in Figure 1).

Decreasing supply and demand mismatches by limiting the amount of excess inventory has shown positive impact on both profitability and sustainability. In terms of profitability, reducing excess inventory reduces capital tied up in warehouses and products salvaged (or eventually destroyed) at very low margins (Seifert et al., 2021). In terms of sustainability, reducing excess inventory saves on energy, material, and water required to produce the unsold

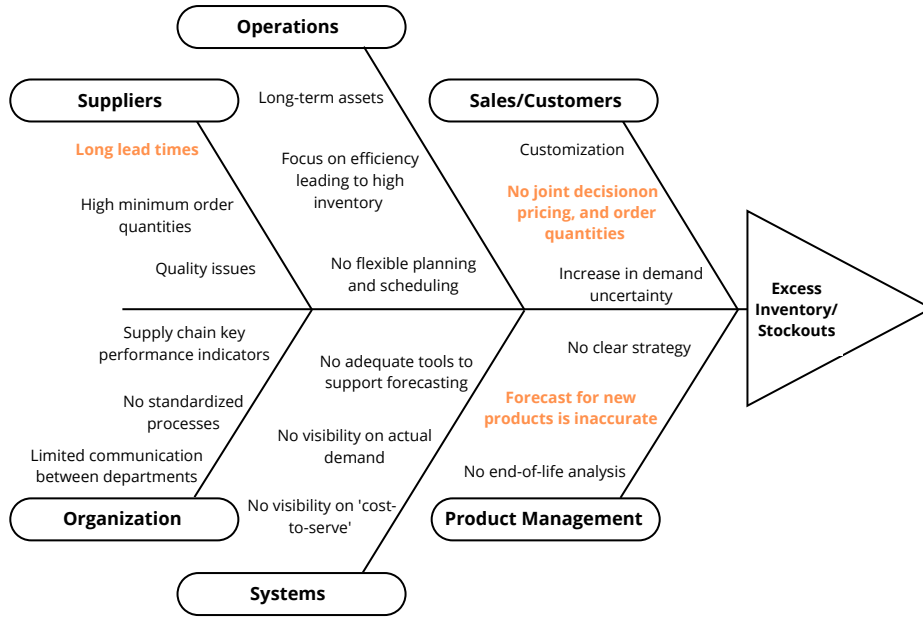


Figure 1: Matching supply and demand by optimizing various supply chain decisions (Desmet, 2016).

items (Winston, 2011). The last chapter of this thesis therefore shows the benefits of decreasing excess stock levels on both profitability and waste reduction.

Background information

For each chapter in the thesis, demand is modeled in a different way, spanning three commonly used forms of demand found in literature. In Chapter 1, demand is merely a single output value for each time-step in the forecast horizon. Demand is the output of the framework developed and the forecasting methods tested in the chapter. To achieve this output, historical sales data is needed. The accuracy of the results depends on the historical sales input data and how well the data fits the forecasting methods. Expressed succinctly, by using y_t to denote the output demand at time t , the demand in the first chapter is formulated as follows:

$$y_t = f(x_1^i, \dots, x_\tau^i, y_{t-1}, \dots, y_{t-\tau}, \mathbf{V}), \quad \forall i \in I, \quad (1)$$

where x_1^i, \dots, x_τ^i denote the input historical sales data of the older products $i \in \mathcal{I}$ offered by the firm apart from the product forecasted for, given τ input week(s), and $y_{t-1}, \dots, y_{t-\tau}$ denotes the historical data of the product being forecasted. \mathbf{V} specifies the different parameters or weights based on the forecasting method used, and f is a single or combination of functions

Introduction

that can be linear (for autoregressive integrated moving average (ARIMA)) or non-linear (for the machine learning (ML) methods).

In Chapter 2, demand is no longer an output that is optimized for, but an input function to optimizing decisions in an inventory and pricing problem. Demand is a function that consists of a deterministic term that is dependent on price and a random term. The random term in the demand function can be modeled in two different fashions: (i) additive demand case or (ii) multiplicative demand case. The demand for a product given the additive case is defined as follows:

$$d(p, \epsilon) = y(p) + \epsilon, \quad (2)$$

and the demand for a product given the multiplicative case is defined as follows:

$$d(p, \epsilon) = y(p) \epsilon, \quad (3)$$

where $y(\cdot)$ denotes the deterministic term in the demand function which is dependent on the price p of the product and ϵ denotes the random term in the demand function that follows a known distribution.

In Chapter 3, demand modeling is used. Demand modeling creates an adaptive demand distribution. It shows a decrease in the level of uncertainty as the selling season approaches, when actual demand becomes known. Demand is used as an input variable. The decision variable optimized for is the choice of an onshore supplier, an offshore supplier, or both, where different suppliers are associated with different costs and lead times. The distribution of demand changes based on the time the order is placed. This corresponds to the chosen supplier's lead time. The multiplicative demand model yields a lognormal distribution for the end demand, which is conditional on the demand forecast at t_j :

$$\ln(D_n)|D_j \sim \mathcal{N}(\ln(D_j) + (v - \zeta^2/2)(t_n - t_j), \zeta \sqrt{t_n - t_j}), \quad \forall j \in \{0, \dots, n-1\}, \quad (4)$$

where D_j is the demand forecast at time t_j prior to the selling season and D_n denotes the end demand. The location parameter of the lognormal distribution is $\ln(D_j) + (v - \zeta^2/2)(t_n - t_j)$, and the scale parameter is $\zeta \sqrt{t_n - t_j}$, where v denotes the drift rate and ζ denotes the volatility parameter.

Contribution to the literature

The individual chapters of this thesis each contribute to different streams of literature. Chapter 1 contributes to the literature of new product forecasting. A common approach used for predicting sales of new products is judgemental forecasting (Kahn, 2002). Practitioners often use their judgement to compare new products to older, similar products, and then predict

how the new products will sell. Recent research on new product forecasting leverages the historical data of comparable, older products to forecast the sales of new products using quantitative methods such as ML (examples found in Meade and Islam (2006), K. Hu et al. (2018), Thomassey and Happiette (2007), and van Steenberg and Mes (2020)) Chapter 1 of this thesis adds to the many recent advancements in the field and extends the data preparation for the quantitative methods to yield more accurate results. Instead of only using the sales history of similar older products through clustering, a simple form of data augmentation is used to increase the number of data points and provide a better fit to the quantitative models. Moreover, two quantitative methods, integration and dynamic time warping (DTW), are used to aid managers in their decision-making process by relating a newly launched product to groups of similar old products. The contribution of this chapter is a framework that enables newly launched products to be forecasted using quantitative methods. In addition, the framework compares the performance of ARIMA, a traditional statistical method, and three deep neural networks (DNNs) under different noise levels in the data. The comparison between different forecasting methods in the paper provides managerial insights on the most ideal forecasting method based on the demand of the product: whether the demand is stable, fluctuating, or the product is experiencing demand shocks (Seifert and Elalem, 2020). Prior to investing in ML for forecasting, managers can use our framework to test ML methods and benchmark their performance against the company's own forecast accuracy measurements and the actual demand that occurred during a given time horizon (Seifert and Elalem, 2020).

Chapter 2 contributes to the literature by providing expressions that enable the calculation of the optimal order quantities and optimal prices for a price-setting newsvendor offering multiple products given a capacity limit. The paper extends price-setting newsvendor studies (examples include Whitin (1955), Karlin and Carr (1962), Younge (1978), and Petruzzi and Dada (1999)) to the multi-product case with the addition of a capacity constraint on the order quantities. We show that the optimal capacity allocation policy is a nested allocation policy. We also show that the pricing policy is more convergent when capacity is more limited. Research in this domain provides solutions that involve dynamic pricing policies, continually updating prices based on changes in exogenous variables such as demand uncertainty (Xu et al. (2010), Salinger and Ampudia (2011), Murray et al. (2012)). Changing prices too often should be justified against the increase in complexity and costs associated to marketing (Feng, 2010) especially that it may result in customer dissatisfaction and attrition (Scott, 2022). Therefore, we contribute to the literature by providing a decision topology for managers to decide on the best pricing policy based on demand uncertainty and price elasticity. The optimal policies differ by being either rigid or capacity-dependent, and from a single pricing policy to a differentiated pricing policy for the different products. We also analyze the effect of changes in the cost structures on the optimal prices and pricing policies.

Chapter 3 contributes to the literature by addressing the gap between research on operational flexibility and environmental sustainability. Studies on operational flexibility focus on its impact on profit (de Treville et al. (2014), Biçer (2015), Bicer and Hagspiel (2016), Biçer and Seifert, (2017), Biçer et al. (2018)). However, operational flexibility also has an impact on

inventory, and excess inventory accounts for waste. We therefore address the environmental value of operational flexibility measured in the form of waste reduction by considering three operational flexibility strategies. We find that operational flexibility strategies that focus on the localization of production are highly effective in increasing profits and reducing leftover inventory waste (Seifert et al., 2021). With emerging trends towards more sustainable operations spanning all scopes (Scopes 1, 2, and 3), our results show that firms can use production localization to decrease their Scope 3 emissions (Seifert and Elalem, 2022).

Thesis overview

This thesis constitutes three articles, two published and one unpublished. These are joint projects with my supervisor Prof. Ralf W. Seifert, and my coauthors Prof. Sebastian Maier (Chapter 1) and Prof. Işık Biçer (Chapters 2 and 3). The thesis consists of three independent chapters. Chapter 1 addresses the challenges of using ML methods in forecasting sales of newly launched products. These methods typically require abundant historical data for proper model-fitting, making it challenging to use them for new products with limited historical data. Chapter 2 optimizes both order quantities and pricing decisions of a newsvendor offering multiple products with a joint capacity constraint. Chapter 3 analyzes the effects of three different operational flexibility strategies on a firm's performance by evaluating the results in terms of profitability and environmental sustainability through waste reduction.

The chapters are self-contained. We summarize the chapters and outline their contributions as follows:

1. **A machine learning-based framework for forecasting sales of new products with short life cycles using deep neural networks**

Demand forecasting is becoming increasingly important as firms launch new products with short life cycles more frequently. This paper provides a framework based on state-of-the-art techniques that enable firms to use quantitative methods to forecast sales of newly launched, short-lived products that are similar to previous products when there is limited availability of historical sales data for the new product. In addition to exploiting historical data using time-series clustering, we perform data augmentation to generate sufficient sales data and consider two quantitative cluster assignment methods. We apply one traditional statistical (ARIMAX) and three machine learning methods based on DNNs – long short-term memory (LSTM), gated recurrent units (GRUs), and convolutional neural networks (CNNs). Using two large data sets, we investigate the forecasting methods' comparative performance and, for the larger data set, show that clustering generally results in substantially lower forecast errors. Our key empirical finding is that simple ARIMAX considerably outperforms the more advanced DNNs, with mean absolute errors up to 21-24% lower; however, when adding Gaussian white noise in our robustness analysis, we find that ARIMAX's performance deteriorates dramatically, whereas the considered DNNs display robust performance. Our results provide insights

for practitioners on when to use advanced deep learning methods and when to use traditional methods.

“Republished with permission of Elsevier Science & Technology Journals, from [A machine learning-based framework for forecasting sales of new products with short life cycles using deep neural networks, Elalem, Maier, and Seifert]; permission conveyed through Copyright Clearance Center, Inc. (“CCC”). (license ID: 1313291-1).”

2. Optimal policies for a price-setting newsvendor with multiple products and a joint capacity constraint

We study a price-setting newsvendor problem in which a decision maker sells multiple products in the market with uncertain demand. The newsvendor procures the products from a supplier with a limited capacity such that the sum of order quantities of all products is restricted to being no more than the capacity limit. Thus, she jointly determines the price and order quantities given the shared capacity constraint. We first develop analytical expressions that make it feasible to derive the optimal price and order quantities under the capacity limit. We show that the optimal policy for capacity allocation is a nested-allocation policy. We also show that the more limited the capacity, the more convergent the pricing policy. We demonstrate that optimal price may increase with the capacity when the demand uncertainty is high and the price elasticity is low. Finally, we develop a decision typology to discuss the managerial insights.

3. Why do companies need operational flexibility to reduce waste at source?

We analyze the environmental benefits of operational flexibility that emerge in the form of less product waste during the sourcing process by reducing overproduction. We consider three different options for operational flexibility: (1) lead-time reduction, (2) quantity-flexibility contracts, and (3) multiple sourcing. We use a multiplicative demand process to model the evolutionary dynamics of demand uncertainty. We then quantify the impact of key modeling parameters for each operational-flexibility strategy on the waste ratio, which is measured as the ratio of excess inventory when a certain operational-flexibility strategy is employed to the amount when an offshore supplier is utilized without any operational flexibility. We find that the lead-time reduction strategy has the maximum capability to reduce waste in the sourcing process of buyers, followed by the quantity-flexibility and multiple-sourcing strategies, respectively. Thus, our results indicate that operational-flexibility strategies that rely on the localization of production are key to reducing waste and improving environmental sustainability at source.

“Republished with permission of Elalem, from [Why Do Companies Need Operational Flexibility to Reduce Waste at Source?, Elalem, Biçer, and Seifert]; permission conveyed through Creative Commons Attribution 4.0 International (CC BY 4.0).”

1 A machine learning-based framework for forecasting sales of new products with short life cycles using deep neural networks

This chapter is based on Elalem YK, Maier S, Seifert RW. A machine learning-based framework for forecasting sales of new products with short life cycles using deep neural networks. *International Journal of Forecasting*. 2022; doi: <https://doi.org/10.1016/j.ijforecast.2022.09.005>.

Republished with permission of Elsevier Science & Technology Journals, from [A machine learning-based framework for forecasting sales of new products with short life cycles using deep neural networks, Elalem, Maier, and Seifert]; permission conveyed through Copyright Clearance Center, Inc. ("CCC"). (license ID: 1313291-1).

1.1 Introduction

Demand forecasting is a crucial element of supply chain management and is becoming increasingly important as firms bring new products with short life cycles to the market more frequently. While forecasting demand for existing products has improved steadily over the past decades, forecasting sales of newly launched, short-lived products remains challenging due to the lack of abundant historical sales data (van Steenbergen and Mes, 2020). Furthermore, with shorter product life cycles (PLCs), the importance of accurate forecasting early on after a product is launched increases substantially (Basallo-Triana et al., 2017).

According to a cross-industry survey (Cooper and Edgett, 2012), new products contribute to an average of 27% of firms' revenues, yet profits from these products lag behind revenues due to the high costs associated with product introduction, partly caused by the difficulty in forecasting sales of new products compared with more stable ones that have been selling regularly in the market (Cecere, 2013). To increase profits from new products, it is crucial to generate more accurate post-launch forecasts.

Several studies show that sales of existing products are forecasted using statistical methods

(Fildes and Goodwin, 2007). The key challenge for new products is the limited availability of historical data, thus preventing the use of more traditional statistical forecasting methods (Burruss and Kuettner, 2003). The dominant forecasting methods for new products are therefore market research, managerial opinions, and sales force input (Kahn, 2002). Although these techniques may be viable for forecasting sales of a new product, quantitative methods have the potential to substantially outperform these methods based on judgment (Sanders and Manrodt, 2003). Here, we seek to exploit the untapped potential of quantitative methods by deploying a range of state-of-the-art techniques in a novel way to forecast new product demand in data-scarce situations.

We develop a framework based on existing techniques that enables practitioners to apply quantitative methods to forecast sales of new products with short life cycles that are similar to previous products. An overview of our proposed framework is shown in Figure 1.1. The

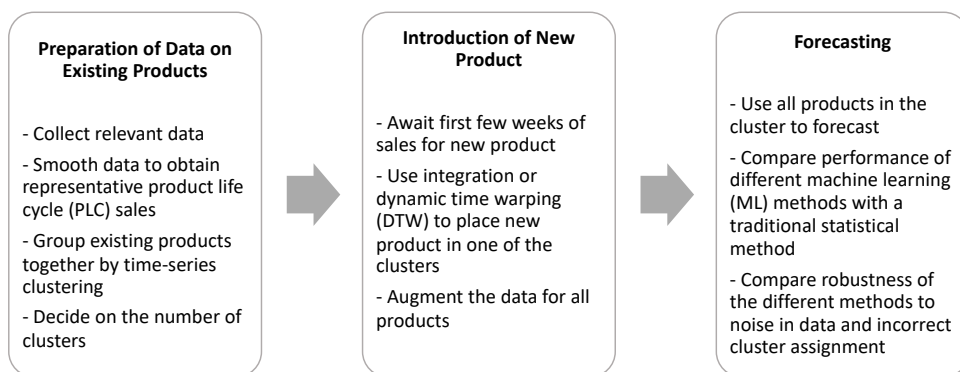


Figure 1.1: Main steps of proposed framework for forecasting sales of new products with short life cycles

framework builds on an approach widely used in research and industry: identify older, similar products to the new product, average their historical sales, and finally use the average sales as a base forecast (Baardman et al., 2018). The three main steps involved are: (1) prepare the sales data for existing products by smoothing the sales over their life cycle to obtain representative PLC sales, and then group similar products by means of clustering; (2) assign the new product to one of the clusters of existing products based on the first few weeks of new product sales using one of the two considered quantitative methods — integration and dynamic time warping (DTW); (3) perform data augmentation on all smoothed existing product sales in the cluster chosen and on the smoothed first few weeks of sales of the new product, then use the data to forecast sales over the rest of the new product's life cycle, applying both statistical and machine learning (ML) methods as well as PLC shape-based methods, and, finally, compare the results under different conditions, including an analysis of the robustness of the quantitative methods to both white noise and an incorrect cluster assignment.

At this point it should be noted that while one sales data point of a new product would already

be enough to apply the proposed forecasting framework, the more sales data available, the higher in general the forecasting accuracy. At the same time, it is also important to have older, similar products for the algorithms to train on. It should also be noted that our framework probably works best on products that sell for at least a couple of months or years before they are replaced by newer technology, such as those found in the electronics industry, including, but not limited to, phones, laptops, cameras, headphones, and speakers. Both data sets used in our study are those of personal computers. Other examples of suitable product categories include fashion products (not fast fashion), books, and movies. Although it could in principle be applied to new products with no similar old products or to fast fashion products that have only a few weeks of sales in the market, our developed forecasting framework should be expected to be less accurate in these situations.

This paper makes several contributions to the literature and practice. We propose a fully quantitative framework for forecasting sales of newly launched, short-lived products that are similar to previous products. More specifically, we describe two quantitative methods – integration and DTW – to position new products with short life cycles in clusters of similar products, rather than relying only on managerial opinions widely used in practice. Furthermore, we demonstrate how data augmentation, a common approach in ML, can be used to generate more data for quantitative forecasting methods. In particular, we use simple interpolation for data augmentation to properly define the statistical and ML models when used with limited data. We apply four different quantitative forecasting methods: one traditional statistical – autoregressive integrated moving average with exogenous variables (ARIMAX) –, and three advanced deep neural networks (DNNs) – long short-term memory (LSTM), gated recurrent units (GRUs) and convolutional neural networks (CNNs). In addition, we apply four PLC shape-based methods, including the well-known Bass model.

To operationalize the proposed framework and to evaluate the comparative performance of the forecasting methods considered, we use two distinct data sets: a publicly available Dell data set that comprises customer orders for 170 complete PLCs, and a second, much larger data set – from Retailer X¹ – that includes complete PLC order history for 843 products. The inputs of the quantitative forecasting methods are the first few weeks of sales of the product being forecasted and, depending on the type of analysis, the weekly time-series that consist of chronological sales data (i.e. weekly customer orders) of all products in either a cluster or the entire data set. Lastly, we introduce additive white Gaussian noise as our basic noise model to test the robustness of the quantitative methods to noise in the data (see Figure 1.2), and we intentionally assign new products to incorrect clusters to analyse the resulting impact on forecasting results.

This paper is organized as follows: Section 1.2 covers the relevant literature related to forecasting sales of new products. In Section 1.3 we describe the two data sets used in this work and the necessary data preparation. Section 1.4 presents our overall approach to time-series

¹Retailer X is the disguised name of a real international electronics retailer. We have used “X” instead of the retailer’s actual name for confidentiality purposes.

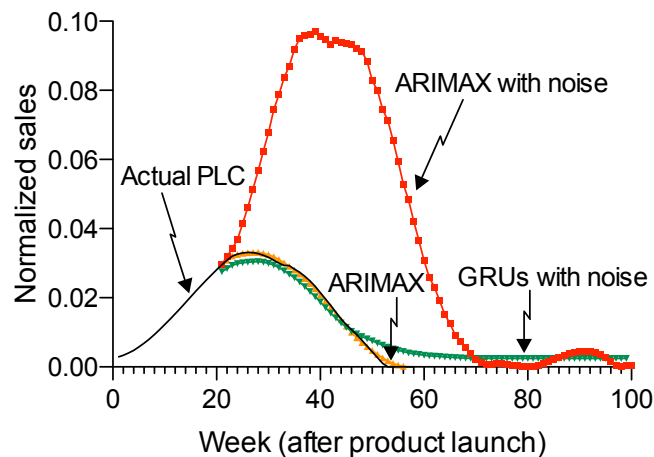


Figure 1.2: Comparison between ARIMAX with and without noise (SD of 0.01) and GRUs with noise after 20 input weeks (Dell data)

clustering, the two cluster assignment methods and a comparison between them. In Section 1.5 we describe the data augmentation step, the different forecasting techniques used and the computational implementation. Section 1.6 presents and discusses the forecasting results for the different methods. Section 1.7 analyzes the robustness of the quantitative methods under additive Gaussian white noise and under incorrect cluster assignments of products. Finally, in Section 1.8 we provide some concluding remarks.

1.2 Literature review

1.2.1 Product life cycle

The notion of product life cycle (PLC) was introduced in 1950 in a *Harvard Business Review* article, “Pricing Policies for New Products”, by Joel Dean. The main idea behind the PLC is that the invention of a new marketable product is first followed by a period where markets are still hesitant to try it, but the product is adopted over time and accepted in the market. Competitor encroachment then arises, and innovations narrow the gap of uniqueness between the product and its substitutes, turning the product into a commodity (Dean, 1976).

In the 1960s, the economist Raymond Vernon developed a more formal model known as Product Life Cycle Stages, or International Product Life Cycle (Vernon, 1966). According to Vernon, each product has a distinctive life cycle that starts with the product’s development and ends with its decline. In general, PLC refers to primarily non-durable consumer goods and represents the sales curve from the time a product is introduced until it is removed from the market (Rink and Swan, 1979). It is usually pictured as a bell-shaped curve (similar to Figure 1.2), and divided into several stages — introduction, growth, maturity, and decline —

with each stage's lifetime ranging from a few weeks to several decades (Vernon, 1966).

Subsequently, many studies on different aspects of the PLC have been published. For example, Kurawarwala and Matsuo (1996) focused on forecasting and inventory management for short life-cycled products. Qi-zhi (2007) demonstrated the application of a diffusion model to increase forecasting accuracy for new products with short life cycles. Interestingly, as in the works cited, most of the research on short life-cycled products is illustrated by consumer electronics data, as they are representative of this class of products.

1.2.2 Quantitative models for forecasting sales of new products

Several growth models have been suggested for forecasting sales during the life cycle of a new product. The most commonly used product growth models are diffusion models. They are a set of stochastic modeling techniques that capture the life cycle dynamics, estimate demand for new product categories, and direct important strategic choices in the pre-launch, launch, and post-launch phases (Kahn, 2014). One of the best known diffusion models was developed by Bass (1969) to estimate how long it will take customers to start purchasing a new product, considering that sales of the product grow to a peak after introduction, before decreasing and leveling off.

Since this early work, several reviews of extensions of diffusion models have been published. Meade and Islam (2006) discussed the importance of modifications of diffusion models to include exogenous variables related to both consumers and the market, and reviewed the works already done in the field. One of the future research directions they identified was forecasting new product diffusion with little or no data. Subsequently, Peres et al. (2010) reviewed works that focus on heterogeneity in customers' willingness to pay and different customer interactions that act as drivers to growth models.

One disadvantage of diffusion models is that, in general, they depend on historical data, which limits their usefulness for forecasting new technology sales (Lee and Lee, 2017). This is the case with the Bass model, where the parameters are usually estimated, which becomes difficult in situations where little is known about how the product will perform in the market. This limitation has resulted in many researchers attempting to improve the Bass diffusion model by making it more adaptive over time as more information becomes available (Lee and Lee, 2017).

In particular, Bayesian updating, which revises the parameters estimated by the Bass model once sufficient sales have been recorded in order to provide better forecasting results, has been integrated in several studies. For example, Zhu and Thonemann (2004) developed an adaptive forecasting algorithm of the Bass model. Their algorithm uses structural information about the PLC to model demand of products with short life cycles by combining the knowledge available before launch with the actual demand that becomes known when the product is released to the market. This enabled them to improve demand forecasts by continually updating the

shape parameters of the Bass model.

1.2.3 Machine learning for forecasting sales of new products

Mišić and Perakis (2019) reflected on how research in operations management has shifted from model-based approaches to data-driven analytical approaches that use data to create models, rather than applying known models to the data available. For example, Yildiz et al. (2017) showed that artificial neural networks with Bayesian regulation backpropagation provide the highest forecasting accuracy when compared with other methods used for electricity load forecasting, which is an example of a field with sufficient data to apply such algorithms. Lu and Kao (2016) used K-means clustering to obtain different clusters of similar products. The authors related the sales data of the product they were forecasting to one of the clusters using different types of linkage methods, and then constructed an extreme learning machine model for that cluster. Their approach was applied to products with sufficient historical data and showed superior performance to forecasting without clustering. More recently, Petropoulos et al. (2022) provided a state-of-the-art overview of a wide range of forecasting methods, and also discussed advancements in new product forecasting.

The use of ML models in new product forecasting comes either through unsupervised learning by clustering, or through supervised learning, which is applied after clustering. Techniques such as analogous forecasting may be used to overcome the lack of sufficient historical demand data (Meade and Islam, 2006), but often past similar products are grouped together by means of clustering and then the products in a cluster are used as a base for forecasting sales of a new product. For example, K. Hu et al. (2018) used different types of clustering – feature-based, category-based, and data-driven – to group similar past products. Then, using managerial opinions, they placed a ready-to-launch new product in a particular cluster and applied traditional curve fitting to the average sales data in a cluster to forecast an entire PLC before product launch. By contrast, Thomassey and Happiette (2007) employed neural networks for clustering and classification. The authors obtained prototypes for each cluster, which is the mean life curve of all products in a cluster, and classified a new product using its descriptive characteristics and then used the prototype of that cluster as its sales forecast.

Other relevant works focused on clustering and then using ML methods for time-series forecasting. This allows the ML methods to be applied in a cross-learning manner, which means that they learn from multiple time-series in a cluster to forecast one individual series (Makridakis et al., 2020). For example, Fallah Tehrani and Ahrens (2016) used a probabilistic approach to classify fashion products depending on their sales, and then used a kernel machine approach to predict the units of sales of the new product. Basallo-Triana et al. (2017) proposed an analogue-based demand forecasting model for one-step forecasting by integrating multiple regression with fuzzy clustering. More recently, van Steenbergen and Mes (2020) developed a hybrid method combining K-means clustering, Random Forest, and Quantile Regression Forest to forecast demand of new products (within 18 weeks of introduction) prior to product

launch.

The research most closely related to ours is the recent work of K. Hu et al. (2018), who have used different types of clustering and managerial opinions for cluster assignment. In this paper, we perform a similar clustering step to prepare the sales data for quantitative forecasting methods. However, instead of relying on managerial opinions, which are widely used in practice, we apply two quantitative methods — integration and DTW — and assess their performance in positioning a new product in its respective cluster. We then perform data augmentation, which is a common approach used in ML to generate artificial data points to reinforce the training of ML algorithms (DeVries and Taylor, 2017). To the best of our knowledge, we are the first to apply data augmentation to enhance the limited historical data of newly launched products in order to enable the use of quantitative forecasting methods.

Although K. Hu et al. (2018)’s study used clustering, which is an ML algorithm, they do not proceed to forecast sales of the new products using ML methods. They apply traditional curve fitting to forecast an entire PLC before the product launch. In this work, we use four different quantitative methods — one traditional statistical and three DNN methods — to forecast sales after the new product’s launch date (i.e. after the first demand is realized). Complementing their approach, we evaluate the comparative performance of the different methods under different conditions. Unlike K. Hu et al. (2018), we report actual errors here rather than relative errors based on proprietary forecasts, making our work directly replicable and, as such, comparable with future studies.

Our work is also somewhat related to Szozda (2010), who proposed a forecasting method for newly launched products with initial post-launch sales data, similar to the situation in our study. They compared the newly launched product sales to the time-series of older products. However, the authors do not apply ML methods for sales forecasting, but use the time-series of the closest older product by adjusting its demand patterns both in scale and length.

1.3 Data

1.3.1 Data sets

We use two distinct data sets for this study: the first is from Dell, and the second is from Retailer X (anonymized name). Dell is the third largest producer of personal computers globally (Hamilton and Webster, 2012). The Dell data set is publicly available and consists of make-to-stock (MTS) products spanning multiple product categories such as fixed workstations, laptops, and desktops. Importantly, this data set was also used by K. Hu et al. (2018), and Acimovic et al. (2018) provide further detail by describing the pre-processing steps taken to clean the data and render it ready to use.

The data consists of weekly customer orders for 170 stock-keeping units (SKUs) of personal computers that completed their life cycle from 2013 to 2016 (Acimovic et al., 2018). All the

data is normalized and associated with the North American market. For our analysis, we use the data after filtering out the cancellations and configure-to-orders (CTOs). However, we consider all sales without truncating the end of the life cycle, as we are interested in how the product life cycles evolve even with external forces arising from managerial decisions such as promotions (refer to Acimovic et al. (2018) for more details). This allows us to forecast the life cycle of a product when it is in a market environment, where the PLC shape is generally affected by such factors.

Furthermore, all sales are normalized to a lifetime cumulative value of 1 by dividing the weekly customer orders per product by its total lifetime sales volume. This means that, for example, a value of 0.1 indicates that 10% of the total sales of a product occurred during that week and a value of 0.5 indicates that 50% of the total sales of a product occurred in that week. Normalizing sales is crucial, since it defines the pattern of sales regardless of the actual quantities, thus making the models used less sensitive to different sales magnitudes. The volumes can then be adjusted depending on how the sales evolve. Several ML models train more efficiently in the presence of normalized data when the data has different ranges (Singh and Singh, 2020). Figure 1.3 (a) displays the full time-series of all the 170 Dell products, i.e. the normalized weekly sales for each Dell product, along the selling horizon available and highlights three different PLCs selling in the beginning (in red), middle (in blue), and towards the end of the horizon (in green).

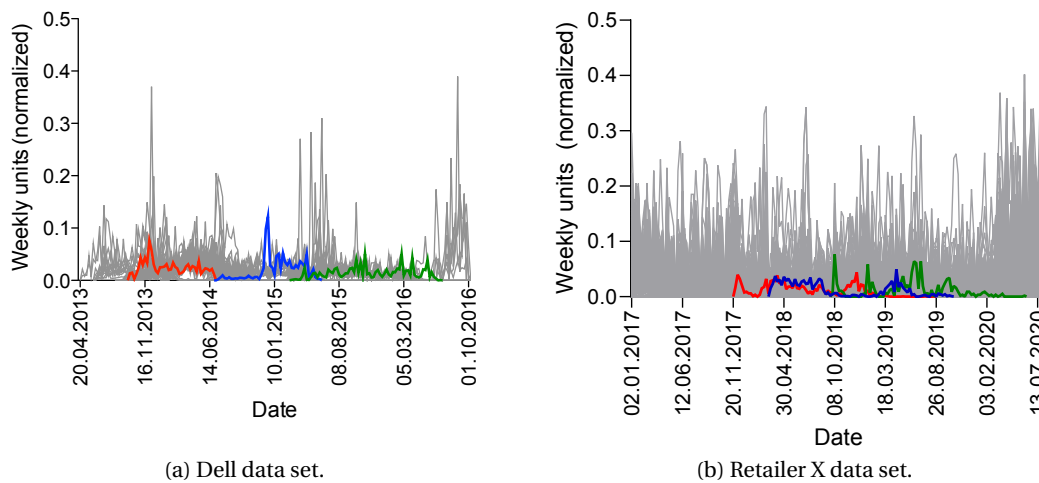


Figure 1.3: Time-series of the two data sets used in this study, with the red, blue and green lines representing weekly normalized sales (before smoothing) for three randomly selected SKUs of each data set.

The second data set is from Retailer X, an international electronics retailer with both physical stores and online sales. The data set consists of normalized weekly sales data of 843 SKUs of personal computers that completed their life cycle from 2017 to mid 2020 in one market. The life cycle of a product in the data set starts when the first purchase of the product occurs and

ends when no more purchases are made. The computers comprise different brands such as Acer, Asus, Apple, HP, and Dell, with different models of the same brand differing by features such as processors and storage capacity. The data has been pre-processed in a similar way to the Dell data set and is also normalized to 1 unit of life cycle sales per SKU. The range of products in the Retailer X data set includes fixed workstations, laptops, and gaming computers, making the product portfolio similar to that of Dell.

The main differences between the two data sets is that Retailer X has almost five times as many SKUs as Dell and it comprises different brands of computers including Dell. The PLC lengths in the Retailer X data set differ greatly compared to the Dell data set, and the PLCs of Retailer X do not necessarily follow the classical PLC shape as some of the products have two maturity stages instead of just one, as exemplary highlighted by the blue PLC curve in Figure 1.3 (b), which shows the full time-series of all the 843 Retailer X products. Although most of the methodology and results shown in our work are for the Dell data, clustering and forecasting are also carried out on Retailer X's data to further demonstrate the consistency of the results obtained by our proposed framework and to strengthen the comparative performance evaluation.

1.3.2 Data preparation

Before clustering the products, it is necessary to have them all aligned at the same start date and to smooth the sales in order to remove displacement effects and noise from the data, respectively. All products are first shifted to a start date of "0" and then the normalized weekly sales for each product are smoothed to obtain the underlying representative PLC, thereby achieving better results for the clustering and forecasting. Using normalized sales is important, since it groups products with similar PLCs together according to PLC shape, rather than sales volumes. Although it can be argued that shifting all PLCs to the same start date ignores the effect of seasonality, we are dealing with short product life cycles lasting for a few months to a few years in the market, so the effect of seasonality may not necessarily be apparent. It should be noted that seasonality effects and trends in the market can be accounted for by adjusting sales projections according to the launch data of the new product.

In order to smooth the PLCs, we applied a smoothing filter that fits polynomials throughout the data points for each time-series. Note that negative values returned by the smoothing filter are set equal to zero because using negative values for smoothed sales would indicate that we account for phenomena such as customer returns and it would lead to negative training data (for the forecasting models), which is outside the scope of this paper. We chose a polynomial of *3rd* degree as it results in each PLC shape being close to a normal curve (see Figure 1.4). This is being supported by a large body of research on the rate of sales of new products with short life cycles which has found that PLC shapes tend to follow normal distributions, with a peak at 50% penetration (for example, see Rogers (1962), Mahajan et al. (1990), and Golder and Tellis (2004)).



Figure 1.4: PLC curve fit by 3rd degree polynomial (SKU150, Dell data)

1.4 Clustering

In this section, we cluster the PLCs of all existing products into different groups depending on similarities in sales patterns and life cycle lengths. We then assign the newly launched product to a group of similar existing products that have already been sold in the market and use the sales data of that group's products as input for the quantitative forecasting techniques. In contrast to the approach of K. Hu et al. (2018), we do not normalize the PLC selling times. This is important as it allows us to forecast more accurately the expected time a product sells in the market, which is critical in order to decide, for example, on when to release new products.

In our study, we apply data-driven, time-series clustering. This type of clustering is capable of identifying hidden product attributes possibly unknown to demand planners and not represented in the raw data itself (K. Hu et al., 2018). More specifically, we deploy agglomerative hierarchical clustering. The main reason for this choice of clustering algorithm is that hierarchical clustering does not only form groups of similar products but it also provides a graphical representation of the data, thereby making the choice of the number of clusters easier and more intuitive for practitioners (Özkoç, 2020) (see online Appendix A.1 for a graphical representation of results). Moreover, Javed et al. (2020) recently presented a time-series clustering benchmark based on 112 data sets using eight popular clustering methods, including agglomerative hierarchical clustering, and found that no method outperformed the others for all data sets. We have also tested several linkage methods: *single*, *complete*, *average*, and *ward*. Since the latter gave the best clustering results, we chose the *ward* linkage method for clustering. The results of the different linkage methods are presented in Appendix A.1 in the supplementary material.

1.4.1 Overall approach

Our overall approach for clustering is as follows:

1. Shift all time-series to the same starting date “0”; otherwise, the clustering step groups the products according to their start dates.
2. Smooth all individual PLCs so that the clustering step groups the products according to the underlying PLC shape and is not affected by weekly sales fluctuations.
3. Apply hierarchical clustering, a data-driven time-series clustering technique.
4. Choose the number of clusters by referring to the dendrogram generated from the clustering step (or from the scree plot).

The result of the clustering step is a dendrogram (see online Appendix A.1) that separates the products into different clusters using the Euclidean distance between each pair of PLCs. The variables used to calculate the Euclidean distance are the smoothed sales points of each time-series. For each pair of products, i and j , the Euclidean distance between their smoothed PLCs, $ED(\tilde{D}^{a,i}, \tilde{D}^{a,j})$, is calculated as follows:

$$ED(\tilde{D}^{a,i}, \tilde{D}^{a,j}) = \sqrt{\sum_{t=0}^n (\tilde{D}_t^{a,i} - \tilde{D}_t^{a,j})^2}, \quad (1.1)$$

where $(\tilde{D}_t^{a,i}, \tilde{D}_t^{a,j})$ is the pair of smoothed sales at time t , and n is the entire length of the time-series.

Note that the Euclidean distance biases the clustering results when PLCs are not aligned to the same starting date. This is because products are clustered predominantly based on their launch dates rather than similarities in sales patterns. So shifting products to the same starting date allows similarities in sales patterns to be the dominant clustering feature. This is important for analogous forecasting, which uses sales information about similar past products to forecast sales of a new product. However, when the time-series are all shifted to the same starting date, the forecaster must pay attention to any seasonality and trend effects of the old time-series and then consider adding similar effects to the newly generated forecasts depending on the selling season.

As the threshold distance between the clusters increases, the number of clusters needed to separate the SKUs decreases, until all products are eventually grouped into one big cluster. The number of clusters can be determined either visually from the dendrogram by determining a horizontal line that cuts through the number of clusters needed, or by producing a scree plot. These curves illustrate the percentage variance in each cluster and how this variance

decreases as the number of clusters increases. We chose four clusters, since this results in a meaningful separation in the dendrogram and explains almost 80% of the variance within each cluster of the Dell data set. Figure 1.5 shows the result of time-series clustering for the Dell data set; the four clusters are distinguished by different PLC behavior, peak sales amounts (note the difference in the y-axis scale), and life cycle lengths. We also chose four clusters for

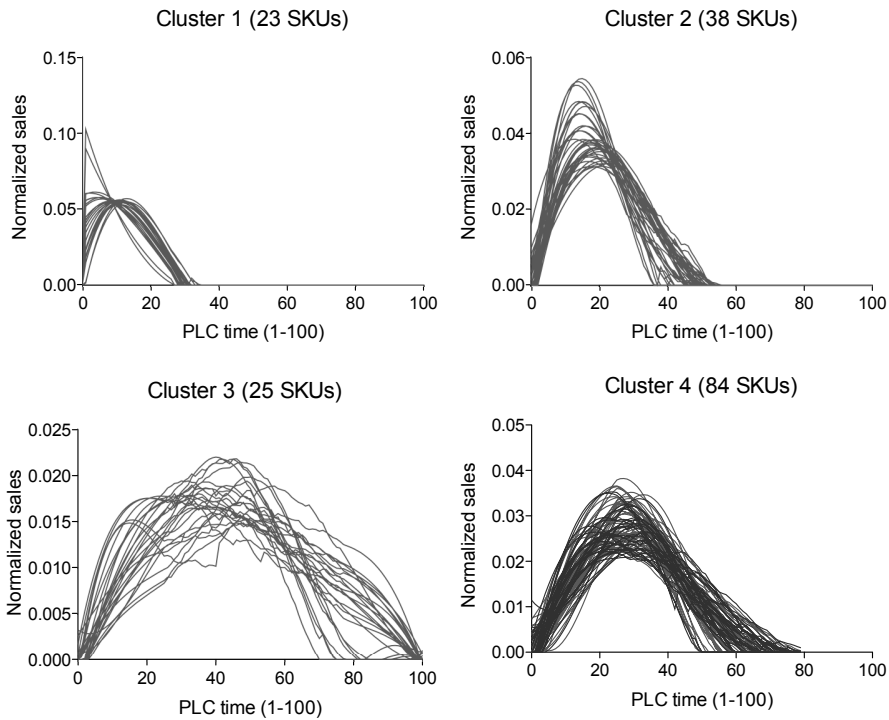


Figure 1.5: Clustering results with individual curves both normalized and smoothed (Dell data)

the Retailer X data set and report all the clustering steps in online Appendix A.1, which reveal not only the larger number of SKUs but also the higher variance in the time-series curves of our second data set. In Appendix A.2 of the supplementary material, we illustrate the results of correlation heat maps between SKUs in each cluster for both data sets to show the similarity between the different PLC curves.

It is important to note that clustering was first tested without the smoothing step. However, the results of the dendrogram were poor as the products were not separated according to their underlying sales patterns. Therefore, the smoothing step is deemed necessary since it removes noise in the data, which may adversely affect clustering.

1.4.2 Assigning products to clusters

After the clusters have been established, we present two quantitative methods — integration and dynamic time warping (DTW) — to assign a new product to one of the four clusters within a few weeks of its launch. In industry, managers typically relate new products to old ones using their judgment (K. Hu et al., 2018). By contrast, we apply two quantitative methods that do not rely on managers' judgment but, in practice, can be used to support managers' opinions and verify their choice.

To test the effectiveness of the two methods in assigning a new product to the correct cluster, we first remove the product we want to test completely from the data set. With the remaining curves already assigned to one of the four clusters, and knowing which cluster the “new” product belonged to before removing it, we can proceed to evaluate the two methods. Specifically, we apply the following five-step approach:

1. Generate an average representative curve for each cluster by calculating the average weekly sales of all products in a cluster at each time-step. This substantially decreases the computational time needed, since the newly launched product is compared with four representative curves, rather than with all the remaining curves in the data set.
2. Wait for the first few weeks of sales of the new product.
3. Use integration or DTW to position the new product in one of the clusters, giving the mean absolute error (MAE) between the new product and each cluster.
4. Assign the new product to the cluster with the lowest MAE.
5. Repeat steps 1-4 for three different groups of 15% of the data set (a percentage commonly used for validation) for each assignment method, and report the average percentage of correct assignments per method for different weeks of sales after product launch. Starting from one week we test up to 20 weeks since some products, especially those in clusters 1 and 2, have entire life cycles of just 30 weeks.

For steps 3 and 4, we repeated the measurements using the mean squared error (MSE) and compared the error measures obtained to those found using the MAE measure but did not find notable differences. The reader is referred to online Appendices A.3 and A.4 for a description of the integration and DTW methods, respectively, and a comparison of them using MSE can be found in online Appendix A.5. In the following section we compare their performance in correctly assigning new products to their respective clusters.

1.4.3 Comparison between integration and DTW

To compare the effectiveness of using integration versus DTW in assigning a new product to its correct cluster, we measure the percentage of correct assignments per method for

different weeks after a new product is launched. Figure 1.6 illustrates the percentage of correct assignments resulting for both methods.

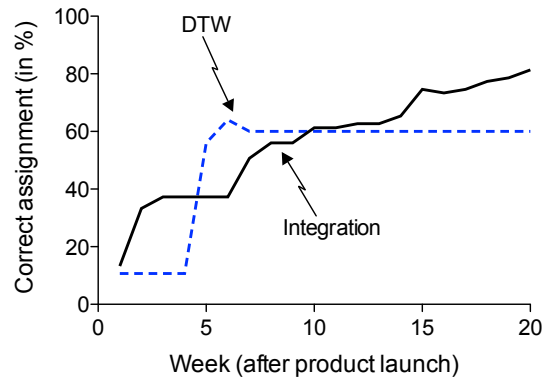


Figure 1.6: Percentage of correct cluster assignment achieved by DTW and integration for various weeks after product launch (Dell data)

The results show that integration follows a more gradual increase in correct assignments than DTW as more weeks of sales data of the new products are available. Because all the data has been shifted to the same start date, and because the PLCs follow a very similar pattern, albeit with different shape parameters, the power of DTW is somewhat limited for this data. As the number of weeks after product launch increases to five, DTW can find some differences between the clusters, but after that, the similarities become too high. As a result, DTW assigns the products randomly, giving the same average percentage of correct assignments even if more weeks are available. Interestingly, the small peak at seven weeks after product launch suggests that at this point in time, the difference in shape between the different clusters is detected; however, beyond this point the clusters continue behaving in a similar way, decreasing the assignment accuracy.

It is important to recall that for DTW, even if there are transformations in amplitude or period of the curves, the algorithm considers them to be similar, unlike the integration method, which uses these transformations to detect differences in the rate of sales. We use the Euclidean distance as a similarity measure for clustering. Since the Euclidean distance measures the similarity between each pair of points at the same time-step between two curves, curves will be grouped together if they have similar slopes. Integration in this case gives better results in assigning products to the correct clusters as it gives a measurement of the rate of sales of the products, which is represented through the slope measures. As more weeks pass, the rate of sales of the new product (i.e., the criterion that integration relies on) becomes more similar to the representative cluster, and that is why integration generally produces better results. For example, 15 weeks after the new product's launch, it is apparent from Figure 1.5 that if the product is in the decline phase at that point, it belongs to cluster 1; if it is in the maturity phase, it belongs to cluster 2; and if it is in the growth phase, it either belongs to cluster 3 or 4. This also explains why there is a bump in the percentage of correct integration assignments at

15 weeks after product launch (Figure 1.6) as the clusters are easily distinguished at that point in time.

When repeating the clustering using DTW as the distance measure instead of the Euclidean distance and then assigning new products to clusters, we find that DTW outperforms integration in cluster assignment (see online Appendix A.6 in the supplementary material). By using DTW as a distance measure, products are clustered based on similarities despite temporal shifts, rather than based on slopes and evolution of sales. Therefore, if two time-series are similar in shape but have different slopes and sales evolution, DTW-based clustering would group them together, while integration-based clustering no longer provides a good performance. As a result, cluster assignment using DTW will start to assign products more correctly. However, due to the nature of DTW, it may be argued that using it for only parts of curves (that is, few weeks of sales after introduction) may be unsuitable. Moreover, the dendrogram results from using the Euclidean distance as a measure show greater separation between the clusters than using DTW as a distance measure, as shown and discussed in online Appendix A.6.

1.5 Forecasting

In this section, we first describe how we enhance the data to make it suitable for the use of quantitative methods, before testing and comparing a traditional statistical forecasting method with three different types of deep learning ML methods for sales forecasting. See online Appendix A.7 for details on implementation and parameter estimation of the quantitative methods. We apply the methods using smoothed sales, with and without the clustering step. Doing so allows us to verify the need for the clustering step by evaluating its importance in the forecasting results. As a benchmark for the shape-free quantitative methods in our comparative performance analysis, we apply three forecasting methods that are based on fitting families of curves to historical PLC shapes. Importantly, for all the methods applied here, we assume at this point that we know to which cluster the new product belongs. The input of the quantitative forecasting methods is in the form of weekly time-series (of chronological sales history) and consists of both the first few weeks of sales of the new product and the historical sales either of all products in the relevant cluster or, if clustering is not used, of all products in the entire data set.

1.5.1 Data augmentation

To ensure that the quantitative methods have sufficient training data, we generate more data using data augmentation (DeVries and Taylor, 2017). This step is necessary when the algorithms require abundant data points to perform well but the training data available is very limited, as in our case with newly introduced products with short life cycles. The PLCs in our data sets have between 30 and 100 data points, meaning products in our data sets experience at most 100 weeks of sales. We generate 9 additional points between each pair of sales by interpolation, before using the quantitative forecasting methods. We chose 9 additional points

because this increases the forecasting accuracy while resulting in an acceptable running time, as described in the following paragraph. These additional points fall on the curve of each product's sales, reinforcing the shape of the sales curve yet not affecting the weekly sales volumes. We perform data augmentation for all the curves input into the forecasting algorithms. However, after using these augmented points to support the training of the algorithms, we disregard them and consider only the actual points of sales when calculating the forecast errors.

By augmenting 9 additional data points, the percentage of original data points to all data points on the augmented new time-series – which consists of original plus augmented ones – is 10%. The choice behind the percentage of actual to augmented data points involves an important tradeoff: While data augmentation may be necessary to properly train ML algorithms and avoid over-fitting with limited data, too much use of data augmentation may lead to under-fitting, which is also non-desirable (Park et al., 2019). To avoid under-fitting, the models would have to be more complex, which, however, increases computational costs. A ratio of 10% represents an appropriate tradeoff between forecasting accuracy and computational time.

Many of the data augmentation techniques for time-series increase the number of time-series for classification purposes. These include: *Jittering*, where white noise is added to some of the time-series to generate new time-series; *scaling*, where for a given time-series the values are scaled by a given factor to generate new time-series; *permutations*, where segments of the time-series are rearranged to produce new time-series; and, *averaging and interpolation*, where a new time-series is created from the combination of two time-series and located between the original time-series (see Iwana and Uchida (2021) for more data augmentation techniques for time-series). However, since our data sets have sufficient time-series for classification, but limited data points per time-series for forecasting, we use a data augmentation technique that increases the amount of data points per time-series, rather than the total number of time-series available.

We note that the data augmentation step is done only for the quantitative methods but not for the PLC shape-based methods. The reason is that the latter rely on curve-fitting, meaning that they optimize fitting as many data points through curves of different parameters, until the optimal parameters are found. So if we used curve-fitting with data augmentation, we would be forcing the curves to fit through the augmented data points, giving less degrees of freedom to fit through the actual data points. Figure 1.7 illustrates the results of data augmentation considering the first few weeks of sales of an SKU in the Dell data set, showing the actual smoothed sales data in red (left) and the same data points with the additional augmented data points in blue (right). After data augmentation, the curve is supported by more curve-defining data points and the shape of the PLC curve is better defined, which enhances the forecasting performance.

Importantly, we ensured that the augmented points are equidistant both from one another and from the original sales data when interpolating these additional data points. The reason

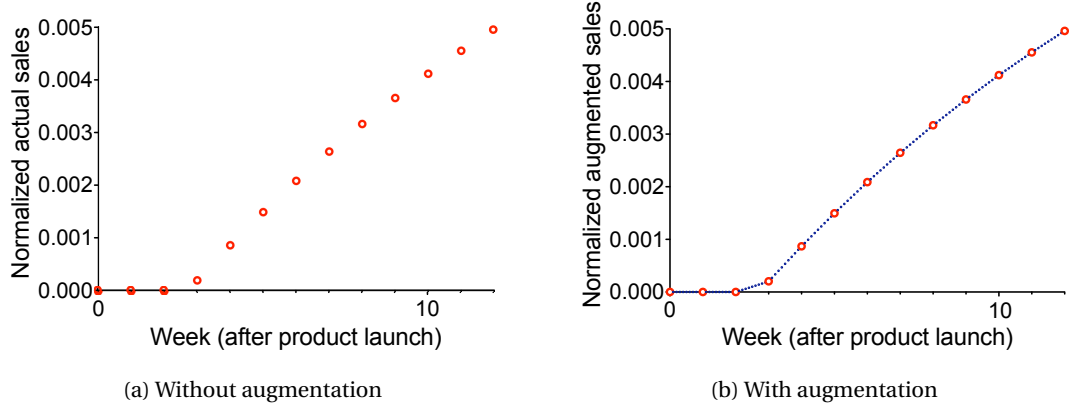


Figure 1.7: Example of data augmentation results on Dell data set (first weeks of SKU 105)

behind positioning all (augmented and original) data points equidistant from one another is that the augmented data points and the original sales data act together as the new input to the forecasting algorithms. If the input data points were not equidistant from one another, it would translate into giving the algorithms data with different x -positions, and therefore different frequencies (e.g. some daily and some monthly data points). This would alter the skewness of the curves as the algorithms consider the time-intervals between the points equidistant from one another, requiring the forecaster to reconcile the different frequencies into one before using the data for forecasting.

1.5.2 Traditional statistical model

We apply a version of AutoRegressive Integrated Moving Average (ARIMA) called ARIMAX as the traditional statistical forecasting method and as a baseline for comparison with the DNN methods. ARIMA is best suited to short-term forecasting for around 12 months ahead (Stellwagen and Tashman, 2013), making it suitable for forecasting sales of short-lived products. ARIMAX allows the use of exogenous (or explanatory) variables (the X in ARIMAX) within the algorithm. Examples of exogenous variables are prices and weather conditions. In our case, the exogenous variables are the smoothed sales values of all (older) products in either a cluster or the entire data set.

Extending the ARIMA model, the mathematical formulation of ARIMAX is given by:

$$y_t^* = \sum_{i=1}^n \beta_i x_t^i + \mu + \sum_{i=1}^p \phi_i y_{t-i}^* + \sum_{i=1}^q \theta_i \epsilon_{t-i} + \epsilon_t, \quad (1.2)$$

where β_i is the coefficient associated with the smoothed sales of product i at time t , x_t^i ; n denotes the number of older products or curves present in the cluster; μ denotes the intercept coefficient; $y_t^* = \Delta^d y_t$ denotes the observations of the target time-series values differenced d times to fulfill the stationary requirements; ϕ and θ are the coefficients of the AR and MA

parts, respectively; and, finally, ϵ denotes the prediction error.

As a general rule of thumb, ARIMA typically requires 50-100 data points as input (Box and Tiao, 1975). In Section 1.6 we forecast with a minimum input of 10 and 6 introduction weeks for the Dell and Retailer X data sets, respectively. Therefore, adding 9 data points between each pair of sales data points augments the 10 weeks of sales to 100 data points and the 6 weeks of sales to 60 data points, making the augmented data set suitable for use with ARIMAX.

To validate our choice of 9 additional data points, we analyze the distribution of the residuals (errors) of the fitted ARIMAX models. If the ARIMAX model is correctly specified, then the residuals should be normally distributed with mean zero. In Figure 1.8, we show an example of the distribution of the residuals considering SKU 10 of the Dell data set. The figure shows

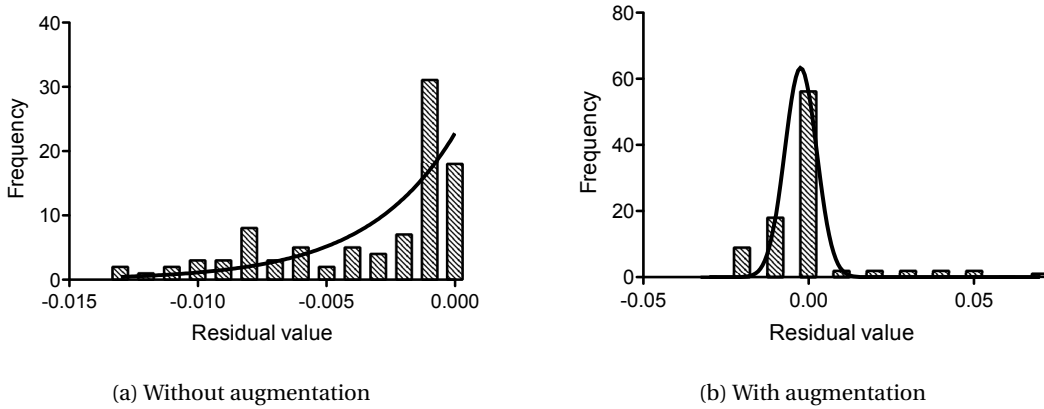


Figure 1.8: Example of histogram of residuals without and with data augmentation (SKU 10, Dell data)

the difference between the distribution before and after the data augmentation step, and how adding 9 data points changes the distribution of the errors, justifying that the augmented data set is appropriate for the model. In Table 1.1, we report the best-fit values for fitting a normal distribution to the two histograms. It can be seen that fitting a Gaussian curve without

Table 1.1: Best-fit values for Gaussian Distribution (SKU 10, Dell data)

Metric	Without augmentation	With augmentation
Mean	0.1761	-0.002387
Standard deviation	0.02453	0.004813
Amplitude	3.548e+012	63.29
R^2	n/a	0.9649

data augmentation does not converge (no R^2), whereas applying data augmentation leads to a sufficiently high R^2 of 0.9646, which confirms that our choice of ninefold augmented data points is appropriate.

1.5.3 Machine learning models

For the ML methods, we apply three different types of DNNs: long short-term memory (LSTM), gated recurrent units (GRUs), and convolutional neural networks (CNNs). The main reason for choosing DNNs is that they, like ARIMAX, can be multivariate, meaning that sales of other (e.g. old) products can be fed into the algorithms when forecasting sales of the new product. DNNs are an extension of single-layered artificial neural networks to multiple layers and have been tested and compared with other time-series forecasting methods in many studies, often giving the best results (see Cao et al. (2012), Zhang et al. (2017), Paliari et al. (2021), and L. Hu et al. (2021)). We have also tested XGBoost but it did not provide meaningful forecasting results for the considered numbers of sales weeks of the new product, so we excluded it from our report. Recently, Kraus et al. (2020) discussed the importance of deep learning in the field of business analytics (including sales forecasting) and argued that DNNs are able to identify previously unknown, potentially useful patterns more accurately than other widely used predictive models such as support vector machines and random forests.

Recurrent neural networks (RNNs) such as LSTM and GRUs take sequential data as input and are comprised of a network $f_{\text{RNN}} : \mathbb{X} \rightarrow \mathbb{Y}$, where the sequences in \mathbb{X} have the form of $\mathbf{x}^i = [x_1^i, x_2^i, \dots, x_\tau^i]$, with x^i representing the input variable, such as sales time-series of one old product or the new product, given τ input week(s). The network also has internal hidden layers, denoted by h_1^i, \dots, h_τ^i . The knowledge of the sequence is accumulated in these hidden states. For a given time $t \in \{1, \dots, \tau\}$, the input to the network is the concatenation of x_t^i and the previous hidden state h_{t-1}^i . The output of the RNN is therefore a computation of the following:

$$f_{\text{RNN}}(x_1^i, \dots, x_\tau^i) = f_{\text{DNN}}([x_\tau^i, f_{\text{DNN}}([x_{\tau-1}^i, \dots, f_{\text{DNN}}([x_1^i; W, b]; \dots]; W, b]; W, b)], W, b), \quad (1.3)$$

where f_{DNN} is a deep layer of neural networks composed of k layers of single neural networks, given by:

$$f_{\text{DNN}}(\mathbf{x}) = \underbrace{f_{\text{NN}}(\dots f_{\text{NN}}(\mathbf{x}))}_{k}. \quad (1.4)$$

Each neural network f_{NN} is computed via an activation function and a linear combination as follows:

$$f_{\text{NN}}(\mathbf{x}; W, b) = \sigma(W\mathbf{x} + b), \quad (1.5)$$

where σ is the activation function, which in our case is the rectified linear unit (*ReLU*) function given by $\sigma(x) = \max(0, x) \in [0, \infty)$; W is the weight matrix; and b is the intercept. We refer the reader to Kraus et al. (2020) for more information regarding the architecture of RNNs and CNNs and the optimization of model parameters.

Long short-term memory. LSTM is a type of RNN that addresses the problem of vanishing

gradients faced by ordinary RNNs (Hochreiter and Schmidhuber, 1997). It does so by learning to bridge time intervals without the loss of short time lag capabilities, allowing the network to *remember* both long- and short-term patterns in the data. The architecture consists of several gated units and memory cells to facilitate information storage over time, in addition to the other layers found in RNNs.

Gated recurrent units. GRUs are the second most common type of RNN behind LSTM. GRUs are similar to LSTM in that they operate through gates to overcome the vanishing gradient problem, but they differ in the type of gates used. GRUs use reset and update gates, whereas LSTM uses input, output, and forget gates. Many studies demonstrate the applications of GRUs, with Kumar et al. (2018) showing that both LSTM and GRUs perform best when used for forecasting.

Convolutional neural networks. CNNs are a type of artificial neural network commonly used in image processing. The aim is to form spatial filters and convolve these filters over each channel in an image (Rudin and Carlson, 2019), before passing the input to the next layer. Therefore, the number of filters in a CNN should be optimized. Although their use mostly involves image processing, some studies have applied CNNs to forecast time-series. For example, Liu et al. (2017) used a CNN-based model to forecast foreign exchange rates and showed that its performance for long-term forecasts outperforms other ML models, even when compared with GRUs. We therefore include it in our study.

1.5.4 PLC shape-based methods

To complement our analysis, we also apply three forecasting methods that are based on fitting families of curves – Bass diffusion, piece-wise linear, and polynomial curves – to PLC shapes. Using these families of curves was proposed by K. Hu et al. (2018), and the methods' mathematical formulation is summarised by Table 1.2.

We use the methods on our two data sets after the smoothing and clustering steps. As in the work of K. Hu et al. (2018), we apply two different approaches to fit the models' parameters: “(a) *Taking the average of similar curves (GenerateAvg)*” for each cluster; and “(b) *Fitting the best curve through the data points (GenerateFit)*” in each cluster. For *GenerateAvg*, we simply calculate the average weekly sales of all the products in a cluster at each time-step (similar to Section 1.4.2). For *GenerateFit*, we generate a representative curve in a similar fashion to K. Hu et al. (2018) by formulating an optimization problem that depending on the curve (bass, polynomial, triangle and trapezoid) finds the parameter values that minimize the sum of squared errors across all products in the cluster over all time-steps.

Note that for the Bass diffusion model, we denote the forecasting results (in Tables 1.3 and 1.4) obtained through the two approaches described above by Bass-I. We then use the Bass model in a different manner by applying it to only one PLC – the PLC of the most similar older product in the cluster – instead of applying it to the average PLC curve per cluster. More

Table 1.2: Summary of PLC shape-based methods

Base curve family	Forecasted demand at time t	Parameters
Bass	$\hat{D}_t^{Bass} = \frac{p(p+q)^2 e^{-(p+q)t}}{(p+q e^{-(p+q)t})^2}$	(p, q) : shape parameters
Triangle	$\hat{D}_t^{triangle} = \begin{cases} at + b & 0 < t < t_1 \\ c(t - t_1) + (at_1 + b) & t_1 \leq t \leq T \end{cases}$	(a, b, c) : shape parameters t_1 : turning point T : life cycle length
Trapezoid	$\hat{D}_t^{trapezoid} = \begin{cases} at + b & 0 < t < t_1 \\ at_1 + b & t_1 \leq t < t_2 \\ c(t - t_2) + (at_1 + b) & t_2 \leq t \leq T \end{cases}$	(a, b, c) : shape parameters t_1 : beginning of maturity stage t_2 : end of maturity stage T : life cycle length
Polyn	$\hat{D}_t^{polyn} = \sum_{j=0}^n a_j t^j$	a_j : shape parameters for $j = 0, \dots, n$

specifically, we first calculate the MSE between the data of the first few weeks of sales of the new product and the first few weeks of each PLC in the cluster. The PLC in the cluster with the lowest MSE value is assumed to have the highest similarity to the new product. After finding the most similar PLC of an old product, we fit the Bass parameters to that PLC and use it for forecasting. Denoted by Bass-II, this approach is inspired by Zhu and Thonemann (2004) in which over time, new products are compared to old products and the parameters of the Bass model are updated accordingly. Importantly, by using Bass-II, the Bass model uses the new product’s sales data available after a few weeks of its launch and compares it to sales of older products, thus taking advantage of the sales information available at that point.

In Figure 1.9, we provide a small example to demonstrate how the Bass-II approach works. Assume we place the ‘new product’ after 20 weeks of launch in cluster 3, which has only two

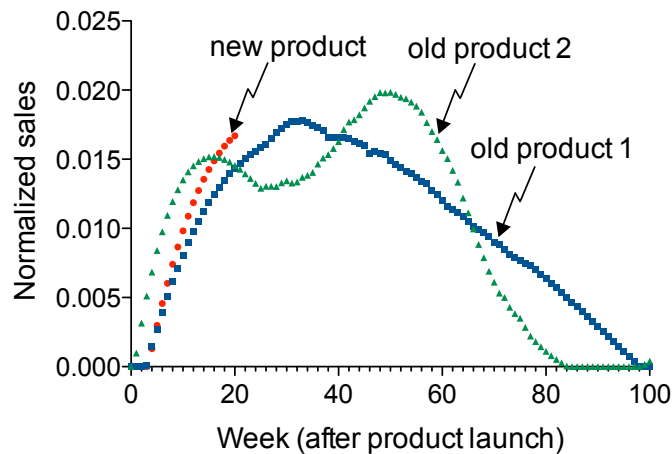


Figure 1.9: Reference to how the Bass-II model works

older products: ‘old product 1’ and ‘old product 2’. Considering the first 20 weeks, the MSE measure between new product and old product 1 is lowest, so the sales of the new product are closest to old product 1. We therefore fit the Bass-II model parameters, p and q , to the sales time-series of old product 1, and use those parameters to forecast the rest of the PLC of the new product.

It should be noted, however, that since the Bass-II model parameters are obtained based on estimates from a single old PLC, the results can be extremely poor if it turns out that the old PLC has a very different sales pattern afterwards. In our case, the old and new PLCs are very similar, so this approach can be used. But if the new product is expected to behave very differently from the old products, this approach should not be considered as it relies on the sales of a single old product during forecasting, and does not benefit from learning sales patterns from different PLCs.

We also tried for the Triangle and Trapezoid to use the additional information by fitting their a and b parameters to the first few weeks of sales data, but this gave poor results since these few data points are, in general, insufficient to estimate these parameters accurately.

For each PLC shape-based method, we calculate the parameters for both *GenerateAvg* and *GenerateFit* twice: (1) with respect to the *GenerateAvg* and *GenerateFit* of each cluster independent of the number of weeks of sales of the new product; and (2) with respect to *GenerateAvg* and *GenerateFit* of each cluster starting from the number of weeks of sales already encountered until the end of the new product’s life cycle. We apply the second strategy to make use of the additional sales information we have available – by truncating the beginning (up to the new product’s sales weeks encountered) of the representative curves generated through the *GenerateAvg* and *GenerateFit* approaches – and report the results of the more accurate parameter calculation strategy. Superscripts 1 and 2 in Tables 1.3 and 1.4 denote which of the two strategies performed best, and the results for both strategies are reported in online Appendix A.8.

1.6 Forecasting results

This section reports the forecasting results on both the Dell and Retailer X data sets, using ARIMAX and the three considered DNNs — LSTM, GRUs, and CNNs — with and without clustering, considering different weeks after the introduction of the new product. It also includes a comparison of the forecast accuracy of the quantitative methods and the PLC shape-based methods. Note that the results in terms of forecast accuracy reported here are calculated between the forecasted sales and the actual (unsmoothed) sales of the product over its life cycle, across three groups of 15% of the data set, equal to 25 SKUs per group. To evaluate the out-of-sample forecast accuracy, we measure the mean absolute scaled error (MASE) and the sum of absolute errors (SAE). These two error measurements are widely and commonly used in the related literature. The errors are calculated per product between the forecasted and actual (unsmoothed) sales at each time step, and are then averaged over the

group of SKUs tested, as described in online Appendix A.9. Note that for both data sets the parameters of ARIMAX and those of the two Bass models are statistically significant at the 5% level. In online Appendix A.10, we show the error distributions over the different SKUs for each forecasting method used.

1.6.1 Dell data set

We start with the Dell data set and evaluate the sales forecast accuracy over the new product's entire life cycle, at 10, 15, and 20 weeks after a new product is launched. We chose 20 weeks as the maximum number of weeks because products in clusters 1 and 2 would already be in their decline stage after that. The forecast errors in Table 1.3 show that ARIMAX achieved the highest forecast accuracy. When compared to the best-performing DNN – either LSTM or GRUs – the forecast errors in terms of MASE (SAE) achieved by ARIMAX are 24.32% (22.86%), 23.64% (37.14%), and 21.10% (39.29%) lower for 10, 15, and 20 input weeks, respectively. On the other hand, when compared with the worst-performing DNN – CNNs –, ARIMAX achieved even higher reductions of 38.69% (47.06%), 34.38% (46.34%), and 37.23% (50%), respectively. Even though ARIMAX without clustering gives the best results, LSTM and GRUs also perform relatively well. However, CNNs exhibited the poorest performance compared with the two RNN methods, in contrast to what was found by Liu et al. (2017) in the context of long-term forecasting of exchange rates.

It can be observed from Table 1.3 that the PLC shape-based methods produce relatively accurate forecasts and, on average, perform better than the DNNs. In fact, when ARIMAX is compared to the best-performing curve-fitting method – either poly2 (Avg) or Bass-II – the forecast errors in terms of MASE (SAE) achieved by ARIMAX are 9.68% (25%), 10.64% (31.25%), and 11.34% (34.61%) lower for 10, 15, and 20 input weeks. This can be explained by the relatively small size of the Dell data set and by the PLCs of the included products, which follow the classical life cycle pattern with well-defined PLC stages.

Figure 1.10 displays the forecasting results of ARIMAX and GRUs for an example SKU with 10 weeks of post-launch sales data. Note that the forecasting is done before clustering for ARIMAX and after clustering for GRUs as this gives the best results. It can be seen that the ARIMAX outperforms GRUs because the forecasted sales points of ARIMAX lie closer to the actual sales curve than those of GRUs. Furthermore, ARIMAX shows a clear end of PLC as the sales drop to zero, while GRUs shows that the life cycle is longer than expected. In Figure 1.11, we illustrate the forecasting results of ARIMAX and Bass-II on another example SKU with 10 weeks of post-launch sales data. As can be seen, the results show that the forecasting performance of ARIMAX is superior to Bass-II, which produces a PLC with lower peak and extended selling period, as opposed to the actual PLC and the results of ARIMAX.

Although K. Hu et al. (2018) did not report actual forecast errors, which would have allowed us to compare our results with theirs, it is evident that the relative performance of the curve families differs slightly. In their study of pre-launch forecasts, even though poly2 performed

Table 1.3: Forecast errors for new products for different weeks after introduction (Dell data)

Quantitative method	Time-series clustering	MASE			SAE		
		10 weeks	15 weeks	20 weeks	10 weeks	15 weeks	20 weeks
LSTM	No	1.14	1.13	1.18	0.43	0.36	0.30
	Yes	1.13	1.13	1.09	0.35	0.35	0.28
GRUs	No	1.16	1.10	1.10	0.44	0.35	0.32
	Yes	1.11	1.11	1.12	0.42	0.35	0.29
CNNs	No	1.37	1.23	1.20	0.51	0.38	0.30
	Yes	1.32	1.28	1.37	0.49	0.41	0.34
ARIMAX	No	0.84	0.84	0.86	0.27	0.22	0.17
	Yes	0.97	0.99	1.06	0.37	0.32	0.26
Base curve family	Cluster curve generation	MASE			SAE		
		10 weeks	15 weeks	20 weeks	10 weeks	15 weeks	20 weeks
Bass-I	Fit ¹	0.99	1.01	1.05	0.38	0.33	0.28
	Avg ¹	0.97	1.00	1.04	0.38	0.33	0.28
poly2	Fit ¹	1.07	1.10	1.18	0.41	0.36	0.30
	Avg ¹	0.93	0.94	0.98	0.36	0.32	0.26
poly3	Fit ¹	1.10	1.14	1.21	0.42	0.37	0.31
	Avg ¹	0.97	0.98	1.02	0.37	0.32	0.27
poly4	Fit ^{1,2}	1.11	1.15	1.23	0.43	0.38	0.32
	Avg ²	0.98	1.00	1.04	0.38	0.33	0.28
Trapezoid	Fit ¹	1.16	1.19	1.26	0.44	0.38	0.32
	Avg ¹	1.11	1.13	1.14	0.42	0.36	0.30
Triangle	Fit ¹	1.12	1.14	1.22	0.42	0.37	0.31
	Avg ¹	1.10	1.14	1.21	0.42	0.37	0.31
Bass-II ^{1,2}	-	1.08	0.94	0.97	0.38	0.34	0.26

Note: Double boxes denote the best values in each column, while single boxes denote the second best values. Superscripts 1 and 2 denote that the best results for the curve-fitting methods were obtained by fitting the parameters to the entire *GenerateAverage/GenerateFit* curves per cluster and only to the part after the introduction weeks, respectively.

well, Trapezoid and Triangle performed best. Unlike K. Hu et al. (2018), we find that the Bass model gives very good results. This difference in relative performance may well be explained by differences in the smoothing step or the fact that we consider the Dell products' entire life cycle length, rather than truncate the end of the life cycles like K. Hu et al. (2018). Also, it is important to note that we do not normalize the selling period and therefore our error calculations penalize the results if the methods do not predict the life cycle length accurately.

It is interesting to see, however, that while all considered methods exhibit improved performance in terms of SAE when more weeks of sales are included, the quantitative forecasting methods do not always benefit from time-series clustering. The forecasts generated by ARIMAX show substantially worse performance when using clustering, whereas the DNNs with

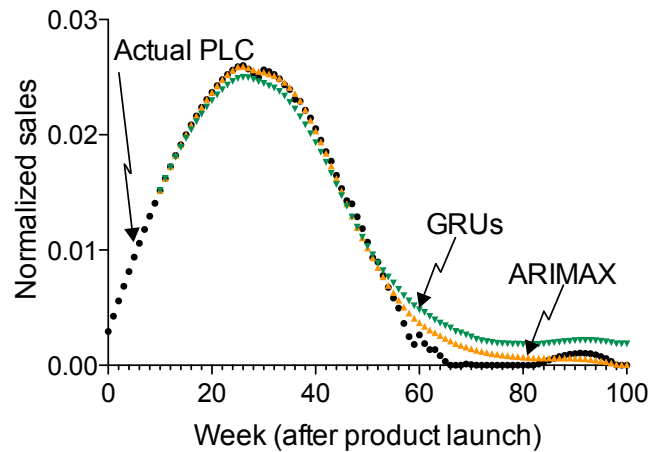


Figure 1.10: Comparison of forecasting results of ARIMAX and GRUs (SKU 112 from the Dell data set, with 10 weeks of introduction)

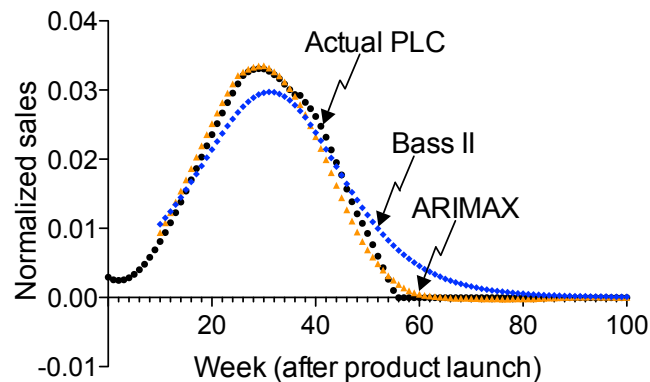


Figure 1.11: Comparison of forecasting results of ARIMAX and Bass-II (SKU 24 from the Dell data set, with 10 weeks of introduction)

clustering have often lower forecast errors (and, hence, higher accuracy), with reductions in MASE (SAE) of up to 7.63% (18.6%), 4.31% (9.38%), and 3.65% (3.92%) under LSTM, GRUs, and CNNs, respectively. The intuition behind this is that after clustering the variance between the sales of products in the clusters is minimized, removing possible noise coming from PLCs that are different and could affect the forecasting accuracy. The reductions are rather modest as the Dell data set is composed of only 170 SKUs, so separating the data into clusters reduces the amount of data available for the DNN algorithms to train. We should therefore expect greater improvements in accuracy by clustering when applying our framework to the Retailer X data set as it is almost five times larger.

1.6.2 Retailer X data set

We now repeat the clustering and forecasting steps for the second data set — Retailer X — to analyse the performance of our proposed framework on a second, much larger data set which not only consists of almost 850 SKUs, but it also features irregular PLCs such as those with two maturity phases. The accuracy of post-launch forecasts is evaluated at 6, 8, and 10 weeks after a new product is launched. We chose 10 weeks as the maximum, since the majority of products have already started their decline stage by then.

The results in terms of forecast accuracy are displayed in Table 1.4 and show that the MASE generally does not follow any trend when more weeks are introduced, whereas the SAE virtually always decreases as more data becomes available. While the latter confirms that the considered

Table 1.4: Forecast errors for new products for different weeks after introduction (Retailer X data)

Quantitative method	Time-series clustering	MASE			SAE		
		6 weeks	8 weeks	10 weeks	6 weeks	8 weeks	10 weeks
LSTM	No	1.40	1.46	1.52	0.50	0.46	0.46
	Yes	1.19	1.19	1.36	0.43	0.39	0.33
GRUs	No	1.41	1.46	1.50	0.52	0.49	0.45
	Yes	1.47	1.18	1.32	0.45	0.39	0.33
CNNs	No	1.65	1.61	1.56	0.57	0.52	0.46
	Yes	1.47	1.38	1.71	0.53	0.44	0.39
ARIMAX	No	1.26	1.53	1.47	0.47	0.44	0.42
	Yes	1.08	1.05	1.22	0.40	0.35	0.30

Base curve family	Cluster curve generation	MASE			SAE		
		6 weeks	8 weeks	10 weeks	6 weeks	8 weeks	10 weeks
Bass-I	Fit ²	1.20	1.20	1.30	0.56	0.50	0.49
	Avg ^{1,2}	1.10	1.17	1.24	0.53	0.49	0.45
poly2	Fit ¹	1.40	1.46	1.57	0.67	0.62	0.58
	Avg ²	1.15	1.40	1.54	0.54	0.50	0.47
poly3	Fit ²	2.11	2.85	3.30	0.99	0.83	0.71
	Avg ²	1.14	1.39	1.54	0.54	0.50	0.46
poly4	Fit ²	1.61	1.82	2.14	0.78	0.71	0.64
	Avg ^{1,2}	1.13	1.38	1.53	0.54	0.51	0.46
Trapezoid	Fit ²	1.29	1.36	1.34	0.61	0.58	0.51
	Avg ^{1,2}	1.29	1.30	1.25	0.58	0.54	0.47
Triangle	Fit ²	1.28	1.36	1.39	0.60	0.55	0.50
	Avg ²	1.20	1.25	1.26	0.56	0.51	0.47
Bass-II ²	–	1.11	1.17	1.24	0.55	0.49	0.47

Note: Double boxes denote the best values in each column, while single boxes denote the second best values. Superscripts 1 and 2 denote that the best results for the curve-fitting methods were obtained by fitting the parameters to the entire *GenerateAverage/GenerateFit* curves per cluster and only to the part after the introduction weeks, respectively.

forecasting methods typically perform better with more data, the improvement is not reflected in the MASE due to the fact that we introduce only two more weeks at a time. It is now more apparent that clustering improves the forecast accuracy of the quantitative methods, with average MASE (SAE) improvements of 14.67% (19.16%), 8.97% (20.18%), 5.19% (12.54%), and 20.89% (21.31%) under LSTM, GRUs, CNNs, and ARIMAX, respectively, which is in line with the best results obtained for the Dell data set for the DNN methods. Due to the increased amount of data for Retailer X, the reduction in forecasting errors over no clustering is much more substantial than for the Dell data set, so clustering can be regarded as an important step, especially when historical PLC data is plentiful.

The forecast errors in Table 1.4 support the results of the Dell data set, with ARIMAX still generating the best results, but the difference between ARIMAX and the three DNNs is now reduced. In fact, when compared with the best-performing DNN – again, either LSTM or GRUs –, the percentage reductions in MASE (SAE) achieved by ARIMAX with clustering are 9.24% (6.98%), 11.02% (10.26%), and 10.29% (9.09%) for 6, 8, and 10 input weeks, respectively. On the other hand, when compared with CNNs, ARIMAX achieved again higher improvements of 26.53% (24.53%), 23.91% (20.45%), and 28.65% (23.07%), respectively. The standard deviation within the time-series clusters of Retailer X is higher than that of Dell’s, meaning the products have less similar PLC curves and the linear dependencies between the PLCs are minimized, which may explain why the forecast errors for ARIMAX are almost always higher now.

It can also be observed from Table 1.4 that although the PLC shape-based methods still perform relatively well, they are generally outperformed by the DNN methods in terms of SAE. Comparing the performance of ARIMAX to the best-performing PLC shape-based method – either Bass-I (Avg) or Bass-II – the percentage reductions in MASE (SAE) achieved by ARIMAX are 1.82% (24.53%) 10.26% (28.57%), and 1.61% (33.33%) for 6, 8, and 10 input weeks, respectively. The improved comparative performance of the ML methods may again be explained by the PLC curves in the Retailer X data set, which are less similar and thus show a higher degree of variability than those of Dell, and some of which even follow irregular life cycles with two maturity stages.

It is interesting to note at this point that the performance gap between ARIMAX and the DNNs is reduced as both the variability in the historical PLC data increases and the similarity between past products’ life cycles decreases. Since PLC curves of new products are not smooth but rather tend to fluctuate substantially, we smoothed the historical PLC curves before generating forecasts (see Figure 1.4). In the next sub-section, we introduce noise to the (smoothed) input data for the new product to portray the difference in the accuracy of the quantitative forecasting techniques – DNNs with and ARIMAX without clustering – for situations with various degrees of sales fluctuations. We do not test the robustness of the shape-based methods because, apart from Bass-II, adding noise to the sales of the newly launched product does not change the fact that we fit the base curve shapes only to the PLC curves of past products (albeit truncated at the beginning under superscript 2 strategy), which are unaffected by this noise.

1.7 Robustness of the quantitative methods

1.7.1 Impact of adding Gaussian white noise

Up until now we have used smoothed first weeks of sales data as input to the different forecasting methods for the new products. The first few weeks, however, may not be sufficient to generate an optimal fitted curve that would represent the entire new PLC. So unsmoothed or slightly smoothed sales of the new product might be used as input into the forecasting algorithms.

To test how robust the quantitative methods are to noise in the input data for new products, we introduce noise with different standard deviations and evaluate the impact on the forecast accuracy of each method. In particular, Gaussian noise with zero mean ($\mu = 0$) and different levels of standard deviation (σ) is introduced to disrupt the sales during the first few weeks after the new product is launched. Introducing Gaussian white noise (GWN) is a common strategy to test the robustness of different forecasting techniques, and we use it as our basic noise model. It is generally known that DNNs may be prone to over-fitting, so adding noise will provide further validation of this characteristic. We add noise after the smoothing step to be able to measure and control its level, because introducing noise to the unsmoothed sales data can have the opposite effect by decreasing the actual amount of noise in the input.

Let $\tilde{D}^{a,i} = (\tilde{D}_t^{a,i})_{t=1}^{\tau}$ denote the vector of the smoothed sales already encountered for product i . The length of the vector corresponds to the number of weeks we choose to introduce, with each component $\tilde{D}_t^{a,i}$ representing the sales for week t , where $t \in \{1, 2, \dots, \tau\}$ and τ is the number of weeks considered after a new product's launch. Then, the vector of disrupted weekly input sales of the new product i after adding GWN, $\tilde{y}^{(i)}$, is given by:

$$\tilde{y}^{(i)} = \tilde{D}^{a,i} + \epsilon, \quad (1.6)$$

where ϵ is a vector of the same length as $\tilde{D}^{a,i}$ that consists of components that are chosen randomly (i.i.d.) from a zero-mean normal distribution: $\epsilon_t \sim \mathcal{N}(0, \sigma^2) \forall t \in \{1, 2, \dots, \tau\}$.

The results of the robustness testing are reported in Figure 1.12. We chose to test standard deviations $\sigma \in \{0.0001, 0.001, 0.01, 0.1\}$ and considered $\tau = 20$ weeks of product introduction. The results show that all the methods are robust to GWN for σ -values up to 0.001. However, when σ is increased to 0.01, the average MASE reported for ARIMAX is substantially increased. The three DNNs, by contrast, are resistant and therefore robust to this noise level. Although introducing noise with a σ of 0.1 causes the forecast errors for all methods to increase, only ARIMAX's performance deteriorates drastically and its forecast error increases exponentially. The intuition behind this behavior is that ARIMAX detects linear relationships between the sales data, whereas the DNN methods detect non-linear relationships. Because noise standard

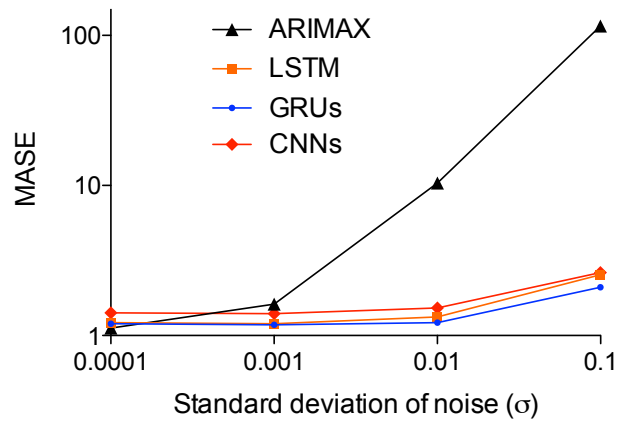


Figure 1.12: Forecast errors (MASE) after adding Gaussian white noise to Dell data (with a log-scale vertical axis)

deviation is inversely correlated to linearity, ARIMAX’s performance decreases substantially.

The robustness results indicate that ARIMAX is considerably more sensitive to high levels of noise in the sales data than the three DNNs under consideration. When compared with the worst-performing DNN – CNNs –, the forecast errors (MASE) achieved by ARIMAX are 15.71%, 578.43%, and 4,291.63% higher for standard deviations 0.001, 0.01, and 0.1, respectively. On the other hand, when compared with either LSTM or GRUs, ARIMAX achieved even higher increases of 36.14%, 715.64%, and 4,923.62%, respectively. Different levels of noise can represent different sales patterns that may occur in real world situations. In fact, we also tested the forecasting methods using unsmoothed sales as input for the new products and obtained results very similar to those reported above for low levels of σ . This similarity suggests that low noise levels correspond to sales fluctuations that can be reasonably expected to be present in “real-world” sales data. Higher values of σ can represent larger unexpected fluctuations such as sharp spikes and/or drops in sales that may occur due to economic booms and busts, recessions, wars, and pandemics.

1.7.2 Effect of incorrect cluster assignment

Although we have used integration and DTW to assist managers in placing the new product in a cluster, the results in Figure 1.6 show that, on average, both cluster assignment methods applied provide only around 60% correct assignments given the input weeks considered in our study. Up until now we have assumed that we know to which cluster each new product belongs to when forecasting. In this section, we assign the new products to (in)correct clusters and report the forecasting errors in terms of MASE. In the tables presented in this section, the rows represent the clusters the products being forecasted belong to, and the columns represent the clusters the new products have been (intentionally incorrectly) assigned to. All results are obtained using 10 weeks of input data, which corresponds to about 60% correct

cluster assignment for both integration and DTW (see Figure 1.6). The forecast errors are calculated for 15% of the data of each cluster and averaged over three repetitions, using the same set of randomly selected products for the different methods in each repetition to ensure comparability of results. MASE results obtained using all products in the Dell data set for forecasting, rather than solely those of the assigned cluster, are reported for comparison in online Appendix A.11.

The results in Table 1.5 show an increase in the average MASE for ARIMAX over all clusters with incorrect cluster assignment when compared to the highest possible accuracy achieved by either forecasting with correct cluster assignment (diagonal entries in Table 1.5) or forecasting without clustering (see ARIMAX column of Table A.5 in Appendix A.11). Note that while the

Table 1.5: Average MASE results for (in)correct cluster assignment using ARIMAX on Dell data

From \ To	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Cluster 1	1.77	1.77	1.77	1.77
Cluster 2	91.56	0.86	0.87	0.86
Cluster 3	23.15	1.86	1.07	1.07
Cluster 4	38.68	1.75	1.05	0.85

results indicate that incorrectly assigning products of cluster 1 (to clusters 2, 3 or 4) results in equally accurate forecasts than correct assignments (to cluster 1), reporting results to more than two decimal places would actually show that using all historical data (i.e. without clustering) produces slightly more accurate forecasts. It can also be seen that incorrectly assigning a product to cluster 1 results in substantial increases in MASE values. This is because the properties of cluster 1 are very different due to the high initial sales volumes of the PLCs and short life cycles, and since ARIMAX captures linear relationships, the errors considerably increase if the products in the cluster are very different from the new products' initial sales. Assuming forecasters do not misplace a product in cluster 1 given the distinct characteristics of the clusters, ARIMAX with incorrect cluster assignment results in an average MASE of 1.42, which is substantially lower than in the three worst case incorrect cluster assignments.

Tables 1.6, 1.7, and 1.8 show the results of (in)correct cluster assignment for LSTM, GRUs, and CNNs, respectively. It can be observed that except for cluster 1, an incorrect cluster assignment

Table 1.6: Average MASE results for (in)correct cluster assignment using LSTM on Dell data

From \ To	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Cluster 1	1.84	1.78	1.50	1.72
Cluster 2	1.17	0.97	1.45	1.16
Cluster 3	1.91	1.72	1.20	1.47
Cluster 4	1.58	1.19	1.39	0.92

virtually always results in higher MASE values compared to a correct assignment. A correct cluster assignment, however, does not necessarily lead to the most accurate forecast. In fact, unlike for products from clusters 2, 3 and 4, forecasts for products from cluster 1 are generally

Table 1.7: Average MASE results for (in)correct cluster assignment using GRUs on Dell data

From \ To	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Cluster 1	1.87	1.77	1.50	1.73
Cluster 2	1.17	0.98	1.45	1.16
Cluster 3	2.02	1.80	1.20	1.48
Cluster 4	1.61	1.20	1.39	0.93

Table 1.8: Average MASE results for (in)correct cluster assignment using CNNs on Dell data

From \ To	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Cluster 1	1.78	1.86	1.46	1.63
Cluster 2	1.23	1.01	1.46	1.28
Cluster 3	2.29	1.81	1.42	1.45
Cluster 4	1.73	1.23	1.42	1.02

more accurate when all PLCs in the data set are used for forecasting (see Table K.13). Somewhat counterintuitively, correctly assigning products from cluster 1 (to cluster 1) generally results in poorer forecasts than incorrectly assigning them to clusters 2-4, which may be explained by the small size and the distinct characteristics of cluster 1. Although the forecast accuracy of DNNs also typically decreases through incorrect assignments, the percentage increases in MASE values are lower than for ARIMAX, with CNNs showing the least percentage increase, which suggests that their performance is more robust to incorrect cluster assignments.

It is interesting to note that the considered DNNs share the same worst case incorrect cluster assignment across clusters. In fact, as can be seen from Tables 1.6-1.8, for all three DNNs the worst incorrect assignment in terms of average MASE is found by assigning products from clusters 1, 2, 3 and 4 to clusters 2, 3, 1 and 1, respectively. Based on the results of this robustness analysis, it can be argued that if forecasters are not sure to which cluster a new product belongs, they may want to consider using all the old PLCs rather than risking an incorrect cluster assignment which can then lead to poor forecasts. Indeed, as shown in this section, forecasting without clustering often results (especially for products from cluster 1) in lower forecast errors than those obtained using incorrect cluster assignments. Forecasters can then apply the ML-based forecasting framework proposed in this paper and follow all the steps, apart from the clustering step, to obtain more accurate sales forecasts of newly launched, short-lived products.

1.8 Conclusions

In this paper, we have developed a framework to forecast sales of new, but not completely new, products with short life cycles after their launches using quantitative methods. To accomplish this, our framework involves a number of important steps based on state-of-the-art techniques. Firstly, we apply smoothing to obtain representative product life cycle (PLC) sales and then use time-series clustering to group similar products. In order to assign the newly launched

product to one of the clusters, we consider two alternative quantitative methods – integration and dynamic time warping (DTW). Subsequently, we perform data augmentation, which is a common approach in machine learning (ML), to generate artificial data points to support the training of the forecasting algorithms. In addition to three deep neural networks (DNNs) – long short-term memory (LSTM), gated recurrent units (GRUs), and convolutional neural networks (CNNs) –, we apply a traditional statistical approach, autoregressive integrated moving average with exogenous variables (ARIMAX), in our framework’s forecasting step. We also apply three forecasting methods that are based on fitting families of curves – Bass diffusion, polynomial and piece-wise linear curves – to historical PLC shapes.

We have illustrated the applicability of our framework using the publicly available Dell data set comprising complete PLC order history for 170 products, and have additionally evaluated the comparative performance of the forecasting methods using a second, much larger data set – from Retailer X – that includes customer orders for 843 complete PLCs. We investigated the accuracy of both integration and DTW in correctly assigning a new product to its cluster, and found that the effectiveness of the latter is somewhat limited. In our empirical analysis, we found that ARIMAX gives the best forecasting results for the three different number of input weeks considered and, for our larger data set, showed that clustering generally results in substantially lower forecast errors. In fact, we have shown that, when compared with the best-performing DNN – either LSTM or GRUs – the forecast errors (MASE) achieved by ARIMAX using Dell data are 24.32%, 23.64%, and 21.20% lower for 10, 15, and 20 input weeks, respectively. When compared with the best-performing PLC shape-based method – either poly2 (Avg) or Bass-II – on the other hand, ARIMAX achieved comparatively lower reductions of 9.68%, 10.64%, and 11.34%, respectively. We obtained consistent results for our larger data set, but found that the performance gap between the DNNs and ARIMAX is reduced and that the former generally outperformed the PLC shape-based methods.

We have also investigated the robustness of the quantitative forecasting methods to noise in the input data and to an incorrect cluster assignment. We added Gaussian white noise with different levels of standard deviation to represent different sales patterns that may occur in real-life situations and found that ARIMAX’s performance deteriorates drastically as the level of noise increases, whereas the three DNN methods’ forecast accuracy remains relatively unaffected. In fact, we have found that, when compared with the worst-performing ML method – CNNs – the forecast errors (MASE) achieved by ARIMAX are more than 15%, 578%, and 4,291% higher for standard deviations of 0.001, 0.01, and 0.1, respectively. This means that the three DNNs are considerably more robust to noise in data sets and, as such, more suitable for forecasting sales of newly launched, short-lived products when there are sufficiently large demand fluctuations such as sudden spikes and drops in sales. We assigned new products to incorrect clusters to analyse the resulting impact on forecasting results and found that errors generally increased considerably. Our results suggest that forecasting may be carried out without the clustering step using all data if it is unclear to which cluster a new product belongs.

The framework proposed in this paper enables practitioners to use quantitative methods to forecast demand in data-scarce situations, and the managerial insights provided support practitioners to decide when to apply ML-based forecasting methods and when to apply traditional methods. In fact, we demonstrated how state-of-the-art techniques that are simple and readily applicable can be combined in a powerful framework in order to enable companies to exploit the untapped potential of quantitative forecasting methods when there is limited availability of historical sales data. For example, integration for cluster assignment – by determining the area under the curve – and data augmentation for the generation of sufficient training data – by performing interpolation – are both simple and straightforward methods that can be readily applied by practitioners.

There are several important directions for future research. For example, our framework could be used in tandem with existing prediction tools that provide pre-launch forecasts to improve overall accuracy by forecast updating. Alternatively, by assuming that it is known to which cluster a new product belongs to before the first demand is realized, our framework could also be used to generate a forecast before a product's launch date. It could also be interesting to evaluate other quantitative methods such as Exponential Smoothing or Multi-Layer Perceptron, to incorporate additional information such as pricing and product reviews, as well as to explore the application of our proposed framework to forecast sales of old or long-lived products.

2 Optimal policies for a price-setting newsvendor with multiple products and a joint capacity constraint

This chapter is based on Elalem YK, Biçer I, Seifert RW. Optimal policies for a price-setting newsvendor with multiple products and a joint capacity constraint. *(Not published yet)*.

2.1 Introduction

Challenges in inventory management continue to persist. Retailers such as Walmart and Target Corp. stored excess inventory in 2022, expecting demand for products such as apparel and furniture to continue increasing post-pandemic, especially during the holiday seasons. The outcome was Walmart discounting goods for months, still ending with around \$1.5 billion worth of inventory, and Target Corp. facing a 90% fall in net earnings, expecting a further \$200 million in expenses related to offloading excess inventory (Nassauer and Terlep, 2022). Inventory glut at the end of the season leads to firms offering high promotional discounts unaccounted for to accommodate for the increase in price-sensitive consumers, leading to a decrease in profits (Ojea, 2022). The problem pertains to many industries. For instance, tech demand surged post-pandemic, causing supply shortages and price increases in semiconductors. With market leaders such as Samsung increasing inventory buildup for its memory business, which comprises about 27% of the semiconductor industry's estimated revenue (Wong, 2023), the firm faces a profit drop of 69% from the previous year as consumer demand unexpectedly drops in 2022. Therefore, coordinating inventory and pricing decisions remains of crucial importance in improving a firm's profit by better matching supply and demand (He and Lu, 2020).

Pricing strategies in firms can range from the use of complex dynamic pricing algorithms that determine optimal prices in a near real-time basis to the use of simple practical methods (Bertini and Koenigsberg, 2021). While companies in different industries, including advertising, insurance, entertainment, and travel, have employed dynamic pricing, the outcome has shown varying degrees of success (Bertini and Koenigsberg, 2021). Other firms rely on more practical pricing methods to minimize complexity for the company and for the consumers. Research

shows that many people delay making decisions or back out of them altogether when decisions become more complex (Dholakia, 2015), with some firms reporting customer dissatisfaction and customer attrition due to implementing dynamic pricing (Scott, 2022). Firms thus have to consider price changes that maximize revenues yet keep customers loyal to the firm (Bertini and Koenigsberg, 2021). It is therefore important for firms to decide on when to use a dynamic pricing policy and when to use a rigid pricing policy instead.

In this work, we study a price-setting newsvendor problem where the decision maker offers multiple products with uncertain demand. The newsvendor jointly determines the price and order quantities for the products which share a capacity constraint on the total order quantity, with the objective of maximizing total profits for a single selling period. We contribute to the literature by addressing the following research questions:

1. What are the optimal prices and order quantities in a price-setting newsvendor problem with multiple products given a shared capacity constraint?
2. How are the optimal results affected by salient modeling parameters such as demand uncertainty, capacity limit, and price elasticity?

For the first research question, we develop analytical expressions that enable us to derive optimal order quantities and prices for the multiple products offered by the retailer under the capacity limit. The analytical expressions are a result of the extension of existing price-setting newsvendor studies (Whitin (1955), Karlin and Carr (1962), Petruzzi and Dada (1999)), by considering multiple products and a capacity constraint. The optimal values are found by iterating over the two sub-problems. For the first sub-problem, given a set of prices, we show that the optimization problem for quantities can be written as a constrained stochastic program (SP). For the second sub-problem, given a set of order quantities, we show that the expected profit equation is concave in prices. We show that the optimal policy is a nested-allocation policy and demonstrate that with only a few iterations between the two sub-problems, the solutions converge.

For the second research question, we analytically demonstrate the impact of changes in demand uncertainty, capacity limit, price elasticity, and cost structures of the products on the optimal prices. We show that the more limited the capacity, the more convergent the pricing policy. We also demonstrate that the optimal price may increase with capacity when the demand uncertainty is high and the price elasticity is low. Furthermore, we show that for a limited capacity, an increase in the order cost of one product causes optimal prices of the other products to decrease. We define a threshold capacity that indicates a change from restrictive capacity to excess capacity, and we demonstrate how this threshold value is affected by changes in order costs, price elasticity, and demand uncertainty. The threshold capacity is an indication of whether the optimal prices and quantities of different products are correlated or not.

Finally, based on our pricing analysis and numerical experiments, we translate the numerical pricing results into managerial insights to be used in practice. We do so by defining a decision topology that indicates to firms the most suitable pricing policy depending on price elasticity and demand uncertainty of the products. Based on the quadrant the products lie in, managers can decide between using a single or differentiated pricing policy, and whether the policy is capacity-dependent or not, therefore setting guidelines that take into account the necessity of price changes based on the product and demand attributes.

2.2 Literature review

Our research naturally connects to the field of revenue management and pricing. One stream within revenue management closely related to our study focuses on jointly determining price and order quantities to maximize profit. Studies in the literature focus on a price-setting newsvendor, with Whitin (1955) being the first to show a fundamental connection between price theory and inventory control. Whitin first determines the optimal order quantity as a function of price and then the corresponding optimal price by establishing a sequential system of expressions. Several extensions of his work have been published. Mills (1959) reformulates the demand in Whitin's study, explicitly specifying mean demand as a function of price. Other extensions later follow using different forms of demand in the formulation and analysis (Karlin and Carr (1962), Younge (1978), Petruzzi and Dada (1999)). Elmaghraby and Keskinocak (2003) provide a review of the literature conducted in this field. Petruzzi and Dada (1999) provide a more integrated framework, investigate different formulations of the problem, and extend the problem to the multi-period setting. Common to these papers is that a single product is considered in the study. More recently, several studies contributed to the field by jointly determining prices and order quantities for multiple products. In particular, Aydin and Porteus (2008) solve for a unique vector of prices and inventory levels for an assortment of substitutable products. Shi et al. (2011), Murray et al. (2012) include capacity constraints to the problem and solve for the inventory and pricing decisions. Our contribution to this literature is the development of analytical expressions that make it feasible to derive the optimal order quantities and prices under a capacity limit on the sum of order quantities. We show that the optimal policy is a nested-allocation policy. We also show that the pricing policy is affected by the capacity limit, and that the more limited the capacity, the more convergent the pricing policy.

Studies in the price-setting newsvendor literature focus on the effects of demand uncertainty on optimal prices. Authors study how different demand models and different levels of demand uncertainty affect the optimal price (Xu et al. (2010), Salinger and Ampudia (2011), and Murray et al. (2012)). In all these studies, firms are expected to use a dynamic pricing policy, updating prices based on changes in exogenous variables to maximize profits. While research shows the importance of dynamic pricing in maximizing profit, firms show that in practice customers expect prices of some products to be more stable. Changing prices too often should be justified against the increase in complexity and costs associated to marketing (Feng, 2010) especially as

it may result in customer dissatisfaction and attrition (Scott, 2022). We contribute to the extant literature by defining different pricing policies based on demand uncertainty, price elasticity, and cost structures for a price-setting newsvendor selling multiple products. Based on the capacity limit, we show that a more convergent pricing policy should be used when capacity is more limited. The cost structures of the products affect the results. An increase in the order cost of one product causes a decrease in the optimal prices of the other products, decreasing the extent of using a convergent pricing policy. We also describe a topology that, based on the demand uncertainty and price elasticity, defines the best choice of pricing policy. For products with identical cost structures and a salvage value of zero, we show that the optimal pricing policy can change from capacity-dependent to rigid, and from a single pricing policy for all products to a differentiated pricing policy based on how high or low the values of demand uncertainty and price elasticities are for the products.

2.3 Model preliminaries

Consider a price-setting newsvendor selling n products in the coming season. The newsvendor has to decide on the optimal price and optimal order quantity for each product prior to the season to maximize expected profits. Let q_i denote the order quantity for product i , and p_i denote its price. The sum of order quantities q_i for $i \in \{1, \dots, n\}$ is constrained by a capacity limit K .

We use $d_i(p_i, \epsilon_i)$ to denote the demand of product i for $i \in \{1, \dots, n\}$. The demand is a function that consists of a deterministic term that depends on price and a random term. The random term in the demand function can be modeled either in an additive or multiplicative fashion. Let ϵ_i denote the randomness in demand for $i \in \{1, \dots, n\}$. The demand in the additive case is defined as $d_i(p_i, \epsilon_i) = y_i(p_i) + \epsilon_i$ and the demand in the multiplicative case is defined as $d_i(p_i, \epsilon_i) = y_i(p_i) \epsilon_i$, where $y_i(p_i)$ is the deterministic term that depends on price and is a decreasing function of price. In this study, we use the multiplicative demand case. While both demand models are commonly used in literature and fit empirical data well, the multiplicative demand model provides a better fit when uncertainty in demand is relatively high (Gallego et al., 2007), making it a better fit for our study where products can have high or low demand uncertainties.

2.3.1 Multi-product model

We develop our model with multiple products based on the methodology given by Petruzzi and Dada (1999), where the authors provide the following analytical derivations for a single-product model:

$$z = F^{-1}\left(\frac{p+g-c}{p+g-s}\right), \quad (2.1)$$

$$p = \frac{bc}{b-1} + \frac{b}{b-1} \left[\frac{(c-s)A(z) + g\Theta(z)}{\mu - \Theta(z)} \right], \quad (2.2)$$

where p is the selling price, z is the ratio of quantity ordered to the deterministic, price-dependent term in the demand function, $y(p)$, c is the per unit order cost, g is the per unit loss-of-goodwill when shortages occur, and s is the per unit salvage value when inventory is left over. The authors denote $A(z) = \int_L^z (z-u) f(u) du$ and $\Theta(z) = \int_z^U (u-z) f(u) du$, where $f(\cdot)$ denotes the probability density function of the random term in the demand function, $F(\cdot)$ denotes its cumulative distribution function, and L and U denote the lower and upper limits, respectively, that ϵ , the random term in demand, is defined on. Expressions (2.1) and (2.2) help in calculating the optimal order quantities and optimal prices.

We extend the single-product model to the multi-product case and develop the analytical expressions for quantities and prices given a shared capacity limit on the sum of order quantities. We define $y_i(p_i) = a_i p_i^{-b_i}$ ($a_i > 0$, $b_i > 1$), as it combines more naturally with the multiplicative demand case (Petruzzi and Dada, 1999). We assume the random term ϵ_i follows a lognormal distribution:

$$\ln(\epsilon_i) \sim \mathcal{N}(-\zeta_i^2/2, \zeta_i), \quad \forall i \in \{1, \dots, n\}, \quad (2.3)$$

where $-\zeta_i^2/2$ is the location parameter and $\zeta_i > 0$ is the scale parameter.

Demand then follows a lognormal distribution:

$$\ln(d_i(p_i, \epsilon_i)) \sim \mathcal{N}(\ln(y_i(p_i)) - \zeta_i^2/2, \zeta_i), \quad \forall i \in \{1, \dots, n\}. \quad (2.4)$$

Let c_i denote the order cost per unit of product i for $i \in \{1, \dots, n\}$. For each unit of unsold inventory, a salvage value of s_i is obtained. For each unit of lost demand, a loss-of-goodwill cost of g_i is encountered. The newsvendor's expected profit is given by:

$$E[\Pi(\mathbf{Q}, \mathbf{p})] = E\left\{ \sum_{i=1}^n [p_i \min(d_i(p_i, \epsilon_i), q_i) + s_i (q_i - d_i(p_i, \epsilon_i))^+ - g_i (d_i(p_i, \epsilon_i) - q_i)^+ - c_i q_i] \right\}, \quad (2.5)$$

where \mathbf{Q} denotes the vector of order quantities and \mathbf{p} denotes the vector of prices for products $i \in \{1, \dots, n\}$, respectively.

The retailer's optimization problem is a maximization of the expected profit given by Equation (2.5) with a constraint on the sum of order quantities. The decision variables are both order quantities and prices of the products. We reformulate the newsvendor's optimization problem

as two separate sub-problems: (i) a constrained stochastic program (SP) to determine order quantities for a given set of prices and (ii) a price expression for a given quantity based on the first-order condition of the expected profit equation with respect to price. Starting with a set of prices, the retailer solves for order quantities using sub-problem (i), calculates the z_i values ($z_i = q_i / y_i(p_i)$) corresponding to those quantities and prices, then uses sub-problem (ii) to calculate prices. Iterating between both sub-problems, the retailer reaches the optimal quantities and prices under the capacity limit imposed when the solutions converge.

Optimal order quantities

To solve for quantities, we formulate the first subproblem as an SP model (Shapiro et al., 2009, Ch. 1):

$$\begin{aligned} \max_Q z = & \sum_{i=1}^n (p_i - c_i) d_i(p_i, \mu_i) \\ & - \left\{ \sum_{i=1}^n (c_i - s_i) \mathcal{W}_i(q_i, d_i(p_i, \epsilon_i)) + (p_i - c_i + g_i) \mathcal{V}_i(q_i, d_i(p_i, \epsilon_i)) \right\} \end{aligned} \quad (2.6)$$

subject to:

$$\sum_{i=1}^n q_i \leq K, \quad (2.7)$$

$$q_i \geq 0, \quad \forall i \in \{1, \dots, n\}. \quad (2.8)$$

where $\mathcal{W}_i = \max\{0, q_i - d_i(p_i, \epsilon_i)\}$ denotes the expected leftovers, $\mathcal{V}_i = \max\{0, d_i(p_i, \epsilon_i) - q_i\}$ denotes the expected shortages, and μ_i denotes the mean of the random term in the demand function of product i , for $i \in \{1, \dots, n\}$. Constraint (2.7) ensures that the sum of all order quantities does not exceed the available capacity limit K . Constraint (2.8) guarantees that order quantities cannot be negative. We follow the same approach as Shapiro et al. (2009, Ch. 1–3) and transform the SP model into a linear program (LP). We analyze the LP model and its dual, and accordingly partition the demand space and determine the shadow prices (Biçer et al., 2021). We provide details on the solution of the mathematical problem (2.6)–(2.8) in Appendix B.1. The optimal policy is a nested allocation policy that depends on the marginal value of ordering an additional unit of each of the products, and is satisfied when all the products have the same marginal value of ordering one additional unit (Biçer et al., 2021).

Theorem 1 *The optimal order quantity under a joint capacity constraint given a fixed set of prices is given by:*

$$q_i = \begin{cases} q_i^* & \text{if } \sum_{i=1}^n q_i^* < K \\ \hat{q}_i & \text{if } \sum_{i=1}^n q_i^* \geq K \end{cases} \quad (2.9)$$

where q_i^* satisfies the following expression:

$$\begin{aligned} g_i(q_i, d_i(p_i, \epsilon_i)) &= -(c_i - s_i) \int_0^{q_i} (q_i - d_i(p_i, \epsilon_i)) f_i(d_i(p_i, \epsilon_i)) \partial(d_i(p_i, \epsilon_i)) \\ &+ (p_i - c_i + g_i) \int_{q_i}^{\infty} (d_i(p_i, \epsilon_i) - q_i) f_i(d_i(p_i, \epsilon_i)) \partial(d_i(p_i, \epsilon_i)) = 0. \end{aligned} \quad (2.10)$$

Proof: Found in Appendix B.1.

Equations (2.9) and (2.10) show that the solution for optimizing order quantities given a fixed set of prices reduces to that of a single-product newsvendor model when the capacity is non-binding. If the capacity is binding, the capacity is allocated to the products by comparing their marginal profits given by Equation (2.10), and by assigning capacity to the products with the highest marginal profit first. This solution is similar to that of Biçer et al. (2021) for the optimal ordering policy at each decision epoch in a product proliferation setting.

We provide a simplified example for demonstration: assume a retailer sells three products, A , B , and C , with marginal profits of 200, 100, and 50, respectively. Capacity allocation follows a nested-allocation policy when capacity is limited. Capacity is first allocated to the product with the highest marginal profit, i.e., product A with marginal profit equal to 200. The marginal profit of product A decreases with the allocated capacity. Once the marginal profit of product A equates the next highest marginal profit, capacity is allocated to both the products with the current highest marginal profit level, i.e., products A and B . This process is performed until all available capacity is depleted.

Optimal prices

To solve for prices, we define the variable $z_i = q_i / y_i(p_i)$ which is the ratio of the order quantity to the deterministic term of the demand function for $i \in \{1, \dots, n\}$ (Ernst (1970), Thowsen (1975), Petruzzi and Dada (1999)). To calculate the z_i values, we use the order quantities obtained from the first sub-problem and the prices used in the first sub-problem. We rewrite the expected profit equation in terms of z_i :

$$E[\Pi(\mathbf{z}, \mathbf{p})] = \sum_{i=1}^n (p_i - c_i) y_i(p_i) \mu_i - \sum_{i=1}^n y_i(p_i) [(c_i - s_i) \Lambda_i(z_i) + (p_i - c_i + g_i) \Theta_i(z_i)], \quad (2.11)$$

where \mathbf{z} is the vector of z_i values. Similar to Petruzzi and Dada (1999), we define $\Lambda_i(z_i) = \int_0^{z_i} (z_i - u_i) f_i(u_i) du_i$ and $\Theta_i(z_i) = \int_{z_i}^{\infty} (u_i - z_i) f_i(u_i) du_i$, where $f_i(\cdot)$ is the probability density function of the distribution of ϵ_i for $i \in \{1, \dots, n\}$. The expected leftovers and expected shortages for each product are then given by $y_i(p_i) \Lambda_i(z_i)$ and $y_i(p_i) \Theta_i(z_i)$, respectively.

After rewriting the expected profit equation in terms of z_i , we notice that $E[\Pi(z_i, p_i)]$ is concave in p_i for a given z_i (proof found in Appendix B.2). We conclude Lemma 1 (Petruzzi and Dada, 1999):

Lemma 1 *For a fixed value of z_i , the optimal price for product i is determined as a function of z_i :*

$$p_i^* = \frac{b_i c_i}{b_i - 1} + \frac{b_i}{b_i - 1} \left[\frac{(c_i - s_i) \Lambda_i(z_i) + g_i \Theta_i(z_i)}{\mu_i - \Theta_i(z_i)} \right]. \quad (2.12)$$

Proof: The proof can be determined directly from the first- and second- order conditions found in Appendix B.2.

Iterating between both sub-problems, order quantities and prices converge to their optimal values under the capacity limit imposed.

The objective function prior to separating the problem is given by Equation (2.5) and is a function of both quantities and prices. It is therefore important to analyze the effect of the capacity constraint on the pricing problem. We reformulate the maximization problem by using Equation (2.11) as the objective function and by imposing a constraint on the order quantities which we now formulate in terms of z_i and $y_i(p_i)$ (problem reformulation found in Appendix B.3). By formulating the Lagrangian of the problem we conclude the following:

Theorem 2 *The optimal product price is dependent on the quantities and prices of the other products offered by the newsvendor when capacity is restrictive. When capacity is in excess, the optimal price per product is independent from the other products.*

Proof: Found in Appendix B.3.

This theorem defines different pricing policies based on the capacity available. The more limited the capacity, the more dependent the product prices on one another, leading to a convergent pricing policy. There exists a cannibalization effect due to the limited capacity. As capacity increases, this dependency decreases. When capacity is no longer a limiting factor, a differentiated pricing policy is adopted. We define the threshold where capacity changes from binding to unbinding as the critical threshold, K^C .

2.3.2 Properties of the optimal pricing solution

With the characterization of the pricing expression and identifying the relation between the optimal prices and capacity limit, we can use the pricing expression to analyze the impact of

changes in the cost structure of the products on the optimal prices and optimal policies. If everything else remains the same, increasing the salvage value of any product lowers its selling price. The following proposition shows how the prices are affected by changes in the salvage value.

Proposition 1 *When increasing the salvage value of the products in the portfolio, the retailer lowers the selling price of the products to maximize profits.*

Proof: The proof can be directly found by referring to Equation 2.12 where, given a fixed z_i value, the price of a product, p_i , is inversely proportional to its salvage value, s_i .

Increasing the salvage value decreases the overage cost. The newsvendor orders higher quantities in response to this change. Therefore, decreasing the price in this case increases demand for the products to account for the additional quantity ordered. Salvage values also have an effect on the pricing policy. When the salvage values of all products increase and $s_i \rightarrow c_i$ for $i \in \{1, \dots, n\}$, switching from a differentiated pricing policy to a convergent pricing policy when capacity is in excess should increase expected profit.

Theorem 3 *As the difference between the salvage values and the ordering costs of the products decreases, the newsvendor should abandon the differentiated pricing policy and use a convergent pricing policy for all products to maximize profits when capacity is in excess.*

Proof: Found in Appendix B.4.

This theorem provides useful insights on the importance of adopting different pricing policies based on the product cost structures and the capacity limit. Theorem 2 states that products are priced independently when capacity is in excess. Theorem 3 states that when the salvage values of the products increase and are almost equivalent to the order costs, the difference between the optimal prices of the products decreases even if the products are priced independently. This means that when capacity is in excess, the lower the salvage values of the products, the more differentiated the pricing policy.

If everything else remains constant, increasing the order cost of the products decreases their newsvendor order quantities, decreasing the critical threshold. In the following proposition, we demonstrate the impact of an increase in the order costs on the prices.

Proposition 2 *An increase in the order cost of a product causes its optimal price to increase.*

Proof: The proof can be determined directly from the positive relationship between the order cost, c_i , and price, p_i , of a product given by Equation (2.12).

Increasing the selling price of the products when their respective order costs increase compensates for the lost profit margins. If the order cost, however, of only one product i increases,

keeping everything else constant, the optimal prices of all the other products ($j \neq i$) in the portfolio decreases when the capacity is restrictive. The price of product i itself increases as shown in Proposition 2.

Theorem 4 *Given a restrictive capacity, as the order cost of one product in a portfolio increases, the optimal prices of all other products in the portfolio should be decreased to maximize expected profit.*

Proof: Found in Appendix B.5.

The theorem builds on the relationship between different product order costs and selling prices when the capacity is limited. When capacity is limited, it is optimal if the retailer uses all the capacity provided. When the price of one of the products increases, its demand decreases. It is therefore beneficial for the retailer to decrease the prices of other products to increase her overall demand. The retailer can then order a total quantity equal to the available capacity limit and still guarantee enough demand for the products ordered.

2.4 Numerical illustration for the multi-product case

Using an example of a retailer offering three products with identical cost structures but different demand distributions, we first show the convergence of the quantity and pricing solutions over several iterations for a given capacity limit. We then perform numerical analysis and show how the optimal order quantities and optimal prices change with different capacity limits. We demonstrate the changes in the optimal results with changes in the cost structures of the products. Finally, we perform a sensitivity analysis on the optimal prices by changing the price elasticity and demand uncertainties.

2.4.1 Convergence of solutions

Suppose a retailer offers three products, $i \in \{1, 2, 3\}$, and she would like to jointly determine the order quantities and prices to maximize total profit for the coming season. The sum of the order quantities of all products is restricted by a capacity limit set by her supplier. We assume that the deterministic term in the demand function that is dependent on price is the same for all products, and is given by $y(p_i) = 50000 p_i^{-1.1}$. The random term in the demand function follows a lognormal distribution with scale parameters 0.01, 0.02, and 0.03 for products 1, 2, and 3, respectively. We assume that the cost of purchasing each product from the supplier is USD 10 per unit and is identical across the three products, i.e., $c_1 = c_2 = c_3 = \$10$. Unsold inventory is discarded, so the salvage value is set to zero for all the products. We also assume that for each unit of unmet demand, the retailer faces a loss-of-goodwill cost of USD 2 per unit which is equivalent across the three products, i.e., $g_1 = g_2 = g_3 = \$2$. Therefore, the three products have identical cost structures but different demand distributions.

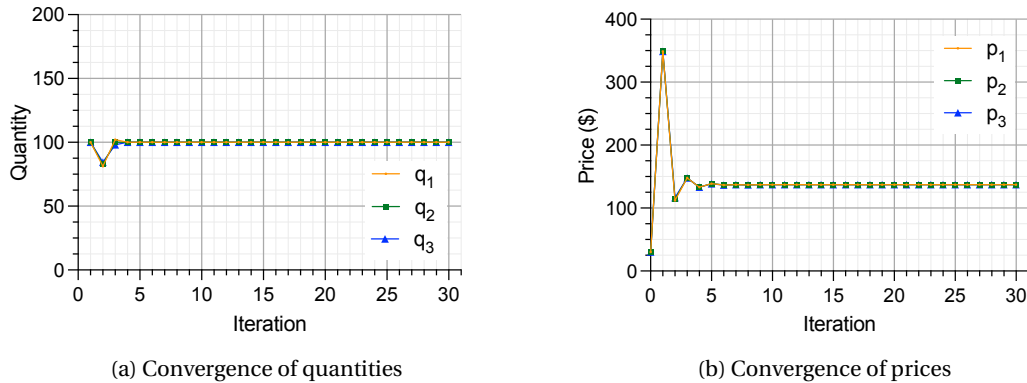


Figure 2.1: Convergence of quantities and prices over 30 iterations of the two sub-problems with a capacity constraint of 300 units.

Starting with an initial price of USD 30 per unit of product, we present the convergence of the order quantity and price solutions in Figure 2.1 when iterating between the two sub-problems, given a capacity limit of 300 units. We present the convergence of the optimal order quantities in Figure 2.1a. The x -axis represents the iteration number, over 30 iterations in total. The y -axis represents the order quantity per product. After four iterations, all quantities converge. We present the convergence of the prices in Figure 2.1b, where the y -axis now represents the price per product. The results again show similar behavior to the convergence of the quantities. Reporting the results to three decimal places, it requires 16 iterations for all prices to converge in our example. The optimal values as a result of solving the two sub-problems iteratively are $q_1 = 100$, $q_2 = 100$, $q_3 = 100$, $p_1 = \$137.065$, $p_2 = \$137.065$, and $p_3 = \$137.065$ for a capacity limit of 300 units, where the prices are reported to three decimal places. The results show that it requires only a few iterations between the two sub-problems for the solutions to converge. The results also show that for a capacity limit of 300 units, optimal quantities and prices are equivalent across the three products. The products are identical in cost structures but differ slightly in their demand distribution.

2.4.2 Identical cost structures

We present the optimal order quantities and prices of the three products for different capacity limits in Figure 2.2. The results for each capacity limit are reported after 30 iterations. The capacity limits are imposed on the sum of the order quantities of all products. We report the results for the following capacity limits: $K = \{300, 500, 700, 900, 1100, 1300\}$. We present the optimal order quantities in Figure 2.2a for each capacity limit. The x -axis represents the capacity limit imposed and the y -axis represents the optimal order quantity for each product. When the capacity is restrictive, as shown in the figure when $K \leq 700$, the optimal order quantities increase with capacity. This is because the retailer orders quantities less than the expected demand when capacity is restrictive, leading to shortages. Once more capacity is available, the retailer orders more quantity to diminish shortages, which in turn increases the

profit. When the capacity is no longer restrictive, the retailer orders the newsvendor quantities that maximize profits based on overage and underage costs. Since this is the optimal quantity, the retailer keeps ordering it regardless of the capacity available when capacity is excessive, which is presented by the plateau in Figure 2.2a. We refer to the point where capacity changes from restrictive to excessive as the critical threshold, K^C , and mark it in red on the x -axes of the figures.

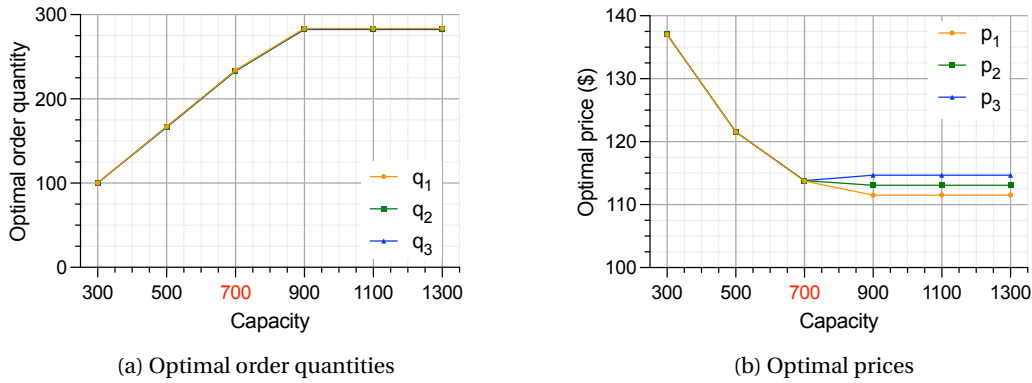


Figure 2.2: Impact of capacity on optimal order quantities and optimal prices.

We present the optimal prices for different capacity limits in Figure 2.2b. The y -axis represents the optimal selling price for each product. As capacity increases, Figure 2.2b shows that the optimal price per product decreases. Then, once the capacity is no longer restrictive, the optimal prices remain constant as they are dependent on the optimal quantities from the first sub-problem, which do not change after the capacity becomes in excess. The results in Figure 2.2b also show that the more limited the capacity, the more convergent the pricing policy. When capacity is restrictive, i.e., $K \leq 700$, the prices of all products offered are equal. When the critical threshold is surpassed and the capacity becomes in excess, a differentiated pricing policy for the products is optimal.

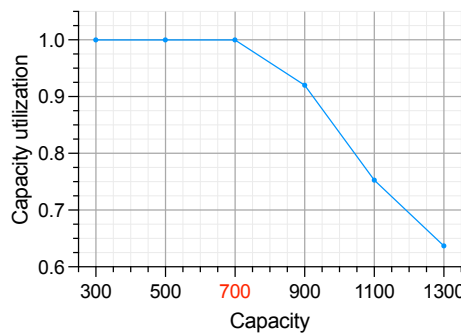


Figure 2.3: Impact of capacity on capacity utilization.

In Figure 2.3, we show the change in capacity utilization as capacity increases. Capacity utilization is the ratio of the sum of order quantities of all the products to the capacity available.

When capacity exceeds 700 units and is no longer restrictive, the retailer has some operational flexibility, and a differentiated pricing policy results in higher profits in this example. Before that point when the retailer has no operational flexibility with respect to capacity, a convergent pricing policy results in the maximum profit.

In Figure 2.4 we present the expected leftovers and expected sales corresponding to each capacity limit. The y-axis in Figure 2.4a shows the expected leftovers per product that the retailer will dismiss at a salvage value of zero in this example. When the capacity is lower than the critical threshold, the retailer utilizes the capacity to its fullest, yet still cannot satisfy demand fully, resulting in shortages, and leading to no leftover inventory. For capacity higher than the critical threshold, the retailer can order the optimal quantities resulting in only a few leftover units to counterbalance the additional demand that can happen due to demand uncertainty.

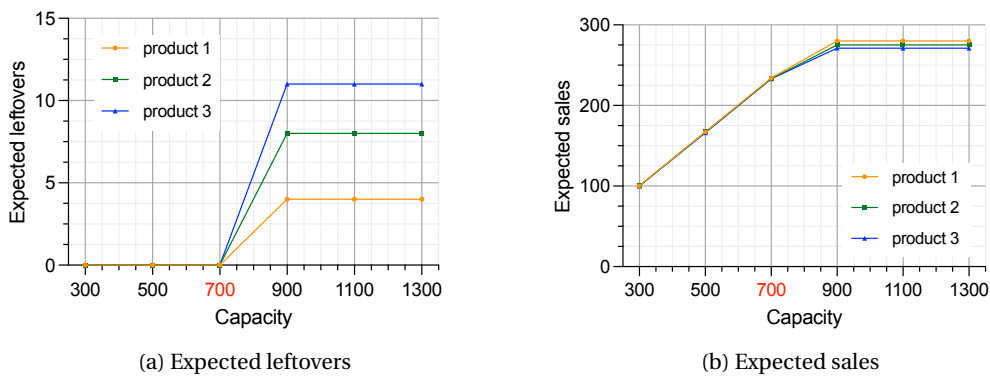


Figure 2.4: Impact of capacity on expected leftovers and expected sales.

In Figure 2.4b the y-axis represents the expected sales corresponding to each capacity limit. Expected sales are equivalent to the optimal order quantities found in Figure 2.2a when capacity is restrictive. Once capacity becomes in excess, the expected sales differ from the optimal quantities. In this example, the expected sales are lower than the optimal quantities, which is reflected by the amount of leftover inventory in Figure 2.4a.

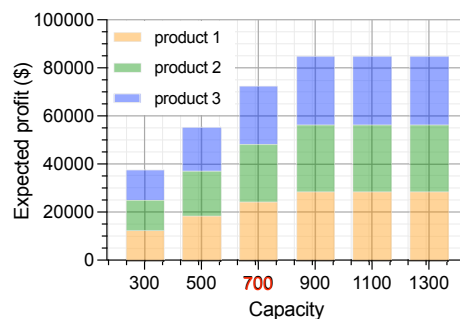


Figure 2.5: Impact of capacity on expected profit.

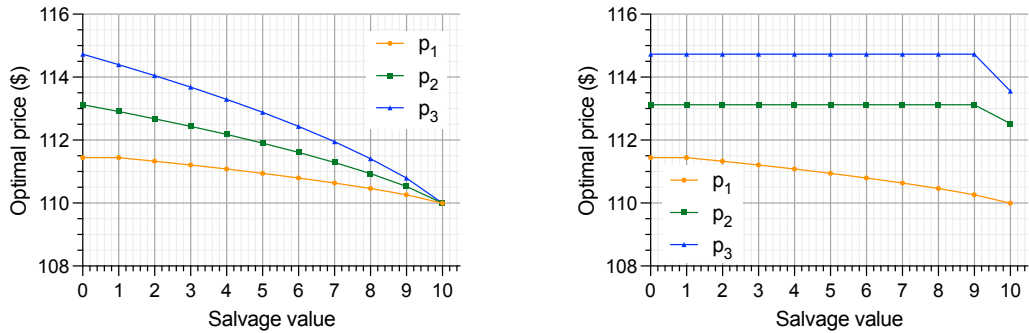
In Figure 2.5, we present the impact of capacity on expected profit. The y -axis represents the sum of expected profit from all products in the portfolio. The results show that when capacity is limited, expected profit increases with capacity. When the capacity exceeds the critical threshold, the retailer orders the newsvendor order quantities and sells them for their respective prices, independent of capacity. Profit no longer increases with capacity when capacity is in excess. Therefore, when capacity is restrictive, as the retailer gains operational flexibility through increased capacity, profit increases. However, when capacity is no longer restrictive, there is no need to invest in higher levels of operational flexibility, especially if additional operational flexibility comes with a cost, as there is no profit improvement beyond the critical threshold. Figure 2.5 also shows the breakdown of the total profit into the profit resulting from each of the three products, where in this example, each product contributes to around 33% of the profit irrespective of the capacity constraint.

2.4.3 Variations in cost structures

We change the cost structure of the products and discuss changes in the optimal pricing results. We first increase the salvage value for all three products incrementally from zero to the order cost of the products, i.e., to 10 USD. In Figure 2.6a, we present the optimal prices as the salvage values for all three products increase simultaneously. The x -axis represents the salvage value for all three products, where $s_1=s_2=s_3$, for a capacity limit of 1100 units, meaning that capacity is in excess. We notice that as the salvage values of all the products increase, the optimal prices decrease. This result follows directly from Proposition 1 that states that when increasing the salvage value of the products in a portfolio, the retailer lowers the selling price of the products to maximize profits. Moreover, we notice from Figure 2.6a that as the salvage value increases and approaches the ordering cost, i.e., as $s_i \rightarrow c_i$, the prices of the products converge to the same price value when there is excess capacity available. The results show that a more convergent pricing policy should be used in the presence of excess inventory as the salvage values increase (Theorem 3).

In Figure 2.6b, we present the optimal prices as the salvage value of product 1 increases incrementally. The salvage values of the two other products are kept constant at a value of zero. The results are again reported for a capacity limit of 1100 units. Up to a salvage value $s_1 = \$9$, the optimal price of product 1 decreases but the optimal prices of the two other products remain unaffected. At a salvage value of $s_1 = c_1 = \$10$, the optimal order quantity of product 1 tends to infinity, as the product can be salvaged at its order cost with no losses in case of leftover inventory. Therefore, when $s_i = c_i$ for any $i \in \{1, \dots, n\}$, infinite capacity is required, and the critical threshold tends to infinity, $K^C \rightarrow \infty$. This effect is reflected in Figure 2.6b, where at $s_1 = c_1 = \$10$ the optimal pricing policy becomes convergent as capacity changes from excessive to restrictive.

In Figure 2.7, we present the optimal prices as the order cost of product 1 changes. The order costs of the other two products are kept constant at a value of 10 USD. The results are reported



(a) Identical change in salvage values of all three products. (b) Change in salvage value of product 1 only.

Figure 2.6: Impact of changes in products' salvage values on optimal prices with a capacity constraint of 1100 units.

for a capacity limit of 300 units. The x -axis represents the order cost of product 1, c_1 . The y -axis represents the optimal prices. The results show that as the order cost of product 1 increases, its selling price increases. This is expected as the increase in selling price compensates for the lost margin due to the increase in the order cost of the product. The optimal prices of the other products decrease when the capacity is limited, although their cost structures do not change. The results follow directly from Theorem 4, showing the negative relationship between optimal selling prices and order costs of different products when the capacity is limited.

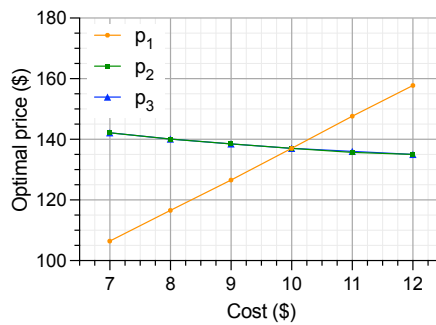


Figure 2.7: Changes in optimal prices with changes in the order cost of product 1 for a capacity constraint of 300 units.

Finally, increasing the order costs of all three products increases the selling price of all products. An increase in the order cost has an effect on the critical threshold, which means that the point at which the differentiated pricing policy should be used changes. In Figure 2.8, we present the changes in the critical threshold with changes in order costs. In Figure 2.8a, we present the change in the critical threshold when the order costs of all three products change by the same amount. The x -axis represents the order cost of the products where $c_1 = c_2 = c_3$. The y -axis shows the critical threshold for each cost value. The results show that the higher

the order costs, the lower the critical threshold. The higher the order costs, the lower the newsvendor order quantities, and therefore less capacity is required. Figure 2.8b presents the results when only the order cost of product 1 changes, keeping the order costs of the other products constant at 10 USD per unit. The x -axis represents the order cost of product 1, c_1 . The results are similar to those of Figure 2.8a, with the critical threshold decreasing with an increase in c_1 . The critical threshold, however, decreases at a lower rate when the order cost of one product increases compared to increasing the order costs of all three products simultaneously.



(a) Identical change in ordering costs of all three products.



(b) Change in ordering cost of product 1 only.

Figure 2.8: Impact of changes in products' ordering costs on the critical capacity K^C

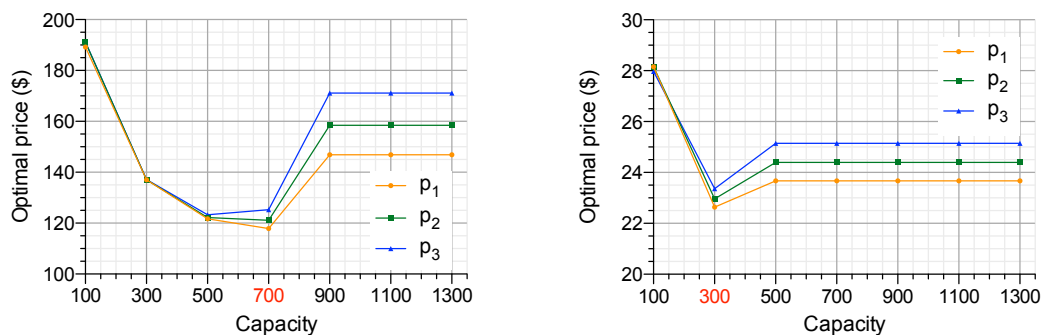
The results show that different cost structures impact the optimal prices and pricing policies. When the products have identical cost structures, the more limited the capacity the more convergent the pricing policy. When capacity is not restrictive, a differentiated pricing policy should be used. Increasing the salvage value of one of the products makes the pricing policy more differentiated when capacity is excessive, while increasing the salvage value of all the products decreases the degree to which a differentiated pricing policy is used. An increase in the order cost of one product causes the optimal price of the product to increase, but decreases the optimal prices of the other products when capacity is restrictive. An increase in the order cost of one or multiple products causes the critical threshold to decrease, which means that the firm should start using a differentiated pricing policy for lower capacity levels.

2.4.4 Sensitivity analysis

We conduct a sensitivity analysis on the optimal prices as a result of changes in consumer price elasticity and in demand uncertainty. Both increases in price elasticity and demand uncertainty are common changes firms face, having severe negative effects on profits if not managed correctly (Bertini and Koenigsberg, 2021). We perform the analysis using different cases of price elasticities and demand uncertainties: (i) low price elasticity and high demand uncertainty, (ii) high price elasticity and high demand uncertainty, (iii) low price elasticity and low demand uncertainty, and (iv) high price elasticity and low demand uncertainty. To test

low price elasticity, we set the b parameter in the demand function to 1.1. To test high price elasticity, we set the b parameter to 1.9. We note that the value of b should always be greater than 1. We show the results for three products with the same cost structure as in Section 2.4.2, i.e., for products having identical cost structures with a salvage value of zero.

In Figure 2.9, we show the results when the demand uncertainty for the products is high, with scale parameters of 0.2, 0.25, and 0.3 for products 1, 2, and 3, respectively. Figure 2.9a shows the optimal prices with low price elasticity and Figure 2.9b shows the results with high price elasticity. The y -axes represent the optimal prices for each capacity limit, where the x -axes represent the capacity limit imposed. Figure 2.9a shows a steep decrease in optimal prices with an increase in capacity from 100 units to 500 units. For example, as the capacity limit increases from $K = 100$ to $K = 300$ units, the optimal price for each product drops by around \$54. Beyond a capacity of 500 units, optimal prices start increasing with capacity. Feng (2010) and Feng et al. (2013) show similar results where the optimal price is increasing with inventory level. This is because the interaction of the price with the demand distribution has a significant impact on the optimal results (Feng et al., 2013). For some parameters, if the price continues to decrease, the newsvendor order quantities increase and the gap between the newsvendor order quantities and capacity limit continues to persist even with high capacity limits. In this case, an increase in price is optimal to maximize profits (please refer to Appendix B.6 for a more detailed explanation). The difference between the optimal prices of the products in Figure 2.9a as capacity becomes less limited is greater and more visible than when the demand uncertainty is low (Figures 2.10a and 2.10b) or the price elasticity is high, as shown in Figure 2.9b. In Figure 2.9b, the results show that the optimal prices are slightly affected with increases in capacity. As the capacity limit increases from $K = 100$ to $K = 300$ units, the optimal price of each product drops by only around \$5.5. The differences in the optimal prices after the critical threshold is surpassed are also minimal. The optimal prices are lower for the higher price elasticity as customers are more sensitive to changes in price.

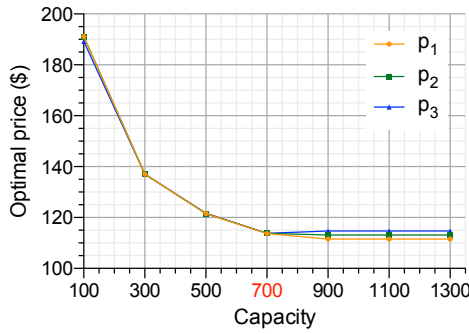


(a) Low price elasticity, $b = 1.1$.

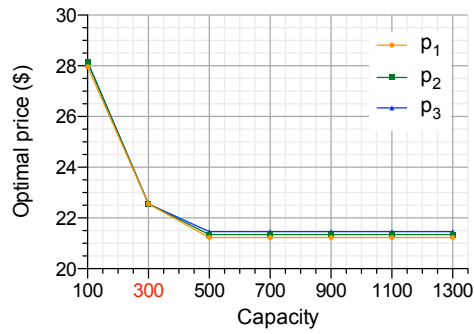
(b) High price elasticity, $b = 1.9$.

Figure 2.9: Sensitivity analysis of price elasticity on optimal prices of products with high demand uncertainty.

In Figure 2.10, we repeat the same analysis as in Figure 2.9 but use low demand uncertainty,



(a) Low price elasticity, $b = 1.1$.



(b) High price elasticity, $b = 1.9$.

Figure 2.10: Sensitivity analysis of price elasticity on optimal prices of products with low demand uncertainty.

with scale parameters of 0.01, 0.02, and 0.03 for products 1, 2, and 3, respectively (same demand uncertainties as in Section 2.4). Similar to the results with high demand uncertainty, low price elasticity results in higher optimal prices when compared to high price elasticity. We present the results for low and high price elasticities in Figures 2.10a and 2.10b, respectively. Figure 2.10a shows that the optimal prices are highly affected with an increase in capacity when the capacity is restrictive. The results also show that although optimal prices are differentiated after the critical threshold, the differences in the optimal prices are minimal. Figure 2.10b shows that for high price elasticities, the difference in the optimal prices of the products does not change considerably with an increase in capacity limit.

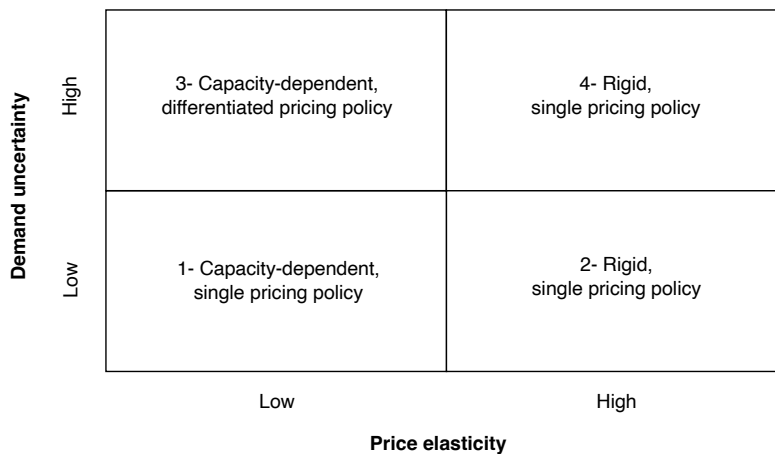


Figure 2.11: Pricing policy topology based on demand uncertainty and price elasticity.

Based on the sensitivity analysis results, we develop a decision typology in Figure 2.11 that shows the pricing policies managers should choose based on the demand uncertainty and

price elasticity of the products sold by the firm. We note that for this topology, products have identical cost structures with a salvage value of zero. We categorize the products along two dimensions: the first dimension is demand uncertainty and the second dimension is price elasticity. While firms can invest in better forecasting and decrease demand uncertainty, products are still classified into having (i) low demand uncertainty and (ii) high demand uncertainty. Price elasticity is classified into (i) low price elasticity and (ii) high price elasticity.

1. *Bottom-left quadrant (Capacity-dependent, single pricing policy)*: These products are essential to consumers and are regarded as staple products. Product categories falling into this quadrant include oil, tobacco, and food products such as rice and flour. Figure 2.10a indicates that the capacity-dependent, single pricing policy is appealing for low price elasticity and low demand uncertainty. If retailers offering a category of these products sell products at very different price levels, most consumers will simply buy the cheapest product when the quality is similar. However, the prices of these products are highly affected by the available supply. Taking the example of rice, different rice options on a supermarket shelf vary slightly in price. With different crises affecting the supply of rice, rice options now cost 15% more than a year ago (Wood, 2022). This makes the single pricing policy dependent on supply (capacity limit) most appealing for these products.
2. *Bottom-right quadrant (Rigid, single pricing policy)*: These product categories can be easily substituted by other products. Although their demand uncertainty is low, an increase in their price can lure customers away. Frozen and junk food are examples of this category. Figure 2.10b indicates that the rigid, single pricing policy is appealing for high price elasticity and low demand uncertainty. For example, if the prices of pizza suddenly increase, customers will simply seek other comfort food within their price range. Therefore, retailers selling products in this category should adopt a fixed pricing policy to maintain their customer base and to make sure all the products they offer sell well.
3. *Upper-left quadrant (Capacity-dependent, differentiated pricing policy)*: These products have highly uncertain seasonal demand. They are essential in some seasons and therefore have low price elasticities. Pharmaceuticals and some chemical products are examples of such product categories. Figure 2.9a indicates that the capacity-dependent, differentiated pricing policy is appealing for low price elasticity and high demand uncertainty. For instance, hand sanitizer prices spiked during the pandemic (Whalen et al., 2020), with some Purell bottles being sold for \$54.99 and higher. Therefore, for such product categories, retailers can use a capacity-dependent, differentiated pricing policy and still ensure that their products will sell.
4. *Upper-right quadrant (Rigid, single pricing policy)*: These products have high demand uncertainty and high price elasticity as they face a lot of competition. Fast fashion products and non-alcoholic beverages are examples of product categories that lie in

this quadrant (Dyer et al. (2014), Karaian and Majerol (2022)). Figure 2.9b indicates that the rigid, single pricing policy is appealing for high price elasticity and high demand uncertainty. If for example the price of a Coke can increases, customers either buy fewer cans or switch to cheaper competitor products (Karaian and Majerol, 2022). Therefore, retailers selling products in this category are ideally suited for using a single pricing strategy for their products and should not reflect capacity constraints on the selling price. It is important to note here that the capacity limits are assumed to have no effect on the order costs. If they do, retailers may want to increase prices to accommodate for lost profit margins.

Based on the demand uncertainty, price elasticity, and cost structures of the products sold, managers can choose the best pricing policy that increases profit yet maintains a loyal customer base for their products.

2.5 Conclusion

In this study, we develop analytical expressions to derive the optimal order quantities and optimal prices given a shared capacity limit for a price-setting newsvendor selling multiple products. We show that the optimal policy for capacity allocation is a nested allocation policy, where products are assigned capacity based on their marginal profits. We also show that when products have identical cost structures with zero salvage value but different demand uncertainties, the pricing policy is more convergent when capacity is more limited. We demonstrate that optimal prices may increase with the capacity when demand uncertainty is high and price elasticity is low. We show that the reason is due to the interaction between the price and demand distribution having a significant impact on the optimal results.

Furthermore, we assess how the optimal prices are affected by varying modeling parameters such as demand uncertainty, capacity limit, price elasticity, and products' cost structures. We show that as the products' salvage values increase and approaches the order cost of the products, the newsvendor should adopt a single pricing policy for the products when capacity is in excess. We also show that an increase in the order cost of one product decreases the optimal prices of the other products offered when capacity is limited. We perform a sensitivity analysis to test the effect of different demand uncertainties and price elasticities on the optimal pricing policy. Utilizing our results, we develop a decision topology to help managers choose the most suitable pricing policy based on the products' price elasticity and demand uncertainty. The analytical expressions derived in this study can be easily used by managers and readily incorporated into a firm's existing operating platform. Based on the cost structures of the products, demand uncertainty, and price elasticity, managers can determine which pricing policy matches their needs best.

There are different directions for future research. One direction is solving the same problem of jointly deciding on order quantities and prices when the retailer does not have access

to complete demand information. Current research on this topic is still limited to a single-product setting (He and Lu, 2020). Another possible research direction involves extending the problem to the multi-period setting and introducing demand-forecast evolution, similar to the work of Biçer and Seifert (2017), to account for changes in the demand distribution.

3 Why do companies need operational flexibility to reduce waste at source?

This chapter is based on Elalem YK, Biçer I, Seifert RW. Why Do Companies Need Operational Flexibility to Reduce Waste at Source? Sustainability. 2022; 14(1):367.

Republished with permission of Elalem, from [Why Do Companies Need Operational Flexibility to Reduce Waste at Source?, Elalem, Biçer, and Seifert]; permission conveyed through Creative Commons Attribution 4.0 International (CC BY 4.0).

3.1 Introduction

Improving sustainability on the production and consumption sides of product life cycles has proven to be critical in reducing the carbon footprint and combating global warming (Martí and Seifert, 2013). For this reason, one of the United Nations Sustainable Development Goals (i.e., Goal #12) explicitly addresses the problems associated with unsustainable production and consumption (<https://www.un.org/sustainabledevelopment/sustainable-development-goals/>). Many manufacturers shift production to low-cost and distant countries to benefit from low production costs, but the long production and shipping lead times between production and the market bases contribute to significant amounts of excess inventory (Biçer et al., 2018) that risk going to waste in retail stores without ever reaching consumers. The cost of excess inventory in the retail industry was estimated to be \$471 billion in 2014 (Gustafson, 2015). In other words, the earth's resources to a value of \$471 billion are wasted in producing goods that are never sold, and hence never used, by any consumer.

Let's consider the apparel industry, which is responsible for 8-10% of global carbon emissions (Niinimäki et al., 2020). The industry is dominated by strong brands that outsource production to contract manufacturers in offshore countries that rely on coal-fueled power plants. These contract manufacturers sometimes even outsource production to yet other countries to further reduce production costs and increase their capacity to fulfill increasing global demand (Donahue, 2018). These offshoring waves have severe effects on the environment. The industry is reported to be responsible for around 35% of oceanic microplastic pollution, 20%

of industrial water pollution, and more than 8% of global carbon emissions (Niinimäki et al., 2020). Despite this environmental destruction, for 30-40% of clothes produced there is no customer demand (Hausman and Thorbeck, 2010), resulting in a loss of profit for the apparel brands. Therefore, 30-40% of the environmental disaster could be eliminated by avoiding holding excess inventory, which is also appealing for retailers because it helps them increase their profits.

Operational flexibility has been proposed by scholars as an effective method for minimizing mismatches between supply and demand under *demand uncertainty* (Biçer et al., 2018). If demand exceeds supply, companies incur the opportunity cost of losing the demand. If demand falls short of supply, companies end up with excess inventory and incur inventory holding costs. In addition to the negative impact on profits, excess inventory has a catastrophic impact on the environment due to the carbon emissions and pollution that arise during the production and logistics operations for goods that are not even demanded by customers. It is therefore important to conduct a comprehensive study of operational and environmental trade-offs arising from the interaction between different supply chain processes such as procurement and inventory management (Martí et al., 2015). In the extant literature, the merits of operational flexibility are quantified from the perspective of its impact on profits (de Treville et al. (2014), Biçer (2015), Bicer and Hagspiel (2016), Biçer and Seifert (2017), Biçer et al. (2018)). However, its benefits for environmental sustainability have not been addressed yet. In this research, we aim to fill this gap in the literature by addressing the following two questions:

- What is the environmental value of operational flexibility measured in the form of waste reduction?
- What types of operational-flexibility strategies are highly effective in increasing profits while improving environmental sustainability?

We consider three different operational-flexibility strategies. The first is lead-time reduction, which can be achieved by localizing production near the market bases. Lead-time reduction allows a buyer to postpone ordering decisions until credible information from the market about the final demand has been collected. Therefore, decision makers base their decisions on accurate demand forecasts and hence are able to reduce supply-demand mismatches (Biçer et al., 2018). Second, we analyze quantity flexibility whereby an offshore supplier offers the buyer flexibility to update the initial order quantity, within some limits, after the buyer has improved its demand forecasts (Bicer and Hagspiel, 2016). Compared with lead-time reduction, quantity flexibility does not require the localization of production near the market bases. Finally, we consider multiple sourcing, in which a buyer employs a domestic supplier and an offshore supplier to exploit the market responsiveness of the domestic supplier and the cost efficiency of the offshore supplier at the same time (Biçer, 2015). Although it has been well established in the extant literature that these three strategies are highly effective in reducing supply-demand mismatches, their impact on reducing waste is not well known.

We assume a profit-oriented buyer who aims to maximize profit and employs operational-flexibility strategies just to reduce mismatch costs. Based on the profit-maximizing decisions of the buyer, we quantify the secondary positive impacts of the operational-flexibility strategies on environmental sustainability.

Following Biçer et al. (2021), we use a multiplicative demand process to model the evolutionary dynamics of demand uncertainty. Then, we quantify the impact of key modeling parameters for each operational-flexibility strategy on the waste ratio, which is measured as the ratio of excess inventory when a certain operational-flexibility strategy is employed to the amount when an offshore supplier is utilized without any operational flexibility. Suppose, for example, the expected excess inventory is 100 units if a buyer sources products from an offshore supplier. Then, the supplier offers the buyer quantity flexibility, helping the buyer reduce the expected excess inventory to 60 units. For the quantity-flexibility strategy employed, the waste ratio obtained is $60/100 = 60\%$. Our results show that the lead-time reduction strategy has the maximum capability to reduce waste in the sourcing process of buyers, followed by the quantity-flexibility and multiple-sourcing strategies, respectively. Therefore, operational-flexibility strategies that rely on the localization of production are key to reducing waste and improving environmental sustainability at source.

We organize the remainder of the paper as follows. In Section 3.2, we position our research by reviewing the extant literature on circular operations management and operational flexibility. We present the model preliminaries in Section 3.3. Then, we analyze each operational strategy and present some numerical examples in Section 3.4, where we also discuss the environmental implications further in Section 3.5. Finally, we provide concluding remarks and envision future research directions in Section 3.6.

3.2 Literature review

Our research is connected to two streams of the operations management literature: (1) circular operations management and (2) operational flexibility. One of the fundamental problems in the circular operations management literature is how to transform the linear “take-make-dispose” operational model to a circular structure, so that products can stay in the market after their lifetime to minimize waste on the consumption side (Abbey and Guide Jr (2017), Atasu et al. (2021)). The phenomenon of circular operations management is also known as closed-loop supply chain management (CLSC). There are three different layers of CLSC, which aim to minimize product waste on the consumption side. We depict these layers in Figure 3.1.

The first layer of the CLSC is *reusing*, which focuses on strategies to extend the consumption length of products (Bras (2009), Agrawal et al. (2018), Atasu et al. (2021)). If a product is damaged or loses its functionality, it must be repaired to increase the length of the consumption period. When a customer loses interest in using a product, it must be sold in the secondary market or shared with other people. Therefore, the ease of repairing, resharing, and selling in secondary markets are key elements of the first layer (Agrawal et al. (2018), Atasu et al. (2021)).

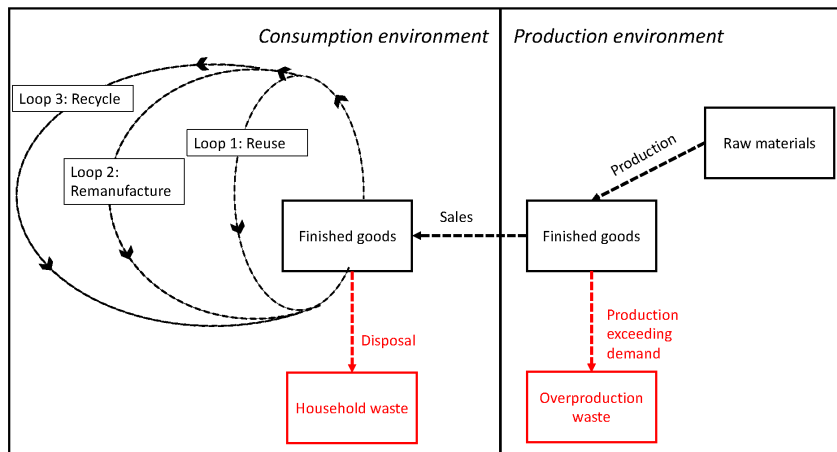


Figure 3.1: Closing the loop in fashion industry

One of the successful applications of reuse is the online marketplace of Patagonia, an outdoor apparel firm, where customers can exchange their clothes when they lose interest in them (Casadesus-Masanell et al., 2010). Another example is the product ownership program of Xerox whereby the printing company retains ownership of the printers and leases them to customers (Atasu et al., 2021). When a customer terminates its contract, Xerox leases the product to a new customer, so the company’s products are shared over their lifetime.

The second layer of the CLSC is *remanufacturing* whereby a set of refurbished and new components are used to manufacture products (Toktay et al. (2000), Debo et al., (2006), Guide Jr and Van Wassenhove (2009)). There are two main challenges regarding the implementation of remanufacturing. The first is uncertainty about the flow of used components, which will later be refurbished for use in the manufacturing process. The second challenge is cannibalization of the original items because introducing the remanufactured products to the market would result in lower sales of the original ones, leading to lower profits. Toktay et al. (2000) address the first challenge by developing a queuing-theory model that dynamically estimates the flow of used products and then applying an aggregate base-stock policy to optimize the inventory policy. To address the second challenge, Debo et al. (2006) develop a diffusion model and categorize products depending on their market diffusion and purchase frequency. The authors outline a decision typology that shows the product categories with the maximum potential for remanufacturing.

The last layer of the CLSC is *recycling* whereby products at the end of their life go through a series of operations to manufacture new items. Well-known examples of recycling are paper and plastic recycling, which are observed in the recycling centers of municipalities of big cities. Yet, the most important challenge of recycling remains the collection of products from households. Atasu et al. (2021) give the example of Norway where the recycling rate for plastic bottles is impressively high—97%. Norwegians achieve this by providing government funds to support retail stores in collecting plastics via reverse vending machines (the same system

can be observed in other western European countries such as the Netherlands). Another approach to increasing the recycling rate is to mandate manufacturers to develop collection and recycling mechanisms for their products, which is popularly known as extended producers responsibility (EPR) (Alev et al., 2019). EPR has been popularized in the electronics industry, with an example being the Minnesota Electronics Recycling Act (Alev et al., 2018). According to this act, the state of Minnesota imposes strict collection and recycling targets on producers as a percentage of their total sales volume (Alev et al., 2018).

Studies in the extant literature successfully address the most important problems related to improving sustainability on the consumer side. Once a product reaches the market, keeping it in the loop of the CLSC has certain environmental benefits. Studies regarding circular operations management are also being tackled for developing and emerging countries with a recent study on recycling and reusing challenges and opportunities for sustainable solid waste management (Bui et al., 2021). However, the extant literature does not quantify the environmental impact of overproduction nor develop remedies for that problem. We contribute to the literature by filling this gap.

Our research is also related to a second stream of literature that prices the value of operational flexibility. Companies establish operational flexibility in different ways, such as lead-time reduction (de Treville et al. (2014), Biçer et al. (2018)), quantity-flexibility contracts (Bicer and Hagspiel, 2016), and multiple sourcing (Cattani et al. (2008), Biçer (2015)). These operational-flexibility strategies make it possible for buyers to determine order quantities after the partial or full resolution of demand uncertainty, helping them to better match supply with uncertain demand. One of the challenges in the extant literature is related to demand modeling because the demand model should involve the time element in order to quantify the benefits of delaying the ordering decision. In practice, manufacturers often employ demand planning teams that collect credible information from customers and update demand forecasts over time. Thus, demand forecasts are improved over time as a result of such efforts. For this reason, the modeling approaches used in the operational-flexibility literature incorporate the evolutionary dynamics of demand forecasts in order to price the value of operational flexibility. de Treville et al. (2014) use a multiplicative demand process to price the value of lead-time reduction. Biçer et al. (2018) later extend the multiplicative demand model by incorporating sudden changes in the demand forecasts and show that the value of lead-time reduction increases with positive jumps in the demand forecasts. Bicer and Hagspiel (2016) use a multiplicative demand model to price the value of quantity flexibility and show that the value is jointly affected by the order-adjustment flexibility and the time when the order adjustments are made. Cattani et al. (2008) develop a tailored capacity model, which is analogous to the multiple-sourcing model that we consider in this research. In their model, a buyer utilizes a speculative capacity under demand uncertainty, but also reserves a reactive capacity that can be utilized once the demand is known. Biçer (2015) extends Cattani et al. by using the extreme-value theory, so the tailored capacity model can be applied to a wider selection of product categories.

Our contribution to the operational-flexibility literature is that we quantify the environmental benefits of operational flexibility that appear in the form of reduced product waste during the sourcing process. The studies in the extant literature are based on a profit-oriented view of the firm such that a reduction in the supply-demand mismatch costs determines the value of operational flexibility. We hopefully expect that companies will be less concerned about increasing their profits in the future, focusing rather on understanding the environmental impact of their operations. Our research aims to fill this gap in the literature by showing how operational flexibility can help minimize product waste during the sourcing process.

3.3 Model preliminaries

To quantify the impact of operational flexibility on excess inventory, we model the evolutionary dynamics of demand forecasts. There are two types of demand models that can be used for such a purpose: (1) the additive demand model and (2) the multiplicative demand model (Oh and Özer (2013), Biçer et al. (2021)). The *difference* between the successive demand forecasts follows a normal distribution in the additive demand model, whereas the *ratio* of the successive demand forecasts follows a normal distribution in the multiplicative demand model. It has been well established in the literature that the additive demand model fits the empirical data well when the forecast horizon is short and the demand uncertainty is low. However, the multiplicative demand model fits the empirical data well when the forecast horizon is long and the demand uncertainty is high (Oh and Özer, 2013). In this paper, we consider the multiplicative demand model because the lead times are expected to be long when companies source from offshore suppliers. Additionally, we focus on products with high demand uncertainty because the magnitude of excess inventory is more pronounced for products with high demand uncertainty than for those with low demand uncertainty.

We use D_i to denote the demand forecast at time t_i such that $t_0 \leq t_i \leq t_n$. The forecast-updating process starts at time t_0 and ends at t_n . We fix t_n to the time when the actual demand is realized, so the final demand is fully known at t_n . Therefore, the length of the forecast horizon is $t_n - t_0$. The demand forecasts are updated at each time epoch t_i for $i \in \{0, 1, \dots, n\}$. According to the multiplicative demand model, the demand forecast D_i is formulated as follows:

$$D_i = D_0 e^{(\nu(t_i - t_0) + \varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_i)}. \quad (3.1)$$

The ν term denotes the drift rate, and ε terms are the forecast adjustments that follow a normal distribution:

$$\varepsilon_i \sim \mathcal{N}(-\zeta^2/2, \zeta), \quad \forall i \in \{1, \dots, n\}, \quad (3.2)$$

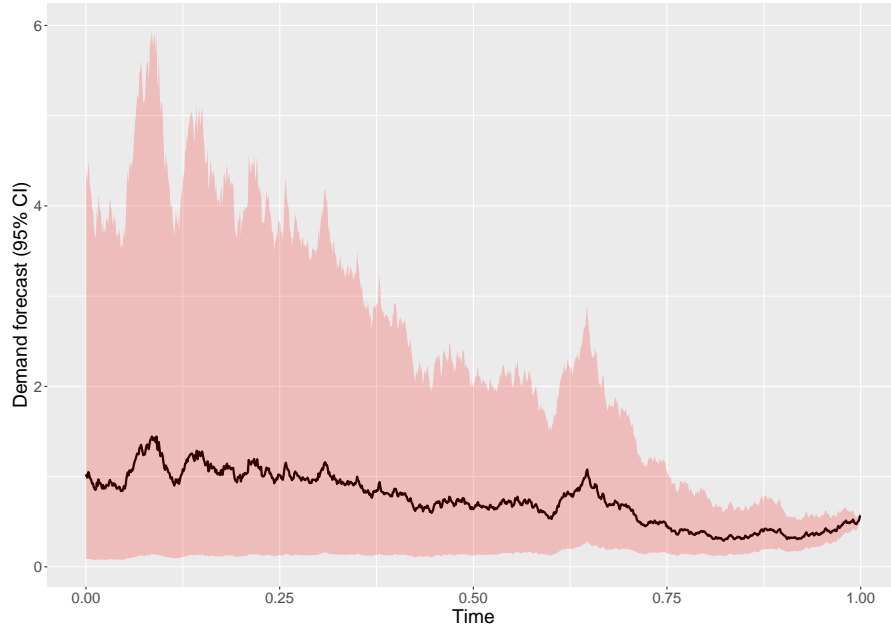


Figure 3.2: Evolution of demand forecasts for the multiplicative process with $D_0 = 1$, $v = 0$, and $\zeta = 1$

where ζ is the volatility parameter.

The drift rate can take non-zero values depending on the forecast-updating process. Biçer et al. (2021) give an example of a forecast-updating process such that demand planners use only the advance demand information to update the demand forecasts, which is modeled by a multiplicative demand model with a positive drift rate. When the forecasts are updated based on an unbiased judgmental demand process, the drift rate should be set equal to zero (Biçer et al., 2021). The multiplicative demand model, given by Equation (3.1), yields a lognormal distribution for the end demand, which is conditional on the demand forecast at t_i :

$$\ln(D_n)|D_i \sim \mathcal{N}(\ln(D_i) + (v - \zeta^2/2)(t_n - t_i), \zeta\sqrt{t_n - t_i}), \quad \forall i \in \{0, \dots, n-1\}. \quad (3.3)$$

The location parameter of the lognormal distribution is $\ln(D_i) + (v - \zeta^2/2)(t_n - t_i)$, and the scale parameter is $\zeta\sqrt{t_n - t_i}$.

In Figure 3.2, we present an example of the multiplicative demand model with a drift rate of zero and a volatility parameter of one. We normalize the initial demand forecast to one and scale the length of the forecast horizon to one. Thus, $D_0 = 1$, $t_0 = 0$, and $t_1 = 1$. We simulate a random path of the evolution of demand forecasts and calculate the 95% confidence interval over the forecast horizon. The black curve represents the demand forecasts, and the pink area shows the 95% confidence interval. For example, the demand forecast at $t = 0$ is equal to one, and the actual demand is expected to be between zero and four at $t = 0$ given by the limits of

the pink area. As shown in the figure, the distance between the limits of the confidence interval decreases over time. This observation indicates that the accuracy of the demand forecasts improves over time as the time for the realization of the final demand approaches, which is consistent with practice. Therefore, the multiplicative model is very effective in capturing the dynamics of demand-updating mechanisms in practice (Oh and Özer (2013), Biçer et al. (2021)).

We now apply the multiplicative demand model given by Equation (3.3) to develop the expected profit, optimal order quantity, and expected excess inventory derivations. We consider the classical newsvendor model such that a buyer sells the products in a market with uncertain demand. We use p to denote the selling price of a product per unit. The buyer incurs a purchasing cost of c per unit. Unsold inventory is salvaged at a salvage value of s per unit. The salvage value can be negative in some industries where companies pay to throw away the excess inventory. In the pharmaceutical industry, for example, unsold drugs must be destroyed after their shelf life because of strict regulations, making the salvage value negative for pharmaceutical companies. The critical-fractile solution was developed to determine the optimal order quantity in the classic paper of Arrow et al. (1951):

$$\beta = \frac{p - c}{p - s}, \quad (3.4)$$

where β is known as the critical fractile or the critical ratio. When the demand follows the lognormal distribution given by Equation (3.3), the optimal order quantity is found by:

$$Q^* = e^{\ln(D_i) + (v - \zeta^2/2)(t_n - t_i) + \Phi^{-1}(\beta)\zeta\sqrt{t_n - t_i}}, \quad (3.5)$$

where $\Phi^{-1}(\cdot)$ is the inverse of the standard normal distribution function $\Phi(\cdot)$.

To find the expected profit, we first need to derive the standardized order quantity. When the buyer orders Q units, the standardized order quantity becomes:

$$z_Q = \frac{\ln(Q/D_i) - (v - \zeta^2/2)(t_n - t_i)}{\zeta\sqrt{t_n - t_i}}. \quad (3.6)$$

Then, the expected profit for an order quantity of Q units is given by Bicer and Hagspiel (2016):

$$E(\Pi(Q)|D_i) = (p - c)Q - (p - s) \left[Q\Phi(z_Q) - D_i e^{v(t_n - t_i)} \Phi(z_Q - \zeta\sqrt{t_n - t_i}) \right]. \quad (3.7)$$

The first term on the right-hand side of Equation (3.7) gives the total profit when all the units ordered are sold in the market at the selling price. However, the demand is uncertain, and it can be less than Q units. The second term on the right-hand side of the expression can be

considered as the cost of an insurance policy that fully hedges the excess inventory risk. The term in brackets is the expected excess inventory:

$$E(\text{Excess Inventory} | Q, D_i) = Q\Phi(z_Q) - D_i e^{v(t_n - t_i)} \Phi(z_Q - \zeta \sqrt{t_n - t_i}). \quad (3.8)$$

The last expression indicates that the excess inventory (hence the waste) can be reduced using two different approaches. First, postponing the ordering decision leads to a reduction in the time window $t_n - t_i$, which in turn helps decrease the expected excess inventory. Second, reducing the order quantity results in a decrease in the expected excess inventory.

These results provide useful insights regarding the use of operational flexibility to improve sustainability by reducing waste in the sourcing process. The lead-time reduction and quantity-flexibility practices make it possible for the buyer to postpone their ordering decision. Therefore, these two operational-flexibility strategies help decrease excess inventory. Utilizing multiple sources (one offshore supplier and one domestic supplier), the buyer can reduce the quantity ordered from an offshore supplier. Thus, multiple sourcing also helps reduce the excess inventory.

3.4 Analysis of the impact of operational flexibility on excess inventory

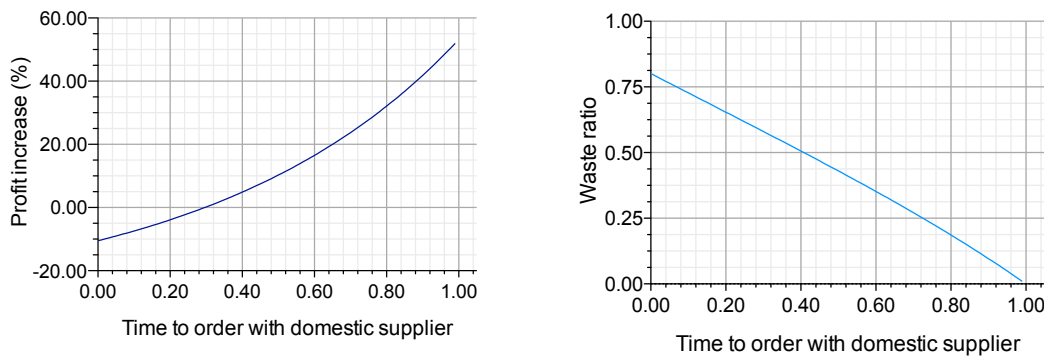
We now look at the impact of the three operational-flexibility strategies (i.e., lead-time reduction, quantity flexibility, and multiple sourcing) on the excess inventory in order to quantify the environmental benefits of operational flexibility. The analytical derivations of the optimal policies for these strategies are given in detail in de Treville et al. (2014) for lead-time reduction, Bicer and Hagspiel (2016) for quantity-flexibility contracts, and Biçer (2015) for multiple sourcing. The following subsections first discuss the analytical derivations of the optimal policies for these strategies and then quantify the impact of operational flexibility on excess inventory.

3.4.1 Lead-time reduction

Suppose that a buyer purchases products from an offshore supplier and sells them in a market with uncertain demand. The buyer places the purchase order at time t_l and the products are delivered at time t_n such that $t_0 \leq t_l \leq t_n$. The buyer sells the products in the market at time t_n . This setting applies to fashion apparel brands that use contract manufacturers to make their clothes and sell them to retail stores at the beginning of each selling season. The length of $t_n - t_l$ is the decision lead time, which is the time elapsed between when the ordering decision is made and when the actual demand is observed. The demand is highly uncertain at time t_l , which in turn exposes the buyer to excess inventory risk.

We assume that the selling price is $\$p$ per unit, and the salvage value is $\$s$ per unit. We use c_l to denote the cost of ordering from the offshore supplier per unit. Then, the optimal order quantity can be found by Equation (3.5). When the buyer places the optimal order quantity, its expected profit and expected excess inventory can be found by Equations (3.7) and (3.8) conditional on the optimal order quantity.

We now consider the case that the buyer aims to reduce the lead time by switching to a local responsive supplier. This makes it possible for the buyer to place the order at time t_s such that $t_s \geq t_l$. However, the buyer incurs a higher purchasing cost when buying products from the local supplier. We use c_s to denote the cost per unit of ordering from the local supplier such that $c_s \geq c_l$. Therefore, the buyer is exposed to a trade-off between postponing the ordering decision and incurring a higher ordering cost. This trade-off has a significant impact on the buyer's profits and the excess inventory.



(a) Impact of lead-time reduction on profit increase. (b) Impact of lead-time reduction on waste ratio.

Figure 3.3: Lead-time reduction

In Figures 3.3a and 3.3b, we present an example of a buyer who would like to decide whether to purchase products from an offshore or a domestic supplier. The selling price of the product is $\$300$ per unit; the cost of purchasing from the offshore supplier is $\$40$ per unit; the cost of purchasing from the domestic supplier is $\$50$ per unit. Unsold inventory is thrown away, so the salvage value is set equal to zero. We normalize the initial demand forecast to one ($D_0 = 1$) and change the demand parameters accordingly. The drift rate of the multiplicative demand model is set equal to zero, and the volatility is equal to one. We also normalize the long lead time (i.e., when an order is placed with the offshore supplier) to one such that $t_n - t_l = 1$ by setting $t_n = 1$ and $t_l = 0$.

We present the percentage change in profit in Figure 3.3a when the buyer switches from the offshore supplier to the domestic supplier. The x-axis represents the time of ordering with the domestic supplier (i.e., t_s), and the y-axis represents the percentage increase in profits.

When the domestic supplier is not responsive enough, the benefits of local sourcing disappear, resulting in a loss of profit. As shown in the figure, the profit increase is negative when $t_s < 0.29$. In this case, it is not advantageous for the buyer to order from the domestic supplier, so a buyer aiming to maximize profit would continue to source from the offshore supplier. If the domestic supplier is responsive enough to let the buyer postpone the ordering decision to later than $t = 0.29$ — i.e., $t_s > 0.29$ — the buyer would increase their profit by switching from the offshore supplier to the domestic supplier. If the lead time is reduced by half (i.e., $t_s = 0.5$), Figure 3.3a shows that ordering from the domestic supplier leads to a profit increase of around 10%. If the lead time is reduced by 90% so that $t_s = 0.9$, the buyer can increase their profit by around 40%.

In addition to these economic benefits, the lead-time reduction helps the buyer reduce waste, thus having a positive environmental impact on the sourcing process. Figure 3.3b shows the waste ratio, which is the ratio of excess inventory when the buyer orders from the domestic supplier to the excess inventory when they order from the offshore supplier. The x-axis represents the t_s value, and the y-axis represents the waste ratio. When $t_s = 0$, the lead time for ordering from the offshore supplier is the same as the lead time for ordering from the domestic supplier. Even if the lead times are the same for both sourcing alternatives, the waste ratio is lower than one for $t_s = 0$, meaning that local sourcing helps reduce waste even in the absence of a lead-time reduction. The waste ratio of 0.8 for $t_s = 0$ is a result of the difference in ordering costs between the domestic and offshore suppliers. The cost of ordering from the domestic supplier is more than ordering from the offshore supplier ($c_s > c_l$). This leads to lower ordering levels when the domestic supplier is used rather than the offshore supplier. Therefore, the 20% reduction in waste for $t_s = 0$ can only be attributed to the lower ordering levels, which is independent of the benefits of a lead-time reduction. However, this improvement is not attainable because Figure 3.3a shows that the buyer prefers the offshore supplier over the domestic one when $t_s = 0$.

When the t_s value increases, Figure 3.3b shows that the waste ratio decreases. Therefore, the buyer can reduce waste during the sourcing process by cutting the lead time with the domestic supplier. When sourcing from the domestic supplier makes it possible to reduce the lead time substantially, the buyer reaches alignment between the economic and environmental incentives of local sourcing. On the one hand, the buyer can increase their profit due to better matching between supply and demand. On the other hand, they can also reduce the waste at source by minimizing the excess inventory. Apart from these direct benefits, promoting local production may also help improve the extent of remanufacturing and recycling because it increases product know-how in local markets.

3.4.2 Quantity-flexibility contracts

When the buyer does not have the possibility to implement local sourcing, a quantity-flexibility contract can be used to increase profits and reduce excess inventory. Under a quantity-

flexibility agreement, the buyer determines the initial order quantity at time t_l and the flexibility percentage. We use Q_l to denote the initial order quantity and following the terminology in Bicer and Hagspiel (2016) we use α to denote the flexibility percentage. Then, the buyer determines the final order quantity Q_f at time $t_s > t_l$ within some limits:

$$(1 - \alpha)Q_l \leq Q_f \leq (1 + \alpha)Q_l. \quad (3.9)$$

If the demand forecasts are updated upward from t_l to t_s , the buyer would increase the order quantity up to $(1 + \alpha)Q_l$ units. Otherwise, the buyer would decrease the order quantity down to $(1 - \alpha)Q_l$ units. Thus, the final order quantity depends on the demand forecast at time t_s , the initial order quantity, and the flexibility percentage. It is given by Bicer and Hagspiel (2016):

$$Q_f = \begin{cases} Q_l(1 - \alpha) & \text{if } D_s < D_{s1}, \\ Q_f^* = D_s e^{(v-\zeta^2/2)(t_n-t_s) + \Phi^{-1}(\beta)\zeta\sqrt{t_n-t_s}} & \text{if } D_{s1} \leq D_s \leq D_{s2}, \\ Q_l(1 + \alpha) & \text{if } D_{s2} < D_s, \end{cases} \quad (3.10)$$

where

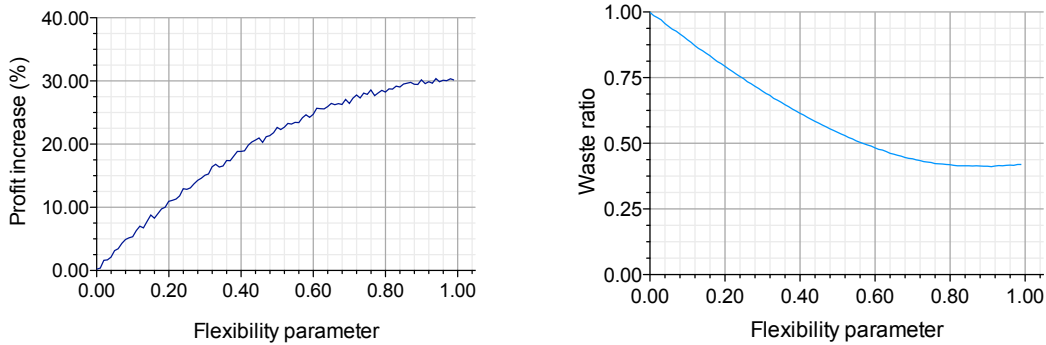
$$D_{s1} = Q_l(1 - \alpha) e^{-(v-\zeta^2/2)(t_n-t_s) - \Phi^{-1}(\beta)\zeta\sqrt{t_n-t_s}}, \quad (3.11)$$

$$D_{s2} = Q_l(1 + \alpha) e^{-(v-\zeta^2/2)(t_n-t_s) - \Phi^{-1}(\beta)\zeta\sqrt{t_n-t_s}}. \quad (3.12)$$

The D_{s1} and D_{s2} terms can be interpreted as the lower and upper critical values for the demand forecast at time t_s . If the demand forecast D_s turns out to be higher than D_{s2} , the buyer should order the maximum allowable quantity based on the quantity-flexibility contract, which is equal to $Q_l(1 + \alpha)$ units. If the demand forecast D_s turns out to be lower than D_{s1} , the buyer should reduce the order quantity to the minimum allowable level, which is equal to $Q_l(1 - \alpha)$ units. If the demand forecast D_s is between these limits, the buyer should set the order quantity to the profit-maximizing level. Based on the final order quantity, the expected profit and the expected excess inventory can be calculated by Equations (3.7) and (3.8), respectively.

In Figures 3.4a and 3.4b, we present an example of a buyer who orders products from an offshore supplier and has the flexibility to update the initial order quantity based on a quantity-flexibility contract. The cost parameters are the same as above: The selling price is \$300 per unit, the cost of purchasing from the offshore supplier is \$40 per unit, and there is no salvage value for unsold inventory. Likewise, the demand parameters are the same as above. The demand forecast at t_0 is normalized to one. The drift rate and the volatility are equal to zero and one, respectively. The initial order quantity is determined at the very beginning such that $t_l = 0$. The final order quantity is determined at t_s within the quantity-flexibility limits.

Figure 3.4a shows the impact of flexibility on the percentage profit increase. The x-axis



(a) Impact of flexibility parameter on profit increase. (b) Impact of flexibility parameter on waste ratio.

Figure 3.4: Quantity-flexibility contracts

represents the flexibility percentage α , and the y-axis represents the percentage increase in profits as a result of the order-adjustment flexibility. To calculate the values of the profit increase, we generate 100,000 random demand paths for each α value. We compare the demand realization at t_s along each sample path with D_{s1} and D_{s2} limits to determine the final order quantity. Then, the expected profit is calculated using Equation (3.7). Figure 3.4a demonstrates that the percentage change in profit increases with a decreasing rate as the flexibility increases. When $\alpha = 0.4$, the buyer can achieve around 20% profit increase compared with the no-flexibility case.

Figure 3.4b depicts the waste ratio as a function of the flexibility parameter. We calculate the waste ratio as the ratio of the expected excess inventory when the buyer has the flexibility to update the initial order quantity to the expected excess inventory when the buyer has no flexibility. Therefore, the waste ratio is close to one when the flexibility percentage is near zero. As the flexibility percentage increases, the waste ratio decreases with a decreasing rate. The curve becomes flatter for high α values such that the waste ratio cannot be reduced below 40%. These results indicate that quantity flexibility has positive economic and environmental impacts on the sourcing process of the buyer. However, its environmental impact is more limited than what can be achieved with lead-time reduction.

3.4.3 Multiple sourcing

In the lead-time reduction and quantity-flexibility practices, the buyer can source products from only one supplier. In the former case, the buyer can source products from either an offshore supplier or a domestic supplier, but not from both at the same time. In the latter case, the buyer can only order from an offshore supplier that provides the flexibility to adjust the initial order quantity in a later time epoch. We now consider an alternative strategy whereby the buyer can source products from two different suppliers at the same time: One is the offshore supplier and the other the domestic.

The multiple sourcing strategy is very effective in mitigating the risk of supply-demand mismatches (Biçer, 2015). By utilizing an offshore supplier, the buyer benefits from the cost advantages of offshore production. If the buyer orders lower quantities from the offshore supplier, the excess inventory risk can also be minimized. If the demand turns out to be unexpectedly high, the buyer then utilizes the domestic supplier to meet the surplus demand. Therefore, a multiple-sourcing strategy allows the buyer to benefit from the cost advantages of the offshore supplier and the responsiveness of the domestic supplier at the same time. However, one of the implementation challenges of this strategy is that the domestic supplier may not always be utilized at a high level. If the demand turns out to be low, the quantity ordered from the domestic supplier would not be high enough to fully utilize its available capacity. Therefore, domestic suppliers are exposed to the risk of capacity underutilization when a multiple-sourcing strategy is employed. To compensate for this risk, domestic suppliers often charge their buyers a capacity reservation fee.

To capture these dynamics, we consider a multiple-sourcing setting with one buyer, one offshore supplier, and one domestic supplier. The buyer determines the order quantity of Q_l units from the offshore supplier and reserves a capacity of K units with the domestic supplier at time t_l . We use c_l and c_k to denote the cost of ordering from the offshore supplier and the capacity reservation cost at the domestic supplier per unit, respectively. At time $t_n > t_l$, the buyer observes the final demand and determines the final order quantity of Q_s units from the domestic supplier such that $Q_s \leq K$. The domestic supplier is additionally paid c_s per each unit ordered. The formulation of Q_s is given by Biçer (2015):

$$Q_s = \max(\min(D_n, Q_l + K), Q_l) - Q_l, \quad (3.13)$$

where D_n is the final demand for the product, which is observed at time t_n . Following up from Biçer (2015), we first define two ratios to derive the optimal values of Q_l and K . They are:

$$\beta_1 = \frac{c_s + c_k - c_l}{c_s - s}, \quad (3.14)$$

$$\beta_2 = \frac{p - c_s - c_k}{p - c_s}. \quad (3.15)$$

Then, the optimal values are given by Biçer (2015):

$$Q_l^* = D_l e^{(v-\zeta^2/2)(t_n-t_l) + \Phi^{-1}(\beta_1)\zeta\sqrt{t_n-t_l}}, \quad (3.16)$$

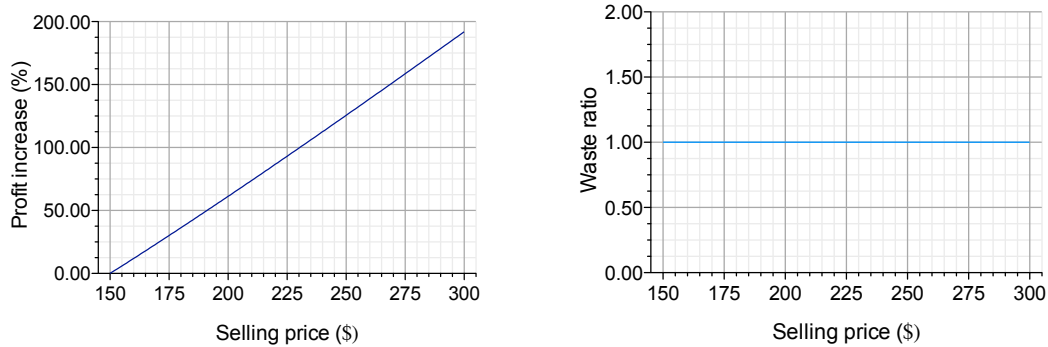
$$K^* = D_l e^{(v-\zeta^2/2)(t_n-t_l)} [e^{\Phi^{-1}(\beta_2)\zeta\sqrt{t_n-t_l}} - e^{\Phi^{-1}(\beta_1)\zeta\sqrt{t_n-t_l}}]. \quad (3.17)$$

Then, the profit in the multiple-sourcing setting is formulated as follows:

$$\Pi(Q_l, K, Q_s) = p \min(D_n, Q_l + K) + s \max(Q_l - D_n, 0) - c_l Q_l - c_k K - c_s Q_s. \quad (3.18)$$

Using this formula and simulating the demand paths, the expected inventory for the optimal Q_l and K levels can be found. The expected excess inventory can also be calculated by plugging Q_l into Equation (3.8).

We now present an example to demonstrate the impact of multiple sourcing on the buyer's profits and excess inventory. We assume the same demand parameters as the examples given above. The cost of purchasing from the offshore supplier is \$40 per unit, and the cost of purchasing from the domestic supplier is \$50 per unit. The domestic supplier also charges \$15 for each unit of capacity reserved. Unlike the lead-time reduction and quantity-flexibility examples, we cannot vary the key decision parameter that determines the magnitude of operational flexibility in the multiple-sourcing setting. In the multiple-sourcing setting, operational flexibility is directly influenced by the reactive capacity K . However, the reactive capacity is not a control variable. It is a decision variable that has to be optimized depending on the cost and demand parameters. For that reason, we vary the selling price between \$150 and \$300 in our analysis because Equation (3.17) indicates that the optimal capacity level increases with the selling price. In other words, we can observe the impact of responsiveness on profits and the waste ratio by varying the selling price because the selling price is directly correlated with the capacity level.



(a) Impact of selling price on profit increase with multiple sourcing.

(b) Impact of selling price on waste ratio with multiple sourcing.

Figure 3.5: Multiple Sourcing

Figure 3.5a shows the impact of the selling price on the percentage profit increase such that the profit increases with the selling price. Doubling the selling price from \$150 to \$300 increases the profit by almost 190%. Figure 3.5b demonstrates that the waste ratio does not change depending on the selling price. This is because the optimal order quantity from the offshore supplier does not depend on the selling price but only on the cost parameters. For this reason, there is no environmental benefit to improving operational flexibility with the

multiple-sourcing strategy. Both the profit increase and waste ratios in Figures 3.5a and 3.5b are calculated with respect to the resulting profit and waste when the selling price is set at \$150.

We summarize the results of the three operational-flexibility strategies in Table 3.1 below. For each strategy, the profit increase and waste ratio reported are calculated by setting the parameter of the strategy in the table to the maximum value in comparison to setting the parameter to the minimum value within the given range. Results show that lead-time reduction has the highest potential in reducing waste while improving the profits of companies. In Section 3.5, we further elaborate on and discuss the results of the three strategies.

Table 3.1: Summary of results of the operational-flexibility strategies

Strategy	Parameter	Range of parameter values	Profit increase	Waste ratio
Lead-time reduction	lead time with domestic supplier (t_s)	$t_s \in [0, 0.99]$	51.94%	0.01
Quantity-flexibility contracts	flexibility parameter (α)	$\alpha \in [0, 0.99]$	29.98%	0.42
Multiple sourcing	selling price (p)	$p \in [150, 300]$	191.98%	1.00

3.5 Discussion

So far, we have shown the effect on the change in profit and the waste ratio of three supply chain flexibility strategies: (1) lead-time reduction, (2) quantity-flexibility contracts, and (3) multiple sourcing. The numerical analysis of lead-time reduction shows that when a firm reduces its lead time by utilizing an onshore supplier instead of an offshore supplier, the firm can achieve both profit increase and waste reduction due to the decrease in demand uncertainty at the order time. When t_s is greater than 0.29, the firm starts experiencing a positive change in profit, which exceeds 50% as t_s increases to 1. As for the waste, even if $t_s = t_l = 0$, the firm still experiences waste reduction of 20% due to the increased cost, which lowers the newsvendor order quantity. The waste ratio decreases to 0 as t_s increases to 1, since actual demand is known at this point and the firm does not operate under demand uncertainty anymore.

If firms cannot order from an onshore supplier, they can still benefit from the flexibility of an offshore supplier with a quantity flexibility contract where, up to time t_s , the firm can still adjust its order quantities by a factor of α . For a fixed t_s , our analysis shows that firms

can achieve a 30% increase in profit, and waste reduction can reach more than 50% as the flexibility parameter increases to 1.

For multiple sourcing, the numerical analysis is performed based on the selling price of the product. It shows that doubling the selling price from \$150 to \$300 results in a profit increase of almost 190%. The waste, however, does not change with different pricing as it is calculated based on the offshore order quantity, which is independent of price.

The selection of a given flexibility strategy is also dependent on the risk associated with it. On the one hand, although single sourcing with lead-time reduction has the highest effect in reducing waste while still increasing a firm's profits, it can amplify the exposure to risk in the presence of uncertainty. On the other hand, multiple sourcing can present higher costs, and this is not accounted for as a result of having to manage several suppliers (Costantino and Pellegrino, 2010). It is therefore critical for firms to assess all the benefits and risks arising from choosing a specific strategy in pursuit of improved sustainability and profits.

3.6 Conclusion

Improving sustainability in every aspect of our lives is vital to safeguard a durable future for the next generation. Manufacturers ought to pay special attention to sustainability because what they produce, how they produce, and where they produce has a substantial impact on the carbon footprint. The extant literature on sustainable operations management mainly focuses on the consumption side of product life cycles, with an intention to extend the product lifetime and reduce household waste. However, significant inefficiencies exist on the production side of product life cycles whereby manufacturers overutilize the earth's limited resources and generate carbon emissions when producing products in excessive amounts. A significant number of these products may never reach consumers. Prominent examples reported in the media include Amazon destroying thousands of unsold TVs and laptops in one of its warehouses (Wood, 2021).

In this research, we aim to fill the gap in the literature by demonstrating how operational flexibility can help organizations achieve sustainability at source. We focus on three different operational-flexibility strategies: (1) lead-time reduction, (2) quantity-flexibility contracts, and (3) multiple sourcing. Our results indicate that lead-time reduction has the highest potential to reduce waste while improving the profits of companies. Therefore, operational-flexibility strategies that promote local production are key to reducing waste and improving sustainability.

In particular, in our numerical analysis, where ordering time varies between 0 (start of planning horizon) to 1 (beginning of sales season), we show that lead-time reduction can result in a profit increase when orders are placed with onshore suppliers at a time t_s greater than 0.29, compared with being placed with an offshore supplier at time $t_l = 0$. The profit increase can go up to around 40% when orders are placed at $t_s = 0.9$, even if ordering costs are higher. Waste

in terms of excess unsold inventory also decreases, even when $t_s = t_l = 0$, when the onshore supplier is used and results in a 20% waste reduction compared with ordering from an offshore supplier as a result of the higher ordering costs, which result in lower order quantities. The waste decreases as t_s increases and finally approaches zero when orders can be placed during the selling season, when $t_s = 1$.

The results of our research offer some useful insights regarding the development of effective environmental policies. Because lead-time reduction is the most effective strategy, environmental policies should target cutting lead times, not only for inbound but also for outbound logistics. Increasing the import and export tariffs and imposing trade barriers would force countries to promote local production, which in turn leads to shorter lead times and lower waste. Although such policies conflict with the free trade and economic development ideas, we envision that environmental concerns would be highly dominant in the near future and governments would incrementally pass some regulations to reduce the volume of imports and exports. One of the side benefits of local production would be to establish a close connection between local manufacturers and local authorities such that recycling and remanufacturing can be easily implemented near the market bases. Therefore, local production may also help increase the product life cycle along the stages of the closed loop supply chains (Figure 3.1).

One of the limitations of our research is that we mainly focus on the dynamics on the production side of the product life cycle without connecting it with the consumption side. We mainly consider a newsvendor setting such that the buyer sells the products in the market at a certain price. The classical utility theory (Lancaster, 1966) suggests that the life cycle of a product is positively associated with the price paid for it. We believe that there is a need for empirical research that investigates the relationship between price and the length of a product's lifetime in the retail industry for different categories. We envision that this would be an interesting avenue for further research.

Another direction for future research is how digital transformation could contribute to lead-time reduction. Recent research focuses on lean manufacturing in Industry 4.0 (Tseng et al., 2021). Studies on lean manufacturing show that lead-time reduction is an important factor that enhances reliability and flexibility while decreasing inventory carrying costs, and that integrating lean management and collaboration in the supply chain has important social, environmental, and economic benefits (Tseng et al., 2020). Therefore, we foresee that studies of the effect of Industry 4.0 on decreasing waste and increasing profit through lead-time reduction are a potential field for future work.

Conclusion

In this thesis, we address different factors that mitigate supply and demand mismatches. Increasing forecasting accuracy for newly launched products constitutes the first chapter of the thesis. By increasing forecasting accuracy, managers can coordinate inventory levels with more accurate demand information, and therefore reduce supply and demand mismatches. We develop a framework to forecast sales of newly launched products with limited data using statistical and machine learning (ML) methods. As the base statistical method, we use autoregressive integrated moving average with exogenous variables (ARIMAX). For the ML methods, we use three deep neural networks (DNNs): long short-term memory (LSTM), gated recurrent units (GRUs), and convolutional neural networks (CNNs). The framework groups older, similar products together using time-series clustering, and then relates the newly launched product to one of the clusters using integration or dynamic time warping (DTW). We then perform data augmentation on all time-series to generate additional data points. Finally, we compare the performance of ARIMAX and the ML methods using two data sets. The results show that ARIMAX gives better forecasting results compared to the ML methods when there is no noise in the data. Once we perform a robustness analysis and add Gaussian white noise to the input data, the forecasting accuracy of ARIMAX deteriorates, and the ML methods give more accurate results when compared to ARIMAX. The framework we propose in this chapter allows forecasters to use quantitative forecasting methods in data-scarce situations. Furthermore, our study and results provide support to practitioners to decide on when to apply ML-based forecasting methods and when to apply traditional methods.

The second chapter of the thesis addresses a newsvendor jointly deciding on order quantities and pricing decisions given a capacity limit on the sum of order quantities. Price affects demand, and demand affects order quantities. Therefore, optimizing both order quantities and prices simultaneously improves a firm's performance by reducing inventory glut or shortages at the end of the season. With factors such as pandemics and geopolitical conflicts, supply shortages increase and firms should adjust their decision making according to these constraints. Therefore, in this chapter we develop analytical expressions to derive the optimal order quantities and optimal prices given a shared capacity limit for a price-setting newsvendor selling multiple products. We show that the optimal policy is a nested allocation policy, where products are assigned capacity based on their marginal profits. We show that the more limited the capacity, the more convergent the pricing policy. We also develop a decision topology for

managers to choose the most suited pricing policy depending on price elasticity and demand uncertainty. While dynamic pricing is very important, constant price changes may cause customer dissatisfaction and attrition. Therefore, based on the price elasticity and demand uncertainty, we show using our numerical results that the pricing policy can change from rigid to capacity-dependent, and from a single pricing policy to a differentiated pricing policy for the different products. In addition, we demonstrate the effect of changes in the cost structures on the optimal prices and pricing policies.

The third and final chapter of the thesis analyzes the environmental benefits of three different operational flexibility strategies: lead-time reduction, quantity-flexibility contracts, and multiple sourcing. By ordering from a domestic supplier, offshore supplier, or both, we analyze the changes in profits and waste in terms of leftover inventory. Our results show that operational-flexibility strategies that rely on the localization of production are key to decreasing leftover inventory. This is explained by the reduction in lead time that leads to a better match between supply and demand. Managers can therefore utilize operational-flexibility strategies that can increase profits and decrease waste production at the same time.

Limitations of our work include the following. In the first chapter, the framework developed requires the new product to have at least one data point of historical sales. This limits the applicability of our work to new products that have not been launched yet. In the second chapter, we assume that the demand distribution for a given price is known, which may not be the case in practice. Assuming a wrong demand distribution can have negative impacts on profits. Finally, in the third chapter, we assume the lead times are constant. Lead times come with high uncertainties especially in times such as pandemics and geopolitical conflicts. Although uncertainty will probably still favor localization of production, it is important to consider it in future studies.

Various future research directions can result from our work. For the first chapter, our framework can be extended for use as a prediction tool in the pre-launch phase. Once the new product starts to sell in the market, it can be used for improving the forecasting accuracy by forecast updating. In addition, different quantitative methods such as exponential smoothing or multi-layer perceptron could be tested using the same framework. As pricing affects demand, including planned prices and promotions into the algorithms in future studies could also help in improving forecast accuracy. For the second chapter, the joint decision of order quantities and prices can be solved in the presence of incomplete demand information. Using Fourier transform as an example, future studies can use the characteristic function for transactional data instead of assuming a known demand distribution. Another possibility for future research would be extending the problem discussed in the second chapter to the multi-period setting and introducing demand-forecast evolution. As the third chapter focuses on the dynamics of the production side, one research direction could be incorporating the effects of changes in the consumption side on profitability and environmental sustainability. Finally, analyzing the effects of how digital transformation could enhance reliability and flexibility could be studied in the future, as it is expected to positively impact inventory decisions.

A Supplementary material for Chapter 1

Republished with permission of Elsevier Science & Technology Journals, from [A machine learning-based framework for forecasting sales of new products with short life cycles using deep neural networks, Elalem, Maier, and Seifert]; permission conveyed through Copyright Clearance Center, Inc. (“CCC”). (license ID: 1313291-1).

A.1 Clustering results

The graphical result of the clustering step is a dendrogram. Figure A.1 shows the resulting dendrogram from time-series clustering of the Dell data. It can be visually identified (see

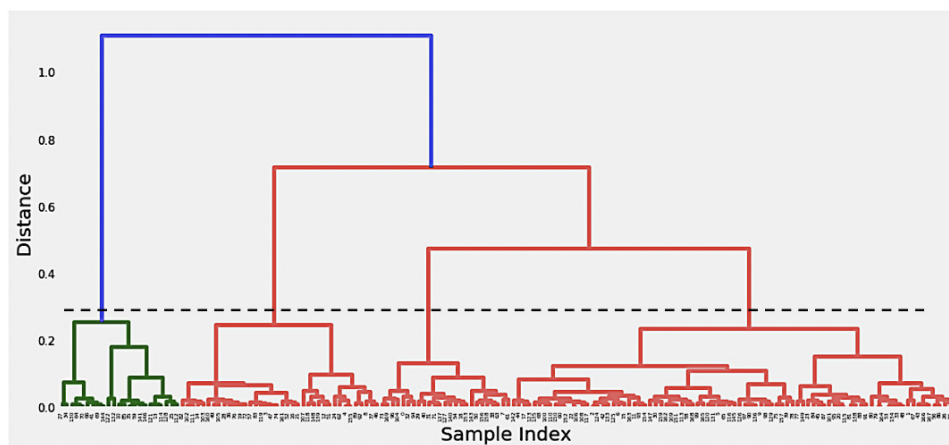


Figure A.1: Result of dendrogram analysis using the Dell data

dashed horizontal line) that four clusters separate the data in a meaningful way. In general, the lower the horizontal line, the more clusters there are, as the threshold distance between the clusters decreases. The resulting scree plot is displayed in Figure A.2 and shows a relatively low variance with four clusters, which validates our choice.

We now show the clustering results for the Retailer X data set, which is larger than the Dell data

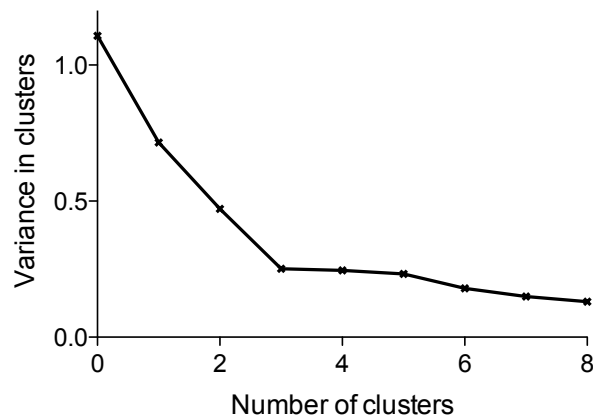


Figure A.2: Scree plot of the Dell data

set. Although we tried different types of clustering for both data sets before reaching the final results, the analysis of the effect of different linkage methods on the clustering is presented here for only Retailer X. Applying the Euclidean distance, Figures A.3, A.4, A.5, and A.6 show the results of the hierarchical agglomerative clustering using the single, complete, average, and ward method, respectively.

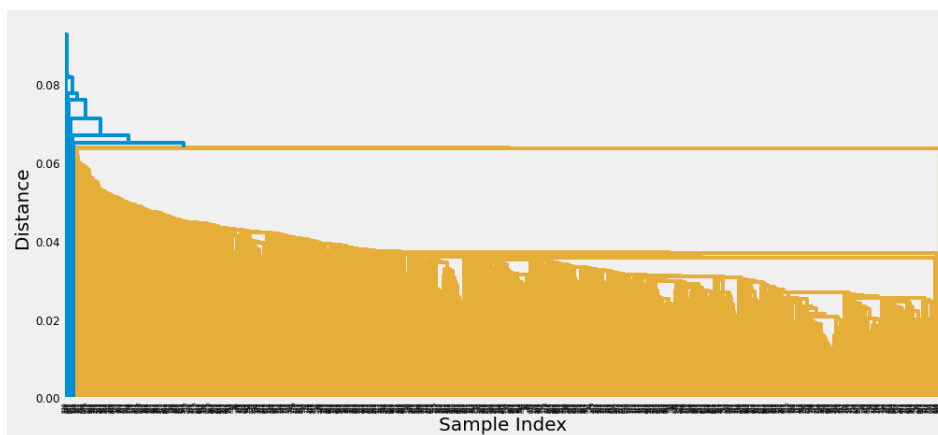


Figure A.3: Result of dendrogram analysis using the single method on Retailer X data

By comparing the effects of the different linkage methods, we find that the ward method provides the best separation in terms of both grouping similar products together and providing well separated clusters. As a result, we chose the ward method in the clustering step. Furthermore, based on the results of the dendrogram analysis by Figure A.6, we also chose four clusters for the Retailer X data since it separates the data in a meaningful way.

The resulting scree plot from using the ward method for the Retailer X data is displayed in Figure A.7. As can be seen, choosing four clusters leads to a relatively low variance, which

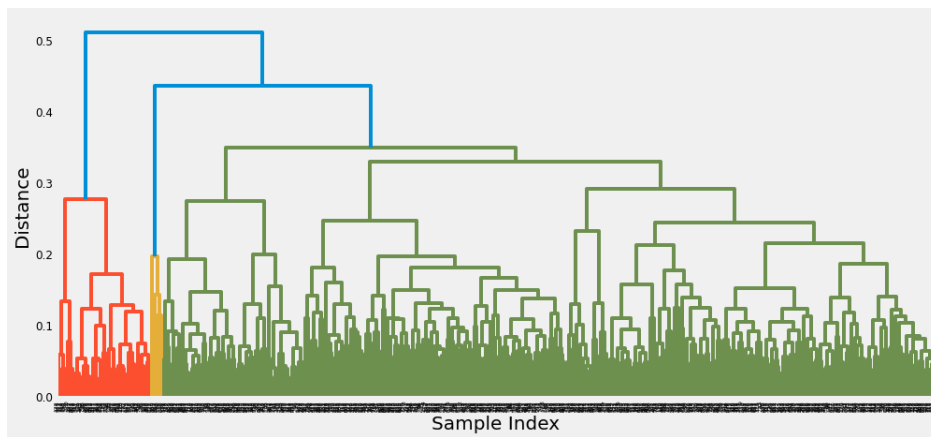


Figure A.4: Result of dendrogram analysis using the complete method on Retailer X data

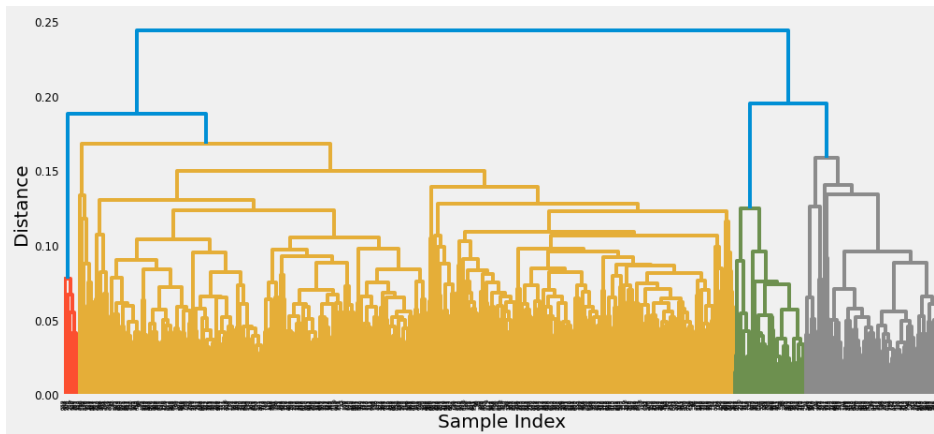


Figure A.5: Result of dendrogram analysis using the average method on Retailer X data

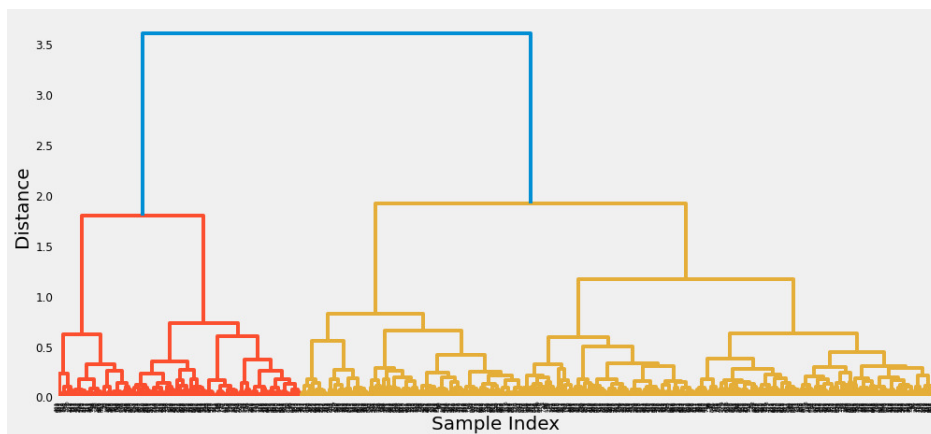


Figure A.6: Result of dendrogram analysis using the ward method on Retailer X data

validates our choice. The scree plot also shows that increasing the number of clusters beyond four results in low decreases in the cluster variance.

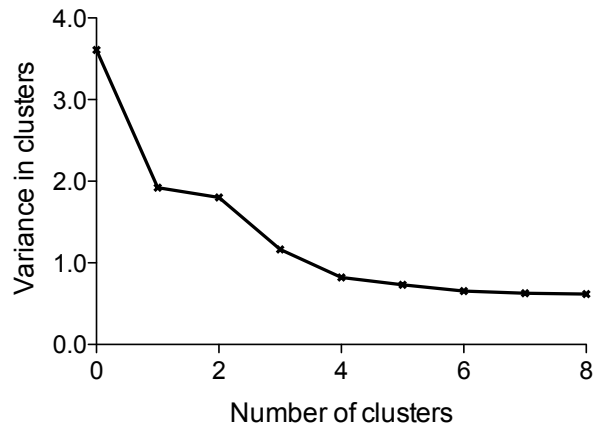


Figure A.7: Scree plot of the Retailer X data

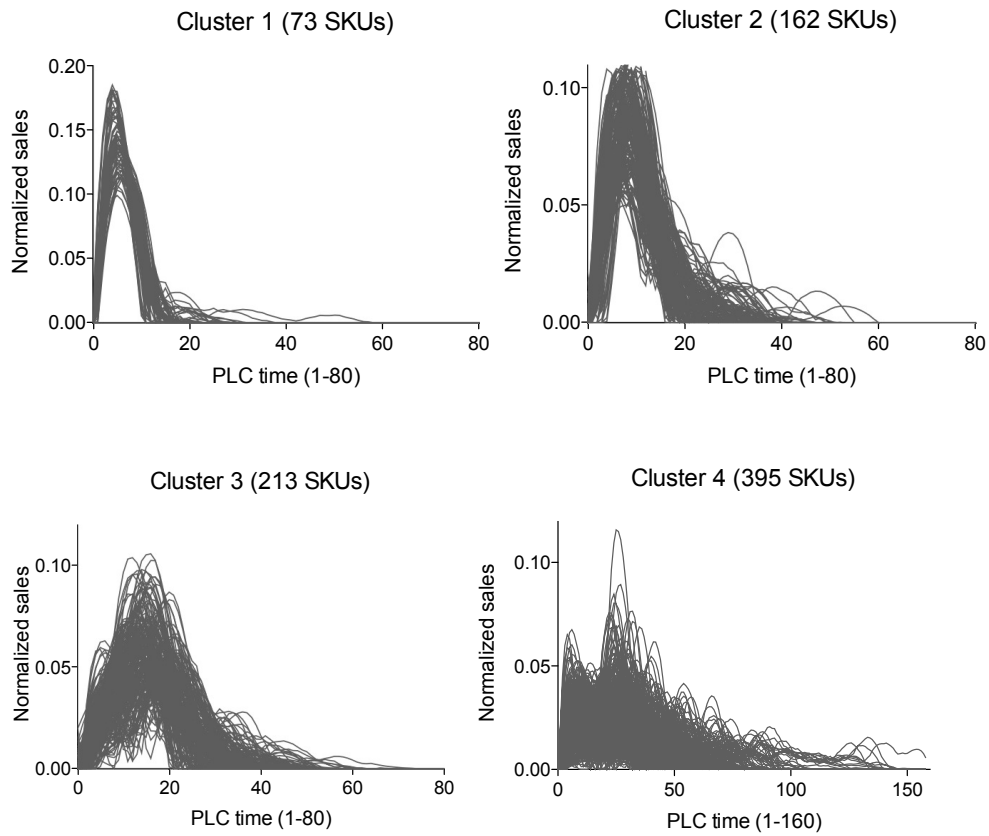


Figure A.8: Clustering results with individual curves both normalized and smoothed (Retailer X data)

Appendix

Figure A.8 shows the result of time-series clustering for the Retailer X data set. It can be seen that the four clusters are distinguished by different product life cycle (PLC) behavior, peak sales amounts (note the difference in the x - and y -axes scales), and life cycle lengths. Moreover, the clusters have more variance in their time-series data and more SKUs compared to the Dell data set.

A.2 Exploratory data analysis

In Figure A.9, we show the correlation in terms of heat maps for the SKUs with smoothed sales in each cluster of the Dell data set. The color bar on the right of each subfigure indicates the

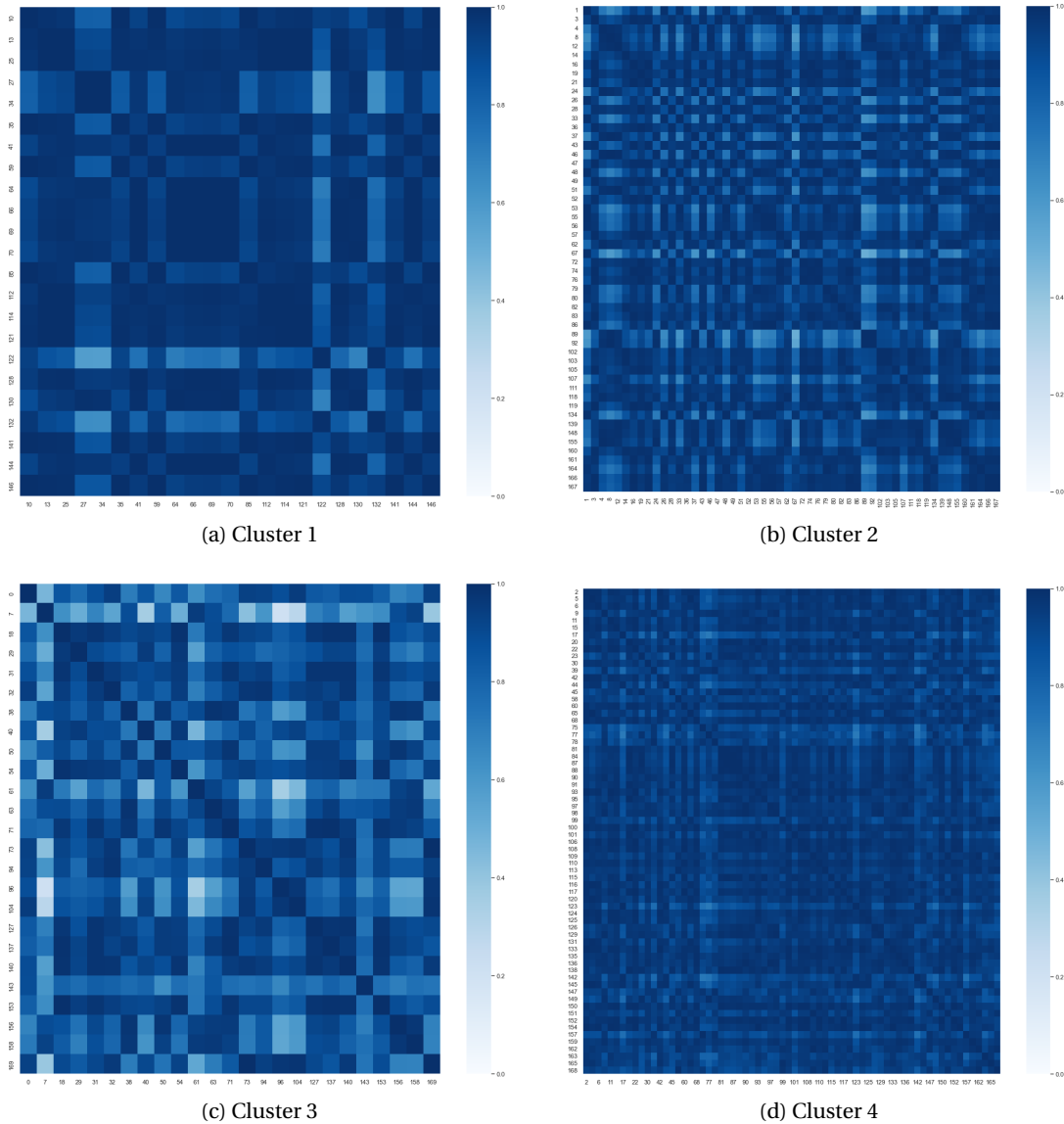


Figure A.9: Correlation heat maps for Dell data set clusters

degree of correlation, with darker color indicating higher correlation. The navy blue color indicates high correlation between the SKUs, which is what we would expect when they are grouped together. We notice however, for cluster 3 in Figure A.9 (c), between some SKUs the correlation is low, as indicated by the very light blue color. We notice that this is because cluster 3 in Figure 1.5 in this thesis has higher variation between sales patterns of the SKUs, which is reflected in the heat map by the lighter color.

Appendix

In Figure A.10, we show the correlation heat maps of 15% randomly chosen SKUs in each cluster of Retailer X. The reason behind the choice of 15% random SKU is because of the large

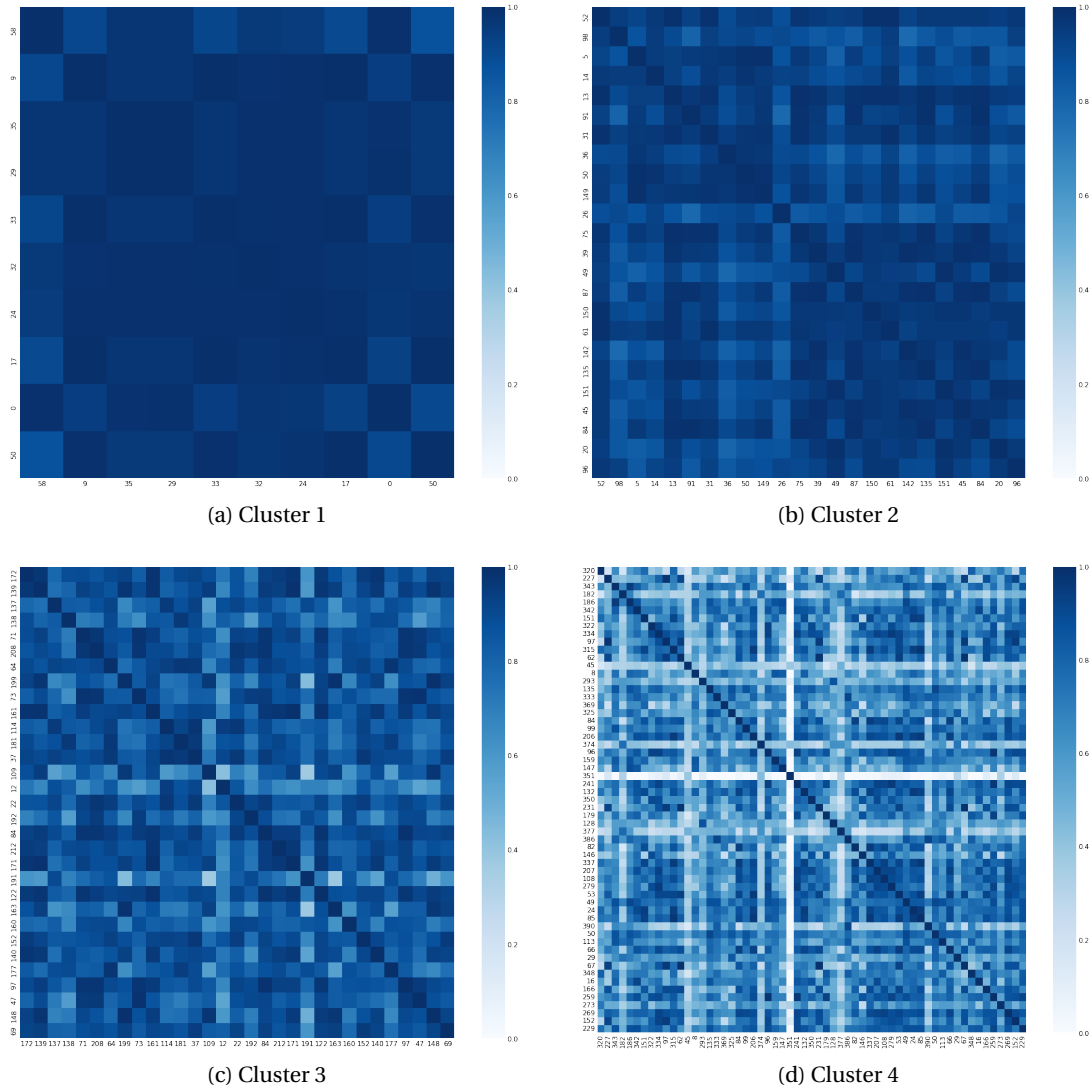


Figure A.10: Correlation heat maps for Retailer X data set clusters (15% of SKUs)

number of curves in each cluster, which makes the display difficult and unclear. However, as can be seen from the results, 15% is sufficient to show that SKUs in cluster 1 are highly correlated, and the correlation between the curves decreases, yet is still high, as we move to clusters 2, 3, and 4, respectively (please refer to the clustering results of Retailer X shown in Figure A.8). Cluster 4 has very high variations in the PLCs, and that is why some of the SKUs show low correlation.

A.3 Integration

The integration method refers to the mathematical concept of continually calculating the area under a curve as more data becomes available. The intuition behind this method rests on the fact that products in different clusters have different life cycle lengths and growth rates, so integrating sales over time would show the rate at which products in each respective cluster are sold. The value of integrating each PLC will equal to 1 since each time-series is normalized to 1. For each average representative curve per cluster and for the new product introduced, we integrate under the sales curve up to time τ , which gives the cumulative sales, CS , at that time and that is defined as follows:

$$CS_{i,\tau} = \int_0^{\tau} \tilde{D}_{i,t}^a dt, \quad (\text{A.1})$$

where i refers to the representative cluster or new product and $\tilde{D}_{i,t}^a$ is the life cycle curve we are integrating (for each cluster this is the average of the smoothed curves within the cluster). We note that due to the discrete time nature of the weekly sales, the integration in our case amounts to the summation of the discretized areas under the sales curves.

Figure A.11 shows the integration results of the sales curve of a randomly selected SKU (red curve) compared with the average representative curves for each cluster (gray curves). The

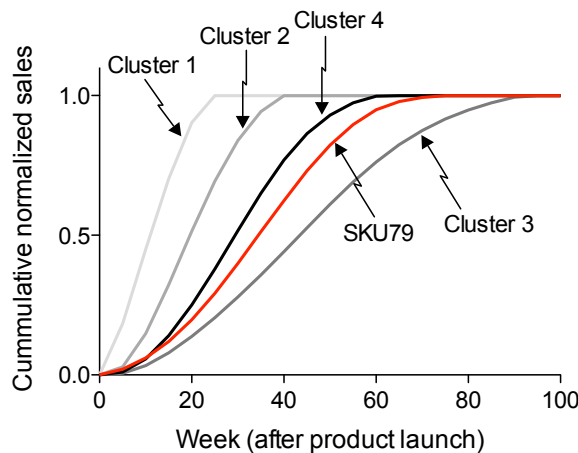


Figure A.11: Assigning products to clusters by integration (Dell data)

cumulative normalized sales curves shown in the figure are the result of integrating the average PLCs per cluster, and the PLC of the new product. To obtain these results, we have considered different values of τ between 0 and 100 (to cover the maximum life cycle length in the data set) and calculated the integration results accordingly. Importantly, however, these curves are not used for assigning products to clusters using DTW nor as an input to the forecasting methods.

It can be observed that sales of the newly launched product behave in a similar way to those of

Appendix

cluster 4's average curve and we would therefore assume that the new product will be assigned to cluster 4. The result of the MAE measure between the behavior of the individual SKU and each of the other representative curves gives a quantitative method of assigning the new product to one of the clusters. Interestingly, some products show similarities to a particular cluster initially, but after waiting for more sales, they start behaving like another cluster. This reinforces the importance of measuring the assignment using different weeks of sales after product launch.

A.4 Dynamic time warping

Dynamic time warping (DTW) is a procedure that measures the similarities between two time-series even when shifts exist in the temporal data or when the time series are of different lengths (Neamtu et al., 2018). Müller (2007) provided a detailed description of how the algorithm works in aligning and measuring the distances between the data points of two time-series. Its main use is in speech recognition to determine whether one phrase matches another, even if the phrases are spoken with wide variations in timings and pronunciations (Berndt and Clifford, 1994). More recently, DTW has been used in other fields such as data mining, information retrieval, bioinformatics, chemical engineering, signal processing, and computer graphics (see the work of Keogh and Ratanamahatana (2005) and the literature cited therein).

The objective of DTW is to find the optimal global alignment between two time-series while exploiting temporal distortions between them. Unlike the standard Euclidean distance, which only measures the vertical distance by aligning the i th point of one time-series with the i th point of another time-series (and therefore creates a poor similarity measure since it cannot detect shifts or transformations), DTW measures the distance between a point of one series and all points of the other series, not just with respect to the vertical point aligned with it. Crucially, it can therefore match similar shapes that are out of phase on the time axis.

In order to do so, DTW considers all the points on the two series. It then calculates the Euclidean distance between the first point of the first time-series and every point of the second time-series, storing the distances calculated in a matrix. These calculations are repeated for all points of both time-series. Subsequently, DTW determines the minimum distance by creating a path from the bottom-left to the upper-right corner of the distance matrix previously calculated. This is done for the curve of the new product compared with each of the average cluster curves, depending on the number of weeks after launch chosen. Finally, the minimum distance found by comparing the newly launched product with each of the four clusters determines which cluster the new product belongs to.

To illustrate the application of dynamic time warping (DTW), we show an example of two time-series from the Dell data set and how DTW finds similarities between them. The example considered is shown in Figure A.12 and is between SKU107 and SKU115, representing two SKUs that have a similar shape yet are phased in time. Figure A.12 shows both the distance matrix (heat map) created between all the points and the warping path (red), which passes through the minimum distances between all the points on both curves. The lighter the color on the heat map, the less distance there is between the two points — this is why the warping path is generated through the lighter regions. Each axis in Figure A.12 represents the sales points of each curve when comparing the two curves. Therefore, if the curves look very similar and are perfectly aligned with each other, the i th point on one curve will match the i th point on the other curve. The warping path (red) is then a straight diagonal line.

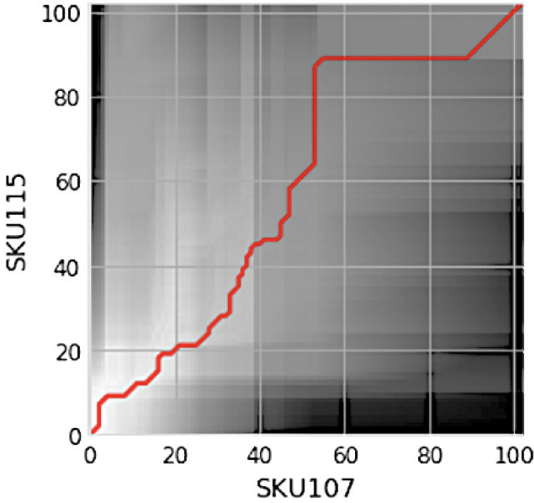


Figure A.12: DTW performance when comparing SKU107 and SKU115 (Dell data)

A.5 Comparison of DTW and Integration using MSE

In order to verify that using the mean absolute error (MAE) for DTW and Integration does not produce biased assignment results, we repeat the cluster assignment steps using the mean squared error (MSE) as an error measure. As can be seen from Figure A.13, the results are almost identical to those reported in the main paper, which suggests that there is no noticeable bias in the correct cluster assignment.

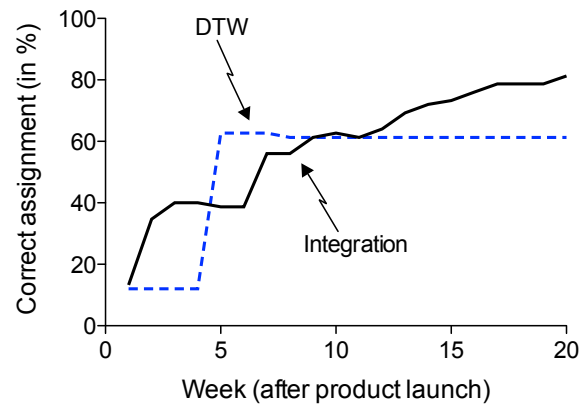


Figure A.13: Results of DTW and integration in percentage of correct assignment using MSE instead of MAE (Dell data)

A.6 Results of using DTW in clustering and cluster assignment

We have repeated the clustering using DTW as the distance measure, rather than the Euclidean distance. The results of correctly assigning products to clusters using DTW (DTW-D) then outperform that of integration (Integration-D), as shown in Figure A.14. We also show the results of using the Euclidean distance for clustering and then assigning the products using Integration (Integration-E) and DTW (DTW-E) on the same figure. The Euclidean distance

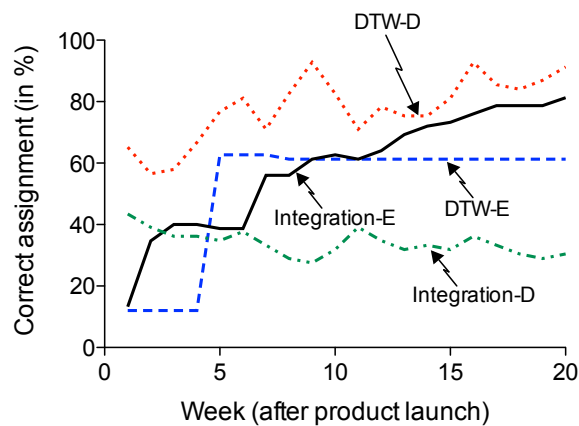


Figure A.14: Comparison of cluster assignment performance of DTW and Integration for various weeks after product launch when clustering on the Dell data set is done using either the DTW (-D) or the Euclidean distance (-E) measure

measures the similarity between each pair of points at the same time-step between two curves, meaning that if curves have similar slopes then they will be grouped together. In this case, integration gives better results in assigning products to the correct clusters as it gives a measurement of the rate of sales of the products, which is represented through such a slope measure. A steeper slope means that the sales evolve faster at the beginning of the PLC. So, the new product is related to a cluster depending on both the rate of sales of the new product and the rate of sales of the representative curve of each cluster.

By contrast, DTW accounts for the similarity between different PLC curves when there are temporal shifts (on the x -axis), which are not taken into account by the Euclidean distance. So, when DTW is used as a distance measure, the products are clustered based on similarities despite temporal shifts, rather than based on slopes and evolution of sales. Therefore, if two time-series are similar in shape but have different slopes and sales evolution, DTW-based clustering would group them together, while the integration method no longer provides good results. As a result, cluster assignment using DTW will start to assign products more correctly as more sales data of the new product are available – the more the shape is known, the higher the percentage of correct cluster assignment (as shown by the upward trend of DTW assignment in Figure A.14).

However, matching normalized curves shifted to zero may defy the purpose of DTW, which is to match time-series with temporal shifts. This means that even if curves occur at different time points, the algorithm will be able to find similarities between them. Therefore, we keep the clustering based on Euclidean distance. For completeness, however, we show the clustering results of the Dell data set using DTW in Figures A.15 and A.16. Comparing the two

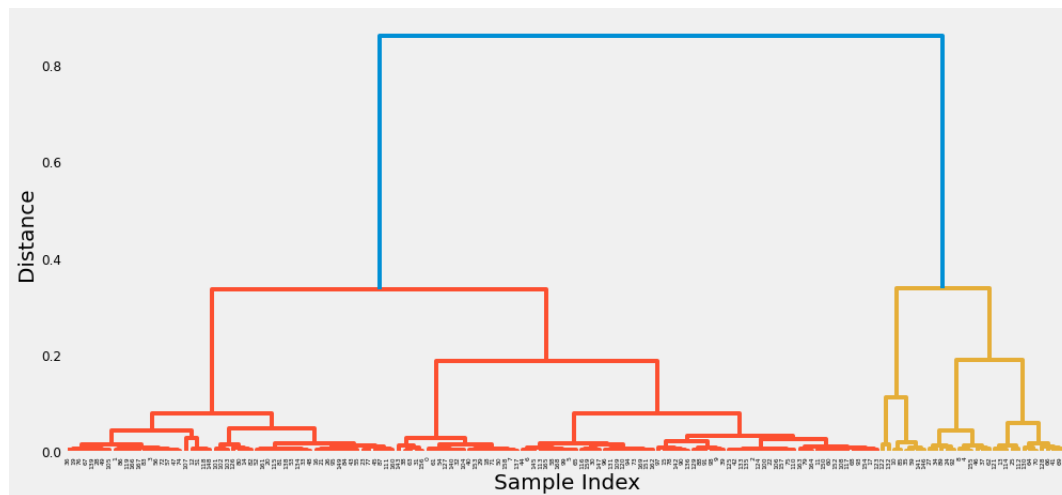


Figure A.15: Dendrogram result from using DTW as a distance measure on the Dell data set

dendrograms in Figures A.15 (when DTW is used) and A.1 (when the Euclidean distance is used), we see that although the dendrogram using DTW in Figure A.15 shows good separation between the time-series in this data set, the separation with the Euclidean distance in Figure A.1 is better due to the greater distance separation between the clusters given a chosen number of clusters. Indeed, the distance between the clusters, which is indicated by the y -axis of the dendrograms, shows greater separation when the Euclidean distance is used.

Apart from the dendrogram results, considering the clusters obtained using DTW (see Figure A.16), it can be observed that cluster 3 has 8 old products only, which may be too little for the ML algorithms. The results also show that the variation between the different products in cluster 4 is somewhat high when DTW is used, with some old products starting with either zero sales or very high sales. These products should be placed in different clusters as they exhibit very different behavior. Based on these results and our analysis, we keep the clustering using the Euclidean distance in the main part of our study.

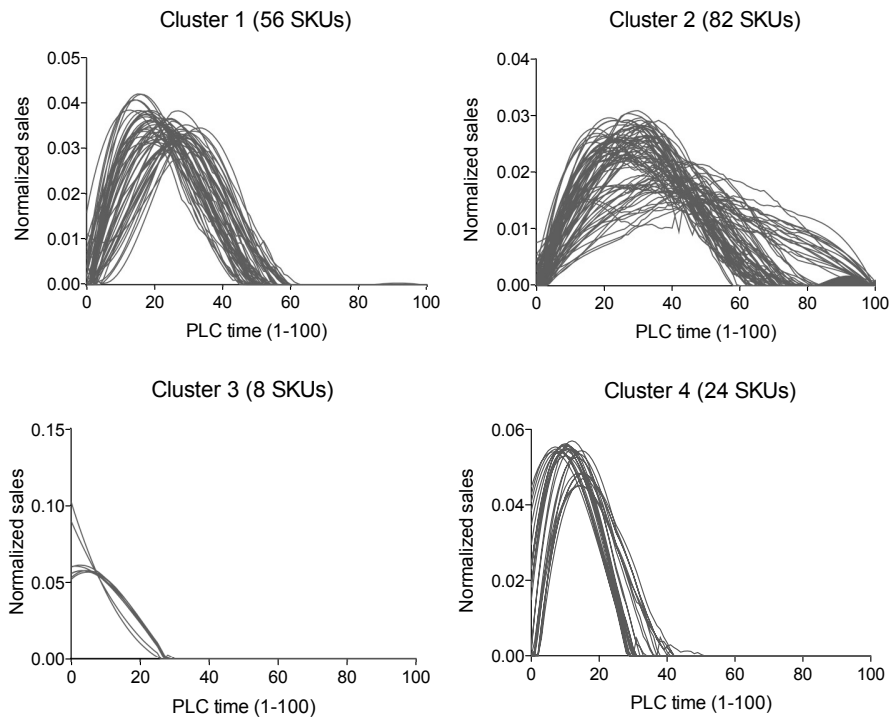


Figure A.16: Clustering result from using DTW as a distance measure on the Dell data set

A.7 Estimation and implementation details

To estimate the parameters for each of the models, we conduct hyper-parameter tuning on a selection of parameters and choose those that yield the smallest error values after cross-validation. In Table A.1, we report both the tuning parameters and ranges used in our grid search for the quantitative methods.

Table A.1: Grid search for hyperparameter tuning

Model	Tuning parameter	Tuning range
ARIMAX	Auto-regressive lags	0, 1, 2, 3, 10, 20, 30, 50
	Difference iterations	0, 1, 2, 3
	Moving-average terms	0, 1, 2, 3
LSTM & GRUs	Number of neurons in hidden layer	10, 50, 100, 200, 500, 700
	Step size	1, 3, 20, 30, 50
CNNs	Number of filters	8, 10, 16, 32, 64
	Step size	1, 3, 20, 30, 50

For the autoregressive integrated moving average with exogenous variables (ARIMAX) model, we use the `statsmodels.tsa.statespace.sarimax` library in Python. To specify the model for our data, we input the different sets of hyperparameters, the first few weeks of smoothed sales of the new product we are testing for, and the smoothed sales values for all other products in the cluster, over their entire life cycle. The model would then fit the first few weeks of sales of the new product and of the older products to the hyperparameters and provide the Akaike Information Criteria (AIC) for each order of parameters – the number of autoregressive terms p , the parameter for difference iterations d , and for moving-average q . We choose the order of parameters with the lowest AIC value. For that order, the model would then use the input data of the new product and of the old products to forecast the future sales of the new product.

For the deep neural networks (DNNs), we implement all the code in Python using the Keras library installed with TensorFlow (details of Python packages found in Appendix E of supplementary materials). More specifically, the DNN models are implemented as follows. We use *Sequential* models for all the DNNs considered here: long short-term memory (LSTM), gated recurrent units (GRUs), and convolutional neural networks (CNNs). The activation function we use is *ReLU*. We add several layers of hidden neurons and define the three-dimensional input shape based on the step size we chose, and then add one *Dense* layer. Subsequently, we compile the model using the *Adam* optimizer and fit the model on the X_{train} and y_{train} , which represent the training (with validation) set and training vector, respectively. For CNNs, we add a *Conv1D* layer with filters, a *MaxPooling1D* layer, a *Flatten* layer, and more *Dense* layers.

We tune the parameters of the DNN algorithms using cross-validation and, following general conventions, select 15% of the training samples for validation. We use grid search and, on the

Appendix

validation set, we measure which parameters for the number of steps, layers, and filters (filters for CNNs only) give the lowest mean absolute scaled error (MASE). We use those parameters for the model. After the model is defined, we use the X_{test} matrix, which represents the test set, as input to the prediction part to forecast the rest of the new PLC. Lastly, the results of the forecasting are compared to the testing vector, y_{test} , and the resulting forecast errors are determined and reported.

Table A.2 reports the packages used in Python (version 3.8.5) for the implementation of the main algorithms deployed in our study.

Table A.2: Python packages used for implementation

Step	Package	Version
Clustering	scipy.cluster.hierarchy (hac, dendrogram)	1.5.2
Cluster assignment (DTW)	fastdtw	0.3.4
Data augmentation	scipy.interpolate (interpld)	1.5.2
ARIMAX	statsmodels.tsa.statespace.sarimax (SARIMAX)	0.12.0
ML methods	tensorflow.keras, tensorflow.keras.layers (Sequential, LSTM, GRU, Conv1D, Dense, Flatten)	2.4.0

A.8 Error measurements of PLC shape-based methods

In this section, we report the complete set of forecast errors of the PLC shape-based methods. Superscripts 1 and 2 denote that the results of the curve-fitting methods were obtained by fitting the parameters to the entire *GenerateAverage/GenerateFit* curves per cluster and only to the part after the introduction weeks, respectively.

For the Dell data set (see Table A.3), we notice that for the polyn base curves both variants – i.e. 1 and 2 – perform very similarly. This is due to the fact that the smoothed PLCs of the Dell data

Table A.3: Forecast errors of PLC shape-based methods for new products for different weeks after introduction (Dell data)

Base curve family	Cluster curve generation	MASE			SAE		
		10 weeks	15 weeks	20 weeks	10 weeks	15 weeks	20 weeks
Bass-I	Fit ¹	0.99	1.01	1.05	0.38	0.33	0.28
	Fit ²	1.20	1.47	1.71	0.44	0.45	0.42
	Avg ¹	0.97	1.00	1.04	0.38	0.33	0.28
	Avg ²	1.22	1.48	1.72	0.45	0.45	0.42
poly2	Fit ¹	1.07	1.10	1.18	0.41	0.36	0.30
	Fit ²	1.09	1.12	1.21	0.42	0.37	0.31
	Avg ¹	0.93	0.94	0.98	0.36	0.32	0.26
	Avg ²	0.96	0.98	1.09	0.37	0.32	0.27
poly3	Fit ¹	1.10	1.14	1.21	0.42	0.37	0.31
	Fit ²	1.10	1.15	1.23	0.42	0.38	0.32
	Avg ¹	0.97	0.98	1.02	0.37	0.32	0.27
	Avg ²	0.98	0.99	1.04	0.38	0.33	0.28
poly4	Fit ¹	1.11	1.15	1.23	0.43	0.38	0.32
	Fit ²	1.11	1.15	1.23	0.43	0.38	0.32
	Avg ¹	0.98	1.00	1.05	0.38	0.33	0.28
	Avg ²	0.98	1.00	1.04	0.38	0.33	0.28
Trapezoid	Fit ¹	1.16	1.19	1.26	0.44	0.38	0.32
	Fit ²	1.16	1.64	1.59	0.44	0.52	0.39
	Avg ¹	1.11	1.13	1.14	0.42	0.36	0.30
	Avg ²	1.18	2.09	1.38	0.44	0.66	0.36
Triangle	Fit ¹	1.12	1.14	1.22	0.42	0.37	0.31
	Fit ²	1.14	1.26	1.34	0.43	0.39	0.34
	Avg ¹	1.10	1.14	1.21	0.42	0.37	0.31
	Avg ²	1.16	1.20	1.32	0.44	0.38	0.33
Bass-II ¹	-	1.09	0.94	0.97	0.38	0.34	0.26
Bass-II ²	-	1.08	0.99	1.05	0.38	0.34	0.28

set follow a pattern that can be easily fitted using either the entire *GenerateAverage/GenerateFit*, or just part of them. On the other hand, for the Bass, Trapezoid, and Triangle families, fitting to the entire PLC generally results in more accurate forecasts as the shapes of those curve fitting-based methods are better represented when considering the entire PLC, rather than just part of it.

For the Retailer X data set (see Table A.4), we notice that, in general, fitting the parameters

Appendix

according to variant 2 gives better results, except for Bass-I Avg and poly2 Fit. This can be

Table A.4: Forecast errors of PLC shape-based methods for new products for different weeks after introduction (Retailer X data)

Base curve family	Cluster curve generation	MASE			SAE		
		6 weeks	8 weeks	10 weeks	6 weeks	8 weeks	10 weeks
Bass-I	Fit ¹	1.73	1.87	2.11	0.83	0.75	0.65
	Fit ²	1.20	1.20	1.30	0.56	0.50	0.49
	Avg ¹	1.10	1.25	1.25	0.53	0.49	0.45
	Avg ²	1.16	1.17	1.24	0.54	0.50	0.48
poly2	Fit ¹	1.40	1.46	1.57	0.67	0.62	0.58
	Fit ²	1.70	2.13	2.43	0.83	0.95	0.99
	Avg ¹	1.22	1.74	2.01	0.54	0.50	0.47
	Avg ²	1.15	1.40	1.54	0.54	0.50	0.47
poly3	Fit ¹	2.31	3.21	3.83	1.09	1.03	0.97
	Fit ²	2.11	2.85	3.30	0.99	0.83	0.71
	Avg ¹	1.23	1.66	1.92	0.56	0.52	0.48
	Avg ²	1.14	1.39	1.54	0.54	0.50	0.46
poly4	Fit ¹	2.30	3.22	4.12	1.06	1.00	0.93
	Fit ²	1.61	1.82	2.14	0.78	0.71	0.64
	Avg ¹	1.16	1.45	1.61	0.55	0.51	0.47
	Avg ²	1.13	1.38	1.53	0.54	0.50	0.46
Trapezoid	Fit ¹	1.97	2.16	2.58	0.93	0.87	0.82
	Fit ²	1.29	1.36	1.34	0.61	0.58	0.51
	Avg ¹	1.29	1.75	2.03	0.58	0.54	0.50
	Avg ²	1.29	1.30	1.25	0.60	0.54	0.47
Triangle	Fit ¹	1.42	1.96	2.21	0.65	0.62	0.58
	Fit ²	1.28	1.36	1.39	0.60	0.55	0.50
	Avg ¹	1.31	1.85	2.24	0.59	0.54	0.50
	Avg ²	1.20	1.25	1.26	0.56	0.51	0.47
Bass-II ¹	-	1.13	1.24	1.26	0.56	0.51	0.47
Bass-II ²	-	1.11	1.17	1.24	0.55	0.49	0.47

explained due to the fact that some of the PLCs in the Retailer X data set follow irregular (i.e. non-classical) life cycles with two maturity stages. So fitting the parameters starting from a point of sales closer to the PLC's end typically gives better results. To provide some intuition into this, consider the example of a PLC with two peaks, which would not be represented well by a triangle. If we fit the parameters right after the first peak, however, then a triangular representation may well give a better forecast of the forthcoming sales.

A.9 Evaluating out-of-sample forecast accuracy

The formulas for calculating the mean absolute scaled error (MASE) and the sum of absolute errors (SAE) per product $i \in \mathcal{I}$, where \mathcal{I} is the set of products, are given by (A.2) and (A.3), respectively.

$$\text{MASE}(i) = \frac{\sum_{t=\tau+1}^{T_i} |\hat{D}_t^{p,i} - D_t^{a,i}|}{(T_i / (T_i - 1)) \sum_{t=\tau+2}^{T_i} |D_t^{a,i} - D_{t-1}^{a,i}|}, \quad \forall i \in \mathcal{I}, \quad (\text{A.2})$$

$$\text{SAE}(i) = \sum_{t=\tau+1}^{T_i} |\hat{D}_t^{p,i} - D_t^{a,i}|, \quad \forall i \in \mathcal{I}, \quad (\text{A.3})$$

where $\hat{D}_t^{p,i}$ and $D_t^{a,i}$ denote the forecasted and actual (unsmoothed) sales of product i at time t , respectively, τ is the time a forecast is generated, and T_i denotes the time product $i \in \mathcal{I}$ sells in the market (i.e. its life cycle length). The errors are calculated per product i and then averaged, that is, $\text{MASE} = 1/|\mathcal{I}| \sum_{i \in \mathcal{I}} \text{MASE}(i)$ and $\text{SAE} = 1/|\mathcal{I}| \sum_{i \in \mathcal{I}} \text{SAE}(i)$.

A.10 Error distributions

We plot the box and whiskers of the errors of ARIMAX and the three ML methods at 20 weeks after product launch in Figure A.17 to illustrate their performance with respect to evaluation metrics other than the mean. More specifically, we report both the results of running the

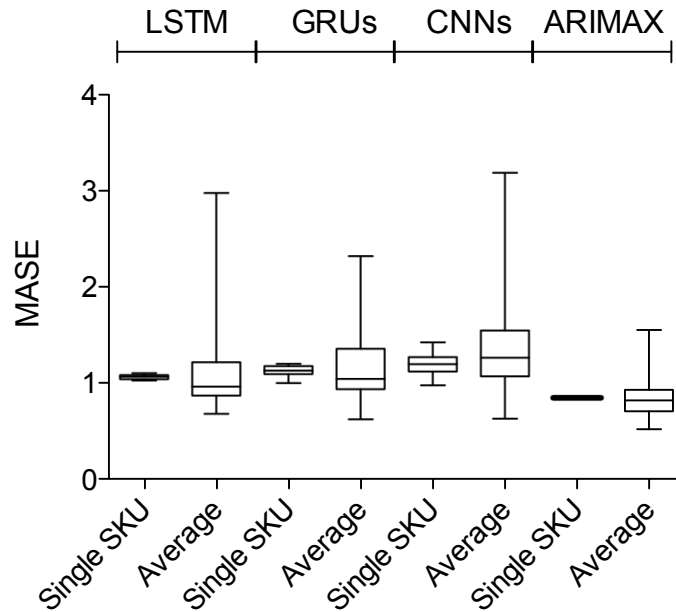


Figure A.17: Forecast errors (MASE) of the quantitative methods using the Dell data set with 20 weeks of input

forecasting methods on a single SKU ten times and the results of the average errors obtained for the ML methods with clustering and for the ARIMAX without clustering using the Dell data. Considering the variance in the distribution of forecast errors, we observe that for a single SKU, using LSTM, GRUs, or ARIMAX gives very similar results over the runs, whereas CNNs are less stable and analyzing the average error of several runs per SKU should therefore be considered.

In Figure A.18, we plot the box and whiskers for the MASE of the PLC shape-based methods at 20 weeks after product introduction using the Dell data set. We only report here the results of the average errors obtained for each method but not those of the average runs of a single SKU as these always give the same results, independent of the number of runs. Superscripts *1* and *2* denote that the results for the PLC shape-based methods were obtained by fitting the parameters to the entire *GenerateAverage/GenerateFit* curves per cluster and only to the part after the introduction weeks, respectively. For the sake of compactness, in Figure A.18 we plot only the error distributions for the superscripts with the best results for 20 weeks after product launch according to Table A.3.

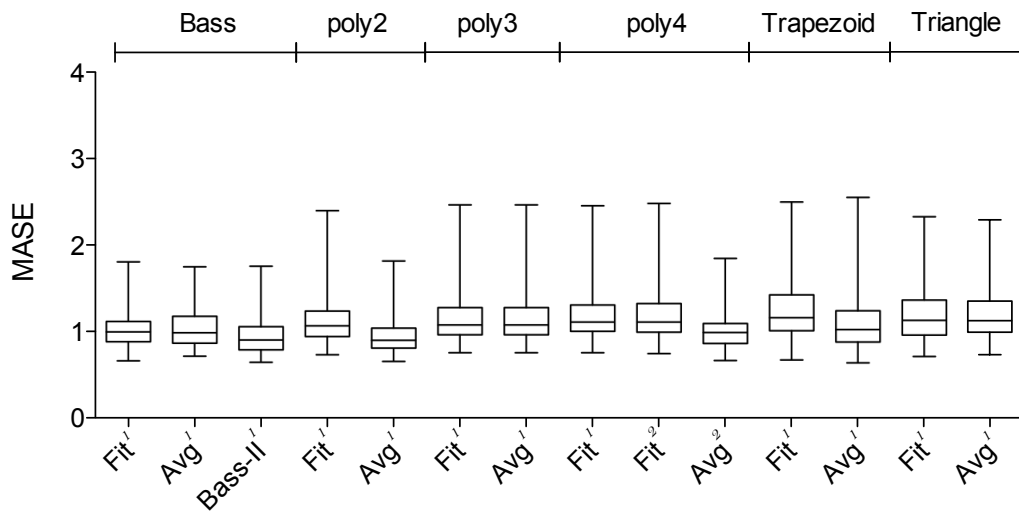


Figure A.18: Forecast errors (MASE) of the PLC shape-based methods using the Dell data set with 20 weeks of input

A.11 Additional results for incorrect cluster assignment analysis

Considering the same three sets of randomly selected SKUs as in Section 7.2 but using all products in the Dell data set (rather than just those of a cluster) when forecasting sales of products from Clusters 1-4, we obtain the average MASE results shown in Table A.5.

Table A.5: Average MASE results when forecasting using all products in the Dell data set (i.e. without cluster assignment)

Product from	Quantitative method			
	ARIMAX	LSTM	GRUs	CNNs
Cluster 1	1.77	1.36	1.48	1.30
Cluster 2	0.85	1.34	1.35	1.49
Cluster 3	1.03	1.60	1.55	1.71
Cluster 4	0.86	1.31	1.30	1.46

B Supplementary material for Chapter 2

B.1 Proof of Theorem 1

To solve for the order quantities in the mathematical problem (9)–(11), we transform the SP model into an LP model as demonstrated by Shapiro et al. (2009). Let $\epsilon_{i,r}$ denote a realization of ϵ_i for product $i \in \mathcal{I}$ such that $r \in \mathcal{R}$, where \mathcal{I} is the set of n products offered by the retailer and \mathcal{R} is defined as a large finite set of integers that represent randomness in demand occurring with positive probabilities. We use \mathcal{W}_i^r to denote the overage quantity realized for product i , such that $\mathcal{W}_i^r = \mathcal{W}_i(q_i, d_i(p_i, \epsilon_{i,r}))$, where $\mathcal{W}_i = \max\{0, q_i - d_i(p_i, \epsilon_{i,r})\}$ and \mathcal{V}_i^r to denote the underage quantity realized for product i , such that $\mathcal{V}_i^r = \mathcal{V}_i(q_i, d_i(p_i, \epsilon_{i,r}))$, where $\mathcal{V}_i = \max\{0, d_i(p_i, \epsilon_{i,r}) - q_i\}$. Then, we write the SP model (9)–(11) as a large-scale LP model:

$$\max_{\mathbf{Q}, \forall i \in \mathcal{I}} z = \sum_{i=1}^n (p_i - c_i) d_i(p_i, \mu) \quad (\text{B.1})$$

$$- \sum_{i=1}^n y(p_i) \left\{ (c_i - s_i) \sum_{r \in \mathcal{R}} \text{Pr}(\epsilon_{i,r}) \mathcal{W}_i^r + (p_i - c_i + g_i) \sum_{r \in \mathcal{R}} \text{Pr}(\epsilon_{i,r}) \mathcal{V}_i^r \right\} \quad (\text{B.2})$$

$$\text{subject to: } \mathcal{W}_i^r - (q_i - d_i(p_i, \epsilon_{i,r})) \geq 0, \quad \forall i \in \mathcal{I}, \quad \forall r \in \mathcal{R} \quad (\text{B.3})$$

$$\mathcal{V}_i^r - (d_i(p_i, \epsilon_{i,r}) - q_i) \geq 0, \quad \forall i \in \mathcal{I}, \quad \forall r \in \mathcal{R} \quad (\text{B.4})$$

$$\sum_{i=1}^n q_i \leq K, \quad (\text{B.5})$$

$$q_i \geq 0, \quad \mathcal{W}_i^r \geq 0, \quad \mathcal{V}_i^r \geq 0, \quad \forall i \in \mathcal{I}, \quad \forall r \in \mathcal{R}. \quad (\text{B.6})$$

We add constraints (B.3) and (B.4) to satisfy the conditions $\mathcal{W}_i = \max\{0, q_i - d_i(p_i, \epsilon_{i,r})\}$ and

$\mathcal{V}_i = \max\{0, d_i(p_i, \epsilon_{i,r}) - q_i\}$, respectively. We then formulate the LP in its standard form:

$$\max_{\mathbf{Q}, \forall i \in \mathcal{I}} z = \sum_{i=1}^n (p_i - c_i + g_i) d_i(p_i, \mu) \quad (\text{B.7})$$

$$- \sum_{i=1}^n y(p_i) \left\{ (c_i - s_i) \sum_{r \in \mathcal{S}} \Pr(\epsilon_{i,r}) \mathcal{W}_i^r + (p_i - c_i) \sum_{r \in \mathcal{S}} \Pr(\epsilon_{i,r}) \mathcal{V}_i^r \right\} \quad (\text{B.8})$$

$$\text{subject to: } -\mathcal{W}_i^r + q_i \leq d_i(p_i, \epsilon_{i,r}), \quad \forall i \in \mathcal{I}, \quad \forall r \in \mathcal{S} \quad (\text{B.9})$$

$$-\mathcal{V}_i^r - q_i \leq -d_i(p_i, \epsilon_{i,r}), \quad \forall i \in \mathcal{I}, \quad \forall r \in \mathcal{S} \quad (\text{B.10})$$

$$\sum_{i=1}^n q_i \leq K, \quad (\text{B.11})$$

$$q_i \geq 0, \quad \mathcal{W}_i^r \geq 0, \quad \mathcal{V}_i^r \geq 0, \quad \forall i \in \mathcal{I}, \quad \forall r \in \mathcal{S}. \quad (\text{B.12})$$

The dual of the problem is:

$$\min_{\beta_{i,r}, \rho_{i,r}, \gamma} w = \sum_{i=1}^n \sum_{r \in \mathcal{S}} d_i(p_i, \epsilon_{i,r}) (\beta_{i,r} - \rho_{i,r}) + \gamma K \quad (\text{B.13})$$

$$\text{subject to: } \beta_{i,r} \leq y(p_i) (c_i - s_i) \Pr(\epsilon_{i,r}), \quad \forall i \in \mathcal{I}, \quad \forall r \in \mathcal{S} \quad (\text{B.14})$$

$$\rho_{i,r} \leq y(p_i) (p_i - c_i + g_i) \Pr(\epsilon_{i,r}), \quad \forall i \in \mathcal{I}, \quad \forall r \in \mathcal{S} \quad (\text{B.15})$$

$$\sum_{r \in \mathcal{S}} (\beta_{i,r} - \rho_{i,r}) + \gamma \geq 0, \quad \forall i \in \mathcal{I}, \quad \forall r \in \mathcal{S}, \quad (\text{B.16})$$

$$\beta_{i,r} \geq 0, \quad \rho_{i,r} \geq 0, \quad \gamma \geq 0, \quad \forall i \in \mathcal{I}, \quad \forall r \in \mathcal{S}, \quad (\text{B.17})$$

Similar to Biçer et al. (2021), the values of $\rho_{i,r}$ and $\beta_{i,r} \forall i \in \mathcal{I}$ are found by the parametric analysis:

- $\beta_{i,r} = y(p_i) (c_i - s_i) \Pr(\epsilon_{i,r})$ and $\rho_{i,r} = 0$ for $i \in \mathcal{I}$ when $d_i(p_i, \epsilon_{i,r}) \leq q_i$.
- $\beta_{i,r} = 0$ and $\rho_{i,r} = y(p_i) (p_i - c_i + g_i) \Pr(\epsilon_{i,r})$ when $d_i(p_i, \epsilon_{i,r}) > q_i$.

Then, constraint (B.16) becomes:

$$y(p_i) (c_i - s_i) \Pr(d_i(p_i, \epsilon_{i,r}) \leq q_i) - y(p_i) (p_i - c_i + g_i) \Pr(d_i(p_i, \epsilon_{i,r}) > q_i) + \gamma \geq 0, \quad \forall i \in \mathcal{I} \quad (\text{B.18})$$

We set the dual variable γ to:

$$\gamma = -y(p_i) (c_i - s_i) \Pr(d_i(p_i, \epsilon_{i,r}) \leq q_i) + y(p_i) (p_i - c_i + g_i) \Pr(d_i(p_i, \epsilon_{i,r}) > q_i) = g_i(q_i, d_i(p_i, \epsilon_{i,r})), \quad \forall i \in \mathcal{I}, \quad (\text{B.19})$$

where $g_i(q_i, d_i(p_i, \epsilon_{i,r})) = \partial E[II(q_i, p_i)] / \partial q_i$. The objective function value for the dual prob-

lem becomes:

$$\begin{aligned}
 w = & \\
 & \sum_{i=1}^n d_i(p_i, \epsilon_{i,r}) \left[y(p_i) (c_i - s_i) Pr(d_i(p_i, \epsilon_{i,r}) \leq q_i) - y(p_i) (p_i - c_i + g_i) Pr(d_i(p_i, \epsilon_{i,r}) > q_i) \right] + \\
 & \sum_{i=1}^n q_i \left[-y(p_i) (c_i - s_i) Pr(d_i(p_i, \epsilon_{i,r}) \leq q_i) + y(p_i) (p_i - c_i + g_i) Pr(d_i(p_i, \epsilon_{i,r}) > q_i) \right],
 \end{aligned} \tag{B.20}$$

which simplifies to:

$$\begin{aligned}
 w = & - \sum_{i=1}^n \left\{ y(p_i) (c_i - s_i) \int_0^{q_i} (q_i - d_i(p_i, \epsilon_i)) f_i(d_i(p_i, \epsilon_i)) \partial(d_i(p_i, \epsilon_i)) + \right. \\
 & \left. y(p_i) (p_i - c_i + g_i) \int_{q_i}^{\infty} (d_i(p_i, \epsilon_i) - q_i) f_i(d_i(p_i, \epsilon_i)) \partial(d_i(p_i, \epsilon_i)) \right\}.
 \end{aligned} \tag{B.21}$$

Equation (B.21) is equivalent to the solution of the primal problem for the feasible q_i values $\forall i \in \mathcal{I}$, where the value of w in brackets is equivalent to the marginal profit of a product when ordering one additional unit of it (Biçer and Seifert (2017), Biçer et al. (2021)). From Equation (B.21), we characterize the optimal allocation policy given the total available capacity, K , for products $i \in \mathcal{I}$ as a nested allocation policy if the capacity constraint is binding. With the nested allocation policy, products are distinguished by their marginal profits (the expression inside the curly brackets) and ranked from the highest marginal profit to the lowest. The capacity available to different products follows a hierarchical manner. Products with the highest marginal profit level are allocated capacity first, until their marginal profit decreases and is equal to the marginal profit of the next class of products (second highest marginal profit). Then, all products with the current highest marginal profit are allocated capacity until their marginal profit equates the next marginal profit level, and so on, until the capacity is fully utilized.

From strong duality, the optimal solution satisfies Equation (B.19). Hence, the optimality conditions are:

$$\gamma = g_i(q_i, d_i(p_i, \epsilon_i)) \geq 0, \quad \forall i \in \mathcal{I} \tag{B.22}$$

$$\sum_{i=1}^n q_i \leq K, \tag{B.23}$$

It follows that if constraint (B.11) is non-binding, then the optimal solution reduces to the solution of n independent newsvendor problems, where n is the number of products offered by the retailer. Else, the optimal solution corresponds to the point at which the marginal value of ordering one unit is equivalent for all products $i \in \mathcal{I}$ (Biçer and Seifert (2017), Biçer et al. (2021)).

For each product $i \in \mathcal{I}$, the optimal quantity to order under the capacity constraint is:

$$q_i = \begin{cases} q_i^* & \text{if } \sum_{i \in \mathcal{I}} q_i^* < K \\ \hat{q}_i & \text{if } \sum_{i \in \mathcal{I}} q_i^* \geq K \end{cases} \quad (\text{B.24})$$

where q_i^* satisfies the following expression:

$$\begin{aligned} g_i(q_i, d_i(p_i, \epsilon_i)) &= -(c_i - s_i) \int_0^{q_i} (q_i - d_i(p_i, \epsilon_i)) f_i(d_i(p_i, \epsilon_i)) \partial(d_i(p_i, \epsilon_i)) \\ &+ (p_i - c_i + g_i) \int_{q_i}^{\infty} (d_i(p_i, \epsilon_i) - q_i) f_i(d_i(p_i, \epsilon_i)) \partial(d_i(p_i, \epsilon_i)) = 0. \end{aligned} \quad (\text{B.25})$$

Equation (B.24) implies that the problem for optimizing the quantities given a capacity constraint reduces to a single product newsvendor model if the quantity constraint is not binding. If the constraint is binding, the optimal solution exists at the point where the marginal value of ordering one additional unit is the same for all products $i \in \mathcal{I}$ (Biçer and Seifert (2017), Biçer et al. (2021)).

B.2 Proof of Concavity

Taking the first- and second-order derivatives of the expected profit equation given by Equation (2.11) with respect to p_i for a given z_i , we obtain the following:

$$\begin{aligned} \frac{\partial E[\Pi(q_i, p_i)]}{\partial p_i} = & \\ (b_i - 1) \frac{y_i(p_i)}{p_i} [\mu_i - \Theta_i(z_i)] \cdot & \left\{ \frac{b_i c_i}{b_i - 1} + \frac{b_i}{b_i - 1} \left[\frac{(c_i - s_i) \Lambda_i(z_i) + g_i \Theta_i(z_i)}{\mu_i - \Theta_i(z_i)} \right] - p_i \right\}, \end{aligned} \quad (\text{B.26})$$

$$\begin{aligned} \frac{\partial^2 E[\Pi(q_i, p_i)]}{\partial p_i^2} = (b_i - 1) \frac{y(p_i)}{p_i} [\mu_i - \Theta_i(z_i)] & \\ \cdot \left\{ \frac{(-b_i - 1)}{p_i} \cdot \left\{ \frac{b_i c_i}{b_i - 1} + \frac{b_i}{b_i - 1} \left[\frac{(c_i - s_i) \Lambda_i(z_i) + g_i \Theta_i(z_i)}{\mu_i - \Theta_i(z_i)} \right] - p_i \right\} - 1 \right\}, & \end{aligned} \quad (\text{B.27})$$

which prove that the expected profit is concave in p_i for a given z_i .

B.3 Proof of Theorem 2

Expected profit maximization with respect to price can be formulated as follows:

$$\max_{\mathbf{p}} E[\Pi(\mathbf{z}, \mathbf{p})] = \sum_{i=1}^n (p_i - c_i) y_i(p_i) \mu_i - \sum_{i=1}^n y_i(p_i) [(c_i - s_i) \Lambda_i(z_i) + (p_i - c_i + g_i) \Theta_i(z_i)], \quad (\text{B.28})$$

subject to:

$$\sum_{i=1}^n y_i(p_i) z_i \leq K. \quad (\text{B.29})$$

Assuming the retailer sells three products, i.e., $i \in \{1, 2, 3\}$, and assuming the products have the same cost and demand parameters as products 1, 2, and 3 in the numerical example in Section 4.3 in this paper, we expand the mathematical problem (B.28) – (B.29):

$$\begin{aligned} & \max_{\mathbf{p}} E[\Pi(\mathbf{z}, \mathbf{p})] \\ & = (p_1 - c) y(p_1) \mu_1 + (p_2 - c) y(p_2) \mu_2 + (p_3 - c) y(p_3) \mu_3 \\ & \quad - y(p_1) [(c - s) \Lambda_1(z_1) + (p_1 - c + g) \Theta_1(z_1)] \\ & \quad - y(p_2) [(c - s) \Lambda_2(z_2) + (p_2 - c + g) \Theta_2(z_2)] \\ & \quad - y(p_3) [(c - s) \Lambda_3(z_3) + (p_3 - c + g) \Theta_3(z_3)], \end{aligned} \quad (\text{B.30})$$

subject to:

$$y_1(p_1) z_1 + y_2(p_2) z_2 + y_3(p_3) z_3 \leq K. \quad (\text{B.31})$$

Using the Lagrangian formulation, a candidate optimal solution must satisfy the following:

$$\begin{aligned} & \frac{\partial E[\Pi(\mathbf{z}, \mathbf{p})]}{\partial p_1} = \\ & (b-1) \frac{y(p_1)}{p_1} [\mu_1 - \Theta_1(z_1)] \cdot \left\{ \frac{bc}{b-1} + \frac{b}{b-1} \left[\frac{(c-s) \Lambda_1(z_1) + g \Theta_1(z_1)}{\mu_1 - \Theta_1(z_1)} \right] - p_1 \right\} - \frac{\lambda b y(p_1) z_1}{p_1} = 0 \end{aligned} \quad (\text{B.32})$$

$$\begin{aligned} & \frac{\partial E[\Pi(\mathbf{z}, \mathbf{p})]}{\partial p_2} = \\ & (b-1) \frac{y(p_2)}{p_2} [\mu_2 - \Theta_2(z_2)] \cdot \left\{ \frac{bc}{b-1} + \frac{b}{b-1} \left[\frac{(c-s) \Lambda_2(z_2) + g \Theta_2(z_2)}{\mu_2 - \Theta_2(z_2)} \right] - p_2 \right\} - \frac{\lambda b y(p_2) z_2}{p_2} = 0 \end{aligned} \quad (\text{B.33})$$

$$\begin{aligned} \frac{\partial E[\Pi(\mathbf{z}, \mathbf{p})]}{\partial p_3} = \\ (b-1) \frac{y(p_3)}{p_3} [\mu_3 - \Theta_3(z_3)] \cdot \left\{ \frac{bc}{b-1} + \frac{b}{b-1} \left[\frac{(c-s)\Lambda_3(z_3) + g\Theta_3(z_3)}{\mu_1 - \Theta_3(z_3)} \right] - p_3 \right\} - \frac{\lambda b y(p_3) z_3}{p_3} = 0 \end{aligned} \quad (\text{B.34})$$

$$\lambda(y(p_1)z_1 + y(p_2)z_2 + y(p_3)z_3 - K) = 0 \quad (\text{B.35})$$

When $\lambda > 0$, i.e., there is a binding capacity constraint, $y(p_1)z_1 + y(p_2)z_2 + y(p_3)z_3 = K$, and the optimal price values are given by:

$$p_1^* = \frac{bc}{b-1} + \frac{b}{b-1} \left[\frac{(c-s)\Lambda_1(z_1) + g\Theta_1(z_1)}{\mu_1 - \Theta_1(z_1)} \right] - \frac{b\lambda z_1}{(b-1)[\mu_1 - \Theta_1(z_1)]}, \quad (\text{B.36})$$

$$p_2^* = \frac{bc}{b-1} + \frac{b}{b-1} \left[\frac{(c-s)\Lambda_2(z_2) + g\Theta_2(z_2)}{\mu_2 - \Theta_2(z_2)} \right] - \frac{b\lambda z_2}{(b-1)[\mu_2 - \Theta_2(z_2)]}, \quad (\text{B.37})$$

$$p_3^* = \frac{bc}{b-1} + \frac{b}{b-1} \left[\frac{(c-s)\Lambda_3(z_3) + g\Theta_3(z_3)}{\mu_3 - \Theta_3(z_3)} \right] - \frac{b\lambda z_3}{(b-1)[\mu_3 - \Theta_3(z_3)]}. \quad (\text{B.38})$$

Since the z_i values of all three products are related through Equation (B.35), p_i is dependent on p_j through z_j for $j \neq i$. Furthermore, by solving for λ from Equations (B.36) – (B.38), we obtain:

$$\begin{aligned} \lambda &= \frac{(b-1)[\mu_1 - \Theta_1(z_1)]}{bz_1} \cdot \left\{ \frac{bc}{b-1} + \frac{b}{b-1} \left[\frac{(c-s)\Lambda_1(z_1) + g\Theta_1(z_1)}{\mu_1 - \Theta_1(z_1)} \right] - p_1^* \right\} \\ &= \frac{(b-1)[\mu_2 - \Theta_2(z_2)]}{bz_2} \cdot \left\{ \frac{bc}{b-1} + \frac{b}{b-1} \left[\frac{(c-s)\Lambda_2(z_2) + g\Theta_2(z_2)}{\mu_2 - \Theta_2(z_2)} \right] - p_2^* \right\} \\ &= \frac{(b-1)[\mu_3 - \Theta_3(z_3)]}{bz_3} \cdot \left\{ \frac{bc}{b-1} + \frac{b}{b-1} \left[\frac{(c-s)\Lambda_3(z_3) + g\Theta_3(z_3)}{\mu_1 - \Theta_3(z_3)} \right] - p_3^* \right\}, \end{aligned} \quad (\text{B.39})$$

which shows that when capacity is limited, i.e., for a fixed $\lambda > 0$, a change in one of the prices affects the other prices of the products.

When $\lambda = 0$, i.e., the capacity constraint is non-binding, the product prices are no longer

related to one another through Equation (B.35), and the optimal price values are given by:

$$p_1^* = \frac{bc}{b-1} + \frac{b}{b-1} \left[\frac{(c-s)\Lambda_1(z_1) + g\Theta_1(z_1)}{\mu_1 - \Theta_1(z_1)} \right], \quad (\text{B.40})$$

$$p_2^* = \frac{bc}{b-1} + \frac{b}{b-1} \left[\frac{(c-s)\Lambda_2(z_2) + g\Theta_2(z_2)}{\mu_2 - \Theta_2(z_2)} \right], \quad (\text{B.41})$$

$$p_3^* = \frac{bc}{b-1} + \frac{b}{b-1} \left[\frac{(c-s)\Lambda_3(z_3) + g\Theta_3(z_3)}{\mu_3 - \Theta_3(z_3)} \right], \quad (\text{B.42})$$

where p_i is independent of z_j for $j \neq i$, causing the prices to diverge from one another when operational flexibility is available through unsaturated capacity utilization. This means that when capacity is non-binding, the optimal pricing solution reduces to that of Petruzzi and Dada (1999) for a single product.

B.4 Proof of Theorem 3

When capacity is abundant, the price of a product is determined based on the expected profit maximizing pricing equation:

$$p_i^* = \frac{b_i c_i}{b_i - 1} + \frac{b_i}{b_i - 1} \left[\frac{(c_i - s_i) \Lambda_i(z_i) + g_i \Theta_i(z_i)}{\mu_i - \Theta_i(z_i)} \right]. \quad (\text{B.43})$$

As salvage value increases, the probability that the retailer overstocks increases, resulting in higher leftovers and lower shortages (proof obtained directly from the newsvendor optimal solution by evaluating the relation between critical ratio and salvage value). Therefore, from Equation (B.43), $\Lambda(z_i)$ is expected to be much higher than $\Theta(z_i)$ when $s_i > g_i$. As $s_i \rightarrow c_i$, the pricing equation becomes less dependent on the expected leftovers, and is mostly affected by expected shortages, where $\Theta(z_i) \rightarrow 0$ as $s_i \rightarrow c_i$ and $s_i > g_i$. The difference in pricing between the different products therefore decreases as the salvage values increase.

B.5 Proof of Theorem 4

From Equation (B.39), we have the following relationship:

$$\begin{aligned} & \frac{(b-1)[\mu_1 - \Theta_1(z_1)]}{bz_1} \cdot \left\{ \frac{bc_1}{b-1} + \frac{b}{b-1} \left[\frac{(c_1 - s)\Lambda_1(z_1) + g\Theta_1(z_1)}{\mu_1 - \Theta_1(z_1)} \right] - p_1^* \right\} \\ &= \frac{(b-1)[\mu_2 - \Theta_2(z_2)]}{bz_2} \cdot \left\{ \frac{bc_2}{b-1} + \frac{b}{b-1} \left[\frac{(c_2 - s)\Lambda_2(z_2) + g\Theta_2(z_2)}{\mu_2 - \Theta_2(z_2)} \right] - p_2^* \right\}, \end{aligned} \quad (\text{B.44})$$

which shows the relationship between the prices of two products, p_1 and p_2 , given a binding capacity constraint and different ordering costs (c_1 and c_2 instead of c). Assuming we increase c_1 and want to study its effect on p_2 , we formulate p_2 in terms of c_1 and p_1 and write down the derivative with respect to c_1 , which gives the following:

$$\frac{\partial p_2}{\partial c_1} = -\frac{b}{b-1} \left\{ \frac{\Lambda_1(z_1)}{\mu_1 - \Theta_1(z_1)} \right\}. \quad (\text{B.45})$$

From Equation (B.45) we find that an increase in c_1 decreases p_2 . From Equation (B.39), we see that this effect applies for a portfolio of more than two products as well.

B.6 Detailed explanation of possible price increase with capacity

In Section 2.4.4, we notice from Figure 2.9a that for some parameters, price does not necessary always decrease when capacity increases. This effect can be directly explained by referring to the optimal pricing equation given by Equation (2.12). As capacity increases, expected leftovers, $\Lambda(z_i)$, increase. As capacity tends to infinity, $\Lambda(z_i)$ tends to infinity and the expected shortages, $\Theta(z_i)$, tend to zero, causing the price to approach infinity. Therefore, for certain parameters, optimal prices calculated by using Equation (2.12) may increase with capacity, as the relationship between p_i and z_i is nonmonotonic.

To provide a more intuitive interpretation of this phenomenon, in Figure B.1 we show the demand distribution of product 3 in Figure 2.9a for different capacity constraints and prices. Product 3 has a scale parameter of 0.3 (with $b = 1.1$). We now analyze the increase in price between a capacity of 500 units and 700 units, i.e., for $K = 500$ and for $K = 700$. Both those capacities are binding constraints in this case. Each of the graphs in Figure B.1 shows the histogram of 50,000 simulated demands given the optimal price for a capacity of 500 in the top figure, a decreased (non-optimal) price for a capacity of 700 in the middle figure, and an increased (optimal) price for a capacity of 700 in the lower figure, respectively. When restricted in capacity, product 3 shares the capacity with products 1 and 2. Therefore, for a capacity limit of 500 units, the order quantity of product 3 is limited to 166 units, and for a capacity limit of 700 units, the order quantity of product 3 is limited to 232 units (shown by the red dashed vertical lines on the graphs in Figure B.1). The green dashed vertical lines represent the optimal newsvendor order quantities for the prices chosen. We notice that by decreasing the price from the top figure to the middle figure, the mean demand increases. By increasing the price between the top and bottom figures, the mean demand decreases. Decreasing the price also increases the newsvendor order quantity, while increasing the price decreases it (shown by the shift in the green dashed vertical line between the figures). We note that the optimal newsvendor order quantities are high compared to the mean values of the demand distributions. This is expected due to the low cost and high margin of the product, which favors stocking more as the underage cost is much higher than the overage cost.

For some parameters, decreasing the price does not close the gap between the red and green dashed vertical lines, and as prices decrease, this gap continues to exist. Therefore, we notice in most of the pricing results in this paper that although prices decrease with capacity, they decrease at a decreasing rate. For some parameters, in order to close this gap, it is necessary to increase the price of the product.

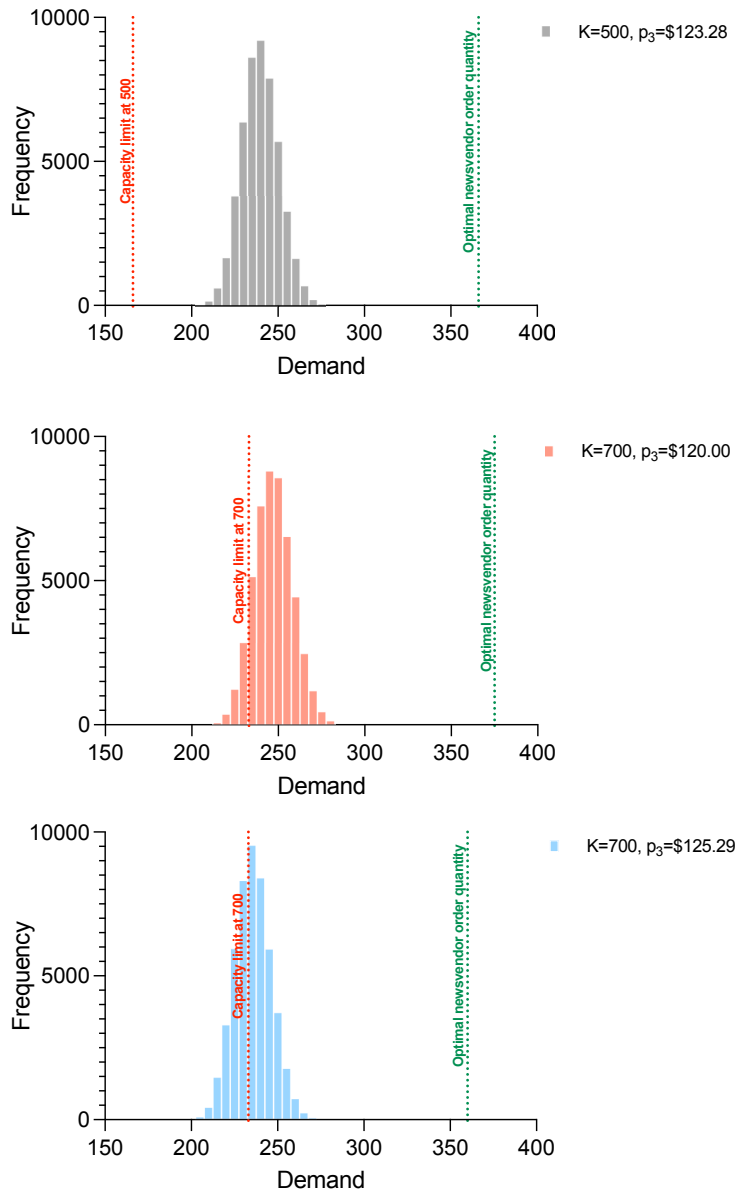


Figure B.1: Explanation of price increase with capacity for some parameters.

Bibliography

- Abbey, J. D., & Guide Jr, V. D. R. (2017). Closed-Loop Supply Chains: A Strategic Overview. In Y. Bouchery, C. Corbett, J. Fransoo, & T. Tan (Eds.), *Sustainable supply chains* (pp. 375–393). Springer, Cham.
- Acimovic, J., Erize, F., Hu, K., Thomas, D. J., & Van Mieghem, J. A. (2018). Product life cycle dataset: Raw and cleaned data of weekly orders for personal computers. *Manufacturing & Service Operations Management*, 21(1), 171–176.
- Agrawal, V. V., Atasu, A., & Van Wassenhove, L. N. (2018). OM Forum – New opportunities for operations management research in sustainability. *Manufacturing & Service Operations Management*, 21(1), 1–12.
- Alev, I., Agrawal, V. V., & Atasu, A. (2019). Extended producer responsibility for durable products. *Manufacturing & Service Operations Management*, 22(2), 364–382.
- Alev, I., Huang, X., Atasu, A., & Toktay, L. B. (2018). A case discussion on market-based extended producer responsibility: The Minnesota Electronics Recycling Act. *Journal of Industrial Ecology*, 23(1), 208–221.
- Arrow, K. J., Harris, T., & Marschak, J. (1951). Optimal inventory policy. *Econometrica: Journal of the Econometric Society*, 19(3), 250–272.
- Atasu, A., Dumas, C., & Van Wassenhove, L. N. (2021). The circular business model [Accessed on: December 20, 2021]. *Harvard Business Review*.
- Aydin, G., & Porteus, E. L. (2008). Joint inventory and pricing decisions for an assortment. *Operations Research*, 56(5), 1247–1255.
- Baardman, L., Levin, I., Perakis, G., & Singhvi, D. (2018). Leveraging comparables for new product sales forecasting. *Production and Operations Management*, 27(12), 2340–2343.
- Basallo-Triana, M. J., Rodríguez-Sarasty, J. A., & Benitez-Restrepo, H. D. (2017). Analogue-based demand forecasting of short life-cycle products: A regression approach and a comprehensive assessment. *International Journal of Production Research*, 55(8), 2336–2350.
- Bass, F. M. (1969). A new product growth for model consumer durables. *Management Science*, 15(5), 215–227.
- Berndt, D., & Clifford, J. (1994). Using dynamic time warping to find patterns in time series. *AAAIWS'94: Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining*, 359–370.

- Bertini, M., & Koenigsberg, O. (2021). The pitfalls of pricing algorithms [Accessed on: January 9, 2023]. *Harvard Business Review*.
- Bicer, I., & Hagspiel, V. (2016). Valuing quantity flexibility under supply chain disintermediation risk. *International Journal of Production Economics*, 180, 1–15.
- Biçer, I. (2015). Dual sourcing under heavy-tailed demand: an extreme value theory approach. *International Journal of Production Research*, 53(16), 4979–4992.
- Biçer, I., Hagspiel, V., & de Treville, S. (2018). Valuing supply-chain responsiveness under demand jumps. *Journal of Operations Management*, 61, 46–67.
- Biçer, I., Lücker, F., & Boyacı, T. (2021). Beyond retail stores: Managing product proliferation along the supply chain. *Production and Operations Management*, 31(3), 1135–1156.
- Biçer, I., & Seifert, R. W. (2017). Optimal dynamic order scheduling under capacity constraints given demand-forecast evolution. *Production and Operations Management*, 26(12), 2266–2286.
- Box, G. E. P., & Tiao, G. C. (1975). Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association*, 70(349), 70–79.
- Bras, B. (2009). Sustainability and product life cycle management - Issues and challenges. *International Journal of Product Lifecycle Management*, 4(1-2-3).
- Bui, T.-D., Tseng, J.-W., Tseng, M.-L., & Lim, M. K. (2021). Opportunities and challenges for solid waste reuse and recycling in emerging economies: A hybrid analysis. *Resources, Conservation & Recycling*, 177, 105968.
- Burruss, J., & Kuettner, D. (2003). Forecasting for short-lived products: Hewlett-Packard's journey. *The Journal of Business Forecasting Methods & Systems*, 21(4), 9–14.
- Cao, Q., Ewing, B. T., & Thompson, M. A. (2012). Forecasting wind speed with recurrent neural networks. *European Journal of Operational Research*, 221(1), 148–154.
- Casadesus-Masanell, R., Kim, H., & Reinhardt, F. L. (2010). Patagonia. *Harvard Business School Case#: 711-020*.
- Cattani, K. D., Dahan, E., & Schmidt, G. M. (2008). Tailored capacity: Speculative and reactive fabrication of fashion goods. *International Journal of Production Economics*, 114(2), 416–430.
- Cecere, L. (2013). New products: More costly and more important [Accessed on: May 16, 2020]. *Forbes*.
- Colback, L. (2022). How technology can help redraw the supply chain map [Accessed on: January 12, 2023]. *Financial Times*.
- Cooper, R. G., & Edgett, S. J. (2012). Best practices in the idea-to-launch process and its governance. *Research-Technology Management*, 55(2), 43–54.
- Costantino, N., & Pellegrino, R. (2010). Choosing between single and multiple sourcing based on supplier default risk: A real options approach. *Journal of Purchasing and Supply Management*, 16(1), 27–40.
- de Treville, S., Bicer, I., Chavez-Demoulin, V., Hagspiel, V., Schürhoff, N., Tasserit, C., & Wager, S. (2014). Valuing lead time. *Journal of Operations Management*, 32(6), 337–346.

Bibliography

- Dean, J. (1976). Pricing policies for new products [Accessed on: February 21, 2019]. *Harvard Business Review*.
- Debo, L. G., Toktay, L. B., & Van Wassenhove, L. N. (2006). Joint life-cycle dynamics of new and remanufactured products. *Production and Operations Management*, 15(4), 498–513.
- Desmet, B. (2016). My biggest mistake: Inventory levels are the thermometer for the management of the company [Accessed on: January 6, 2023]. *Arkieva*.
- DeVries, T., & Taylor, G. W. (2017). Dataset augmentation in feature space. *5th International Conference on Learning Representations, ICLR 2017 - workshop track*.
- Dholakia, U. M. (2015). The risks of changing your prices too often [Accessed on: January 9, 2023]. *Harvard Business Review*.
- Donahue, B. (2018). China is turning Ethiopia into a giant fast-fashion factory [Accessed on: December 20, 2021]. *Bloomberg BusinessWeek*.
- Dyer, J., Furr, N., & Lefrandt, C. (2014). The Industries Plagued by the Most Uncertainty [Accessed on: January 12, 2023]. *Harvard Business Review*.
- Elmaghraby, W., & Keskinocak, P. (2003). Dynamic pricing in the presence of inventory considerations: Research overview, current practices, and future directions. *Management Science*, 49(10), 1287–1309.
- Ernst, R. L. (1970). *A linear inventory model of a monopolistic firm* (Ph.D. Dissertation). Department of Economics, University of California, Berkeley, CA.
- Fallah Tehrani, A., & Ahrens, D. (2016). Enhanced predictive models for purchasing in the fashion field by using kernel machine regression equipped with ordinal logistic regression. *Journal of Retailing and Consumer Services*, 32, 131–138.
- Feng, Q. (2010). Integrating dynamic pricing and replenishment decisions under supply capacity uncertainty. *Management Science*, 56(12), 2154–2172.
- Feng, Q., Luo, S., & Zhang, D. (2013). Dynamic inventory–pricing control under backorder: Demand estimation and policy optimization. *Manufacturing & Service Operations Management*, 16(1), 149–160.
- Fildes, R., & Goodwin, P. (2007). Against your better judgment? How organizations can improve their use of management judgment in forecasting. *Interfaces*, 37(6), 570–576.
- Gallego, G., Katircioglu, K., & Ramachandran, B. (2007). Inventory management under highly uncertain demand. *Operations Research Letters*, 35(3), 281–289.
- Golder, P. N., & Tellis, G. J. (2004). Growing, growing, gone: Cascades, diffusion, and turning points in the product life cycle. *Marketing Science*, 23(2), 207–218.
- Guide Jr, V. D. R., & Van Wassenhove, L. N. (2009). The evolution of closed-loop supply chain research. *Operations Research*, 57(1), 10–18.
- Gustafson, K. (2015). Retailers are losing \$1.75 trillion over this [Accessed on: December 20, 2021]. CNBC.
- Hamilton, L., & Webster, P. (2012). *The international business environment* (2nd ed.). Oxford University Press.
- Hausman, W. H., & Thorbeck, J. S. (2010). Fast fashion: Quantifying the Benefits. In T. Cheng & T. M. Choi (Eds.), *Innovative quick response programs in logistics and supply chain management* (pp. 315–329). Springer, Berlin, Heidelberg.

- He, R., & Lu, Y. (2020). A Robust Price-Setting Newsvendor Problem. *Production and Operations Management*, 30(1), 276–292.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
- Hu, K., Acimovic, J., Erize, F., Thomas, D. J., & Van Mieghem, J. A. (2018). Forecasting new product life cycle curves: Practical approach and empirical analysis. *Manufacturing & Service Operations Management*, 21(1), 66–85.
- Hu, L., Wang, C., Ye, Z., & Wang, S. (2021). Estimating gaseous pollutants from bus emissions: A hybrid model based on GRU and XGBoost. *Science of the Total Environment*, 783, Article 146870.
- Iwana, B. K., & Uchida, S. (2021). An empirical survey of data augmentation for time series classification with neural networks. *PLoS ONE*, 16(7), e0254841.
- Javed, A., Lee, B. S., & Rizzo, D. M. (2020). A benchmark study on time series clustering. *Machine Learning with Applications*, 1, Article 100001.
- Kahn, K. B. (2002). An exploratory investigation of new product forecasting practices. *Journal of Product Innovation Management*, 19(2), 133–143.
- Kahn, K. B. (2014). Solving the problems of new product forecasting. *Business Horizons*, 57(5), 607–615.
- Karaian, J., & Majerol, V. (2022). Why Are C.E.O.s suddenly obsessed with ‘Elasticity’? [Accessed on: January 12, 2023]. *The New York Times*.
- Karlin, S., & Carr, C. R. (1962). Studies in applied probability and management science. In K. J. Arrow, S. Karlin, & H. Scarf (Eds.), *Prices and optimal inventory policy* (pp. 159–172). Stanford University Press, Stanford, CA.
- Keogh, E., & Ratanamahatana, C. A. (2005). Exact indexing of dynamic time warping. *Knowledge and Information Systems*, 7, 358–386.
- Kraus, M., Feuerriegel, S., & Oztekin, A. (2020). Deep learning in business analytics and operations research: Models, applications and managerial implications. *European Journal of Operational Research*, 281(3), 628–641.
- Kumar, S., Hussain, L., Banarjee, S., & Reza, M. (2018). Energy load forecasting using deep learning approach-LSTM and GRU in spark cluster. *2018 Fifth International Conference on Emerging Applications of Information Technology (EAIT)*, 1–4.
- Kurawarwala, A. A., & Matsuo, H. (1996). Forecasting and inventory management of short life-cycle products. *Operations Research*, 44(1), 131–150.
- Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of Political Economy*, 74(2), 132–157.
- Lee, C.-Y., & Lee, M.-K. (2017). Demand forecasting in the early stage of the technology’s life cycle using a Bayesian update. *Sustainability*, 9(8), 1378.
- Liu, C., Hou, W., & Liu, D. (2017). Foreign exchange rates forecasting with convolutional neural network. *Neural Processing Letters*, 46, 1095–1119.
- Lu, C.-J., & Kao, L.-J. (2016). A clustering-based sales forecasting scheme by using extreme learning machine and ensembling linkage methods with applications to computer server. *Engineering Applications of Artificial Intelligence*, 55, 231–238.

Bibliography

- Mahajan, V., Muller, E., & Bass, F. M. (1990). New product diffusion models in marketing: A review and directions for research. *Journal of Marketing*, 54(1), 1–26.
- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, 36(1), 54–74.
- Martí, J. M. C., & Seifert, R. W. (2013). Assessing the comprehensiveness of supply chain environmental strategies. *Business Strategy and the Environment*, 22, 339–356.
- Martí, J. M. C., Tancrez, J.-S., & Seifert, R. W. (2015). Carbon footprint and responsiveness trade-offs in supply chain network design. *International Journal of Production Economics*, 166, 129–142.
- Meade, N., & Islam, T. (2006). Modelling and forecasting the diffusion of innovation – A 25-year review. *International Journal of Forecasting*, 22(3), 519–545.
- Mills, E. S. (1959). Uncertainty and price theory. *The Quarterly Journal of Economics*, 73(1), 116–130.
- Mišić, V. V., & Perakis, G. (2019). Data analytics in operations management: A review. *Manufacturing & Service Operations Management*, 22(1), 158–169.
- Müller, M. (2007). Dynamic Time Warping. In *Information Retrieval for Music and Motion*. Springer, Berlin, Heidelberg.
- Murray, C. C., Gosavi, A., & Talukdar, D. (2012). The multi-product price-setting newsvendor with resource capacity constraints. *International Journal of Production Economics*, 138(1), 148–158.
- Nassauer, S., & Terlep, S. (2022). Inventory pileup, uneasy shoppers put retailers in jeopardy [Accessed on: January 7, 2023]. *The Wall Street Journal*.
- Neamtu, R., Ahsan, R., Rundensteiner, E. A., Sarkozy, G., Keogh, E., Anh Dau, H., Nguyen, C., & Lovering, C. (2018). Generalized Dynamic Time Warping: Unleashing the warping power hidden in point-wise distances. *2018 IEEE 34th International Conference on Data Engineering (ICDE)*, 521–532.
- Niinimäki, K., Peters, G., Dahlbo, H., Perry, P., Rissanen, T., & Gwilt, A. (2020). The environmental price of fast fashion. *Nature Reviews Earth & Environment*, 1, 189–200.
- Niranjan, T. T., Mathur, S., Kumar, G. N., & Gavirneni, N. (2022). Inflating orders can worsen supply chain uncertainty [Accessed on: January 12, 2023]. *Harvard Business Review*.
- Oh, S., & Özer, Ö. (2013). Mechanism design for capacity planning under dynamic evolutions of asymmetric demand forecasts. *Management Science*, 59(4), 987–1007.
- Ojea, S. (2022). Hasbro's Sales Drop 15% as High Prices Jolt Consumers [Accessed on: January 7, 2023]. *The Wall Street Journal*.
- Özkoç, E. E. (2020). Clustering of time-series data. (Ed.) *Data mining - methods, applications and systems*, IntechOpen.
- Paliari, I., Karanikola, A., & Kotsiantis, S. (2021). A comparison of the optimized LSTM, XG-BOOST and ARIMA in time series forecasting. In *12th International Conference on Information, Intelligence, Systems & Applications (IISA)*, 1–7.
- Park, D. S., Chan, W., Zhang, Y., Chiu, C.-C., Zoph, B., Cubuk, E. D., & Le, Q. V. (2019). SpecAugment: A simple data augmentation method for automatic speech recognition. *Inter-speech 2019*.

- Peres, R., Muller, E., & Mahajan, V. (2010). Innovation diffusion and new product growth models: A critical review and research directions. *International Journal of Research in Marketing*, 27(2), 91–106.
- Petropoulos, F., Apiletti, D., Assimakopoulos, V., & et al. (2022). Forecasting: Theory and practice. *International Journal of Forecasting*, 38(3), 705–871.
- Petruzzi, N. C., & Dada, M. (1999). Pricing and the newsvendor problem: A review with extensions. *Operations Research*, 47(2), 183–194.
- Qi-zhi, S. (2007). Forecasting for products with short life cycle based on improved BASS model. *Industrial Engineering and Management*.
- Rink, D. R., & Swan, J. E. (1979). Product life cycle research: A literature review. *Journal of Business Research*, 7(3), 219–242.
- Rogers, E. M. (1962). *Diffusion of innovations*. Free Press of Glencoe, New York.
- Rudin, C., & Carlson, D. (2019). The secrets of machine learning: Ten things you wish you had known earlier to be more effective at data analysis. *INFORMS Tutorials in Operations Research*.
- Salinger, M., & Ampudia, M. (2011). Simple economics of the price-setting newsvendor problem. *Management Science*, 57(11), 1996–1998.
- Sanders, N. R., & Manrodt, K. B. (2003). The efficacy of using judgmental versus quantitative forecasting methods in practice. *Omega*, 31(6), 511–522.
- Scott, C. L. (2022). Shoppers are caught off guard as prices on everyday items change more often [Accessed on: January 11, 2023]. *The Wall Street Journal*.
- Seifert, R. W., & Elalem, Y. K. (2020). Can AI deliver tangible benefits to demand forecasting? 3 questions to help managers decide [Accessed on: January 21, 2023]. *IMD*.
- Seifert, R. W., & Elalem, Y. K. (2022). Sustainable supply chains: time to get serious [Accessed on: January 24, 2023]. *IMD*.
- Seifert, R. W., Elalem, Y. K., & Biçer, I. (2021). Take the circular route to curbing over-production [Accessed on: January 21, 2023]. *IMD*.
- Seifert, R. W., & Markoff, R. (2021). How a semiconductor shortage was a surprise late entry to the pandemic chaos [Accessed on: January 12, 2023]. *I by IMD*.
- Shapiro, A., Dentcheva, D., & Ruszczyński, A. (2009). *Lectures on stochastic programming: Modeling and theory*. SIAM: Society for Industrial; Applied Mathematics, Philadelphia, PA.
- Shi, J., Zhang, G., & Sha, J. (2011). Jointly pricing and ordering for a multi-product multi-constraint newsvendor problem with supplier quantity discounts. *Applied Mathematical Modelling*, 35(6), 3001–3011.
- Singh, D., & Singh, B. (2020). Investigating the impact of data normalization on classification performance. *Applied Soft Computing*, 97, Part B, Article 105524.
- Stellwagen, E., & Tashman, L. (2013). ARIMA: The models of Box and Jenkins. *Foresight: The International Journal of Applied Forecasting, International Institute of Forecasters*, (30), 28–33.
- Szozda, N. (2010). Analogous forecasting of products with a short life cycle. *Decision Making in Manufacturing and Services*, 4(2), 71–85.

Bibliography

- Thomassey, S., & Happiette, M. (2007). A neural clustering and classification system for sales forecasting of new apparel items. *Applied Soft Computing*, 7(4), 1177–1187.
- Thowsen, G. T. (1975). A dynamic, nonstationary inventory problem for a price/quantity setting firm. *Naval Research Logistics Quarterly*, 22(3), 461–476.
- Toktay, L. B., Wein, L. M., & Zenios, S. A. (2000). Inventory management of remanufacturable products. *Management Science*, 46(11), 1412–1426.
- Tseng, M.-L., Tran, T. P. T., Ha, H. M., Bui, T.-D., & Lim, M. K. (2021). Sustainable industrial and operation engineering trends and challenges Toward Industry 4.0: a data driven analysis. *Journal of Industrial and Production Engineering*, 38(8), 581–598.
- Tseng, M.-L., Tran, T. P. T., Wu, K.-J., Tan, R. R., & Bui, T. D. (2020). Exploring sustainable seafood supply chain management based on linguistic preferences: collaboration in the supply chain and lean management drive economic benefits. *International Journal of Logistics Research and Applications*, 1–23.
- van Steenbergen, R. M., & Mes, M. R. K. (2020). Forecasting demand profiles of new products. *Decision Support Systems*, 139, Article 113401.
- Vernon, R. (1966). International investment and international trade in the product cycle. *The Quarterly Journal of Economics*, 80(2), 190–207.
- Whalen, J., Bhattarai, A., & Greene, J. (2020). Purell prices are spiking on Amazon, as sanitizer speculation becomes a cottage industry [Accessed on: January 11, 2023]. *The Washington Post*.
- Whitin, T. (1955). Inventory control and price theory. *Management Science*, 2(1), 61–68.
- Winston, A. (2011). Excess Inventory Wastes Carbon and Energy, Not Just Money [Accessed on: January 3, 2023]. *Harvard Business Review*.
- Wong, J. (2023). How Samsung Could Climb Out of the Chip Price Pit [Accessed on: January 7, 2023]. *The Wall Street Journal*.
- Wood, Z. (2021). Amazon faces MPs' scrutiny after destroying laptops, tablets and books. *The Guardian*.
- Wood, Z. (2022). 'People are ripping up contracts': price of rice boils over into new territory in UK [Accessed on: January 10, 2023]. *The Guardian*.
- Xu, M., Chen, Y., & Xu, X. (2010). The effect of demand uncertainty in a price-setting news vendor model. *European Journal of Operational Research*, 207(2), 946–957.
- Yildiz, B., Bilbao, J. I., & Sproul, A. B. (2017). A review and analysis of regression and machine learning models on commercial building electricity load forecasting. *Renewable and Sustainable Energy Reviews*, 73, 1104–1122.
- Younger, L. (1978). Price, inventory and the structure of uncertain demand. *New Zealand Operations Research*, 6(2), 157–177.
- Zhang, X., Shen, F., Zhao, J., & Yang, G. (2017). Time series forecasting using GRU neural network with multi-lag after decomposition. In D. Liu, S. Xie, Y. Li, D. Zhao, & E. (El-Alfy (Eds.), *Neural Information Processing. ICONIP 2017. Lecture Notes in Computer Science*. Springer, Cham.
- Zhu, K., & Thonemann, U. W. (2004). An adaptive forecasting algorithm and inventory policy for products with short life cycles. *Naval Research Logistics*, 51(5), 633–653.



Yara Kayyali El Alem

With a current concentration in supply chain management, I am looking to contribute to optimising supply chain operations while making them more sustainable. I am a quick-learner, a multitasker, and a people's person.

Phone:
+41 79 642 34 31

Email:
yara.kayyalialelem@gmail.com

Address:
Avenue de la Gare 2,
Chavannes-près-Renens 1022.
Switzerland

Work Experience

Research Assistant

Sept 2018 - March 2023

EPFL | Switzerland

- Forecasted new product sales using machine learning and ARIMA
 - Paper accepted: A machine learning-based framework for forecasting sales of new products with short life cycles using deep neural networks
 - Managerial article: Can AI deliver tangible benefits to demand forecasting?, IMD
- Worked on sustainable supply chain sourcing by mathematically showing the benefits of different operational flexibility strategies on both profits and waste reduction
 - Paper published: Why do companies need operational flexibility to reduce waste at source?
 - Managerial article: Take the circular route to curbing over-production, I by IMD
 - Managerial article: Sustainable supply chains: time to get serious, I by IMD
- Optimized inventory and pricing decisions simultaneously for multiple products with capacity constraint using Newsvendor model (paper in progress)

Junior Applied Scientist

June 2021 - Dec 2021

Zalando | Switzerland

Worked on the machine learning algorithm that produces daily demand forecasts for all Zalando products for warehouse optimization purposes. Conducted sensitivity analysis on the machine learning algorithm

Intern Information Protection Governance

Feb 2018 - Aug 2018

Philip Morris | Switzerland

Followed a design thinking approach and analyzed results of employee interviews to create a more user-friendly platform on Tableau to access necessary company procedure documentation

Research Assistant

June 2015 - Aug 2015

University of Belgrade | Serbia

Researched steel structures for stadiums to provide proper rigidity while minimizing environmental impacts to develop more sustainable outcome

Additional Experiences

Master Thesis Supervision

Sept 2018 - Present

EPFL | Switzerland

Supervised master thesis of EPFL students. All projects are 6 months on supply chain and sustainability topics and include the following companies: Rolex, Panerai, and Tetra Pak

Teaching Assistant, Supply Chain Management

Sept 2018 - July 2022

EPFL | Switzerland

Changemakers Startup Challenge

June 2021 - Dec 2021

EPFL | Switzerland

PhD Student Representative & Event Organizer

Sept 2019 - Sept 2021

EPFL | Switzerland

Course on Lean Management

Sept 2015 - Feb 2016

American University of Beirut | Lebanon

Education

PhD in Technology & Operations Management:

Sept 2018 - Present

Concentration in Supply Chain Management

EPFL | Switzerland

Master in Management, Technology & Entrepreneurship

Sept 2016 - Aug 2018

EPFL | Switzerland

Bachelors in Civil & Environmental Engineering

Sept 2012 - June 2016

American University of Beirut | Lebanon

Languages

English (Fluent)

Arabic (Fluent)

French (B1)

Computer Skills

Python

Machine Learning in Python

Microsoft Office

AutoCAD

Tableau

Canva

Soft Skills

Team work

Leadership

Problem solving

Project management

Presentation skills

Achievements

Best teaching assistant award for three years, EPFL, 2020, 2021, 2022

First place Times Competition, Zurich, 2016

Dean's Honor List, Lebanon, 2014

First on Bachelors, 2014

First on Mathematics A Levels, Saudi Arabia, 2011

Class Valedictorian, Saudi Arabia, 2012

Hobbies

Jewelry making

Puzzle making

Painting

Swimming

Spinning