Université de Lille
Doctoral School MADIS
Laboratoire Paul Painlevé UMR 8524

ESTIMATING SUBSTITUTION FOR OPTIMISED REPLENISHMENT WITH SLOW MOVERS PRODUCTS

AXEL POTIER

Doctorate CIFRE in Applied Mathematics
Faculté des Sciences et Technologies de Lille
Date of defense 12 - November - 2024

Christophe Biernacki	Inria Lille	Director	Professor
Vincent Vandewalle	Université Côte d'Azur	Co-supervisor	Professor
Matthieu Marbac Lourdelle	INSA Rennes	Co-supervisor	Associate Professor
Julien Favre	Adeo Service	Supervisor Adeo	Engineer
Joseph Ngatchou-Wandji	Université de Rennes	Referees	Professor
Marie Chavent	Université de Bordeaux	Referees	Professor
Cristian Preda	Inria Lille	Chief examiner	Professor

Université de Lille
Ecole Doctorale MADIS
Laboratoire Paul Painlevé UMR 8524

ESTIMATION DE LA SUBSTITUTION POUR L'OPTIMISATION DU RÉAPPROVISIONNEMENT DES PRODUITS À FAIBLE ROTATION

AXEL POTIER

Thèse de Doctorat CIFRE en Mathématiques Appliquées
Faculté des Sciences et Technologies de Lille
Date de soutenance 12 - Novembre - 2024

Christophe Biernacki	Inria Lille	Directeur de Thèse	Professeur
Vincent Vandewalle	Université Côte d'Azur	Co-encadrant	Professeur
Matthieu Marbac Lourdelle	INSA Rennes	Co-encadrant	Professeur associé
Julien Favre	Adeo Service	Encadrant Adeo	Ingénieur
Joseph Ngatchou-Wandji	Université de Rennes	Rapporteur	Professeur
Marie Chavent	Université de Bordeaux	Rapporteur	Professeur
Cristian Preda	Inria Lille	Président du jury	Professeur

ABSTRACT

In Retail, inventory optimisation is a common problem targeting a trade-off between the risk of stock-out and the risk of overstocking, in order to reach an optimal global profit. This inventory optimisation task is however very challenging in case of products that are sold in low quantity (so- called slow movers in Retail). Indeed, estimating properly the related future sales, especially after discretisation, usually suffers from high relative standard deviation, from which the optimal replenishment solution inherits to the point of being usefulness in practice. Nevertheless, slow movers in many companies, as ADEO (French holding company selling consumer goods for DIY and decoration), are sufficiently numerous to represent a significant portion of the sales and of the stock. Consequently, concerned items are difficult to be optimally replenished, and even a small improvement of the replenishment process may have a significant positive effect on the whole global profit.

For bridging the gap, this thesis reformulates the slow movers optimal replenishment problem as a substitution probability estimation problem between items. When a product is out of stock, a client may alternatively choose another item (a so-called substitute product), avoiding to definitively loose the intended initial sale. As a consequence, instead of choosing the optimal quantity to replenish separately for each item (classical approach), we leverage that additional information that a group of items products can be substituted for each other allows to more efficiently estimate the optimal replenished quantity of the whole group of items. Obviously, such a substitution process occurs only with a certain probability, (1) that we have to estimate and (2) then we have to properly use through the optimal replenishment calculation. Notice fundamentally that the discrete nature of the stock quantity is expected to provide a better relative improvement in terms of profit in the case of slow movers compared to the case of fast movers (the contrary of slow movers), which justifies the special interest of our approach for slow movers.

For estimating the substitution probabilities within a group of substitutable products, we reformulate a specific existing model. This model is however only based on poor observed data since limited to sales and stocks transactions in the store. In particular, the initial demand of the client, and also the lost sales, are not observable. We circumvent this missing data issue by adapting an EM algorithm for the estimation of the probabilities of substitution. We pay also attention to the identifiability of such a model for applying properly the maximum likelihood paradigm, establishing some theoretical hard constraints on the size of the group of substitutable products. Some experiments on synthetic and real data sets (from ADEO) allow to measure the variability of the substitution probability estimation (which is quite large in this large missing data case) but illustrate however that the quality of estimates when merging sales and stock data from several stores allows to reach very valuable and useful inference on products substitution and their replenishment. The last step of our work consists to discover the groups of substitutable products, ideally from a large set of a raw products list providing from the store. For this purpose, we propose a specific clustering of products relying on the light hypothesis that most of products have zero probability of substitution. We then adapt a hierarchical clustering algorithm, allowing to estimate the targeted groups in a very fast manner and we apply it to a consequent real data set from the ADEO company.

Keywords. Replenishment Optimisation, Substitutable Products, Missing Data, Identifiability, EM Algorithm, Clustering

RÉSUMÉ

Dans le commerce de détail, l'optimisation des stocks est un problème courant qui consiste à trouver un compromis entre le risque de rupture de stock et le risque de surstockage, afin d'atteindre un profit global optimal. Cependant, cette tâche d'optimisation des stocks est très difficile dans le cas des produits vendus en faible quantité (les fameux ßlow movers"dans le commerce de détail). En effet, estimer correctement les ventes futures de ces produits, notamment après discrétisation, souffre généralement d'une déviation standard relative élevée, dont la solution de réapprovisionnement optimale hérite au point de devenir inutilisable en pratique. Néanmoins, les slow movers, dans de nombreuses entreprises comme ADEO (une holding française vendant des biens de consommation pour le bricolage et la décoration), sont suffisamment nombreux pour représenter une part significative des ventes et des stocks. En conséquence, les articles concernés sont difficiles à réapprovisionner de manière optimale, et même une petite amélioration du processus de réapprovisionnement peut avoir un effet positif significatif sur le profit global.

Pour combler cette lacune, cette thèse reformule le problème de réapprovisionnement optimal des slow movers en un problème d'estimation de la probabilité de substitution entre les articles. Lorsqu'un produit est en rupture de stock, un client peut choisir un autre article en alternative (un produit dit substitut), évitant ainsi de perdre définitivement la vente initialement prévue. En conséquence, au lieu de choisir la quantité optimale à réapprovisionner séparément pour chaque article (approche classique), nous tirons parti de l'information supplémentaire qu'un groupe d'articles peut être substitué les uns aux autres, ce qui permet d'estimer plus efficacement la quantité optimale à réapprovisionner pour l'ensemble du groupe d'articles. Évidemment, un tel processus de substitution ne se produit qu'avec une certaine probabilité, (1) que nous devons estimer et (2) que nous devons ensuite utiliser correctement lors du calcul du réapprovisionnement optimal. Il est important de noter que la nature discrète de la quantité de stock devrait fournir une meilleure amélioration relative en termes de profit dans le cas des slow movers par rapport aux fast movers (le contraire des slow movers), ce qui justifie l'intérêt particulier de notre approche pour les slow movers.

Pour estimer les probabilités de substitution au sein d'un groupe de produits substituables, nous reformulons un modèle existant spécifique. Cependant, ce modèle ne repose que sur des données observées limitées puisqu'il se limite aux transactions de ventes et de stocks en magasin. En particulier, la demande initiale du client, ainsi que les ventes perdues, ne sont pas observables. Nous contournons ce problème de données manquantes en adaptant un algorithme EM pour l'estimation des probabilités de substitution. Nous prêtons également attention à l'identifiabilité d'un tel modèle pour appliquer correctement le paradigme de maximum de vraisemblance, en établissant certaines contraintes théoriques strictes sur la taille du groupe de produits substituables. Des expériences sur des ensembles de données synthétiques et réelles (d'ADEO) permettent de mesurer la variabilité de l'estimation des probabilités de substitution (qui est assez grande dans ce cas de données manquantes importantes) mais illustrent néanmoins que la qualité des estimations lorsqu'on fusionne les données de ventes et de stocks de plusieurs magasins permet d'atteindre des inférences très précieuses et utiles sur la substitution des produits et leur réapprovisionnement.

La dernière étape de notre travail consiste à découvrir les groupes de produits substituables, idéalement à partir d'une vaste liste brute de produits provenant du magasin. Pour ce faire, nous proposons un regroupement spécifique de produits en nous appuyant sur l'hypothèse légère que la plupart des produits ont une probabilité de substitution nulle. Nous adaptons ensuite un algorithme de clustering hiérarchique, permettant d'estimer les groupes ciblés de manière très rapide et nous l'appliquons à un ensemble de données conséquent provenant de l'entreprise ADEO.

Mots clés. Optimisation du Réapprovisionnement, Produits substituables, Données Manquantes, Identifiabilité, Algorithme EM, Partitionnement.

PUBLICATIONS

Potier, Axel, Christophe Biernacki, Vincent Vandewalle, Marbac-Lourdelle Matthieu, and Julien Favre (2022). "Modélisation de la Substitution entre Articles pour Optimiser leur Réapprovisionnement," pp. 101–106. URL: https://jds22.sciencesconf.org/data/pages/LivretJdS22_version_longue.pdf.

ACKNOWLEDGMENTS

I wish to express my gratitude to all those who supported me throughout my thesis journey.

First and foremost, I extend my thanks to the examiners for reviewing this manuscript and recognizing the novelty of the results, which ultimately led to the awarding of this doctorate.

I am especially gratefull to my research supervisors Christophe Biernacki, Matthieu Marbac-Lourdelle and Vincent Vandewalle for their kindness, support and continuous availability from the internship phase to the completion of this work, as well as for their invaluable advices regarding the core models and designing numerical experiments.

I would also like to thank Maxime Hemery for the opportunity that fostered the partnership between Adeo and Inria. I am gratefull to Julien Favre for overseeing my Phd at Adeo offering perspective and advices regarding coding and retail practices. I am thankfull to Florian Deze for coding advices and to Luc de Ricke and Guillaume Bonningue for sharing their expertise in the retail field from Adeo's perspective.

My sincere thanks to the MODAL team at Inria for enriching discussions during coffee breaks and for their camaraderie.

My gratitude also goes to Adeo for the opportunity of working on such a rich subject.

Finally, i thank my family and friends for their emotional support as well as at times for their technical advice.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LISTINGS

ACRONYMS

NVM Newsvendor Model

OOS Out Of Stock

SKU Stock Keeping Unit

PIT Probability Integral Transform

CDF Cumulative Density Function

NOTATIONS

NOTATIONS COMMON TO ALL CHAPTERS

- $\hat{}$ estimated quantities.
- $\top$ transpose.
- $\mathbb{E}$ expectation operator.
- $\mathbb{V}$ variance operator.
- $\mathbb{P}$ probability measure depending on a random variable.
- $\mathcal{P}$ Poisson law.
- $\mathcal{M}$ Multinomial law.
- $\mathcal{G}$ Geometrical law.
- $\mathcal{N}$ Gaussian law.
- K number of products.
- t continuous variable denoting the time.
- k, ℓ index of products.
- $\pi_{\ell k}$ probability of substitution from product ℓ to k.
- $\boldsymbol{\pi}_\ell = (\pi_{\ell 1}, \ldots, \pi_{\ell K})^\top$ probabilities of substitution of product ℓ.
- $\boldsymbol{\pi} = (\boldsymbol{\pi}_1^\top, \ldots, \boldsymbol{\pi}_K^\top)^\top$ vector of probabilities of substitution.
- $S_K = \left\{ \mathbf{a} \in \mathbb{R}^K,\ 0 \leqslant a_k \leqslant 1,\ \sum_{k=1}^K a_k = 1 \right\}$ is the simplex of dimension K.

NOTATIONS RELATED TO CHAPTER 3

- H number of periods.
- h index of the period of interest.
- n time horizon.
- $\mathbf{D}_h = (D_{h1}, \ldots, D_{hK})$ demand of products over the period h.
- $\tilde{\mathbf{D}}_t = (\tilde{D}_{t1}, \ldots, \tilde{D}_{tK})$ vector of demand at time t.
- $\mathbf{V}_t = (V_{t1}, \ldots, V_{tK})$ vector of sale at time t.
- $\mathbf{P} = (P_1, \ldots, P_K)$ vector of Replenishment Policies.
- θ parameters of the distribution of $\mathbf{D}_h$.
- $\mathbf{Q} = (Q_1, \ldots, Q_K)$ ordering quantity.
- $\hat{D}_{h+\tau|h}$ demand's forecast made at period h for the period $h + \tau$.
- ρ_k stochastic profit associated to product k.
- c^+ underage cost.
- c^- overage cost.

NOTATIONS RELATED TO CHAPTER 4 AND CHAPTER 5

- j variable for the configuration of availability.
- J (integer) number of configurations of availability.
- n time horizon.
- $\boldsymbol{\omega}_j = (\omega_{j1}, \ldots, \omega_{jK})^\top$ vector of availability of products.
- $\mathbf{X}_n = (X_{n1}, \ldots, X_{nJ})^\top$ random variable of the time spent in the configurations.
- x_n observation associated to $\mathbf{X}_n$
- $\mathbf{Z}_n$ random variable of the initial demand of products (hidden variable).
- z_n observation associated to $\mathbf{Z}_n$.
- $\mathbf{Y}_n$ random variable associated to the sales.
- y_n observation associated to $\mathbf{Y}_n$.
- μ_k Poisson intensity parameters of arriving customers for product k.
- θ whole parameters related to the model of substitution $(\mu_1, \ldots, \mu_K, \boldsymbol{\pi}_1^\top, \ldots, \boldsymbol{\pi}_K^\top)^\top$.

Part I

FROM MULTI-PRODUCT REPLENISHMENT TO MULTI-PRODUCT SUBSTITUABILITY

1 INTRODUCTION

1.1 MOTIVATION FOR MULTI-PRODUCT REPLENISHMENT AT ADEO

1.1.1 *General information about Adeo*

Adeo is a French holding in the retail sector including Leroy Merlin, Bricoman, Weldom, Zodio, Kbane, Alice Delice, Adeo Service. These brands are specialised in the do-it-yourself (DIY) and decoration sector for inhabitants/professionals. Adeo serves 500 millions consumers and has about 150,000 employees. With a turnover of 24.2 billion euros in 2024 and 1,000 points of sale distributed in 20 countries, Adeo is in first place in the European market and at the third place on the global. Adeo SERVICE[1] supports the global strategy of Adeo by providing with its digital department optimisations along the supply chain.

1.1.2 *Importance and difficulties associated to infrequently purchased products*

Adeo's strategy is based on products' variety. Indeed, the large assortment of products is an incentive for clients to select a product suiting a home project and buying complementary products. Providing such a large assortment implies that some of the products are sold in small quantities (slow movers). An exploratory analysis (Figure 1) showed that Leroy Merlin France has 50% of the couples store products that are sold less than once each ten days, and it amounts to 33% of turnover. The total stock of these products is around 150 millions euros. The importance of slow movers is common to the retail sector, as described in surveys such as Valery Lukinskiy, Vladislav Lukinskiy, and Sokolov, 2020 which attest the rich body of literature related to slow movers.

The demand for slow movers, denoted by y_i for a product i, is uncertain as a consequence of the many zeros in the time series, implying difficulties in terms of forecasting and replenishment. As an example, Figure 2 illustrates the decreasing quality of the forecasting as a function of the mean frequency of sales. The measure is the Symmetrical Weighted Absolute Percentage Error given by SWAPE $= \frac{\sum_i |y_i - \hat{y}_i|}{\sum_i |y_i + \hat{y}_i|}$ which is scaled independent, where $\hat{y}_i$ is an estimate of y_i. The turnover generated by the slow movers can be seen as 33% whereas fast movers with the same amount generate 75%. Improving the inventory policy even by on the margins would yield substantial gains.

1 https://www.ADEO.com/

Figure 1: Distribution of the number of products and turnover per daily frequency of sales

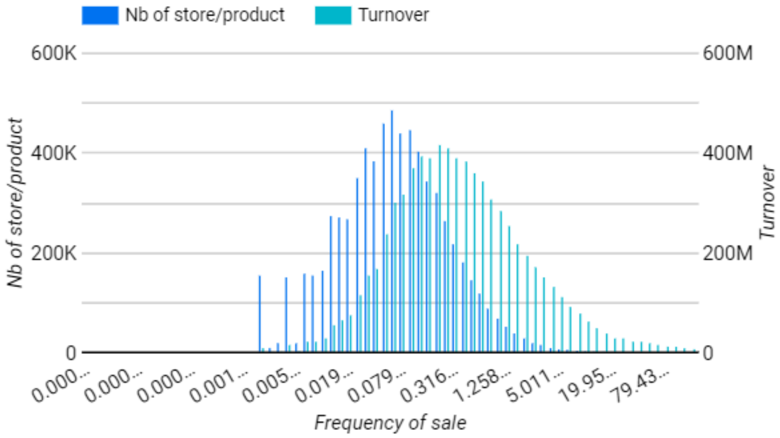
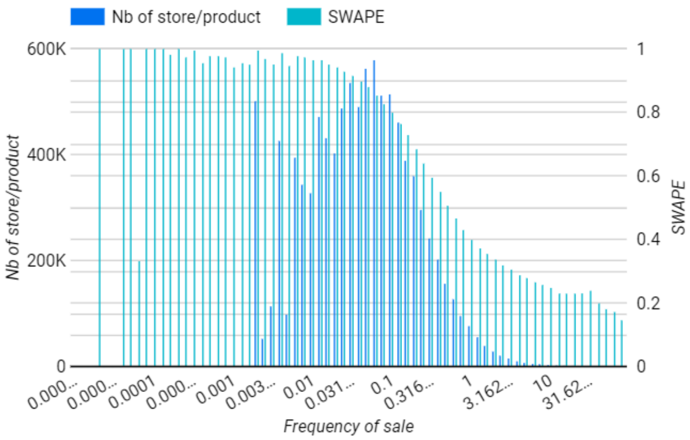
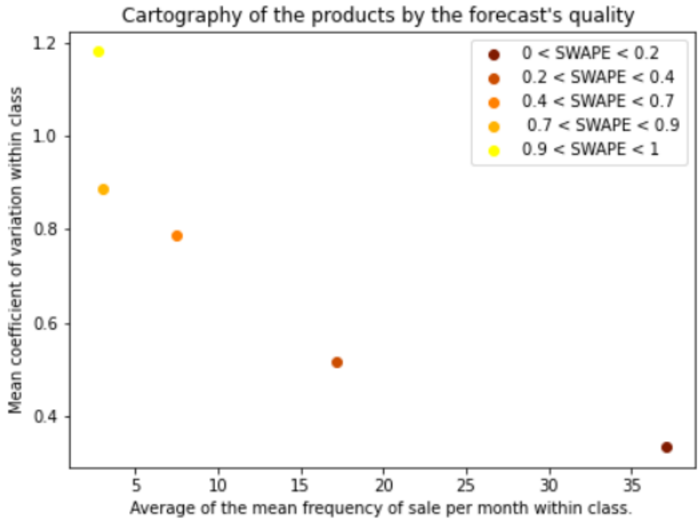


Figure 2: Mean SWAPE per daily frequency of sale



Another study we conducted on 300 electrical accessories showed that a lower frequency of sales per month is associated with a higher coefficient of variation and with lower quality of forecast (Figure 3). The metric used was SWAPE and the prediction was made by the RAP software in the year 2018 for Bricoman France. An increase in the coefficient of variation has impacts on the quality of the inventory (Zotteri, 2000). At a given fixed mean frequency of sale, an increase in uncertainty results in a lower quality of replenishment. Due to the quantity of slow movers, this problem scales.

Figure 3: Cartography of the products by the forecast's quality



The financial importance, the difficulty associated with replenishment, and the commonality of the problem make the optimisation of slow movers a subject for research that has been extensively studied and that is still problematic.

1.2 OBJECTIVE AND CONTRIBUTIONS

The seminal intention of this thesis was then to promote a method for improving the replenishment of slow movers. Our approach is to use the notion of substitution between products as additional information that could serve as a lever of optimization for the optimal replenishment quantity. However, this notion of substitution happens to be critical and unknown for Adeo. Currently, the demand of a client that is not met by sufficient stock has a probability of $\frac{1}{3}$ to be lost, $\frac{1}{3}$ to be backlogged (postponed in the future) and $\frac{1}{3}$ to be substituted. This quantity is valuable in itself because it yields information about the assortment and can be used in various ways, including replenishment. Hence, the probability of substitution is the key quantity to estimate.

In this PhD thesis, our first contribution to the literature is to provide a modelling of the substitution between products close to other works in the field with an EM (Expectation Maximization) estimation procedure and a necessary and sufficient condition for the identifiability of the probabilities of substitution. Estimating substitution based on out of stock is a hard problem because of the sparse data and the high number of parameters to estimate. It implies a high variance of the estimators, hence we provided a statistical test of the significance of the substitution and tested the estimation procedure both on simulated data close to Adeo's framework and also on Adeo's datasets. A second contribution was the proposition of a model that accounts for the numerous zero probabilities of substitution between products to create a first pre-clustering based on a constrained agglomerative hierarchical clustering algorithm. This enables us to account for the problems of identifiability and execution time problems that arise when the number of products becomes larger. The results on real datasets were promising in the case of the small datasets, however not fully informative for larger datasets, even if encouraging. A last chapter concludes our study.

1.3 ORGANISATION OF THE DOCUMENT

Chapter 2 introduces some replenishment models and times series forecasting methods. It presents the state of the art on the topic of slow movers with the difficulties related to the forecast and the replenishment and the actual improvements proposed in the literature. Chapter 3 introduces the notion of assortment and substitution, and more specifically the state of the art in the literature on the estimation of the substitution. It presents how the additional information of substitution can be used to improve replenishment, and how it is especially beneficial for slow movers. In Chapter 4, we introduce a first full model of demand and sales with a numerical experiment that shows the gain in terms of profit of including the substitution in the replenishment process. A second model of demand and sales with its modelling of the substitution is provided along with a necessary and sufficient condition for the identifiability. Follows an estimation procedure and some numerical experiments both on simulated and real data for a low number of products. Chapter 5 presents the sparse model of substitution and the pre-clustering using a constrained agglomerative hierarchical clustering algorithm along with numerical experiments both on simulated and real datasets.

2 ISSUES OF REPLENISHMENT PROCESSES ASSOCIATED TO SLOW MOVERS

In this chapter, we introduce a general overview of an assortment planning process and provide some common techniques for forecasts and replenishment. A definition of slow movers is provided along with arguments showing that there are classic problems associated with slow movers' forecasts and inventory level optimization. We look at the state of the art about forecasts and replenishment procedures and position our contribution in this field by adding multiproduct substitution effects. The notion of substitution between products will be described in Chapter 3.

2.1 GENERAL OVERVIEW OF MONO-PRODUCT MANAGEMENT

2.1.1 *Overview*

In order to optimize its profit, a company chooses a strategy, and from this stems the construction of an assortment of products. An assortment has a certain number of categories (breadth) and a number of products (depth) within these categories. Some firms such as Lidl are more based on a model that provides a wide breadth but small depth, meaning that for a specific usage, there are few choices. Other firms such as Toys R Us have low breadth but large depth in that the categories are restricted to games and for a given type of game there are a lot of choices. These strategies drive in their way a flow of incoming customers. The flow of incoming customers depends on these strategies and so does the supply chain. At the technical level of logistics, a Stock Keeping Unit (SKU) corresponds to the unique identifier of a product which enables the management of the stocks.

Hubner, 2017 conducted on the subject of the main processes in category management a survey of 6 firms including full assortment retailers, discounters, drugs store and mixed grocery/ department retailers weighing more than 10 billion annual sales. They report that category management could be summarised in a 4 part hierarchical model that spans from mid-term decisions to short-term decisions: category sales planning, assortment planning, shelf space planning, and in-store replenishment planning.

In the category of sales planning, the firm takes in the overarching decisions made at the strategical level and makes a midterm sales plan based on midterm predictions about possible income in a particular market and see if it is beneficial. Category planning includes the selection of the categories, the definition of their role, midterm forecasting, and total category shelf space.

Then comes the assortment planning, which is the choice of the SKUs (depth) that will be integrated into stores. Here the product line is designed via forecasting and taking into account the similarity, complementarity, and price ranges according to prerequisite of the category plan.

The shelf planning takes into account information about the products to include and the total category shelf space. It assigns the location of the product on the shelf, which influences the consumer's demand.

On the operational part, the inventory replenishment planning fixes the cycles of replenishment and the quantity to meet a target "on-shelf availability". The inventory level planning is composed of two elements. The first is how much to purchase, and the second is when to make the order.

In this section, we assume that once the assortment is created, the replenishment policy is dealt with in a mono-product way. That is, a choice of a policy is made independently for each product by experts.

2.1.2 *A description of inventory replenishment planning with approximate modelling*

In the following sections of this document, we refer to the probability measure as $\mathbb{P}$, to the expectation operator as $\mathbb{E}$, and to the variance operator as $\mathbb{V}$. Random variables and their realisations are referred to by upper/lower class letters capital such as X, x. Vectors are referred to by a bold letter, such as $\mathbf{X}$. The distribution and cumulative density function of a random variable X is f_X and F_X.

We provide here a modelling of the choice of the inventory policy. Let K be the set of products commercialised at a store. At time $t \in [0, n]$ a client chooses a basket of products among the assortment of K products. The demand at time t is $\tilde{\mathbf{D}}_t = (\tilde{D}_{t1}, \dots, \tilde{D}_{tK})$ with $\tilde{D}_{tk} \in \{0, 1\}$. When a client has a demand for the product k, $\tilde{D}_{tk} = 1$.

The time horizon $[0, n]$ is split into H equal sections of length n/H and over the period $[t_h, t_{h+1}]$ with $h \in \{1, \dots, H\}$. The demand of a product over that period is $D_{hk} = \int_{t=t_h}^{t_h+1} \tilde{D}_{tk} d\delta$ with δ being the Dirac measure.

Let the stock at the time t be $\mathbf{S}_t = (S_{t1}, \dots, S_{tK})$. If a client wants a product at time t and it is available, then it is purchased $V_{tk} = 1$. The random vector of sales at time t is $\mathbf{V}_t = (V_{t1}, \dots, V_{tK})$. The stock is then updated by subtracting the sale hence $S_{t+k} = S_{tk} - V_{tk}$.

A replenishment policy P_k associated with product k is a set of rules that defines when and how much of a product is purchased by the retailer to its supplier. Let $\mathbf{P} = (P_1, \dots, P_K)$ be the vector of policies. A description of the classic policies is given in Section 2.1.3. A policy of replenishment makes an arbitration between the transaction costs such as transportation or reception, costs of Out Of Stock (OOS) when products are unavailable to meet demand, and holding costs induced by the stock itself. The costs of transactions include some variable costs such as transportation, reception, or administration fees according to Vandeput, 2020.

We chose to restrict the costs involved in our framework to $\mathbf{c}^\top = (\mathbf{c}_1^\top, \dots \mathbf{c}_k^\top)^\top$ the costs associated with each product. The costs for product k, $\mathbf{c}_k^\top = (c_{kp}, c_{kc}, c_{ksn}, c_{kh}, c_{kr}, c_{ko})^\top$ includes respectively the selling price, the cost of purchase of the retailer, the salvage cost which is the value of the product at the end of the considered period n, the holding cost of a unit of stock for a given period n, the transaction cost and the OOS cost.

The profit associated with a trajectory of the demand, sales, and replenishment policy $\rho(\tilde{\mathbf{D}}, \mathbf{V}; \theta, \mathbf{c}, \mathbf{P})$ is a function of the parameters $\mathbf{c}$ that includes various costs and prices of each product. An appropriate policy would optimize a measure of the profit $m_\theta \rho(\tilde{\mathbf{D}}, \mathbf{V}; \theta, \mathbf{c}, \mathbf{P})$ where θ includes the parameters of the distributions of $\tilde{\mathbf{D}}, \mathbf{V}$. We provide some details about the prices and inventory policies in Section 2.1.3.

The profit being a complex function, we chose a set of assumptions to describe first the *mono-product inventory model* which is composed of Assumptions 1 and 2. Assumption 1 states that the demands for the products are independent. Assumption 2 states that a sale is deterministic given the demand and the stock of product k. As a consequence, the sales trajectories of each product $\{V_{tk}\}_{k\in[K]}$ are then mutually independent. Recall that $[K] = \{1, \dots, K\}$

Assumption 1 *(Independence of the initial demands) The demand $\{\tilde{D}_{kt}\}_{k\in[K]}$ are mutually independent.*

Assumption 2 *(Sales for the* mono-product inventory model*) A sale is made if there is enough stock, else the sale is lost.* $V_{tk} = 1$ *if and only if* $\tilde{D}_{tk} = 1$ *and* $S_{tk} > 0$.

Assumption 3 *(Stationary costs)* $\mathbf{c}$ *is fixed.*

Remark 1 *In real situations, from the retailer perspective the independence Assumptions 1 and 2 do not result in completely independent inventory policies. The gain of uniting the policies is key for an efficient supply chain because the allocation of transport or supplier relationships is not mono-product. For example, choosing to order several products from a supplier would reduce costs, and pooling products in terms of transportation would also improve the supply chain.*

A *mono-product inventory policy* considers the quantity and the time to replenish a product independently of another product. Despite Remark 1, we believe that it is reasonable and choose in this chapter to adopt it. We define as an *optimal mono-product inventory policy* as a *mono-product inventory policy* that optimizes a certain measure of the profit $\rho_k(\{\tilde{D}_{tk}\}_{t\in[0,n]}, \{V_{tk}\}_{t\in[0,n]}; \theta_k, \mathbf{c}_k, P_k)$ for a specific product k on a time frame $t \in [0, n]$. Note that since the demands are independent in the *mono-product inventory policy*, we can refer to $\theta = (\theta_1, \dots, \theta_K)^\top$ as the parameters of the distribution of the demands. In this framework, the distribution of the demand for product k has parameters $\theta_k \in \mathbb{R}^p$. The total profit for the assortment would then be $m_\theta \rho(\tilde{\mathbf{D}}, \mathbf{V}; \theta, \mathbf{c}, \mathbf{P}) = \sum_{k=1}^K m_\theta \rho_k(\{\tilde{D}_{tk}\}_{t\in[0,n]}, \{V_{tk}\}_{t\in[0,n]}; \theta_k, \mathbf{c}_k, P_k)$.

Nevertheless, explicating a profit function including an inventory policy is complex, hence in practice, the choice of the policy is done in two steps, first a choice of a policy to treat the frequency of the purchases of product k denoted as r_k. The closer the replenishment dates, the less quantity is needed. The further away, the more stock is needed, and hence stock cost increases. The problem is the choice of the number of transactions with the supplier that can be expected over a period of time and the cost of holding the stocks.

Second, the quantity to replenish $Q_k(r_k)$, the retailer mitigates the variation of the demand between the two dates. It is a compromise between the OOS and the excessive stock cost associated respectively to a lack and excess of stock to meet the demand. The lead time L is the time between the order to the supplier and the delivery. This second part can be referred to as choosing a safety stock. According to that modelling, $P_K = (r_k, Q_k(r_k))$ summarises the *mono-product inventory policy*. We refer to the optimal policy as

$$P_k^* = \text{argmax}_{P_k} m_{\theta_k} \rho_k(\{\tilde{D}_{tk}\}_{t\in[0,n]}, \{V_{tk}\}_{t\in[0,n]}; \theta_k, \mathbf{c}_k, P_k). \quad (1)$$

The safety stock to replenish given r_k^* is

$$Q_k^*(r_k^*) = \text{argmax}_Q m_{\theta_k} \rho_k(\{\tilde{D}_{tk}\}_{t\in[0,n]}, \{V_{tk}\}_{t\in[0,n]}; \theta_k, \mathbf{c}_k, (r_k^*, Q_k(r_k^*))). \quad (2)$$

In the following sections, once r_k is defined for the notation we use $Q_k \in \mathbb{R}^+$ instead for the quantity to replenish. These two aspects of the *mono-product inventory policy* are described in Section 2.1.2. Note that the cost of OOS is difficult to model due to global effects, such as the loss of a customer induced by dissatisfaction. This is why experts usually use other types of rules to account for the OOS. Such as the service level on a cycle which is the probability of having no stock out on the period considered (Definition 1) or the fill rate which is the proportion of the demand that is met (Definition 2). A price for stock out c_{ko} can be leveraged along with a holding cost c_{kh} to make a compromise between expected unit short over the period h defined above $\mathbb{E}[(D_{hk} - Q_k)^+]$ versus expected excess of stock $\mathbb{E}[(Q_k - D_{hk})^+]$, two quantities computed from the same distribution.

Definition 1 *The fill rate* β *is defined as the expected part of the demand for product* k *that will be met with the on hand inventory over cycle* h*:*

$$\beta = 1 - \frac{\mathbb{E}[(D_{hk} - Q_k)^+]}{\mathbb{E}(D_k)}.$$

Definition 2 *The service level for product* k *over cycle* h *for a stock is defined as*

$$\alpha = \mathbb{P}(D_{hk} > Q_k).$$

Assuming that the service level proposed is fixed based on an expert judgment, the optimal quantity for a specified service level $\alpha(\theta_k)$ that determines the right quantity to replenish is

$$Q^*_{hk} = \inf\{Q \in \mathbb{R}^+ : \mathbb{P}(D_{hk} \leqslant Q) \geqslant \alpha\}. \tag{3}$$

In the retail sector, $\alpha = 0.95$ is a common value for the safety stock.

2.1.3 *Mono-product inventory policies*

We provide in the following subsections a review of the most classical inventory policies. These models can be found in Vandeput, 2020, Caplice and Ponce, 2020, Schoot, Heuts, and Strijbosch, 2000, Choi, 2012. The objective here is related to Problem 1 which is to choose the right policy that maximises the profit. As said earlier, in practice the problem is solved by first choosing a policy that will set the orders frequency and then how much is replenished.

We introduce first the Economic Order Quantity which is modelled with a deterministic demand. Then two policies (s, Q) and (R, S) that target service level rather than incorporate explicit values of the costs. The last model introduced is the newsvendor model (N̦) which is based on explicit values of the costs.

2.1.3.1 *Economic order quantity (EOQ) model*

The EOQ provides the quantity to order and how much re-order points equally spaced we need. The EOQ model is simple (Vandeput, 2020), but is based on restrictive assumptions. The assumption is that the demand is deterministic. It is constant over the H periods, and the total demand is D_k over the chosen horizon. A re-order is made when the stock reaches 0 and there is no lead time (no time between the time of re-order and the delivery). The EOQ for the product k is a solution of a cost minimisation problem including the holding cost of the stock c_{kh}, a backorder cost c_{kb}, a transaction cost c_{kt} and the delivery cost per unit c_{kd}. The fixed order quantity

$$Q_k^* = \text{argmax}_{Q_k} c_{kh} \frac{Q_k}{2} + c_{kk} \frac{D_k}{Q_k} + c_{kt} D_k$$

which yields $Q_k^* = \sqrt{\frac{2c_{kk}D_k}{c_{kh}}}$. The EOQ also implies the transaction period

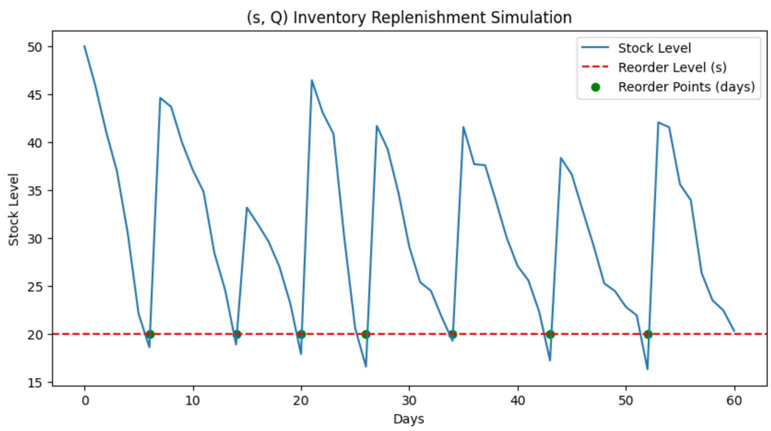
$$T_k^* = \frac{Q_k^*}{D_k} = \sqrt{\frac{2c_{kk}}{c_{kh}D_k}}.$$

The EOQ model has good mathematical properties. If the quantity Q_k is close to Q_k^* then the total cost will be close to the optimal one (mathematical proof is provided in Vandeput, 2020).

2.1.3.2 *The continuous review policy* (s, Q)

The main characteristic of the (s, Q) method (Vandeput, 2020, Caplice and Ponce, 2020) is that an order can be made at any time and the stock is reviewed continuously. A threshold s called reorder point is set and when the stock falls below an order is made for Q. The value of Q can be set by an EOQ, Q^* following the method described in the previous section, including the cost of transaction and the cost of holding the stock. A forecast of the demand is made over a prolonged period of time (such as a year) and then given the cost of transaction and the cost of holding stock we use the EOQ model to fix Q^*. Fixing the period at $H = n/L$ where L is the lead time, we get that the quantity s makes a trade-off between the under-stock and the over-stock over the fixed lead time L. Figure 4 is an example of a (s, Q) continuous policy with $s = 20$, $Q = 30$ on a horizon of 60. Generally, using $s = Q_k^*$, we have Q_k^* solution of Problem 3.

Figure 4: Replenishment policy (s,Q)

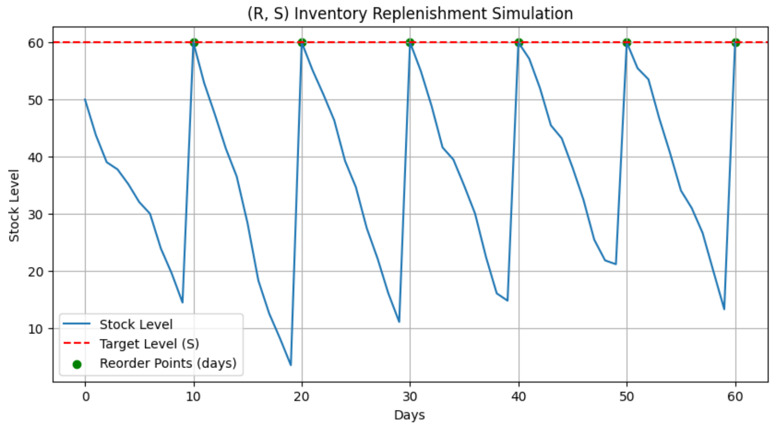


2.1.3.3 *The periodic review policy (R,S)*

Policy (R,S) is a re-order level policy with an order that can be placed each R period. A level S is chosen and in each R period the difference between the current stock and S is ordered. The periodicity R is computed in the same way as T does in the EOQ optimization problem based on holding and transaction costs. The target level

S makes a trade-off between the OOS cost and the holding cost associated with the demand over the period between two replenishments $R + L$ where R is the re-order interval and L is the lead time. Q_k^* can be computed based on Problem 3 for a chosen service level α. Figure 5 is an example of such a policy with $S = 60$, $R = 10$, a time horizon of 60 periods, and no lead time $L = 0$.

Figure 5: Replenishment policy (R,S)



2.1.3.4 *The base stock policy*

The base stock policy is a periodic review policy (R, S) with the objective to restore the base stock S each period $R = 1$. The base stock S is computed to mitigate the OOS cost and holding stock over the lead time L, hence $S = Q_k^*$ is the solution of Problem 3 where α is the target service level chosen by the retailer.

2.1.3.5 *Newsvendor model*

The newsvendor model Newsvendor Model (NVM) first introduced by Kenneth J. Arrow and Marschak, 1951 is a model of the policy of replenishment based on the random variables of the demand $\mathbf{D}, \mathbf{V}$ and explicit values of the costs. In its simplest form, introduced in this section, NVM is a single period model of the replenishment associated with a problem whose optimal quantity of stock is a solution to Problem 2 where the measure m_θ is the expectation $\mathbb{E}_\theta$ operator.

The usual presentation of the model refers to a newsvendor wanting to know the quantity of newspaper he has to buy in order to maximize his expected profit. He knows its selling price c_{kp} and purchasing cost c_{kc} along with the value at the end of the period (salvage value) c_{ks}. Let D_{hk} be the demand associated with the newspaper (product $k = 1$). The quantity purchased by the newsvendor is the inventory level Q_k. The quantity of unit short is $(D_k - Q_k)^+$. The quantity of unit in excess is $(Q_k - D_k)^+$. The optimal inventory level for the period h is given by

$$Q_{hk}^* = \text{argmax}_{Q_{hk} \in \mathbb{N}} \mathbb{E}\left[c_{kp} \min(Q_{hk}, D_{hk}) + c_{ks} \max(0, Q_{hk} - D_{hk}) - c_{kc} Q_{hk}\right]. \tag{4}$$

NVM is a model that has been thoroughly studied and has various extensions, see Choi, 2012 for an exhaustive presentation of the extensions. D_{hk} may be discrete or continuous. We provide here a model that will be extended for substitutions and related analysis in Chapter 3.

THE CHOSEN NEWSVENDOR MODEL We consider the replenishment over a single period that we identify with $[0, n]$. The formula of the profit is a function of the aggregated sales D_{hk} hence for this paragraph we drop the reference to the period h for ease of reading. The explicit chosen costs are the holding cost c_{kh}, the salvage value c_{ks}, the purchase price c_{hp}, the cost of OOS c_{ho} and the cost of the product c_{hc}. The total cost associated respectively to hold the stock, to the OOS is linear in the associated cost hence

$$\rho_k(\{\tilde{D}_{tk}\}_{t\in[0,n]}, \{V_{tk}\}_{t\in[0,n]}; \theta_k, Q_k) = c_{kp}\min(Q_k, D_k) - (c_{kh} - c_{ks})\max(0, Q_k - D_k) \\ - c_{ko}\max(0, D_k - Q_k) - c_{kc}Q_k.$$

The optimal replenishment in NVM is then given by

$$Q_k^* = \text{argmax}_{Q_k\in\mathbb{R}}\rho_k(\{\tilde{D}_{tk}\}_{t\in[0,n]}, \{V_{tk}\}_{t\in[0,n]}; \theta_k, Q_k). \tag{5}$$

Here the cost of one unit shortage is the lost margin plus the additional cost of not fulfilling the order ($c_k^- = c_{ko} + c_{kp} - c_{kc}$). The unit cost for an excess of one unit (excessive stock cost) here is $c_k^+ = c_{kh} + c_{kc} - c_{ks}$, and we have

Proposition 1 *Problem 5 is equivalent to*

$$Q_k^* = \mathit{argmax}_{Q_k\in\mathbb{R}}\mathbb{E}_{\theta_k}\left((c_{kp} - c_{kc})D_k - c_k^-(D_k - Q_k)^+ - c_k^+(Q_k - D_k)^+\right) \tag{6}$$

which then yields

$$Q_k^* = \mathit{argmin}_{Q_k\in\mathbb{R}}c_k^+\mathbb{E}_{\theta_k}((Q_k - D_k)^+) + c_k^-\mathbb{E}_{\theta_k}((D_k - Q_k)^+). \tag{7}$$

Proof 1 is provided Appendix.

As mentioned in the other *mono-product inventory policies* the trade-off between OOS and excessive stock is key for the computation of the replenishment quantity. Proposition 1 states it for the NVM. The following result is meaningful, it relates the service level to the costs of OOS and excessive stock in a simple formula

$$\alpha = \frac{c_k^-}{c_k^- + c_k^+}. \tag{8}$$

Proposition 2 *Let D_k be a continuous random variable with distribution function f and cumulative distribution function F. The optimal inventory level is given by $F(Q_k) = \frac{c_k^-}{c_k^- + c_k^+}$.*

This last results provides the relation between the service level and the costs for a discrete variable. The proof is not given but can be found in Vandeput, 2020.

Proposition 3 *Let D_k be a discrete random variable. The optimal inventory level is the lowest Q_k such that $\mathbb{P}(D_k \geqslant Q_k) \geqslant \frac{c_k^-}{c_k^- + c_k^+}$. Proof 2 is provided in Appendix.*

Example 1 *If the cost of an out of stock is 95 euros and a cost per period is five euros the optimal service level would be $\alpha = 95\%$.*

The optimal quantity is expressed in a tractable way as can be seen both from the Proposition 3 and Example 1. The intuition is that the most costly a shortage is compared to an excess, the more inventory will be held. The quantile will in practice depend upon the type of product sold: the fresh product has a high excess cost, the spare part a high shortage cost.

2.1.4 *Probabilistic and point forecasts*

2.1.4.1 *Motivation for the estimation and demand forecast*

In various processes including the *mono-product inventory policy*, the information about the demand is the input for making the best decisions. The demand being stochastic, giving a specific number as a forecast results in an impoverishment of the information at hand hence the need to retrieve a full distribution. A probabilistic forecast is a forecasted distribution for a given period and since it is dependent on observations it is random. In order to provide some theoretical aspects we refer to the work of Gneiting and Ranjan, 2013, we chose to define the forecast as a probability measure on the observations Y when dealing with the theory and D_{hk} when dealing with the demand framework. $\mathcal{A}$ is the set of events with $\mathcal{A}_1 \subset \mathcal{A}$ being a sub σ-algebra containing the information at hand such as data and expertise.

Definition 3 *A prediction space is defined as $(\Omega, \mathcal{A}, \mathbb{Q})$ with $\mathcal{A}_1 \subset \mathcal{A}$ being a sub σ-algebra. Elements of Ω are (F, Y) and $\mathbb{Q}$ is the joint distribution of the two. It verifies :*

- *F is the random Cumulative Density Function (CDF) measurable on $\mathcal{A}_1$.*
- *Y is a real valued random variable.*

Definition 4 *Let $\mathcal{L}$ be a conditional law. A random CDF is ideal with respect to the sub σ-algebra $\mathcal{A}_1$ if $F = \mathcal{L}(Y|\mathcal{A}_1)$.*

An ideal CDF can be seen as optimal based on the information provided by $\mathcal{A}$.

In the case of the time series such as D_{hk}, the probabilistic forecast for the future period $H+\tau$ with $\tau \in \mathbb{N}$ based on observations $\mathbf{D}_k = (D_{1k}, \dots, D_{hk})^\top$ is referred as $\hat{F}_{H+\tau|H}(\bullet; \mathbf{D}_k)$. A point forecast $\hat{D}_{h+\tau|h}$ is a forecast of the value the demand would have in τ days according to a forecast method m.

2.1.4.2 *Famous distributions of the demand*

We review the most famous parametric distributions of the demand. It is classical to model the arrival of customers by a Poisson law. The clear link between the incoming flow of clients and the demand contributes to make this law one of the most used (Vandeput, 2020, Anupindi, Dada, and Gupta, 1998). The distribution can be found in Table 1. The Poisson law is a one parameter distribution with equal variance and mean. We use this distribution of the demand in our contributions. According to Agrawal and Smith, 1996, the negative binomial is a better choice. It is also a popular distribution for the distribution of demand used in (Smith et al., 2000). The demand is positive and often shows a skewness on the right tail which is due to the low probability of having a high demand. Moreover the demand in the retail sector also has to deal with unobservable lost sales due to the OOS. Figure 6 shows the right tail is longer for negative binomial.

The demand can also be approximated by continuous distribution such as the normal law which exhibit nice properties such as a formula for the quantiles $F_{D_{hk}}^{-1}(\alpha) = \mu + \Phi^{-1}(\alpha)\sigma$ often used for safety stocks computation 2.1.3. Φ and ϕ in the rest of the document are respectively the cumulative distribution function and the density of the normal centred and scaled law. But it has some limitations such as the fact that the distribution is symmetrical (it can be a limiting assumption) and the fact that the

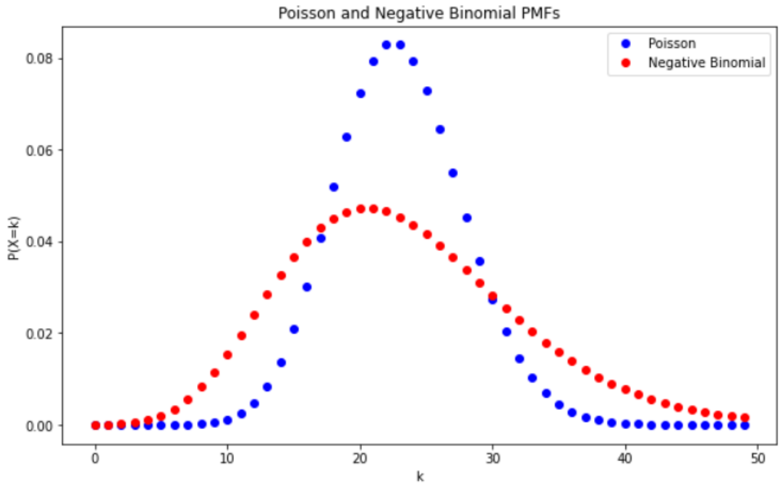


Figure 6: Figure of the Poisson and Negative Binomial Distributions

demand can be negative which is not possible in practice. However Schoot, Heuts, and Strijbosch, 2000 uses the gamma law which is more convenient, it has a positive support and two parameters that account for the skewness of the right tail. If the smallest value of the dataset is strictly positive this law can be shifted by the minimum value. AZZALINI, 1985 proposes a skew normal distribution which takes three parameters accounting the mean, variance and skewness of the distribution and is a generalization of the normal law.

After choosing the model for the demand, parameters are estimated based on observed data. Assuming the stationarity of the demand enables to see the past observations $(D_{hk})_{h\in[H]}$ as i.i.d samples of the distribution. Then via classical techniques such as maximum likelihood maximization the parameters of the demand can be retrieved. Semi parametric and non parametric distributions such as bootstraping are an alternative.

For example, in the case of the Gaussian distribution, the maximum likelihood estimator is $\hat{\mu} = \frac{1}{H}\sum_t D_{hk}$ and the variance is $\hat{\sigma}^2 = \frac{1}{H}\sum_{h\in[\![0,H]\!]}(D_{hk} - \frac{1}{H}\sum_t D_{hk})^2$.

Table 1: Common distribution of demand

Distribution	distribution
$\mathcal{P}(\lambda)$	$\frac{\lambda^k}{k!}\exp(-\lambda)$
$\mathcal{NG}(r,p)$	$\binom{k+r-1}{k}p^r(1-p)^k$
$\mathcal{N}(\mu,\sigma)$	$\frac{1}{\sqrt{2\pi}\sigma}e^{\frac{(x-\mu)^2}{2\sigma^2}}$
$\mathcal{G}(k,\theta)$	$\frac{1}{\Gamma(k)\theta^k}x^{k-1}e^{-\frac{x}{\theta}}$
$\mathcal{SN}(\xi,\omega,\alpha)$	$\frac{2}{\omega}\phi(\frac{x-\xi}{\omega})\Phi(\alpha(\frac{x-\xi}{\omega}))$

In next Section 2.1.4.3, we provide some forecasting techniques that assume the demand may be non stationary and we provide next an introduction to the calibration and the score functions that enables the evaluation of the quality of the forecasted distribution and permits to define a link between the probabilistic forecasts, point forecasts and their accuracy measures. This is motivated by the fact that point forecasts are the provided information for the replenishment.

2.1.4.3 *Some usual time series forecasting methods*

When the demand is not stationary and the $\tilde{\mathbf{d}}$ are not iid, a time series forecast takes as input the data over a period of time $(D_h)_{h\in[H]}$ and outputs a value at the period h for a specific future period τ $\hat{D}_{h+\tau|h}$.

A first naive method assumes a future period has the value of the current period $\hat{D}_{h+\tau|h} = D_h$.

The simple exponential (SES) was investigated around 1950-1960. The seminal work can be found in Brown, 1959, Holt, 1957 and Winters, 1960. Given a smoothing parameter α, the forecasted demand specified by $\hat{D}_{h+1|h}$ is computed via fixing a_0 arbitrarily and computing

$$
\begin{aligned}
a_h &= \alpha D_h + (1-\alpha)a_{h-1} \\
\hat{D}_{h+\tau|h} &= a_h.
\end{aligned}
\tag{9}
$$

Equation 9 reflects the fact that exponential smoothing is a flat forecast with $\hat{D}_{h+\tau|h} = \hat{D}_{h+1|h}$ and not tailored for time series with trends or seasonality. The parameter $\alpha \in [0,1]$ assigns more weight to the new values when close to 1. The adjective "exponential" refers to the fact that the weight attached to a D_h is decreasing exponentially. Another version of the estimator is the weighted average formula:

$$
a_h = \sum_{j=0}^{h-1} \alpha(1-\alpha)^j D_{h-j} + \alpha(1-\alpha)^h a_0.
$$

In pratice, values of different α can be evaluated in terms of performances to retrieve the best value.

The double exponential smoothing (also called Holt's linear trend method Holt, 1957) is based on the same principle and adds a linear trend term. Let $\alpha \in]0,1[$ and $\gamma \in]0,1[$ be the smoothing parameters associated respectively to the data and to the trend. The intialization of the variables are $a_0 = D_0$ and $b_0 = D_1 - D_0$.

$$
\begin{aligned}
a_h &= \alpha D_h + (1-\alpha)(a_{h-1} + b_{h-1}), \\
b_h &= \gamma(a_h - a_{h-1}) + (1-\gamma)b_{h-1}, \\
\hat{D}_{h+\tau|h} &= a_h + h b_h.
\end{aligned}
$$

Another alternative is the Brown's method which is similar and can be found in Brown, 1959. The triple exponential method (also called Holt Winters method) adds a seasonal term. It involves smoothing parameters for the data, trend and seasonality respectively named α, γ, ζ. Two versions exist, the first with additive seasonality and the second with multiplicative. We provide the additive version for a cycle of length L. Let $a_0 = D_0$, we have

$$
\begin{aligned}
a_h &= \alpha(D_h - D_{t-L}) + (1-\alpha)(a_{h-1} + b_{h-1}), \\
b_h &= \gamma(a_h - a_{h-1}) + (1-\gamma)b_{h-1}, \\
c_h &= \zeta(D_h - a_{h-1} - b_{h-1}) + (1-\zeta)D_{t-L}, \\
\hat{D}_{h+\tau|h} &= a_h + h b_h + D_{t-L+((h-1) \mod L)}.
\end{aligned}
$$

Other methods can involve exterior information and more practices can be found in Rob J Hyndman and Athanasopoulos, 2018 and Kolassa, 2020 for an exhaustive review of the forecasting in retail. If we relax the assumption of stationarity of the demand, it is possible to combine forecasting techniques with standard estimation of the demand's probability distribution in order to compute the future demand. For example, evaluating the mean on past data using exponential smoothing and the standard error with the RMSE (the root of the empirical risk introduced in Equation 2) can be used with a Gaussian distribution.

2.1.4.4 *Calibration of the estimation*

In the case of probabilistic forecast the candidate distribution is evaluated in terms of the calibration and the concentration. The former is the property of how well the distribution fits the data and the second refers to the pread of the possible value around a point forecast.

The probability integral transform is a statistic based on the observation and the forecast distribution that enables the calibration.

Definition 5 *Let $V \sim \mathcal{U}[0,1]$ independent from the forecast CDF F and Y the observations. The Probability Integral Transform (PIT) of F is defined as*

$$Z_F = \lim_{x \to Y^-} F(x) + V(F(Y) - \lim_{x \to Y^-} F(x)).$$

For a function F that is continuous and if $Y \sim F$ then Z_F is uniform. Based on this property we define the notion of calibration as

Definition 6 *The forecast F is stochastically calibrated if the PIT Z_F has a standard uniform distribution.*

Definition 7 *The forecast F is marginally calibrated if $\mathbb{E}_Q[F(Y)|Y = y] = Q(Y \leqslant y)$.*

Associated to the notion of calibration there is the notion of dispersion. A forecast is over dispersed if the histogram of the PIT is inverse U shaped and under dispersed if it is U shaped.

Definition 8 *The forecast is underdispersed if $\mathbb{V}(Z_F) < \frac{1}{12}$ and overdispersed is $\mathbb{V}(Z_F) > \frac{1}{12}$.*

It is also possible to compare the PIT of two candidate distributions F, G .

Definition 9 *F is more dispersed than G if $\mathbb{V}(Z_F) \leqslant \mathbb{V}(Z_G)$.*

According to Gneiting, Balabdaoui, and Raftery, 2007 the PIT does not permit to discriminate in all situations whether a candidate distribution is preferable over an other. He provides the example of 4 distributions candidate and the PIT histogram is similar for all despite the fact that the true distribution is included. The authors then propose to add a second criterion through the notion of sharpness.

2.1.5 *Scoring rule and score functions*

Drawing from Gneiting and Katzfuss, 2014, the concept of score is here defined and some of its properties and use cases are explored. The evaluation of the quality of the forecasted distribution to the data $\tilde{d}$ may be evaluated with a scoring rule. The scoring rule evaluates both the callibration and the sharpness. Let $\mathcal{F}$ be a convex class of distribution functions on $\mathbb{R}$ and Ω be the set of values of $\mathbf{y}$. A scoring rule is a function $S : \mathcal{F} \times \Omega \to \bar{\mathbb{R}}$ with $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ the completed real line.

Definition 10 *Let* $F \in \mathcal{F}$ *be a forecast cumulative density function. Let* $S(F, G) = \mathbb{E}_G[S(F, Y)]]$ *be the expected score associated to the function* F *given that* G *is the real distribution. A proper scoring function* S *for the class of function* $\mathcal{F}$ *verifies*

$$S(G, G) \leqslant S(F, G), \forall (F, G) \in \mathcal{F} \times \mathcal{F}.$$

It is strictly proper if $S(G, G) = S(F, G)$ implies that $F = G$. We provide a formulation characterising a proper scoring rule S.

Theorem 1 *The scoring rule* S *is proper relative to* S *if the expected score function* $e(F) = S(F, F)$ *is concave and* $S(F, \bullet)$ *is a super gradient of* e *at the point* F, $\forall F \in \mathcal{F}$.

A super gradient is a generalization of the gradient for functions that are not smooths. It verifies the following property.

Definition 11 *Let* $C \subset \mathbf{R}^m$ *be a convex set, and let* $F : C \to \mathbf{R}$ *be concave. A vector* p *is a supergradient of* F *at the point* x *if for every* y *it satisfies the supergradient inequality,*

$$F(x) + p \cdot (y - x) \geqslant F(y)$$

For concave F*, the set of all supergradients of* F *at* x *is called the superdifferential of* F *at* x*, and is denoted* $\partial f(x)$.

Among the scoring rules we have the logarithmic score which is defined as

$$LS(f, y) = -\log(f(y))$$

which is associated to the maximum likelihood estimation of the distribution. The quadratic score

$$QS(f, y) = -f(y) + \int_{\mathbb{R}} f^2(x)dx$$

which is associated to the density functions that are square integrable which corresponds to the least square estimation. A list of other scoring rules is available in the paper Gneiting and Katzfuss, 2014.

In the retail sector, for the decisions of the replenishment it is necessary to compute a point forecast which is defined as a value in $\mathbb{R}$ optimal for a given distribution.

Definition 12 *A scoring function or loss function is a function* s *that verifies* $s : \Omega \times \Omega \to \mathbb{R}^+$.

Definition 13 *Let* F *be the forecast distribution ("predictive" in the paper) and* $s(x, y)$ *be a non negative loss function / scoring function which assigns the loss to the point forecast* x *given that* y *is the realized value. The Bayes rule or optimal point forecast is then given by* $\hat{x} = argmin_x \mathbb{E}_F[s(x, Y)]$ *where* Y *follows* F.

A statistical functional is defined as $T : \mathcal{F} \to \mathcal{P}(\mathbb{R})$ for which the expectation, quantiles are examples. A scoring function s is consistent for the functional T relative to the class $\mathcal{F}$ if $\mathbb{E}_F[s(t, Y)] \leqslant \mathbb{E}_F[s(x, Y)]$, $\forall F \in \mathcal{F}$, $t \in T(F)$ and $x \in \mathbb{R}$. It is strictly consistent if the equality implies that $x \in T(F)$.

A consistent scoring function generates a proper scoring rule.

Theorem 2 *Let* s *be a scoring function consistent for* T *relative to a convex class* $\mathcal{F}$*. Let* $t_F \in T(F)$ *then* $S(F, y) = s(t_f, y)$ *is a proper scoring rule relative to the class* $\mathcal{F}$.

2.1.5.1 *Ellicitable functionals, point forecasts and accuracy measures*

The value t_f is a point forecast.

Definition 14 *Functionals* T *which have a scoring function strictly consistent are called ellicitable.*

Among the functionals that are elicitable there is the expectation and α-quantile which have the related scoring functions respectively $s(x, y) = (x - y)^2$ and $s(x, y) = (\mathbb{1}_{y<x} - \alpha)(x - y)$.

In practice, a point forecast is a quantity that answers to a specific question such as, "What will be the level of sale of this specific product in 6 months?" from the demand perspective and "Which stock should i order to cover a one month period?" from a replenishment perspective. To each of these questions there is associated a loss function that enables the evaluation of the quality of the response.

A point forecast stemming from a distribution $F_{D_{hk}}$ can be expressed in terms of a functional $T : \mathcal{F} \to \mathcal{P}(\bar{\mathbb{R}})$. For example, the expected sale for the period h is expressed as $T(F_{D_{hk}}) = \{\mathbb{E}_{\theta_k}(D_{hk})\}$, and the variance $T(F_{D_{hk}}) = \{\mathbb{V}_{\theta_k}(D_{hk})\}$. As seen in the previous section in Theorem 2 there is an association between a functional and a loss function.

In practice, there is a risk function associated to the loss function, an empirical approximation of it. Examples 2, 3, 4 show some loss functions used in practice.

Example 2 *The risk defined by the mean squared error* $\mathbb{E}[[Y_h - \hat{Y}_h]^2]$ *is approximated by the in-sample mean squared error* $H^{-1}\sum_{h=1}^{H}[Y_h - \hat{Y}_h]^2$ *and the loss function is quadratic* $s(x, y) = (x - y)^2$. *The associated functional is the expectation meaning that the optimal value in terms of risk is the mean of the distribution.*

Example 3 *The risk defined by the mean absolute error* $\mathbb{E}[|Y_h - \hat{Y}_h|]$ *is associated with the empirical mean absolute error* $H^{-1}\sum_{h=1}^{H}|Y_h - \hat{Y}_h|$. *The loss function is* $s(x, y) = |x - y|$ *and the functional is the median operator.*

Example 4 *The risk defined by the loss function* $s(x, y) = (\mathbb{1}_{y<x} - \alpha)(x - y)$ *(also called pinball function (Biau and Patra, 2000) is associated with the* α *quantile functional. The risk associated to it is* $\mathbb{E}[\mathbb{1}_{Y_h<\hat{Y}_h} - \alpha)(\hat{Y}_h - Y_h)]$. *The empirical risk is* $H^{-1}\sum_{h=1}^{H}[\mathbb{1}_{Y_h<\hat{Y}_h} - \alpha)(\hat{Y}_h - Y_h)]$.

2.1.5.2 *Other forecast accuracy measures*

Additional to the empirical risk provided in the last section, some other empirical risks are worth stating because of their relation to the retail forecasting. Following Rob J. Hyndman and Koehler, 2006, there are 4 different empirical risks: the scale dependent metrics, the percentage error metrics, the mean ratio of error between two methods, the scale independent methods.

In the scale dependent empirical risk, we cite the mean absolute error (MAE) from Example 3, the mean squared error (MSE) from Example 2, the empirical risk associated to the pinball loss function from Example 4 and the geometric mean absolute error $\left(\prod_{h=1}^{H}|Y_h - \hat{Y}_h|\right)^{\frac{1}{H}}$. The scale dependent empirical risks are a relevant for evaluating different forecasting methods on the same time series but are irrelevant across time series with different scales.

The percentage errors include the mean absolute percentage error (MAPE) MAPE $= \frac{100\%}{T} \sum_{h=1}^{H} \frac{|Y_h - \hat{Y}_h|}{Y_h}$ which is easy to interpret in term of percentage of error and assigns more importance to positive errors than negative errors (Rob J. Hyndman and Koehler, 2006). So an alternative has been suggested which is the symmetric mean absolute percentage error (SMAPE) equal to $\frac{100\%}{H} \sum_{h=1}^{H} \frac{|Y_h - \hat{Y}_h|}{(Y_h + \hat{Y}_h)/2}$. However it can have a negative value. These measures suffer from their denominators when the observations Y_h have numerous zero values.

It is possible to compare errors between different forecasting methods. Let $\tilde{Y}_t$ be the forecast for a second method. Possible measures are the median $\left(\frac{|Y_h - \hat{Y}_h|}{|Y_h - \tilde{Y}_t|}\right)$. This measure is scale independent so it is possible to compare accross multiple time series. This type of measure is not suited for intermittent demand because low values of errors yields poor accuracy.

The last one is scale independent methods such as the MASE $\left(\frac{\frac{1}{H}\sum_{h=1}^{H} |Y_h - \hat{y}_h|}{\frac{1}{T-1}\sum_{t=2}^{T} |Y_h - Y_{h-1}|}\right)$ which is recommended in the case Rob J. Hyndman and Koehler, 2006 because it does not have the limitations of the other categories. We can also cite the SWAPE $\left(\frac{\sum_t |Y_h - \hat{Y}_h|}{\sum_t |Y_h + \hat{Y}_h|}\right)$ that is used at Adeo.

2.2 STATE OF THE ART ON SLOW MOVERS

In the assortment of Adeo, the frequencies of sale and the prices of products are characteristics driving the importance of products in terms of turnover nevertheless forecasting time series with different frequencies can be challenging. Time series of slow movers or fast movers need different tools both in terms of forecasting techniques and in terms of choice of empirical risk. In this section we provide the most accepted definition of a slow movers, derive cut-off points to distinguish from fast movers and show the shortcomings both in terms of forecast and replenishment.

2.2.0.1 *A partition of the products in four types*

A slow mover also referred to as "product with intermittent demand" or "unfrequently purchased" (A. A. Syntetos, Boylan, and Croston, 2005, Miller et al., 2010) is defined in relation to a given time granularity. For a fixed temporal horizon $[0, n]$, the subdivision into H periods with H increasing increases the number of zeros in the time serie $(\tilde{d}_{1k}, \ldots, \tilde{d}_{Hk})$.

Remark 2 *The definitions of slow movers in practice tend to be inconsistent (Kwan, 1991) from a company to another due to variations of volume of demand, frequency of sale or lead time. We provide some information about it in Section 2.2.1.*

Definition 15 characterises slow movers by the probability of no sales on a given period.

Definition 15 *Given a threshold* η*, a slow mover is a product so that* $\mathbb{P}(D_{hk} = 0 \mid \theta_k) \geqslant \eta$.

In the literature Croston, 1972, the threshold η should be defined in terms of difference best forecasting method. However, the demonstration of the value of this cut-off point is also related to the coefficient of variation of the quantity once a sale is made thus including a more precise partition of the types of products. A partition which separates products as follows: the lumpy demand has long inter demand interval (p) and high coefficient of variation for the quantity (CV_q) when a demand occurs. Erratic demand is characterised with short inter demand interval and high coefficient CV_q. Intermittent demand has long inter demand interval and low CV_q and smooth demand is the last class. We can note that for lumpy items, the inter demand intervals can be long and yet the number of sales high. Often this is not the case.

The demonstration of the cut-off values along the two dimensions build on a modelling of the demand introduced by Croston, 1972, the choice between four proposed methods of forecasting and the demonstration in a theoretical framework of the cut-off values.

CROSTON MODELLING OF INTERMITTENT DEMAND Croston, 1972 proposed a novel modelling of the products demand relevant to the infrequently purchased products. Let $\theta_k = (\mu, \sigma, p)$ be the parameters of D_{hk} such that $D_{hk} = z_{hk} * x_{hk}$ with $z_{hk} \sim \mathcal{N}(\mu, \sigma^2)$ denoting the quantity demanded in the period t and $x_{hk} \sim \mathcal{B}(1/p)$ being 1 if a demand occurs and 0 otherwise. The related inter demand interval is denoted by $p_{hk} \sim \mathcal{G}(p)$ with $\mathcal{G}$ being the geometric law.

CANDIDATE FORECASTING METHOD Based on the past data of demand $((D_{hk})_{h \in [\![0,H]\!]}$, we introduce the simple exponential smoothing

$$\hat{D}_{h+1|h} = \alpha \tilde{D}_h + (1-\alpha)\hat{\tilde{D}}_{h|h-1}$$

where α is the smoothing parameter. The forecast verifies $\hat{D}_{h+\tau|h} = \hat{D}_{h+1|h}$ for $h \in \mathbb{N}$.

Croston, 1972 proposed a method for forecasting based on similar technique as SES: let $\alpha \in]0,1[$ be a smoothing parameter. p'_h is the exponentially smoothed interdemand interval and z'_h is the exponentially smoothed demand quantity.

$$\hat{p}_h = \begin{cases} \alpha p_h + (1-\alpha)\hat{p}_{h-1} & \text{if } \tilde{D}_h \neq 0 \\ \hat{p}_{h-1} & \text{if } \tilde{D}_h = 0, \end{cases}$$

and

$$\hat{z}_h = \begin{cases} \alpha z_h + (1-\alpha)\hat{z}_{h-1} & \text{if } \tilde{D}_h \neq 0 \\ \hat{z}_{h-1} & \text{if } \tilde{D}_h = 0. \end{cases}$$

The estimator of the demand is $\hat{D}_{h+1|h} = \frac{\hat{z}_h}{\hat{p}_h}$. According to A A Syntetos, 2001 this method is biased and he provided a modified version $\hat{D}_{h+1|h} = (1-\frac{\alpha}{2})\frac{\hat{z}_h}{\hat{p}_h}$. We refer to this method by SBA (Syntetos Boylan Approximation). The two methods verify $\hat{D}_{h+\tau|h} = \hat{D}_{h+1|h}$ for $h \in \mathbb{N}^*$.

GENERAL DEFINITION OF DEMAND PATTERNS ALONG 2 DIMENSIONS The criteria used in the literature to discriminate between different demand patterns including slow movers is based on two dimensions: the coefficient of variation of the size of products sold when a sale happens which is $CV(\tilde{D}_h|\tilde{D}_h > 0) = \frac{\mathbb{E}(\tilde{D}_h|\tilde{D}_h>0)}{\mathbb{V}(\tilde{D}_h|\tilde{D}_h>0)^{\frac{1}{2}}}$ and the mean inter demand interval $\mathbb{E}(p_h)$. Which is in the case of the modelling of Croston $CV(z_h)$ and the mean inter demand interval $\mathbf{E}(p_h)$. A cut-off point along these two dimensions separates fast movers, intermittent demand, erratic demand and lumpy demand.

IDEA OF THE DEMONSTRATION OF THE CUT-OFF VALUES Historically, the forecasting of intermittent demand was percieved as best performed by the Croston method. Therefore the criteria for the two thresholds have been defined in A. A. Syntetos, Boylan, and Croston, 2005 by theoretical outperformances of the SBA versus the SES forecasting method.

The demand D_h follows the model of Croston. $\hat{D}_{h+\tau|h}$ refers to one of the three methods. Each of the methods has an auto-correlation error stemming from the property $\hat{D}_{h+\tau|h} = \hat{D}_{h+1|h}$ for $h \in \mathbb{N}$. The evaluation of the forecast methods must include it. Let L be the number of periods over which the demand is forecasted. The error is then $\sum_{\tau=1}^{L}(D_{h+\tau} - \hat{D}_{h+\tau|h})$. The performance of each prediction method is evaluated via the mean squared error:

$$\text{MSE}_{\text{forecasting method}} = \mathbb{E}([\sum_{\tau=1}^{L}(D_{h+\tau} - \hat{D}_{h+\tau|h})]^2).$$

The following lemma gives the formula of the $\text{MSE}_{\text{forecasting method}}$.

Lemma 1 *The mean squared error verifies :*

$$MSE_{forecasting\ method} = L\{L\mathbb{V}(\hat{D}_{h+1|h}) + L\,BIAS^2 + \mathbb{V}(D_h)\}.$$

Which yields the following results when applied to the different methods.

$$\begin{aligned}\text{MSE}_{\text{SES}} =& L\left\{L\frac{\alpha}{2-\alpha}\left[\frac{p-1}{p^2}\mu^2 + \frac{\sigma^2}{p}\right]\right.\\ &\left.+\left[\frac{p-1}{p^2}\mu^2 + \frac{\sigma^2}{p}\right]\right\}\end{aligned}$$

Croston in Croston, 1972 derived a formula for $\text{MSE}_{\text{CROSTON}}$ that appeared to be false according to numerical application. It was later corrected by A A Syntetos, 2001 yielding

$$\begin{aligned}\text{MSE}_{\text{CROSTON}} \approx L \quad & \left\{L\tfrac{\alpha}{2-\alpha}\left[\tfrac{p(p-1)}{p^4}\left(\mu^2 + \tfrac{\alpha}{2-\alpha}\sigma^2\right) + \tfrac{\sigma^2}{p^2}\right]\right.\\ &\left.+L\left[\tfrac{\alpha}{2-\alpha}\mu\tfrac{(p-1)}{p^2}\right]^2 + \left[\tfrac{p-1}{p^2}\mu^2 + \tfrac{\sigma^2}{p}\right]\right\}.\end{aligned}$$

In A A Syntetos, 2001, the authors also showed that

$$\begin{aligned}&\text{MSE}_{\text{SBA}} \approx L\left\{L\frac{\alpha(2-\alpha)}{4}\left[\frac{(p-1)}{p^3}\right.\right.\\ &\left.\times\left(\mu^2 + \frac{\alpha}{2-\alpha}\sigma^2\right) + \frac{\sigma^2}{p^2}\right]\\ &\left.+L\left[\frac{\alpha}{2}\frac{\mu}{p^2}\right]^2 + \left[\frac{p-1}{p^2}\mu^2 + \frac{\sigma^2}{p}\right]\right\}.\end{aligned}$$

The authors chose to assume that the smoothing constant α is the same for each method which is acceptable.

The proofs can be found in A. A. Syntetos, Boylan, and Croston, 2005, theoretically the method of Croston performs better than SES in term of MSE at least when $p > 1$. A categorization has been derived along the two dimensions. The intermittent demand verifies $CV_q < 0.49$ and $p > 1.32$, Erratic demand $CV_q > 0.49$ and $p < 1.32$, lumpy demand $CV_q > 0.49$ and $p > 1.32$. Croston performs better than Syntetos and Boylan on the smooth demand and worst on the intermittent, erratic and lumpy demand.

Remark 3 *According to the results of A. A. Syntetos, Boylan, and Croston, 2005, the accepted Definition 15 of slow movers is* $\eta = 1 - 1/1.32 \approx 0.24$.

OUR APPROXIMATE VERSUS THEORETICAL THRESHOLD In Section 1 we introduced the boundary of slow movers versus fast movers in terms of turnover versus number of products. We have considered two distributions, each product k has a unique mean frequency of sale μ_k and we can describe the K products as a number of products per mean frequency of sale. We can also describe the turnover per mean frequency of sale as a distribution. The cutoff point along the mean frequency of sale that separates the region of superior performance of turnover versus the number of products was empirically found to be $\theta_k = 0.1$. Assuming that $\tilde{D}_{\nu,emp} = \mathcal{P}(\mu_k)$, the empirical cutoff point $\mathbb{P}(\tilde{D}_{\nu,emp} = 0; \theta_k) = \exp(-0.1) = 0.9$ which is above the threshold $\eta = 0.24$ verifying the slow movers Definition 15.

2.2.1 *Slow movers in industry*

The fast movers have attracted a lot of attention since Brown, 1959. The intermittent demand inventory control literature dates at least back to 1960 concerning engineering parts (Mitchel 1960). These spare parts are generally sold/ used every few months and can be needed in amount that exceeds a unit. In the above categorisation, they are considered as lumpy demand or intermittent. The aerospace and machine maintenance companies are concerned by that problematic. A threshold cited in (Gelders and Van Looy, 1978), a slow moving product has a demand less than 2 units in a year. The management of spare parts is a complex problem that has some overlap with the extreme value theory (Zhu, 2021). Associated issues are the high cost of shortage, high price, low information about the highest quantiles and obsolescence problems. For example, the shortage cost associated to one hour of downtime of an air plane can amount to 8,000 dollars according to Zhu, 2021. Yet, the market of spare parts is large because of the pervasive presence of machines. Aris A Syntetos, M Zied Babai, and Jr, 2015 cites a benchmark of 1.5 trillion dollars for the combined revenue of the largest manufacturing companies and about 26% of it is services. In the Introduction, we pointed out the high proportion of the stock associated to the slow movers which was already noted by Aris A Syntetos, M Zied Babai, and Jr, 2015 with 60% of the stock value associated to it in the spare parts sector.

Time series are less intermittent for durable and non-durable goods as well as for the retail sector. In the case of non durable goods, Albert Heijn is a deutch company with local grocery stores which according to A. G. Kok and Fisher, 2007 reports that a slow moving product is sold less than 10 units a day. This observation is coherent with Remark 2 observing the high disparity of criteria for slow movers. Another type of industry is the one with durable goods such as Adeo. According to Valery Lukinskiy, Vladislav Lukinskiy, and Sokolov, 2020, around 30% to 70% of products in retail and service experience low demand and 90% of the logistic cycle time (from vendor to end customer) is spent in storage and expensive products generally have a low demand. This implies too that there is capital that is immobilised.

2.2.2 *Difficulties associated to forecasts*

Slow movers are important from the business perspective however forecasting and replenishment issues stem from the sparse nature of its demand patterns. In this section, we shed some light on the forecast aspect.

In inventory management, forecasting provides insight about the demand pattern. It provides insights at a mid term about the trends and seasonality. As introduced in Section 2.1.4, demand forecasting can be separated into two related fields the analytical methods such as SES, Croston and probabilistic method which are composed of the estimation of the demand and then according to a specific metric computes the value wanted for the forecast. The assessment of the quality of forecast in the case of slow movers is difficult for two reasons: the relative error of forecasts and the interpretation based on empirical risks.

HIGH RELATIVE ERRORS OF FORECASTS It is unanimous that the slow moving items are difficult to forecast. The underlying problem being the fact that the available information are sparse because of the numerous zero sales. From the stand point of the distribution of the demand we have seen that there is a concentration of the probability that is close to 0. According to M Z Babai, Tsadiras, and Papadopoulos, 2020, it is recurrent that given a specific metric the best value forecasted via a probabilistic method is 0, Example 5 is an illustration. A forecast at 0 is not necessarily interesting for the company from a replenishment perspective, a probabilistic forecast is more valuable.

Example 5 *An example is a Bernouilli process with a probability of* 0.9 *to have 0 demand and* 0.1 *to have a unit demand. The best point forecast is* 0 *according to the MAE (see Example 3). For the forecast value being zero,* $\mathbb{E}[|D_{hk} - 0|] = 0.9|0-0| + 0.1|1-0| = 0.1$ *and the forecast at 1 gives* $\mathbb{E}[|D_k - 1|] = 0.9|0-1| + 0.1|1-1| = 0.9$.

An error of one unit in forecast is high relatively to the mean frequency of sale for slow movers. Considering the large proportion of the slow movers in assortments such as Adeo we conclude that this relative error scales and becomes an issue.

Teunter and Duncan, 2009 considers a real application to spare parts in the Royal Air Force, where the time series have many zeros and few positive demand. In the application the point forecast at 0 was the best forecasting method. More-over in Kolassa, 2020 examples have been provided of point forecast. It is possible to see that the point forecast depends largely on the metric and their relative size are substantial. A difference of 1 for a product that is sold unfrequently is a high relative error.

A critical review Pinçe, Turrini, and Meissner, 2021 including a hundred papers with various applications and methods concluded that no method outperforms the others systematically and the performance depends on the industrial. They also conclude that the measures in the case of slow movers are very important and they often are not appropriate. According to Pinçe, Turrini, and Meissner, 2021 a measure of inventory control would be a better choice. The simple method such as SES compete on real data with more complex one such as neural network or the SBA M Z Babai, Tsadiras, and Papadopoulos, 2020 according to the MASE. Pinçe, Turrini, and Meissner, 2021(2021) provides the avenue for future research. It must be noted that at the extreme which is the case for spare parts, time series can be very short having at the granularity of months about a dozen periods (Valery Lukinskiy, Vladislav Lukinskiy, and Sokolov, 2020).

In the next section we investigate the inventory model for slow movers. In the case of reorder "reorder up to level" and "reorder point", a safety stock is computed for a target service level. This quantity in mono-product optimization of stock is a quantile of the demand distribution. Hence, an estimation procedure is necessary.

CHOICE OF AN APPROPRIATE EMPIRICAL RISK The use for empirical risks can be in view of comparing different forecast techniques but also across different products. For the comparison on the same time series, the most interpretable empirical risk is the MAPE which gives a percentage error however this choice is disabled by the numerous zeros in the time series. The use of the MAE and MSE is a viable option.

When comparing products including slow movers based on a forecast method it has to be scale independent and have a formula with a denominator strictly positive. Hence the MAPE ($\sum_{h=1}^{H} \frac{|D_h - \hat{D}_h|}{D_h}$), and SMAPE ($\frac{100\%}{H} \sum_{h=1}^{H} \frac{|Y_h - \hat{Y}_h|}{(Y_h + \hat{Y}_h)/2}$) are not recommended. An example of candidate formula is the symmetric weighted absolute percentage error (SWAPE)($\frac{\sum_{h=1}^{H} |D_h - \hat{D}_h|}{\sum_{h=1}^{H} |D_h + \hat{D}_h|}$) used at adeo which has a no denominator problem.

2.2.3 *Difficulties associated to replenishment*

2.2.3.1 *Modelling of comparable demand of fast mover and slow movers based on aggregation*

We consider a set of K products whose demand are iid variables D_{hk}. We define as fast mover of order K as the aggregation of K products. Its demand verifies $\tilde{D}_{h,(K)} = \sum_{k=1}^{K} D_{hk}$. In particular $\tilde{D}_{h,(1)}$ follows the same distribution as D_{hk}. In this section, we chose to view a slow mover as being a fast mover of order 1 with suitably chosen properties over the distribution which is an approximation that we believe is acceptable if products are assumed to have a Poisson distribution.

2.2.3.2 *Comparison of replenishment performance between fast movers and slow movers*

For each product, the best replenishment quantity Q is computed in order to optimize the expected profit over a single period. The profit is modeled by the single period mono-product newsvendor model (NVM) from Section 2.1.3.5.

The retailer has to make a compromise between out of stock and excessive stock. All products in our model have the same costs. The revenue per product sold is c_{kp} and its purchasing cost is c_{kc}. The excessive stock cost is denoted as c_{ko} and out of stock cost is c_{ku}. Let $\mathbf{c} = (c_{kp}, c_{kc}, c_k^-, c_k^+)$ be the cost parameter. Since the costs are all the same, we drop the index k. The formula of the profit in the NVM is

$$\rho_{(K)}(D_{(K)}; \theta, Q) = \mathbb{E}\big((c_p - c_c)D_{(K)} - c^-(D_{(K)} - Q_k)^+ - c^+(Q_k - D_{(K)})^+\big).$$

The optimal replenishment becomes

$$Q^*_{(K)} = \operatorname{argmax}_{Q_k \in \mathbb{N}} \rho_{(K)}(D_{(K)}; \theta, Q_k). \tag{10}$$

The space of constraint over the quantities to replenish is $\mathbb{N}$. Since the demand of a fast mover of order K is equal to K fast movers of order 1, we can compare the optimal profit and optimal replenishment quantity of the two. We define the gap of profit between a fast mover of order K and a fast mover of order 1 as $\Delta\rho(K, D_k) =$

$\mathbb{E}[\rho_{(K)}(D_{(K)};\theta,Q)] - K * \mathbb{E}[\rho_{(1)}(D_{(1)};\theta,Q)]$ and the gap of total stock as being $\Delta Q(K,f) = K * Q_{(1)} - Q_{(K)}$.

In Section 2.2.3.3, we show properties of the gap functions in the Gaussian case. In Section 2.2.3.4, we provide properties of the gap functions in the case of demand following a Poisson distribution.

2.2.3.3 *Relaxation of the optimization problem to continuous replenishment quantity and Gaussian demand*

As a first step, relaxing the replenishment to a continuous quantity ($\mathbb{R}$) enables to use a close form for the optimum $Q_{(K)}$ when the demand is normal. Let $D_k \overset{\text{iid}}{\sim} \mathcal{N}(\mu,\sigma)$ with probability mass function f. We define the critical quantile as being $\alpha = \frac{c^-}{c^-+c^+}$. The optimal quantity $Q_{(K)}$ does a compromise between the OOS cost and excessive stock costs and a classic result in the case of continuous distributions yields $Q_{(K)} = F_K^{-1}(\alpha)$. Since $D_{(K)} \sim \mathcal{N}(K\mu, K\sigma^2)$, we have $Q_{(K)} = K\mu + z_\alpha K^{\frac{1}{2}}\sigma$ where $z_\alpha = F^{-1}_{\mathcal{N}(0,1)}(\alpha)$ the cumulative density function (CDF) of the standard normal distribution . We then have that the gap of total stock $\Delta Q(K,f) = z_\alpha \sigma K(1-K^{-\frac{1}{2}})$. The gap between the stocks is expanding linearly in the Gaussian case. We also have that the profit gap is linearly increasing such as in Lemma 2.

Lemma 2 *We have*

$$\mathbb{E}[\rho_{(K)}(D_{(K)};\theta,Q)] = (c_p - c_c)K\mu - (c^- + c^+)K^{\frac{1}{2}}\sigma f_{\mathcal{N}(0,1)}(z_\alpha) \tag{11}$$

and

$$\Delta\rho(K,D_k) = (c^- + c^+)\sigma K[1 - K^{-\frac{1}{2}}]f_{\mathcal{N}(0,1)}(z_\alpha). \tag{12}$$

The proof is provided in Appendix A.2.

2.2.3.4 *Properties of the profit under Poisson demand*

In the Poisson case, $D_k \overset{\text{iid}}{\sim} \mathcal{P}(\lambda)$ with pmf P. We have $D_{(K)} \sim \mathcal{P}(\lambda K)$ and $\mathcal{Q} = \mathbb{N}$. A classic result from the newsvendor literature is that $Q_{(K)} = \inf\{Q \mid \mathbb{P}(D_{(K)} \leqslant Q_{(K)}) \geqslant \frac{c^-}{c^+ + c^-}\}$. In this case in addition to the same conclusion that the profit is optimum if $Q_{(k)}$ is exactly equal to the α quantile, we also have the rounding to an integer value for $Q_{(K)}$ that provides additional costs. We derive bounds in the lemma 3 for the optimal profit. The upper bound is attained when $F_{(K)}(Q_{(K)}) = \alpha$. The lower bound is optimal and describes the case when the quantile of $F_{(K)}(Q_{(K)} - 1)$ is close to α. The closer it gets the bigger the effect of the rounding. We show that these bounds are converging asymptotically to the bound of the Gaussian case and the difference is decreasing in $O((\lambda K)^{-1})$.

Lemma 3 *We have that*

$$\begin{aligned} -(c^- + c^+)Q_{(K)}\mathbb{P}(D_{(K)} = Q_{(K)}) < \mathbb{E}[\rho_{(K)}(D_{(K)};\theta,Q)] - (c_p - c_c)K\lambda \\ \leqslant -(c^+ + c^-)K\lambda\mathbb{P}(D_{(K)} = Q_{(K)}), \end{aligned} \tag{13}$$

with the asymptotic convergence of the two bounds being

$$-(c^- + c^+)(\lambda K)^{\frac{1}{2}} f_{\mathcal{N}(0,1)}(z_\alpha) + O(1). \tag{14}$$

The difference between the two bounds converges to the constant $-(c^{+} + c^{-})z_{\alpha} f_{\mathcal{N}(0,1)}(z_{\alpha})$.
The proof is provided in Appendix A.2.

Conclusion The difference between the inventory quality of a slow mover versus a fast mover evolves in $K^{\frac{1}{2}}$ for the Poisson modelling with an additional rounding effect in the case of discrete demand.

3 ASSORTMENT AND REPLENISHMENT OPTIMIZATION WITH MULTIPRODUCT EFFECTS

3.1 ASSORTMENT PLANNING WITH INFORMAL MULTI-PRODUCT ANALYSIS

We introduced the *mono-product inventory replenishment* in the last chapter which is based on Assumption 1 which states that demands are independent from a product to another. The inventory policies considered were chosen as independent. However the demands in real situation are dependent as some products may be complementary or substitutable. This dependence of demands $\{\tilde{D}_{tk}\}_{k\in[K]}$ with $[K] = \{1, \dots, K\}$ can be used as an additional lever for the optimisation of the stocks. In this section we chose to focus on the substitution by first providing in Section 3.1.1 a version of the NVM introduced in Section 2.1.3.5 including the substitution effects. The profit function to optimise is the function $\rho(\tilde{\mathbf{D}}, \mathbf{V}; \theta, \mathbf{c}, \mathbf{P})$ with either a modelling of the substitution based on the demand $\{\tilde{D}_{tk}\}_{k\in[K]}$ or on the sales $\{\tilde{V}_{tk}\}_{k\in[K]}$.

3.1.1 *Profit function under substitution*

In this section we provide a subcase of the modelling of Section 2.1.2 and more specifically it is an extension to multiproduct of the news-vendor model from Section 2.1.3.5. We consider only one period h and choose for ease of reading to note D_k the demand of the product k. Let the probability of substitution be $\pi_{\ell k}$ between the product ℓ and k. We assume that there is no substitution from one product to himself which is equivalent to $\pi_{\ell\ell} = 0$. The vector of probability of substitution is $\boldsymbol{\pi} = (\boldsymbol{\pi}_{\ell k})_{\ell,k\in[K]}$. The $B_{\ell k}$ is the spill over from product ℓ to product k adopting the convention that $B_{\ell\ell}$ is the quantity of lost sale and that there is no substitution from a product to itself. The probability distribution of the spill over quantity conditionally on the stocks Q_k is multinomial, thus

$$(B_{\ell 1}, \dots, B_{\ell K})' \mid Q_\ell, D_\ell \sim \mathcal{M}\big((D_\ell - Q_\ell)^+, \boldsymbol{\pi}_{\ell 1}, \dots, \pi_{\ell k}\big), \tag{15}$$

where $B_{\ell\ell}$ is the amount that does not substitute.

Let $\mathbf{r} = (r_1, \dots, r_K)^\top$ be the revenue per sale per product. $\mathbf{p} = (p_1, \dots p_K)^\top$ is the purchasing cost of the product. The salvage cost at the end of the period considered is $\mathbf{s} = (s_1, \dots, s_K)^\top$. They verify the assumption $r_k > p_k > s_k \geqslant 0, \forall k \in [K]$ which is that the revenue of the product is higher than the purchasing cost and that the purchasing cost is higher than the salvage cost. These costs are referred to by $\theta = (\mathbf{r}^\top, \mathbf{p}^\top, \mathbf{s}^\top)^\top$. Let $c_{uk} = r_k - p_k$ and $c_{ok} = p_k - s_k$ be respectively the underage cost and the overage cost.

The profit expressed via a newsvendor model is

$$\mathbb{E}[\rho(\mathbf{D}; \boldsymbol{\pi}, \mathbf{Q}, \theta)] = \sum_{k\in[K]} \mathbb{E}\big((r_k - p_k)D_k^s - c_{uk}(D_k^s - Q_k)^+ - c_{ok}(Q_k - D_k^s)^+\big) \tag{16}$$

where $D_k^s = D_k + \sum_{\ell\neq k} B_{\ell k}$ is the sum of the primary demand and the spillover where 's' stands for the accounted substitution effects.

Let $\boldsymbol{\pi}^0$ be the no substitution set of probabilities of substitution hence $\pi^0_{\ell\ell} = 1$ and $\pi^0_{\ell k} = 0, \forall k \neq \ell$. The quantity $B_{\ell k} = 0$ for $k \neq \ell$. Hence we have $D^s_k = D_k$ which results in the following profit expression:

$$\mathbb{E}[\rho(\mathbf{D}; \boldsymbol{\pi}^0, \mathbf{Q}, \theta)] = \sum_{k \in [K]} \mathbb{E}\big((r_k - p_k)D_k - c_{uk}(D_k - Q_k)^+ - c_{ok}(Q_k - D_k)^+\big). \tag{17}$$

It corresponds to the classical newsvendor profit function applied to the mono-product case introduced previously.

3.1.1.1 *Optimization of the profit function*

The objective is to choose the best replenishment quantity $\mathbf{Q}$ in terms of the expected profit. It is expressed by

$$\mathbf{Q}^{s*} = \text{argmax}_{\mathbf{Q} \in \mathbb{N}^K} \mathbb{E}[\rho(\mathbf{D}; \boldsymbol{\pi}, \mathbf{Q}, \theta)]. \tag{18}$$

If we denote by $\mathbf{Q}^{0*}$ the optimal replenishment quantity when there is no substitution we have

$$\mathbf{Q}^{0*} = \text{argmax}_{\mathbf{Q} \in \mathbb{N}^K} \mathbb{E}[\rho(\mathbf{D}; \boldsymbol{\pi}^0, \mathbf{Q}, \theta)]. \tag{19}$$

We then have by definition of Q^*_k and Q^{0*}_k that given specific values of $\boldsymbol{\pi}$

$$\mathbb{E}[\rho(\mathbf{D}; \boldsymbol{\pi}, \mathbf{Q}^{S*}, \theta)] \geqslant \mathbb{E}[\rho(\mathbf{D}; \boldsymbol{\pi}, \mathbf{Q}^{0*}, \theta)]. \tag{20}$$

3.1.1.2 *Illustration of the relative gain for slow movers in a simple scenario*

The inclusion of the substitution when deriving the best quantity to replenish in the case of the newsvendor model for two products is especially interesting for slow movers. We conducted an experiment comparing the ratio of the profits, expressed by

$$R_\rho(\mathbf{D}, \boldsymbol{\pi}, \theta) = \frac{\mathbb{E}[\rho(\mathbf{D}; \boldsymbol{\pi}, \mathbf{Q}^{s*}, \theta)]}{\mathbb{E}[\rho(\mathbf{D}; \boldsymbol{\pi}, \mathbf{Q}^{0*}, \theta)]},$$

across different parameterisations of the distribution of $\mathbf{D}$ and use to illustrate the relative gain as a function of the mean frequencies of sales.

We consider a group of two products equals in term of costs θ, probability of substitution $\boldsymbol{\pi}$ and identically distributed which follows $D_1, D_2 \sim \mathcal{P}(\lambda)$. Let $\alpha = \frac{c_u}{c_u + c_o}$. According to Section 2.1.3.5, $\mathbf{Q}^{0*} = \inf\{\mathbf{Q}; F_D(\mathbf{Q}) >= \alpha\}$. When $F_D(\mathbf{Q}^{0*} - 1) \approx \alpha$ the inclusion of substitution has a rounding effect that is higher for slow movers. We chose to illustrate this fact with $\alpha \in \{0.91, 0.93, 0.95, 0.97, 0.99\}$ which corresponds to classical service level used by retailers and $(p_k, s_k) = (110, 100)$, thus implying that $r_k = (-\alpha s + p)/(1 - \alpha)$. We then searched extensively for the optimal $\mathbf{Q}^*$ and retrieved the ratio. Figure 7 shows the function $R_\rho(\lambda, \alpha)$. Figure 8 shows that the intuition for slow movers is effective in this scenario. The relation between the relative gain and the value α is valuable and the ratio at the peaks is higher for low values of λ hence for slow movers and then decreases.

To understand what precisely happens, we assume that we consider the case where the costs of underage is higher than the overage which means that α is closer to 1. In

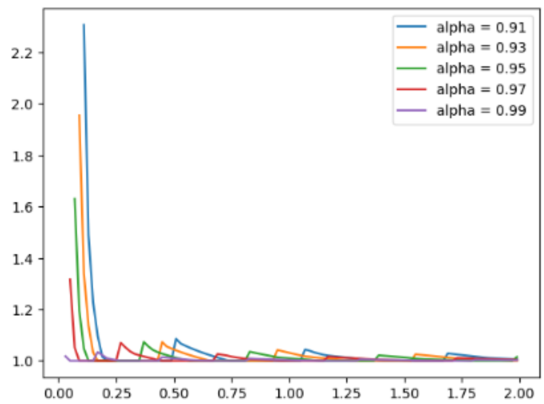


Figure 7: Relative gain of including the substitution.

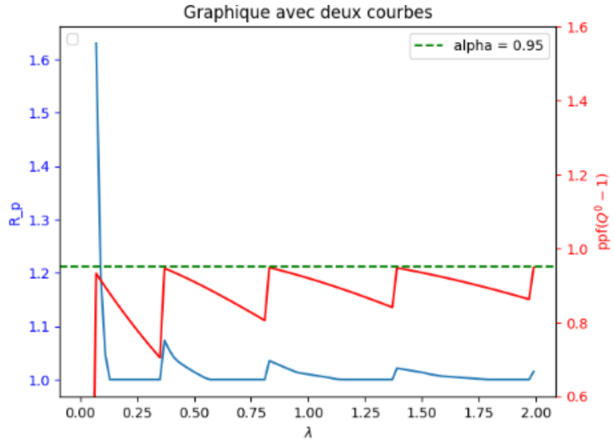


Figure 8: Relative gain of including the substitution.

this case the substitution at the level of a group of substitution reduces the quantity of stock or doesn't modify it. Assuming that replenishing a product at $\mathbf{Q}^0$ is optimal then if $\mathbf{Q}^0 - 1$ is lower and sufficiently close to the percentile α the incentive produced by the substitution induces a reduction of the stock. The closer $\mathbf{Q}^0 - 1$ to the percentile α the bigger the gap between the expected profit evaluated at. The relative gap is bigger for low values of λ.

Let $\{\lambda_1, < \lambda_2 < \dots\}$ be the values of λ such that $F_{D_1}^{-1}(\alpha)$ is an integer. In Figure 8 they are visible where the green line of ordinate α meets the red line which is the quantile associated to the optimal replenishment quantity $F_{D_1}^{-1}(Q^{0*}(\lambda, \alpha)$. Note that since $Q^{0*}(\lambda, \alpha)$ is an integer we have $F_{D_1}(Q^{0*}(\lambda, \alpha)) \leqslant \alpha$. Additionaly $Q^{0*}(\lambda, \alpha)$ is increasing in λ. We refer to the abscissa of downward pikes of the CDF associated to each $Q^{0*}(\lambda, \alpha)$ as $\{\tilde{\lambda}_1 < \tilde{\lambda}_2 < \dots\}$ and conjecture that the profit associated to these pikes is decreasing. For $\epsilon > 0$ small enough, we can define an approximation of these $\tilde{\lambda}$ by $\lambda_i + \epsilon$ $Q^0(\lambda_i + \epsilon, \alpha) = Q^0(\lambda_i, \alpha) + 1$.

At this step, we can thus provide the following conjecture:

Conjecture 1 *For $\epsilon > 0$ sufficiently low, let $\tilde{\lambda}_i(\epsilon) = \lambda_i + \epsilon$. Then for $i < j$ we have* $R_\rho(\tilde{\lambda}_i(\epsilon), \alpha) > R_\rho(\tilde{\lambda_j}(\epsilon), \alpha)$.

3.2 ASSORTMENT AND REPLENISHMENT PLANNING USING MULTIPRODUCT SUBSTITUTION EFFECTS

3.2.1 *Introduction to the concept of substitution*

A client comes with a specific primary demand for an article to a store. If this article is not present, then the client may choose to abandon the purchase, to postpone it in the future or he may substitute for another available article that meets the same needs. It could be an article that is in the same store or a competitor. An exhaustive study of the extent, causes and consumer response of the out of stocks can be found in Gruen, 2002.

The process of substitution gives information about customer behavior but also about the market and the relation between products. Quantifying it yields value that can be leveraged by firms. We have chosen to focus on the subject of estimating the probability of substitution between articles i.e the probability for a client to purchase an article in place of another unavailable.

This PhD Cifre is part of a partnership between the team MODAL at Inria and the group Adeo which is the leader of the DIY market in Europe. For Adeo, the knowledge of the probability of substitution at the scale of the 50,000 articles in store is valuable.

Firstly, currently they have a heuristic for that quantity which is 1/3 lost, 1/3 postponed in the future and 1/3 substitution. But this is very inaccurate and so that can't be used for an optimization.

Secondly, substitution can also be viewed as a similarity score for articles: articles that have the same usage may be highly substitutable. Hence it could be used for the browser of their website in the proposition of similar articles.

Thirdly, the knowledge of the substitution can be precious for having insights about assortment decisions. If in a group, articles are very substitutable then one of them may be deleted saving some fix costs and holding costs for example.

Another example of use is the following. If a client wants a specific article that is not available the customer support may propose the best alternative in term of substitution.

A last example of use is the replenishment, substitution is an additional information that induces a relation between the stocks that can be leveraged in order to improve marginally profits. In the case of Adeo, it could be substantial because due to their business model they have a lot of articles that are sold in small quantities. The definition we adopt here is a product that is sold less than 1 unit per 10 days at the store level. Other definitions can be found in A. A. Syntetos, Boylan, and Croston, 2005. In A. G. Kok and Fisher, 2007 slow movers at Albert Heijn a Deutch grocery company are products sold less than 10 units a day. In addition the quantity sold can also be a source of variability (lumpiness). For Adeo, slow movers represent 25% of the revenue according to our exploratory analysis. These articles are difficult to forecast (Croston, 1972) due to their intermittent nature and sometimes are associated with higher coefficient of variation (Garrett Van Ryzin and Siddharth Mahajan, 1999). This leads to over stock and under-stock that could be improved via the use of substitution in tuning their replenishment.

From the academic point of view, we address a subject that is not recent. The first article that deals with the estimation of substitution can be tracked to Anupindi, Dada, and Gupta, 1998. There is a stream in the operational literature related to this topic. This is because the issues addressed before are common to the retail sector and gave rise to several different approach that we describe in the literature review. We believe that there is still room for improvement. This thesis is part of the field of statistics and probability so we address notions that are not well documented (to our knowledge). The notion of identifiability is not addressed in the literature, nor is the scaling of the estimation method to the whole articles. Exception made for the restriction of the groups at the subcategory granularity. Hence, an objective may be to find substitutes across categories. We want to provide algorithms that are tractable for large numbers of articles and that have good estimator properties such as robustness, convergence.

3.2.2 *Relation of the substitution to the assortment*

The right concept to understand relations between products is the notion of assortment. substitutable articles have a similarity in term of use. The group of articles that are substitutable meets the same need. This is an assumption made by Smith et al., 2000, Salameh et al., 2014. A consequence is that the market shares and their demandss are related. Their demand are also related via the price of the product, the increase of the price of an item increases the demand for variants. Integrating a variant in an assortment may increase the total market share, it could induce cannibalisation and it could introduce recapture of the demand in case of out of stocks. Several models have been proposed for these relations (Wan et al., 2018,Garrett Van Ryzin and Siddharth Mahajan, 1999).

The notion of substitution is linked to the unavailability that a customer could experience. This initial demand is also referred to as the "favourite product","first choice", "primary demand". There is two types of unavailability according to Campo, Gijsbrechts, and Nisol, 2004, a permanent assortment reduction (PAR) which refers to the fact that the store does not carry the variant in the assortment and the out of stock (OOS) substitution which happens when the article is carried in the store's assortment but has no more inventory left.

According to Campo, Gijsbrechts, and Nisol, 2004 these two kinds of unavailability have strong similarities in that customer behavior is close but it also has its differences. The postponement of the article in a PAR is not possible. It can represent around 15% according to Gruen, 2002. Campo, Gijsbrechts, and Nisol, 2004 showed via a survey that customers who would have postpone in the case of OOS are more prone to going to the competitors or abandoning the sale.

Study shows that via the modelling of OOS and the estimation of the probabilities of substitution it is possible to obtain accurate information about the assortments. This is a reason why we focus on OOS substitution.

In Section 3.2.3, we provide the mathematical tools that are used in the literature and in our work. In Section 3.2.4.1, we introduce several modelling and in Section 3.2.5 estimation procedure.

3.2.3 *Mathematical tools*

3.2.3.1 *EM algorithm*

In the case of parametric estimation procedures, the objective is to find the best parameters that fit the data. A classic method is to compute the parameters via a maximum likelihood estimation. But often there are cases such as missing data, for which the likelihood function is intractable.

The EM algorithm proposes an alternative. It enables the estimation by proposing an iterative procedure that re-computes the parameters at each step. The algorithm assures that at each step the likelihood is increasing. When it reaches stationarity the estimators end up being a local maximum of that likelihood. Several runs with random initialization of the algorithm may converge to the global optimum. The core idea is that when the likelihood of the full data has a formula that is tractable, a replacement of the hidden variables by its expectation conditional on the current parameter and the observed variable yields a function that can be maximized.

The EM algorithm was used in a variety of framework before being generalized by A. Dempster, N. Laird, and D. Rubin, 1977. This algorithm is a corner stone for the estimation literature. It has been used in various situation such as missing data, censored data, mixtures, hyperparameter estimation,... All fields benefit from it. A comprehensive literature about the applications and properties can be found in Balakrishnan, Wainwright, and Yu, 2017.

The model

Let $(\mathbf{X}, \mathbf{Z})$ be random variables taking values in the sample space $\mathcal{X} \times \mathcal{Y}$. The joint distribution is $f(\mathbf{X}, \mathbf{Z}; \theta^*)$ that belongs to some parameterized family $\{f(\bullet; \theta) \mid \theta \in \Theta\}$. Θ is a non empty convex set of parameters. $\mathbf{X}$ are observed. $\mathbf{Z}$ are some hidden variables. The marginal distribution of the observable variables is $f(\mathbf{X}; \theta)$. The conditional distribution of the hidden variables given the observed variables is $f(\mathbf{Z}|\mathbf{X}; \theta)$. The functions are related by

$$f(\mathbf{X}, \mathbf{Z}; \theta) = f(\mathbf{X}; \theta) f(\mathbf{Z}|\mathbf{X}; \theta). \tag{21}$$

The objective is to retrieve θ^* that maximizes the observed log-likelihood $l_{obs}(\theta; \mathbf{X}) := \ln f(\mathbf{X}; \theta)$. The framework includes the cases of the i.i.d samples $(\mathbf{X}_i, \mathbf{Z}_i)_{i \in [\![1;N]\!]}$ by substituting $\mathbf{X} = (\mathbf{X}_i)_{i \in [\![1;N]\!]}$, $\mathbf{Z} = (\mathbf{Z}_i)_{i \in [\![1;N]\!]}$ and $f(\mathbf{X}, \mathbf{Z}; \theta) = \prod_{i=1}^{n} f(\mathbf{X}_i, \mathbf{Z}_i; \theta)$.

Let $l_{comp}(\theta; \mathbf{X}) := \ln f(\mathbf{X}, \mathbf{Z}; \theta)$ be the complete log-likelihood. We have the following relationship:

$$l_{obs}(\theta; \mathbf{X}) = l_{comp}(\theta; \mathbf{X}, \mathbf{Z}) - \ln(f(\mathbf{Z}|\mathbf{X}; \theta)). \tag{22}$$

Which implies after taking the expectation conditionally on the observed variables and a value of the parameters θ':

$$l_{obs}(\theta; \mathbf{X}) = \mathbb{E}(l_{comp}(\theta; \mathbf{X}, \mathbf{Z}) \mid \mathbf{X}; \theta') - \mathbb{E}(\ln(f(\mathbf{Z}|\mathbf{X}; \theta)) \mid \mathbf{X}; \theta') \tag{23}$$

$$= G(\theta, \theta') + H(\theta, \theta'). \tag{24}$$

It can be shown that $H(\theta, \theta') \geqslant H(\theta', \theta')$. $G \in \Theta \times \Theta \to \mathbb{R}$. This implies that in order to increase the observed likelihood, we only have to increase $G(\bullet, \theta')$. The EM algorithm is composed of 2 steps at each iteration p, the E step and the M step. At the E step, we retrieve the $G(\bullet, \theta^{(p-1)})$ function by taking the conditional expectation of the complete likelihood. At the M step, we compute parameters $\theta^{(p)}$ that increase $G(\bullet, \theta^{(p-1)})$. Two types of EM algorithm are usually used: the standard EM which retrieves $\theta^{(p)} = \text{argmax}_{\theta \in \Theta} G(\theta, \theta^{(p-1)})$ and the first order EM algorithm such that under certain regularity conditions $\theta^{(p)} = \theta^{(p-1)} + \alpha \nabla G(\bullet, \theta^{(p-1)})(\theta^{(p-1)})$. Some other variations are also possible when G is not tractable but it is not mentioned here.

Algorithm 1 Standard EM algorithm

```
procedure EM
    Input: X
    Initialize θ^(0) ∈ Θ
    p ← 0.
    while Δ > ε do
        Step E: compute G(θ, θ^(p−1))
        Step M: θ^(p) = argmax_{θ∈Θ} G(θ, θ^(p−1))
        Δ = l_obs(θ^(p); X) − l_obs(θ^(p−1); X)
        p ← p + 1
    Output: θ^(p)
```

3.2.4 *Modelling the customer choice and the multiproduct substitution*

In this section we delve in the details of the important concepts. Section 3.2.4.1 deals with modelling of the primary demand. In Section 3.2.4.2 we introduce some generalities about the out of stock. In Section 3.2.4.3 we give the main families of model for the substitution. We introduce in Section 3.2.4.4 the utility framework, in Section 3.2.4.6 the endogenous models, in Section 3.2.4.7 the exogenous models, in Section 3.2.4.8 the different types of substitution patterns and in Section 3.2.4.9 the identifiability of the models in the literature.

3.2.4.1 *Common assumptions on primary demand*

The primary demand for an article is the quantity of articles initially sought by the clients over an arbitrary period of time. If the primary demand meets enough stocks, then it becomes a sale. The EOQ (economic order quantity) model introduced in Section 2.1.3.1, treats the demand as being deterministic (Vandeput, 2020) but generally it is modeled with a random variable that is either continuous (Honhon et al., 2010) or discrete (Smith et al., 2000, Anupindi, Dada, and Gupta, 1998). Laws such as gamma and normal can be used (Vandeput, 2020).

When an estimation of the substitution is needed, all primary demand models are discrete in the literature. It is decomposed in an arrival of customer and a choice among an assortment. The arrivals and choices can be modeled simultaneously with a random law (Anupindi, Dada, and Gupta, 1998). Or the arrival can be modeled and then the choice such as in A. G. Kok and Fisher, 2007, Talluri and Garrett Van Ryzin, 2004. A first model that is very natural for the arrival is the Poisson law (Anupindi, Dada, and Gupta, 1998). Agrawal and Smith, 1996 provides arguments for choosing a negative binomial instead of the Poisson or normal. This model is also used in Smith et al., 2000.

The parameters of the law can be provided additional information such as relative market share, prices, or some other variables such as the temperature and product characteristics (A. G. Kok and Fisher, 2007).

The demand can be sequential (Honhon et al., 2010, Talluri and Garrett Van Ryzin, 2004) in that we have each client successively. Or it can be aggregated over a period of time (Anupindi, Dada, and Gupta, 1998).

It has been shown that the stock out of a product implies an underestimation of the demand by right censoring. Estimating the demand without taking into account the out of stocks of closely related articles with high substitution rate toward the article yields an overestimation (Anupindi, Dada, and Gupta, 1998).

3.2.4.2 *The out of stock concept*

Out of stock is a core idea in our framework in the sense that it provides the opportunity to observe a shift in the demand by the reduction of available articles (i.e the substitution). But the out of stock is also a key point in the replenishment optimization because it has an economic cost for the retailer in terms of retention of the consumers and in term of cost of opportunity. In short, the cost of out of stock is not currently fully understood.

Gruen, 2002 gives a thorough study of the out of stocks for fast moving products in retail encompassing 52 previous studies. In the worldwide market of the retail, 8.4% of the SKU are stocked out. Around 70 to 75% is due to store practices. 47% is due to store ordering and forecasting, 25% in the store and not on shelf and 28% due to upstream causes.

Out of stock has a cost for the consumer. Its response to out of stock are: buy in another store, defer the purchase, substitute in the same brand, substitute by another brand and cancel the purchase. Campo, Gijsbrechts, and Nisol, 2004 investigate via a survey the customer behavior and enumerate the costs for the customer. For example, if the favourite article is not available and the client chooses to cancel there is an opportunity cost associated because he won't be able to fulfill his need. Transaction costs could occur because he has to search for the alternative item. These costs are repercuted on the utility by reducing some utilities associated to the articles available or the other options including the cancellation.

That level of complexity is not translated in the assumptions of the modellings. The literature on the estimation of the probability of substitution do not for instance incorporate the postponement of a sale to our knowledge.

3.2.4.3 *Overview of the response to substitution*

The demand models are parametric in our setting. The literature of customer choices of substitution due to OOS can be partitioned in three main models:

The utility based which treats the substitution in an endogenous manner. There are parameters associated to the utilities and the substitution does not incorporate more parameters meaning that the substitution is related to the products success. Articles related are Talluri and Garrett Van Ryzin, 2004, Wan et al., 2018.

The location models, which assigns a probability based on the distance between the products. That distance is computed based on the product's characteristic (Gaur and Honhon, 2006).

The exogenous models incorporate additional parameters that describe the susbtitution behavior (Anupindi, Dada, and Gupta, 1998, Smith et al., 2000, A. G. Kok and Fisher, 2007, Fisher and Vaidyanathan, 2009, Wan et al., 2018). Our model is related to this stream.

3.2.4.4 *Utility model for the primary demand*

Let $[K] = \{1, \dots, K\}$ be the set of decisions.

Definition 16 *A decision maker* i *associate a utility to decision* $k \in [K]$

$$U_{ik} = V_k + \epsilon_{ik} \tag{25}$$

with V_k *fixed and* $(\epsilon_{ik})_{(i,k)\in[I]\times[K]}$ *is i.i.d random variables following the Gumbell law (extreme value distribution). The probability for a random decision maker to choose decision* k *is* $P_k = \mathbb{P}(U_{ik} > U_{ik'}, \forall k' \neq k)$.

Definition 17 *The Gumbell law is a type I extreme value distribution with probability mass function* $f(\epsilon_{ij}) = e^{-\epsilon_{ij}} e^{e^{-\epsilon_{ij}}}$ *and cumulative distribution function* $F(\epsilon_{ij}) = e^{e^{-\epsilon_{ij}}}$.

Proposition 4 *Let* $\epsilon^*_{ijj'} = \epsilon_{ij} - \epsilon_{ij'}$ *then* $F(\epsilon^*_{ijj'}) = \frac{e^{\epsilon^*_{ijj'}}}{1+e^{\epsilon^*_{ijj'}}}$.

Proposition 5 *The decision maker chooses alternative* k *with probability* $P_k = \frac{e^{V_k}}{\sum_{k'\in[K]} e^{V_{k'}}}$.

Proof 1

$$\begin{aligned} P_k &= \mathbb{P}(V_{ik} + \epsilon_{ik} > V_{ik'} + \epsilon_{ik'}) \\ &= \mathbb{P}(V_{ik} + \epsilon_{ik} - V_{ik'} > \epsilon_{ik'}). \end{aligned}$$

Since $\mathbb{P}(V_{ik} + \epsilon_{ik} - V_{ik'} > \epsilon_{ik'} | \epsilon_{ik}) = \prod_{k\neq k'} e^{e^{-\epsilon_{ik}+V_{ik}-V_{ik'}}}$ *we then have that*

$$P_k = \int (\prod_{k\neq k'} e^{e^{-\epsilon_{ik}+V_{ik}-V_{ik'}}}) e^{-\epsilon_{ik} e^{-\epsilon_{ik}}} d_{\epsilon_{ik}}$$

which yields $P_k = \frac{e^{V_{ik}}}{\sum_{k'\in[K]} e^{V_{ik'}}}$ *after some algebraic manipulations.*

This result will come by often in the modelling literature. Variations of this model are used in A. G. Kok and Fisher, 2007, Garrett Van Ryzin and Siddharth Mahajan, 1999, Musalem et al., 2010, Wan et al., 2018, Talluri and Garrett Van Ryzin, 2004.

In the mean utility it is possible to incorporate some more variables related to product characteristics such as prices of the products, average price in the category, promotions A. G. Kok and Fisher, 2007. The vector of attributes of choice k is z_k (price, indicator variables for product restrictions) β is a vector of weight over these attributes. An example would be that the mean utility is $v_{ik} = \beta^T z_k$.

3.2.4.5 *Endogenous and exogenous models for the first choice*

In the endogenous models there is a set of parameters u_i that characterize the utilities. Once known it yields the probability of choice within an assortment $P_j(N)$. In the case of the exogenous model, the additional parameters are the probability itself $\mathbf{p}$.

3.2.4.6 *Endogenous model*

In Talluri and Garrett Van Ryzin, 2004, Wan et al., 2018, Garrett Van Ryzin and Siddharth Mahajan, 1999 the substitution is endogenous. In Talluri and Garrett Van Ryzin, 2004, for a given assortment $S \subset N$ the probabilities of choosing an article in the assortment becomes

$$P_j(S) = \frac{e^{u_j}}{\sum_{i\in S} e^{u_i} + e^{u_0}}. \tag{26}$$

In Wan et al., 2018, they model the traditional high density small shops. These traditional stores are similar in the sense that they provide the same products. The substitution between couples of store/product (i, j) is modeled. Here i is the store, j is the product. The probability for a customer to choose a product within $\mathcal{J}$ in a given store i conditionally on the availability of the products has the form of a nested logit model.

$$P_{tnij}(\boldsymbol{\nu}, \gamma \mid \mathbf{a}_{tn}) = \frac{a_{tnij} \exp(\nu_{ij}) (\sum_{j' \in \mathcal{J}} a_{tnij'} \exp(\nu_{ij'}))^{\gamma - 1}}{1 + \sum_{i' \in \mathcal{I}} (\sum_{j' \in \mathcal{J}} a_{tni'j'} \exp(\nu_{i'j'}))^{\gamma}} \tag{27}$$

where $a_{tnij} = 1$ if consumer n at time t finds the product i at the store j available. γ is an additional scalar parameter and $\boldsymbol{\nu}$ is a preference weight vector associated to all store/product couple.

A probability of substitution can be computed by:

$$\alpha_{ij \to i'j'} = \frac{p_{i'j'}^{(ij)} - p_{i'j'}}{p_{ij}} \tag{28}$$

with $p_{i'j'}^{(ij)}$ being the probability of choosing i', j' when only product j at store i is unavailable.

Musalem et al., 2010 models customer i's choice of a product j during time period t at store m. A random coefficient model is used. The formula for the utility is another variation of 25:

$$U_{ijtm} = \beta'_{itm} X_{jtm} + \xi_{jtm} + \epsilon_{ijtm}. \tag{29}$$

The individual preferences are captured by β_{itm}, a random multivariate normal variable of mean $\theta' Z_m$ and covariance matrix Σ. X_{jtm} is a time varying vector of the product's characteristics. Here $\boldsymbol{\xi}_{tm} = (\xi_{1tm} \dots \xi_{Jtm})$ is a multivariate Gaussian random variable that accounts for the unobserved market factors. It has zero mean and variance $\sigma_{\xi}^2 I_J$.

The substitution is also endogenous with the following probability of choosing within the available products:

$$P(y_{itm} = j \mid \beta_{itm}, a_{itm}, \xi_{tm}, X) = \frac{a_{itm}^{j} \exp(\beta'_{itm} X_{jtm} + \xi_{jtm})}{1 + \sum_{k \in \mathcal{J}} a_{itm}^{k} \exp(\beta'_{itm} X_{ktm} + \xi_{ktm})}. \tag{30}$$

The limit of this type of model is that it lacks of flexibility: a probability of choice within the available articles cannot be zero. It is a problem because in real settings some substitutions do not account for the popularity of a substitute but for other characteristics such as the size or the price. We describe some of these substitution patterns in Section 3.2.4.8. The substitution probabilities are correlated with the first choice probabilities in the endogenous models.

3.2.4.7 *Exogenous models*

The exogenous models incorporate a set of additional parameters that account for the substitution.

The first article on the estimation of the probability of substitution is Anupindi, Dada, and Gupta, 1998. The context are vending machines that sell cans. There is no modelling of the demand. The focus is entirely on the sales. Each can's sale follows a Poisson law of parameter $\lambda_\bullet$. In the notation of the article, λ_A stands for the mean sales of the article A and $\lambda_{A\bar{B}}$ for the mean frequency of sale when A is available and B is not. These parameters are subject to the following constraints:

$$\lambda_A \leqslant \lambda_{A\bar{B}} \leqslant \lambda_A + \lambda_B$$

and

$$\lambda_B \leqslant \lambda_{A\bar{B}} \leqslant \lambda_A + \lambda_B$$

.

The probability of substitution can be computed from these rates. For example, for an out of stock of only B, the probability of substitution from B to A is $\alpha_{BA} = \frac{\lambda_{A\bar{B}} - \lambda_A}{\lambda_B}$. Each parameter has to be estimated. We discuss the estimation issues in Section 3.2.5.

Smith et al., 2000 propose a modelling with sales following the negative binomial. d_i is item i's demand including substitution effects. f_i is the probability of an incoming customer to have i as favourite product. d_i has then for distribution:

$$\psi(d_i|h_i(x))) = \binom{N + d_i - 1}{N - 1} y_i^N (1 - y_i)^{d_i}$$

with $y_i = \frac{p}{p + h_i(x)(1-p)}$ and $h_i(x) = f_i + \sum_{i \neq j} f_j(1 - x_j)\alpha_{ji}$ and N and p are parameters of the binomial used. Here the substitution matrix $\boldsymbol{\alpha}$ can be of any form accounting for a lot of different substitution patterns.

In A. G. Kok and Fisher, 2007, the assumption is that there is a single substitution attempt. The probability for a client t to purchase a product z depends on the set of products available at that time $S(t)$ and is given by the following formula: $\bar{p}_{z(t),t} + \sum_{k \in \bar{S}(t)} \alpha_{k,z(t)} \bar{p}_{kt}$. $\boldsymbol{\alpha}$ are the substitution's parameters. The probability of first choice $\bar{p}_{z(t),t}$ is utility based. Here $z(t)$ is the first choice of customer t. A. G. Kok and Fisher, 2007 includes parameters that capture the dynamic of arrival of customer in relation to the weather, the holidays, the day or the weekend. It also links the probability of choosing a category and an article in that category to other variables such as promotions and prices.

Wan et al., 2018 propose an exogenous model based on probabilities of first choice $\mathbf{p}$ and of substitution $\boldsymbol{\alpha}$. The indexes are the same as in the endogenous model they proposed in the previous subsection. The probability of purchasing is then given by the following formula:

$$P_{tnij}(\mathbf{p}, \boldsymbol{\alpha} \mid \mathbf{a}_{tn}) = \begin{cases} p_{ij} + \sum_{(i'j') \in C} (1 - a_{tni'j'}) p_{i'j'} \alpha_{i'j' \to ij} & , a_{tnij} = 1 \\ 0 & , a_{tnij} = 0. \end{cases} \tag{31}$$

The difference between this modelling and A. G. Kok and Fisher, 2007 is that this one has first choice that is exogenous whereas the other is endogenous multinomial logit model.

Another approach is proposed by Fisher and Vaidyanathan, 2009. Each article has a vector of categorical characteristics and the probability of substitution is associated to the relation of these vectors in an exogenous manner. Here i is a product, A is the number of attributes and i_a is the level of the attribute a for the product i. The probability of substituting the favourite article i by the article j is given by $\pi_{ij}^s = \prod_{a=1}^{A} \pi_{a i_a j_a}^s$. It is composed of the probabilities $\pi_{a i_a j_a}$ that the customer would substitute the attribute level i_a by j_a. These probabilities are exogenous. f_i is the share of consumers whose favourite product is i. The probability for choosing an article within the assortment S is:

$$F_j(S) = f_j + \sum_{i \notin S, j=j(i,S)} f_i \pi_{ij}. \tag{32}$$

It is more a PAR than an OOS substitution.

3.2.4.8 *Structure of the substitution*

An assumption is generally adopted about substitution: the single attempt assumption states that if the favourite article is out of stock the customer chooses another article and if this one is unavailable then the sale is lost. This is common to Smith et al., 2000, A. G. Kok and Fisher, 2007, Netessine and Rudi, 2003, Wan et al., 2018. A. Kok, 2003 argue that a multi attempt substitution model can be approximated with a single-attempt model by a higher substitution probability.

In Section 3.2.4.6, the formula of substitution in the case of the endogenous model is not flexible. A potential substitute is chosen on the base of its initial success in the assortment. It is in fact proportional to the probability to choose the article if the whole assortment was available.

In practice there could be different substitution patterns. Substitution in the exogenous model can be expressed in term of the substitution matrix $\boldsymbol{\alpha}$. In A. G. Kok and Fisher, 2007, the matrix is proposed with two forms: $\alpha_{kj} = \delta \frac{1}{|N|}$ with random substitution among the variants and with a probability of lost sale associated with $1-\delta$, or $\alpha_{kj} = \delta \frac{d_j}{\sum_{l/k} d_l}$ for the substitution proportional to their primary demand. Other types of matrices are proposed in Caro and Gallien, 2007 and Smith et al., 2000 such as respectively the one item substitution, the random substitution matrix and the adjacent matrix. We provide 3 matrices of substitution where the out of stock product is on the row and the substitute is in the column. The first matrix displayed on the left refers to a super attractor, the second to a uniform substitution (considering lost sale $1-\delta$). The last one is composed of a substitution to two products with the order taken into account which relates to assortment with products order by sizes for example.

$$\begin{pmatrix} 0 & 0 & \delta & 0 & 0 \\ 0 & 0 & \delta & 0 & 0 \\ 0 & 0 & \delta & 0 & 0 \\ 0 & 0 & \delta & 0 & 0 \\ 0 & 0 & \delta & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & \frac{\delta}{4} & \frac{\delta}{4} & \frac{\delta}{4} & \frac{\delta}{4} \\ \frac{\delta}{4} & 0 & \frac{\delta}{4} & \frac{\delta}{4} & \frac{\delta}{4} \\ \frac{\delta}{4} & \frac{\delta}{4} & 0 & \frac{\delta}{4} & \frac{\delta}{4} \\ \frac{\delta}{4} & \frac{\delta}{4} & \frac{\delta}{4} & 0 & \frac{\delta}{4} \\ \frac{\delta}{4} & \frac{\delta}{4} & \frac{\delta}{4} & \frac{\delta}{4} & 0 \end{pmatrix} \begin{pmatrix} 0 & \delta & 0 & 0 & 0 \\ \frac{\delta}{2} & 0 & \frac{\delta}{2} & 0 & 0 \\ 0 & \frac{\delta}{2} & 0 & \frac{\delta}{2} & 0 \\ 0 & 0 & \frac{\delta}{2} & 0 & \frac{\delta}{2} \\ 0 & 0 & 0 & \delta & 0 \end{pmatrix} \tag{33}$$

A constraint on $\boldsymbol{\alpha}$ can be applied but it is not systematic in the literature.

3.2.4.9 *Identifiability of the models*

There are very few remarks about the identifiability in the literature. In Talluri and Garrett Van Ryzin, 2004 they state that their model is not identifiable. In the endogenous model the identifiability is associated to the property of the utility and the MNL model. It can be found in McFadden, 1974, Train, 2009. In the case of the utility because it is an ordinal statistics, a constant and a scale parameter may be added without changing the choice process of the client. Hence normalization are commonly used. In Train, 2009 they state that it is a complex subject.

3.2.5 *Estimation*

We have chosen parametric statistical models for the modelling of our demand. The initial objective of our approach is to quantify the probability of substitution and estimate the parameters related to the primary demand. A general and important remark is that the endogenous models have fewer parameters and they can be estimated over periods when there is no out of stock. The number of parameters typically is the number of products (K). Whereas in the exogenous models, in addition to the parameters linked to the primary demand, there is additional parameters associated to the substitution. These parameters link each couples of articles. There is $K \times (K-1)$ parameters only for the substitution requiring that the right availability/ unavailability configurations must be observed in the input data.

This statement is at the core of the identifiability of the model's parameters. The estimation in the case of the exogenous model can be confronted with two problems: the sparseness of the data associated to stock-out and the complexity in time that increases in relation to the number of availability configurations in the group studied. In Anupindi, Dada, and Gupta, 1998, the mean frequency of sale that is the parameter of the Poisson must be estimated for all the possible configurations. This leads to about $N^2 \times 2^N + N$ parameters which is not tractable. Anupindi, Dada, and Gupta, 1998 observe the fact that only 10% of the machine days have 2 or more product simultaneously out of stock. They provide an approximation that translates in a constraint on the parameters associated to two simultaneous stock out and they ignore higher order of substitution. However A. G. Kok and Fisher, 2007 does not have the problem of increasing time complexity in their algorithm.

The estimation of the parameters require input data. These data in real framework are a product of the systems that tracks the transaction and the stocks. For perpetual review systems the retailer has the time of sales and the stock at that time. Hence they have the configuration of availability at each sale and they also have the time of stock out (Anupindi, Dada, and Gupta, 1998, A. G. Kok and Fisher, 2007). For periodic review systems, the system records at periodic time the stocks and the sales. In this case data are aggregated by article and by period. The time of stock out is not known and the availability configuration of products is not known either (Anupindi, Dada, and Gupta, 1998, Musalem et al., 2010, Wan et al., 2018).

It is important to note that real stocks and system records are not necessarily the same. A classic way to remedy that for the retailer is to check the stock periodically. So even if the review system is perpetual the estimation procedure can include that inaccuracy by assuming that the review system is in fact periodic (Musalem et al., 2010).

Depending on the modelling, the data may not be complete. For instance, in the case of perpetual system review the primary demand is not known. The lost sales are not observed. The data may be incomplete in the case of periodic review systems. The time of stock out is essential to estimate the substitution parameters.

All the papers procedure for optimization rely on methods derived from the maximum likelihood estimation such as Maximum Likelihood Estimate (MLE), algorithm EM and MCMC. The last two are adequate for the missing data framework. The maximum likelihood procedure is used in Anupindi, Dada, and Gupta, 1998, Fisher and

Vaidyanathan, 2009. The algorithm EM is used in Anupindi, Dada, and Gupta, 1998, Talluri and Garrett Van Ryzin, 2004, A. G. Kok and Fisher, 2007 because the complete likelihood is linear in missing variables. The algorithm MCMC is used in Musalem et al., 2010, Wan et al., 2018.

3.2.6 *Optimization of the assortment*

The assortment optimization and the inventory level can be further optimized by incorporating the substitution between products. In the literature, we find that some models optimize the two in a single procedure. In effect, having a proposition of zero stock for a product means that it is not included in the assortment. We separate the models that propose a replenishment quantity with or without the shelf allocation and the related space constraint imposed at higher level.

Most of the models including stochastic demand are based in the newsvendor problem.

In Hubner, 2017, they solve a problem of assortment based on the NVM, shelf constraint and substitution behaviors. It is called capacitated assortment and shelf problem (CASP). The parameters related to the products are their price of purchase, selling price, salvage cost and shortage cost. The benchmark method for assessing the performance is called a sequential planning (SP): given a set of article they retrieve the optimal quantity based on the quantile solution of the NVM, they round it to an integer, then they rank each products according to profit $p_i - c_i$ and they allocate shelf in the rank order until there is no more shelf space. The problem is a mixed integer non linear problem and the number of possible solution is $\binom{N}{S}$. They provide an analysis of the effect of the level of substitution on the profit gain.

4 SUBSTITUABILITY MODELLING BASED ON SALES AND OUT-OF-STOCKS

4.1 INTRODUCTION

In this chapter, we build upon Chapter 2 and Chapter 3 which have reinterpreted the initial optimised replenishment task as a new products substitution task. Indeed, we have seen that the additional information that substitution provides can be included in the replenishment optimisation in order to yield more profits. The novel question is thus now to propose a valuable substitution model, in conjunction with an efficient estimation process.In Section 4.2, we introduce a general modelling of the substitution first based on the trajectory of the demand and sales for a given assortment of products. This model integrates the continuity of time and its indexing of the incoming customers in order to provide the variable of demand and sales. We provide also an illustration of the potential gains using the Newsvendor model introduced in Section 2.1.3.5. In Section 4.3, we proposed however a simpler modelling of substitution relying on an aggregation of the demand and of the sales at the availability configurations level. This choice facilitates in particular the proof of a nice property of identifiability of the model. Retaining definitively this latter model, in Section 4.4, we detail the associated estimation strategy based on an EM algorithm. Then, we provide in Section 4.5 a quite dense numerical study based on simulated datasets to illustrate the good properties, and also the limits, of our proposed model. The last section (Section 4.6) refers to the application of our estimation algorithm on two real datasets from the Adeo company, revealing its promising availability for detecting substituability of products in a real situation..

4.2 AN INTRODUCTIVE MODEL FOR PRODUCT SUBSTITUABILITY

4.2.1 *Model presentation*

Building on the notations from Section 2.1.2, we focus on a period of length n during which, we suppose that no replenishment is made. Customer ξ may have a demand for a product. We note $\xi \in \{1, \dots, T\}$ the customer and t_ξ its time of arrival during the period. Importantly also, T customers arrive sequentially with a (unique) initial unit demand $\tilde{\mathbf{D}}_{t_\xi} = (D_{t_\xi 1}, \dots, D_{t_\xi K})^\top$ for a single item, where $D_{t_\xi k} = 1$ if customer ξ has an initial demand for product k and $D_{t_\xi k} = 0$ otherwise, with $\sum_{k=1}^{K} D_{t_\xi k} = 1$. Since the initial demand of customers is not always satisfied (*i.e.*, the product is unavailable), a customer may switch to another product or make no purchase, resulting in the "trajectory" of demands $\mathbf{D} = (\tilde{\mathbf{D}}_t)_{t\in[0,n]}$ and $\tilde{\mathbf{D}}_t = (D_{t1}, \dots, D_{tK})$. It is important to note that the latter (the demand trajectory) is not practically observable. We also denote $\mathbf{V} = (\mathbf{V}_t)_{t\in[0,n]}$ as the "sales trajectory," and $\mathbf{V}_{t_\xi} = (V_{t_\xi 0}, V_{t_\xi 1}, \dots, V_{t_\xi K})^\top$ with $V_{t_\xi 0} = 1$ if customer ξ does not buy any product and $V_{t_\xi k} = 1$ if customer ξ purchases product k, and $\sum_{k=0}^{K} V_{t_\xi k} = 1$. Since the initial demands are not always satisfied, we generally have $\mathbf{V} \neq \mathbf{D}$. Furthermore, only the stock evolutions are observed. The order of sales made for $k \geqslant 1$ is observed, unlike T or the indices ξ of customers who did not make any purchases.

Here, we consider that a customer's initial demand remains unsatisfied only if the product is out-of-stock. Given an initial stock $\mathbf{S} = (S_1, \dots, S_K)^\top$, where $S_k \in \mathbb{N}$ is the stock of product k at the beginning of the period of interest, the vector $\mathbf{u}(t_\xi) = (u_1(t_\xi), \dots, u_K(t_\xi))^\top$ indicates the products still accessible to customer ξ when he arrives in the store. Specifically, $u_k(t_\xi) = 1$ if product k is accessible to customer ξ (*i.e.*, $S_k - \sum_{\xi'=1}^{\xi-1} V_{t_{\xi'}k} \geqslant 1$, with $\sum_{\xi=1}^{0} V_{t_\xi k} = 0$), and $u_k(t_\xi) = 0$ otherwise (*i.e.*, $S_k = \sum_{\xi'=1}^{\xi-1} V_{t_{\xi'},k}$). Therefore, for customer ξ with an initial demand for product k, if the product is in stock, the customer purchases it (*i.e.*, $V_{t_\xi k} = 1$ if $D_{t_\xi k} = 1$ and $u_k(t_\xi) = 1$), while if it is not in stock, a substitution mechanism dependent on the initial demand and the available products is implemented. Since the sale $\mathbf{V}_{t_\xi}$ is only dependent upon the initial demand of the client $\tilde{\mathbf{D}}_{t_\xi}$, the availability of the products $\mathbf{u}(t_\xi)$ and not the past, we get that the distribution of $\mathbf{V}_{t_\xi} \mid \mathcal{F}_{t_\xi - 1}, \tilde{\mathbf{D}}_{t_\xi}, \mathbf{u}(t_\xi)$ is equal to $\mathbf{V}_{t_\xi} \mid \tilde{\mathbf{D}}_{t_\xi}, \mathbf{u}(t_\xi)$, where $\mathcal{F}_{t_\xi - 1}$ is the natural filtration. This mechanism results in the sale of another product (substitution) or no sale at all (the customer leaves without any product). The customer will then switch to product $k \in [\![0, K]\!]$ (the case $k = 0$ indicates a lost sale) with a probability $\lambda_{k\ell}(\mathbf{u}(t_\xi); \boldsymbol{\pi})$, where for all $\ell \in [\![1, K]\!]$, $\sum_{k=0}^{K} \lambda_{k\ell}(\mathbf{u}(t_\xi); \boldsymbol{\pi}) = 1$. The sales switching probabilities $\lambda_{k\ell}(\mathbf{u}(t_\xi); \boldsymbol{\pi})$ depends on the configuration of availability $\mathbf{u}(t_\xi)$ and on $\boldsymbol{\pi} \in \mathbb{R}^{K^2}$. $\boldsymbol{\pi}$ defines the probabilities of substituting product k with other products unconditionally, regardless of stock levels and product availability (as determined by $\mathbf{u}(t_\xi)$) Thus, the conditional distribution of sales, given the initial demand and stock, is defined as follows:

$$\forall k \in [\![1, K]\!],\ \mathbb{P}(V_{t_\xi k} = 1 \mid \tilde{\mathbf{D}}_{t_\xi}, \mathbf{u}(t_\xi)) = \mathbf{u}_k(t_\xi) D_{t_\xi k} + \sum_{\ell=1}^{K} (1 - \mathbf{u}_\ell(t_\xi)) D_{t_\xi \ell} \lambda_{k\ell}(\mathbf{u}(t_\xi), \boldsymbol{\pi}_\ell)$$

and

$$\mathbb{P}(V_{t_\xi 0} = 1 \mid \tilde{\mathbf{D}}_{t_\xi}, \mathbf{u}(t_\xi)) = \sum_{\ell=1}^{K} (1 - u_\ell(t_\xi)) D_{t_\xi \ell} \lambda_{0\ell}(\mathbf{u}(t_\xi); \boldsymbol{\pi}_\ell).$$

The profit generated over the time period is determined by the function $\rho(\mathbf{D}, \mathbf{V}; \theta, \mathbf{S})$ where θ groups the store-specific parameters (purchase and storage costs, selling prices, customer dissatisfaction costs when the initial demand is not met, etc.; see S. Mahajan and G. Van Ryzin, 2001 for examples). It's important to note that for a fixed initial stock, the profit obtained at the end of the period under consideration, $\rho(\mathbf{D}, \mathbf{V}; \theta, \mathbf{S})$, is a random variable. In a stock management framework, the goal is to determine the optimal initial stock $\mathbf{S}_{\boldsymbol{\pi}}^{\star}$ that maximizes a certain statistical measure $m_{\eta,\pi}$ (e.g., expectation, q-quantile, etc.) of $\rho(\mathbf{D}, \mathbf{V}; \theta, \mathbf{S})$. This measure depends on the distribution of initial demands defined by η and of the product substituability defined by $\boldsymbol{\pi} = (\boldsymbol{\pi}_1^\top, \dots, \boldsymbol{\pi}_K^\top)^\top$. Therefore, the objective is to determine the optimal value

$$\mathbf{S}_{\boldsymbol{\pi}}^{\star} = \operatorname{argmax}_{\mathbf{S}} m_{\eta,\pi} \left[\rho(\mathbf{D}, \mathbf{V}; \theta, \mathbf{S})\right].$$

4.2.2 *Numerical illustration of the model's benefits*

INSTANTIATION OF PROFIT In retail, it is common (S. Mahajan and G. Van Ryzin, 2001) to consider various costs, such as purchase cost, selling cost, salvage cost, holding cost per unit and per period, and stockout cost. We will denote these different costs as c_p, c_k, s_k, c_{kh}, and c_{ok}, respectively, and group them in the parameter vector $\mathbf{c} = (c_{kp}, c_{kc}, c_{ks}, c_{kh}, c_{ko})_{k\in[K]}$.

This parameter will influence the profit. $\rho_t(\tilde{\mathbf{D}}_t, \mathbf{V}_t; \theta)$ is the "instantaneous" gain/loss associated to a demand that occurs at time t. If there is no demand at time t it implies that there is no gain/loss hence $\rho_t(\tilde{\mathbf{D}}_t, \mathbf{V}_t; \theta) = 0$. For customer ξ

$$\rho_{t_\xi}(\tilde{\mathbf{D}}_{t_\xi}, \mathbf{V}_{t_\xi}; \theta) = \sum_{k=1}^{K} \mathbb{1}_{D_{t_\xi k}=1} \left[\sum_{\ell=1}^{K} c_{pk} V_{t_\xi \ell} - c_{ok} V_{t_\xi 0} \right].$$

Taking into consideration that either a sale is made with an associated profit or the sale is lost with the resulting financial penalty, this process is repeated for each customer ξ. At the end, there is a certain amount of remaining stock for product k, denoted as S_k^{end} (this quantity is deduced from the entire sales process), which incurs a holding cost minus the salvage value. Also considering the purchase cost incurred by replenishment at the beginning of the period, we then obtain the overall profit across all customers associated with the "trajectory" $(\mathbf{D}, \mathbf{V})$ of customer demands and actual sales:

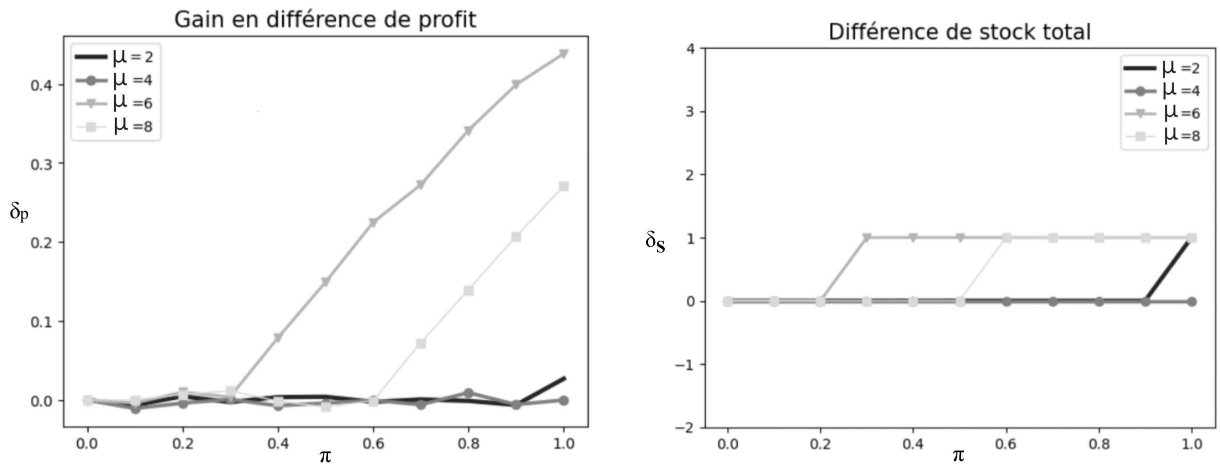
$$\rho(\mathbf{D}, \mathbf{V}; \theta, \mathbf{S}) = \left\{ \int_0^n \rho_t(\tilde{\mathbf{D}}_t, \mathbf{V}_t; \theta) d_\delta(t) - \sum_{k=1}^{K} (c_{kh} - c_{ks}) S_k^{\text{end}} - \sum_{k=1}^{K} \omega_k S_k \right\}$$

where $d_\delta(t)$ is the Dirac measure. $\rho(\mathbf{D}, \mathbf{V}; \theta, \mathbf{S})$ builds upon $\rho_t(\tilde{\mathbf{D}}_t, \mathbf{V}_t; \theta)$. The benefit of the proposed substitution model, in contrast to a naive method that does not account for substitutions even when they occur in reality, can be assessed through the following two key metrics. Firstly, the profit gain, denoted as $\delta_\rho(\boldsymbol{\pi}) = m_{\eta,\pi}[\rho(\mathbf{D}, \mathbf{V}; \theta, \mathbf{S}^\star \boldsymbol{\pi})] - m_{\eta,\pi}[\rho(\mathbf{D}, \mathbf{V}; \theta, \mathbf{S}_0^\star)]$. Secondly, the difference in initial replenishment, denoted as $\delta_S(\boldsymbol{\pi}) = \sum_{k=1}^{K} [(\mathbf{S}_\pi^\star)_k - (\mathbf{S}_0^\star)_k]$.

Note that in the following, we will consider the mean for the statistical measure $m_{\eta,\pi}$. Furthermore, the calculation and optimization of the profit $m_{\eta,\pi}[\rho(\mathbf{D}, \mathbf{V}; \theta, \mathbf{S})]$ are generally far from being straightforward. In this preliminary work, we will content ourselves with approximating this quantity using a simple Monte Carlo method (with 10^5 samples) and then optimizing it using a standard general-purpose procedure available in R.

SCENARIO 1: PRODUCTS WITH LOW DEPRECIATION. In the first scenario, the values (in €) have been set to represent a scenario similar to those found in the company Adeo, specializing in the home market: $c_p = 15$, $c_k = 10$, $s_k = 9$, $c_{kh} = 0.05$ and $c_{ok} = 1$. It can be observed that the stockout cost has a more significant impact than the holding cost. The results of the numerical application for two products can be found in Figure 9. The left sub-figure illustrates that the profit gain δ_ρ, at a fixed mean frequency of demand here denoted by the coefficient μ, consistently increases as the substitution probability $\pi_1 = \pi$ ($\pi \in [0,1]$) increases. The right sub-figure (related to δ_S) indicates that this profit gain is achieved through slight understocking.

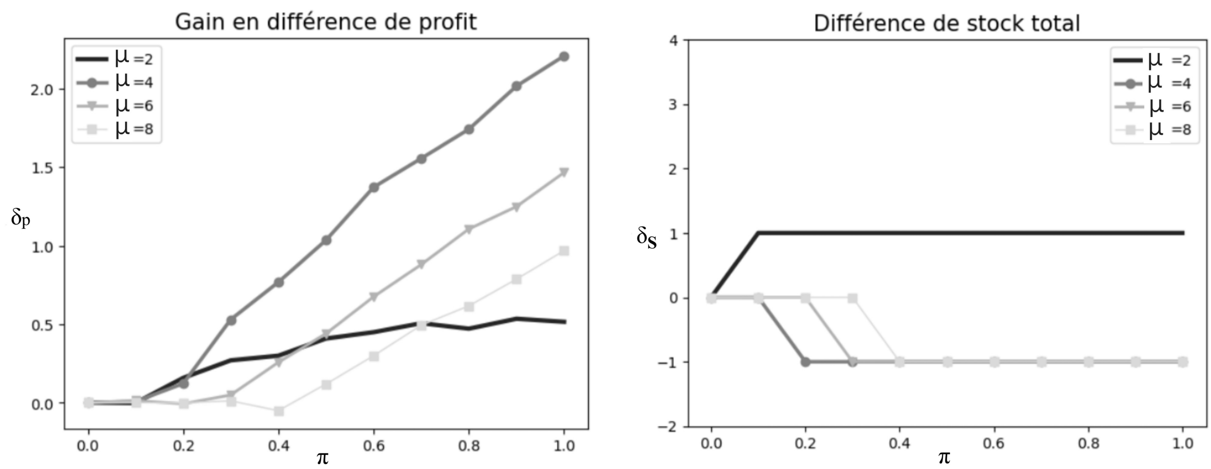
SCENARIO 2: PRODUCTS WITH HIGH DEPRECIATION. The second scenario is associated with perishable goods, which are quite different from the home market described in scenario 1. In this scenario, there is a significant product depreciation over the period, resulting in a substantial loss of stock value. To represent this situation, the only difference from the previous scenario is that $c_s = 0$. The results of this new numerical application can be found in Figure 10. Similar to the previous experiment,



(a) Gain in profit (b) Difference of replenishment quantity.

Figure 9: Case of low devaluation (Scenario 2).

with fixed μ, the left sub-figure shows an increase in profit gain as a function of π. The right sub-figure indicates that this gain is achieved by either slightly increasing or decreasing the stock.



(a) Gain in profit. (b) Difference of replenishment quantity.

Figure 10: Case of high devaluation (Scenario 1).

4.3 THE PROPOSED SUBSTITUTION MODEL

The previous model was just a firt attempt for modelling product substituability. However, it was quite complex to implement and is was also challenging to obtain some identifiability properties. Indeed, we expect that identifiability is a inescapable task since the available data is poor, because limited to the availability of the configurations. In addition, concerning model complexity, we expect to limit it by simplifying the global approach with the sequence of customers and their time of arrival by replacing with aggregated sales and demands over subperiods of $[0, n]$ with same availability configurations of products. Additionally, we will see that such an aggregation strategy opens the door for an algorithm that will have an acceptable execution time.

4.3.1 *A new sale model considering product substitutability*

We still consider a set of K products. At time t, the availability of the products is defined by the binary vector $\mathbf{u}(t) = (u_1(t), \ldots, u_K(t))^\top$ introduced in Section 4.2. There are $J = 2^K$ possible configurations of product availability where the configuration of product availability j is defined by the binary vector $\boldsymbol{\omega}_j = (\omega_{j1}, \ldots, \omega_{jK})^\top$ such that $\omega_{jk} = 1$ if the stock of product k is not zero and $\omega_{jk} = 0$ if the stock of product k is zero, for $j = 1, \ldots, J$. For instance, the configuration of availability $\boldsymbol{\omega}_j = (0, 1)^\top$ for $K = 2$ products refers to the case where product 1 is out-of-stock when product 2 is available. We study the sales during a period of time of length $n > 0$ and we denote by $\mathbf{X}_n = (X_{n1}, \ldots, X_{nJ})^\top$ the times spent in each configuration, where $X_{nj} = \int_0^n \mathbb{1}_{\{\boldsymbol{\omega}_j = \mathbf{u}(t)\}} dt$ denotes the time spent in configuration of product availability j, for $j = 1, \ldots, J$. Note that some configurations could be not observed during the period of study, so X_{nj} can be equal to zero for some j and, by construction, $\sum_{j=1}^J X_{nj} = n$.

Example 6 *We provide an example of a situation where $K = 2$. The time spent in the first configuration $\boldsymbol{\omega}_1 = (1, 1)$ where the two products are available is ten days ($X_{n1} = 10$). The time for the case where the first product is unavailable, i.e. $\boldsymbol{\omega}_2 = (0, 1)$, is 20 days $X_{n2} = 20$. The total time is thus $n = 30$.*

We consider that each consumer comes with an initial demand for a single product. The model assumes that the initial demands of all the products follow independent univariate homogeneous Poisson processes where $\mu_k > 0$ denotes the intensity of the Poisson process modelling the initial demand of product k. If the product is available then the consumer buys it. If the product is not available, then the consumer can randomly decide buying another product among the available products or leaving the shop without buying anything. For any $k \in \{1, \ldots, K\}$, we denote by $Z_{njk\ell} \in \mathbb{N}$ the number of products k sales despite the fact that the initial demands of the consumers were for product ℓ, under the configuration $\boldsymbol{\omega}_j$ and during the time period of length n. Moreover, with the notation $k = 0$, $Z_{nj0\ell} \in \mathbb{N}$ denotes the number of lost sales for product ℓ because product ℓ is not available. From Section 4.2 we have that $Z_{njk\ell} = \int_0^n V_{tk} D_{t\ell} \mathbb{1}_{\boldsymbol{\omega}_j = \mathbf{u}(t)} d_\delta(t)$ with d_δ being the dirac measure. The previous assumption of independent initial demands leads to $Z_{njk\ell}$ being conditionally independent given the times of configurations X_{nj} such that

$$P(\mathbf{Z}_n \mid \mathbf{X}_n) = \prod_{j=1}^{J} \prod_{k=0}^{K} \prod_{\ell=1}^{K} P(Z_{njk\ell} \mid X_{nj}),$$

where $\mathbf{Z}_n = (\mathbf{Z}_{n1}^\top, \ldots, \mathbf{Z}_{nJ}^\top)^\top$, $\mathbf{Z}_{nj} = (\mathbf{Z}_{nj1}^\top, \ldots, \mathbf{Z}_{njK}^\top)^\top$ and $\mathbf{Z}_{njk} = (Z_{njk1}, \ldots, Z_{njkK})^\top$. The random variable $Z_{njk\ell} \mid X_{nj}$ follows a Poisson distribution $\mathcal{P}(X_{nj}\lambda_{k\ell}(\boldsymbol{\omega}_j; \boldsymbol{\theta}))$ with $\lambda_{k\ell}(\boldsymbol{\omega}_j; \boldsymbol{\theta})$ describing the substitution rate from ℓ to k under the availability configuration $\boldsymbol{\omega}_j$ and depending on the parameters $\boldsymbol{\theta}$ that we will describe in detail later.

Example 7 *Following Example 6, one substitution from product 1 to product 2 in the configuration $\boldsymbol{\omega}_2$ is noted $Z_{n221} = 1$. In this configuration, there is no substitution from product 2 to 1 which is denoted by $Z_{n212} = 0$. The number of sales of product 2 stemming from its initial demand is $Z_{n222} = 1$.*

The variable $\mathbf{Z}_n$ is a latent variable since we only have access to the number of sales of each product under each configuration, and thus the initial demands of the consumers are generally unknown. Therefore, we denote by $\mathbf{Y}_n = (\mathbf{Y}_{n1}^\top, \dots, \mathbf{Y}_{nJ}^\top)^\top \in \mathbb{N}^p$, with $p = JK$, the number of sales (whatever be the product) for each configuration where $\mathbf{Y}_{nj} = (Y_{nj1}, \dots, Y_{njK})^\top$ indicates the number of sales under configuration j and $Y_{njk} = \sum_{\ell=1}^{J} Z_{njk\ell}$ is the number of sales of product k under configuration j. From Section 4.2, we have that $Y_{njk} = \int_0^n V_{tk}\mathbb{1}_{\omega_j = \mathbf{u}(t)} d_\delta(t)$. Note that, for any configuration j, the number of people leaving the shop without buying anything (*i.e.*, $\sum_{\ell=1}^{K} Z_{nj0\ell}$) is not observed. The observed variables $\mathbf{Y}_{nj}$ are independent given the time spent in each configuration, leading to

$$P(\mathbf{Y}_n \mid \mathbf{X}_n) = \prod_{k=1}^{K}\prod_{j=1}^{J} P(Y_{njk} \mid X_{nj}), \tag{34}$$

where each Y_{njk} given X_{nj} follows a homogeneous Poisson distribution $\mathcal{P}(X_{nj}\lambda_k(\boldsymbol{\omega}_j;\boldsymbol{\theta}))$ and $\lambda_k(\boldsymbol{\omega}_j;\boldsymbol{\theta}) = \sum_{\ell=1}^{K} \lambda_{k\ell}(\boldsymbol{\omega}_j;\boldsymbol{\theta})$. Thus, the probability mass function of $\mathbf{Y}_n$ given $\mathbf{X}_n$ is

$$f(\mathbf{y}_n \mid \mathbf{x}_n;\boldsymbol{\theta}) = \prod_{J=1}^{J}\prod_{k=1}^{K} p(y_{njk}; x_{nj}\lambda_k(\boldsymbol{\omega}_j;\boldsymbol{\theta})), \tag{35}$$

where $p(.;\tau)$ is the probability mass function of a Poisson distribution with parameter τ (when $\tau = 0$ this corresponds to a Dirac distribution in zero).

Example 8 *Following the previous example, the number of sales of product 2 in configuration* $j = 2$ *is* $Y_{n22} = 2$.

4.3.2 *A first modelling of substitution*

4.3.2.1 *The model*

This first modelling assumes that the demand for a product that is not available is transferred both on available products and the abandon of the sales in proportion to their respective probability of substitution. For any $\ell \in \{1, \dots, K\}$ the substitution rate in the configuration $\boldsymbol{\omega}_j$ from ℓ to k is expressed as

$$\lambda_{k\ell}(\boldsymbol{\omega}_j;\boldsymbol{\theta}) = \begin{cases} \mu_\ell \omega_{j\ell} & \text{if } k = \ell, \\ \mu_\ell(1-\omega_{j\ell})\beta_{\ell k}(1+\boldsymbol{\omega}_j^\top \boldsymbol{\beta}_\ell)^{-1}\omega_{jk} & \text{if } k \neq \ell,\ k \neq 0\,, \\ \mu_\ell(1-\omega_{j\ell})(1+\boldsymbol{\omega}_j^\top \boldsymbol{\beta}_\ell)^{-1} & \text{if } k = 0 \end{cases} \tag{36}$$

where $\boldsymbol{\theta}$ groups all the parameters, $\boldsymbol{\beta}_\ell = (\beta_{\ell 1}, \dots, \beta_{\ell K})^\top$ is the vector of probabilities of sales reporting from ℓ to the other products such that $\beta_{\ell k}$ is a parameter associated to the probability of buying product k, when it is available, with an initial demand of product ℓ. Here $\beta_{\ell\ell} = 0$. Note that when other products are not available the consumer cannot consider them for its sale reporting. Detailing now the parameter $\boldsymbol{\theta}$, we have configuration $\boldsymbol{\omega}_j$, the probability of leaving without buying anything, when a consumer comes with an initial demand for the product ℓ that is not available (*i.e.*, $\omega_{j\ell} = 0$) is $(1+\boldsymbol{\omega}_j^\top \boldsymbol{\beta}_\ell)^{-1}$. Thus, $\boldsymbol{\theta} = (\mu_1, \dots, \mu_K, \boldsymbol{\beta}_1^\top, \dots, \boldsymbol{\beta}_K^\top)^\top \in \Theta$ where $\Theta = (\mathbb{R}^{+*})^K \times (H_\ell)_{\ell \in [K]}$ where $H_\ell = \{\boldsymbol{\beta}_\ell \in (\mathbb{R}^+)^K : \beta_{\ell\ell} = 0\}$.

4.3.2.2 *Identifiability property of the model*

Model 36 has $K + K(K-1)$ free parameters. A natural question is to state sufficient conditions that ensure the identifiability of the parameters given $\mathbf{X}_n$. Let $J_n = \{j \in \{1, \ldots, J\} : X_{nj} > 0\}$.

Assumption 4 *(Poisson intensity) For $k = 1, \ldots, K$, then $0 < \mu_k < \infty$.*

Assumption 5 *(The configuration where every product is available is observed) let $\boldsymbol{\omega}_{\tilde{j}} = \mathbf{1}_K$, $\tilde{j} \in J_n$.*

Assumption 6 *(All the configurations of availabilities where a unique product is unavailable are observed) i.e.: Let j verifies $\exists \ell \in [K] = \{1, \ldots, K\}$, $\forall k \neq \ell$, $\omega_{j\ell} = 0$ and $\omega_{jk} = 1$ then $j \in J_n$.*

Assumption 7 *(A subset of configurations of availability) Let $H \leqslant K$ and $\boldsymbol{\omega}_{j_h}$ be an availability configuration so that $\omega_{j_h k} = 0$ if $k < h$ and $\omega_{j_h k} = 1$ if $k \geqslant h$. Then for $h \leqslant H$, $j_h \in J_n$.*

Remark 4 *The identifiability of the parameters are related to the observed configurations of availability where observed means that $x_{nj} > 0$ for the specific configurations. Assumption 4 assumes that each product has a strictly positive demand ratio which is natural. Assumption 5 and 5 assume that the configuration where every product is available and configurations where only one product is unavailable are observed. Assumption 7 for $H \leqslant K-2$ may be visualized by a matrix*

$$\begin{bmatrix} \boldsymbol{\omega}_{j1}^\top \\ \boldsymbol{\omega}_{j2}^\top \\ \vdots \\ \boldsymbol{\omega}_{jK-2}^\top \end{bmatrix} = \begin{bmatrix} 1 & 1 & \ldots & 1 & 1 \\ 0 & 1 & \ldots & 1 & 1 \\ 0 & 0 & \ddots & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

Assumption 5 is restrictive in the sense that among the 2^K possible observed configurations of availability only K verifies it. Assumption 7 is less restrictive because it can be verified on more subset of the 2^K possible configurations.

Proposition 6

1. *Assumptions 4, 5, and 6 are sufficient conditions for the parameters of the model to be identifiable.*
2. *Assumptions 4 and 7 are sufficient conditions for the parameters $\beta_{\ell k}$ such that $1 \leqslant k \leqslant H$ and $\ell \geqslant k$. to be identifiable.*

Proofs can be found in Appendix B.1.

Remark 5 *Assumption 4 and 7 may often yield a partial identifiability of the parameters.*

However, the conditions are only sufficient conditions. Consequently, we propose another modelling of the substitution for which we are able to derive necessary and sufficient conditions. **It will lead to our most accomplished, thus final, model.**

4.3.3 *A second modelling of substitution*

4.3.3.1 *The model*

This second modelling is a little bit simpler than the previous one and thus allows to obtain finer theoretical properties. It assumes that the demand for a product that is not available is entirely transfered to the no sale case. For any $\ell \in \{1, \dots, K\}$ the substitution rate in the configuration $\boldsymbol{\omega}_j$ from ℓ to k is expressed as

$$\lambda_{k\ell}(\boldsymbol{\omega}_j; \theta) = \begin{cases} \mu_\ell \omega_{j\ell} & \text{if } k = \ell, \\ \mu_\ell (1 - \omega_{j\ell}) \pi_{\ell k} \omega_{jk} & \text{if } k \neq \ell,\ k \neq 0\,, \\ \mu_\ell (1 - \omega_{j\ell}) \boldsymbol{\pi}_\ell^\top (\mathbf{1}_K - \boldsymbol{\omega}_j) & \text{if } k = 0 \end{cases} \tag{37}$$

where $\mathbf{1}_K$ is the vector of ones of length K, θ groups all the parameters, $\boldsymbol{\pi}_\ell = (\pi_{\ell 1}, \dots \pi_{\ell K})^\top$ is the vector of probabilities of sales reporting from ℓ to the other products such that $\pi_{\ell k}$ is the probability of buying product k, when it is available, with an initial demand of product ℓ and $\pi_{\ell\ell}$ is the probability of buying no product with an initial demand of product ℓ when all the products but product ℓ are available. Note that when other products are not available the consumer cannot consider them for his sale reporting. Thus, under configuration $\boldsymbol{\omega}_j$, the probability of leaving without buying anything, when a consumer comes with an initial demand for product ℓ that is not available (*i.e.*, $\omega_{j\ell} = 0$) is $\boldsymbol{\pi}_\ell^\top (\mathbf{1}_K - \boldsymbol{\omega}_j)$. Concerning notations, $\theta = (\mu_1, \dots, \mu_K, \boldsymbol{\pi}_1^\top, \dots, \boldsymbol{\pi}_K^\top)^\top \in \Theta$ where $\Theta = (\mathbb{R}^{+*})^K \times (S_K)^K$ where S_K is the simplex of dimension K.

4.3.3.2 *Identifiability property of the model*

Model 37 involves $K + K(K-1)$ free parameters. A natural question is to state sufficient conditions that ensure the identifiability of these parameters given $\mathbf{X}_n$. Considering conditions presented in Assumption 8, Proposition 7 states the identifiability of the model parameters.

Assumption 8 *(Observed configurations of availability)* $\boldsymbol{\Gamma}_k = \{\mathbf{1}_K - \boldsymbol{\omega}_j : j \in J_n$ *and* $\omega_{jk} = 1\}$ *and* $\mathbf{R}_k$ *be the matrix of size card*$(\boldsymbol{\Gamma}_k) \times K$ *where each row corresponds to the sum between one element of* $\boldsymbol{\Gamma}_k$ *and* $\mathbf{e}_k$, $\mathbf{e}_k$ *being the vector of length* K *composed of zeros except its coordinate* k *that is equal to one. For any* $k \in \{1, \dots, K\}$, $\mathbf{R}_k$ *is such that it contains a sub-matrix* $\mathbf{S}_k$ *having rank* K.

Example 9 *Given* $K = 3$ *and the observed configuration of availability* $J_n = \{j : \boldsymbol{\omega}_j \in \{(1,1,1), (1,1,0)\ (1,0,0)\}\}$,

$$\boldsymbol{\Gamma}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

Proposition 7 *Assumption 8 is a necessary and sufficient condition for the parameters of model 37 to be identifiable leading to any* $\theta \in \Theta$ *and* $\tilde{\theta} \in \Theta$

$$\forall \mathbf{y}_n \in \mathbb{N}^p, f(\mathbf{y}_n \mid \mathbf{x}_n; \theta) = f(\mathbf{y}_n \mid \mathbf{x}_n; \tilde{\theta}) \Rightarrow \theta = \tilde{\theta}.$$

The proofs can be found in Appendix B.2.

If the number of products is large, it is unlikely that the sub-matrix $\mathbf{S}_k$ has rank K for any $k \in \{1, \ldots, K\}$ as required by Assumption 8. Thus, when K is large, it is unlikely that the observed configurations (*i.e.,* $\boldsymbol{\omega}_j$ such that $X_{nj} > 0$) permit to satisfy Assumption 8, and so the identifiability of the model parameters is not ensured. It is the reason why we will address specifically the issues related to K large in Chapter 5.

The availability of a necessary and sufficient condition for the identifiability of the model makes it the one we selected for the continuation of our work.

4.3.3.3 *Asymptotic property of the model*

This property permits to state an asymptotic control of the normalized difference between the number of sales of a product given the fact that another product is not available and the number of sales of the same product given the fact that the second product is available. Let $V_{nk\ell}$ and $W_{nk\ell}$ be the number of sales of product k given the fact that product ℓ is available and is not available respectively,

$$V_{nk\ell} := \sum_{\{j:\omega_{jk}=1,\, \omega_{j\ell}=1\}} Y_{njk} \quad \text{and} \quad W_{nk\ell} := \sum_{\{j:\omega_{jk}=1,\, \omega_{j\ell}=0\}} Y_{njk}$$

where Y_{njk} is defined in Section 4.3. Under Assumption 4, 9, 10 , Lemma 4 shows that the asymptotic normalized difference between $W_{nk\ell}$ and $V_{nk\ell}$ depends on the substitution of product ℓ by product k. Moreover, when product ℓ cannot be substituted by product k, then this difference converges to zero. Assumption 4 and Assumption 9 are standard since they consider that the intensity μ_k are finite and strictly positive and that the proportion of time spent in each configuration satisfies a central limit theorem. Finally, Assumption 10 considers that the asymptotic times spent in each configuration is a product of the asymptotic times of availability of each product.

Assumption 9 *The time spent in each configuration satisfies a central limit theorem leading to, for any $j \in \{1, \ldots, J\}$,*

$$\frac{X_{nj}}{n} = \tau_j + O_{\mathbb{P}}(n^{-1/2}). \tag{38}$$

Assumption 10 *Let $0 < \rho_k < 1$ be the proportion of time where product k is available, then proportion τ_j satisfies*

$$\tau_j = \prod_{j=1}^{K} \rho_k^{\omega_{jk}} (1-\rho_k)^{1-\omega_{jk}}. \tag{39}$$

Lemma 4 *Let $\Delta_{nk\ell} := W_{nk\ell}/(n_k - n_{k\ell}) - V_{nk\ell}/n_{k\ell}$. Under Assumption 4,9,10 we have*

$$\Delta_{nk\ell} := \mu_\ell \pi_{\ell k} + O_{\mathbb{P}}(n^{-1/2}),$$

where $n_k = \sum_{j=1}^{J} X_{nj}\omega_{jk}$ denotes the time where product k was available and $n_{k\ell} = \sum_{j=1}^{J} X_{nj}\omega_{jk}\omega_{j\ell}$ denotes the time where product k and product ℓ were available simultaneously. The proof can be found in Appendix Proof 8.

4.4 PARAMETER ESTIMATION OF THE SECOND MODEL

4.4.1 *Parameter estimation*

The *observed-data log-likelihood* function (*i.e.*, the log-likelihood function computed on the observed data) given the conditional distribution of $\mathbf{Y}$ described in Equation 34 is defined by

$$\ln L(\theta; \mathbf{y}_n | \mathbf{x}_n) = \sum_{j=1}^{J} \sum_{k=1}^{K} y_{njk} \ln \left[x_{nj}\lambda_k(\boldsymbol{\omega}_j; \theta)\right] - x_{nj}\lambda_k(\boldsymbol{\omega}_j; \theta) - \ln \left[y_{njk}!\right].$$

However, the maximization of the *observed-data log-likelihood* does not lead to a closed form of $\hat{\theta}$ (the corresponding maximum likelihood estimate), hence optimization algorithms are needed. Since the model defined by 37 implies latent variables, it is natural to use an Expectation-Maximization algorithm (EM algorithm; A. P. Dempster, N. M. Laird, and D. B. Rubin (1977) and McLachlan and Krishnan (2007)) to achieve the maximization of the *observed-data log-likelihood* on θ. This algorithm considers the *complete-data log-likelihood* function (*i.e.*, the log-likelihood function computed on both of the observed and latent data) that is defined by

$$\ln L_c(\theta; \mathbf{z}_n, \mathbf{y}_n | \mathbf{x}_n) = \sum_{j=1}^{J} \sum_{k=1}^{K} \ln \left[\mathbb{1}_{y_{njk} = \sum_{\ell}^{K} z_{njk\ell}}\right] + \sum_{j=1}^{J} \sum_{k=0}^{K} \sum_{\ell=1}^{K} z_{njk\ell} \ln \left[x_{nj}\lambda_{k\ell}(\boldsymbol{\omega}_j; \theta)\right] - x_{nj}\lambda_{k\ell}(\boldsymbol{\omega}_j; \theta) - \ln \left[z_{njk\ell}!\right].$$

The EM possibly achieves a local maxima for a specific initialization. Thus, in order to retrieve a global maxima, N_{init} initializations are generated so that the best estimator in terms of *observed-data log-likelihood* is retrieved.

An initialization of the EM algorithm starts with an initial value of the parameters $\theta^{[0]} = (\boldsymbol{\mu}^{[0]}, \boldsymbol{\pi}^{[0]})$ sampled in Θ. The asymptotic relation between the statistic $\Delta_{nk\ell}$ and parameters μ_ℓ, $\boldsymbol{\pi}_\ell$ from Lemma 4 supports the choice of the initialization of $\boldsymbol{\mu}^{[0]}$ as $\mu_\ell^{[0]} = (\pi_{\ell\,\ell+1}^{[0]})^{-1}\Delta_{n\ell+1\ell}$ for $\ell \in \{1, \ldots, K-1\}$ and $\mu_K^{[0]} = (\pi_{K1}^{[0]})^{-1}\Delta_{n1K}$. The algorithm alternates between the computation of the conditional expectation of the *complete-data log-likelihood* given the observed data and the current parameters, and the updating of the parameters by maximizing this conditional expectation with respect to θ under the constraint Θ. The EM algorithm ensures that the *observed-data log likelihood* increases at each iteration and converges to a value. Thus, at iteration $(r > 0)$, the algorithm performs the following two steps:

- E-step: computation of the conditional expectation of the *complete-data log-likelihood*

  $$z_{njk\ell}^{[r]} = \mathbb{E}[Z_{njk\ell} \mid \mathbf{y}_n, \mathbf{x}_n; \theta^{[r-1]}].$$

  For any $k \in \{1, \ldots, K\}$, elementary properties of the Poisson distributions and Lemma 7 presented in Appendix applied with $U = Z_{njk\ell}$, $\mu = x_{nj}\lambda_{k\ell}(\boldsymbol{\omega}_j; \theta^{[r-1]})$, $V = Y_{njk} - Z_{njk\ell}$ and $\nu = x_{nj}(\lambda_k(\boldsymbol{\omega}_j; \theta^{[r-1]}) - \lambda_{k\ell}(\boldsymbol{\omega}_j; \theta^{[r-1]}))$ lead to

  $$z_{njk\ell}^{[r]} = y_{njk} \frac{\lambda_{k\ell}(\boldsymbol{\omega}_j; \theta^{[r-1]})}{\lambda_k(\boldsymbol{\omega}_j; \theta^{[r-1]})}.$$

Moreover, the independence between $Z_{nj0\ell}$ and $Z_{njk\ell}$ for $k \neq 0$ conditionally on $\mathbf{x}$ implies independence between $Z_{nj0\ell}$ and $\mathbf{Y}_n$. Therefore, $\mathbb{E}[Z_{nj0\ell} \mid \mathbf{y}_n, \mathbf{x}_n; \theta^{[r-1]}] = \mathbb{E}[Z_{nj0\ell} \mid \mathbf{x}_n; \theta^{[r-1]}]$, leading to

$$z_{nj0\ell}^{[r]} = x_{nj}\lambda_{0\ell}(\boldsymbol{\omega}_j; \theta^{[r-1]}).$$

- M-step: the maximization of the conditional expectation of the *complete-data log-likelihood*

  $$\theta^{[r]} = \text{argmax}_{\theta \in \Theta} \ln L(\theta; z_n^{[r]}, \mathbf{y}_n | \mathbf{x}_n)$$

  is equivalent to maximizing separately K independent problems of optimisation

  $$(\mu_\ell^{[r]}, \boldsymbol{\pi}_\ell^{[r]}) = \text{argmin}_{\boldsymbol{\pi}_\ell \in S_K, \mu_\ell > 0} \tilde{F}_\ell(\mu_\ell, \boldsymbol{\pi}_\ell) \tag{40}$$

  where

  $$\tilde{F}_\ell(\mu_\ell, \boldsymbol{\pi}_\ell) = -\left[\sum_{j=1}^{J}\sum_{k=0}^{K} z_{njk\ell}^{[r]} \ln[x_{nj}\lambda_{k\ell}(\boldsymbol{\omega}_j; \theta)] - x_{nj}\lambda_{k\ell}(\boldsymbol{\omega}_j; \theta)\right].$$

  Since we have (see Appendix B.2)

  $$\mu_\ell^{[r]} = \frac{1}{n}\sum_{j=1}^{J}\sum_{k=0}^{K} z_{njk\ell}^{[r-1]},$$

  we compute

  $$\boldsymbol{\pi}_\ell^{[r]} = \text{argmin}_{\boldsymbol{\pi}_\ell \in S_K} \tilde{F}_\ell(\mu_\ell^{[r]}, \boldsymbol{\pi}_\ell).$$

  The Hessian $\tilde{H}_\ell(\mu_\ell^{[r]}, \boldsymbol{\pi}_\ell) = \nabla^2_{\boldsymbol{\pi}_\ell} \tilde{F}_\ell(\mu_\ell^{[r]}, \boldsymbol{\pi}_\ell)$ is positive semi definite (see in the appendix Lemma 9 for an explicit value of the Hessian) hence the problem is then convex and the solution is unique. The optimization algorithm used is SLSQP (Sequential Least Squares Quadratic Programming). The function chosen in Python is `optimize.minimize` from the package `SciPy`. The stopping criteria for the optimization problem are $\tilde{F}_\ell(\mu_\ell^{[r]}, \boldsymbol{\pi}_\ell^{[s-1]}) - \tilde{F}_\ell(\mu_\ell^{[r]}, \boldsymbol{\pi}_\ell^{[s]}) < \epsilon_M$ and $s > N_M$ where $s > 0$ is the iterations of the SLSQP. The setting of the parameters is discussed in Section 4.5.1.2.

- The EM algorithm stops when $\ln L(\theta^{[r]}; \mathbf{y}_n | \mathbf{x}_n) - \ln L(\theta^{[r-1]}; \mathbf{y}_n | \mathbf{x}_n) < \epsilon_{em}$.

The code of the EM algorithm will soon be available on GitHub for the reproduction of the following experiences.

4.5 NUMERICAL EXPERIMENTS

The numerical experiments proposed in this section illustrate the good properties of our algorithm on synthetic data and on a real datasets from the Adeo company.

4.5.1 *Tuning and evaluation of the EM algorithm on synthetic data*

4.5.1.1 *Realistic synthetic data scenarios in the Adeo company context*

The objective is twofold: motivate the settings of the EM algorithm and evaluate the quality of the estimators in situations close to Adeo's data.

Let $\mathcal{S} = (K, n, \mu, a_K, c)$ be the scenario associated respectively to a number of products (K), time horizon (n), mean frequencies of sales (μ), probability of lost sale (a) and group heterogeneity (c). Let $n \in \{180, 365, 730\}$ respectively stands for 6 months, one year and two years of daily sales. The number of products varies as $K \in \{2, 5\}$. In real situations, the mean frequency of sales is often not of the same order of magnitude from a product to another. This remark motivates to define the coefficient $c \in \{1, 10\}$ referring to the mean frequency of sale heterogeneity among a group of susbtitution. Products have the same mean frequency of sales μ for products 2 to K and the first product has $\mu_1 = c\mu$. In addition, slow and fast movers cases are expressed by $\mu \in \{0.1, 2\}$. The probabilities of substitution are defined by the parameter $a \in \{\textit{perfect susbt}, \textit{unif}, \textit{no subst}\}$ and respectively represents the case with perfect substitution, uniform substitution and no substitution among the group by the following formulas $\pi_{\ell k} \in \{(K-1)^{-1}, K^{-1}, 0\}\ \forall \ell \neq k$ and $\pi_{\ell\ell} = 1 - \sum_{k \neq \ell} \pi_{\ell k}$. Let $\theta^{\mathcal{S}}$ be the unique parameter associated to each scenario $\mathcal{S}$ as defined in Section 4.3.3.1. $N_{samp} = 100$ samples $(\mathbf{x}_n^{i,\mathcal{S}}, \mathbf{y}_n^{i,\mathcal{S}}, \boldsymbol{\omega}^{i,\mathcal{S}})_{i \in [N_{samp}]}$ are generated from a parameter $\theta^{\mathcal{S}}$. The time in each configuration of availability j is the same on all the possible configurations hence $x_{ij}^{\mathcal{S}} = 2^{-K}, \forall j \in [2^K]$. The elements of $\mathbf{y}_n^{i,\mathcal{S}}$ are sampled from the Poisson distribution with parameters based on $\mathbf{x}_n^{i,\mathcal{S}}$ and $\theta^{\mathcal{S}}$.

Let also $\mathcal{A} = (\epsilon_{em}, N_M, N_{init}) \in \{10^{-5}, 10^{-3}, 10^{-1}\} \times \{2, 5, 10\} \times \{10, 25, 50\}$ and a fixed $\epsilon_m = 10^{-6}$ be the parameters of the EM algorithm to be tested. Then, the algorithm for a sample $(\mathbf{x}_n^{i,\mathcal{S}}, \mathbf{y}_n^{i,\mathcal{S}}, \boldsymbol{\omega}^{i,\mathcal{S}})$ and a setting $\mathcal{A}$ retrieves $\hat{\theta}_i^{\mathcal{S},\mathcal{A}}$ composed of $\hat{\mu}_{\ell,i}^{\mathcal{S},\mathcal{A}}$ and $\hat{\boldsymbol{\pi}}_{\ell,i}^{\mathcal{S},\mathcal{A}}$ $\forall \ell \in [K]$.

4.5.1.2 *Tuning hyperparameters of EM in the defined scenarios*

We evaluated Settings $\mathcal{A}$ in the grid defined above. The conclusions we draw have been checked to all scenarios mentioned in the preceding subsection, where $K \in \{2, 5\}$ and with the case of homogeneous groups of substitution ($c = 1$). We drop this last variable from the tuple $\mathcal{S}$ and mention it later when we will study heterogeneity. To illustrate, we present a subset of scenarios $\mathcal{S}$ in Tables 2, 3, 4, 5 which vary in terms of data quantity and provide the rest of the results in Table 49 to 63 for the case $K = 2$ and Table 64 to 78 for $K = 5$ in Appendix B.3. Specifically, we examine an unfavourable scenario $\mathcal{S}_1 = (2, 180, 0.1, 0.5)$ for slow movers over a short period and a more favourable scenario $\mathcal{S}_2 = (2, 730, 2, 0.5)$ for fast movers over a longer period of time. The remaining two tables provide additional information. Our aim is to compare the quality of estimators and their computational times across different settings. We use the reference estimator denoted as $\hat{\theta}_i^{\mathcal{S},\bar{\mathcal{A}}} = (\hat{\boldsymbol{\mu}}_i^{\mathcal{S},\bar{\mathcal{A}}\top}, \hat{\boldsymbol{\pi}}_i^{\mathcal{S},\bar{\mathcal{A}}\top})$, with $\bar{\mathcal{A}} = (\epsilon_{em} = 10^{-5}, N_M = 10, N_{init} = 50)$ because it relates to the smallest ϵ_{em}, and the largest number of iterations and initializations.

The quality is evaluated by the mean of three criteria. The average difference of *data-observed log-likelihood* between the estimator of reference and the estimator associated to $\mathcal{A}$ is

$$\Delta_{\ell}^{\mathcal{S},\mathcal{A}} = N_{\text{samp}}^{-1} \sum_{i=1}^{N_{\text{samp}}} \left(\ln L(\hat{\theta}_i^{\mathcal{S},\bar{\mathcal{A}}}; y_n^{i,\mathcal{S}} | x_n^{i,\mathcal{S}}) - \ln L(\hat{\theta}_i^{\mathcal{S},\mathcal{A}}; y_n^{i,\mathcal{S}} | x_n^{i,\mathcal{S}}) \right).$$

The mean distances of the estimators to the references $\mathcal{A}$ have the following formulas

$$\Delta_{\hat{\pi}_1}^{\mathcal{S},\mathcal{A}} = N_{\text{samp}}^{-1} \sum_{i=1}^{N_{\text{samp}}} \|\hat{\pi}_{1,i}^{\mathcal{S},\bar{\mathcal{A}}} - \hat{\pi}_{1,i}^{\mathcal{S},\mathcal{A}}\| \quad \text{and} \quad \Delta_{\hat{\mu}_1}^{\mathcal{S},\mathcal{A}} = N_{\text{samp}}^{-1} \sum_{i=1}^{N_{\text{samp}}} \|\hat{\mu}_{1,i}^{\mathcal{S},\bar{\mathcal{A}}} - \hat{\mu}_{1,i}^{\mathcal{S},\mathcal{A}}\|.$$

An algorithm with Settings $\bar{\mathcal{A}}$ is ran. The estimators and likelihoods are recorded at each step. At $N_M = 10$ fixed, the results associated to other settings $\mathcal{A}$ are computed from these recordings by sampling without replacement among the initializations and by clipping the recordings associated at the iteration where the criteria ϵ_{em} is met. This enables $\Delta_{\ell}^{\mathcal{S},\mathcal{A}}$ to be positive most of the time. More specifically, it is assured to be positive when $N_M = 10$. Results among $N_M \in \{2, 5\}$ are produced similarly.

The choice among Settings $\mathcal{A}$ makes a compromise between computation time and precision in terms of distance of the maximum *observed-data log-likelihood* to the benchmark generated by the set of parameters $\bar{\mathcal{A}}$. As the EM is more demanding in terms of precision where $K > 2$, we propose here two settings, the first in the case of $K = 2$ and complement it with the experiences on $K = 5$ to provide another set of parameters that can be used with any K.

We chose to keep the $\Delta_{\pi_1}^{\mathcal{S},\mathcal{A}} < 0.05$ and $\Delta_{\ell}^{\mathcal{S},\mathcal{A}} \leqslant 10^{-3}$ for the limitation of the estimator's variability induced by the setting of parameters. We provide first our analysis of the settings of parameters in the scenarios where $K = 2$ (Table 2, 3). All cases from the tables show that $\mathcal{A}$ where $\epsilon_{em} = 10^{-5}$ are possible candidates and all cases related to $\epsilon_{em} = 0.1$ are excluded. According to Table 2 candidates are similar in terms of accuracy for $\epsilon_{em} > 0.1$. $N_M = 2$. Table 3 in the most favourable case is coherent with these conclusions. Since $N_M = 2$ is of the same order of magnitude in terms of time of execution, we leave it aside for the analysis of the case $K = 5$.

Table 4 shows that only the cases where $\epsilon_{em} = 10^{-5}$ verifies $\Delta_{\ell}^{\mathcal{S},\mathcal{A}} \leqslant 10^{-3}$ and all of the accuracies are acceptable. **We conclude that Setting $\mathcal{A} = (10^{-5}, 5, 10)$ (denoted now as $\mathcal{A}^*$) is optimal and should be used in general since it has the lowest time of execution.**

4.5.1.3 *About the negligeability of the non increasing log-likelihood cases*

Throughout the experiences, some EM runs did not validate the increasing property of the *observed-data log-likelihood*. It is however marginal and possibly due to numerical difficulties encounters in the border of the parameter space (when substitution probabilities are close to zero). We isolated the results for the advised parametrization in the cases $K \in \{2, 5\}$ and found out that no cases were reported for $K = 2$. On the scenarios where $(K, \mu) = (5, 2)$, we reported that among the 9,000 different initializations only 0.003% had a non increasing *observed-data log-likelihood*. In these non increasing cases 90% had a cumulative error relative to the max log likelihood under 2% and 60% had a number of non increasing step relative to the total number of step lower than 5%. From these observations, we judged that it is negligible in terms of occurences and implemented a condition of increasing loglikelihood to accept the result of an initialization.

Table 2: Results from the numerical applications in an unfavorable case $\mathcal{S} = (2, 180, 0.1, 0.5)$

ϵ_{em}	N_M	N_{init}	$\Delta_{\ell}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\pi_1}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\mu_1}^{\mathcal{S},\mathcal{A}}$(std)	mean execution time [s]	mean number iterations
1e-05	2	10	0.000 (0.000)	0.017 (0.132)	0.000 (0.000)	3.271	83.760 (30.398)
		25	0.000 (0.000)	0.009 (0.063)	0.000 (0.000)	8.380	77.080 (33.542)
		50	0.000 (0.000)	0.003 (0.009)	0.000 (0.000)	16.550	76.810 (37.987)
	5	**10**	**0.000(0.000)**	**0.017(0.132)**	**0.000(0.000)**	**3.576**	**69.800(33.153)**
		25	0.000 (0.000)	0.010 (0.081)	0.000 (0.000)	9.107	59.580 (28.399)
		50	-0.000 (0.000)	0.002 (0.009)	0.000 (0.000)	17.949	58.340 (35.646)
	10	10	0.000 (0.000)	0.017 (0.132)	0.000 (0.000)	3.701	71.450 (35.623)
		25	0.000 (0.000)	0.001 (0.004)	0.000 (0.000)	9.429	60.570 (30.417)
		50	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	18.559	56.820 (31.235)
0.001	2	10	0.007 (0.003)	0.052 (0.136)	0.001 (0.001)	1.348	31.380 (15.878)
		25	0.006 (0.003)	0.036 (0.067)	0.001 (0.001)	3.455	29.400 (19.057)
		50	0.005 (0.003)	0.026 (0.030)	0.001 (0.001)	6.807	27.150 (18.396)
	5	10	0.005 (0.003)	0.051 (0.136)	0.001 (0.001)	1.413	24.330 (10.866)
		25	0.004 (0.003)	0.038 (0.090)	0.001 (0.001)	3.607	21.620 (12.333)
		50	0.004 (0.003)	0.026 (0.031)	0.001 (0.001)	7.103	20.880 (12.594)
	10	10	0.005 (0.003)	0.053 (0.137)	0.001 (0.001)	1.468	24.620 (12.262)
		25	0.004 (0.003)	0.029 (0.031)	0.001 (0.001)	3.740	22.030 (12.939)
		50	0.004 (0.003)	0.025 (0.030)	0.001 (0.001)	7.372	21.280 (13.322)
0.1	2	10	0.182 (0.142)	0.260 (0.218)	0.009 (0.005)	0.264	5.050 (1.951)
		25	0.120 (0.114)	0.151 (0.155)	0.007 (0.005)	0.668	4.660 (1.823)
		50	0.100 (0.101)	0.155 (0.166)	0.007 (0.005)	1.336	4.420 (1.505)
	5	10	0.133 (0.083)	0.203 (0.172)	0.008 (0.006)	0.337	5.470 (2.207)
		25	0.091 (0.073)	0.121 (0.117)	0.007 (0.005)	0.857	4.990 (2.133)
		50	0.079 (0.071)	0.118 (0.111)	0.007 (0.005)	1.713	4.830 (2.687)
	10	10	0.130 (0.084)	0.197 (0.160)	0.008 (0.006)	0.343	5.560 (2.393)
		25	0.089 (0.073)	0.121 (0.117)	0.007 (0.005)	0.870	5.120 (2.447)
		50	0.076 (0.069)	0.112 (0.107)	0.007 (0.005)	1.736	4.800 (2.285)

In bold: results associated to the advised settings for $K = 2$.

Table 3: Results from the numerical applications in a more favorable case $\mathcal{S} = (2, 730, 2.0, 0.5)$

ϵ_{em}	N_M	N_{init}	$\Delta_{\ell}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\pi_1}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\mu_1}^{\mathcal{S},\mathcal{A}}$(std)	mean execution time [s]	mean number iterations
1e-05	2	10	0.000 (0.000)	0.001 (0.001)	0.001 (0.001)	3.777	85.490 (29.023)
		25	0.000 (0.000)	0.001 (0.001)	0.001 (0.001)	9.481	75.750 (37.031)
		50	-0.000 (0.000)	0.001 (0.001)	0.001 (0.001)	19.031	64.640 (34.078)
	5	**10**	**0.000(0.000)**	**0.001(0.001)**	**0.001(0.001)**	**4.605**	**99.000(33.494)**
		25	0.000 (0.000)	0.001 (0.001)	0.000 (0.001)	11.680	87.630 (34.208)
		50	0.000 (0.000)	0.000 (0.001)	0.000 (0.001)	23.390	75.870 (40.566)
	10	10	0.000 (0.000)	0.001 (0.001)	0.001 (0.001)	4.740	96.250 (34.569)
		25	0.000 (0.000)	0.000 (0.001)	0.000 (0.001)	12.042	92.750 (38.011)
		50	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	24.089	76.930 (40.360)
0.001	2	10	0.006 (0.002)	0.009 (0.004)	0.006 (0.002)	2.199	36.020 (24.208)
		25	0.004 (0.002)	0.007 (0.004)	0.005 (0.003)	5.512	22.710 (13.944)
		50	0.003 (0.002)	0.005 (0.004)	0.004 (0.003)	11.118	18.360 (4.614)
	5	10	0.006 (0.002)	0.010 (0.004)	0.006 (0.003)	2.840	45.030 (26.421)
		25	0.004 (0.002)	0.007 (0.005)	0.005 (0.003)	7.227	28.140 (22.053)
		50	0.003 (0.002)	0.006 (0.004)	0.004 (0.003)	14.514	21.600 (12.830)
	10	10	0.006 (0.002)	0.010 (0.004)	0.006 (0.002)	2.935	47.940 (32.006)
		25	0.004 (0.002)	0.007 (0.005)	0.005 (0.003)	7.500	26.970 (18.885)
		50	0.003 (0.002)	0.006 (0.004)	0.004 (0.003)	15.027	22.460 (17.178)
0.1	2	10	0.162 (0.148)	0.036 (0.031)	0.029 (0.020)	0.848	10.880 (4.323)
		25	0.073 (0.042)	0.017 (0.015)	0.019 (0.012)	2.136	10.140 (2.581)
		50	0.058 (0.020)	0.014 (0.011)	0.017 (0.010)	4.283	10.120 (2.840)
	5	10	0.228 (0.204)	0.047 (0.038)	0.034 (0.022)	1.180	11.620 (9.300)
		25	0.097 (0.068)	0.023 (0.022)	0.020 (0.014)	3.006	9.450 (2.598)
		50	0.072 (0.029)	0.015 (0.013)	0.017 (0.012)	6.027	8.870 (2.493)
	10	10	0.226 (0.201)	0.046 (0.036)	0.035 (0.023)	1.228	12.330 (10.966)
		25	0.097 (0.068)	0.023 (0.022)	0.020 (0.014)	3.160	9.450 (2.598)
		50	0.071 (0.029)	0.015 (0.013)	0.017 (0.012)	6.315	8.910 (2.542)

In bold: results associated to the advised settings for $K = 2$.

Table 4: Results from the numerical applications for $\mathcal{S} = (5, 180, 0.1, 1.0)$

ϵ_{em}	N_M	N_{init}	$\Delta_{\ell}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\pi_1}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\mu_1}^{\mathcal{S},\mathcal{A}}$(std)	mean execution time [s]	mean number iterations
1e-05	**5**	**10**	**0.000(0.000)**	**0.018(0.077)**	**0.000(0.000)**	**34.003**	**164.850(57.164)**
		25	0.000 (0.000)	0.021 (0.079)	0.000 (0.000)	84.970	159.190 (59.341)
		50	0.000 (0.000)	0.016 (0.051)	0.000 (0.000)	170.519	155.700 (59.021)
	10	10	0.000 (0.000)	0.021 (0.082)	0.000 (0.000)	37.728	162.960 (56.026)
		25	0.000(0.000)	0.016 (0.059)	0.000(0.000)	94.202	159.670 (58.172)
		50	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	188.571	155.750 (59.654)
0.001	5	10	0.015 (0.006)	0.073 (0.110)	0.002 (0.002)	10.665	50.050 (9.157)
		25	0.015 (0.007)	0.065 (0.087)	0.002 (0.002)	26.828	50.590 (11.630)
		50	0.014 (0.007)	0.064 (0.104)	0.002 (0.001)	53.853	50.280 (10.831)
	10	10	0.015 (0.006)	0.069 (0.112)	0.002 (0.002)	12.300	50.070 (9.376)
		25	0.014 (0.006)	0.065 (0.104)	0.002 (0.001)	30.943	50.270 (10.166)
		50	0.013 (0.006)	0.068 (0.116)	0.001 (0.001)	62.175	51.330 (11.499)
0.1	5	10	0.444 (0.076)	0.287 (0.199)	0.009 (0.007)	2.462	12.000 (1.929)
		25	0.425 (0.069)	0.250 (0.154)	0.010 (0.007)	6.169	12.000 (1.510)
		50	0.413 (0.066)	0.259 (0.178)	0.009 (0.007)	12.362	11.850 (1.519)
	10	10	0.435 (0.090)	0.276 (0.176)	0.011 (0.009)	2.690	11.130 (1.831)
		25	0.414 (0.084)	0.257 (0.155)	0.010 (0.007)	6.753	10.960 (1.624)
		50	0.396 (0.079)	0.250 (0.157)	0.010 (0.007)	13.544	10.970 (1.791)

In bold: results associated to the advised settings for $K = 5$

Table 5: Results from the numerical applications for $\mathcal{S} = (5, 730, 2.0, 0.2)$

ϵ_{em}	N_M	N_{init}	$\Delta_{\ell}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\pi_1}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\mu_1}^{\mathcal{S},\mathcal{A}}$(std)	mean execution time [s]	mean number iterations
1e-05	**5**	**10**	**0.000(0.000)**	**0.002(0.002)**	**0.001(0.001)**	**88.047**	**328.950(103.552)**
		25	0.000 (0.000)	0.002 (0.002)	0.001 (0.001)	222.208	324.290 (111.880)
		50	0.000 (0.000)	0.001 (0.002)	0.001 (0.001)	443.346	321.840 (115.294)
	10	10	0.000 (0.000)	0.001 (0.002)	0.001 (0.001)	109.585	391.480 (138.332)
		25	0.000(0.000)	0.001(0.002)	0.001(0.001)	275.024	378.210(149.868)
		50	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	550.069	374.170 (166.943)
0.001	5	10	0.025 (0.009)	0.015 (0.008)	0.010 (0.006)	40.371	142.740 (33.886)
		25	0.023 (0.008)	0.013 (0.008)	0.008 (0.006)	101.479	141.880 (48.509)
		50	0.021 (0.007)	0.012 (0.007)	0.008 (0.006)	203.089	142.950 (54.234)
	10	10	0.026 (0.009)	0.015 (0.009)	0.010 (0.007)	54.468	183.040 (70.276)
		25	0.023 (0.008)	0.013 (0.008)	0.008 (0.007)	136.373	180.410 (84.693)
		50	0.021 (0.007)	0.012 (0.007)	0.007 (0.006)	274.045	181.820 (87.495)
0.1	5	10	1.300 (0.172)	0.085 (0.037)	0.082 (0.060)	13.733	45.160 (8.229)
		25	1.191 (0.165)	0.082 (0.037)	0.071 (0.050)	34.275	43.920 (9.715)
		50	1.119 (0.153)	0.081 (0.040)	0.063 (0.045)	68.770	42.820 (10.079)
	10	10	1.419 (0.235)	0.095 (0.044)	0.078 (0.059)	18.944	52.510 (11.158)
		25	1.257 (0.195)	0.086 (0.040)	0.063 (0.042)	47.137	52.030 (10.482)
		50	1.178 (0.160)	0.083 (0.042)	0.064 (0.041)	95.053	50.080 (10.786)

In bold: results associated to the advised settings for $K = 5$

4.5.1.4 *Evaluation of the EM estimates quality in the defined scenarios*

In this section, the parameters are set to the advised $\mathcal{A}^*$ and for ease of reading, the superscript is dropped. According to the generation of $\boldsymbol{\pi}_i^{\mathcal{S}}$ and $\boldsymbol{\mu}_i^{\mathcal{S}}$ where $i \in [1, N_{\text{samp}}]$, the properties of the estimators $\hat{\pi}_{\ell k,i}^{\mathcal{S}}$ and $\mu_\ell^{\mathcal{S}}$ are symmetric for ℓ, k when $k \neq \ell$. Hence first measures provided are the mean and variance of $\hat{\pi}_{\ell k,i}^{\mathcal{S}}$ for $\ell = 1, k = 2$ and $\hat{\mu}_1$. The multivariate aspect is assessed via the mean squared error

$$\text{mse}(\hat{\boldsymbol{\pi}}_i^{\mathcal{S}}) = N_{\text{samp}}^{-1} \sum_{i=1}^{N_{\text{samp}}} \|\boldsymbol{\pi}_i^{\mathcal{S}} - \hat{\boldsymbol{\pi}}_i^{\mathcal{S}}\|^2 \quad \text{and} \quad \text{mse}(\hat{\boldsymbol{\mu}}_i^{\mathcal{S}}) = N_{\text{samp}}^{-1} \sum_{i=1}^{N_{\text{samp}}} \|\boldsymbol{\mu}_i^{\mathcal{S}} - \hat{\boldsymbol{\mu}}_i^{\mathcal{S}}\|^2$$

with $\|\|$ being the matrix Euclidean norm.

In the following tables, we included the value of π_{12} which is determined by a. According to Table 6, results show an improvement in terms of all measures as n increases which is due to the consistance of the estimators. In terms of the quality of the estimators $\hat{\pi}_{12,i}^{\mathcal{S}}, \hat{\mu}_{1,i}^{\mathcal{S}}$, the cases related to slow movers show a significantly higher deviation from the true value both in terms of the bias and the variance compared to fast movers. In comparison, Table 7 and 8 show an increase in terms of mse which is the consequence of the augmentation of the number of parameters to be estimated. The variance and bias of $\hat{\pi}_{12}$ alone are not significantly impacted by increasing K. The impact of substitution parameters on the border of the parameter space concerning the quality the quality do not show any type of recurrent behavior exception made for the augmentation of the bias of μ_1 as K increases in the no substitution case. The high standard deviation of the substituability parameters in the tables, especially for slow movers ($\mu = 0.1$) indicates that the signal of substitution is weak for the chosen time frames n.

4.5.1.5 *Assessment of the quality of the estimators for heterogeneous selling frequency*

In real situations, a group of products may have different mean frequency of sales, thus a natural question is to ask how it affects the quality of the various estimators. We compare the results of the previous subsection with the new results. Scenarios $\mathcal{S}$ remain exactly the same except that we now fix $c = 10$ meaning that the first product is sold more and has $\mu_1 = 10\mu$. Hence π_{12} refers to the substitution of a fast mover by a slow mover. π_{21} refers to the substitution of a slow mover by a fast mover and π_{23} refers a slow mover by a slow mover.

From Table 9 we derive a conclusion both in terms of bias and variance. The conclusions are derived from first a comparison of the column of Table 8 in the last subsection. According to it, the quality of the estimation of the substitution from a slow mover to a slow mover is of the same degree of precision as in the homogeneous case hence we use it as a benchmark for the analysis of the two remaining columns. The quality of the estimators from a slow mover to a fast mover is severely degraded. The substitution from a fast mover to a slow mover however is more precise. This last two observations may stem from the difference of signals meaning that if the slow mover is available and not the fast mover there should be some apparent spikes if the two are substituable. However, if the fast mover is available and not the slow movers, the spillover may not be distinguishable from the variability of the demand of the fast mover. We chose now to provide a statistical test to discriminate cases where the substitution can be captured.

Table 6: Numerical applications for $\mathcal{S}$ where $K = 2$

a	π_{12}	μ	n	mean($\hat{\pi}_{12}$) (std)	mean($\hat{\mu}_1$) (std)	mse($\hat{\boldsymbol{\pi}}$)	mse($\hat{\boldsymbol{\mu}}$)
perfect subst	1.00	0.10	180	0.68 (0.35)	0.12 (0.04)	1.12	0.00
			365	0.78 (0.31)	0.11 (0.03)	0.60	0.00
			730	0.77 (0.28)	0.11 (0.03)	0.48	0.00
		2.00	180	0.91 (0.13)	2.04 (0.20)	0.09	0.09
			365	0.93 (0.09)	2.08 (0.13)	0.05	0.04
			730	0.95 (0.07)	2.04 (0.09)	0.03	0.02
unif	0.50	0.10	180	0.45 (0.39)	0.10 (0.04)	0.64	0.00
			365	0.45 (0.37)	0.10 (0.03)	0.55	0.00
			730	0.55 (0.31)	0.10 (0.02)	0.41	0.00
		2.00	180	0.52 (0.19)	1.99 (0.23)	0.14	0.10
			365	0.52 (0.12)	1.98 (0.14)	0.06	0.05
			730	0.50 (0.08)	2.01 (0.11)	0.03	0.02
no subst	0.00	0.10	180	0.26 (0.37)	0.09 (0.03)	0.70	0.00
			365	0.16 (0.26)	0.09 (0.03)	0.45	0.00
			730	0.15 (0.21)	0.10 (0.02)	0.27	0.00
		2.00	180	0.06 (0.10)	1.98 (0.16)	0.04	0.06
			365	0.05 (0.08)	1.95 (0.13)	0.03	0.04
			730	0.03 (0.04)	1.97 (0.09)	0.01	0.02

Table 7: Numerical applications for $\mathcal{S}$ where $K = 3$

a	π_{12}	μ	n	mean($\hat{\pi}_{12}$) (std)	mean($\hat{\mu}_1$) (std)	mse($\hat{\boldsymbol{\pi}}$)	mse($\hat{\boldsymbol{\mu}}$)
perfect subst	0.50	0.10	180	0.40 (0.38)	0.11 (0.05)	1.53	0.01
			365	0.32 (0.32)	0.11 (0.03)	1.36	0.00
			730	0.42 (0.29)	0.11 (0.03)	0.95	0.00
		2.00	180	0.42 (0.14)	2.03 (0.23)	0.26	0.19
			365	0.47 (0.10)	2.07 (0.17)	0.15	0.11
			730	0.47 (0.07)	2.05 (0.11)	0.07	0.05
unif	0.33	0.10	180	0.26 (0.35)	0.10 (0.05)	1.34	0.01
			365	0.32 (0.33)	0.10 (0.03)	1.05	0.00
			730	0.38 (0.28)	0.10 (0.03)	0.81	0.00
		2.00	180	0.35 (0.16)	1.98 (0.26)	0.34	0.20
			365	0.33 (0.12)	2.01 (0.15)	0.18	0.09
			730	0.34 (0.08)	2.00 (0.13)	0.09	0.05
no subst	0.00	0.10	180	0.20 (0.32)	0.08 (0.04)	2.27	0.01
			365	0.15 (0.25)	0.08 (0.02)	1.51	0.00
			730	0.14 (0.20)	0.09 (0.02)	1.01	0.00
		2.00	180	0.08 (0.09)	1.88 (0.19)	0.16	0.14
			365	0.04 (0.07)	1.92 (0.13)	0.08	0.07
			730	0.03 (0.04)	1.94 (0.09)	0.04	0.04

Table 8: Quality of estimators for $\mathcal{S}$ where $K = 5$

a	π_{12}	μ	n	mean($\hat{\pi}_{12}$) (std)	mean($\hat{\mu}_1$) (std)	mse($\hat{\boldsymbol{\pi}}$)	mse($\hat{\boldsymbol{\mu}}$)
perfect subst	0.25	0.10	180	0.20 (0.32)	0.10 (0.06)	2.75	0.02
			365	0.23 (0.29)	0.11 (0.04)	2.18	0.01
			730	0.21 (0.22)	0.11 (0.03)	1.72	0.01
		2.00	180	0.20 (0.17)	2.11 (0.31)	0.78	0.64
			365	0.21 (0.09)	2.12 (0.24)	0.48	0.35
			730	0.25 (0.09)	2.09 (0.20)	0.26	0.18
unif	0.20	0.10	180	0.21 (0.31)	0.10 (0.05)	2.51	0.01
			365	0.17 (0.24)	0.11 (0.04)	2.02	0.01
			730	0.19 (0.22)	0.10 (0.03)	1.55	0.00
		2.00	180	0.18 (0.17)	2.02 (0.33)	0.72	0.49
			365	0.18 (0.11)	2.02 (0.24)	0.45	0.29
			730	0.19 (0.08)	2.02 (0.16)	0.29	0.14
no subst	0.00	0.10	180	0.20 (0.32)	0.08 (0.04)	6.09	0.01
			365	0.17 (0.26)	0.08 (0.03)	4.84	0.01
			730	0.16 (0.24)	0.08 (0.02)	3.74	0.01
		2.00	180	0.08 (0.12)	1.75 (0.26)	1.04	0.58
			365	0.04 (0.06)	1.86 (0.16)	0.41	0.25
			730	0.04 (0.05)	1.89 (0.12)	0.17	0.13

Table 9: Numerical applications for $K = 5$ and heterogeneous mean frequencies of sale

a	π_{12}	μ	n	mean($\hat{\pi}_{12}$) (std)	mean($\hat{\pi}_{21}$) (std)	mean($\hat{\pi}_{23}$) (std)	mean($\hat{\mu}_1$) (std)	mse($\hat{\pi}$)	mse($\hat{\mu}$)
perfect subst	0.25	0.10	180	0.21 (0.10)	0.23 (0.29)	0.23 (0.30)	1.00 (0.09)	1.91	0.03
			365	0.22 (0.07)	0.23 (0.29)	0.21 (0.27)	1.03 (0.08)	1.62	0.02
			730	0.25 (0.05)	0.21 (0.26)	0.21 (0.22)	1.01 (0.06)	1.31	0.01
		2.00	180	0.24 (0.02)	0.17 (0.21)	0.22 (0.17)	20.24 (0.66)	0.87	1.07
			365	0.24 (0.02)	0.18 (0.17)	0.22 (0.13)	20.22 (0.44)	0.61	0.60
			730	0.25 (0.01)	0.21 (0.16)	0.22 (0.09)	20.18 (0.31)	0.36	0.30
unif	0.20	0.10	180	0.19 (0.11)	0.24 (0.33)	0.14 (0.24)	0.98 (0.10)	1.81	0.03
			365	0.20 (0.08)	0.19 (0.28)	0.18 (0.25)	1.00 (0.09)	1.57	0.02
			730	0.21 (0.06)	0.27 (0.28)	0.20 (0.26)	0.99 (0.06)	1.16	0.01
		2.00	180	0.20 (0.03)	0.27 (0.26)	0.16 (0.16)	19.86 (0.67)	0.84	0.91
			365	0.20 (0.02)	0.22 (0.20)	0.17 (0.13)	19.99 (0.60)	0.55	0.64
			730	0.20 (0.01)	0.19 (0.15)	0.19 (0.09)	19.95 (0.38)	0.36	0.29
no subst	0.00	0.10	180	0.03 (0.04)	0.28 (0.32)	0.22 (0.30)	0.97 (0.11)	5.11	0.03
			365	0.02 (0.03)	0.33 (0.31)	0.15 (0.22)	0.97 (0.08)	4.51	0.02
			730	0.01 (0.02)	0.29 (0.28)	0.17 (0.22)	0.95 (0.05)	4.08	0.01
		2.00	180	0.01 (0.01)	0.19 (0.26)	0.08 (0.12)	19.21 (0.62)	1.89	1.45
			365	0.00 (0.01)	0.15 (0.25)	0.04 (0.07)	19.51 (0.51)	0.94	0.71
			730	0.00 (0.01)	0.13 (0.18)	0.03 (0.05)	19.59 (0.40)	0.53	0.44

4.5.2 *Statistical test of significance of substitution*

4.5.2.1 *Definition and parameterization of a statistical test for discriminating substitution and non substitution*

The signal of substitution between two products $\pi_{\ell k}$ might not be retrievable accurately because of the variability of the estimators. Hence we provide here a first statistical test that assesses if the signal of substitution is significant. The hypothesis $H_0 : \pi_{\ell k} = 0$ is tested against the alternative $H_1 : \pi_{\ell k} > 0$ via an empirical likelihood ratio test. In order to do so a maximum likelihood ratio statistic is computed and compared to the empirical distribution of the statistic under H_0 with a risk α. Let $\hat{\theta}_i^{\mathcal{S}} = \text{argmax}_{\theta \in \Theta} \ln L(\theta; y_i^{\mathcal{S}} | x_i^{\mathcal{S}})$ and $\hat{\theta}_{i|H_0}^{\mathcal{S}} = \text{argmax}_{\theta \in \Theta_0} \ln L(\theta; y_i^{\mathcal{S}} | x_i^{\mathcal{S}})$ where $i \in [N_{\text{samp}}]$ be the maximum likelihood estimator retrieved by the EM procedure where $\Theta_0 = \Theta \cap \{\pi_{\ell k} = 0\}$. The likelihood ratio statistic is

$$\text{LRT}_i^{\mathcal{S}} = -2\left(\ln L(\hat{\theta}_{i|H_0}^{\mathcal{S}}; y_i | x_i) - \ln L(\hat{\theta}_i^{\mathcal{S}}; y_i | x_i) \right).$$

The distribution of the statistic $\text{LRT}_{i|H_0}^{\mathcal{S},h}$ under H_0 is evaluated by bootstrap on $h \in [H]$ samples y_i^h generated from the model described in Section 4.3 given $x_i^{\mathcal{S}}$ and $\hat{\theta}_{i|H_0}^{\mathcal{S}}$.

Let the proportion of rejection of H_0 at risk α for a scenario $\mathcal{S}$ be

$$p_{\text{reject}}^{\alpha,\mathcal{S}} = N_{\text{samp}}^{-1} \quad \#\left\{ i \in [N_{\text{samp}}] : \quad H^{-1} \sum_{h=1}^{H} (1_{\text{LRT}_i^{\mathcal{S}} - \text{LRT}_{i|H_0}^{\mathcal{S},h}}) > 1 - \alpha \right\}.$$

In order to evaluate the quality of the confidence interval, we also provide the proportion of cases where the true value of $\pi_{12}^{\mathcal{S}}$ is present in the interval of confidence of the estimator hence

$$p_{prst}^{\alpha,\mathcal{S}} = N_{samp}^{-1} \quad \#\left\{ i \in [N_{samp}] : \quad \pi_{12}^{\mathcal{S}} \in CI(\alpha, \hat{\pi}_{12,i}^{\mathcal{S}}) \right\}$$

where $CI(\alpha, \hat{\pi}_{12,i}^{\mathcal{S}})$ is the empirical interval of confidence with level $1-\alpha$ computed with bootstrap.

Remark 6 *Since the statistic $LRT_i^{\mathcal{S}}$ for the empirical likelihood test are called* H *times, there is an incentive to provide the initialization at the parameter that generated the bootstrap instead of initializing several times. This procedure introduce bias in the estimators.*

Let the estimators $\hat{\theta}_{i\bullet}^{\mathcal{S},h,\bar{\mathcal{M}}}$ be the one that are produced with *multiple* initializations and $\hat{\theta}_{i\bullet}^{\mathcal{S},h,\mathcal{M}}$ for the estimators associated to the single initialization at $\hat{\theta}_{i|H_0}^{\mathcal{S}}$. Let the difference in *observed-data log-likelihood* for the constrained and unconstrained estimators be

$$\Delta_{\ell|H_0}^{\mathcal{S},\mathcal{M}} = \frac{1}{HN_{samp}} \sum_{i=1}^{N_{samp}} \sum_{h=1}^{H} \ln L(\hat{\theta}_{i|H_0}^{\mathcal{S},h,\bar{\mathcal{M}}}; \mathbf{y}_i^h | \mathbf{x}_i^h) - \ln L(\hat{\theta}_{i|H_0}^{\mathcal{S},h,\mathcal{M}}; \mathbf{y}_i^h | \mathbf{x}_i^h)$$

and

$$\Delta_{\ell}^{\mathcal{S},\mathcal{M}} = \frac{1}{HN_{samp}} \sum_{i=1}^{N_{samp}} \sum_{h=1}^{H} \ln L(\hat{\theta}_{i}^{\mathcal{S},h,\bar{\mathcal{M}}}; \mathbf{y}_i^h | \mathbf{x}_i^h) - \ln L(\hat{\theta}_{i}^{\mathcal{S},h,\mathcal{M}}; \mathbf{y}_i^h | \mathbf{x}_i^h).$$

Let

$$\Delta_{\hat{\pi}_1}^{\mathcal{S},\mathcal{M}} = \frac{1}{HN_{samp}} \sum_{i=1}^{N_{samp}} \sum_{h=1}^{H} ||\hat{\boldsymbol{\pi}}_{1,i}^{\mathcal{S},h,\bar{\mathcal{M}}} - \hat{\boldsymbol{\pi}}_{1,i}^{\mathcal{S},\mathcal{M}}||$$

and

$$\Delta_{\hat{\pi}_1|H_0}^{\mathcal{S},\mathcal{M}} = \frac{1}{HN_{samp}} \sum_{i=1}^{N_{samp}} \sum_{h=1}^{H} ||\hat{\boldsymbol{\pi}}_{1,i|H_0}^{\mathcal{S},h,\bar{\mathcal{M}}} - \hat{\boldsymbol{\pi}}_{1,i|H_0}^{\mathcal{S},\mathcal{M}}||.$$

We chose to evaluate the quality on $N_{samp} = 20$, $H = 10$ and $\mathcal{S} \in \{(2, 0.5, 2, 730), (5, 0.2, 2, 730)\}$. Results shown in Table 10 indicate that there is an order of magnitude in terms of the time of computation. There is a slight decrease on the quality of estimation. In the case of the constrained estimator for the scenario with $K = 2$ the same results are retrieved for both. According to the table, the loss in terms of accuracy is more pronounced for the scenarios of $K = 5$ on both the estimators of the constrained and unconstrained with a value of about 0.08. Hence, this test should be computed with all its initializations despite the mean execution time.

Table 10: Random initialization versus single initialization

K	a	π_{12}	μ	n	$\Delta_{\ell}^{\mathcal{S},\mathcal{M}}$ (std)	$\Delta_{\ell\|H_0}^{\mathcal{S},\mathcal{M}}$ (std)	$\Delta_{\hat{\pi}_1}^{\mathcal{S},\mathcal{M}}$ (std)	$\Delta_{\hat{\pi}_1\|H_0}^{\mathcal{S},\mathcal{M}}$(std)	mean execution time $\mathcal{M}$ [s]	mean execution time $\bar{\mathcal{M}}$ [s]
2	*unif*	0.5	2.0	730	0.27(0.53)	5e-6(1e-5)	0.052(0.07)	2.3e-17 (4.5e-17)	19	382
5	*unif*	0.2	2.0	730	3.58(9.9)	0.31(0.8)	0.08(0.1)	0.08(0.10)	699	8,552

Table 11: Numerical applications for $\mathcal{S}$ where $(K, c) = (2, 1)$

a	π_{12}	μ	n	$p_{prst}^{\alpha,\mathcal{S}}$	$p_{reject}^{\alpha,\mathcal{S}}$
perfect subst	1.00	0.10	180	0.93	0.38
			365	0.88	0.66
			730	0.81	0.84
		2.00	180	0.88	1.00
			365	0.88	1.00
			730	0.84	1.00
unif	0.50	0.10	180	0.98	0.11
			365	0.97	0.24
			730	0.97	0.49
		2.00	180	0.84	0.93
			365	0.91	0.99
			730	0.91	1.00
no subst	0.00	0.10	180	0.96	0.05
			365	0.93	0.06
			730	0.89	0.07
		2.00	180	0.94	0.05
			365	0.91	0.09
			730	0.94	0.04

4.5.2.2 *Application of the statistical tests to the defined scenarios*

Results of the experiment is provided for $K = 2$ in Table 11. According to the column $p_{reject}^{\alpha,\mathcal{S}}$, the test rejects the non substitution case when there is substitution in the fast movers case ($\mu = 2$). The same test does not reject the null hypothesis when there is no substitution hence concluding that for fast movers the signal is detectable. In the slow movers case where $\mu = 0.1$, the signal is weaker. A maximum proportion of true positive for medium substitution ($\pi_{12} = 0.5$) is 0.5 meaning that even in the best case where $n = 730$ we may not detect the signal systematically. For perfectly substituable ($\pi_{12} = 1$) slow movers on a short period of time, only 0.37 pass the test. In the case of slow movers and short period of time the signal is weak and attains around 0.13 of proportion of case where the substitution is detected. We expect that on real data, the probability of detecting substitution for slow movers may be partialy retrievable if enough data are available meaning at least 365 days of sales. The second column $p_{prst}^{\alpha,\mathcal{S}}$ provides the assurance that the CIs are accurate except for a slight deviation in the case of the full substitution case.

4.6 APPLICATION TO REAL DATASETS

4.6.1 *Description of the datasets and the designs of experiments*

The dataset is composed of 3 putting knives of size 16, 14 and 12 cm, a set of 2 tape rolls and a padlock shown in Figure 11. The usage of the first 3 overlap and should be substitutable whereas the set of tapes rolls or the padlock is not substituable with the 3 putting knife. According to the data the padlock is of the same order of mean frequency of sales whereas the set of tapes is much more successful in sales. This choice of products enables the observation of effects of heterogeneous/homogeneous mean frequency of sale on the EM results. The dataset includes the time series of sales and stocks of all stores of Leroy Merlin over a period of $n = 1,613$ days.

We refer later to the *desired structure* as the case where the 16 cm is expected to substitute more to the 14 than to the 12 cm and is expected to not substitute to the set of tape. Moreover, the 14 cm may substitute to the 16 cm and the 12 cm and not the set of tapes. The 12 cm may substitute more to the 14 cm than the 16 cm and not to the set of tape. Finally, the set of tapes does not substitute with other products.

First, we provide the numerical results and interpretation for the 3 putting knives and the set of tapes over a year of data on a single store and delve into shortcomings and the reasons associated to it. Then we provide the analysis on the full period of time and all stores. Finally we provide the results of the numerical application in the case of the second dataset with the 3 putting knives and a padlock. Advantage of this second dataset is that it features homogeneous mean frequency of sales.

All tables that follow should be interpreted from the substitution probability $\pi_{\ell k}$ point of view. In more details, rows are associated to out-of-stock products and columns correspond to the substitute, as it is displayed in the example given by Table 12.

Table 12: $\hat{\theta}$ on dataset

	$\hat{\mu}_k$	1	2	3	4
1	$\hat{\mu}_1$	$\hat{\pi}_{11}$	$\hat{\pi}_{12}$	$\hat{\pi}_{13}$	$\hat{\pi}_{14}$
2	$\hat{\mu}_2$	$\hat{\pi}_{21}$	$\hat{\pi}_{22}$	$\hat{\pi}_{23}$	$\hat{\pi}_{24}$
3	$\hat{\mu}_3$	$\hat{\pi}_{31}$	$\hat{\pi}_{32}$	$\hat{\pi}_{33}$	$\hat{\pi}_{34}$
4	$\hat{\mu}_4$	$\hat{\pi}_{41}$	$\hat{\pi}_{42}$	$\hat{\pi}_{43}$	$\hat{\pi}_{44}$

4.6.2 *Pretreatment of the time series*

The time series undergo a preprocessing composed of the imputation of the negative stock, the imputation of the outliers, the imputation of the zero stock when a sale occurs and the rounding of the values. We explain now the reasons of such actions.

The imputation of the negative stock is zero if the day after or previous a sequence of negative stock is zero; else it equals to 1. It may introduce bias but it is a reasonable assumption that enables to capture more out-of-stocks which is the relevant signal. We chose to address this point by fixing the stock to a value of 1 since the information we need is only the availability of the product. The outliers may not be superior to the value of the 0.001 and 0.99 percentile in order to avoid some effects on the estimators.

Figure 11: Illustration of the selected products

(a) Putting knife

(b) set of tapes

(c) Padlock

Zero stocks when a sale occurs is a mistake in the database. The rounding is necessary because of the assumption of discreteness of the demand.

4.6.3 *Numerical application in the case of 3 putting knives and a set of tapes*

4.6.3.1 *Test of homogeneity*

Since our modelling assumes the homogeneity of the Poisson process in each configuration of availability, we evaluated the homogeneity assumption of the $K \times J_n$ Y_{nkj} on both the separated stores. For this, let $d_{kj1}, \ldots, d_{kjD_k}$ be the duration between sales of the product k on connected sections of $(0, n)$ where $\mathbf{u}(t) = \omega_j$. We conducted a test of Kolmogorov to compare the cumulated inter arrival time to the uniform law over each store and each configuration of availability. Since there are $K \times J_n \times \mathcal{U}$ tests, we deal with the multiplicity by applying a correction using the false discovery rate correction explained in Benjamini and Hochberg, 1995. Table 13 shows that most of the configurations validate the assumption, even if not all.

4.6.3.2 *Application of the EM on short period of time and single store*

Table 14 shows the matrix $\hat{\theta}$ of the EM's results. We can see that structure emerges from the estimators. The 16 cm does substitute with the 14 cm and not with the 12

Table 13: Homogeneity test on all stores for putting knives and the set of tapes

product	significant	nb config
16 cm	0	350
	1	15
14 cm	0	391
	1	16
12 cm	0	328
	1	15
Tapes	0	427
	1	77

cm and the tapes. The 14 cm substitutes both with the 16 and the 12 cm. The 12 cm does not substitute with the 16 cm. However we see an error of estimation: the 12 cm does substitute according to the data with the tapes. The substitution rate from the set of tapes to the other products is low (< 0.12) and most of it is lost which is coherent. Results show that the mean frequency of sale of the 12 cm is less sold than the 14 cm and 16 cm and the tapes. It also points out that the error may be due to the low substitution rate of the 12 cm.

Table 14: $\hat{\theta}$ on a single store

	$\hat{\mu}_k$	16 cm	14 cm	12 cm	Tapes
16 cm	0.22	0.42	0.42	0.16	0.00
14 cm	0.21	0.73	0.01	0.26	0.00
12 cm	0.11	0.00	0.49	0.01	0.49
Tapes	0.54	0.03	0.03	0.06	0.88

Interpretation: The substitution rate from the 16 cm to the 14 cm is 0.42.

The error reveals an impact of the randomness in the time series on the quality of the estimation.

A random split of the dataset into two parts, each of about 180 days, and an estimation of the parameters yield Tables 15 and 16. Since the mean frequency of product sales is close to 0.1 and $n = 180$, the estimators exhibit high variance, as discussed in Section 4.5.2.1. We observe differences between the estimators and the one estimated with $n = 365$. Specifically, $\hat{\pi}_{32}$ has values of 0 for the first subdataset and 1 for the second. The time series of the stock of the putting knives and the sales of the set of tapes are shown in Figure 12. In the first subdataset, sales happen to be concentrated during the out-of-stock period, while in the second, they are not, leading to the disparity between the estimators. Hence we isolate two factors of error: the variance associated to the amount of information (μ_k, n) and the interaction of the time series of sales and out-of-stock periods. In the latter, it is especially influenced when the period of out-of-stock is short.

Table 15: $\hat{\theta}$ on Subdataset 1

	$\hat{\mu}_k$	16 cm	14 cm	12 cm	Tapes
16 cm	0.18	0.71	0.29	0.00	0.00
14 cm	0.19	0.90	0.00	0.10	0.00
12 cm	0.11	0.00	0.00	0.34	0.66
Tapes	0.22	0.26	0.65	0.09	0.00

Interpretation: The substitution rate from the 16 cm to the 14 cm is 0.369.

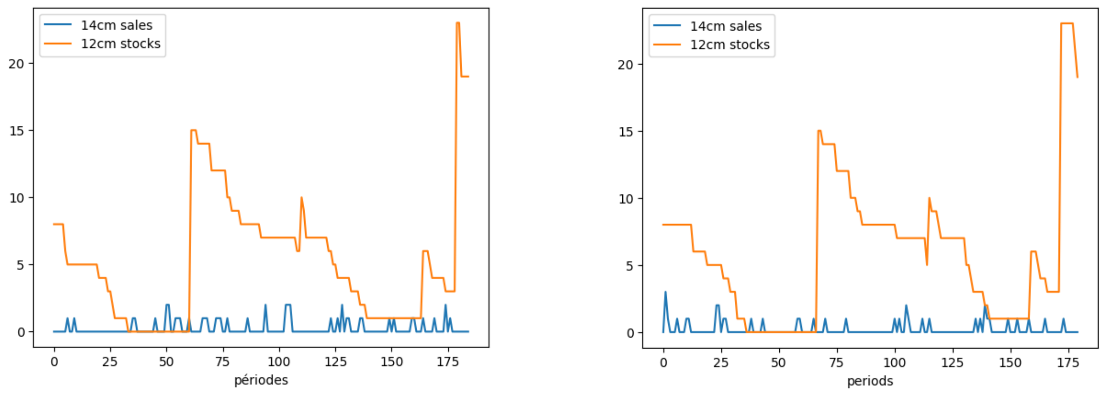
Table 16: $\hat{\theta}$ on Subdataset 2

	$\hat{\mu}_k$	16 cm	14 cm	12 cm	Tapes
16 cm	0.23	0.62	0.38	0.00	0.00
14 cm	0.25	0.00	0.90	0.11	0.00
12 cm	0.09	0.00	1.00	0.00	0.00
Tapes	0.32	0.02	0.00	0.06	0.92

Interpretation: The substitution rate from the 16 cm to the 14 cm is 0.729.

Remark 7 *The values of substitution are estimated in a situation with sparse data so it suffers from a high variability. Moreover an unrelated product may have sales deviating from its normal frequency over a period where another product is out-of-stock.*

Figure 12: Time series of stocks and sales for two subdatasets



(a) Subdataset 1 (b) Subdataset 2

A natural question can the be the following: do the statistical tests in this example provide faith to the values of substitution between putting knives and back up the intuition of non substituability between the 12 cm and the set of tapes? We now try to answer to it.

4.6.3.3 *Application of the EM based on multiple stores*

Having more data is expected to provide a better estimation hence we provide two experiences based on 150 stores. The first experience relies on aggregation of the estimators and the second on the aggregation in a single time series. The demand could be different from one store to the other and introduce bias in the estimators. In this section, the set of stores is refered as $\mathcal{U}$ and an estimator over a store is $\hat{\theta}^{(u)}$.

The first estimator is defined as $\tilde{\theta} = \frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}}\hat{\theta}^{(u)}$ with $\mathcal{U}$ being the set of stores. Table 17 shows that the associated variance is high. The spillover despite the high variance shows the *desired structure* among the putting knives. However the spillover from the putting knives to the set of tapes is high. Note that the set of tapes is sold in higher amounts than the putting knives. According to Section 4.5.1.5 the estimators are impacted by it, hence the conclusion is similar and the set of tapes do not substitute to the other products whereas the putting knives substitute wrongly to the set of tapes .

For the second experience, we first chose to concatenate all the time series of sale and stocks with the 4 concerned products. Table 18 shows that a large proportion of the availability configurations do not respect the assumption of homogeneity.

According to Table 19, the EM procedure retrieved exactly the group of substitution associated to the putting knife and the singleton of the set of tapes. However the *desired structure* is not apparent since the 16 cm substitutes more with the 12 cm than with the 14 cm.

Table 17: Mean and standard deviation of the aggregated estimators

	$\hat{\mu}_k$	16 cm	14 cm	12 cm	Tapes
16 cm	0.30 (0.20)	0.28 (0.37)	0.31 (0.33)	0.16 (0.25)	0.24 (0.36)
14 cm	0.28 (0.16)	0.21 (0.32)	0.28 (0.37)	0.24 (0.32)	0.26 (0.38)
12 cm	0.21 (0.12)	0.15 (0.27)	0.34 (0.39)	0.27 (0.38)	0.23 (0.36)
Tapes	2.29 (1.45)	0.04 (0.10)	0.03 (0.06)	0.01 (0.03)	0.91 (0.16)

Interpretation: The substitution rate from the 16 cm to the 14 cm is 0.31.

Table 18: Homogeneity test on pooled data for putting knife and the set of tapes

product	significant	nb config
16 cm	0	4
	1	3
14 cm	0	4
	1	4
12 cm	0	3
	1	5
Set of tapes	0	3
	1	5

4.6.4 *Numerical application in the case of 3 putting knives and a padlock*

The dataset is composed now of the putting knives and the padlock over 152 stores. We first introduce the estimators over single stores and the pooled data which is

Table 19: Numerical results for the pooled stores

	$\hat{\mu}_k$	16 cm	14 cm	12 cm	Tapes
16 cm	0.31	0.00	0.19	0.80	0.01
14 cm	0.22	0.00	0.81	0.19	0.00
12 cm	0.28	0.08	0.00	0.92	0.00
Tapes	2.41	0.00	0.00	0.01	0.99

Interpretation: The substitution rate from the 16 cm to the 14 cm is 0.19.

composed of a fictional store with all the time series aggregated. Table 20 shows that most of stores availability configurations verifies the homogeneity characteristic.

The experience over the single stores yields Table 21 which shows a similar pattern as for the previous dataset, but the variance of the estimators is high. The signal of substitution from the padlock to the putting knives is coherent with a no substitution situation. However some substitution from the putting knives to the padlock is wrongly detected.

The homogeneity test over configurations of availability in the case of the pooled time series yields Table 22 which confirms that the assumption of homogeneity of the Poisson process is not verified in most cases. Pooling the data in a single time series however yields Table 23. The EM retrieves the signal of substitution between the putting knives and no substitution from the putting knives to the padlock. Moreover the substitution from the 14 cm and 12 cm to the other products is coherent with a high probability of lost sale and a substitution coherent with the *desired structure*. There is however a problem of ordering, the putting knife of 16 cm should substitute to the 14 cm and does it to the 12 cm and that with a high probability. This experience suggests that pooling the data yields mostly coherent results as in the case of the previous dataset in spite of not respecting the homogeneity of sales on availability configurations.

Table 20: Homogeneity test on all stores for putting knives and the padlock

product	significant	nb configurations
16 cm	0	407
	1	14
14 cm	0	376
	1	16
12 cm	0	463
	1	13
Padlock	0	286
	1	15

Table 21: Mean (standard deviation) of $\tilde{\theta}$

	$\hat{\mu}_k$	16 cm	14 cm	12 cm	Padlock
16 cm	0.3 (0.18)	0.28 (0.36)	0.29 (0.31)	0.18 (0.27)	0.24 (0.33)
14 cm	0.27 (0.15)	0.22 (0.31)	0.37 (0.38)	0.24 (0.3)	0.16 (0.28)
12 cm	0.21 (0.10)	0.15 (0.25)	0.38 (0.37)	0.29 (0.37)	0.17 (0.29)
Padlock	0.41 (0.25)	0.03 (0.07)	0.07 (0.15)	0.04 (0.09)	0.84 (0.23)

Interpretation: The mean substitution rate from the 16 cm to the 14 cm is 0.31.

Table 22: Homogeneity test on pooled data for putting knives and the padlock

product	significant	nb config
16 cm	0	4
	1	4
14 cm	0	3
	1	5
12 cm	0	3
	1	5
Padlock	0	5
	1	3

Table 23: Numerical applications for the pooled stores

	$\hat{\mu}_k$	16 cm	14 cm	12 cm	Padlock
16 cm	0.31	0.00	0.27	0.73	0.00
14 cm	0.22	0.00	0.88	0.12	0.00
12 cm	0.29	0.08	0.09	0.83	0.00
Padlock	0.41	0.00	0.00	0.00	1.00

Interpretation: The substitution rate from the 16 cm to the 14 cm is 0.23.

5 SPARSE SUBSTITUABILITY MODELLING IN CASE OF A LARGE NUMBER OF PRODUCTS

As exhibited in the previous chapter, the proposed model for product substituability estimation suffers from both complexity in time of the computations and also from the identifiability problems, as soon as K increases. In particular, it seems difficult to exceed 10 product (sometimes even 5), thus strongly limiting the practical interest of our model since a full assortment is composed . Consequently, we provide here an additional modelling that builds upon the model of the previous section. We introduce a sparse version of the model (37) that builds upon the natural numerous non substitutability that occurs in real situations. This sparsity assumption is then reformulated within the clustering paradigm, where obtained groups of products are non substitutable to each other.

5.1 THE PROPOSED SPARSE MODELLING

We recall that if the number of products K is large, it is unlikely that the sub-matrix $\mathbf{S}_k$ has rank K for any $k \in \{1, \ldots, K\}$ as required by Assumptions 8 (see the previous chapter) and thus there is no guarantee on the identifiability of the model parameters. In return, when the sales of many products are considered (*i.e.*, K is quite large), it is likely that many products are strongly different and so cannot be substituted. It leads to the natural assumption that many substituability product probabilities $\pi_{\ell k}$ are null. Therefore, in this section, we introduce a sparse version of sale model considering product substituability, in order to take into account that many products (most of them in fact. . .) cannot be substituted by all of them.

More precisely, we consider B groups of products such that two products belonging to the same group of products can be substituted while two products belonging to different groups of products cannot be substituted. The group of products k is indicated by the binary vector $\boldsymbol{\gamma}_k = (\gamma_{k1}, \ldots, \gamma_{kB})^\top$ where $\gamma_{kb} = 1$ if product k belongs to group b and $\gamma_{kb} = 0$ otherwise ($k = 1, \ldots, K$, $b = 1, \ldots, B$). We consider that each product belongs to exactly one group. The information about the group of products is stored into the $B \times K$ matrix $\boldsymbol{\gamma} = [\boldsymbol{\gamma}_1 \ldots \boldsymbol{\gamma}_K]$. We consider then the sparse simplex of size K indexed by $\boldsymbol{\gamma}$ and defined by

$$S_K(k; \boldsymbol{\gamma}) = \left\{ \mathbf{a} :\in S_K, \forall \ell \in \{\ell : \boldsymbol{\gamma}_k^\top \boldsymbol{\gamma}_\ell = 0\}, a_\ell = 0 \text{ and } a_k = 1 \text{ iif } \forall \ell \neq k, \boldsymbol{\gamma}_k^\top \boldsymbol{\gamma}_\ell = 0 \right\}.$$

The proposed sparse sale model considers that each vector defining the probability of reporting the initial demand of product ℓ to another product belongs to the sparse simplex of size K indexed by $\boldsymbol{\gamma}$, thus such that $\boldsymbol{\pi}_\ell \in S_K(\ell; \boldsymbol{\gamma})$. Hence, the resulting matrix $\boldsymbol{\gamma}$ defines a sparse version of the seminal sale model of product substituability since it imposes many probabilities $\pi_{\ell k}$ to be equal to zeros. Consequently, the initial model defined by (37) is crossed with the additional constraints on the parameter space $\boldsymbol{\theta} \in \Theta_{\boldsymbol{\gamma}}$, where

$$\Theta_{\boldsymbol{\gamma}} = (\mathbb{R}^{+*})^K \times S_K(1; \boldsymbol{\gamma}) \times \ldots \times S_K(K; \boldsymbol{\gamma}).$$

These parsimonious constraints can be summarized by the undirected graph $G = (V, E)$ where $V = \{1, \dots, K\}$ is the set of vertices (products) and where the edges E are composed by the couples of products that can be substituted leading that $E = \{(k, \ell) : \pi_{k\ell} + \pi_{\ell k} > 0\}$. Thus, a model is defined by the components of G since $\gamma_k^\top \gamma_\ell = 1$ if and only if the vertices k and ℓ belong to the same component of the graph. It thus leads to groups of substituability, that we need to estimate now.

5.2 ESTIMATING THE GROUPS OF SUBSTITUABILITY

5.2.1 *Invoking a specific HAC algorithm for estimating the groups*

A model γ is defined by a partition of the K products, leading to possible intra-group substituability, but impossible inter-group substituability. We now motivate and explain the choice of a specific constrained hierarchical agglomerative clustering (HAC) allowing to estimate such a group structure in an efficient way.

Estimating the groups of products that can be substituted will rely on the computation of an estimator of G defined, for any positive value $\tau_n \in \mathbb{R}^+$, by $\widehat{G}_{n,\tau_n} = (V, \widehat{E}_{n,\tau_n})$ where $\widehat{E}_{n,\tau_n}$ is the estimator of the vertices defined by

$$\widehat{E}_{n,\tau_n} := \{(k, \ell) \in \{1, \dots, K\} \times \{1, \dots, K\},\ \Delta_{nk\ell} + \Delta_{n\ell k} > \tau_n\}.$$

Then, the following proposition shows that if the threshold τ_n follows an appropriate rate, then $\widehat{E}_{n,\tau_n}$ consistently estimates E. Hence, the groups of G, and so the model, can be consistently estimated by the groups within $\widehat{G}$.

Conjecture 2 *Under Assumptions 9, 10, if τ_n tends to zero as n tends to infinity and that $\tau_n n^{1/2}$ tends to infinity as n tends to infinity then*

$$\lim_{n\to\infty} \mathbb{P}(\widehat{E}_{n,\tau_n} = E) = 1.$$

An idea of the proof is provided in the Appendix C.1.

Proposition 2 ensures the consistency of the estimator of γ, as well as $\tau_n = Cn^{-\alpha}$ with $C > 0$ and $0 < \alpha < 1/2$. However, for a finite time n, the choices of C and α can impact the selected model. In othert words, coefficients $\Delta_{nk\ell} + \Delta_{n\ell k}$ for $(\ell, k) \in [K]^2$ retrieve a signal of substitution which, combined with a clustering algorithm, is able to provide a set of candidate partitions γ. We retain an HAC for this task, and more precisely a *constrained* HAC process for preserving the model identifiability as we describe now.

The HAC paradigm relies on a dissimilarity matrix D for providing a set of K possibles embedded models $\gamma^{(1)}, \dots, \gamma^{(K)}$, corresponding to K sequential partitions. Usually, it enables an acceptable time of execution. However, we need to constrain the proposed partitions by the identifiability requirement of the parameters, as described in more detail in Appendix C.2. The whole algorithm is also available at the same place.

The dissimilarity matrix matrix we retain is obtained by the following transformation of the matrix $\Delta := [\Delta_{nk\ell} + \Delta_{n\ell k}]_{k=1,\dots,K}^{\ell=1,\dots,K}$, defining $D(i,j) = \max(\bar{\Delta}) - \bar{\Delta}(i,j)$, $\forall (i,j) \in [K]^2$ so that $i \neq j$ and $D(i,i) = 0$ with $\bar{\Delta} = \Delta - \min(\Delta)$. Concerning now the aggregation criterion, we provide a numerical analysis in Section 5.3.1.2 for choosing within the different classical ones (Ward, single and complete linkage).

5.2.2 *A non asymptotic model selection method*

Our estimation aims at maximizing the *data-observed log-likelihood* hence we consider quite naturally a selection process of the model $\widehat{\gamma}$ based on a penalized log-likelihood criterion. In this context, BIC and AIC are common criteria for model selection. However, we have a doubt on the definition of the asymptotic quantities which are involved in the specific case of our model. This is the reason why, we provide only a heuristic of BIC (Schwarz, 1978) and we decide to locate it in Appendix C.3 with related numerical applications.

However, contrary to BIC, there exists some non asymptotic criteria avoiding this difficulty. It is the reason why we chose to use mainly the heuristic slope paradigm initially proposed by Massart (Birgé and Massart, 2007, Baudry et al., 2010), which is a penalized method for the selection of the best model with the interesting property to be non asymptotic. This related penalization is known up to a constant C and its shape should to be of the form $C\frac{\nu_{\gamma^{(k)}}}{2}$ as successfully studied in various situations by Arlot, 2019. In our case, this criterion becomes

$$\widehat{\gamma} = \text{argmax}_{k=1,\ldots,K} \log L(\widehat{\theta}_{\gamma^{(k)}}; \gamma^{(k)}, y_n | x_n) - C\, \frac{\nu_{\gamma^{(k)}}}{2},$$

where $\nu_{\gamma^{(k)}} = K + \sum_{k'=1}^{K} \sum_{\ell \neq k'} \gamma_k^{(k)\top} \gamma_\ell^{(k')}$ is the number of parameters involved by model γ, $\widehat{\theta}_\gamma$ is the maximum likelihood estimate (see Section 4.4.1 for details on its estimation). To the log-likelihood $\log L$ introduced in Section 4.3, a new parameter γ is added that expresses the sparsity of the parameter space.

In practice, there exists some methods to automatically detect the elbow of the log-likelihood (see Arlot, 2019). However, in our work, we prefer to visually detect the change of slope of the function $(\nu_{\gamma^{(k)}}, \log L(\widehat{\theta}_{\gamma^{(k)}}; \gamma^{(k)}, y_n | x_n))$. It corresponds to a heuristic implementation of the seminal method of Massart but which works well as we will illustrate below along our experiments.

5.3 NUMERICAL APPLICATION ON SYNTHETIC AND REAL DATASETS

5.3.1 *Evaluation of the model selection on synthetic data*

5.3.1.1 *Realistic synthetic data scenarios in the Adeo company context*

We provide scenarios close to the situation at Adeo in order to assess the quality of the model selection on larger groups of products. Similar to the scenarios defined in the previous chapter, let $\tilde{S} = (B_5, B_2, n, \mu, a, c)$ be scenarios where B_5, B_2 represent respectively the number of groups of 5 and 2 substitutable products. The total number of products is then $K = 5B_5 + 2B_2$. We chose the scenarios where $(B_5, B_2) \in \{(10, 50), (50, 10), (100, 0), (200, 0)\}$. $\pi_{\ell k}$ is fixed based on the number of products in a substitution group and the choice of a. Obviously, $\pi_{\ell k} = 0$ for all products ℓ and k belonging to different groups. We chose to consider a time frame $n \in \{10^3, 10^4, 10^5\}$ where $\mathbf{u}$ from Section 4.3 is generated by subdividing $[0, n]$ in $1,000$ periods and choosing the configuration of availability uniformly from the 2^K possible configurations. Additionally, we chose $\mu = 2$ for the mean frequencies of sale.

Each scenario defines a unique $\theta^{\tilde{S}}$ which generates N_{samp} samples $(\mathbf{x}_n^i, \mathbf{y}_n^i, \boldsymbol{\omega}^i)_{i \in [N_{samp}]}$. For instance, the matrix of substitution associated to $\tilde{S}$ such that $(B_5, B_2, a) = (0, 2, \textit{unif})$ is

$$\boldsymbol{\pi}_i^{\tilde{S}} = \begin{pmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0.5 & 0.5 \end{pmatrix}. \tag{41}$$

The true partition of products in groups is denoted by $\boldsymbol{\gamma}^{\tilde{S}}$ as described in Section 5.1.

The model selection proceeds first in a proposal of at most K partitions $(\boldsymbol{\gamma}_i^{(k),\tilde{S}})_{k \in [K]}$ of products in subgroups of sustitutability via the constrained HAC provided in Appendix C.2. The EM applied to the subgroups associated to $\boldsymbol{\gamma}_i^{(k),\tilde{S}}$ provides the estimators $\hat{\theta}^{\tilde{S}}_{i,\boldsymbol{\gamma}_i^{(k),\tilde{S}}}$. Finally a selection of the best partition is made with the slope heuristic method we described before.

5.3.1.2 *Evaluation of the HAC without model selection*

In order to evaluate the quality of the HAC and chose within single, Ward and complete linkage we chose a subset of the scenarios $\tilde{S}$. This expriment does not include the model selection and the application of the EMs. The subset of the scenarios is restricted to $n = 10^4$ where the groups of substitutable products are composed of groups of 5 products and a varying number of groups $B_5 \in \{10, 50, 100, 200\}$. We computed only one sample ($N_{samp} = 1$) for each $\tilde{S}$, since the different number of groups we study already gives a sufficient idea about the HAC variability estimation for our present purpose.

We have computed the ARI for each partition produced by the HAC and retrieved the one with the highest. In Table 24, 25, the ARI is perfect for both the single and the Ward linkage. Table 26 associated to the complete linkage "suffers" in the case of heterogeneous mean frequency of sales. This application gives us confidence in the constrained HAC and we chose the Ward linkage in the following sections.

We then conducted a full set of experiences where $N_{samp} = 10$ with Ward linkage in the case of realistic probabilities of substitution $a_k = \textit{realistic}$. Table 27 shows that the clustering quality in terms of ARI for $n \in \{10^4, 10^5\}$ is high. For the same values of n, the ARI decreases slightly from $c = 1$ to 10 and for increasing K where c defined in the previous chapter is the multiplicative coefficient defining the heterogeneity of sale within a group. The ARI for $n = 10^3$ drops significantly from $B_5 = 10$ to 200.

In terms of proportion of real groups retrieved, values are close to 1 until and including $B_5 = 50$. For homogeneous mean frequencies of sales ($c = 1$) all the groups are retrieved whereas for $c = 10$ the values drop to 0.53 for $n = 10^4$. It thus suggests that the heterogeneous mean frequency of sales is detrimental to the estimation performance. In the case of $n = 10^3$ the best scenario for $B_5 = 10$ is at 0.51 and for $B_5 = 50$ no groups are retrieved. This experience provides confidence in the HAC and concludes additionnally that $n = 10^3$ for $\mu = 2$ is not sufficient in terms of amount of data.

Table 24: HAC with single linkage

a	B_2	n	K	B_5	c	Proportion group retrieved	ARI
realistic	0	10000	50	10	1	1.	1.
					10	1.	1.
			250	50	1	1.	1.
					10	1.	1.
			500	100	1	1.	1.
					10	1.	1.
			1000	200	1	1.	1.
					10	1.	1.

Table 25: HAC with Ward linkage

a	B_2	n	K	B_5	c	Proportion group retrieved	ARI
realistic	0	10000	50	10	1	1.	1.
					10	1.	1.
			250	50	1	1.	1.
					10	1.	1.
			500	100	1	1.	1.
					10	1.	1.
			1000	200	1	1.	1.
					10	1.	1.

5.3.1.3 *Evaluation of model selection on synthetic data*

The quality of the model selection is now evaluated by retrieving the best partition using the slope heuristic method. The change of slope of the *data-observed log-likelihood* corresponds to the retained estimate $\hat{\gamma}$. However, we have to note some specific cases where this elbow in the slope is not present. For instance, in Figure 18 there is no significant change of the slope because the limit number of the groups in the HAC avoids to visualize it. In that case, we retain the beginning of the curve at the first group (100 groups), corresponding to the maximum number of groups available under the constraint HAC. In addition, note that, strictly speaking, the slope criterion should be based upon the change of the slope of the *data-observed log-likelihood* as a function of the number of the parameters of the model associated to γ. However, we prefer to use instead the number of groups in the horizontal axis since it is more explicit for the reader while having no consequence on the results (the number of parameters is a decreasing function of the number of groups). We illustrate this property in Figure 13 and 14(a).

Table 26: HAC with complete linkage

a	B_2	n	K	B_5	c	Proportion group retrieved	ARI
realistic	0	10000	50	10	1	1.	1.
					10	0.5	0.90
			250	50	1	1.	1.
					10	0.26	0.76
			500	100	1	1.	1.
					10	0.25	0.77
			1000	200	1	1.	1.
					10	0.22	0.74

Each partition is chosen based on Figure 14, 15, 16, 17, 18 which show the evolution of the log-likelihood for homogeneous mean frequency of sales in (a) and the heterogeneous mean frequencies of sales in the subfigure (b). The selected partition is compared to the true groups using the ARI (adjusted rand index) (Hubert and Arabie, 1985). Then, best groups are retrieved based on the evaluated $\hat{\pi}_{\ell k}^{\tilde{S}}$ according to a criterion such as the group with the highest substitution probability. The substitution in groups can be seen as an oriented weighted graph with nodes as products and edges weighted by the substitution rate.

According to Table 28, the selection of model is perfect from the standpoint of the proportions of groups retrieved and the ARI both for heterogeneous and homogeneous frequencies of sales (c) and for different sizes of groups.

Figure 13: Data-observed log-likelihood for $\tilde{S} = (10, 50, \textit{perfect substitution}, c)$
Degrees of freedom of the true partition: 600

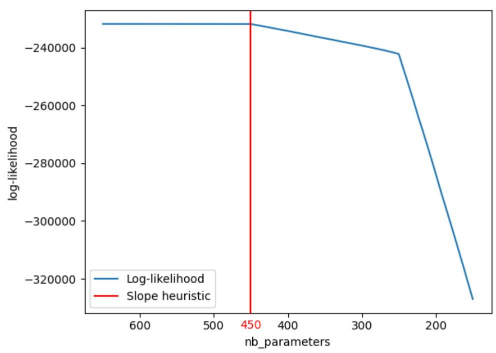
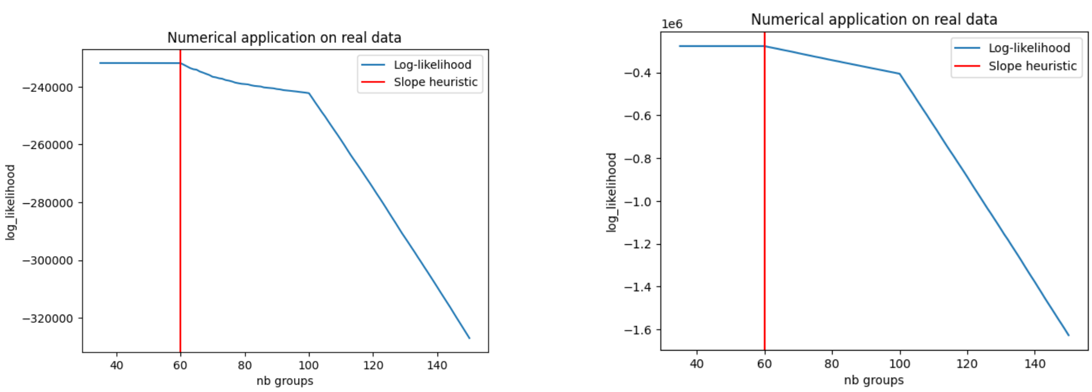


Table 27: Evaluation of the HAC using Ward's linkage

a	B_2	K	B_5	n	c	Proportion group retrieved (std)	ARI (std)
realistic	0	50	10	1000	1	0.37 (0.23)	0.76 (0.13)
					10	0.51 (0.2)	0.87 (0.06)
				10000	1	1.0 (0.00)	1.0 (0.00)
					10	0.92 (0.13)	0.98 (0.02)
				100000	1	1.0 (0.00)	1.0 (0.00)
					10	1.0 (0.00)	1.0 (0.00)
		250	50	1000	1	0.04 (0.03)	0.42 (0.05)
					10	0.09 (0.03)	0.65 (0.03)
				10000	1	1.0 (0.00)	1.0 (0.00)
					10	0.86 (0.08)	0.98 (0.01)
				100000	1	1.0 (0.00)	1.0 (0.00)
					10	0.87 (0.09)	0.98 (0.02)
		500	100	1000	1	0.02 (0.01)	0.32 (0.04)
					10	0.07 (0.02)	0.6 (0.03)
				10000	1	1.0 (0.00)	1.0 (0.00)
					10	0.69 (0.07)	0.94 (0.01)
				100000	1	1.0 (0.00)	1.0 (0.00)
					10	0.86 (0.08)	0.98 (0.01)
		1000	200	1000	1	0.0 (0.00)	0.21 (0.02)
					10	0.04 (0.01)	0.53 (0.02)
				10000	1	1.0 (0.00)	1.0 (0.00)
					10	0.53 (0.04)	0.9 (0.01)
				100000	1	1.0 (0.00)	1.0 (0.00)
					10	0.73 (0.04)	0.95 (0.01)

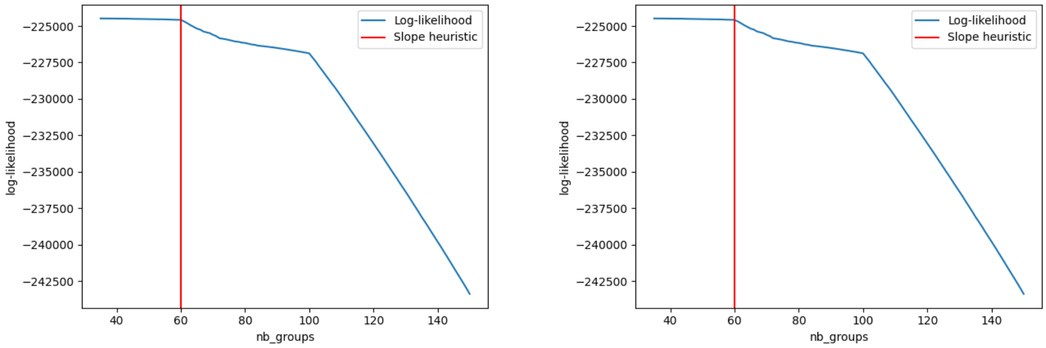
Figure 14: Data-observed log-likelihood for $\tilde{S} = (10, 50, \textit{perfect substitution}, c)$



(a) $\tilde{S}$ with $c = 1$
True number of groups: 60

(b) $\tilde{S}$ with $c = 10$
True number of groups: 60

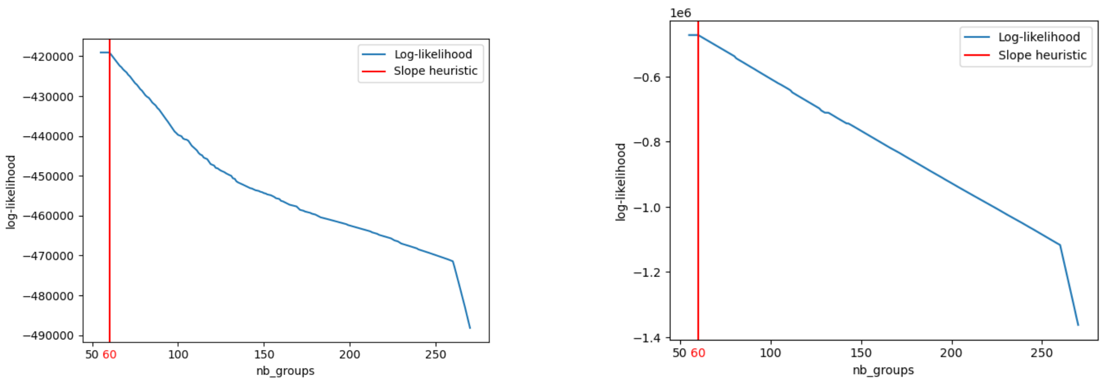
Figure 15: Data-observed log-likelihood for $\tilde{\mathcal{S}} = (10, 50, \textit{realistic}, c)$



(a) $\tilde{\mathcal{S}}$ with $c = 1$
True number of groups: 60

(b) $\tilde{\mathcal{S}}$ with $c = 10$
True number of groups: 60

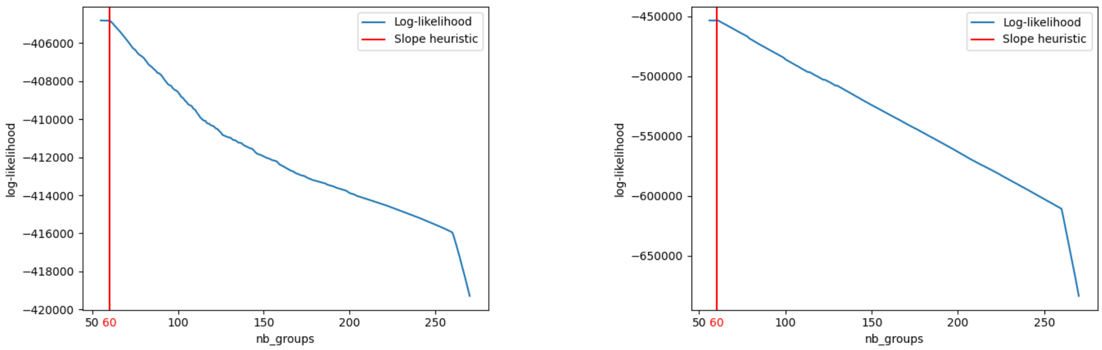
Figure 16: Data-observed log-likelihood for $\tilde{\mathcal{S}} = (50, 10, \textit{perfect substitution}, c)$



(a) $\tilde{\mathcal{S}}$ with $c = 1$
True number of groups: 60

(b) $\tilde{\mathcal{S}}$ with $c = 10$
True number of groups: 60

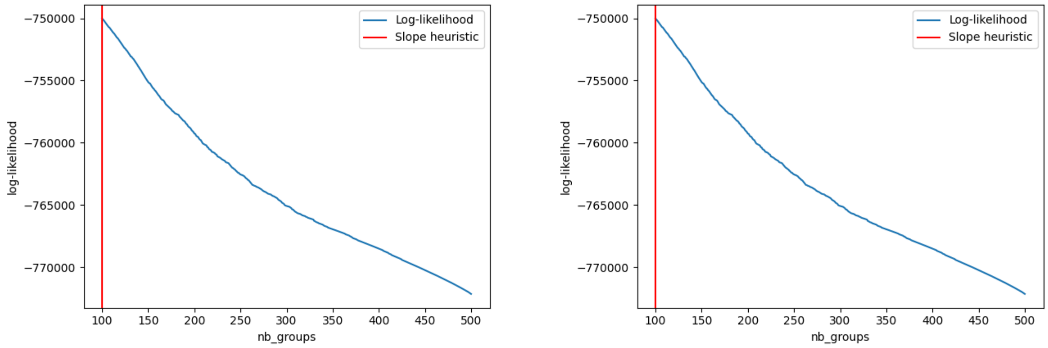
Figure 17: Data-observed log-likelihood for $\tilde{\mathcal{S}} = (50, 10, \textit{realistic}, c)$



(a) $\tilde{\mathcal{S}}$ with $c = 1$
True number of groups: 60

(b) $\tilde{\mathcal{S}}$ with $c = 10$
True number of groups: 60

Figure 18: Data-observed log-likelihood for $\tilde{S} = (100, 0, \textit{realistic}, c)$



(a) $\tilde{S}$ with $c = 1$
True number of groups: 100

(b) $\tilde{S}$ with $c = 10$
True number of groups: 100

5.3.2 *Application on Adeo data*

5.3.2.1 *Description of the four real datasets*

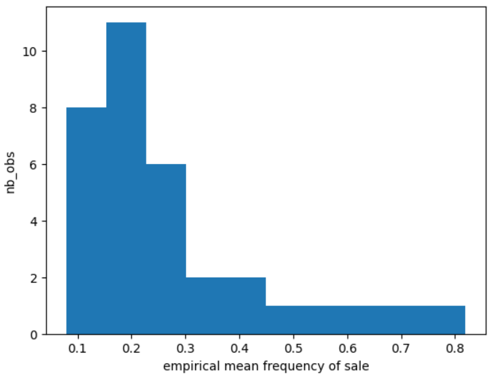
The objective for Adeo is to retrieve the probability of substitution at scale. A common store references about 40,000 products. In our experiments, the products were chosen such that to limit impacts on the mean frequency of sales that are not related to the substitution. In particular, we select products not impacted by seasonality. Moreover, the price promotions have been identified in order to discard their impact on the sales.

In practice, Adeo's categories are a hierarchical structure in the order of the department, subdepartment, type and subtype (Category 1, 2, 3, 4). The retained types of products belong to the following distinct categories: paintings, plumbing and electricity, tools workshop, ironsmith. The products between two different types are obviously not substitutable which makes it possible to evaluate the quality of the partitioning proposed by our unsupervised method. We thus expect that substitutable products are in the same subtypes. We chose products that are sold in a sufficient quantity sold in quantity which is the category A and that are automatically replenished. We described now four datasets which follow globally all these requirements.

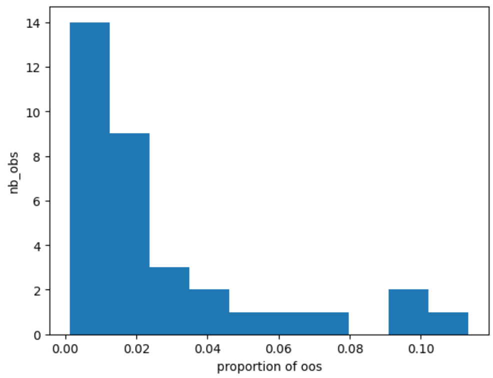
Dataset 1 is composed of the three putting knifes (16cm, 14cm, 12cm) and the set of tapes already studied in the previous chapter on all stores where they are available simultaneously.

In **Dataset 2**, 34 products in the hand tools department (Table 29) have been selected on a single store. We chose for this application a group of 70 products on a single store over a period of 1,613 days. Since there is changes in the assortment, the similarity matrix wasn't calculable, thus we provided the biggest subset of products. According to Figure 19, products have a mean frequency of sale under 0.8 per day and concentrated around 0.2. The proportion of out of stock periods is lower than 10% with most products around 2%.

Figure 19: Descriptive statistics of **Dataset 2**



(a) Mean frequency of sales



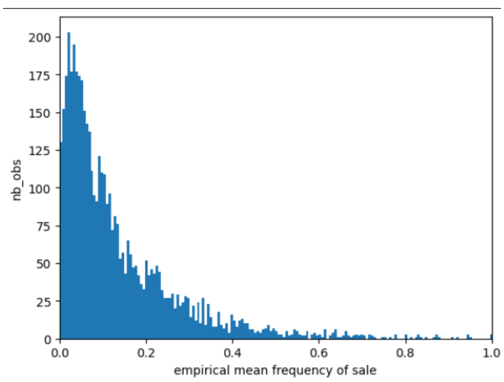
(b) Proportion of OOS

Table 29: Categories of **Dataset 2**

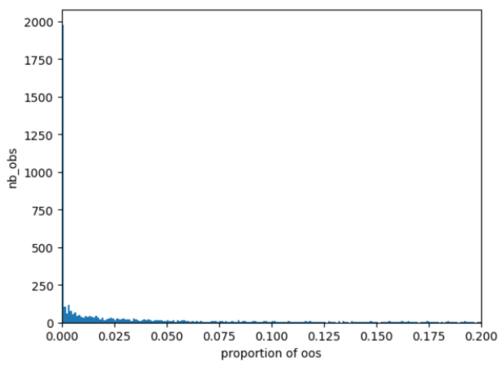
Category 1	Category 2	Category 3	Category 4	NUM_ART
OUTILLAGE	OUTILLAGE A MAIN	CLES ET DOUILLES	CLES A MOLETTES ET JEUX	8
			CLES PLATES ET JEUX	10
			CLES TUBES ET JEUX	2
			MALL. CLES, DOUILLES, ACCES.	2
		TOURNEVIS	T.VIS ELECTRICIEN GAINE	6
			T.VIS ELECTRICIEN NON GAINE	5
			TOURNEVIS MECANICIENS	1

Dataset 3 is composed of 33 products from several subcategories in the tools department over all stores of Leroy Merlin France. Because of the high assortment choice every products were not available simultaneously in every store. Consequently, we retained 3 stores for the first model selection where 20 of these proposed products were available. Detail of categories can be found in Table 30.

Figure 20: Descriptive statistics of **Dataset 3**



(a) Mean frequency of sales



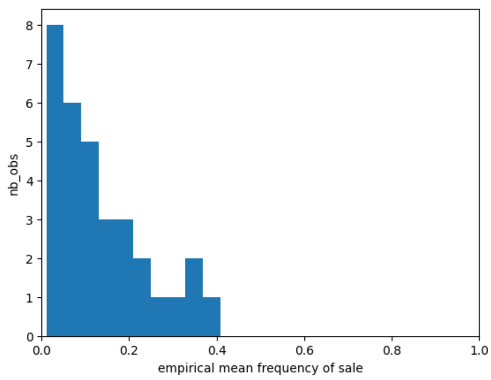
(b) Proportion of OOS

Table 30: Categories of **Dataset 3**

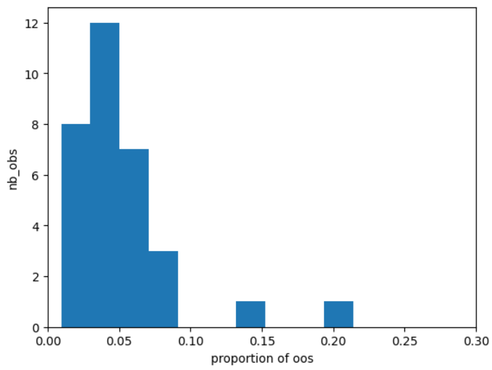
Category	Nb_prod Store 1	Nb_prod Store 2	Nb_prod Store 3
DEFONCEUSES	2	2	2
PERFORATEURS,BURINEUR,MARTEAU	7	10	7
RABOTS	2	1	2
SCIES SAUTEUSES	8	7	8
TOURNEVIS SANS FIL	1	0	1

Dataset 4 is a set of 32 products in the tools department with the information of all the store of Leroy Merlin France. This dataset is composed of 33 products from several subcategories in the tools department over all stores of Leroy Merlin France. We selected non substitutable products accross the finest category. The length of the pooled time series is 238,367. According to Table 21, the mean frequencies of sales is composed of slow movers and other products with a value $\mu_k = 0.4$ at its maximum. Additionnaly, the proportion of out of stock is around 5% for most of the products meaning that some substitution signal is detectable. It must be noted that since all products were not substitutable simultaneously, we inputed zeros in sales and stocks and this introduces biases since there is differences in terms of mean frequencies of sale that violate the assumption of homogeneity of the Poisson process underlying Y_{njk} (see Section 4.3).

Figure 21: Descriptive statistics of the **Dataset 4**



(a) Mean frequency of sales



(b) Proportion of OOS

Concerning the pre-processing of the raw time series, we imputed stock values of adjacent consecutive negative values by zero if the value before or after was zero and else we imputed a positive value. The other negative value were inputed by 1. We clipped the outliers at the quantiles 0.01 and 0.99. We kept all products that were in the intersection of the products within the stock data frame and the sale data frame. We fixed at one the stock of products that were sold on the day. In the case of the pooled time series of stores for which the assortment did not match the selected products we have added zeros both for the stocks and the sales. However, this imputation may have a compromising effect between the amount of data and an introduction of bias. We will discuss later this possible consequence on results.

5.3.2.2 *Application to* ***Dataset 1***

In this section, we apply the model selection to the putting knifes and the tape. We estimated $\hat{\rho}_k = \frac{1}{n}(\sum_{j=1}^{J} x_{nj}\omega_{jk})$ and $\hat{\tau}_j = \prod_{j=1}^{K} \hat{\rho}_k^{\omega_{jk}}(1-\hat{\rho}_k)^{1-\omega_{jk}}$ Assuming that τ_j verifies Equation 39 in Assumption 9 and that each day of the n days has a probability of ρ_k to have the product k available then $X_{nj} \sim \mathcal{B}(n, \tau_j)$, we applied a test of equality of the proportions with $H_0 : \hat{\tau}_j = \tau_j$ against $H_1 : \hat{\tau}_j \neq \tau_j$. The statistics $\frac{X_{nj}/n - \hat{\tau}_j}{\sqrt{\hat{\tau}_j(1-\hat{\tau}_j)/n}}$ shown in Table 31 are all rejecting the null hypothesis and the assumption is not verified for all the configurations of availability X_{nj}. As a consequence the results of the numerical experiments on real datasets may be biased by Assumption 9.

The model selection applied to the real data shows that the HAC with Ward's linkage succeeds in retrieving the right partition (Figure 22).

Table 31: Statistical test of the proportion of availability

1	2	3	4	X_{nj}/n	$\hat{\tau}_j$	Statistics	p-value
1	1	1	1	0.887	0.842	59.166	1.000
1	1	1	0	0.043	0.057	-29.359	0.000
1	0	1	1	0.010	0.025	-45.974	0.000
0	1	1	1	0.022	0.040	-44.410	0.000
1	1	0	1	0.012	0.026	-42.118	0.000
1	1	0	0	0.000	0.002	-14.752	0.000
1	0	1	0	0.001	0.002	-10.977	0.000
1	0	0	1	0.001	0.001	4.322	1.000
0	1	1	0	0.005	0.003	24.260	1.000
0	0	1	1	0.002	0.001	7.605	1.000
0	0	0	1	0.001	0.000	100.178	1.000
0	1	0	1	0.001	0.001	0.014	0.505
0	1	0	0	0.000	0.000	3.468	1.000
0	0	1	0	0.000	0.000	1.500	0.933
1	0	0	0	0.000	0.000	12.520	1.000
0	0	0	0	0.014	0.000	4123.293	1.000

Interpretation: Product 1 (16cm), Product 2 (14cm), Product 3 (12cm), Product 4 (tapes)

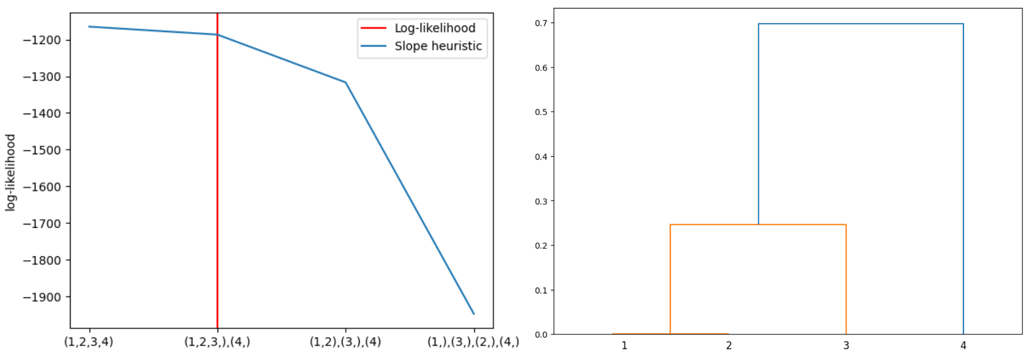
$$D = \begin{bmatrix} 0.000 & 0.000 & 0.217, & 0.283] \\ 0.000 & 0.000 & 0.210 & 0.676] \\ 0.217 & 0.210 & 0.000 & 0.682] \\ 0.283 & 0.676 & 0.682 & 0.000 \end{bmatrix} \tag{42}$$

is obtained from the similarity matrix

$$\Delta = \begin{bmatrix} - & 0.292 & 0.074 & 0.009 \\ 0.292 & - & 0.082 & -0.384 \\ 0.074 & 0.0818 & - & -0.391 \\ 0.009 & -0.384 & -0.391 & - \end{bmatrix} \tag{43}$$

(where the "-" stands for the values that can't be computed).

Figure 22: **Dataset 1** Selection model for Ward linkage
Interpretation: Product 1 (16cm), Product 2 (14cm), Product 3 (12cm), Product 4 (tapes)

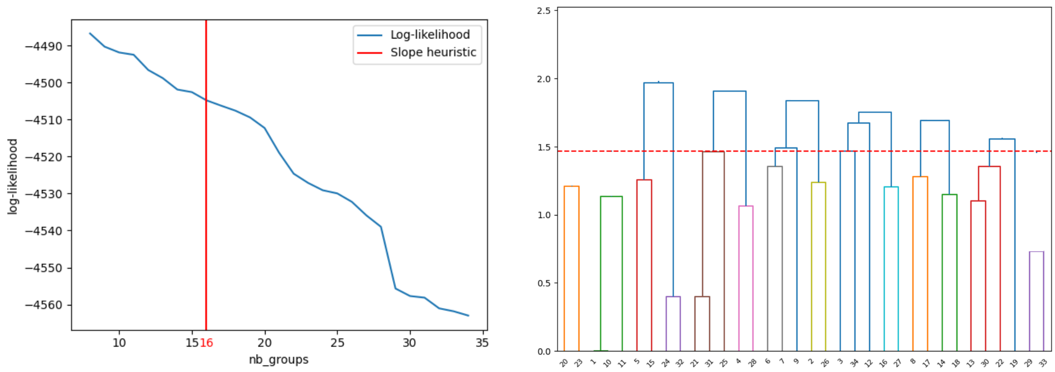


(a) HAC with Ward linkage (b) Dendrogram with Ward linkage

5.3.2.3 *Application to* ***Dataset 2***

The model selection applied to the dataset leads to Figure 23. The slope of the *data-observed log-likelihood* does not have stark changes however it seems to indicate a possible partition with 16 groups. According to this retained partition, the ARI in terms of $\mathtt{Category3}$ and $\mathtt{Category4}$ is 0.02 and 0.10, respectively. These values are close to 0, thus meaning that such clustering solutions are not very far from a random partition. The three groups with maximum substitution for the partition where also retrieved. The analysis of Table 32, 33 and 34 leads to a mix of promising and less promising results. For instance, the wrenchs could substitute with a screw driver or a pipe wrench from an uncompatible size. Thus, the substitutions do not yield reliable information in this case. We guess that it could be due to the lack of information in data (data coming just from one store). It is the reason why the possibility to aggregate data from multiple stores will be studied later. However some values could yield some more valuable information such as the high probability of substitution between the open end wrench of size 18x19mm or 30x32mm to the set of wrenches in Table 32.

Figure 23: **Dataset 2** Selection model for Ward linkage



(a) Log-likelihood (b) Dendrogram of the constrained HAC

Table 32: **Dataset 2** Estimators Group (21,25,31)

	$\hat{\mu}_k$	1	2	3
1	0.19	0.00	0.00	1.00
2	0.21	0.38	0.00	0.62
3	0.24	0.98	0.02	0.00

Interpretation: 1: Open end wrench 18 x 19 mm,
2 : Open end wrench 30x32mm,
3 : Set of 6 open end wrenches

Table 33: **Dataset 2** Estimators Group (20,23)

	μ_k	1	2
1	0.12	0.00	1.00
2	0.20	0.62	0.38

Interpretation: 1: Open end wrench 16 x 17 mm,
2 : Open end wrench 24 x 26 mm

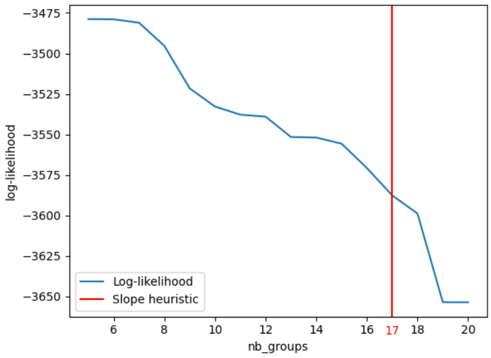
Table 34: **Dataset 2** Estimators Group (13,22,30)

	μ_k	1	2	3
1	0.19	0.00	0.00	1.00
2	0.21	0.38	0.00	0.62
3	0.24	0.98	0.02	0.00

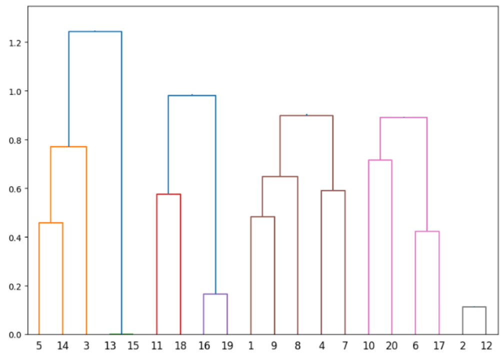
Interpretation: 1: screw drivers, L.80 mm
2 : Open end wrench 21 x 23 mm
3 : pipe wrench 8x9mm

5.3.2.4 *Application to* ***Dataset 3***

Figure 24: **Dataset 3** Store 1

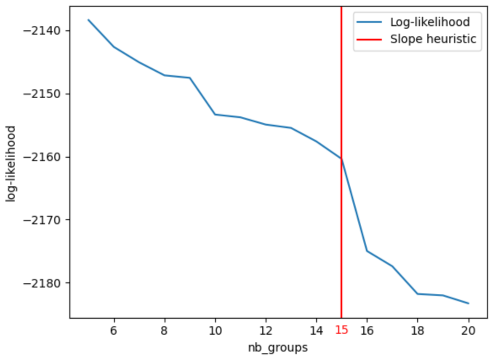


(a) Data-observed log-likelihood

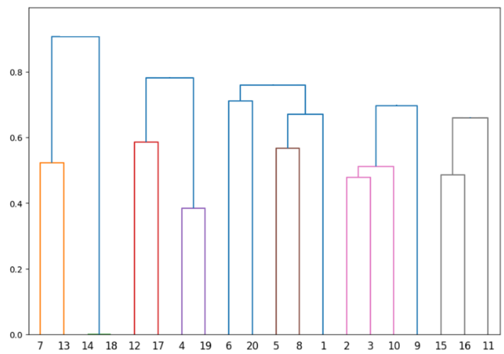


(b) Dendrogram

Figure 25: **Dataset 3** Store 2

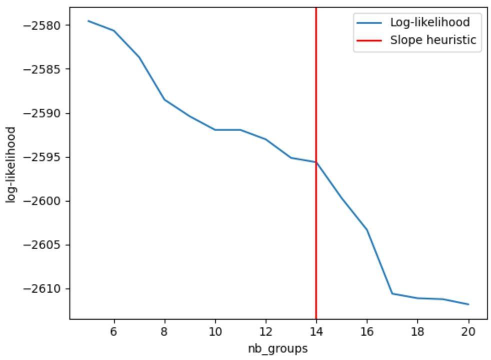


(a) Data-observed log-likelihood

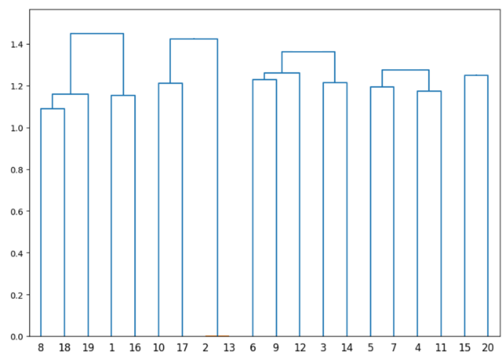


(b) Dendrogram

Figure 26: **Dataset 3** Store 3



(a) Data-observed log-likelihood



(b) Dendrogram

With this new dataset, based on Figure 24, 25, 26, we select again the partition with the slope heuristic but now for each of the three stores. We chose to present the groups in terms of their aggregation order in the constrained HAC. The experiences on simulated data Section 4.5.1.5 showed that the variance of the estimator $\hat{\pi}_{\ell k}$ has a higher variance when product ℓ is sold on a order of magnitude less than k. This last observation will shine light on several false detections of substitution signal.

Table 35, 36, 37 show the estimated probabilities for the three (non-singleton) groups provided by the partition. Products that do not appear in the partition do not have substitute. The first group is composed of two jigsaws. There is a high substitution from the first to the second and no substitution in the other way. The second group appears to be coherent in terms of the two products which are jigsaws with cables with different powers. However there is a high value of substitution from the second product which is less powerful to the first with a perfect substitution. This could be interpreted as an upselling induced by the out of stock which happens when a client choses a more valuable item in place of another. The last group is composed of two products (the jigsaw and the perforator) that are not subtitutable. However, we notice that the first and the second groups have a significant difference in terms of the mean frequency of sale; this fact could explain such an erroneous substitution estimation.

Table 38, 39, 40, 41 provide the results for the second store. The first group aggregates two electric drill perforators. The perfect substitution from the 680W drill perforator to the 1500W indicates an upselling. However the difference of magnitude between the two mean frequencies of sale (corresponding to a biased modelling) could be the reason of such a mistake. The two products of the third group are theoretically not substitutable but the estimation indicates there is a perfect substitution from the first product to the second. This error can be due (again) to the difference of magnitude between the mean frequencies of sales. The third group provides an interesting signal of substitution between two jigsaws with an upselling relationship from the first to the second. The fourth group has two substitutable electric routers which are sold in the same order of mean frequencies of sale and a jigsaw which is not substitutable with the other. The jigsaw is sold twice as much as the other which could indicate why a perfect substitution from the electric router to the jigsaw is detected. However there is some interesting signal such as the fact that there is almost no substitution from the jigsaw to the other products and some substitution between the second electric router to the first.

The analysis of the last store yields Table 42, 43, 44, 45. The first and second groups are examples of the erroneous retrieval of the substitution between products due to the difference of magnitude of mean frequencies of sales (argument already used before). The third group includes two electric drill perforators and a jigsaw. The two perforators can be substituted by each other and do not substitute to the electric jigsaw. However a signal of substitution is retrieved from the jigsaw to the perforator. The last group of the partition produced by the HAC is one including two products that are not substitutable.

Additionally, we can see that the screw drivers and planes were always correctly separated. It must be noted nevertheless that we chose a group of products where there is a disproportion in the products in each category in favor of the electric drill perforator and jigsaw as can be seen in the previous Table 30.

In conclusion, some signals retrieved do have some information about the products. But, the difference of mean frequencies of sales could have an impact on the quality of the groups by aggregating early in the HAC the products that have a difference of order of magnitude of the mean frequency of sales. A further improvement of the algorithm could address this issue and should try to produce different groups including this information of heterogeneity. We note that respectively the presence of jigsaw and of electric router represents a substantial proportion of the products in the dataset yielding groups that could be produced randomly.

Table 35: Store 1 group 1

μ_k	13	15
0.04	0.00	1.00
0.64	0.00	1.00

Interpretation: Scie sauteuse 20V WORX , WX543.9 (Sans batterie ni chargeur)
Scie sauteuse filaire DEXTER Dp5 500.0 W

Table 36: Store 1 group 2

μ_k	2	12
0.48	0.81	0.19
0.59	1.00	0.00

Interpretation: Scie sauteuse filaire BOSCH Pst 650 500 W
Scie sauteuse filaire 400 W

Table 37: Store 1 group 3

μ_k	16	19
0.21	0.00	1.00
1.20	0.11	0.89

Interpretation: Scie sauteuse filaire DEXTER Dp5 750js3-100.5 750.0 W
Perforateur DEXTER Dp5, 800 W

Table 38: Store 2 group 1

μ_k	14	18
0.07	0.00	1.00
0.39	0.00	1.00

Interpretation: Marteau perforateur burineur filaire SDS plus RYOBI RSDS680KA2, 680 W
Perforateur sds plus DEXTER 1500rh2-50.5, 1500 W

Table 39: Store 2 group 2

μ_k	4	19
0.03	0.00	1.00
0.18	0.03	0.97

Interpretation: Scie sauteuse filaire MAKITA Jv0600j 650 W
Perforateur DEXTER Dp5, 800 W

Table 40: Store 2 group 3

μ_k	15	16
0.25	0.20	0.8
0.16	0.15	0.85

Interpretation: Réf 75348084 Scie sauteuse filaire DEXTER Dp5 500.0 W
Scie sauteuse filaire DEXTER Dp5 750js3-100.5 750.0 W

Table 41: Store 2 group 4

μ_k	2	3	10
0.29	0.91	0.06	0.03
0.15	1.00	0.00	0.00
0.10	0.06	0.94	0.00

Interpretation: Scie sauteuse filaire BOSCH Pst 650 500 W
Défonceuse électrique BOSCH Pof 1400 ace + coffret 6 fraises, 1400 W
Défonceuse électrique DEXTER POWER Dp4, 1300.0 W

Table 42: Store 3 group 1

μ_k	2	13
0.05	0.00	1.00
0.19	0.00	1.00

Interpretation: Réf 75348084 Rabot électrique filaire RYOBI Epn 7582 nhg, 750 W
Scie sauteuse filaire 400 W

Table 43: Store 3 group 2

μ_k	1	16
0.02	0.00	1.00
0.23	0.00	1.00

Interpretation: Réf 75348084 Perforateur sds plus AEG Kh28 super xek, 1010 W
Scie sauteuse filaire DEXTER Dp5 750js3-100.5 750.0 W

Table 44: Store 3 group 3

μ_k	8	18	19
0.11	0.00	0.24	0.76
0.45	0.0	0.64	0.36
0.15	0.04	0.11	0.85

Interpretation: Scie sauteuse filaire BOSCH Pst9500pel 620.0 W
Perforateur sds plus DEXTER 1500rh2-50.5, 1500 W
Perforateur DEXTER Dp5, 800 W

Table 45: Store 3 group 4

μ_k	4	11
0.20	0.83	0.17
0.11	0.90	0.10

Interpretation: Défonceuse électrique BOSCH Pof 1400 ace + coffret 6 fraises, 1400 W
Scie sauteuse filaire BOSCH Pst9500pel 620.0 W

5.3.2.5 *Application to **Dataset 4***

Using the heuristic of the method of Massart, we select the best partition based on the change of slope from the *observed-data log-likelihood*.

The three groups show substitution between non substitutable products. In Figure 46, a drill perforator and a router which are note substitutable are aggregated. Figure 47 shows a group composed of a jigsaw and an electric screw driver. Figure 48 groups an electric jigsaw and a drill perforator. We note however that there is a factor of two between the mean frequencies of sales of the two products in the last two groups. This observation has already been drawn for **Dataset 3**. This dataset being based on the aggregation of the time series over multiple stores, the difference of mean frequencies of sales between store introduces some bias to the model. Another cause of bias would be the imputation of zeros in the data over stores that do not carry one of the products in its assortment. The increase of data that the aggregation of the time series provides did not improve the quality of the results, the experience on the **Dataset 3** was more informative.

Table 46: **Dataset 4** 1rst group of substitution

μ_k	6	11
0.22	1.00	0.00
0.20	1.00	0.00

Interpretation: 6 Défonceuse électrique BOSCH Pof 1400 ace + coffret 6 fraises, 1400 W
11 Perforateur sds plus BOSCH Pbh 2500 sre, 600 W

Table 47: **Dataset 4** 2nd group of substitution

μ_k	7	31
0.07	0.00	1.00
0.17	0.01	0.99

Interpretation: 7 Scie sauteuse filaire MAKITA Jv0600j 650 W
31 Tournevis sans fil DEXTER 3.6 V 2.0 Ah

Table 48: **Dataset 4** 3rd group of substitution

μ_k	18	26
0.17	0.67	0.33
0.40	0.37	0.63

Interpretation: 18 Scie sauteuse filaire 400 W
26 Perforateur sds plus DEXTER 1500rh2-50.5, 1500 W

K	B_5	B_2	a	c	proportion group retrieved	ARI	max substitution	overall max substitution	
150	10	50	perfect subst	1	1.	1.	(39, 40),1.	(39, 40),1.	
150	10	50	perfect subst	10	1.	1.	(75, 76),1.	(75, 76),1.	
270	50	10	perfect subst	1	1.	1.	(13, 14), 1.	(13, 14), 1.	
270	50	10	perfect subst	10	1.	1.	(9, 10),1.	(9, 10),1.	
150	10	50	realistic	1	1.	1.	(59, 60),0.45	(59, 60),0.54	
150	10	50	realistic	10	1.	1.	(49, 50),0.5174586004770456	(49, 50),0.5274274568134114	
270	50	10	realistic	1	1.	1.	(19, 20),0.42	(19, 20),0.56	
270	50	10	realistic	10	0.95	0.99	(3, 4),0.45	(3, 4),0.66	
500	100	0	realistic	1	1.	1.	(261, 262, 263, 264, 265),0.17	(261, 262, 263, 264, 265),0.56	
500	100	0	realistic	10	-853704.052	0.80	0.97	(276, 277, 278, 279, 280),0.27	(276, 277, 278, 279, 280),0.64

Table 28: Model selection for $\tilde{S}$ where $(\mu, n) = (2, 10^4)$

6 CONCLUSION

In this thesis, we provided a substitution model with a necessary and sufficient condition for the identifiability of the parameters which was not found in the literature's models. We proposed an EM algorithm that enables the estimation of products in an acceptable run time for a small number of products. The numerical results on realistic simulated data showed the good properties of our algorithm and the results for Adeo's datasets were coherent and interpretable. For a larger number of products, the parameters may not be identifiable and the time of computation of the algorithm was untractable. We proposed a sparse model of substitution that accounts for the numerous zero probability of sales. This additional layer enabled the proposition of embedded models described by partitions of products obtained by a constrained agglomerative clustering that accounts for the identifiability within the groups. The best partition is then selected based on a heuristic of the slope method of Massart. The numerical applications on simulated data showed that the constrained HAC and the selection model perform well. On real data, the small dataset clustered the products in the right way. However on large datasets the results were mitigated in the sense that some groups of products and substitution probabilities were interpretable but some substitution were falsely detected. The experiments showed that the data are sparse and this either leads to a high variance of the estimators on single stores or the aggregated data were not coherent with the assumption of homogeneity of the sales process which led to a bias in the model and some poor results. We also point out the effect of the difference of order of magnitude of the mean frequencies of sales within a group both on the quality of the estimators and on the groups retrieved by the model selection.

As a research perspective, a new model could account for the heterogeneity of the mean frequencies of sales within groups in the objective to reduce the variance of the estimators. Another avenue would be to propose a model that accounts for the non-homogeneous Poisson process underlying the demand and the sales such as some models in the literature and to propose a condition of identifiability of the parameters. We also point out that Adeo uses an heuristic concerning the demand that meets an out of stock which is that $\frac{1}{3}$ is backlogged, $\frac{1}{3}$ is substituted and $\frac{1}{3}$ is lost hence it would be natural to propose a model that accounts for the backlog.

Part II

APPENDIX

A
COMPLEMENTS RELATED TO CHAPTER 2

A.1 MATHEMATICAL DETAILS RELATED TO THE NEWSVENDOR MODEL

Proof 1 (of Proposition 1) *We have*

$$\min(Q_k, \tilde{D}_k) = \tilde{D}_k + \min(Q_k - \tilde{D}_k, 0) = \tilde{D}_k - max(\tilde{D}_k - Q_k, 0) = \tilde{D}_k - (\tilde{D}_k - Q_k)^+.$$

We can rewrite the profit as

$$\begin{aligned}\rho_k(\{D_{tk}\}_{t\in[0,n]}, \{V_{tk}\}_{t\in[0,n]}; \theta, Q_k) &= c_{kp}(\tilde{D}_k - (\tilde{D}_k - Q_k)^+) - (c_{kh} - c_{ks})(Q_k - \tilde{D}_k)^+ \\ &\quad - k(\tilde{D}_k - Q_k)^+ - c(Q_k - \tilde{D}_k) - c_{kc}\tilde{D}_k \\ &= (c_{kp} - c_{kc})\tilde{D}_k - (c_{kh} - c_{ks} + c_{kc})(Q_k - \tilde{D}_k)^+ \\ &\quad - (k + c_{kp} - c_{kc})(\tilde{D}_k - Q_k)^+.\end{aligned}$$

Here the cost of one unit shortage is the lost margin plus the additional cost of not fulfilling the order ($c_k^- = c_{ku} + c_{kp} - c_{kc}$). The unit cost for an excess of one unit (overage cost) here is $c_k^+ = h + c - s$, which is the value of the stock and the holding cost. We obtain then the following formula for the profit:

$$\rho_k(\tilde{D}_k; \theta, Q_k) = (c_{kp} - c)\tilde{D}_k - c_k^+(Q_k - \tilde{D}_k)^+ - c_k^-(\tilde{D}_k - Q_k)^+.$$

We get also the expected profit:

$$\mathbb{E}[\rho_k \tilde{D}_k; \theta, Q_k)] = (c_{kp} - c)\mu - c_k^+\mathbb{E}((Q_k - \tilde{D}_k)^+) - c_k^-\mathbb{E}((\tilde{D}_k - Q_k)^+),$$

where μ is the mean of the demand. Optimizing this function is equivalent to optimize the cost $c_k^+\mathbb{E}((Q_k - \tilde{D}_k)^+) + c_k^-\mathbb{E}((\tilde{D}_k - Q_k)^+)$.

Proof 2 (of Proposition 2) *In the case of continuous $\tilde{D}_k$, an optimum of this function in Q_k^* verifies $\frac{\partial\mathbb{E}[\rho_k(\{D_{tk}\}_{t\in[0,n]}, \{V_{tk}\}_{t\in[0,n]}; \theta_k, Q_k^*)]}{\partial Q_k} = 0$. We have*

$$\begin{aligned}&\frac{\partial\mathbb{E}((Q_k - \tilde{D}_k)^+)}{\partial Q_k} = \frac{\partial}{\partial Q_k}\int_0^{Q_k}(Q_k - x)f_{\tilde{D}_k}(x)d_x \\ &= F_{\tilde{D}_k}(Q_k) + Q_k f_{\tilde{D}_k}(Q_k) - Q_k f_{\tilde{D}_k}(Q_k) = F_{\tilde{D}_k}(Q_k).\end{aligned}$$

We also have

$$\begin{aligned}&\frac{\partial\mathbb{E}((\tilde{D}_k - Q_k)^+)}{\partial Q_k} = \frac{\partial}{\partial Q_k}\int_0^{+\infty}(x - Q_k)f_{\tilde{D}_k}(x)d_x \\ &= (Q_k f_{\tilde{D}_k}(Q_k) - (1 - F_{\tilde{D}_k}(Q_k)) - Q_k(f_{\tilde{D}_k}(Q_k)) = F_{\tilde{D}_k}(Q_k) - 1.\end{aligned}$$

We then obtain

$$\begin{aligned}&c_0 F_{\tilde{D}_k}(Q_k) + c_u(F_{\tilde{D}_k}(Q_k) - 1) = 0 \\ &\Leftrightarrow F_{\tilde{D}_k}(Q_k) = \frac{c_u}{c_u + c_0}.\end{aligned}$$

A.2 PROOFS RELATED TO THE REPLENISHMENT OF SLOW MOVERS IN SECTION 2.2.3

Lemma 5 *Let $X \sim \mathcal{N}(\mu, \sigma^2)$, and f, F be respectively the probability mass function and the cumulative distribution function then $\mathbb{E}((X-Q)^+) = \sigma^2 f(Q) + (\mu - Q)(1 - F(Q))$.*

Proof 2 (of Lemma 5) *We have*

$$\mathbb{E}((X-Q)^+) = \int_Q^{+\infty} (x-Q) f_{\mathcal{N}(0,1)}(x) dx,$$

which is equal to

$$\begin{aligned} &= \int_{x=Q}^{\infty} x \cdot f(x) dx - Q \int_{x=Q}^{\infty} f(x) dx \\ &= \text{Int1} - Q[1 - F(Q)]. \end{aligned}$$

The following proof can be found in Vandeput, 2020. In order to solve Int1, we will use the definition of $f(x)$ for a normal distribution:

$$f(x) = \frac{1}{\sqrt{2n\sigma^2}} e^{-\frac{x-\mu^2}{2\sigma^2}}.$$

So that

$$\begin{aligned} \text{Int1} &= \int_{x=Q}^{\infty} x f(x) dx \\ &= \int_Q^{\infty} \frac{x}{\sqrt{2n\sigma^2}} \exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right) dx \\ &= \frac{1}{\sqrt{2n\sigma^2}} \int_Q^{\infty} x \exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right) dx. \end{aligned}$$

We will integrate this by defining

$$u = \frac{x-\mu}{\sigma}; du = \frac{1}{\sigma} dx \iff d = u\sigma + \mu; dx = \sigma du.$$

So that,

$$\begin{aligned} \int_{x=Q}^{\infty} x \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx &= \int_{u=\frac{Q-\mu}{\sigma}}^{\infty} \sigma(u\sigma + \mu) \cdot \exp\left(\frac{-(u\sigma)^2}{2\sigma^2}\right) du \\ &= \sigma \int_{u=\frac{Q-\mu}{\sigma}}^{\infty} (u\sigma + \mu) \cdot \exp\left(\frac{-u^2}{2}\right) du \end{aligned}$$

$$\begin{aligned} &= \sigma^2 \int_{u=\frac{Q-\mu}{\sigma}}^{\infty} u \cdot \exp\left(\frac{-u^2}{2}\right) du + \mu\sigma \int_{u=\frac{Q-\mu}{\sigma}}^{\infty} \exp\left(\frac{-u^2}{2}\right) du \\ &= \sigma^2 \cdot \text{Int}\, 2_1 + \mu \cdot \text{Int}\, 2_2. \end{aligned}$$

$\text{Int}\, 2_1$,

$$\mathit{Int}\ 2_1 = \int_{u=\frac{Q-\mu}{\sigma}}^{\infty} u \cdot \exp\left(\frac{-u^2}{2}\right) du.$$

We integrate by defining v as

$$v = \frac{-u^2}{2}; dv = -u du \Leftrightarrow u = \sqrt{-2v}.$$

The limits of the integral will change as

$$u = \infty \Rightarrow v = -\infty \text{ and } u = \frac{Q-\mu}{\sigma} \Rightarrow v = -\left(\frac{Q-\mu}{\sqrt{2}\sigma}\right)^2.$$

We transform our integral in

$$\begin{aligned}
\int_{v=-\left(\frac{Q-\mu}{\sqrt{2}}\sigma\right)^2}^{-\infty} u \cdot \exp\left(\frac{-u^2}{2}\right) du &= -\int_{v=-\left(\frac{Q-\mu}{\sqrt{2}\sigma}\right)^2}^{-\infty} e^{v} dv \\
&= -\left(e^{-\infty} - e^{-\left(\frac{Q-\mu}{\sqrt{2}\sigma}\right)^2}\right) \\
&= e^{-\left(\frac{Q-\mu}{\sqrt{2}}\right)^2}.
\end{aligned}$$

According to the definition of the normal density function (see Eq. B.6), we can transform into

$$\exp\left(-\left(\frac{Q-\mu}{\sqrt{2}\sigma}\right)^2\right) = \frac{\sqrt{2\pi\sigma^2}}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{(Q-\mu)^2}{2\sigma^2}\right) = \sqrt{2\pi\sigma^2} \cdot f(Q).$$

We finally solve $\text{Int}\, 2_1$*:*

$$\text{Int}\, 2_1 = \sqrt{2\pi\sigma^2} \cdot f(Q).$$

We now solve $\text{Int}\, 2_2$*,*

$$\text{Int2}_2 = \sigma \int_{u=\frac{Q-\mu}{\sigma}}^{\infty} \exp\left(\frac{-u^2}{2}\right) du.$$

We can revert to x *instead of* u *and we obtain*

$$\sigma \int_{x=Q}^{\infty} \frac{1}{\sigma} \cdot \exp\left(\frac{-(x-\mu)^2}{\sigma^2}\right) dx = \int_{x=Q}^{\infty} \cdot \exp\left(\frac{-(x-\mu)^2}{\sigma^2}\right) dx = \sqrt{2\pi\sigma^2}(1-F(Q)).$$

So that,

$$\text{Int}\, 2_2 = \sqrt{2\pi\sigma^2}(1-F(Q)).$$

$$\begin{aligned}
\textit{Int2} &= \int_{x=Q}^{\infty} x \cdot f(x) dx \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{x=Q}^{\infty} x \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \left(\sigma^2 \cdot \text{Int}\, 2_1 + \mu \cdot \text{Int}\, 2_2\right) \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \left(\sigma^2 \cdot \sqrt{2\pi\sigma^2} \cdot f(Q) + \mu \cdot \sqrt{2\pi\sigma^2}(1-F(Q))\right) \\
&= \sigma^2 f(Q) + \mu(1-F(Q)).
\end{aligned}$$

We can then solve our initial loss function:

$$\begin{aligned}
\mathbb{E}[(x-Q)^+] &= \int_{x=Q}^{\infty} (x-Q) f(x) dx \\
&= \int_{x=Q}^{\infty} x \cdot f(x) - Q \int_{x=Q}^{\infty} f(x) dx \\
&= \textit{Int1} - Q \cdot \textit{Int2} \\
&= \sigma^2 f(Q) + \mu(1-F(Q)) - Q(1-F(Q)) \\
&= \sigma^2 f(Q) + (\mu - Q)(1-F(Q)).
\end{aligned}$$

Proof 3 (of Lemma 2) *We have that* $Q_{(K)} = F^{-1}_{(K)}(\alpha)$ *and* $F_{(K)}$ *is the cumulative distribution function of* $\mathcal{N}(K\mu, K\sigma^2)$. *Hence* $Q_{(K)} = K\mu + z_\alpha\sqrt{K}\sigma$ *with* $z_\alpha = F^{-1}_{N(0,1)}(\alpha)$. *We then have that the gap of total stock can be written as* $\Delta Q(K,f) = K(\mu + z_\alpha\sigma) - (K\mu + z_\alpha\sigma\sqrt{K}) = z_\alpha\sigma K(1 - K^{-\frac{1}{2}})$.

The formula of the profit is

$$\Pi_{(K)}(Q_{(K)}, f) = (r-c)\mathbb{E}(D_{(K)}) - c_u\mathbb{E}((D_{(K)} - Q_{(K)})^+) - c_o\mathbb{E}((Q_{(K)} - D_{(K)})^+). \tag{44}$$

Since $\mathbb{E}((Q_{(K)} - D_K)^+) = \mathbb{E}(Q_{(K)} - D_{(K)}) + \mathbb{E}((D_{(K)} - Q_K)^+)$ *we have*

$$\Pi_{(K)}(Q_{(K)}, f) = (r-c)\mathbb{E}(D_{(K)}) - (c_u + c_o)\mathbb{E}((D_{(K)} - Q_{(K)})^+) - c_o\mathbb{E}(Q_{(K)} - D_{(K)}).$$

According to Lemma 5, we have

$$\begin{aligned}\Pi_{(K)}(Q_{(K)}, f) &= (r-c)K\mu - (c_u + c_o)[K\sigma^2 f_{(K)}(Q_{(K)}) + (K\mu - Q_{(K)})(1 - F_{(K)}(Q_{(K)}))] \\ &\quad - c_o(z_\alpha\sqrt{K}\sigma) \\ &= (r-c)K\mu - (c_u + c_o)[K\sigma^2 f_{(K)}(Q_{(K)}) - z_\alpha\sqrt{K}\sigma)(\frac{c_o}{c_u + c_o})] - c_o(z_\alpha\sqrt{K}\sigma) \\ &= (r-c)K\mu - (c_u + c_o)K\sigma^2 f_{(K)}(Q_{(K)}). \end{aligned} \tag{45}$$

Since

$$\begin{aligned} f_{(K)}(Q_{(K)}) &= \frac{1}{\sqrt{2\pi K}\sigma}\exp\left(-\frac{1}{2}(\frac{Q_{(k)} - K\mu}{\sqrt{K}\sigma})^2\right) && (46) \\ &= \frac{1}{\sqrt{2\pi K}\sigma}\exp\left(-\frac{(z_\alpha)^2}{2}\right) && (47) \\ &= \frac{1}{\sqrt{K}\sigma} f_{\mathcal{N}(0,1)}(z_\alpha). \end{aligned}$$

The profit is then

$$\Pi_{(K)}(Q_{(K)}, f) = (r-c)K\mu - (c_u + c_o)\sqrt{K}\sigma f_{\mathcal{N}(0,1)}(z_\alpha). \tag{48}$$

Hence we obtain $\Delta\Pi(K, f)$ *by taking the difference between* $\Pi_{(K)}(Q_{(K)}, f)$ *and* $K\Pi_{(1)}(Q_{(1)}, f)$.

A.3 ADDITIONAL CONTENT RELATED TO THE POISSON CASE

Lemma 6 *Let* $D_{(K)} \sim \mathcal{P}(K\lambda)$, *and* $Q = \inf\{Q \mid \mathbb{P}(D_{(K)} \leqslant Q) \geqslant \alpha\}$. *Then* $\beta(\alpha, K) = \mathbb{P}(D_{(K)} \leqslant Q_{(K)})$ *verifies* $\beta(\alpha, K) \xrightarrow{K} \alpha$.

Proof 4 (of Lemma 6) *We have that by definition of* $Q^*_{(K)}$ *that* $0 \geqslant \beta - \alpha$. *We also have that* $\mathbb{P}(D_{(K)} \leqslant Q_{(K)} - 1) < \alpha$ *so we have* $\beta - \alpha < \mathbb{P}(D_{(K)} \leqslant Q_{(K)}) - \mathbb{P}(D_{(K)} \leqslant Q_{(K)} - 1)$ *which is* $\beta - \alpha < \mathbb{P}(D_{(K)} = Q_{(K)})$. *At the limit we have* $\mathbb{P}(D_{(K)} = Q_{(K)}) \approx f_{\mathcal{N}(\lambda K, \lambda K)}(Q_{(K)})$ *which is decreasing to* 0.

Proof 5 (of Lemma 3) ***Proof of the bounds*** *We have that* $\mathbb{E}[(Q_{(K)} - D_{(K)})^{+}) = \mathbb{E}[Q_{(K)} - D_{(K)}] + \mathbb{E}[(D_{(K)} - Q_{(K)})^{+}]$ *which yields*

$$\Pi_{(K)}(Q_{(K)}, P) = (r-c)K\lambda - (c_u + c_o)\mathbb{E}[D_{(K)} - (Q_{(K)})^{+}] - c_o\mathbb{E}[Q_{(K)} - D_{(K)}].$$

We have that $\mathbb{E}[(D_{(K)} - Q_{(K)})^{+}] = \sum_{d=Q_{(K)}+1}^{\infty} d\mathbb{P}(D_{(K)} = d) - Q_{(K)}(1 - \mathbb{P}(D_{(K)} \leqslant Q_{(K)})$ *and*

$$\sum_{d=Q_{(K)}+1}^{\infty} d\mathbb{P}(D_{(K)} = d) = \sum_{d=Q_{(K)}+1}^{\infty} d\frac{(K\lambda)^{d}\exp(-K\lambda)}{d!} = K\lambda(1 - \mathbb{P}(D_{(K)} \leqslant Q_{(K)} - 1)).$$

Then we have

$$\begin{aligned}\Pi_{(K)}(Q_{(K)}, P) = (r-c)K\lambda - (c_u + c_o)\big[K\lambda(1 - \mathbb{P}(D_{(K)} \leqslant Q_{(K)} - 1) - Q_{(K)}(1 - \mathbb{P}(D_{(K)} \leqslant Q_{(K)})\big] \\ - c_o\mathbb{E}[Q_{(K)} - D_{(K)}].\end{aligned}$$

By rearranging the terms we get

$$= (r-c)K\lambda - (c_u + c_o)\big[K\lambda(\frac{c_u}{c_u + c_o} - \mathbb{P}(D_{(K)} \leqslant Q_{(K)} - 1) + Q_{(K)}(\mathbb{P}(D_{(K)} \leqslant Q_{(K)}) - \frac{c_u}{c_u + c_o})\big].$$

Since $\mathbb{P}(D_{(K)} \leqslant Q_{(K)}) \geqslant \frac{c_u}{c_u+c_o}$ *we have the upper bound*

$$\begin{aligned}&(r-c)K\lambda - (c_u + c_o)K\lambda(\mathbb{P}(D_{(K)} \leqslant Q_{(K)}) - \mathbb{P}(D_{(K)} \leqslant Q_{(K)} - 1)) \\ &\qquad = (r-c)K\lambda - (c_u + c_o)K\lambda\mathbb{P}(D_{(K)} = Q_{(K)}).\end{aligned} \tag{49}$$

In order to retrieve the strict lower bound we use the inequality $\mathbb{P}(D_{(K)} \leqslant Q_{(K)} - 1) < \frac{c_u}{c_u+c_o}$. *We get*

$$\begin{aligned}&(r-c)K\lambda - (c_u + c_o)Q_{(K)}(\mathbb{P}(D_{(K)} \leqslant Q_{(K)}) - \mathbb{P}(D_{(K)} \leqslant Q_{(K)} - 1))\big] \\ &\qquad = (r-c)K\lambda - (c_u + c_o)Q_{(K)}\mathbb{P}(D_{(K)} = Q_{(K)}).\end{aligned} \tag{50}$$

Proof of the asymptotic behavior of the bounds *Now we compute the asymptotic complexity of the last term of the upper and lower bound. In order to do so we have to compute the complexity of* $\mathbb{P}(D_{(K)} = Q_{(K)})$. *We have that this quantity for large values of* K *can be approximated by* $\frac{1}{\sqrt{2\pi\lambda K}}\exp\big(-\frac{1}{2}\big(\frac{Q_{(K)}-\lambda K}{\sqrt{\lambda K}}\big)^2\big)$. *The quantile of the Poisson law can be approximated by the Cornish Fisher expansion given the ratio associated to* $Q_{(K)}$. *This ratio is here denoted by* $\beta(\alpha, K) = \mathbb{P}(D_{(K)} \leqslant Q_{(K)})$. *We have that*

$$Q_{(K)} = \lambda K + F^{-1}_{\mathcal{N}(0,1)}(\beta(\alpha, K))\sqrt{\lambda K} + O(1).$$

We then have

$$\mathbb{P}(D_{(K)} = Q_{(K)}) \approx \frac{1}{\sqrt{2\pi\lambda K}}\exp\big(-\frac{1}{2}\big(F^{-1}_{\mathcal{N}(0,1)}(\beta(\alpha, K)\big)^2\big). \tag{51}$$

According to Lemma 6 we have the convergence of $\beta(\alpha, K) \xrightarrow{K} \alpha$ *so using the continuity of* $F^{-1}_{\mathcal{N}(0,1)}$ *we get*

$$\mathbb{P}(D_{(K)} = Q_{(K)}) \approx \frac{1}{\sqrt{2\pi\lambda K}} \exp\left(-\frac{1}{2}\left(F^{-1}_{\mathcal{N}(0,1)}(\alpha)\right)^2\right). \tag{52}$$

The upper bound verifies the stated result in the lemma

$$-(c_u + c_o)(K\lambda)\mathbb{P}(D_{(K)} = Q_{(K)}) \approx -(c_u + c_o)(\lambda K)^{\frac{1}{2}} f_{\mathcal{N}(0,1)}(z_\alpha)$$

where $z_\alpha = F^{-1}_{\mathcal{N}(0,1)}(\alpha)$.

The lower bound given the expression 51 is

$$\begin{aligned} -(c_u + c_o)Q_{(K)}\mathbb{P}(D_{(K)} = Q_{(K)}) &\approx -(c_u + c_o)(\lambda K)^{\frac{1}{2}}\Big[1 + F^{-1}_{\mathcal{N}(0,1)}(\beta(\alpha, K))(\lambda K)^{-\frac{1}{2}} \\ &\quad + O((\lambda K)^{-1})\Big] f_{\mathcal{N}(0,1)}(z_\alpha) \\ &\approx -(c_u + c_o)\Big[(\lambda K)^{\frac{1}{2}} f_{\mathcal{N}(0,1)}(z_\alpha) \\ &\quad + F^{-1}_{\mathcal{N}(0,1)}(\beta(\alpha, K)) f_{\mathcal{N}(0,1)}(z_\alpha) + O((\lambda K)^{-\frac{1}{2}})\Big]. \end{aligned}$$

Proof of the asymptotic behavior of the difference of the bounds

We have that the difference between the two bounds are

$$\begin{aligned} -(c_u + c_o)\lambda K\mathbb{P}(D_{(K)} = Q_{(K)}) &+ (c_u + c_o)Q_{(K)}\mathbb{P}(D_{(K)} = Q_{(K)}) \\ &= (c_u + c_o)[Q_{(K)} - \lambda K]\mathbb{P}(D_{(K)} = Q_{(K)}) \end{aligned}$$

Using Formula 51, the quantity $[Q_{(K)} - \lambda K] = F^{-1}_{\mathcal{N}(0,1)}(\beta(\alpha, K))(\lambda K)^{\frac{1}{2}} + O(1)$. *We then get using the normal approximation*

$$= (c_u + c_o)[F^{-1}_{\mathcal{N}(0,1)}(\beta(\alpha, K))(\lambda K)^{\frac{1}{2}} + O(1)] \frac{1}{\sqrt{\lambda K}} f_{\mathcal{N}(0,1)}(z_\alpha).$$

Due to the fact that $\beta(\alpha, K) \xrightarrow{K} \alpha$ *and the continuity of* $F^{-1}_{\mathcal{N}(0,1)}$ *, we get the final result:*

$$= (c_u + c_o)[z_\alpha] f_{\mathcal{N}(0,1)}(z_\alpha) + O\left(\frac{1}{\sqrt{\lambda K}}\right). \tag{53}$$

This concludes the proof.

B COMPLEMENTS RELATED TO CHAPTER 4

B.1 MODEL 1 PROPERTIES

Proof 6 (of Proposition 6) *Let two sets of parameters* θ *and* $\widetilde{\theta}$ *satisfy*

$$\forall \boldsymbol{y}_n \in \mathbb{N}^p, f(\boldsymbol{y}_n \mid \boldsymbol{x}_n; \theta) = f(\boldsymbol{y}_n \mid \boldsymbol{x}_n; \widetilde{\theta}).$$

1. *Since* $x_{nj} > 0$ *for* $\omega_j = \mathbf{1}_K$ *we have that* $\lambda_k(\boldsymbol{\omega}_j; \theta) = \lambda_k(\boldsymbol{\omega}_j; \tilde{\theta})$. *Assumptions 5 implies that* $\boldsymbol{\mu}$ *is identifiable since the equation translates to* $\mu_k \omega_{jk} = \tilde{\mu}_k \omega_{jk}$.

   From the second condition, we chose $\ell \in [K]$ *and then retrieve* $\bar{j} \in [J]$ *such that* $\forall h \in [K], h \neq \ell, \omega_{\bar{j}h} = 1$ *and* $\omega_{\bar{j}\ell} = 0$. *For any* $k \in [K]$ *with* $k \neq \ell$

$$\mu_k + \sum_{h \neq k} \mu_h (1 - \boldsymbol{\omega}_{\bar{j}h}) \omega_{\bar{j}k} \beta_{hk} (1 + \omega_{\bar{j}}^\top \boldsymbol{\beta}_h)^{-1} = \tilde{\mu}_k + \sum_{h \neq k} \tilde{\mu}_h (1 - \omega_{\bar{j}h}) \omega_{\bar{j}k} \tilde{\beta}_{hk} (1 + \omega_{\bar{j}}^\top \tilde{\boldsymbol{\beta}}_h)^{-1}$$

   which becomes

$$\omega_{\bar{j}k} \beta_{\ell k} (1 + \omega_{\bar{j}}^\top \boldsymbol{\beta}_\ell)^{-1} = \omega_{\bar{j}k} \tilde{\beta}_{\ell k} (1 + \omega_{\bar{j}}^\top \tilde{\boldsymbol{\beta}}_\ell)^{-1}. \tag{54}$$

   This equality is verified for all k *hence summing over the* k *yields the equality of* $(1 + \omega_{\bar{j}}^\top \boldsymbol{\beta}_\ell)^{-1} = (1 + \omega_{\bar{j}}^\top \tilde{\boldsymbol{\beta}}_\ell)^{-1}$. *Therefore*

$$\omega_{\bar{j}k} \beta_{\ell k} (1 + \omega_{\bar{j}}^\top \boldsymbol{\beta}_\ell)^{-1} = \omega_{\bar{j}k} \tilde{\beta}_{\ell k} (1 + \omega_{\bar{j}}^\top \boldsymbol{\beta}_\ell)^{-1}$$

   which implies $\beta_{\ell k} = \tilde{\beta}_{\ell k}$.

2. *The proof of the identifiability is recursive. The objective is to show the property at* h *that* $\beta_{hh}, \dots, \beta_{hK}$ *are identifiable. For* $h \in [K]$, *let* $j_h \in [J]$ *be a reference to an availability configuration that verifies* $\omega_{j_h k} = 0$ *for* $k < h$ *and* $\omega_{j_h k} = 1$ *for* $k \geqslant h \in [K]$.

   According to the previous proof, the observation of $\boldsymbol{\omega}_{j_1}$ *implies the identifiability of* $\boldsymbol{\mu}$, *the observation of* $\boldsymbol{\omega}_{j_2}$ *that of* $\boldsymbol{\beta}_1$. *By recursion, we suppose that the property is verified for* $h - 1$. *We choose the observation* $\boldsymbol{\omega}_{j_{h+1}}$ *to demonstrate that* $\beta_{hh}, \dots \beta_{hK}$ *are identifiable. In the following, we drop the subscript* $h+1$ *from* j *for ease of reading. Suppose we have* $\beta_{hh} = \tilde{\beta}_{hh} = 0$. *For* $k > h$ *and from the previous proof, we have that*

$$\begin{aligned} &\lambda_k(\boldsymbol{\omega}_j; \theta) = \lambda_k(\boldsymbol{\omega}_j; \tilde{\theta}) \\ \Leftrightarrow &\mu_k + \sum_{\ell \neq k} \mu_\ell (1 - \omega_{j\ell}) \omega_{jk} \beta_{\ell k} (1 + \omega_j^\top \boldsymbol{\beta}_\ell)^{-1} \\ &= \tilde{\mu}_k + \sum_{\ell \neq k} \tilde{\mu}_\ell (1 - \omega_{j\ell}) \omega_{jk} \tilde{\beta}_{\ell k} (1 + \omega_j^\top \tilde{\boldsymbol{\beta}}_\ell)^{-1}. \end{aligned}$$

   Since $k > h$, *we have that* $\omega_{jk} = 1$ *and for* $\ell > h$ *we have that* $(1 - \omega_{j\ell}) = 0$. *Hence we have*

$$\sum_{\ell \leqslant h} \mu_\ell (1 - \omega_{j\ell}) \omega_{jk} \beta_{\ell k} (1 + \omega_j^\top \boldsymbol{\beta}_\ell)^{-1} = \sum_{\ell \leqslant k} \tilde{\mu}_\ell (1 - \omega_{j\ell}) \omega_{jk} \tilde{\beta}_{\ell k} (1 + \omega_j^\top \tilde{\boldsymbol{\beta}}_\ell)^{-1}.$$

The identifiability of the parameters $\boldsymbol{\mu}$ *and* $\boldsymbol{\beta}_\ell$ *for* $\ell < h$ *implies that* $\forall k > h$

$$\omega_{jk}\beta_{hk}(1+\boldsymbol{\omega}_j^\top\boldsymbol{\beta}_h)^{-1} = \omega_{jk}\tilde{\beta}_{hk}(1+\boldsymbol{\omega}_j^\top\tilde{\boldsymbol{\beta}}_h)^{-1}. \tag{55}$$

Since we sum over $k > h$*, we obtain*

$$\sum_{k>h}\frac{\omega_{jk}\beta_{hk}}{(1+\boldsymbol{\omega}_j^\top\boldsymbol{\beta}_h)} = \sum_{k>h}\frac{\omega_{jk}\tilde{\beta}_{hk}}{(1+\boldsymbol{\omega}_j^\top\tilde{\boldsymbol{\beta}}_h)}.$$

And since $\sum_{k>h}\omega_{jk}\beta_{hk} = \boldsymbol{\omega}_j^\top\boldsymbol{\beta}_h$ *and* $\beta_{hk} \geqslant 0$ *we either have identifiability of the parameters through* $\beta_{hk} = 0, \forall k > h$ *or*

$$\frac{1}{(1+(\boldsymbol{\omega}_j^\top\boldsymbol{\beta}_h)^{-1})} = \frac{1}{(1+(\boldsymbol{\omega}_j^\top\tilde{\boldsymbol{\beta}}_h)^{-1})}$$

leading to $(\boldsymbol{\omega}_j^\top\boldsymbol{\beta}_h) = (\boldsymbol{\omega}_j^\top\tilde{\boldsymbol{\beta}}_h)$.

Since $k \geqslant h$ *we have that* $\omega_{jk} = 1$ *and* $\beta_{hh} = 0$*, equation (55) yields* $\forall k \geqslant h$, $\beta_{hk} = \tilde{\beta}_{hk}$.

B.2 MODEL 2 PROPERTIES

Proof 7 (of Proposition 7) *Let* $\theta \in \Theta$ *and* $\widetilde{\theta} \in \Theta$ *that satisfy*

$$\forall \boldsymbol{y}_n \in \mathbb{N}^p, f(\boldsymbol{y}_n \mid \boldsymbol{x}_n;\theta) = f(\boldsymbol{y}_n \mid \boldsymbol{x}_n;\widetilde{\theta}).$$

Define the following reparameterization of the intensity functions

$$\lambda_k(\boldsymbol{\omega}_j;\boldsymbol{\alpha}_k) = \omega_{jk}\left[\alpha_{kk} + \sum_{\ell\neq k}(1-\omega_{j\ell})\alpha_{k\ell}\right],$$

where $\alpha_{kk} = \mu_k$, $\alpha_{k\ell} = \mu_\ell\pi_{\ell k}$ *and* $\boldsymbol{\alpha}_k = (\alpha_{k1},\dots,\alpha_{kK})^\top$. *Defining* $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^\top,\dots,\boldsymbol{\alpha}_K^\top)^\top$ *and* $\widetilde{\boldsymbol{\alpha}} = (\widetilde{\boldsymbol{\alpha}}_1^\top,\dots,\widetilde{\boldsymbol{\alpha}}_K^\top)^\top$ *the reparameterization of* θ *and* $\widetilde{\theta}$ *respectively, we have that*

$$\begin{aligned}&\forall \boldsymbol{y}_n \in \mathbb{N}^p, f(\boldsymbol{y}_n \mid \boldsymbol{x}_n;\theta) = f(\boldsymbol{y}_n \mid \boldsymbol{x}_n;\widetilde{\theta})\\ \iff &\forall k \in \{1,\dots,K\},\ \forall j \in J_n,\ \lambda_\ell(\boldsymbol{\omega}_j;\boldsymbol{\alpha}_k) = \lambda_\ell(\boldsymbol{\omega}_j;\widetilde{\boldsymbol{\alpha}_k}).\end{aligned}$$

Hence, $\boldsymbol{\alpha}$ *and* $\widetilde{\boldsymbol{\alpha}}$ *satisfy*

$$\forall k \in \{1,\dots,K\},\ \forall j \in J_n,\ \lambda_k(\boldsymbol{\omega}_j;\boldsymbol{\alpha}_k) = \lambda_k(\boldsymbol{\omega}_j;\widetilde{\boldsymbol{\alpha}_k}).$$

Considering only the subsets of index j *that defines matrix* $\boldsymbol{S}_k$, $\boldsymbol{\alpha}$ *and* $\widetilde{\boldsymbol{\alpha}}$ *satisfy*

$$\forall k \in \{1,\dots,K\},\ \boldsymbol{S}_k(\boldsymbol{\alpha}_k - \widetilde{\boldsymbol{\alpha}_k}) = 0_K. \tag{56}$$

Let $\boldsymbol{A}$ *be the matrix of size* $K^2 \times K^2$ *composed of* $K \times K$ *matrices* $\boldsymbol{A}_{k\ell}$ *with*

$$\boldsymbol{A} = \begin{bmatrix}\boldsymbol{A}_{11} & & \cdots & & \boldsymbol{A}_{1K}\\ & \ddots & & & \\ \vdots & & \boldsymbol{A}_{k\ell} & & \vdots\\ & & & \ddots & \\ \boldsymbol{A}_{K1} & & \cdots & & \boldsymbol{A}_{KK}\end{bmatrix},$$

where $\mathbf{A}_{k\ell}$ *is a* $K \times K$ *matrix composed of zeros if* $k \neq \ell$ *and* $\mathbf{A}_{kk} = \mathbf{S}_k$. *By Assumptions 8, all the matrices* $\mathbf{S}_k$ *are invertible and so is* $\mathbf{A}$. *Noting that (56) implies*

$$\mathbf{A}(\boldsymbol{\alpha} - \widetilde{\boldsymbol{\alpha}}) = 0_{K^2},$$

and the invertibility of $\mathbf{A}$ *implies that* $\boldsymbol{\alpha} = \widetilde{\boldsymbol{\alpha}}$ *and so the identifiability of the model parameters holds.*

Proof 8 (of Lemma 4) *First we prove a formula for* $\frac{V_{nk\ell}}{n_{k\ell}}$, *then for* $\frac{W_{nk\ell}}{n_k - n_{k\ell}}$ *and, finally, we prove that the difference verifies the formula of the lemma.*

Let $\mathcal{J}_1(k,\ell) = \{j : \omega_{jk} = 1, \omega_{j\ell} = 1\}$ *be a set the set of configurations where product* ℓ, k *are both available.* Y_{njk} *is the number of realisations of a Poisson point process of intensity* $X_{nj}\lambda_k(\boldsymbol{\omega}_j, \theta)$ *on the temporal horizon* X_{nj}. *Hence the theorem central limit (TCL) for the homogeneous Poisson point process yields*

$$Y_{njk} - X_{nj}\lambda_k(\boldsymbol{\omega}_j;\theta) \underset{d}{\to} \mathcal{N}(0, (X_{nj}\lambda_k(\boldsymbol{\omega}_j;\theta)))$$

where $\mathcal{N}(\mu,\sigma^2)$ *is the normal law with mean* μ *and variance* σ^2.

Since the Y_{njk} *are independent, summing over the* $\mathcal{J}_1(k,\ell)$ *and dividing by* $n_{k\ell}$ *yields*

$$\frac{V_{nk\ell}}{n_{k\ell}} - \sum_{j\in\mathcal{J}_1(k,\ell)} \frac{X_{nj}}{n_{k\ell}}\lambda_k(\boldsymbol{\omega}_j;\theta) \underset{d}{\to} \mathcal{N}\Big(0, \sum_{j\in\mathcal{J}_1(k,\ell)} \frac{X_{nj}}{n_{k\ell}^2}\lambda_k(\boldsymbol{\omega}_j;\theta))\Big). \tag{57}$$

According to the definition of $\lambda_k(\boldsymbol{\omega}_j, \theta)$ *we have that*

$$\begin{aligned}\sum_{j\in\mathcal{J}_1(k,\ell)} \frac{X_{nj}}{n_{k\ell}}\lambda_k(\boldsymbol{\omega}_j;\theta) = \sum_{j\in\mathcal{J}_1(k,\ell)} &\Big[(\mu_k\omega_{jk} + \mu_\ell\pi_{\ell k}(1-\omega_{j\ell}) \\ &+ \sum_{h\neq k,\ell} \mu_h(1-\omega_{jh})\omega_{jk}\pi_{hk}\Big]\frac{X_{nj}}{n_{k\ell}}.\end{aligned}$$

Moreover, since $1 - \omega_{j\ell} = 0$ *for* $j \in \mathcal{J}_1(k,\ell)$, *we have that*

$$= \mu_k \sum_{j\in\mathcal{J}_1(k,\ell)} \frac{X_{nj}}{n_{k\ell}} + \sum_{h\neq k,\ell} \mu_h\pi_{hk} \sum_{j\in\mathcal{J}_1(k,\ell)} (1-\omega_{jh})\omega_{jk}\frac{X_{nj}}{n_{k\ell}}.$$

The definition of X_{nj} *implies that*

$$= \mu_k + \sum_{h\neq k,\ell} \mu_h\pi_{hk} \sum_{j\in\mathcal{J}_1(k,\ell)} (1-\omega_{jh})\omega_{jk}\frac{X_{nj}}{n_{k\ell}}. \tag{58}$$

According to Assumptions 9, $\frac{X_{nj}}{n} = \tau_j + O(n^{\frac{-1}{2}})$ *which implies that*

$$\begin{aligned}\frac{X_{nj}}{n_{k\ell}} &= \frac{X_{nj}}{n}\frac{n}{n_{k\ell}} \\ &= (\tau_j + O_{\mathbb{P}}(n^{\frac{-1}{2}}))\frac{n}{n_{k\ell}}.\end{aligned}$$

The definition of $\mathcal{J}_1(k,\ell)$ *implies that*

$$\frac{n_{k\ell}}{n} = \sum_{j\in\mathcal{J}_1(k,\ell)} \frac{X_{nj}}{n}\omega_{jk}\omega_{j\ell} = \sum_{j\in\mathcal{J}_1(k,\ell)} (\tau_j + O_{\mathbb{P}}(n^{-\frac{1}{2}})).$$

Using the definition of τ_j *in Assumptions 10 and of* $\mathcal{J}_1(k,\ell)$*, we have now that*

$$\frac{n_{k\ell}}{n} = \rho_k\rho_\ell + O_{\mathbb{P}}(n^{-\frac{1}{2}}).$$

Using the development at the second order of $\frac{n}{n_{k\ell}}$ *we obtain*

$$\frac{n}{n_{k\ell}} = (\rho_k\rho_\ell)^{-1}(1 - O_{\mathbb{P}}(n^{-\frac{1}{2}})).$$

Equation 58 then yields

$$\sum_{j\in\mathcal{J}_1(k,\ell)} \frac{X_{nj}}{n_{k\ell}}\lambda_k(\omega_j;\theta) = \mu_k + \sum_{h\neq k,\ell} \mu_h\pi_{hk} \sum_{j\in\mathcal{J}_1(k,\ell)} \Big\{(\rho_k\rho_\ell)^{-1}(1 - O_{\mathbb{P}}(n^{-\frac{1}{2}})) (\tau_j + O_{\mathbb{P}}(n^{-\frac{1}{2}}))(1-\omega_{jh})\omega_{jk}\Big\}.$$

The definition of τ_j *in Assumptions 10 implies that* $\sum_{j\in\mathcal{J}_1(k,\ell)} \tau_j(1-\omega_{jh}) = \rho_k\rho_\ell(1-\rho_h)$ *therefore*

$$= \mu_k + \sum_{h\neq k,\ell} \mu_h\pi_{hk} \sum_{j\in\mathcal{J}_1(k,\ell)} (\rho_k\rho_\ell)^{-1}(1 - O_{\mathbb{P}}(n^{-\frac{1}{2}}))(\tau_j + O_{\mathbb{P}}(n^{\frac{-1}{2}}))$$
$$= \mu_k + \sum_{h\neq k,\ell} \mu_h\pi_{hk}(1-\rho_h) + O_{\mathbb{P}}(n^{\frac{-1}{2}}).$$

Moreover, since $\sum_{j\in\mathcal{J}_1(k,\ell)} \frac{X_{nj}}{n_{k\ell}}\lambda_k(\omega_j;\theta) = O_{\mathbb{P}}(1)$ *and* $\frac{1}{n_{k\ell}} = O_{\mathbb{P}}(n^{-1})$*, we have that the variance from Equation 57 is equal to*

$$\frac{1}{n_{k\ell}} \sum_{j\in\mathcal{J}_1(k,\ell)} \frac{X_{nj}}{n_{k\ell}}\lambda_k(\omega_j;\theta) = O_{\mathbb{P}}(n^{-1}).$$

From Equation 57, we finally prove our first point which is

$$\frac{V_{nk\ell}}{n_{k\ell}} = \mu_k + \sum_{h\neq k,\ell} \mu_h\pi_{hk}(1-\rho_h) + O_{\mathbb{P}}(n^{-\frac{1}{2}}). \tag{59}$$

We apply the same reasoning to retrieve a similar expression for $\frac{W_{nk\ell}}{n_k - n_{k\ell}}$*. Let* $\mathcal{J}_2(k,\ell) = \{j : \omega_{jk} = 1, \omega_{j\ell} = 0\}$*. The TCL gives us*

$$\frac{W_{nk\ell}}{n_k - n_{k\ell}} - \sum_{j\in\mathcal{J}_2(k,\ell)} \frac{X_{nj}}{n_k - n_{k\ell}}\lambda_k(\omega_j;\theta) \underset{d}{\rightarrow} \mathcal{N}\Big(0, \sum_{j\in\mathcal{J}_2(k,\ell)} \frac{X_{nj}}{(n_k - n_{k\ell})^2}\lambda_k(\omega_j;\theta))\Big). \tag{60}$$

We have that

$$\sum_{j\in\mathcal{J}_2(k,\ell)} \frac{X_{nj}}{n_k - n_{k\ell}}\lambda_k(\omega_j;\theta) = \sum_{j\in\mathcal{J}_2(k,\ell)} \frac{n}{n_k - n_{k\ell}}\frac{X_{nj}}{n}\lambda_k(\omega_j;\theta).$$

The definition of $\lambda_k(\omega_j,\theta)$ *yields*

$$= \sum_{j\in\mathcal{J}_2(k,\ell)} \Big[\mu_k\omega_{jk} + \mu_\ell\pi_{\ell k}(1-\omega_{j\ell}) + \sum_{h\neq k,\ell} \mu_h(1-\omega_{jh})\omega_{jk}\pi_{hk}\Big]\frac{X_{nj}}{n}\frac{n}{n_k - n_{k\ell}}. \tag{61}$$

Since $\frac{X_{nj}}{n} = \tau_j + O_{\mathbb{P}}(n^{-\frac{1}{2}})$ *we have*

$$= \sum_{j \in \mathcal{J}_2(k,\ell)} \mu_k \omega_{jk} \frac{X_{nj}}{n_k - n_{k\ell}} + \sum_{j \in \mathcal{J}_2(k,\ell)} \mu_\ell \pi_{\ell k}(1-\omega_{j\ell}) \frac{X_{nj}}{n_k - n_{k\ell}}$$
$$+ \sum_{j \in \mathcal{J}_2(k,\ell)} \Big[\sum_{h \neq k,\ell} \mu_h (1-\omega_{jh}) \omega_{jk} \pi_{hk} \Big] \frac{X_{nj}}{n} \frac{n}{n_k - n_{k\ell}}.$$

According to the definition, $n_k - n_{k\ell} = \sum_{j \in \mathcal{J}_2(k,\ell)} X_{nj}$ *and* $1 - \omega_{j\ell} = 1, \omega_{jk} = 1$ *therefore*

$$= \mu_k + \mu_\ell \pi_{\ell k} + \sum_{h \neq k,\ell} \mu_h \pi_{hk} \sum_{j \in \mathcal{J}_2(k,\ell)} (1-\omega_{jh}) \frac{X_{nj}}{n} \frac{n}{n_k - n_{k\ell}}.$$

Using Assumption 9, we have that

$$= \mu_k + \mu_\ell \pi_{\ell k} + \sum_{h \neq k,\ell} \mu_h \pi_{hk} \sum_{j \in \mathcal{J}_2(k,\ell)} (1-\omega_{jh})(\tau_j + O_{\mathbb{P}}(n^{-\frac{1}{2}})) \frac{n}{n_k - n_{k\ell}} \tag{62}$$

and

$$\begin{aligned} \frac{n_k - n_{k\ell}}{n} &= \sum_{j \in \mathcal{J}_2(k,\ell)} \frac{X_{nj}}{n} \\ &= \sum_{j \in \mathcal{J}_2(k,\ell)} (\tau_j + O_{\mathbb{P}}(n^{-\frac{1}{2}})) \\ &= \rho_k (1 - \rho_\ell) + O_{\mathbb{P}}(n^{-\frac{1}{2}}). \end{aligned}$$

Hence using a limited development we obtain

$$\frac{n}{n_k - n_{k\ell}} = (\rho_k(1-\rho_\ell))^{-1}(1 - O_{\mathbb{P}}(n^{-\frac{1}{2}})).$$

Equation 62 becomes

$$\begin{aligned} &= \mu_k + \mu_\ell \pi_{\ell k} + \sum_{h \neq k,\ell} \mu_h \pi_{hk} \sum_{j \in \mathcal{J}_2(k,\ell)} (1-\omega_{jh})(\tau_j + O_{\mathbb{P}}(n^{-\frac{1}{2}}))(\rho_k(1-\rho_\ell))^{-1}(1 - O_{\mathbb{P}}(n^{-\frac{1}{2}})) \\ &= \mu_k + \mu_\ell \pi_{\ell k} + \sum_{h \neq k,\ell} \mu_h \pi_{hk} (1 - \rho_h) + O_{\mathbb{P}}(n^{-\frac{1}{2}}). \end{aligned}$$

Since $\sum_{j \in \mathcal{J}_2(k,\ell)} \frac{X_{nj}}{n} \lambda_k(\omega_j; \theta)) = O_{\mathbb{P}}(1)$ *and* $\frac{n}{(n_k - n_{k\ell})^2} = O_{\mathbb{P}}(n^{-1})$ *the variance term in Equation 60 is* $O_{\mathbb{P}}(n^{-1})$ *which implies our second objective that:*

$$\frac{W_{nk\ell}}{n_k - n_{k\ell}} - \mu_k + \mu_\ell \pi_{\ell k} + \sum_{h \neq k,\ell} \mu_h \pi_{hk}(1-\rho_h) + O_{\mathbb{P}}(n^{-\frac{1}{2}}) = O_{\mathbb{P}}(n^{-\frac{1}{2}}). \tag{63}$$

Taking the difference of

$$\frac{W_{nk\ell}}{n_k - n_{k\ell}} - \frac{V_{nk\ell}}{n_{k\ell}} = \mu_\ell \pi_{\ell k} + O_{\mathbb{P}}(n^{-\frac{1}{2}}). \tag{64}$$

Lemma 7 *Let* U *and* V *be two independent random variables that follow Poisson distributions with parameters* μ *and* ν *respectively. Then, for any positive integer* c*, the conditional expectation of* U *given* $U + V = c$ *is*

$$\mathbb{E}[U \mid U + V = c] = c \frac{\mu}{\mu + \nu}.$$

Proof 9 (of Lemma 7) *Noting that* $P(U=u \mid U+V=c) = P(U=u)P(V=c-u)/P(U+V=c)$, *we have*

$$\mathbb{E}[U \mid U+V=c] = \frac{c!}{e^{-\mu-\nu}(\mu+\nu)^c}\sum_{u=0}^{c} u\frac{e^{-\mu}}{u!}\mu^u\frac{e^{-\nu}}{(c-u)!}\nu^{c-u}.$$

If $c=0$, *then* $\mathbb{E}[U \mid U+V=c]=0$. *Now, if* c *is an integer that is strictly positive, then*

$$\begin{aligned}\mathbb{E}[U \mid U+V=c] &= \frac{c!}{(\mu+\nu)^c}\mu\sum_{u=1}^{c}\frac{1}{(u-1)!}\mu^{u-1}\frac{1}{(c-1-(u-1))!}\nu^{c-1-(u-1)}\\ &= \frac{c}{(\mu+\nu)^c}\mu\sum_{u=0}^{c-1}\frac{(c-1)!}{u!(c-1-u)!}\mu^u\nu^{c-1-u}\\ &= \frac{c}{(\mu+\nu)^c}\mu(\mu+\nu)^{c-1}.\end{aligned}$$

Lemma 8 *From the optimisation problem 40, we have that :*

$$\mu_\ell^{[r]} = n^{-1}(\sum_{j=1}^{J}\sum_{k=0}^{K} z_{njk\ell}^{[r-1]}).$$

Proof 10 *Since*

$$\lambda_{k\ell}(\boldsymbol{\omega}_j;\theta) = \begin{cases}\mu_\ell\omega_{j\ell} & \text{if } k=\ell\\ \mu_\ell(1-\omega_{j\ell})\pi_{\ell k}\omega_{jk} & \text{if } k\neq\ell,\ k\neq 0\\ \mu_\ell(1-\omega_{j\ell})\boldsymbol{\pi}_\ell^\top(\mathbf{1}_K-\boldsymbol{\omega}_j) & \text{if } k=0,\end{cases}$$

we obtain

$$\frac{\partial}{\partial\mu_\ell}\lambda_{k\ell}(\boldsymbol{\omega}_j;\theta) = \lambda_{k\ell}(\boldsymbol{\omega}_j;\theta)/\mu_\ell.$$

We notice also that $\mu_\ell^{[r]}$ *verifies*

$$\frac{\partial\tilde{F}_\ell(\mu_\ell^{[r]},\boldsymbol{\pi}^{[r]})}{\partial\mu_\ell} = 0.$$

Then we have $\mathbf{z}_n^{[r-1]} = \{z_{njk\ell}^{[r-1]}\}_{j,k,\ell}$, *leading to*

$$\sum_{j=1}^{J}\sum_{k=0}^{K} z_{njk\ell}^{[r-1]}\frac{1}{\mu_\ell^{[r]}} - x_{nj}\lambda_{k\ell}(\boldsymbol{\omega}_j;\theta^{[r]})\frac{1}{\mu_\ell^{[r]}} = 0$$

which yields

$$\begin{aligned}\mu_\ell^{[r]} &= \frac{\sum_{j=1}^{J}\sum_{k=0}^{K} z_{njk\ell}^{[r-1]}}{\sum_{j=1}^{J}\sum_{k=0}^{K} x_{nj}\lambda_{k\ell}(\boldsymbol{\omega}_j;\theta^{[r]})\frac{1}{\mu_\ell^{[r]}}}\\ &= \frac{\sum_{j=1}^{J}\sum_{k=0}^{K} z_{njk\ell}^{[r-1]}}{\sum_{j=1}^{J} x_{nj}[\omega_{j\ell}+\sum_{\substack{k=1\\k\neq\ell}}^{K}(1-\omega_{j\ell})\pi_{\ell k}^{[r]}\omega_{jk}+(1-\omega_{j\ell})\boldsymbol{\pi}_\ell^{[r]\top}(\mathbf{1}_K-\boldsymbol{\omega}_j)]}\\ &= \frac{\sum_{j=1}^{J}\sum_{k=0}^{K} z_{njk\ell}^{[r-1]}}{\sum_{j=1}^{J} x_{nj}[\omega_{j\ell}+(1-\omega_{j\ell})\sum_{k=1}^{K}\pi_{\ell k}^{[r-1]}]}.\end{aligned}$$

Finally, since $\sum_{k=1}^{K}\pi_{\ell k}^{[r]}=1$ *we do have the result.*

Lemma 9 *$\tilde{H}_\ell(\mu_\ell^{[r]}, \boldsymbol{\pi}_\ell)$ is semi definite positive.*

Proof 11 *For $\ell \in [K]$, we have that*

$$\tilde{F}_\ell(\mu_\ell, \boldsymbol{\pi}_\ell) = -\Big[\sum_{j=1}^{J}\sum_{k=0}^{K} z_{njk\ell}^{[r]} \ln[x_{nj}\lambda_{k\ell}(\boldsymbol{\omega}_j;\theta)] - x_{nj}\lambda_{k\ell}(\boldsymbol{\omega}_j;\theta)\Big]. \tag{65}$$

From Model 37, we have that $\lambda_{k\ell}(\boldsymbol{\omega}_j;\theta)$ is linear in $\pi_{\ell k}$ hence its second derivative with respect to $\boldsymbol{\pi}_\ell$ is null. The derivation of the first term with respect to $\pi_{\ell h}$ based on Model 37 yields for $k \neq 0$ if $(h,h') \neq (k,k)$, $\frac{\partial^2 \ln \lambda_{k\ell}(\boldsymbol{\omega}_j,\theta)}{\partial \pi_{\ell h}\partial \pi_{\ell h'}} = 0$. We also have that $\frac{\partial \ln \lambda_{k\ell}(\boldsymbol{\omega}_j,\theta)}{\partial \pi_{\ell k}\partial \pi_{\ell k}} = -\frac{\mu_\ell(1-\omega_{j\ell})\omega_{jk}}{\pi_{\ell k}^2}$ for $k \neq 0$ and $\forall (h,h') \in [K]^2$, $\frac{\partial^2 \ln(\lambda_{0\ell}(\boldsymbol{\omega}_j,\theta)))}{\partial \pi_{\ell h}\partial \pi_{\ell h'}} = -\frac{(1-\omega_{jh})(1-\omega_{jh'})}{[\sum_{k''=0}^{K}\pi_{\ell k''}(1-\omega_{jk''})]^2}$. These equations yield the following formulas

$$\frac{\partial^2}{\partial^2 \pi_{\ell\ell}}\Big(\sum_{k=1}^{K}\ln(\lambda_{k\ell}(\boldsymbol{\omega}_j,\theta))\Big) = -\frac{(1-\omega_{j\ell})}{[\sum_{k=0}^{K}\pi_{\ell k'}(1-\omega_{jk'})]^2} \text{ for } k' \neq k;$$

$$\frac{\partial^2}{\partial^2 \pi_{\ell k}}\Big(\sum_{k=1}^{K}\ln(\lambda_{k\ell}(\boldsymbol{\omega}_j,\theta))\Big) = -\left(\frac{\mu_\ell(1-\omega_{j\ell})\omega_{jk}}{\pi_{\ell k}^2} + \frac{(1-\omega_{jk})(1-\omega_{j\ell})}{[\sum_{k'=0}^{K}\pi_{\ell k'}(1-\omega_{jk'})]^2}\right);$$

$$\frac{\partial^2}{\partial \pi_{\ell h}\partial \pi_{\ell h'}}\left(\sum_{k=1}^{K}\ln(\lambda_{k\ell}(\boldsymbol{\omega}_j,\theta))\right) = -\frac{(1-\omega_{jh})(1-\omega_{jh'})}{[\sum_{k''=0}^{K}\pi_{\ell k''}(1-\omega_{jk''})]^2} \text{ for any } h' \neq h.$$

Hence the second partial derivatives are given by the following formulas:

$$\frac{\partial^2 \tilde{F}_\ell(\mu_\ell,\pi_\ell)}{\partial^2 \pi_{\ell\ell}} = n^{-1}\sum_{j=1}^{J} z_{nj0\ell}\frac{(1-\omega_{j\ell})}{[\sum_{k=0}^{K}\pi_{\ell k'}(1-\omega_{jk'})]^2};$$

$$\frac{\partial^2 \tilde{F}_\ell(\mu_\ell,\pi_\ell)}{\partial^2 \pi_{\ell k}} = n^{-1}\left(\frac{r_{\ell k}}{\pi_{\ell k}^2} + \sum_{j=1}^{J} z_{nj0\ell}\frac{(1-\omega_{jk})(1-\omega_{j\ell})}{[\sum_{k'=0}^{K}\pi_{\ell k'}(1-\omega_{jk'})]^2}\right);$$

$$\frac{\partial^2 \tilde{F}_\ell(\mu_\ell,\pi_\ell)}{\partial \pi_{\ell k}\partial \pi_{\ell k'}} = n^{-1}\sum_{j=1}^{J} z_{nj0\ell}\frac{(1-\omega_{jk})(1-\omega_{jk'})}{[\sum_{k''=0}^{K}\pi_{\ell k''}(1-\omega_{jk''})]^2};$$

$$\frac{\partial^2 \tilde{F}_\ell(\mu_\ell,\pi_\ell)}{\partial \pi_{\ell k}\partial \pi_{\ell\ell}} = n^{-1}\sum_{j=1}^{J} z_{nj0\ell}\frac{(1-\omega_{j\ell})(1-\omega_{jk})}{[\sum_{k'=0}^{K}\pi_{\ell k'}(1-\omega_{jk'})]^2}$$

where $r_{\ell k} = \sum_{j=1}^{J}\sum_{k=1}^{K} z_{nk\ell}$. We conclude since $z_{nk\ell}, \pi_{\ell k}, (1-\omega_{jk})$ are positive and $\forall (k,\ell) \in [K]^2$ the Hessian with respect to $\boldsymbol{\pi}_\ell$ is positive.

B.3 ADDITIONAL NUMERICAL EXPERIMENTS RESULTS

Additional tables related to Section 4.5.1.2. We chose to restrict the $\Delta_{\pi_1}^{\mathcal{S},\mathcal{A}} < 0.005$ and $\Delta_\ell^{\mathcal{S},\mathcal{A}} \leqslant 10^{-3}$. The setting of parameters $\epsilon_{em} = 10^{-5}$ and $N_M \in \{5, 10\}$ is compatible with all the $\mathcal{S}$ with $K = 2$ (Table 49 to 63). Since all the $N_{init} \in \{10, 25, 50\}$ are compatible we chose 10 for its lowest time of computation. In the case $\mathcal{S}$ with $K = 5$ we also have that this setting is compatible with certain values exceeding slightly the threshold for Ninit $= 10$ while still being in the standard deviation. The chosen set of parameters of the EM $\mathcal{A}^* = (10^{-5}, 5, 10)$ is then compatible with the settings of the parameters.

Table 49: Results from the numerical applications for $\mathcal{S} = (2, 180, 0.1, 0.0)$

ϵ_{em}	N_M	N_{init}	$\Delta_{\ell}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\pi_1}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\mu_1}^{\mathcal{S},\mathcal{A}}$(std)	mean execution time [s]	mean number iterations
1e-05	2	10	0.000 (0.000)	0.004 (0.007)	0.000 (0.000)	3.570	86.630 (32.820)
		25	0.000 (0.000)	0.003 (0.005)	0.000 (0.000)	8.584	81.300 (34.773)
		50	0.000 (0.000)	0.002 (0.004)	0.000 (0.000)	17.178	82.760 (41.559)
	5	10	0.000 (0.000)	0.003 (0.007)	0.000 (0.000)	4.038	76.640 (38.082)
		25	0.000 (0.000)	0.003 (0.006)	0.000 (0.000)	9.651	69.120 (35.956)
		50	0.000 (0.000)	0.000 (0.003)	0.000 (0.000)	19.299	65.810 (40.332)
	10	10	0.000 (0.000)	0.004 (0.007)	0.000 (0.000)	4.175	78.560 (39.350)
		25	0.000 (0.000)	0.003 (0.006)	0.000 (0.000)	9.991	70.190 (36.406)
		50	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	19.978	67.100 (41.203)
0.001	2	10	0.007 (0.003)	0.029 (0.034)	0.001 (0.001)	1.472	32.950 (17.566)
		25	0.006 (0.003)	0.029 (0.036)	0.001 (0.001)	3.528	29.200 (20.058)
		50	0.005 (0.003)	0.027 (0.035)	0.001 (0.001)	7.062	26.110 (18.343)
	5	10	0.005 (0.004)	0.034 (0.048)	0.001 (0.001)	1.723	26.970 (12.927)
		25	0.004 (0.003)	0.031 (0.038)	0.001 (0.001)	4.062	23.050 (13.352)
		50	0.004 (0.003)	0.028 (0.036)	0.001 (0.001)	8.136	22.050 (12.936)
	10	10	0.005 (0.004)	0.031 (0.037)	0.001 (0.001)	1.794	27.490 (14.117)
		25	0.004 (0.003)	0.031 (0.038)	0.001 (0.001)	4.239	23.410 (13.920)
		50	0.004 (0.003)	0.028 (0.036)	0.001 (0.001)	8.493	22.380 (13.649)
0.1	2	10	0.181 (0.150)	0.161 (0.137)	0.010 (0.006)	0.332	5.670 (2.973)
		25	0.134 (0.132)	0.121 (0.102)	0.009 (0.006)	0.793	5.140 (2.698)
		50	0.106 (0.106)	0.098 (0.092)	0.008 (0.006)	1.582	4.780 (2.296)
	5	10	0.143 (0.095)	0.145 (0.142)	0.009 (0.006)	0.442	6.090 (3.723)
		25	0.106 (0.078)	0.103 (0.084)	0.008 (0.005)	1.026	5.410 (3.144)
		50	0.087 (0.074)	0.085 (0.073)	0.007 (0.005)	2.061	5.330 (3.344)
	10	10	0.144 (0.096)	0.144 (0.142)	0.009 (0.006)	0.443	6.030 (3.675)
		25	0.106 (0.079)	0.102 (0.084)	0.008 (0.006)	1.047	5.370 (3.104)
		50	0.087 (0.075)	0.084 (0.073)	0.007 (0.006)	2.087	5.210 (3.213)

Table 50: Results from the numerical applications for $\mathcal{S} = (2, 180, 2.0, 0.0)$

ϵ_{em}	N_M	N_{init}	$\Delta_{\ell}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\pi_1}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\mu_1}^{\mathcal{S},\mathcal{A}}$(std)	mean execution time [s]	mean number iterations
1e-05	2	10	0.000 (0.000)	0.001 (0.002)	0.001 (0.001)	8.390	225.930 (108.723)
		25	-0.000 (0.000)	0.001 (0.002)	0.001 (0.001)	20.849	211.350 (118.389)
		50	-0.000 (0.000)	0.001 (0.001)	0.000 (0.001)	41.553	191.820 (118.519)
	5	10	0.000 (0.000)	0.001 (0.002)	0.001 (0.001)	10.019	233.520 (106.616)
		25	0.000 (0.000)	0.001 (0.002)	0.000 (0.001)	24.947	224.830 (111.148)
		50	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	49.702	202.840 (115.154)
	10	10	0.000 (0.000)	0.001 (0.002)	0.000 (0.001)	10.352	234.730 (106.662)
		25	0.000 (0.000)	0.001 (0.002)	0.000 (0.001)	25.741	222.380 (111.703)
		50	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	51.306	200.540 (113.903)
0.001	2	10	0.021 (0.013)	0.020 (0.017)	0.009 (0.007)	3.231	75.850 (36.789)
		25	0.018 (0.012)	0.016 (0.014)	0.008 (0.006)	8.041	61.460 (34.795)
		50	0.016 (0.011)	0.012 (0.012)	0.006 (0.005)	16.001	54.720 (27.883)
	5	10	0.022 (0.012)	0.020 (0.017)	0.010 (0.007)	3.971	79.560 (34.770)
		25	0.019 (0.011)	0.016 (0.014)	0.008 (0.007)	9.889	61.620 (29.585)
		50	0.016 (0.010)	0.012 (0.011)	0.006 (0.005)	19.691	57.930 (29.168)
	10	10	0.022 (0.012)	0.020 (0.017)	0.010 (0.007)	4.122	79.770 (35.333)
		25	0.019 (0.011)	0.016 (0.014)	0.008 (0.007)	10.259	61.760 (29.109)
		50	0.016 (0.010)	0.012 (0.011)	0.006 (0.005)	20.431	57.710 (28.589)
0.1	2	10	0.546 (0.331)	0.118 (0.086)	0.064 (0.039)	0.778	11.620 (7.757)
		25	0.400 (0.254)	0.083 (0.063)	0.054 (0.028)	1.943	8.610 (8.530)
		50	0.328 (0.235)	0.074 (0.061)	0.051 (0.026)	3.859	6.990 (7.429)
	5	10	0.596 (0.356)	0.125 (0.085)	0.068 (0.040)	1.088	11.620 (9.503)
		25	0.436 (0.297)	0.087 (0.062)	0.057 (0.030)	2.699	7.120 (4.330)
		50	0.356 (0.273)	0.077 (0.058)	0.053 (0.029)	5.379	6.010 (3.882)
	10	10	0.594 (0.354)	0.125 (0.086)	0.068 (0.039)	1.151	13.260 (13.690)
		25	0.435 (0.296)	0.087 (0.062)	0.057 (0.030)	2.831	7.670 (7.928)
		50	0.355 (0.272)	0.077 (0.058)	0.053 (0.029)	5.647	6.560 (7.772)

Table 51: Results from the numerical applications for $\mathcal{S} = (2, 180, 2.0, 0.5)$

ϵ_{em}	N_M	N_{init}	$\Delta_{\ell}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\pi_1}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\mu_1}^{\mathcal{S},\mathcal{A}}$(std)	mean execution time [s]	mean number iterations
1e-05	2	10	0.000 (0.000)	0.002 (0.002)	0.001 (0.001)	3.985	86.690 (58.839)
		25	0.000 (0.000)	0.002 (0.002)	0.001 (0.001)	9.973	77.880 (67.292)
		50	-0.000 (0.000)	0.002 (0.002)	0.001 (0.001)	20.069	63.990 (65.493)
	5	10	0.000 (0.000)	0.002 (0.002)	0.001 (0.001)	5.087	101.650 (67.429)
		25	0.000 (0.000)	0.001 (0.002)	0.001 (0.001)	12.677	84.160 (67.725)
		50	0.000 (0.000)	0.000 (0.001)	0.000 (0.000)	25.500	70.180 (66.871)
	10	10	0.000 (0.000)	0.002 (0.002)	0.001 (0.001)	5.298	100.500 (67.118)
		25	0.000 (0.000)	0.001 (0.002)	0.001 (0.001)	13.138	85.930 (72.234)
		50	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	26.475	70.190 (67.090)
0.001	2	10	0.006 (0.006)	0.018 (0.013)	0.011 (0.006)	1.935	30.020 (18.765)
		25	0.004 (0.004)	0.013 (0.010)	0.007 (0.005)	4.815	21.480 (16.904)
		50	0.003 (0.002)	0.009 (0.007)	0.006 (0.004)	9.721	18.130 (6.685)
	5	10	0.007 (0.005)	0.021 (0.012)	0.011 (0.006)	2.591	37.470 (21.425)
		25	0.004 (0.003)	0.015 (0.010)	0.008 (0.005)	6.401	25.050 (20.832)
		50	0.003 (0.002)	0.011 (0.008)	0.007 (0.005)	12.915	20.580 (14.049)
	10	10	0.007 (0.005)	0.021 (0.012)	0.011 (0.006)	2.710	40.090 (24.859)
		25	0.004 (0.003)	0.015 (0.011)	0.007 (0.005)	6.709	25.640 (22.200)
		50	0.003 (0.002)	0.011 (0.008)	0.007 (0.005)	13.535	20.160 (9.886)
0.1	2	10	0.125 (0.108)	0.062 (0.058)	0.044 (0.034)	0.553	8.490 (2.674)
		25	0.075 (0.044)	0.037 (0.034)	0.034 (0.022)	1.382	7.590 (2.926)
		50	0.057 (0.031)	0.029 (0.027)	0.033 (0.020)	2.763	7.320 (3.095)
	5	10	0.174 (0.146)	0.079 (0.065)	0.053 (0.042)	0.743	8.130 (2.855)
		25	0.084 (0.046)	0.043 (0.035)	0.035 (0.024)	1.846	6.880 (2.662)
		50	0.065 (0.034)	0.034 (0.028)	0.035 (0.021)	3.716	6.600 (2.731)
	10	10	0.174 (0.146)	0.079 (0.065)	0.053 (0.042)	0.770	8.130 (2.855)
		25	0.084 (0.046)	0.043 (0.035)	0.035 (0.024)	1.885	6.880 (2.662)
		50	0.065 (0.034)	0.034 (0.028)	0.035 (0.021)	3.799	6.600 (2.731)

Table 52: Results from the numerical applications for $\mathcal{S} = (2, 180, 0.1, 1.0)$

ϵ_{em}	N_M	N_{init}	$\Delta_{\ell}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\pi_1}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\mu_1}^{\mathcal{S},\mathcal{A}}$(std)	mean execution time [s]	mean number iterations
1e-05	2	10	0.000 (0.000)	0.005 (0.020)	0.000 (0.000)	3.362	88.370 (22.613)
		25	0.000 (0.000)	0.003 (0.007)	0.000 (0.000)	8.392	83.220 (24.760)
		50	0.000 (0.000)	0.001 (0.003)	0.000 (0.000)	16.776	76.690 (25.911)
	5	10	0.000 (0.000)	0.005 (0.020)	0.000 (0.000)	3.192	64.330 (32.287)
		25	0.000 (0.000)	0.002 (0.006)	0.000 (0.000)	7.969	58.480 (32.234)
		50	-0.000 (0.000)	0.000 (0.001)	0.000 (0.000)	15.920	50.880 (28.213)
	10	10	0.000 (0.000)	0.005 (0.020)	0.000 (0.000)	3.274	64.760 (33.035)
		25	0.000 (0.000)	0.001 (0.006)	0.000 (0.000)	8.181	57.040 (32.143)
		50	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	16.368	49.400 (29.322)
0.001	2	10	0.007 (0.002)	0.040 (0.058)	0.001 (0.001)	1.404	33.700 (15.047)
		25	0.007 (0.002)	0.033 (0.040)	0.001 (0.001)	3.499	33.390 (15.523)
		50	0.006 (0.003)	0.031 (0.041)	0.001 (0.001)	7.009	30.730 (15.624)
	5	10	0.004 (0.003)	0.033 (0.055)	0.001 (0.001)	1.230	19.230 (7.261)
		25	0.004 (0.003)	0.027 (0.041)	0.001 (0.001)	3.063	18.170 (7.203)
		50	0.003 (0.003)	0.020 (0.031)	0.001 (0.001)	6.128	17.050 (6.635)
	10	10	0.004 (0.003)	0.033 (0.055)	0.001 (0.001)	1.254	19.020 (7.555)
		25	0.004 (0.003)	0.027 (0.041)	0.001 (0.001)	3.127	18.070 (7.445)
		50	0.003 (0.003)	0.020 (0.031)	0.001 (0.001)	6.265	16.730 (6.820)
0.1	2	10	0.219 (0.176)	0.256 (0.222)	0.008 (0.005)	0.227	4.500 (1.091)
		25	0.144 (0.126)	0.161 (0.144)	0.007 (0.004)	0.566	4.220 (0.701)
		50	0.106 (0.100)	0.139 (0.139)	0.006 (0.004)	1.139	4.100 (0.557)
	5	10	0.110 (0.074)	0.169 (0.182)	0.006 (0.004)	0.308	5.410 (2.103)
		25	0.085 (0.060)	0.121 (0.121)	0.006 (0.004)	0.768	4.930 (1.899)
		50	0.065 (0.048)	0.097 (0.100)	0.005 (0.003)	1.539	4.680 (1.618)
	10	10	0.105 (0.076)	0.170 (0.183)	0.006 (0.004)	0.318	5.390 (2.049)
		25	0.082 (0.061)	0.120 (0.122)	0.006 (0.004)	0.790	4.870 (1.641)
		50	0.063 (0.048)	0.094 (0.101)	0.005 (0.003)	1.585	4.720 (1.537)

Table 53: Results from the numerical applications for $\mathcal{S} = (2, 180, 2.0, 1.0)$

ϵ_{em}	N_M	N_{init}	$\Delta_{\ell}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\pi_1}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\mu_1}^{\mathcal{S},\mathcal{A}}$(std)	mean execution time [s]	mean number iterations
1e-05	2	10	-0.000 (0.000)	0.000 (0.001)	0.000 (0.001)	6.585	186.610 (78.270)
		25	-0.000 (0.000)	0.000 (0.001)	0.000 (0.001)	16.519	186.590 (78.345)
		50	-0.000 (0.000)	0.000 (0.001)	0.000 (0.001)	32.974	180.170 (79.140)
	5	10	0.000 (0.000)	0.000 (0.001)	0.000 (0.001)	7.953	193.170 (76.177)
		25	0.000 (0.000)	0.000 (0.001)	0.000 (0.001)	19.856	189.360 (80.970)
		50	-0.000 (0.000)	0.000 (0.000)	0.000 (0.001)	39.608	180.040 (83.740)
	10	10	0.000 (0.000)	0.000 (0.001)	0.000 (0.001)	8.201	194.660 (76.979)
		25	0.000 (0.000)	0.000 (0.001)	0.000 (0.001)	20.479	190.840 (81.113)
		50	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	40.881	182.690 (83.894)
0.001	2	10	0.018 (0.009)	0.013 (0.012)	0.010 (0.008)	2.519	67.560 (16.283)
		25	0.017 (0.009)	0.013 (0.010)	0.010 (0.008)	6.318	64.050 (18.004)
		50	0.016 (0.008)	0.012 (0.010)	0.010 (0.007)	12.580	59.420 (17.285)
	5	10	0.018 (0.008)	0.014 (0.012)	0.011 (0.008)	3.035	68.550 (16.889)
		25	0.017 (0.008)	0.013 (0.010)	0.010 (0.008)	7.581	62.440 (20.031)
		50	0.016 (0.008)	0.012 (0.010)	0.009 (0.008)	15.071	56.580 (16.828)
	10	10	0.018 (0.008)	0.014 (0.012)	0.011 (0.008)	3.133	68.930 (16.411)
		25	0.017 (0.008)	0.013 (0.010)	0.010 (0.008)	7.824	62.940 (19.400)
		50	0.016 (0.008)	0.012 (0.010)	0.009 (0.008)	15.555	56.550 (16.630)
0.1	2	10	0.601 (0.189)	0.111 (0.070)	0.091 (0.047)	0.754	18.620 (5.919)
		25	0.527 (0.186)	0.095 (0.070)	0.084 (0.050)	1.887	16.610 (5.943)
		50	0.478 (0.204)	0.083 (0.073)	0.077 (0.051)	3.736	15.090 (5.870)
	5	10	0.610 (0.210)	0.108 (0.072)	0.092 (0.046)	0.932	17.620 (6.397)
		25	0.531 (0.218)	0.095 (0.069)	0.076 (0.045)	2.320	15.420 (6.223)
		50	0.478 (0.221)	0.088 (0.067)	0.070 (0.047)	4.590	13.270 (5.451)
	10	10	0.611 (0.207)	0.109 (0.072)	0.093 (0.046)	0.966	17.750 (6.674)
		25	0.531 (0.217)	0.095 (0.067)	0.077 (0.045)	2.407	14.990 (5.799)
		50	0.478 (0.220)	0.088 (0.067)	0.070 (0.046)	4.758	13.210 (5.529)

Table 54: Results from the numerical applications for $\mathcal{S} = (2, 365, 0.1, 0.0)$

ϵ_{em}	N_M	N_{init}	$\Delta_{\ell}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\pi_1}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\mu_1}^{\mathcal{S},\mathcal{A}}$(std)	mean execution time [s]	mean number iterations
1e-05	2	10	0.000 (0.000)	0.003 (0.006)	0.000 (0.000)	4.105	102.070 (49.427)
		25	0.000 (0.000)	0.003 (0.005)	0.000 (0.000)	10.341	89.560 (45.302)
		50	-0.000 (0.000)	0.002 (0.005)	0.000 (0.000)	20.714	82.810 (43.965)
	5	10	0.000 (0.000)	0.003 (0.006)	0.000 (0.000)	4.761	104.660 (51.733)
		25	0.000 (0.000)	0.002 (0.005)	0.000 (0.000)	12.022	88.110 (48.774)
		50	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	24.032	81.500 (46.237)
	10	10	0.000 (0.000)	0.003 (0.007)	0.000 (0.000)	4.973	106.930 (57.842)
		25	0.000 (0.000)	0.001 (0.005)	0.000 (0.000)	12.550	90.620 (52.270)
		50	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	25.068	82.520 (47.469)
0.001	2	10	0.009 (0.004)	0.031 (0.031)	0.001 (0.001)	1.567	34.250 (17.088)
		25	0.007 (0.004)	0.025 (0.023)	0.001 (0.001)	3.949	30.740 (19.046)
		50	0.007 (0.004)	0.024 (0.023)	0.001 (0.001)	7.915	30.010 (22.662)
	5	10	0.008 (0.005)	0.033 (0.034)	0.001 (0.001)	1.879	30.260 (15.955)
		25	0.007 (0.005)	0.024 (0.025)	0.001 (0.001)	4.750	26.710 (14.317)
		50	0.006 (0.004)	0.022 (0.025)	0.001 (0.001)	9.520	23.740 (12.093)
	10	10	0.008 (0.005)	0.033 (0.034)	0.001 (0.001)	1.967	30.900 (17.389)
		25	0.007 (0.005)	0.024 (0.024)	0.001 (0.001)	5.001	26.840 (14.514)
		50	0.006 (0.004)	0.022 (0.025)	0.001 (0.001)	10.009	24.060 (12.957)
0.1	2	10	0.205 (0.159)	0.164 (0.121)	0.007 (0.004)	0.368	6.300 (3.189)
		25	0.147 (0.130)	0.116 (0.108)	0.006 (0.004)	0.924	4.910 (2.254)
		50	0.112 (0.096)	0.094 (0.091)	0.005 (0.004)	1.852	4.410 (1.588)
	5	10	0.176 (0.120)	0.139 (0.119)	0.006 (0.005)	0.471	6.220 (3.094)
		25	0.127 (0.093)	0.108 (0.095)	0.005 (0.004)	1.181	4.880 (2.183)
		50	0.100 (0.078)	0.088 (0.074)	0.005 (0.004)	2.348	4.450 (1.841)
	10	10	0.175 (0.117)	0.139 (0.118)	0.006 (0.004)	0.478	6.250 (3.297)
		25	0.127 (0.092)	0.108 (0.097)	0.005 (0.004)	1.203	5.110 (3.313)
		50	0.100 (0.078)	0.088 (0.074)	0.005 (0.004)	2.389	4.450 (1.841)

Table 55: Results from the numerical applications for $\mathcal{S} = (2, 365, 2.0, 0.0)$

ϵ_{em}	N_M	N_{init}	$\Delta_{\ell}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\pi_1}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\mu_1}^{\mathcal{S},\mathcal{A}}$(std)	mean execution time [s]	mean number iterations
1e-05	2	10	0.000 (0.000)	0.001 (0.002)	0.001 (0.001)	11.290	321.510 (138.344)
		25	0.000 (0.000)	0.001 (0.001)	0.000 (0.001)	28.149	294.560 (146.399)
		50	0.000 (0.000)	0.000 (0.001)	0.000 (0.001)	56.292	275.290 (156.234)
	5	10	0.000 (0.000)	0.001 (0.002)	0.001 (0.001)	13.277	330.550 (138.467)
		25	0.000 (0.000)	0.000 (0.001)	0.000 (0.001)	33.120	296.730 (157.933)
		50	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	66.269	281.320 (160.321)
	10	10	0.000 (0.000)	0.001 (0.002)	0.000 (0.001)	13.663	333.300 (140.156)
		25	0.000 (0.000)	0.000 (0.001)	0.000 (0.001)	34.091	299.840 (157.458)
		50	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	68.255	279.650 (158.077)
0.001	2	10	0.032 (0.017)	0.021 (0.013)	0.010 (0.006)	4.125	100.620 (42.170)
		25	0.028 (0.017)	0.016 (0.012)	0.008 (0.006)	10.278	77.740 (38.349)
		50	0.024 (0.016)	0.014 (0.011)	0.006 (0.005)	20.583	66.660 (29.670)
	5	10	0.032 (0.017)	0.020 (0.014)	0.010 (0.006)	4.878	98.390 (43.740)
		25	0.028 (0.017)	0.016 (0.012)	0.007 (0.006)	12.163	80.570 (39.744)
		50	0.024 (0.016)	0.014 (0.011)	0.006 (0.005)	24.374	70.270 (30.610)
	10	10	0.032 (0.017)	0.020 (0.014)	0.010 (0.006)	5.047	99.440 (44.334)
		25	0.028 (0.017)	0.017 (0.012)	0.007 (0.006)	12.577	81.610 (40.906)
		50	0.024 (0.016)	0.014 (0.011)	0.006 (0.005)	25.215	70.160 (30.312)
0.1	2	10	0.785 (0.471)	0.128 (0.071)	0.065 (0.035)	0.961	12.750 (9.712)
		25	0.559 (0.400)	0.090 (0.067)	0.046 (0.029)	2.404	8.410 (5.724)
		50	0.421 (0.296)	0.071 (0.056)	0.040 (0.025)	4.823	7.130 (5.392)
	5	10	0.810 (0.467)	0.127 (0.068)	0.064 (0.034)	1.279	14.190 (13.456)
		25	0.580 (0.407)	0.092 (0.069)	0.048 (0.028)	3.201	8.880 (7.125)
		50	0.461 (0.350)	0.077 (0.061)	0.042 (0.026)	6.419	6.850 (5.070)
	10	10	0.810 (0.467)	0.128 (0.069)	0.063 (0.034)	1.336	15.430 (15.952)
		25	0.580 (0.407)	0.092 (0.069)	0.048 (0.028)	3.346	8.760 (6.524)
		50	0.461 (0.350)	0.077 (0.061)	0.042 (0.026)	6.714	6.850 (5.070)

Table 56: Results from the numerical applications for $\mathcal{S} = (2, 365, 0.1, 0.5)$

ϵ_{em}	N_M	N_{init}	$\Delta_{\ell}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\pi_1}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\mu_1}^{\mathcal{S},\mathcal{A}}$(std)	mean execution time [s]	mean number iterations
1e-05	2	10	0.000 (0.000)	0.004 (0.007)	0.000 (0.000)	3.618	87.820 (42.990)
		25	0.000 (0.000)	0.003 (0.005)	0.000 (0.000)	9.115	84.740 (48.836)
		50	-0.000 (0.000)	0.002 (0.003)	0.000 (0.000)	18.148	81.120 (51.864)
	5	10	0.000 (0.000)	0.004 (0.008)	0.000 (0.000)	4.255	85.770 (43.942)
		25	0.000 (0.000)	0.002 (0.006)	0.000 (0.000)	10.718	75.660 (42.986)
		50	0.000 (0.000)	0.000 (0.001)	0.000 (0.000)	21.354	71.060 (46.918)
	10	10	0.000 (0.000)	0.004 (0.008)	0.000 (0.000)	4.409	85.610 (44.183)
		25	0.000 (0.000)	0.002 (0.005)	0.000 (0.000)	11.124	77.240 (43.233)
		50	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	22.145	72.340 (47.793)
0.001	2	10	0.007 (0.004)	0.030 (0.034)	0.001 (0.001)	1.361	29.400 (15.375)
		25	0.006 (0.004)	0.026 (0.026)	0.001 (0.001)	3.428	26.480 (15.439)
		50	0.005 (0.004)	0.023 (0.024)	0.001 (0.001)	6.810	25.750 (16.826)
	5	10	0.007 (0.004)	0.036 (0.036)	0.001 (0.001)	1.637	27.970 (14.057)
		25	0.006 (0.004)	0.030 (0.029)	0.001 (0.001)	4.113	23.420 (12.557)
		50	0.005 (0.004)	0.026 (0.027)	0.001 (0.001)	8.193	21.570 (11.938)
	10	10	0.007 (0.004)	0.036 (0.036)	0.001 (0.001)	1.694	28.240 (14.601)
		25	0.006 (0.004)	0.030 (0.030)	0.001 (0.001)	4.280	24.530 (14.988)
		50	0.005 (0.004)	0.026 (0.027)	0.001 (0.001)	8.525	22.360 (13.760)
0.1	2	10	0.180 (0.162)	0.165 (0.126)	0.007 (0.005)	0.298	5.990 (2.528)
		25	0.136 (0.130)	0.140 (0.125)	0.005 (0.004)	0.741	5.330 (1.965)
		50	0.099 (0.097)	0.101 (0.089)	0.005 (0.003)	1.485	4.980 (1.903)
	5	10	0.150 (0.112)	0.140 (0.117)	0.006 (0.004)	0.352	6.010 (2.335)
		25	0.114 (0.091)	0.113 (0.091)	0.006 (0.004)	0.869	5.420 (2.036)
		50	0.085 (0.071)	0.080 (0.068)	0.005 (0.003)	1.744	5.070 (2.006)
	10	10	0.150 (0.112)	0.140 (0.117)	0.006 (0.004)	0.358	6.130 (2.834)
		25	0.114 (0.091)	0.113 (0.091)	0.006 (0.004)	0.889	5.420 (2.036)
		50	0.085 (0.071)	0.080 (0.068)	0.005 (0.003)	1.774	5.070 (2.006)

Table 57: Results from the numerical applications for $\mathcal{S} = (2, 365, 2.0, 0.5)$

ϵ_{em}	N_M	N_{init}	$\Delta_{\ell}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\pi_1}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\mu_1}^{\mathcal{S},\mathcal{A}}$(std)	mean execution time [s]	mean number iterations
1e-05	2	10	0.000 (0.000)	0.002 (0.002)	0.001 (0.001)	3.758	85.520 (35.636)
		25	0.000 (0.000)	0.002 (0.001)	0.001 (0.001)	9.391	72.940 (34.774)
		50	-0.000 (0.000)	0.001 (0.001)	0.001 (0.001)	18.790	54.840 (32.864)
	5	10	0.000 (0.000)	0.001 (0.002)	0.001 (0.001)	4.573	90.010 (34.107)
		25	0.000 (0.000)	0.001 (0.001)	0.001 (0.001)	11.439	87.210 (34.694)
		50	0.000 (0.000)	0.000 (0.001)	0.000 (0.001)	22.933	61.730 (35.012)
	10	10	0.000 (0.000)	0.001 (0.002)	0.001 (0.001)	4.773	91.610 (36.013)
		25	0.000 (0.000)	0.001 (0.002)	0.001 (0.001)	11.907	86.140 (36.130)
		50	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	23.860	63.480 (39.475)
0.001	2	10	0.006 (0.003)	0.014 (0.006)	0.008 (0.004)	2.055	35.320 (20.036)
		25	0.004 (0.002)	0.010 (0.006)	0.006 (0.004)	5.144	22.040 (11.634)
		50	0.003 (0.002)	0.007 (0.006)	0.004 (0.003)	10.303	18.220 (7.001)
	5	10	0.006 (0.002)	0.014 (0.006)	0.009 (0.003)	2.678	41.710 (23.682)
		25	0.005 (0.002)	0.011 (0.006)	0.007 (0.004)	6.698	27.800 (19.231)
		50	0.003 (0.002)	0.008 (0.005)	0.005 (0.004)	13.457	20.430 (10.224)
	10	10	0.006 (0.002)	0.014 (0.006)	0.009 (0.003)	2.821	44.160 (28.196)
		25	0.005 (0.002)	0.012 (0.006)	0.007 (0.004)	7.026	28.740 (20.888)
		50	0.003 (0.002)	0.008 (0.005)	0.005 (0.004)	14.112	20.870 (11.834)
0.1	2	10	0.135 (0.112)	0.046 (0.036)	0.034 (0.026)	0.681	9.440 (2.879)
		25	0.074 (0.036)	0.027 (0.021)	0.026 (0.016)	1.684	9.140 (2.793)
		50	0.060 (0.030)	0.022 (0.017)	0.024 (0.012)	3.369	8.950 (2.882)
	5	10	0.173 (0.148)	0.055 (0.044)	0.036 (0.029)	0.958	9.350 (4.112)
		25	0.100 (0.068)	0.035 (0.028)	0.027 (0.018)	2.381	8.330 (2.687)
		50	0.069 (0.031)	0.024 (0.017)	0.023 (0.014)	4.758	7.900 (2.528)
	10	10	0.172 (0.144)	0.053 (0.040)	0.037 (0.030)	1.015	9.950 (7.407)
		25	0.101 (0.068)	0.035 (0.028)	0.027 (0.018)	2.509	8.330 (2.687)
		50	0.069 (0.031)	0.024 (0.017)	0.023 (0.014)	5.024	7.900 (2.528)

Table 58: Results from the numerical applications for $\mathcal{S} = (2, 365, 0.1, 1.0)$

ϵ_{em}	N_M	N_{init}	$\Delta_{\ell}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\pi_1}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\mu_1}^{\mathcal{S},\mathcal{A}}$(std)	mean execution time [s]	mean number iterations
1e-05	2	10	0.000 (0.000)	0.003 (0.005)	0.000 (0.000)	3.653	97.060 (33.509)
		25	0.000 (0.000)	0.002 (0.004)	0.000 (0.000)	9.116	92.820 (35.904)
		50	0.000 (0.000)	0.002 (0.004)	0.000 (0.000)	18.254	91.250 (36.466)
	5	10	0.000 (0.000)	0.002 (0.004)	0.000 (0.000)	3.813	81.300 (39.000)
		25	0.000 (0.000)	0.001 (0.003)	0.000 (0.000)	9.558	73.940 (37.620)
		50	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	19.054	68.300 (38.191)
	10	10	0.000 (0.000)	0.002 (0.004)	0.000 (0.000)	3.942	81.270 (39.030)
		25	0.000 (0.000)	0.001 (0.003)	0.000 (0.000)	9.852	74.460 (38.050)
		50	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	19.696	68.750 (38.186)
0.001	2	10	0.008 (0.003)	0.025 (0.032)	0.001 (0.001)	1.557	38.900 (15.401)
		25	0.007 (0.003)	0.021 (0.025)	0.001 (0.001)	3.877	34.830 (16.362)
		50	0.007 (0.003)	0.020 (0.023)	0.001 (0.001)	7.762	34.540 (17.059)
	5	10	0.006 (0.004)	0.028 (0.036)	0.001 (0.001)	1.517	26.710 (10.416)
		25	0.005 (0.004)	0.019 (0.023)	0.001 (0.001)	3.795	23.140 (8.993)
		50	0.005 (0.003)	0.017 (0.021)	0.000 (0.001)	7.547	21.890 (8.964)
	10	10	0.006 (0.004)	0.028 (0.036)	0.001 (0.001)	1.556	26.760 (10.134)
		25	0.005 (0.004)	0.020 (0.025)	0.001 (0.001)	3.897	22.740 (9.048)
		50	0.005 (0.003)	0.017 (0.021)	0.000 (0.001)	7.779	21.720 (8.849)
0.1	2	10	0.315 (0.222)	0.219 (0.166)	0.008 (0.005)	0.310	5.690 (1.875)
		25	0.206 (0.181)	0.148 (0.124)	0.006 (0.004)	0.780	5.320 (1.933)
		50	0.156 (0.142)	0.123 (0.112)	0.005 (0.003)	1.556	5.000 (1.400)
	5	10	0.180 (0.111)	0.151 (0.127)	0.006 (0.004)	0.373	6.870 (2.618)
		25	0.125 (0.093)	0.104 (0.086)	0.005 (0.003)	0.942	5.700 (1.889)
		50	0.105 (0.079)	0.092 (0.083)	0.004 (0.003)	1.870	5.370 (1.858)
	10	10	0.179 (0.111)	0.150 (0.128)	0.005 (0.004)	0.385	6.810 (2.587)
		25	0.125 (0.092)	0.104 (0.086)	0.005 (0.003)	0.969	5.690 (1.880)
		50	0.105 (0.079)	0.091 (0.083)	0.004 (0.003)	1.932	5.320 (1.827)

Table 59: Results from the numerical applications for $\mathcal{S} = (2, 365, 2.0, 1.0)$

ϵ_{em}	N_M	N_{init}	$\Delta_{\ell}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\pi_1}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\mu_1}^{\mathcal{S},\mathcal{A}}$(std)	mean execution time [s]	mean number iterations
1e-05	2	10	-0.000 (0.000)	0.001 (0.002)	0.000 (0.001)	8.451	245.580 (104.859)
		25	-0.000 (0.000)	0.000 (0.002)	0.000 (0.001)	21.136	245.240 (107.465)
		50	-0.000 (0.000)	0.000 (0.002)	0.000 (0.001)	42.307	238.890 (107.777)
	5	10	0.000 (0.000)	0.000 (0.002)	0.000 (0.001)	10.168	254.020 (102.004)
		25	0.000 (0.000)	0.000 (0.002)	0.000 (0.000)	25.422	248.710 (107.574)
		50	-0.000 (0.000)	0.000 (0.001)	0.000 (0.001)	50.820	238.960 (112.500)
	10	10	0.000 (0.000)	0.000 (0.002)	0.000 (0.001)	10.428	253.410 (102.412)
		25	0.000 (0.000)	0.000 (0.001)	0.000 (0.000)	26.025	249.520 (107.068)
		50	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	52.070	239.980 (110.209)
0.001	2	10	0.025 (0.013)	0.011 (0.009)	0.007 (0.006)	3.016	82.210 (19.502)
		25	0.024 (0.013)	0.011 (0.009)	0.007 (0.006)	7.573	77.710 (22.325)
		50	0.022 (0.012)	0.010 (0.008)	0.007 (0.005)	15.153	71.310 (23.366)
	5	10	0.025 (0.011)	0.011 (0.009)	0.008 (0.006)	3.687	84.200 (20.441)
		25	0.024 (0.011)	0.011 (0.009)	0.007 (0.006)	9.255	77.260 (23.076)
		50	0.022 (0.011)	0.010 (0.008)	0.007 (0.006)	18.489	71.310 (23.791)
	10	10	0.025 (0.011)	0.011 (0.009)	0.008 (0.006)	3.786	86.250 (18.864)
		25	0.024 (0.011)	0.011 (0.009)	0.007 (0.006)	9.486	78.050 (22.568)
		50	0.022 (0.011)	0.010 (0.008)	0.007 (0.006)	18.966	71.240 (23.559)
0.1	2	10	0.718 (0.230)	0.095 (0.049)	0.065 (0.035)	0.850	20.520 (7.920)
		25	0.666 (0.257)	0.095 (0.054)	0.057 (0.036)	2.158	18.530 (7.520)
		50	0.591 (0.254)	0.078 (0.053)	0.055 (0.038)	4.310	16.900 (7.151)
	5	10	0.774 (0.242)	0.097 (0.049)	0.069 (0.036)	1.095	21.040 (7.746)
		25	0.679 (0.291)	0.084 (0.051)	0.063 (0.037)	2.794	17.940 (7.243)
		50	0.608 (0.292)	0.071 (0.054)	0.060 (0.039)	5.574	16.480 (6.642)
	10	10	0.775 (0.240)	0.098 (0.049)	0.069 (0.036)	1.127	21.180 (8.192)
		25	0.683 (0.286)	0.085 (0.052)	0.062 (0.037)	2.874	18.180 (7.492)
		50	0.609 (0.291)	0.071 (0.055)	0.060 (0.039)	5.739	16.480 (6.642)

Table 60: Results from the numerical applications for $\mathcal{S} = (2, 730, 0.1, 0.0)$

ϵ_{em}	N_M	N_{init}	$\Delta_{\ell}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\pi_1}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\mu_1}^{\mathcal{S},\mathcal{A}}$(std)	mean execution time [s]	mean number iterations
1e-05	2	10	0.000 (0.000)	0.002 (0.004)	0.000 (0.000)	4.835	127.880 (68.991)
		25	0.000 (0.000)	0.002 (0.003)	0.000 (0.000)	11.906	108.560 (66.486)
		50	-0.000 (0.000)	0.001 (0.002)	0.000 (0.000)	23.929	104.150 (66.008)
	5	10	0.000 (0.000)	0.002 (0.004)	0.000 (0.000)	5.868	133.410 (69.902)
		25	0.000 (0.000)	0.001 (0.002)	0.000 (0.000)	14.416	114.350 (72.544)
		50	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	29.032	105.720 (68.238)
	10	10	0.000 (0.000)	0.002 (0.004)	0.000 (0.000)	6.133	138.260 (70.871)
		25	0.000 (0.000)	0.001 (0.002)	0.000 (0.000)	15.001	114.100 (70.655)
		50	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	30.193	106.280 (67.606)
0.001	2	10	0.011 (0.007)	0.025 (0.023)	0.001 (0.001)	1.845	39.890 (22.264)
		25	0.009 (0.006)	0.020 (0.019)	0.001 (0.001)	4.490	32.760 (20.633)
		50	0.008 (0.006)	0.019 (0.018)	0.001 (0.000)	9.071	30.340 (19.414)
	5	10	0.011 (0.007)	0.028 (0.024)	0.001 (0.001)	2.329	42.230 (25.529)
		25	0.008 (0.006)	0.020 (0.018)	0.001 (0.001)	5.638	30.890 (18.641)
		50	0.008 (0.006)	0.019 (0.018)	0.001 (0.000)	11.412	29.330 (17.595)
	10	10	0.011 (0.007)	0.028 (0.024)	0.001 (0.001)	2.460	43.680 (27.888)
		25	0.008 (0.006)	0.020 (0.018)	0.001 (0.001)	5.918	32.410 (20.644)
		50	0.008 (0.006)	0.019 (0.018)	0.001 (0.000)	11.952	29.970 (18.292)
0.1	2	10	0.239 (0.181)	0.136 (0.097)	0.005 (0.003)	0.444	6.950 (3.689)
		25	0.158 (0.136)	0.082 (0.070)	0.004 (0.003)	1.062	5.390 (2.231)
		50	0.133 (0.131)	0.083 (0.067)	0.004 (0.003)	2.139	4.890 (2.213)
	5	10	0.239 (0.179)	0.140 (0.103)	0.005 (0.003)	0.601	6.940 (4.440)
		25	0.150 (0.124)	0.088 (0.079)	0.005 (0.002)	1.437	5.510 (2.410)
		50	0.127 (0.120)	0.080 (0.066)	0.004 (0.003)	2.906	4.970 (2.317)
	10	10	0.238 (0.179)	0.139 (0.103)	0.005 (0.003)	0.620	7.000 (4.783)
		25	0.150 (0.124)	0.088 (0.079)	0.005 (0.002)	1.481	5.510 (2.410)
		50	0.127 (0.120)	0.080 (0.066)	0.004 (0.003)	2.981	4.970 (2.317)

Table 61: Results from the numerical applications for $\mathcal{S} = (2, 730, 2.0, 0.0)$

ϵ_{em}	N_M	N_{init}	$\Delta_{\ell}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\pi_1}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\mu_1}^{\mathcal{S},\mathcal{A}}$(std)	mean execution time [s]	mean number iterations
1e-05	2	10	0.000 (0.000)	0.001 (0.001)	0.000 (0.001)	13.022	403.480 (173.329)
		25	0.000 (0.000)	0.001 (0.001)	0.000 (0.001)	32.682	388.830 (175.063)
		50	-0.000 (0.000)	0.000 (0.001)	0.000 (0.000)	65.284	362.250 (189.200)
	5	10	0.000 (0.000)	0.000 (0.001)	0.000 (0.001)	15.790	417.780 (187.494)
		25	0.000 (0.000)	0.000 (0.001)	0.000 (0.001)	39.658	398.660 (193.035)
		50	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	79.284	369.570 (205.599)
	10	10	0.000 (0.000)	0.000 (0.001)	0.000 (0.001)	15.904	415.760 (189.212)
		25	0.000 (0.000)	0.000 (0.001)	0.000 (0.001)	39.883	399.570 (194.325)
		50	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	79.697	372.340 (203.967)
0.001	2	10	0.040 (0.019)	0.017 (0.011)	0.008 (0.005)	4.700	130.810 (51.079)
		25	0.037 (0.020)	0.015 (0.011)	0.007 (0.005)	11.875	113.970 (53.816)
		50	0.033 (0.021)	0.013 (0.010)	0.006 (0.004)	23.681	98.750 (48.356)
	5	10	0.041 (0.021)	0.018 (0.012)	0.008 (0.006)	5.580	127.950 (51.578)
		25	0.038 (0.022)	0.015 (0.012)	0.007 (0.005)	14.094	107.170 (48.642)
		50	0.034 (0.022)	0.013 (0.011)	0.006 (0.004)	28.133	94.020 (44.325)
	10	10	0.041 (0.021)	0.018 (0.012)	0.008 (0.006)	5.633	125.590 (50.996)
		25	0.038 (0.022)	0.015 (0.012)	0.007 (0.005)	14.207	108.990 (50.498)
		50	0.034 (0.022)	0.013 (0.011)	0.006 (0.004)	28.355	94.310 (44.536)
0.1	2	10	1.140 (0.555)	0.114 (0.058)	0.057 (0.027)	1.114	17.210 (13.223)
		25	0.929 (0.502)	0.093 (0.061)	0.047 (0.024)	2.819	11.530 (6.391)
		50	0.710 (0.461)	0.075 (0.052)	0.038 (0.023)	5.632	8.490 (5.681)
	5	10	1.108 (0.532)	0.111 (0.057)	0.055 (0.026)	1.454	18.250 (15.753)
		25	0.909 (0.479)	0.092 (0.061)	0.046 (0.024)	3.669	12.560 (7.530)
		50	0.711 (0.458)	0.078 (0.055)	0.038 (0.022)	7.348	8.940 (6.988)
	10	10	1.107 (0.532)	0.111 (0.057)	0.056 (0.027)	1.473	18.120 (16.075)
		25	0.909 (0.479)	0.092 (0.061)	0.046 (0.024)	3.743	12.560 (7.530)
		50	0.711 (0.458)	0.078 (0.055)	0.038 (0.022)	7.472	8.940 (6.988)

Table 62: Results from the numerical applications for $\mathcal{S} = (2, 730, 0.1, 0.5)$

ϵ_{em}	N_M	N_{init}	$\Delta_{\ell}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\pi_1}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\mu_1}^{\mathcal{S},\mathcal{A}}$(std)	mean execution time [s]	mean number iterations
1e-05	2	10	0.000 (0.000)	0.004 (0.005)	0.000 (0.000)	3.770	94.690 (53.990)
		25	0.000 (0.000)	0.004 (0.005)	0.000 (0.000)	9.477	83.780 (58.680)
		50	0.000 (0.000)	0.002 (0.003)	0.000 (0.000)	18.837	68.230 (51.142)
	5	10	0.000 (0.000)	0.004 (0.006)	0.000 (0.000)	4.680	102.590 (60.426)
		25	0.000 (0.000)	0.002 (0.004)	0.000 (0.000)	11.793	88.730 (62.436)
		50	-0.000 (0.000)	0.000 (0.001)	0.000 (0.000)	23.374	74.070 (57.812)
	10	10	0.000 (0.000)	0.004 (0.006)	0.000 (0.000)	4.731	98.930 (58.948)
		25	0.000 (0.000)	0.002 (0.004)	0.000 (0.000)	11.803	89.080 (63.479)
		50	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	23.502	74.910 (58.348)
0.001	2	10	0.008 (0.006)	0.032 (0.023)	0.001 (0.001)	1.455	28.830 (17.777)
		25	0.006 (0.006)	0.024 (0.022)	0.001 (0.000)	3.676	22.140 (13.752)
		50	0.005 (0.005)	0.020 (0.022)	0.001 (0.000)	7.288	19.910 (14.371)
	5	10	0.008 (0.006)	0.033 (0.022)	0.001 (0.001)	1.888	29.920 (17.782)
		25	0.006 (0.005)	0.025 (0.021)	0.001 (0.001)	4.724	23.560 (16.503)
		50	0.005 (0.005)	0.021 (0.022)	0.001 (0.000)	9.389	19.980 (13.255)
	10	10	0.008 (0.006)	0.033 (0.023)	0.001 (0.001)	1.923	30.170 (18.500)
		25	0.006 (0.005)	0.025 (0.021)	0.001 (0.001)	4.758	24.120 (17.582)
		50	0.005 (0.005)	0.020 (0.022)	0.001 (0.000)	9.491	20.090 (13.258)
0.1	2	10	0.162 (0.149)	0.124 (0.095)	0.004 (0.003)	0.329	6.480 (2.563)
		25	0.099 (0.103)	0.092 (0.083)	0.004 (0.003)	0.826	5.700 (2.711)
		50	0.077 (0.088)	0.077 (0.068)	0.004 (0.003)	1.647	5.290 (2.590)
	5	10	0.157 (0.128)	0.120 (0.098)	0.004 (0.003)	0.416	6.640 (2.571)
		25	0.100 (0.098)	0.081 (0.068)	0.004 (0.003)	1.032	5.470 (2.394)
		50	0.075 (0.085)	0.071 (0.063)	0.003 (0.002)	2.059	4.820 (2.118)
	10	10	0.157 (0.129)	0.121 (0.099)	0.004 (0.003)	0.405	6.630 (2.568)
		25	0.100 (0.098)	0.081 (0.068)	0.004 (0.003)	1.007	5.470 (2.394)
		50	0.075 (0.084)	0.071 (0.063)	0.003 (0.002)	2.017	4.850 (2.211)

Table 63: Results from the numerical applications for $\mathcal{S} = (2, 730, 0.1, 1.0)$

ϵ_{em}	N_M	N_{init}	$\Delta_{\ell}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\pi_1}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\mu_1}^{\mathcal{S},\mathcal{A}}$(std)	mean execution time [s]	mean number iterations
1e-05	2	10	0.000 (0.000)	0.002 (0.004)	0.000 (0.000)	3.885	113.950 (40.651)
		25	-0.000 (0.000)	0.002 (0.003)	0.000 (0.000)	9.614	106.790 (43.284)
		50	-0.000 (0.000)	0.001 (0.002)	0.000 (0.000)	19.253	105.030 (44.586)
	5	10	0.000 (0.000)	0.001 (0.004)	0.000 (0.000)	4.503	109.200 (49.968)
		25	0.000 (0.000)	0.001 (0.002)	0.000 (0.000)	11.120	97.930 (49.506)
		50	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	22.217	94.990 (50.650)
	10	10	0.000 (0.000)	0.001 (0.003)	0.000 (0.000)	4.437	109.580 (50.074)
		25	0.000 (0.000)	0.001 (0.002)	0.000 (0.000)	11.087	97.930 (50.381)
		50	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	22.149	95.210 (51.639)
0.001	2	10	0.010 (0.004)	0.021 (0.020)	0.001 (0.001)	1.628	46.160 (14.943)
		25	0.009 (0.004)	0.018 (0.019)	0.001 (0.001)	4.024	42.050 (15.563)
		50	0.008 (0.004)	0.017 (0.018)	0.001 (0.001)	8.048	39.180 (15.868)
	5	10	0.009 (0.005)	0.022 (0.022)	0.001 (0.001)	1.820	37.940 (13.740)
		25	0.008 (0.005)	0.018 (0.018)	0.001 (0.001)	4.466	34.430 (12.892)
		50	0.007 (0.005)	0.017 (0.019)	0.001 (0.001)	8.913	31.330 (12.147)
	10	10	0.009 (0.005)	0.022 (0.022)	0.001 (0.001)	1.797	37.780 (13.699)
		25	0.008 (0.005)	0.018 (0.018)	0.001 (0.001)	4.459	34.410 (12.895)
		50	0.007 (0.005)	0.017 (0.019)	0.001 (0.001)	8.895	31.250 (12.086)
0.1	2	10	0.410 (0.216)	0.179 (0.124)	0.007 (0.004)	0.387	9.770 (4.964)
		25	0.300 (0.196)	0.160 (0.127)	0.006 (0.003)	0.951	7.350 (3.948)
		50	0.247 (0.179)	0.127 (0.110)	0.005 (0.003)	1.914	6.280 (2.761)
	5	10	0.301 (0.145)	0.149 (0.100)	0.006 (0.004)	0.458	8.410 (3.281)
		25	0.224 (0.126)	0.114 (0.096)	0.005 (0.003)	1.118	7.270 (2.973)
		50	0.191 (0.117)	0.105 (0.092)	0.005 (0.003)	2.238	6.610 (2.742)
	10	10	0.300 (0.145)	0.149 (0.100)	0.006 (0.004)	0.449	8.410 (3.274)
		25	0.224 (0.126)	0.113 (0.097)	0.005 (0.003)	1.115	7.300 (2.978)
		50	0.191 (0.117)	0.105 (0.092)	0.005 (0.003)	2.229	6.610 (2.742)

Table 64: Results from the numerical applications for $\mathcal{S} = (5, 180, 0.1, 0.0)$

ϵ_{em}	N_M	N_{init}	$\Delta_{\ell}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\pi_1}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\mu_1}^{\mathcal{S},\mathcal{A}}$(std)	mean execution time [s]	mean number iterations
1e-05	5	10	0.000 (0.000)	0.011 (0.041)	0.000 (0.000)	39.568	188.470 (56.228)
		25	0.000 (0.000)	0.004 (0.009)	0.000 (0.000)	99.257	187.600 (56.425)
		50	0.000 (0.000)	0.004 (0.013)	0.000 (0.000)	197.873	183.070 (53.202)
	10	10	0.000 (0.000)	0.007 (0.035)	0.000 (0.000)	44.358	187.480 (60.264)
		25	0.000 (0.000)	0.003 (0.012)	0.000 (0.000)	111.588	187.490 (60.796)
		50	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	222.153	181.310 (54.118)
0.001	5	10	0.017 (0.006)	0.047 (0.061)	0.002 (0.003)	12.342	57.090 (10.122)
		25	0.017 (0.006)	0.046 (0.067)	0.002 (0.002)	31.067	56.020 (10.879)
		50	0.016 (0.005)	0.049 (0.075)	0.002 (0.002)	61.964	55.490 (11.494)
	10	10	0.017 (0.006)	0.047 (0.074)	0.002 (0.003)	14.355	59.160 (11.511)
		25	0.016 (0.005)	0.049 (0.088)	0.002 (0.003)	36.072	57.850 (11.079)
		50	0.015 (0.005)	0.049 (0.081)	0.002 (0.002)	71.894	57.210 (12.743)
0.1	5	10	0.528 (0.114)	0.249 (0.151)	0.012 (0.012)	2.791	13.200 (2.232)
		25	0.500 (0.110)	0.249 (0.147)	0.011 (0.011)	7.044	13.080 (1.880)
		50	0.481 (0.105)	0.243 (0.159)	0.011 (0.011)	14.044	13.100 (2.062)
	10	10	0.534 (0.117)	0.271 (0.169)	0.012 (0.011)	3.119	13.030 (2.855)
		25	0.508 (0.117)	0.251 (0.152)	0.011 (0.011)	7.879	12.640 (2.256)
		50	0.486 (0.114)	0.235 (0.151)	0.012 (0.010)	15.698	12.790 (2.511)

Table 65: Results from the numerical applications for $\mathcal{S} = (5, 180, 2.0, 0.0)$

ϵ_{em}	N_M	N_{init}	$\Delta_{\ell}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\pi_1}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\mu_1}^{\mathcal{S},\mathcal{A}}$(std)	mean execution time [s]	mean number iterations
1e-05	5	10	0.000 (0.000)	0.002 (0.003)	0.001 (0.002)	84.344	311.460 (67.957)
		25	0.000 (0.000)	0.001 (0.002)	0.001 (0.001)	209.823	298.790 (69.130)
		50	0.000 (0.000)	0.001 (0.002)	0.001 (0.001)	418.906	300.960 (79.332)
	10	10	0.000 (0.000)	0.002 (0.002)	0.001 (0.001)	96.400	328.840 (81.257)
		25	0.000 (0.000)	0.001 (0.002)	0.001 (0.001)	238.726	317.860 (80.612)
		50	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	477.953	316.640 (94.333)
0.001	5	10	0.027 (0.006)	0.020 (0.013)	0.012 (0.010)	29.167	108.040 (19.478)
		25	0.025 (0.006)	0.019 (0.012)	0.013 (0.010)	72.604	110.520 (21.617)
		50	0.024 (0.006)	0.017 (0.011)	0.012 (0.009)	144.905	113.390 (28.199)
	10	10	0.028 (0.007)	0.022 (0.015)	0.012 (0.012)	35.013	122.410 (33.630)
		25	0.026 (0.007)	0.019 (0.013)	0.013 (0.011)	86.751	120.890 (27.292)
		50	0.024 (0.006)	0.018 (0.013)	0.013 (0.010)	173.320	128.960 (38.170)
0.1	5	10	0.949 (0.127)	0.128 (0.063)	0.088 (0.071)	7.981	29.430 (5.636)
		25	0.881 (0.110)	0.119 (0.058)	0.079 (0.062)	19.868	28.040 (4.665)
		50	0.852 (0.106)	0.115 (0.056)	0.077 (0.059)	39.605	28.070 (5.256)
	10	10	1.046 (0.161)	0.131 (0.064)	0.085 (0.065)	9.588	32.550 (7.990)
		25	0.958 (0.121)	0.117 (0.047)	0.078 (0.063)	23.873	31.530 (6.019)
		50	0.919 (0.119)	0.122 (0.052)	0.084 (0.062)	47.549	31.020 (6.005)

Table 66: Results from the numerical applications for $\mathcal{S} = (5, 180, 2.0, 0.2)$

ϵ_{em}	N_M	N_{init}	$\Delta_{\ell}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\pi_1}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\mu_1}^{\mathcal{S},\mathcal{A}}$(std)	mean execution time [s]	mean number iterations
1e-05	5	10	0.000 (0.000)	0.001 (0.002)	0.001 (0.001)	89.800	331.740 (64.068)
		25	0.000 (0.000)	0.001 (0.002)	0.001 (0.001)	224.278	328.310 (70.949)
		50	0.000 (0.000)	0.001 (0.001)	0.000 (0.001)	448.569	327.910 (72.443)
	10	10	0.000 (0.000)	0.001 (0.002)	0.001 (0.001)	101.660	348.840 (77.102)
		25	0.000 (0.000)	0.001 (0.001)	0.000 (0.001)	252.885	340.890 (76.057)
		50	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	505.921	340.970 (84.459)
0.001	5	10	0.031 (0.007)	0.021 (0.014)	0.013 (0.009)	30.361	110.780 (16.464)
		25	0.029 (0.007)	0.019 (0.013)	0.011 (0.009)	75.822	111.870 (19.228)
		50	0.028 (0.006)	0.020 (0.013)	0.012 (0.009)	151.533	113.150 (20.895)
	10	10	0.031 (0.007)	0.022 (0.015)	0.014 (0.009)	35.760	120.420 (18.678)
		25	0.029 (0.007)	0.020 (0.015)	0.012 (0.009)	89.210	121.720 (22.958)
		50	0.028 (0.007)	0.019 (0.014)	0.012 (0.009)	178.424	125.120 (24.574)
0.1	5	10	1.000 (0.152)	0.138 (0.069)	0.090 (0.060)	7.692	28.930 (5.260)
		25	0.935 (0.136)	0.125 (0.058)	0.078 (0.060)	19.264	29.040 (6.290)
		50	0.908 (0.127)	0.123 (0.055)	0.073 (0.052)	38.288	29.200 (5.848)
	10	10	1.079 (0.166)	0.136 (0.068)	0.087 (0.058)	9.070	31.430 (6.228)
		25	1.013 (0.140)	0.126 (0.057)	0.091 (0.064)	22.675	32.140 (7.185)
		50	0.982 (0.131)	0.117 (0.056)	0.087 (0.060)	45.129	32.850 (6.715)

Table 67: Results from the numerical applications for $\mathcal{S} = (5, 180, 2.0, 1.0)$

ϵ_{em}	N_M	N_{init}	$\Delta_{\ell}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\pi_1}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\mu_1}^{\mathcal{S},\mathcal{A}}$(std)	mean execution time [s]	mean number iterations
1e-05	5	10	0.000 (0.000)	0.001 (0.001)	0.000 (0.001)	102.206	369.930 (52.898)
		25	0.000 (0.000)	0.000 (0.001)	0.000 (0.000)	255.936	368.410 (52.833)
		50	0.000 (0.000)	0.000 (0.001)	0.000 (0.000)	511.591	364.620 (55.262)
	10	10	0.000 (0.000)	0.001 (0.001)	0.000 (0.001)	113.800	377.420 (53.454)
		25	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	284.986	374.460 (56.169)
		50	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	569.106	369.630 (57.445)
0.001	5	10	0.035 (0.006)	0.025 (0.014)	0.014 (0.009)	37.259	133.210 (13.753)
		25	0.034 (0.006)	0.023 (0.013)	0.013 (0.008)	93.222	131.550 (13.131)
		50	0.033 (0.006)	0.022 (0.012)	0.012 (0.008)	186.414	131.080 (13.891)
	10	10	0.036 (0.006)	0.025 (0.015)	0.014 (0.009)	42.930	137.550 (13.327)
		25	0.035 (0.006)	0.024 (0.014)	0.013 (0.008)	107.465	137.450 (13.379)
		50	0.033 (0.006)	0.022 (0.012)	0.014 (0.009)	214.845	136.790 (14.220)
0.1	5	10	1.515 (0.228)	0.217 (0.079)	0.132 (0.059)	9.552	35.090 (5.316)
		25	1.454 (0.217)	0.199 (0.072)	0.122 (0.055)	23.979	34.720 (5.344)
		50	1.402 (0.223)	0.183 (0.069)	0.113 (0.054)	47.888	34.990 (5.617)
	10	10	1.513 (0.218)	0.201 (0.080)	0.126 (0.063)	11.268	37.170 (5.463)
		25	1.438 (0.200)	0.185 (0.076)	0.121 (0.058)	28.303	37.810 (5.112)
		50	1.388 (0.198)	0.178 (0.072)	0.108 (0.056)	56.547	37.850 (5.474)

Table 68: Results from the numerical applications for $\mathcal{S} = (5, 365, 0.1, 0.0)$

ϵ_{em}	N_M	N_{init}	$\Delta_{\ell}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\pi_1}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\mu_1}^{\mathcal{S},\mathcal{A}}$(std)	mean execution time [s]	mean number iterations
1e-05	5	10	0.000 (0.000)	0.003 (0.004)	0.000 (0.000)	47.686	208.320 (50.152)
		25	0.000 (0.000)	0.002 (0.005)	0.000 (0.000)	119.526	200.260 (49.553)
		50	0.000 (0.000)	0.002 (0.003)	0.000 (0.000)	239.191	202.080 (53.141)
	10	10	0.000 (0.000)	0.003 (0.005)	0.000 (0.000)	52.356	211.170 (52.678)
		25	0.000 (0.000)	0.001 (0.004)	0.000 (0.000)	130.871	201.580 (49.992)
		50	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	262.125	199.290 (49.609)
0.001	5	10	0.019 (0.005)	0.040 (0.033)	0.002 (0.001)	15.089	63.870 (9.459)
		25	0.018 (0.005)	0.037 (0.034)	0.001 (0.001)	38.033	63.150 (10.301)
		50	0.018 (0.005)	0.040 (0.035)	0.001 (0.001)	76.048	63.310 (11.051)
	10	10	0.019 (0.005)	0.038 (0.032)	0.001 (0.001)	17.085	65.970 (9.328)
		25	0.018 (0.005)	0.038 (0.034)	0.001 (0.001)	43.044	65.820 (11.504)
		50	0.017 (0.005)	0.038 (0.035)	0.002 (0.001)	86.093	66.260 (12.246)
0.1	5	10	0.587 (0.108)	0.204 (0.117)	0.010 (0.008)	3.603	15.500 (2.508)
		25	0.562 (0.097)	0.212 (0.130)	0.008 (0.007)	9.052	15.610 (2.549)
		50	0.541 (0.096)	0.198 (0.119)	0.009 (0.007)	18.112	15.320 (2.315)
	10	10	0.613 (0.109)	0.216 (0.121)	0.011 (0.008)	4.122	16.200 (2.713)
		25	0.584 (0.095)	0.210 (0.115)	0.009 (0.007)	10.361	16.040 (2.775)
		50	0.566 (0.093)	0.198 (0.107)	0.009 (0.007)	20.735	15.470 (2.590)

Table 69: Results from the numerical applications for $\mathcal{S} = (5, 365, 2.0, 0.0)$

ϵ_{em}	N_M	N_{init}	$\Delta_{\ell}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\pi_1}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\mu_1}^{\mathcal{S},\mathcal{A}}$(std)	mean execution time [s]	mean number iterations
1e-05	5	10	0.000 (0.000)	0.001 (0.002)	0.001 (0.001)	92.291	329.340 (90.550)
		25	0.000 (0.000)	0.001 (0.002)	0.001 (0.001)	229.967	318.490 (91.658)
		50	0.000 (0.000)	0.001 (0.001)	0.001 (0.001)	459.291	316.100 (102.850)
	10	10	0.000 (0.000)	0.001 (0.002)	0.001 (0.001)	109.799	359.070 (115.398)
		25	0.000 (0.000)	0.001 (0.002)	0.000 (0.001)	276.147	365.460 (125.392)
		50	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	550.769	365.210 (143.736)
0.001	5	10	0.024 (0.008)	0.014 (0.010)	0.011 (0.008)	37.696	130.520 (45.480)
		25	0.022 (0.008)	0.013 (0.009)	0.011 (0.008)	93.462	131.700 (44.305)
		50	0.021 (0.008)	0.012 (0.009)	0.011 (0.008)	186.898	139.500 (58.222)
	10	10	0.025 (0.009)	0.016 (0.012)	0.012 (0.009)	49.266	153.020 (62.084)
		25	0.023 (0.008)	0.014 (0.011)	0.011 (0.009)	122.717	155.330 (63.061)
		50	0.021 (0.008)	0.013 (0.010)	0.010 (0.009)	245.668	157.270 (63.781)
0.1	5	10	1.112 (0.219)	0.104 (0.057)	0.084 (0.061)	11.836	39.100 (8.664)
		25	1.012 (0.186)	0.097 (0.049)	0.087 (0.070)	29.679	38.380 (8.898)
		50	0.940 (0.149)	0.080 (0.042)	0.075 (0.057)	59.174	38.920 (9.450)
	10	10	1.234 (0.277)	0.113 (0.060)	0.080 (0.066)	16.032	44.330 (12.743)
		25	1.126 (0.197)	0.105 (0.054)	0.078 (0.066)	40.010	42.770 (10.230)
		50	1.047 (0.161)	0.086 (0.040)	0.066 (0.053)	79.900	41.570 (9.218)

Table 70: Results from the numerical applications for $\mathcal{S} = (5, 365, 0.1, 0.2)$

ϵ_{em}	N_M	N_{init}	$\Delta_{\ell}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\pi_1}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\mu_1}^{\mathcal{S},\mathcal{A}}$(std)	mean execution time [s]	mean number iterations
1e-05	5	10	0.000 (0.000)	0.009 (0.052)	0.000 (0.000)	47.124	208.490 (50.303)
		25	0.000 (0.000)	0.004 (0.007)	0.000 (0.000)	117.750	207.380 (51.699)
		50	-0.000 (0.000)	0.003 (0.006)	0.000 (0.000)	235.441	199.830 (53.316)
	10	10	0.000 (0.000)	0.006 (0.032)	0.000 (0.000)	51.338	207.720 (50.930)
		25	0.000 (0.000)	0.002 (0.005)	0.000 (0.000)	129.126	203.550 (50.081)
		50	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	258.133	202.990 (53.553)
0.001	5	10	0.019 (0.005)	0.046 (0.069)	0.002 (0.002)	14.668	63.170 (9.750)
		25	0.019 (0.006)	0.037 (0.040)	0.001 (0.002)	36.726	63.650 (10.903)
		50	0.018 (0.005)	0.040 (0.047)	0.001 (0.002)	73.356	62.430 (10.458)
	10	10	0.019 (0.006)	0.051 (0.098)	0.002 (0.001)	16.653	65.230 (10.084)
		25	0.018 (0.006)	0.044 (0.078)	0.001 (0.002)	41.803	65.990 (11.333)
		50	0.017 (0.005)	0.044 (0.083)	0.001 (0.001)	83.367	65.260 (11.110)
0.1	5	10	0.584 (0.100)	0.221 (0.144)	0.009 (0.007)	3.408	15.260 (2.265)
		25	0.558 (0.094)	0.210 (0.130)	0.008 (0.007)	8.509	14.920 (2.023)
		50	0.542 (0.091)	0.204 (0.143)	0.008 (0.007)	17.024	14.880 (2.036)
	10	10	0.620 (0.110)	0.238 (0.141)	0.009 (0.007)	3.916	15.760 (2.608)
		25	0.587 (0.096)	0.229 (0.144)	0.009 (0.007)	9.777	15.520 (2.755)
		50	0.570 (0.091)	0.228 (0.146)	0.009 (0.007)	19.546	14.910 (2.413)

Table 71: Results from the numerical applications for $\mathcal{S} = (5, 365, 2.0, 0.2)$

ϵ_{em}	N_M	N_{init}	$\Delta_{\ell}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\pi_1}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\mu_1}^{\mathcal{S},\mathcal{A}}$(std)	mean execution time [s]	mean number iterations
1e-05	5	10	0.000 (0.000)	0.002 (0.002)	0.001 (0.001)	96.404	353.320 (89.766)
		25	0.000 (0.000)	0.001 (0.001)	0.001 (0.001)	240.545	345.760 (83.007)
		50	0.000 (0.000)	0.001 (0.001)	0.001 (0.001)	480.448	340.140 (93.486)
	10	10	0.000 (0.000)	0.001 (0.002)	0.001 (0.001)	110.822	381.430 (99.348)
		25	0.000 (0.000)	0.001 (0.002)	0.000 (0.001)	276.333	370.900 (97.205)
		50	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	552.759	356.730 (98.656)
0.001	5	10	0.030 (0.009)	0.019 (0.012)	0.013 (0.009)	35.125	125.730 (21.248)
		25	0.027 (0.009)	0.018 (0.012)	0.012 (0.009)	87.989	127.760 (29.151)
		50	0.026 (0.008)	0.016 (0.010)	0.011 (0.008)	175.981	130.560 (35.448)
	10	10	0.030 (0.009)	0.020 (0.012)	0.013 (0.009)	43.346	143.610 (44.990)
		25	0.028 (0.010)	0.018 (0.013)	0.011 (0.009)	108.622	149.600 (48.528)
		50	0.026 (0.009)	0.016 (0.011)	0.011 (0.008)	217.191	153.090 (51.692)
0.1	5	10	1.098 (0.215)	0.114 (0.059)	0.093 (0.063)	9.785	36.020 (7.340)
		25	0.999 (0.186)	0.098 (0.047)	0.082 (0.055)	24.643	36.810 (7.713)
		50	0.953 (0.179)	0.094 (0.043)	0.075 (0.054)	49.133	36.470 (7.299)
	10	10	1.216 (0.239)	0.118 (0.063)	0.087 (0.047)	12.205	39.750 (8.153)
		25	1.124 (0.198)	0.109 (0.053)	0.083 (0.054)	30.793	38.240 (7.861)
		50	1.074 (0.187)	0.103 (0.055)	0.071 (0.050)	61.518	38.380 (7.320)

Table 72: Results from the numerical applications for $\mathcal{S} = (5, 365, 0.1, 1.0)$

ϵ_{em}	N_M	N_{init}	$\Delta_{\ell}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\pi_1}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\mu_1}^{\mathcal{S},\mathcal{A}}$(std)	mean execution time [s]	mean number iterations
1e-05	5	10	0.000 (0.000)	0.005 (0.011)	0.000 (0.000)	41.348	190.530 (65.262)
		25	0.000 (0.000)	0.005 (0.014)	0.000 (0.000)	103.632	187.080 (62.082)
		50	0.000 (0.000)	0.008 (0.038)	0.000 (0.000)	207.305	186.070 (62.841)
	10	10	0.000 (0.000)	0.004 (0.014)	0.000 (0.000)	45.562	189.760 (63.304)
		25	0.000 (0.000)	0.006 (0.047)	0.000 (0.000)	114.190	186.120 (63.350)
		50	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	228.581	184.170 (62.585)
0.001	5	10	0.018 (0.007)	0.062 (0.104)	0.001 (0.001)	12.920	58.120 (11.480)
		25	0.017 (0.007)	0.056 (0.091)	0.001 (0.001)	32.375	58.080 (10.428)
		50	0.017 (0.007)	0.056 (0.092)	0.001 (0.001)	64.803	58.980 (12.832)
	10	10	0.018 (0.007)	0.062 (0.109)	0.001 (0.001)	14.913	59.370 (11.872)
		25	0.017 (0.007)	0.055 (0.095)	0.001 (0.001)	37.393	59.670 (11.623)
		50	0.016 (0.007)	0.061 (0.107)	0.001 (0.001)	74.751	59.420 (11.220)
0.1	5	10	0.525 (0.105)	0.256 (0.171)	0.008 (0.007)	3.048	14.190 (2.053)
		25	0.502 (0.101)	0.217 (0.129)	0.007 (0.006)	7.648	14.390 (2.262)
		50	0.489 (0.097)	0.211 (0.133)	0.008 (0.006)	15.288	14.350 (2.422)
	10	10	0.543 (0.108)	0.253 (0.153)	0.008 (0.006)	3.435	13.890 (2.457)
		25	0.516 (0.101)	0.228 (0.142)	0.007 (0.005)	8.622	13.790 (2.160)
		50	0.503 (0.101)	0.222 (0.135)	0.007 (0.005)	17.209	13.790 (2.041)

Table 73: Results from the numerical applications for $\mathcal{S} = (5, 365, 2.0, 1.0)$

ϵ_{em}	N_M	N_{init}	$\Delta_{\ell}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\pi_1}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\mu_1}^{\mathcal{S},\mathcal{A}}$(std)	mean execution time [s]	mean number iterations
1e-05	5	10	0.000 (0.000)	0.000 (0.001)	0.000 (0.000)	131.559	484.900 (70.036)
		25	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	329.092	484.470 (68.602)
		50	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	657.847	480.790 (67.195)
	10	10	0.000 (0.000)	0.000 (0.001)	0.000 (0.000)	148.165	492.840 (70.715)
		25	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	370.301	491.310 (73.410)
		50	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	738.746	488.020 (71.765)
0.001	5	10	0.046 (0.007)	0.015 (0.007)	0.011 (0.006)	48.655	176.200 (13.662)
		25	0.045 (0.008)	0.015 (0.008)	0.010 (0.006)	121.743	176.070 (15.380)
		50	0.044 (0.008)	0.014 (0.007)	0.010 (0.006)	243.373	177.540 (21.513)
	10	10	0.047 (0.008)	0.015 (0.008)	0.011 (0.006)	56.391	181.150 (15.816)
		25	0.046 (0.008)	0.015 (0.008)	0.011 (0.006)	141.224	182.050 (19.259)
		50	0.045 (0.008)	0.014 (0.007)	0.010 (0.006)	282.093	181.910 (19.259)
0.1	5	10	2.028 (0.160)	0.166 (0.055)	0.117 (0.043)	14.201	50.700 (5.199)
		25	1.977 (0.178)	0.153 (0.047)	0.108 (0.039)	35.532	50.270 (5.366)
		50	1.936 (0.176)	0.150 (0.050)	0.104 (0.036)	71.068	50.640 (5.759)
	10	10	2.014 (0.156)	0.164 (0.059)	0.112 (0.041)	16.686	52.760 (4.813)
		25	1.956 (0.165)	0.148 (0.050)	0.105 (0.044)	41.748	52.690 (5.126)
		50	1.903 (0.168)	0.146 (0.052)	0.099 (0.041)	83.497	53.630 (4.795)

Table 74: Results from the numerical applications for $\mathcal{S} = (5, 730, 0.1, 0.0)$

ϵ_{em}	N_M	N_{init}	$\Delta_{\ell}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\pi_1}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\mu_1}^{\mathcal{S},\mathcal{A}}$(std)	mean execution time [s]	mean number iterations
1e-05	5	10	0.000 (0.000)	0.002 (0.002)	0.000 (0.000)	57.935	237.280 (51.792)
		25	0.000 (0.000)	0.002 (0.002)	0.000 (0.000)	145.091	232.910 (55.392)
		50	0.000 (0.000)	0.001 (0.002)	0.000 (0.000)	290.051	228.890 (52.451)
	10	10	0.000 (0.000)	0.002 (0.002)	0.000 (0.000)	64.337	239.800 (50.793)
		25	0.000 (0.000)	0.001 (0.002)	0.000 (0.000)	160.090	234.390 (55.298)
		50	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	320.629	230.620 (50.476)
0.001	5	10	0.023 (0.006)	0.027 (0.024)	0.001 (0.001)	18.997	77.850 (11.952)
		25	0.022 (0.006)	0.025 (0.019)	0.001 (0.001)	47.521	75.700 (12.131)
		50	0.021 (0.006)	0.024 (0.017)	0.001 (0.001)	95.010	75.340 (13.361)
	10	10	0.023 (0.006)	0.026 (0.022)	0.001 (0.001)	21.657	81.030 (12.714)
		25	0.021 (0.006)	0.024 (0.018)	0.001 (0.001)	54.144	78.530 (11.835)
		50	0.021 (0.006)	0.026 (0.018)	0.001 (0.001)	108.174	79.740 (13.796)
0.1	5	10	0.733 (0.118)	0.164 (0.097)	0.007 (0.005)	4.524	18.720 (3.073)
		25	0.693 (0.116)	0.161 (0.076)	0.007 (0.005)	11.244	18.510 (2.844)
		50	0.673 (0.106)	0.154 (0.066)	0.007 (0.004)	22.539	18.580 (3.131)
	10	10	0.757 (0.116)	0.169 (0.103)	0.007 (0.005)	5.314	20.550 (3.996)
		25	0.714 (0.110)	0.155 (0.081)	0.007 (0.005)	13.183	19.870 (3.405)
		50	0.694 (0.106)	0.149 (0.075)	0.007 (0.005)	26.424	19.820 (3.648)

Table 75: Results from the numerical applications for $\mathcal{S} = (5, 730, 2.0, 0.0)$

ϵ_{em}	N_M	N_{init}	$\Delta_{\ell}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\pi_1}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\mu_1}^{\mathcal{S},\mathcal{A}}$(std)	mean execution time [s]	mean number iterations
1e-05	5	10	0.000 (0.000)	0.001 (0.001)	0.001 (0.001)	103.260	350.060 (84.877)
		25	0.000 (0.000)	0.001 (0.002)	0.001 (0.001)	258.618	347.790 (115.054)
		50	0.000 (0.000)	0.001 (0.001)	0.001 (0.001)	516.548	340.540 (127.493)
	10	10	0.000 (0.000)	0.001 (0.002)	0.001 (0.001)	128.346	393.270 (103.521)
		25	0.000 (0.000)	0.001 (0.001)	0.000 (0.001)	321.250	400.010 (149.218)
		50	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	641.873	411.060 (188.371)
0.001	5	10	0.025 (0.007)	0.013 (0.010)	0.008 (0.006)	48.624	148.270 (57.472)
		25	0.023 (0.007)	0.012 (0.009)	0.008 (0.006)	121.495	149.320 (59.852)
		50	0.020 (0.006)	0.010 (0.009)	0.007 (0.006)	243.007	158.390 (79.278)
	10	10	0.025 (0.007)	0.012 (0.009)	0.008 (0.006)	66.517	189.630 (90.292)
		25	0.023 (0.007)	0.011 (0.010)	0.008 (0.006)	165.294	190.890 (95.858)
		50	0.020 (0.006)	0.010 (0.010)	0.007 (0.005)	331.532	204.460 (109.883)
0.1	5	10	1.333 (0.226)	0.088 (0.051)	0.083 (0.062)	17.254	48.560 (11.889)
		25	1.217 (0.209)	0.076 (0.042)	0.071 (0.055)	43.158	45.800 (9.793)
		50	1.142 (0.188)	0.071 (0.039)	0.064 (0.050)	86.525	44.580 (9.418)
	10	10	1.411 (0.287)	0.093 (0.057)	0.077 (0.057)	24.717	57.020 (16.387)
		25	1.256 (0.235)	0.073 (0.042)	0.073 (0.054)	61.733	51.850 (12.857)
		50	1.171 (0.203)	0.067 (0.039)	0.067 (0.046)	123.735	49.390 (10.400)

Table 76: Results from the numerical applications for $\mathcal{S} = (5, 730, 0.1, 0.2)$

ϵ_{em}	N_M	N_{init}	$\Delta_{\ell}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\pi_1}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\mu_1}^{\mathcal{S},\mathcal{A}}$(std)	mean execution time [s]	mean number iterations
1e-05	5	10	0.000 (0.000)	0.003 (0.004)	0.000 (0.000)	55.771	231.000 (50.577)
		25	0.000 (0.000)	0.002 (0.004)	0.000 (0.000)	138.917	223.060 (53.277)
		50	0.000 (0.000)	0.002 (0.004)	0.000 (0.000)	278.267	222.620 (54.930)
	10	10	0.000 (0.000)	0.002 (0.004)	0.000 (0.000)	62.296	234.130 (52.784)
		25	0.000 (0.000)	0.002 (0.004)	0.000 (0.000)	154.856	227.160 (54.929)
		50	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	309.911	224.090 (55.043)
0.001	5	10	0.021 (0.006)	0.033 (0.027)	0.001 (0.001)	18.145	73.520 (10.404)
		25	0.020 (0.005)	0.031 (0.026)	0.001 (0.001)	45.163	73.370 (10.366)
		50	0.019 (0.005)	0.029 (0.023)	0.001 (0.001)	90.475	72.500 (11.732)
	10	10	0.021 (0.006)	0.033 (0.026)	0.001 (0.001)	20.892	77.670 (10.794)
		25	0.019 (0.005)	0.029 (0.022)	0.001 (0.001)	51.910	76.900 (10.736)
		50	0.019 (0.005)	0.029 (0.025)	0.001 (0.001)	103.963	75.620 (12.582)
0.1	5	10	0.699 (0.105)	0.185 (0.101)	0.007 (0.005)	4.430	18.460 (2.762)
		25	0.662 (0.090)	0.173 (0.098)	0.006 (0.005)	11.094	18.830 (2.832)
		50	0.646 (0.087)	0.164 (0.085)	0.007 (0.005)	22.171	18.590 (2.916)
	10	10	0.728 (0.093)	0.193 (0.103)	0.007 (0.005)	5.225	19.450 (3.170)
		25	0.694 (0.089)	0.183 (0.100)	0.006 (0.005)	13.080	19.940 (3.655)
		50	0.675 (0.087)	0.173 (0.090)	0.007 (0.005)	26.134	20.110 (3.715)

Table 77: Results from the numerical applications for $\mathcal{S} = (5, 730, 0.1, 1.0)$

ϵ_{em}	N_M	N_{init}	$\Delta_{\ell}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\pi_1}^{\mathcal{S},\mathcal{A}}$(std)	$\Delta_{\mu_1}^{\mathcal{S},\mathcal{A}}$(std)	mean execution time [s]	mean number iterations
1e-05	5	10	0.000 (0.000)	0.002 (0.004)	0.000 (0.000)	47.776	207.270 (46.567)
		25	0.000 (0.000)	0.002 (0.002)	0.000 (0.000)	119.525	205.560 (50.892)
		50	0.000 (0.000)	0.001 (0.002)	0.000 (0.000)	239.376	203.630 (52.811)
	10	10	0.000 (0.000)	0.002 (0.004)	0.000 (0.000)	53.905	209.300 (45.370)
		25	0.000 (0.000)	0.001 (0.002)	0.000 (0.000)	134.971	208.400 (49.737)
		50	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	270.158	204.310 (52.557)
0.001	5	10	0.020 (0.005)	0.032 (0.021)	0.001 (0.001)	15.732	68.290 (11.345)
		25	0.019 (0.006)	0.029 (0.019)	0.001 (0.001)	39.259	66.470 (11.122)
		50	0.018 (0.005)	0.028 (0.018)	0.001 (0.001)	78.698	67.130 (11.732)
	10	10	0.019 (0.005)	0.030 (0.020)	0.001 (0.001)	18.379	70.900 (11.560)
		25	0.018 (0.005)	0.029 (0.019)	0.001 (0.001)	45.889	69.870 (14.030)
		50	0.017 (0.005)	0.026 (0.020)	0.001 (0.001)	91.808	71.450 (15.018)
0.1	5	10	0.625 (0.114)	0.194 (0.092)	0.006 (0.004)	3.896	17.410 (2.502)
		25	0.587 (0.104)	0.179 (0.093)	0.005 (0.003)	9.701	17.290 (2.197)
		50	0.574 (0.100)	0.175 (0.090)	0.005 (0.004)	19.477	17.130 (2.671)
	10	10	0.657 (0.111)	0.197 (0.099)	0.006 (0.004)	4.622	18.180 (2.744)
		25	0.621 (0.103)	0.204 (0.103)	0.005 (0.004)	11.504	18.110 (2.860)
		50	0.605 (0.102)	0.201 (0.106)	0.005 (0.004)	23.063	18.250 (3.468)

Table 78: Results from the numerical applications for $\mathcal{S} = (5, 730, 2.0, 1.0)$

ϵ_{em}	N_M	N_{init}	$\Delta_{\ell}^{\mathcal{S},\mathcal{A}}$ (std)	$\Delta_{\pi_1}^{\mathcal{S},\mathcal{A}}$ (std)	$\Delta_{\mu_1}^{\mathcal{S},\mathcal{A}}$ (std)	mean execution time [s]	mean number iterations
1e-05	5	10	0.001 (0.002)	0.001 (0.002)	0.001 (0.002)	54.797	608.890 (133.559)
		25	0.001 (0.002)	0.001 (0.003)	0.001 (0.002)	137.300	610.450 (134.781)
		50	-0.000 (0.001)	0.001 (0.003)	0.000 (0.001)	274.368	610.790 (134.592)
	10	10	0.001 (0.002)	0.001 (0.003)	0.001 (0.002)	57.521	607.520 (130.220)
		25	0.001 (0.002)	0.001 (0.002)	0.001 (0.002)	143.610	611.950 (135.164)
		50	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	287.074	610.370 (132.982)
0.001	5	10	0.084 (0.029)	0.013 (0.009)	0.010 (0.006)	28.068	313.820 (40.024)
		25	0.078 (0.025)	0.012 (0.009)	0.010 (0.006)	70.282	315.050 (39.699)
		50	0.073 (0.022)	0.011 (0.008)	0.009 (0.005)	140.131	315.920 (42.450)
	10	10	0.084 (0.029)	0.013 (0.009)	0.010 (0.006)	29.482	313.300 (39.476)
		25	0.078 (0.028)	0.012 (0.009)	0.010 (0.006)	73.569	316.220 (39.974)
		50	0.073 (0.024)	0.011 (0.008)	0.009 (0.005)	146.964	317.880 (41.593)
0.1	5	10	2.465 (0.191)	0.117 (0.035)	0.088 (0.028)	6.690	71.410 (5.078)
		25	2.403 (0.201)	0.115 (0.031)	0.084 (0.029)	16.755	71.580 (5.196)
		50	2.359 (0.200)	0.112 (0.035)	0.083 (0.031)	33.443	71.300 (5.677)
	10	10	2.481 (0.188)	0.118 (0.033)	0.086 (0.026)	7.035	71.970 (4.649)
		25	2.422 (0.196)	0.110 (0.034)	0.083 (0.028)	17.557	71.750 (5.233)
		50	2.386 (0.190)	0.108 (0.033)	0.082 (0.030)	35.075	71.290 (5.645)

C

COMPLEMENTS RELATED TO CHAPTER 5

C.1 CONSISTENCY OF THE ESTIMATION OF THE EDGES E OF THE GRAPH OF SUBSTITUTION

We reming the reader that the substitution graph is composed of products as vertices and edges link products that are substitute.

Proof 12 (Some ideas about a possible future proof of Conjecture 2) *The definition of* G *is entirely defined by the probability of substitution meaning that all edges that are not part of* E *will have a* $\pi_{\ell k} + \pi_{k\ell} = 0$ *and every edge that is part of the graph will have* $\pi_{\ell k} + \pi_{k\ell} > 0$*. The definition of* $E_{n\tau_n}$ *implies that every edge* (ℓ, k) *verifies* $\Delta_{nk\ell} + \Delta_{n\ell k} > \tau_n$*. Lemma 4 implies that* $\lim_{n\to+\infty} \Delta_{nk\ell} = \mu_\ell \pi_{\ell k}$ *at a rate of* $O_{\mathbb{P}}(n^{-\frac{1}{2}})$ *and that* $\lim_{n\to+\infty} \tau_n = 0$ *at a rate slower than* $n^{-\frac{1}{2}}$*.*

The main argument for believing that the conjecture is true is that for $(\ell, k) \in E$*,* $\lim_{n\to+\infty} \mathbb{P}(0 < \tau_n < \pi_{\ell k} + \pi_{k\ell}) = 1$ *and then that* $\lim_{n\to+\infty} \mathbb{P}(\Delta_{nk\ell} + \Delta_{n\ell k} > \tau_n) = 1$*. For* $(\ell, k) \in [K^2]/E$*, for* n *sufficiently large since* $\Delta_{nk\ell} + \Delta_{n\ell k}$ *converges faster to* 0 *than* τ_n*, we would obtain* $\lim_{n\to+\infty} \mathbb{P}(\Delta_{nk\ell} + \Delta_{n\ell k} < \tau_n) = 1$*. These two results imply that we retrieve the signal of substitution for the products that are in the same groups and rule out the edges corresponding to non substituable products.*

C.2 THE CONSTRAINED HIERARCHICAL AGGLOMERATIVE CLUSTERING

Algorithm 2 Proposed constrained HAC Algorithm

Input: $[K]$, $D(.,.)$ ▷ D is the distance matrix.

Output: Constrained HAC tree T

$C \leftarrow \emptyset$
for $k \in [K]$ **do**
  $C \leftarrow C \cup \{k\}$
Compute $D_{k,l} = D(k,l), \forall (k,l) \in [K] \times [K]$
$T \leftarrow C$
while $|C| > 1$ **do**
  $cm_1, cm_2 \leftarrow \text{argmin}_{(c_1,c_2)\in C\times C, \textsc{Is-Identifiable}(c_1\cup c_2)} \textsc{Dist-SingleLink}(c_1, c_2)$
  $C \leftarrow (C/\{cm_1\})/\{cm_2\}$
  $C \leftarrow C \cup \{cm_1 \cup cm_2\}$
  $T \leftarrow T \cup \{cm_1 \cup cm_2\}$
return T

function $\textsc{Dist-SingleLink}(\{k_h\}_{h=1}^{H}, \{l_m\}_{m=1}^{M}) = \min_{h,m} D(k_h, l_m)$

1: **function** Is-Identifiable(c) ▷ Defined by the Proposition 7 given ω associated to c.

C.3 A BIC HEURISTIC PROPOSAL

C.3.1 *The proposed BIC formulation*

We have hesitated between two BIC formulas: the first whose asymptotic property is dependent on n itself and the second which is dependent on the number of sales such as in Choiruddin, Coeurjolly, and Waagepetersen, 2021. Respectively these possible BIC criteria are

$$\widehat{\gamma} = \text{argmax}_{k=1,\dots,K} \log L(\widehat{\theta}_{\gamma^{(k)}}; \gamma^{(k)}, y_n | x_n) - \nu_{\gamma^{(k)}} \frac{\ln n}{2}, \tag{66}$$

and

$$\widehat{\gamma} = \text{argmax}_{k=1,\dots,K} \log L(\widehat{\theta}_{\gamma^{(k)}}; \gamma^{(k)}, y_n | x_n) - \nu_{\gamma^{(k)}} \frac{\ln(\sum_{j=1}^{J} \sum_{k=1}^{K} Y_{njk})}{2},$$

where $\nu_\gamma = K + \sum_{k=1}^{K} \sum_{\ell \neq k} \gamma_k^\top \gamma_\ell$ is the number of parameters involved by model γ, $\widehat{\theta}_\gamma$ is the maximum likelihood estimate (see Section 4.4.1 for details on its estimation). Here the log-likelihood $\log L$ refers to the likelihood introduced in Section 4.3 where we added a new parameter γ that expresses the sparse modelling. We chose only Criterion 66 for the numerical applications that follow.

C.3.2 *Numerical experiments related to the proposed BIC*

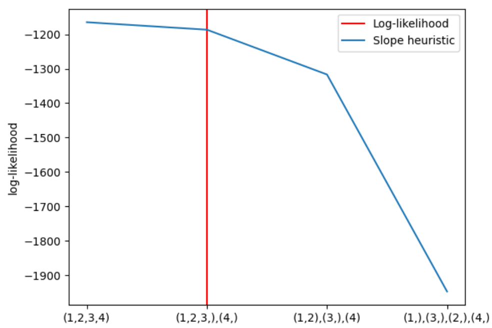
We remind the reader that the constrained HAC produces a partition of at most 5 products hence some groups in the dendrogram are not aggregated with other groups.

Dataset 1

In Figure 27 are the results of the model selection based on a the retained BIC heuristic. The right groups are retrieved with the Ward linkage.

Figure 27: Model selection based on the retained BIC heuristic for the putting knife-tape dataset

Interpretation: Product 1 (16cm), Product 2 (14cm), Product 3 (12cm), Product 4 (tapes)



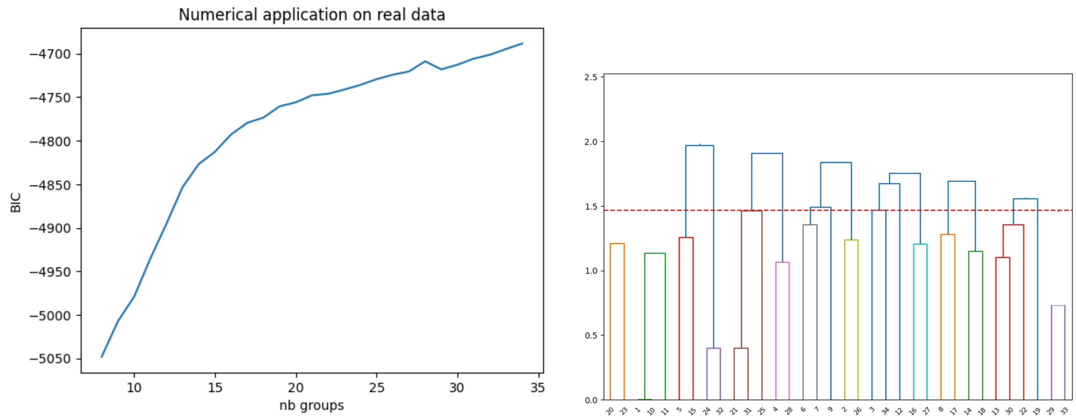
(a) HAC with single linkage

Dataset 2

The model selection applied to the dataset yields the Figure 28. The BIC is best for the partition where all the products are separated meaning when there is no substitution.

However the slope of the *data-observed log-likelihood* changes for a partition where there is 16 groups. According to that true partition, the ARI in terms of LIB_TYP and LIB_STYP is 0.02 and 0.10. These values are close to 0 meaning that the clustering solutions are not better than a random partition. The three groups with maximum substitution for the partition where retrieved. The analysis of Table 32, 33 and 34 is not promising. The wrenchs could substitute with a screw driver or a pipe wrench from an uncompatible size. Substitutions estimators are not reliable in this case; it could be due to the lack of information because we did not aggregate all stores. However some values are of interest such as the high probability of substitution between the open end wrench of size 18x19mm or 30x32mm to the set of wrenches in Table 32.

Figure 28: 35 Products



(a) BIC curve

(b) Dendrogram of the constrained HAC

Dataset 4

Table 29 shows that the best partition is composed of 8 groups. The dendrogram shows that some groups are aggregated early. Since we expect that products for different categories at the TYP and STYP are not substitutable we used it to evaluate the quality of our partition at 0.02 which is close to 0 and reveals that the quality is close to a random clustering. The first 3 groups that have the maximum mean substitution rate (i.e: $K^{-1} \sum_{\ell=1}^{K}(1-\hat{\pi}_{\ell\ell})$) have been studied.

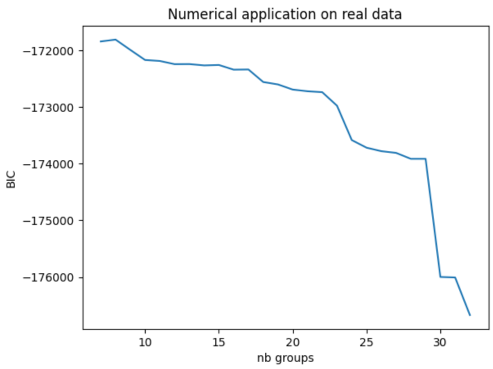
The first group shown in Table 80 is composed of the products (1,2,3,4) which retrieves some interesting signals such as the two corded router (products 1 and 4) can have the same usage. The probability of substitution to the first corded router and to the jigsaw and the hammer drill is at 0 denoting the non substitutability. The same observation can be made for the second corded router. The other associations are however incoherent such as the jigsaw (product 2) substituting to the corded router with a value of 0.66.

The second group shown in Table 81 detects substitution within products that do not have the same usage.

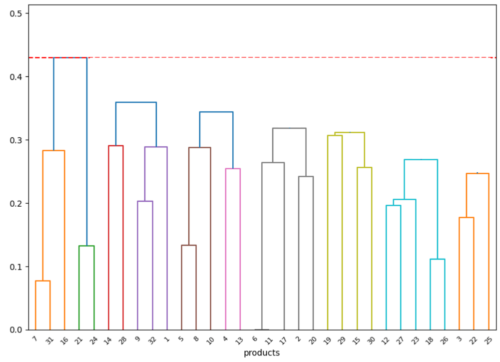
The third group shown in Table 82 is composed of 2 jigsaws that are substitutable and three other products that are not, such as an electric screw driver, a corded planer or a drill. The probabilities of substitution make apparent the substitution from the

1rst jigsaw to the second (respectively products 8 and 12). The electric screw driver does not substitute with other products.

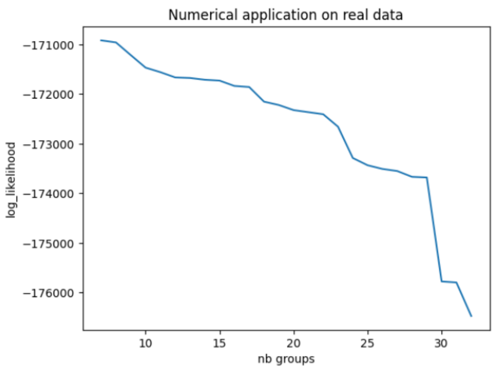
Figure 29: Descriptive statistics



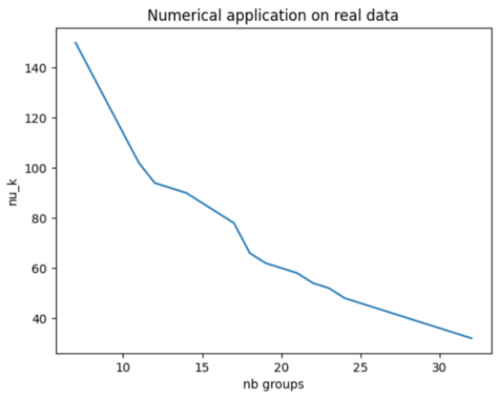
(a) BIC



(b) Dendrogram



(c) Data-observed log-likelihood



(d) Degrees of freedom

Table 79: Product id and information

1	Défonceuse électrique DEXTER POWER Dp4, 1300.0
2	Scie sauteuse 20V WORX , WX543.9 (Sans batterie ni chargeur)
3	Marteau perforateur 3en1 Brushless 20V WORX NITRO, WX380 (2 batteries 4Ah)
4	Défonceuse filaire DEXTER Dp5, 1300 W
5	Scie sauteuse filaire MAKITA Jv0600j 650 W
6	Marteau perforateur burineur sds plus METABO Uhe 2660-2 quick, 800 W
7	Rabot électrique filaire RYOBI Epn 7582 nhg, 750 W
8	Scie sauteuse filaire BOSCH Pst 650 500 W
9	Tournevis sans fil sans fil BOSCH Psr select 3.6 V 1.5 Ah
10	Rabot électrique filaire RYOBI Epn 7582 nhg, 750 W
11	Perforateur sds plus MAKITA Hr2300x9, 720 W
12	Scie sauteuse pendulaire filaire METABO Steb 65 quick 450 W

Table 80: 1st group of substitution

μ_k	1	2	3	4
0.07	0.71	0.03	0.00	0.26
0.02	0.66	0.00	0.17	0.17
0.01	0.86	0.14	0.00	0.00
0.04	0.74	0.01	0.00	0.25

Table 81: 2nd group of substitution

μ_k	5	6	7
0.07	0.00	0.00	1.00
0.06	0.20	0.00	0.80
0.17	0.01	0.04	0.95

Table 82: 3rd group of substitution

μ_k	8	9	10	11	12
0.05	0.00	0.25	0.00	0.10	0.64
0.36	0.00	1.00	0.00	0.00	0.00
0.15	0.09	0.65	0.00	0.03	0.23
0.07	0.05	0.58	0.37	0.00	0.00
0.12	0.06	0.76	0.18	0.00	0.00

BIBLIOGRAPHY

Agrawal, Narendra and Stephen A Smith (1996). "Estimating Negative Binomial Demand for Retail Inventory Management with Unobservable Lost Sales." 43, pp. 839–861.

Anupindi, Ravi, Maqbool Dada, and Sachin Gupta (1998). "Estimation of consumer demand with stock-out based substitution: An application to vending machine products." *Marketing Science* 17.4, pp. 406–423. ISSN: 07322399. DOI: 10.1287/mksc.17.4.406.

Arlot, Sylvain (2019). "Minimal penalties and the slope heuristics: a survey." 160.3. arXiv: 1901.07277. URL: http://arxiv.org/abs/1901.07277.

AZZALINI, A. (1985). "A Class of Distributions which Includes the Normal Ones." *Scandinavian Journal of Statistics* 12.2, pp. 171–178.

Babai, M Z, A Tsadiras, and C Papadopoulos (2020). "On the empirical performance of some new neural network methods for," pp. 1–25. DOI: 10.1093/imaman/dpaa003.

Balakrishnan, Sivaraman, Martin J. Wainwright, and Bin Yu (2017). "Statistical guarantees for the EM algorithm: From population to sample-based analysis." *Annals of Statistics* 45.1, pp. 77–120. ISSN: 00905364. DOI: 10.1214/16-AOS1435. arXiv: 1408.2156.

Baudry, Jean-patrick, Cathy Maugis, Bertrand Michel, Jean-patrick Baudry, Cathy Maugis, Bertrand Michel, and Slope Heuristics (2010). "Slope Heuristics : Overview and Implementation To cite this version : HAL Id : hal-00461639 Slope Heuristics : Overview and Implementation."

Benjamini, Yoav and Yosef Hochberg (1995). "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society: Series B (Methodological)* 57.1, pp. 289–300. ISSN: 0035-9246. DOI: 10.1111/j.2517-6161.1995.tb02031.x.

Biau, G and B Patra (2000). "Sequential Quantile Prediction of Time Series," pp. 1–27.

Birgé, Lucien and Pascal Massart (2007). "Minimal penalties for Gaussian model selection." *Probability Theory and Related Fields* 138.1-2, pp. 33–73. ISSN: 01788051. DOI: 10.1007/s00440-006-0011-8.

Brown, R. G. (1959). "Statistical forecasting for inventory control."

Campo, Katia, Els Gijsbrechts, and Patricia Nisol (2004). "Dynamics in consumer response to product unavailability: Do stock-out reactions signal response to permanent assortment reductions?" *Journal of Business Research* 57.8, pp. 834–843. ISSN: 01482963. DOI: 10.1016/S0148-2963(02)00486-1.

Caplice, P and E Ponce (2020). "MITx MicroMasters Program in SCM Key Concepts."

Caro, Felipe and Jérémie Gallien (2007). "Dynamic assortment with demand learning for seasonal consumer goods." *Management Science* 53.2, pp. 276–292. ISSN: 00251909. DOI: 10.1287/mnsc.1060.0613.

Choi, Tsan-Ming (2012). *Handbook of Newsvendor problems: Models, extensions and applications*. Vol. 176. Springer.

Choiruddin, Achmad, Jean François Coeurjolly, and Rasmus Waagepetersen (2021). "Information criteria for inhomogeneous spatial point processes." *Australian and New Zealand Journal of Statistics* 63.1, pp. 119–143. ISSN: 1467842X. DOI: 10.1111/anzs.12327. arXiv: 2003.03880.

Croston, J. D. (1972). "Forecasting and Stock Control for Intermittent Demands." *Operational Research Quarterly (1970-1977)* 23.3, pp. 289–303. ISSN: 00303623. URL: http://www.jstor.org/stable/3007885.

Dempster, A.P., N.M. Laird, and D.B Rubin (1977). "MAXIMUM LIKELIHOOD ESTIMATION FROM INCOMPLETE DATA VIA EM-TYPE ALGORITHMS."

Dempster, Arthur P, Nan M Laird, and Donald B Rubin (1977). "Maximum likelihood from incomplete data via the EM algorithm." *Journal of the royal statistical society: series B (methodological)* 39.1, pp. 1–22.

Fisher, Marshall L. and Ramnath Vaidyanathan (2009). "An Algorithm and Demand Estimation Procedure for Retail Assortment Optimization." *Management* 2007, pp. 1–36. ISSN: 15265501.

Gaur, Vishal and Dorothée Honhon (2006). "Assortment planning and inventory decisions under a locational choice model." *Management Science* 52.10, pp. 1528–1543. ISSN: 00251909. DOI: 10.1287/mnsc.1060.0580.

Gelders, L.F. and P.M. Van Looy (1978). "An inventory policy for slow and fast movers in a petrochemical plant: a case study." *J Opl Res Soc*, 29, pp. 867–874.

Gneiting, Tilmann, Fadoua Balabdaoui, and Adrian Raftery (2007). "Probabilistic forecasts, calibration and sharpness." *Jrssb* 69.2, pp. 243–268.

Gneiting, Tilmann and Matthias Katzfuss (2014). "Probabilistic Forecasting." *Annual Review of Statistics and Its Application* 1.Volume 1, 2014, pp. 125–151. ISSN: 2326-831X. URL: https://www.annualreviews.org/content/journals/10.1146/annurev-statistics-062713-085831.

Gneiting, Tilmann and Roopesh Ranjan (2013). "Combining predictive distributions." *Electronic Journal of Statistics* 7.1, pp. 1747–1782. ISSN: 19357524. DOI: 10.1214/13-EJS823. arXiv: 1106.1638.

Gruen, Thomas W. (2002). "Retail Out-of-Stocks: A worldwide Examination of Extent, Causes and Consumer Responses."

Holt, Charles C. (1957). "Forecasting seasonals and trends by exponentially weighted moving averages." *International Journal of Forecasting* 20.1, pp. 5–10. ISSN: 01692070. DOI: 10.1016/j.ijforecast.2003.09.015.

Honhon, Dorothée, Vishal Gaur, Sridhar Seshadri, and Dorothée Honhon (2010). "Assortment Planning and Inventory Decisions Under Assortment Planning and Inventory Decisions Under Stockout-Based Substitution." September 2015. DOI: 10.1287/opre.1090.0805.

Hubert, Lawrence and Phipps Arabie (1985). "Comparing partitions." *Journal of Classification* 2.1, pp. 193–218. ISSN: 01764268. DOI: 10.1007/BF01908075.

Hubner, Alexandre (2017). "A decision support system for retail assortment planning Article information :" *International Journal of Retail & Distribution Management*.

Hyndman, Rob J and George Athanasopoulos (2018). "Forecasting : Principles and Practice."

Hyndman, Rob J. and Anne B. Koehler (2006). "Another look at measures of forecast accuracy." *International Journal of Forecasting* 22.4, pp. 679–688. ISSN: 01692070. DOI: 10.1016/j.ijforecast.2006.03.001.

Kenneth J. Arrow, Theodore Harris and Jacob Marschak (1951). "Optimal Inventory Policy." *Econometrica* 19.3, pp. 250–272.

Kok, A. Gurhan and Marshall L. Fisher (2007). "Demand estimation and assortment optimization under substitution: Methodology and application." *Operations Research* 55.6, pp. 1001–1021. ISSN: 0030364X. DOI: 10.1287/opre.1070.0409.

Kok, A.G. (2003). "Management of product variety in retail operations."

Kolassa, Stephan (2020). "Why the "best" point forecast depends on the error or accuracy measure." *International Journal of Forecasting* 36.1, pp. 208–211. ISSN: 01692070. DOI: 10.1016/j.ijforecast.2019.02.017. URL: https://doi.org/10.1016/j.ijforecast.2019.02.017.

Kwan, H.W. (1991). "On the demand distributions of slow moving items." *Unpublished Ph.D. thesis, Lancaster University, UK*.

Lukinskiy, Valery, Vladislav Lukinskiy, and Boris Sokolov (2020). "Control of inventory dynamics: A survey of special cases for products with low demand." *Annual Reviews in Control* 49.xxxx, pp. 306–320. ISSN: 13675788. DOI: 10.1016/j.arcontrol.2020.04.005.

Mahajan, S. and G. Van Ryzin (2001). "Stocking retail assortments under dynamic consumer substitution." *Operations Research* 49.3, pp. 334–351. ISSN: 0030364X. DOI: 10.1287/opre.49.3.334.11210.

McFadden, Daniel (1974). *Conditional logit analysis of qualitative choice behavior*.

McLachlan, Geoffrey J and Thriyambakam Krishnan (2007). *The EM algorithm and extensions*. John Wiley & Sons.

Miller, Christopher M, Stephen A Smith, Shelby H Mcintyre, and Dale D Achabal (2010). "Optimizing and Evaluating Retail Assortments for Infrequently Purchased Products." *Journal of Retailing* 86.2, pp. 159–171. ISSN: 0022-4359. DOI: 10.1016/j.jretai.2010.02.004. URL: http://dx.doi.org/10.1016/j.jretai.2010.02.004.

Musalem, Andrés, Marcelo Olivares, Eric T. Bradlow, Christian Terwiesch, and Daniel Corsten (2010). "Structural estimation of the effect of out-of-stocks." *Management Science* 56.7, pp. 1180–1197. ISSN: 00251909. DOI: 10.1287/mnsc.1100.1170.

Netessine, Serguei and Nils Rudi (2003). "Centralized and Competitive Inventory Models with Demand Substitution." *SSRN Electronic Journal*. DOI: 10.2139/ssrn.303779.

Pinçe, Çerağ, Laura Turrini, and Joern Meissner (2021). "Intermittent demand forecasting for spare parts: A Critical review." *Omega (United Kingdom)* 105. ISSN: 03050483. DOI: 10.1016/j.omega.2021.102513.

Salameh, Moueen K., Ali A. Yassine, Bacel Maddah, and Likaa Ghaddar (2014). "Joint replenishment model with substitution." *Applied Mathematical Modelling* 38.14, pp. 3662–3671. ISSN: 0307904X. DOI: 10.1016/j.apm.2013.12.008. URL: http://dx.doi.org/10.1016/j.apm.2013.12.008.

Schoot, E. H.M. van der, R. M.J. Heuts, and LWG Strijbosch (2000). "A combined forecast—inventory control procedure for spare parts." *Journal of the Operational Research Society* 51.10, pp. 1184–1192. ISSN: 14769360. DOI: 10.1057/palgrave.jors.2601013.

Schwarz, Gideon (1978). "1978 Shwarz." *The Annals of Statistics* 6.2, pp. 461–464. ISSN: 0090-5364.

Smith, Stephen A, Narendra Agrawal, Stephen A Smith, and Narendra Agrawal (2000). "Management of Multi-Item Retail Inventory Systems with Demand Substitution MANAGEMENT OF MULTI-ITEM RETAIL INVENTORY SYSTEMS WITH DEMAND SUBSTITUTION." March 2015.

Syntetos, A A (2001). "Forecasting of intermittent demand." *Dissertation*, pp. 1–495.

Syntetos, A. A., J. E. Boylan, and J. D. Croston (2005). "On the categorization of demand patterns." *Journal of the Operational Research Society* 56.5, pp. 495–503. ISSN: 01605682. DOI: 10.1057/palgrave.jors.2601841.

Syntetos, Aris A, M Zied Babai, and Everette S Gardner Jr (2015). "Forecasting intermittent inventory demands : simple parametric methods vs . bootstrapping."

Journal of Business Research. ISSN: 0148-2963. DOI: 10.1016/j.jbusres.2015.03.034. URL: http://dx.doi.org/10.1016/j.jbusres.2015.03.034.

Talluri, Kalyan and Garrett Van Ryzin (2004). "Revenue Management Under a General Discrete Choice Model of Consumer Behavior." *Management Science* 50.1, pp. 15–33. ISSN: 00251909. DOI: 10.1287/mnsc.1030.0147.

Teunter, R. H. and L. Duncan (2009). "Forecasting intermittent demand: A comparative study." *Journal of the Operational Research Society* 60.3, pp. 321–329. ISSN: 14769360. DOI: 10.1057/palgrave.jors.2602569.

Train, Kenneth E. (2009). *Discrete choice methods with simulation, second edition*. Vol. 9780521766555. January 2009, pp. 1–388. ISBN: 9780511805271. DOI: 10.1017/CBO9780511805271.

Van Ryzin, Garrett and Siddharth Mahajan (1999). "On the relationship between inventory costs and variety benefits in retail assortments." *Management Science* 45.11, pp. 1496–1509. ISSN: 00251909. DOI: 10.1287/mnsc.45.11.1496.

Vandeput, N. (2020). *Inventory Optimization: Models and Simulations*. Business & economics. De Gruyter. ISBN: 9783110673913. URL: https://books.google.fr/books?id=i6VJyAEACAAJ.

Wan, Mingchao, Yihui Huang, Lei Zhao, Tianhu Deng, and Jan C. Fransoo (2018). "Demand estimation under multi-store multi-product substitution in high density traditional retail." *European Journal of Operational Research* 266.1, pp. 99–111. ISSN: 03772217. DOI: 10.1016/j.ejor.2017.09.014.

Winters, Peter R. (1960). "Forecasting Sales by Exponentially Weighted Moving Averages." *Management Science* 6.3, pp. 324–342. ISSN: 0025-1909. DOI: 10.1287/mnsc.6.3.324.

Zhu (2021). *Spare Parts Demand Forecasting and Inventory Management*. ISBN: 9789058926135.

Zotteri, Giulio (2000). "The impact of distributions of uncertain lumpy demand on inventories." *Production Planning and Control* 11.1, pp. 32–43. ISSN: 09537287. DOI: 10.1080/095372800232469.

COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". `classicthesis` is available for both LaTeX and LyX:

https://bitbucket.org/amiede/classicthesis/

Happy users of `classicthesis` usually send a real postcard to the author, a collection of postcards received so far is featured here:

http://postcards.miede.de/