PRODUCTION DECISIONS WITH CONVEX COSTS AND CARBON EMISSION CONSTRAINTS

A DISSERTATION SUBMITTED TO

THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE

OF BILKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

INDUSTRIAL ENGINEERING

By

Ramez Kian

March, 2016

PRODUCTION DECISIONS WITH CONVEX COSTS AND CARBON EMISSION CONSTRAINTS
By Ramez Kian
March, 2016

We certify that we have read this dissertation and that in our opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Ülkü Gürler(Advisor)

E. Emre Berk(Co-Advisor)

Osman Oğuz

Özlem Karsu

Haldun Süral

A. Özgür Toy

Approved for the Graduate School of Engineering and Science:

Levent Onural
Director of the Graduate School

ABSTRACT

PRODUCTION DECISIONS WITH CONVEX COSTS AND CARBON EMISSION CONSTRAINTS

Ramez Kian
Ph.D.in Industrial Engineering
Advisor: Ülkü Gürler
Co-Advisor: E. Emre Berk
March, 2016

In this thesis, different variants of the production planning problem are considered.

We first study an uncapacitated deterministic lot sizing model with a nonlinear convex production cost function. The nonlinearity and convexity of the cost function may arise due to the extra fines paid by a manufacturer for environmental regulations or it may originate from some production functions. In particular, we have considered the Cobb-Douglas production function which is applied in sectors such as energy, agriculture and cement industry. We demonstrate that this problem can be reformulated as a lot sizing problem with nonlinear production cost which is convex under certain assumptions. To solve the problem we have developed a polynomial time dynamic programming based algorithm and nine fast heuristics which rest on some well known lot sizing rules such as Silver-Meal, Least Unit Cost and Economic Order Quantity. We compare the performances of the heuristics with extensive numerical tests.

Next, motivated from the first problem, we consider a lot sizing problem with convex nonlinear production and holding costs for decaying items. The problem is investigated from mathematical programming perspective and different formulations are provided. We propose a structural procedure to reformulate the problem in the form of second order cone programming and employ some optimality and valid cuts to strengthen the model. We conduct an extensive computational test to see the effect of cuts in different formulations.

We also study the performance of our heuristics on a rolling horizon setting. We conduct an extensive numerical study to compare the performance of heuristics and to see the effect of forecast horizon length on their dominance order and to see when they outperform exact solution approaches.

Finally, we study the lot sizing problem with carbon emission constraints. We propose two Lagrangian heuristics when the emission constraint is cumulative over

periods. We extend the model with possibility of lost sales and examine several carbon emission cap policies for a cost minimizing manufacturer and conduct a cost-emission Pareto analysis for each policy.

Keywords: Production planning, Convex cost function, Optimization, Lot sizing.

ÖZET

KONVEKS ÜRETİM MALİYETLERİ VE KARBON EMİSYON KISITLARI ALTINDA ÜRETİM PLANLAMASI

Ramez Kian
Endüstri Mühendisliği, Doktora
Tez Danışmanı: Ülkü Gürler
İkinci Tez Danışmanı: E. Emre Berk
Mart, 2016

Bu tezde üretim planlama probleminin farklı versiyonları ele alınmıştır.

İlk olarak kapasitesiz deterministik kafile büyüklüğü problemi doğrusal olmayan konveks üretim maliyet fonksiyonu ile çalışılmıştır. Maliyet fonksiyonunun doğrusal olmaması ve konveks bir yapıya sahip olması, çevresel düzenlemeler nedeniyle üretici tarafından üstlenilmesi gereken ekstra maliyetler veya bazı üretim fonksiyonlarından kaynaklanabilir. özellikle, enerji, tarım ve çimento sektörlerinde kullanılan Cobb-Douglas üretim fonksiyonu incelenmiştir. Bu problemin kesin varsayımlar altında doğrusal olmayan konveks üretim fonksiyonlu kafile büyüklüğü problemi olarak formülize edilebileceği gösterilmiştir. Problemi çözmek için dinamik programlama temelinde polinom zamanlı bir algoritma ve Silver-Meal, En Az Birim Maliyet ve Ekonomik Parti Büyüklüğü gibi klasik kafile büyüklüğü sezgisellerine dayalı dokuz hızlı sezgisel geliştirilmiştir. Bu sezgisellerin performansları ayrıntılı sayısal testler ile kıyaslanmıştır.

Sonrasında, ilk problemden yola çıkarak, bozulabilir ürünler için konveks doğrusal olmayan üretim ve elde tutma maliyetlerine sahip kafile büyüklüğü problemi ele alınmıştır. Bu problem matematiksel programlama perspektifinden incelenmiştir. Problemi ikinci derece konik programlama formunda yeniden formüle etmek için yapısal bir yöntem önerilmiş ve bazı optimalite ve geçerlilik kesmeleri modeli güçlendirmek için uygulanmıştır. Bu kesmelerin etkilerini farklı formulasyonlarda görmek için kapsamlı bir nümerik test yapılmıştır.

Ayrıca sezgisellerin performansı bir döner ufuklu yapıda çalışılmıştır. Sezgiselleri kıyaslamak, tahminleme dönem uzunluğunun etkisini sezgisellerin performans sıralaması üzerinde görmek ve ne zaman kesin çözüm yöntemini domine edeceklerini belirlemek için kapsamlı bir sayısal çalışma yapılmıştır.

Son olarak, kafile büyüklüğü problemi karbom emisyon kısıtları ile çalışılmıştır.

Emisyon kısıtının periyotlar üzerinde kümülatif olduğu durum için iki Lagrange sezgiseli önerilmiştir. Model, kayıp talep durumu için genişletilmiş, çeşitli karbon emisyon kısıtları altında maliyet minimizasyonunu hedefleyen bir üretici için test edilmiş ve maliyet-emisyon Pareto analizi yapılmıştır.

Anahtar sözcükler: Üretim planlama, Konveks maliyetli fonksiyon, Optimizasyon, Kafile büyüklügü problemi .

Acknowledgement

I would like to express my deepest gratitude to my advisors Prof. Ülkü Gürler and Assoc. Prof. Emre Berk for giving me the opportunity to pursue my PhD research under their supervision. Without their invaluable and continuous support, experience and patience my doctoral study would not be finished. They always devoted their valuable time to enlighten my path in my research with long productive meetings which sometimes continued after midnights. I would like to thank both of them for their understanding and friendly behavior toward me.

I cannot express enough thanks to my committee: Assoc. Prop. Osman Oğuz, Asst. Prof. Özlem Karsu, Prof. Haldun Süral and Assoc. Prof. Özgür Toy for their valuable time devoted to reading each part of my dissertation and for providing me with precious suggestions.

It was an honor to be a member of Bilkent University in Department of Industrial Engineering, and I would like to thank each faculty member and the Chair of the department for providing good working conditions.

I would like to acknowledge that this research was supported by Program 2215 of TÜBİTAK, the scientific and technological research council of Turkey.

I warmly thank all my colleagues and friends at Bilkent University. Namely, I would like to thank my best friend and my office mate Kamyar Kargar for printing this dissertation while I was not in Turkey and for being beside me in the most difficult moments of my life during the research.

Also, I thank my friends Dr. Yahya Saleh, Ece Zeliha Demirci, Burak Paç, Dr. Hatiçe Çalik, Dr. Esra Koca, Nihal Bektaş, Gizem Özbaygin, Merve Merakli, Dr. Okan Arslan, Dr. Vedat Bayram, Dr. Bariş Yıldız, Nazlı Sönmez, Özum Korkmaz, İrfan Mahmutoğulları, Okan Dukkanci, Sinan Bayraktar, Özge Şafak. I am very lucky to have these great friends without whom life would not be cheerful.

I would like to express my thanks to my Iranian friends in Turkey for their support and their friendships. Namely, I would like to thank my friends Behnam Ghassemiparvin, Rahim Bahari, Akbar Alipour, Soheil Taraghinia, Samad Nadimi, Mohammad Tofighi, Mehdi Dabirnia, Sina Rezaei, Mehdi Kazempour, Manouchehr Takrimi, Sepideh Yekani and Parisa Sharif.

I also wish to show my sincere gratitude to Prof. Togla Bektaş for his understanding and for the flexible working environment he has provided in Southampton Business School. Also, my special thanks go to my colleague Dr. Şener Akpınar in Southampton Business School for helping me in Turkish translation of this dissertation's abstract.

I would like to appreciate all the people who tolerated me during my research; in addition, I am indebted more to ones who did not tolerate me but taught me a great lesson in the life which will undoubtedly be more useful than any of the findings in this research in my future life.

Last but not least, I am so grateful to my beloved family for their encouragement and endless support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Production decision making or production planning deals with the planning of the resources which are used in a manufacturing or procurement system in such a way that minimize their utilization cost.

In the competitive industrial environment, companies equipped with better production planning policies gains key advantages in terms of customer satisfaction and cost reduction. A class of production planning problems in which a known demand should be met over a finite or medium term discrete time horizon is specifically known as lot sizing. The problem is to determine the amount that should be produced at each period and it may be subject to capacity limits. Besides the physical limitations, the environmental concerns about global warming also lead to new regulations and possible limitations on carbon footprint of the manufacturing firms. Among those, the most important is the Kyoto protocol which is initially adopted in 1997 in Kyoto, Japan. Under this protocol, countries are classified based on their contribution in global CO_2 emission amounts and their development level. Then, they are assigned a certain amount of emission allowance to specified time period and are obligated to reduce and control their emissions under their assigned emission quota.

As a consequence, manufacturers may be forced to limit their activities to respect the carbon emission limits and/or they are charged for undesirable wastes whose negative ecological impact must be mitigated. As legal penalty rates become progressive, the costs associated with production activities may exhibit a convex behavior.

This study aims at investigating a single item multi-period finite horizon production planning (lot sizing) problem with a special production cost function. The problem is to find a production plan on the smallest possible cost including setup, holding and production costs. In the entire study the production cost is assumed to be a nonlinear convex function which lacks economies of scale. That is, the manufacturer incurs more costs as he produces in larger batch quantities. The assumed non-linearity aims to capture the externalities in production activities that are encountered in a number of industrial settings as briefly discussed below.

(i) Productive assets require maintenance and repair activities over their lifetimes and almost all production processes generate undesirable wastes, which must be disposed of and/or whose negative ecological impact must be mitigated. The costs associated with such auxiliary activities need not be linear and/or concave. On the contrary, As additional resources are required or legal penalty rates become progressive, the costs associated with such auxiliary activities exhibit a convex behavior. Given that public awareness of and concern for production's impact on the environment is increasing worldwide, the importance of such pollution control or ecological impact mitigation efforts and the associated costs is increasing, as well.

(ii) Nonlinear production functions also arise from production activities that use a number of substitutable resources such as materials, labor, machinery, capital, energy, *etc.*. There is a vast literature in microeconomics on economic production functions that relate output to usage of such resources (See eg.[2]). One of the most common production functions is the Cobb-Douglas production function which is introduced in detail in Section 2.3 of the next chapter.

(iii) Another commonly used economic production function is the Leontieff function which is introduced in [3]. This production function also results in nonlinear production cost function. Its structure and applications are also discussed briefly in the next chapter.

(iv) Nonlinearity of cost may also arise in production planning models with price dependent supply. For instance, sellers of refurbished products need to acquire used items from consumers. The amount that suppliers are willing to provide depends on the price that the producer offers and therefore, the total amount that the producer pays is a nonlinear function of the number of items procured [4].

Considering the production resource quantity together with the output quantity reminds a multi-stage model of lot sizing problem, but it is assumed that the resources are ample and as it is shown in Chapter 2 that the model is handled in a single-level uncapacitated form. The main contribution of this work is the convexity and nonlinearity of the production cost. This leads to new optimality structure which differs from the classical models.

The remainder of this dissertation is structured as follows: in Chapter 2 we provide a literature review in three directions including lot sizing, the Cobb-Douglas production function and carbon emission aspect.

In Chapter 3 an uncapacitated lot sizing problem with a nonlinear convex production function is investigated. For motivation purpose, we also show that how a production planning problem with the Cobb-Douglass production function can be reduced to such lot sizing problems. We provide some optimality properties and propose forward and backward dynamic programming based solution methods. Also, we design several fast heuristic algorithms and compare their performance from different aspects.

In Chapter 4 we consider the uncapacitated lot sizing model for deteriorating products. We focus on mathematical programming formulations which result in different Mixed Integer Nonlinear Programming (MINLP) models. We compare the performance of some well known commercial optimization packages to see how they perform against this problem. Then we reformulate the problem in the form of Second Order Conic Programming (SOCP). We use some structural results to provide several optimality and valid cuts to be used in each of the formulations. Then we compare their performances in the CPLEX optimization package.

In Chapter 5 we investigate the performance of the lot sizing heuristics, which are proposed in Chapter 2, in a rolling horizon setting where the demand data revealed gradually as time progresses.

In Chapter 6 we return to our main model with the Cobb-Douglas production function and apply carbon emission constraints. Two exact and heuristic approaches are discussed for this problem. Then the model is extended to address lost sales and examined under different carbon emission cap policies.

At last we summarize all findings, contributions and future research opportunities of this dissertation in Chapter 7.

Chapter 2

Literature Review

The literature is surveyed in three different parts: lot sizing problem, application of the production function, especially the Coob-Douglas production function, and carbon emission studies in Industrial Engineering (IE) and Operation Management (OM) problems.

2.1 Lot sizing

A brief definition of lot sizing problem is "deciding on the optimal *timing* and *level* of production" where the objective is to minimize total cost over a finite time horizon including setup, inventory holding and production costs. Basically, there is a trade off between holding and setup costs. They play the main role on finding the optimal production policy if the production cost is constant over the horizon. The problem can be classified based on the model extensions and the approach of finding optimal solution. For instance considering the problem for a firm which uses the same facilities for different products leads to a multi-item model while the sequence of operations and facility usage can be modeled as a scheduling problem. If the final product is an output of several manufacturing operations and we are concerned with the output of each stage, then the problem is called multi-stage or multi-level. Another classification is to divide the problem to *capacitated* and *uncapacitated* groups. In the uncapacitated models, it is assumed that manufacturer has enough resources and physical facilities to cover any amount of demands, whereas in the capacitated case the maximum production quantity per period is limited due to machines or resources. Table 2.1 from the literature summarizes some extensions of the basic lot sizing model. The asterisked items are addressed directly or indirectly in this dissertation. The related papers have been reviewed extensively in Brahimi et al. [1].

General classification	Extensions	Other extensions
Number of items	Backlogging	* Rolling horizons
* Number of production stages (levels)	Multiple facility	* Speculative motives
* Capacity constraints	* Perishable inventory	
Length of the production periods	Remanufacturing	* Lost sale
	Demand time windows	* Bounded inventory

Table 2.1: Lot sizing model extensions [1]

Wagner and Whitin [5] present a forward dynamic programming algorithm to solve uncapacitated lot sizing problem. Under certain assumptions, it is shown that in an optimal production plan, the inventory level and production amount can not be positive simultaneously and therefore, any positive production amount is equal to accumulative demand of some future periods. Using these optimality conditions, Federgruen and Tzur [6] propose a forward algorithm with complexity order of $T \log(T)$. Silver and Meal [7] present a heuristic algorithm based on minimal average cost per period. Florian and Klein [8] characterize the extreme points of the solution set in the capacitated production planning. They show that an optimal production sequence consists of independent production subplans separated with zero-inventory points –referred to as regeneration points–, and each subplan has a special property called *capacity constrained*; then they have proposed a dynamic programming algorithm for both cases with or without back ordering. The capacitated lot sizing Problem (CLSP) is shown to be NP-hard in Florian et al. [9] and also by Bitran and Yanasse [10] which persuades employment of heuristics in practice. The literature is rich for lot sizing problem both from solution methods and also problem characteristics such as planning horizon, demand type, backlogging allowance.

Various review papers also address the lot sizing problems. Karimi et al. [11] review single level lot sizing problems and classify the solution methods into exact and heuristic. They also branch heuristics to specialized or common-sense based and mathematical programming based ones. Brahimi et al. [1] review *single item* lot sizing trying to build a survey totally dedicated to the single item problems and focus on uncapacitated cases by referring to 110 papers. They identify polynomially solvable cases and discuss four different mathematical programming formulations: Aggregate, No-Inventory, Shortest Path, and Facility Location Formulation. Various solution methods are mentioned in their work.

Jans and Degraeve [12] survey meta-heuristics developed specifically for lot sizing problem and also review some other solution methods such as dedicated heuristics, Lagrangian relaxation, cutting planes, Dantzig-Wolfe decomposition and dynamic programming by going over about 160 papers.

Quadt and Kuhn [13] inspect capacitated lot sizing and scheduling papers and classify them to big bucket and small bucket models. Big bucket models refer to long but with small number of periods while models with short periods called small buckets. They focus on the big bucket type capacitated lot sizing problems and their extensions. They classify the related papers based on assumptions in four factors: (1) back orders, (2) setup carry-over (3) sequencing and (4) parallel machines. Then the solution algorithm for the related works of each type is discussed in their review study.

In the other review work of Jans and Degraeve [14] they concentrate on modeling and operational aspects and refer to more than 240 papers in their comprehensive survey study.

Buschkuhl et al.[15] review four decades of research on capacitated lot sizing and discuss both modeling and solution approaches. They separate lot sizing with sequencing (scheduling) problem. They cite about 140 papers and discuss about the different emerging trends in this research area over time such as mathematical programming and metaheuristic solution approaches.

Ullah et al. [16] also provide an extensive review on inventory lot sizing for both problem configuration and solution aspect and in a very recent work, Bushev et al. [17] provide a review for the review papers of the inventory lot sizing paper.

2.1.1 Lot sizing with deteriorating items

Despite commonplace occurrence of deteriorating products, very little attention has been given to decaying items in discrete time lot sizing. Friedman et al. [18] appears to be the first work where a dynamic lot size model is developed with inventory deterioration.

Hsu [19] considers a dynamic lot size model with age dependent carrying cost with general concave holding and production costs and propose a DP algorithm with complexity order of $O(T^4)$ for T-period problem. The model is generalized to allow for backordering in [20]. Chu et al. [21] propose a polynomial time approximation algorithm for the same problem with a more general cost structure taking into account the economies of scale. Recently, Waterer [22] has provided a generalization with linear costs allowing for inventory gains as well as loss.

For excellent reviews, we refer the reader to Nahmias[23], Raafat [24] and Goyal et al.[25]. Among the existing literature there is no study on lot sizing of perishable inventories in the presence of non-linear production and holding costs.

2.1.2 Lot sizing in rolling horizon environment

A well-known research area in the lot sizing algorithms is the rolling horizon basis where the demand data are available for a limited horizon length and they are gradually revealed as time proceeds. In an early work of Blackburn and Millen [26], the impact of rolling schedule on the performance of three pioneer heuristics has been examined. Their main finding is that Silver-Meal (SM) heuristic can provide cost performance superior to Wagner-whitin (WW) algorithm. Wemmerlöv and Whybark [27] conduct a simulation experiment to evaluate fourteen different lot sizing rules considering the influencing factors such as lead time, forecast error, time between orders and demand variability. They conclude that the ranking of the algorithms is very different under demand uncertainty and then they identify six best of them. Bookbinder et al. [28] introduce a lot sizing rules for probabilistic demands in a rolling horizon which minimizes expected setup and holding costs. Venkataraman and Smith [29] study the production planning problem from the hierarchical standpoint and consider a disaggregation form of aggregate plan in a rolling horizon master production schedule. They examine the impact of forecast windows on the performance of a rolling schedule under the minimum batch-size limitation while Clark et al. [30] deal with a scheduling-lot sizing model with sequence dependent setups in a rolling horizon framework. Stadtler [31] propose a modification method to the lot sizing heuristics in order to look beyond the planning horizon and demonstrate that the modified version of heuristics is fairly insensitive to the planning horizon length. Simpson [32] tests several heuristic lot sizing rule by extensive numerical simulation and argues their performance dominance and weakness. Van Den Heuvel et al. [33] compare four lot sizing methods including Stadler's, SM, WW and its extended variant emphasizing on ending inventory effect in rolling horizon framework. They conclude that when reasonable estimates of future data are available it is better to use these estimates. Chand et al. [34] provide a classified review of the literature in the forecast, solution and rolling horizon problems and vindicate the importance of rolling horizon studies. Similarly, Sahin et al. [35] review the rolling horizon planning literature and also highlight its new research directions on supply chain systems. In a recent work, Toy and Berk [36] examine the performance of the modified counterpart of the classic lot sizing heuristics on a special kind of lot sizing problem called warm/cold process both in static and rolling basis. They identify operating environment characteristics where each particular heuristic is the best or among the best. In another recent work of Baciarello et al. [37] performance of eight well-known heuristics for the classical uncapacitated lot sizing problem is compared with extensive numerical simulations.

The papers discussed above include linear or concave cost lot sizing problems. Some of the few papers with nonlinear convex costs including the one related to this study –which contributed to the literature– are reviewed in the following section.

2.1.3 Lot sizing with convex cost

Despite a large number of different configurations of lot sizing problem in literature, other than discount models (which are nonlinear but piecewise linear functions), there are a few studies with nonlinear production cost function (For example see [38], [39] and [4]). The structure of the optimal schedule of the lot sizing problem with nonlinear convex costs is not known. Hence, the existing works provide either heuristics or focus on improved numerical solution methodologies. The classical dynamic lot sizing problem has been formulated as either a dynamic programming (DP) problem or a mixed-integer mathematical programming problem. When costs involve nonlinear components, both approaches encounter difficulties. The former suffers from the curse of dimensionality and the latter resulting in a mixed integer nonlinear programming (MINLP) problem, which may not be optimally solved using the available generic nonlinear optimizers. Karush [40] studies a production planning model in the presence of general convex costs and absence of setup cost. Then he proposes an incremental solution algorithm. Heck et al. [38] consider a similar setting and propose an improvement heuristic to solve the problem.

Kian et al. [39] consider nonlinear convex production costs in the form of power functions and obtain the optimal policy structure for the case of negligible fixed setup costs and propose several heuristics for the general case based on some structural results.

2.2 Second order cone programming

After the seminal works by Nemirovski [41] and Alizadeh [42], the conic quadratic mathematical programming (also called the second order conic programming, SOCP) models started getting the attention of modelers as they provide an alternative solution methodology to some nonlinear mathematical programming problems. The SOCP models are more general types of formulations which include the linear programming (LP) and quadratically constrained models (QP) as special cases. For more details about the solution approaches and theoretical background, we refer the reader to [43], [44], [41], [45] and [46]. Application of SOCP reformulation to an MINLP may result in efficient improvements in the performance of commercial solvers. However, its application to the classical problems is rare in the literature. Some related studies are as follow. Koca et al. [47] consider a stochastic lot sizing model with controllable process time in which the process can be accelerated with an extra nonlinear convex cost to meet a certain degree of service level in terms of demand satisfaction. They have also employed the conic quadratic reformulation and strengthening method similar to the work of Akturk et al. [48] who use conic formulation in a machine-job assignment problem with a convex machine process cost. In another closely related work Atamturk et al. [49] consider joint location-inventory problems and show how to formulate them in conic form.

2.3 Production function

In this research we investigate the lot sizing problem with nonlinear production functions. One of the motivations behind such settings is the Cobb-Douglas production function which was developed by Paul Douglas and his student, Cobb, in 1927 and it is widely used to represent the relation of output and inputs in economic models [2]. The Cobb-Douglas production function assumes that multiple (m) resources are needed for output, Q and they may be substituted to exploit the marginal cost advantages. In general, it has the form $Q = A\prod_{i=1}^{m} x(i)^{\alpha(i)}$ where A is the technology level for the production process, $x(i)$ denotes the amount of resource i used and $\alpha(i) > 0$ is the resource elasticity. It was first introduced at a macroeconomic level for the US manufacturing industries for the period 1899-1922 but has been widely applied to individual production processes at the microeconomic level, as well. For example, Shadbegian and Gray [50] use the Cobb-Douglas production function to model production processes in the paper, steel and oil industries, Hatirli et al. [51] analyze the relation between inputs and output levels in Turkish agriculture using the Cobb-Douglas production function. Gupta [52] estimates the production factors contribution to output growth in Indian cement industry using the Cobb-Douglas function. Kummel et al. [53] look into economic nature of energy and check their interpretation by econometric analysis of West Germany and USA in previous decades. Solderholm [54] provides a quantitative analysis of innovation and diffusion of wind power in Europe using the Cobb-Douglas function in their model, while Wei [55] applies the Cobb-Douglas production function to analyze the impact of energy efficiency gains on output and energy use, and Khanna [56] examines the cost of meeting the Kyoto Protocol commitments by applying this production function. Finally, Banaeian et al.[57] determine the efficient allocation of energy resources for strawberry production which are modeled by the Cobb-Douglas production function.

Another type of production function with a nonlinear structure which has applications in industry is the Leontieff production function. Its main difference from the Cobb-Douglas function is that it assumes resources are not substitutable but complementary. The applications include Haldi at al. [58] for refining of petroleum and primary metals, Ozaki [59] for large-scale assembly production, Lau et al. [60] for ethylene production, and Nakamura [61] for iron and steel production. The Leontieff production function has the form $Q = \min_i \{{x_i}^{\alpha_i}\}$ for a given set of resources where $x(i)$ denotes the amount of resource i used and $\alpha_i > 0$ is the resource elasticity. Assuming that resource i has unit cost of p_i, the total cost for output Q is given by $\sum_{i=1}^{m} w_i Q^{1/\alpha_i}$ where $w_i = p_i$.Typically, it is assumed that $\alpha_i \leq 1$ so that the variable cost of production is convex in output.

We refer the reader to [2] for more details and other kinds of production functions. To the best of our knowledge there are no studies on the dynamic lot-sizing problem in the presence of Cobb-Douglas or Leontieff production functions. Therefore, one of our contribution is to fill this gap in the literature.

2.4 Sustainability in lot sizing problems

We review the some of the papers in literature related to production and inventory management with environmental considerations. Environmental issues have been considered from different perspectives in the operations management literature which hes led to a stream of research papers in recent years. Hammami et al. [62] provide a nice classification of the recent research.

Heck et al. [38] consider an uncapacitated single item lot sizing problem with a cost function based on ecological considerations. In their model, the production cost consists of three components; *power usage*, *carbon dioxide emission* and *water consumption* costs which are assumed to be nonlinear. They develop three WW based and three Part Period Balancing based heuristics to their model.

In a recent work, Benjaafar et al.[63] highlight an emerging research area and try to attract attention of researchers on the carbon emission concerns by integrating it with operational decision making. They review and classify the related literature into economic, measurement and technical segments and conclude that the literature is very sparse in operational management's area of studies. Then they represent three models of single item lot sizing: under strict carbon caps, under Cap-and-Trade assumptions and a multiple firms configuration. Based on their model, they show the impact of the tighter caps on total cost, effect of collaboration and carbon offset on emission and cost, effect of carbon price on Cap-and-Trade configuration, etc.

Hua et al. [64] employ the idea in the classical EOQ model and discuss the impacts of carbon emission trading on optimal order quantity. They compare the optimal order size under classical EOQ, cap-and-trade, and minimal emission assumptions and provide some conditions to sort them. Also they investigate sensitivity of the order size over the parameters. Helmrich et al.[65] investigate a lot-sizing problem with a cumulative carbon emission constraint. They prove that the problem is NP-hard and propose a Lagrangian heuristic solution algorithm. Anutariya et al. [66] study a manufacturing system with rework under carbon emission allowance. They consider a profit maximizer approach and perform sensitivity analysis. Absi et al. [67] present single item lot sizing model with three different carbon constraints. They provide some complexity result in their paper. Yuyang et al. [68] look into a carbon emission constrained lot sizing model with different production modes each with different emission and cost. They classify the polynomial solvable case. Helmrich [69] in his PhD dissertation titled "Green Lot sizing" studies the lot sizing models including remanufacturing, and with minimum batch size. Yu et al.[68] consider the uncapacitated lot sizing model with multi mode production under cumulative carbon emission where emission is a linear function of production quantity.

Chapter 3

Uncapacitated lot sizing with nonlinear convex production cost

3.1 Introduction

In this chapter, we consider the problem of dynamic lot sizing with nonlinear convex production cost function. The so-called classical dynamic lot sizing problem was first analyzed by [5]. They established that, in an optimal plan with positive fixed setup costs and linear production and holding costs, production is done in a period only if its net demand (actual demand less inventories) is positive, and a period's demand is satisfied entirely by production in a single period. This property is called Zero Inventory Order (ZIO). Most of the existing works on the dynamic lot-sizing problem deal with linear and/or concave production functions rather than convex functions. We demonstrate that when the total resource elasticity parameter, r_t, is smaller than 1 depending on whether there is diminishing returns to resources, we confront (after reformulation) a lot sizing problem with convex production costs function. For convex cost functions and zero setup costs, a parametric algorithm was developed by Veinott [70] for the problem, which can be solved by an incremental approach satisfying each unit of demand as cheaply as possible. In the remainder of this chapter problem definition, structural results, backward and forward DP algorithms and additional fast solution heuristics, their comparison and sensitivity analysis are presented, respectively.

3.2 Model assumptions and formulation

We consider a single item. The length of the problem horizon, T is finite and known. The demand amount in period t is denoted by d_t $(t = 1, \ldots, T)$. All demands are non-negative and known, but may be different over the planning horizon. No shortages are allowed; that is, the amount demanded in a period has to be produced in or before its period. The amount of production in period t is denoted by q_t and is uncapacitated. Production in any period t incurs a fixed cost (of setup) $K_t(\geq 0)$ and a variable cost component. Any period in which $q_t > 0$ is called a *production* period; otherwise, it is a *no-production* period. The inventory on hand at the end of period t is denoted by I_t; each unit of ending inventory in the period is charged with a unit holding cost of h_t. Without loss of generality, the initial inventory level, I_0, is assumed to be zero. The objective is to find a production plan that determines the timing and amount of production (q_t) such that total cost of production and holding over the horizon is minimized. For the sub-horizon consisting of periods $\{u, u+1, \ldots, v\}$, ($[u, v]$ in short), let $P_{u,v}$ denote the production planning problem, $D_{u,v} = d_u + d_{u+1} + \ldots + d_v$ denote the total demand, $Q_{u,v} = (q_u, \ldots, q_v)$ denote the production plan and $F_{u,v}$ denote the corresponding total cost.

We formulate the problem as the following MINLP problem, (P0):

$$\text{(P0)} \qquad \min = \sum_{t=1}^{T} \left[K_t y_t + h_t I_t + w_t q_t^{r_t} \right] \tag{3.1a}$$

s.t.

$$I_0 = 0, \tag{3.1b}$$

$$I_t = I_{t-1} + q_t - d_t, \qquad t = 1, \ldots, T \tag{3.1c}$$

$$q_t \leq M y_t, \qquad t = 1, \ldots, T \tag{3.1d}$$

$$I_t \geq 0, q_t \geq 0, y_t \in \{0, 1\}, \qquad t = 1, \ldots, T \tag{3.1e}$$

Equation (3.1a) shows the objective function including setup, holding, and production costs; (3.1b)and (3.1c) correspond to inventory balance equation while (3.1d) is set for setup cost detection in which M is a sufficiently large number as $M = \sum_{i=1}^{T} d_t$.

The last constraint sets, (3.1e), determine types of the variables and ensure that there should not be any unsatisfied demand at the end of periods.

3.2.1 Application: lot sizing with the Cobb-Douglas production function

The assumed nonlinearity in the production cost in (P0) aims to capture the externalities in production activities that are encountered in a number of industrial settings. In this section, as a special case, we demonstrate how a lot sizing problem with the Cobb-Douglas production function can be reformulated to a lot sizing with nonlinear production cost as (P0). It is assumed that there are m required inputs to produce the product and the production quantity at each period t is equal to the Cobb-Douglas production function of the inputs. That is, $q_t = A_t \prod_{i=1}^{m} x_i^{\alpha_{it}}$ where x_{it} denotes the amount of resource i used in period t for production with the output elasticity of α_{it}. We let $r_t = 1/\sum_{i=1}^{m} \alpha_{it}$ and we keep all other notations the same as the previous model.

$$
\begin{aligned}
&\text{(P1)} \qquad \min \sum_{t=1}^{T} \Big(K_t y_t + h_t I_t + \sum_{i=1}^{m} c_{it} x_{it} \Big) && && \text{(3.2a)}\\
&\text{s.t.} && &&\\
&I_0 = 0, && && \text{(3.2b)}\\
&I_t = I_{t-1} + q_t - d_t, && t = 1, ..., T && \text{(3.2c)}\\
&q_t = A_t \prod_{i=1}^{m} x_{it}^{\alpha_{it}}, && t = 1, ..., T && \text{(3.2d)}\\
&q_t \le M y_t, && t = 1, ..., T && \text{(3.2e)}\\
&I_t, q_t \ge 0, y_t \in \{0,1\}, && t = 1, ..., T && \text{(3.2f)}\\
&x_{it} \ge 0, && t = 1, ..., T; i = 1, ..., m && \text{(3.2g)}
\end{aligned}
$$

Equation (3.2a) shows the objective function including setup, holding, resource usage costs; (3.2c) and (3.2b) correspond to inventory balance equation while the constraint sets (3.2d) is for input-output production relation; and (3.2e) is set for setup cost detection in which M is a sufficiently large number as $M = \sum_{i=1}^{T} d_t$. The last two constraints sets, (3.2f) and (3.2g), determine types of the variables and ensure that there should not be any unsatisfied demand at the end of periods. All the constraints of the model (P1) except (3.2d), which belongs to the production function, have the similar structure of standard uncapacitated lot sizing model. This constraints set makes the problem highly nonlinear because the q_t variables are nonlinearly dependent on x_{it}, $i = 1, \ldots, m$. Note that for any certain x_{it}, the variable q_t is specified uniquely and we can easily replace the q_t variables with their x_{it} equivalence and eliminate all of the q_t variables form the model (P1). However, it makes the model more nonlinear and in contrast, we tend to write x_{it} variables in terms of q_t but a certain value of q_t can be represented by different values of x_{it}; However, for a given q_t there exist a unique assignment of x_{it} which minimizes the resource usage cost. Thus, we use this idea to define uniquely the Γ function as

$$
\Gamma(x_{it}) = \arg \min_{\substack{A_t \prod_{i=1}^{m} x_{it}^{\alpha_{it}} = q_t \\ x_{it} \ge 0}} \left\{ \sum_{i=1}^{m} c_{it} x_{it} \right\}
$$

to write x_{it} in terms of q_t. So we need to solve the subproblem (SP1) below,

$$\text{(SP1)} \qquad \min \sum_{i=1}^{m} c_{it} x_{it} \tag{3.3a}$$

s.t.

$$q_t = A_t \prod_{i=1}^{m} x_{it}^{\alpha_{it}}, \qquad t = 1, ..., T \tag{3.3b}$$

$$x_{it} \geq 0, \qquad t = 1, ..., T; i = 1, ..., m. \tag{3.3c}$$

If $q_t = 0$ then obviously $x_{it}^* = 0$ and otherwise, to solve (SP1) for a given positive q_t, we make Lagrangian function and check the first order necessary conditions of optimality.

$$\mathcal{L}(\vec{x}, \vec{\lambda}, \vec{\mu}) = \sum_{t=1}^{T} \sum_{i=1}^{m} c_{it} x_{it} + \sum_{t=1}^{T} \lambda_t \Big(q_t - A_t \prod_{i=1}^{m} x_{it}^{\alpha_{it}} \Big) - \sum_{t=1}^{T} \sum_{i=1}^{m} \mu_{it} x_{it} \tag{3.4}$$

Let $\partial \mathcal{L} / \partial w$ denote derivative of the Lagrangian function, $\mathcal{L}$, with respect to w. Then the necessary optimality conditions are: (3.3b), (3.3c), (3.5) and (3.6).

$$\mu_{it} \geq 0 \tag{3.5}$$

$$\frac{\partial \mathcal{L}}{\partial x_{it}} = 0 \quad \Leftrightarrow \quad c_{it} - \lambda_t \big(\alpha_{it} x_{it}^{-1} \underbrace{A_t \prod_{i=1}^{m} x_{it}^{\alpha_{it}}}_{q_t} \big) = 0 \quad \forall i, t \tag{3.6}$$

Since (3.6) is valid for all i, we can express λ_t as $\lambda_t = \frac{c_{it}}{\alpha_{it} x_{it}^{-1} q_t}$, which also implies that $\lambda_t = \frac{c_{1t} x_{1t}}{\alpha_{1t} q_t}$ by letting $i = 1$. Now, plugging this into the expression for x_{it} in (3.6) we obtain

$$x_{it} = x_{1t} \Big(\frac{\alpha_{it}}{\alpha_{1t}} \Big) \Big(\frac{c_{1t}}{c_{it}} \Big) \quad \forall i, t. \tag{3.7}$$

Using (3.7) in (3.3b) gives us

$$\begin{aligned} q_t &= A_t x_{1t}^{\alpha_{1t}} \Big(x_{1t} \frac{\alpha_{2t}}{\alpha_{1t}} \frac{c_{1t}}{c_{2t}} \Big)^{\alpha_{2t}} \Big(x_{1t} \frac{\alpha_{3t}}{\alpha_{1t}} \frac{c_{1t}}{c_{3t}} \Big)^{\alpha_{3t}} \ldots \Big(x_{1t} \frac{\alpha_{nt}}{\alpha_{1t}} \frac{c_{1t}}{c_{nt}} \Big)^{\alpha_{nt}} \\ &= A_t x_{1t}^{(1/r_t)} \Big(\frac{c_{1t}}{\alpha_{1t}} \Big)^{(1/r_t)} \prod_{j=1}^{m} \Big(\frac{\alpha_{jt}}{c_{jt}} \Big)^{\alpha_{jt}}. \end{aligned} \tag{3.8}$$

Using (3.8) to derive x_{it} in terms of q_t results in

$$x_{it} = q_t^{r_t} \underbrace{\left(\frac{\alpha_{it}}{c_{it}}\right) \left[\frac{1}{A_t} \prod_{j=1}^{m} \left(\frac{\alpha_{jt}}{c_{jt}}\right)^{-\alpha_{jt}}\right]^{r_t}}_{p_{it}} = p_{it} q_t^{r_t}. \tag{3.9}$$

Now we are ready to reformulate (P1) as a problem in terms of q_t. To do so we replace I_t with $\sum_{s=1}^{t} q_s - \sum_{s=1}^{t} d_s$ and y_t with indicator type notation. Also, using (3.9), we obtain $\sum\limits_{i=1}^{m} c_{it} x_{it} = \sum\limits_{i=1}^{m} c_{it} p_{it} q_t^{r_t} = w_t q_t^{r_t}$. So we obtain reformulated problem $(P_{u,v})$ for the time segment $[u, v]$. This allows us to establish certain structural properties of the optimal solution.

$$(P_{u,v}) \qquad \min_{q_u,\ldots,q_v} F_{u,v} = \sum_{t=u}^{v} \left[(K_t y_t + \left(\sum_{i=t}^{v} h_i \right) q_t + w_t q_t^{r_t} \right] - \sum_{t=u}^{v} h_t D_{u,t} \tag{3.10a}$$

s.t.

$$\sum_{i=u}^{t} q_i \geq D_{u,t}, \qquad t = u, \ldots, v \tag{3.10b}$$

$$q_t \geq 0, \qquad t = u, \ldots, v \tag{3.10c}$$

$$q_t \leq D_{t,v} y_t, \qquad t = u, \ldots, v \tag{3.10d}$$

$$y_t \in \{0, 1\}, \qquad t = u, \ldots, v \tag{3.10e}$$

where y_t denotes the binary variable for a setup. The first set of constraints (3.10b) ensure that all demands will be met and (3.10c) are nonnegativity constraints. The optimization problem at hand is finding $Q_{1,T}^* = (q_1^*, \ldots, q_T^*)$ and $F_{1,T}^*$ for $P_{1,T}$ over the horizon $[1, T]$, where we use $(*)$ to indicate optimality for all entities. In the analysis that follows, we assume, for convenience, that production quantities are non-negative real numbers.

The nonlinear convex production cost is the key difference between our model and the classical well-known model introduced by Wagner and Whitin [5] which is a Mixed integer Programming (MIP) model. The fundamental properties of the optimal solution for $r_t \leq 1$ are that, in an optimal plan, *(i)* production may occur in period t only if $I_{t-1} = 0$ and *(ii)* the entire demand in a period is covered by production in a single period (demand integrality is preserved). For $r_t > 1$, these properties do not hold. This makes the production planning problem in the presence of convex production costs challenging and interesting. To illustrate this point, consider $P_{1,T}$ for the following simple example. For $T = 2$, $K_t = K = 700$, $h_t = h = 1$, $m = 1$, $w_t^1 = w = 0.01$, $r_t^1 = r = 2$ for $1 \leq t \leq T$ and $\mathbf{d} = (100, 300)$. As will be established later, the optimal plan for this problem gives $q_1^* = 175$ and $q_2^* = 225$. Note that neither of the two properties holds; $I_1^* \times q_2^* \neq 0$ and $0 < q_2^* < d_2$. The cost function is plotted in Figure 3.1.

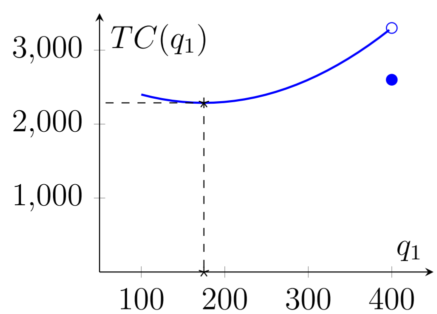


Figure 3.1: The changes of total cost for different feasible q_1 values

Formally, we note that the feasible solution set is convex. A concave function attains its minimum over a convex set at an extreme point. Thus, whenever the cost functions in a lot sizing model is concave, the optimal solution lies on the extreme points. On the other hand, a convex function may attain its minimum in an interior point of the feasible region (as in the example above). Such an interior point solution is called a non-integral plan since the production quantity in each period is not exactly equal to the demand summed over one or more future periods. Our main contribution is to characterize such non-integral solutions (if any) and the related structural results which are provided in the next section.

3.3 Structural results

In this section, we present structural results on the optimal production plan for the dynamic lot-sizing problem $P_{u,v}$ introduced above. In particular, we introduce the key concept of a generation and related definitions; establish the decomposition properties for production subplans in terms of inventory levels and generations, and the characteristics of a production plan for a generation; and, based on these, we characterize the optimal production plan structure. For the special case of $K = 0$, we also provide a planning horizon that rests on merging of generations as problem horizon extends. We begin with the definitions and key concepts.

Definition 1 *In a given production plan, Q_{ij} for periods $\{i, \ldots, j\}$,*

(1) period t is a regeneration point if $I_{t-1} = 0$;

(2) a sequence of periods $\{u, u+1, \ldots, v\}$, for $i \leq u \leq v \leq j$, is a generation, denoted by $\langle u, v \rangle$, if $I_{u-1} = I_v = 0$ and $I_t > 0$ for all $t \in \{u, u+1, \ldots, v-1\}$;

(3) the production plan of a generation is called a production sequence.

Theorem 1 *(Inventory Decomposition Property) Suppose that the constraint*

$$I_k = 0 \qquad \text{for some } k \in \{1, \cdots, t-1\}, \tag{3.11}$$

is added to problem $P_{1,t}$, then, an optimal solution to the original problem can be found by independently finding solutions to the problems for the first k periods and for the last $t-k$ periods.

Inventory decomposition has direct implications on the structure of an optimal production plan. Based on this property, it suffices to consider only production sequences to find the optimal solution to problem $P_{u,v}$ as stated below.

Corollary 1 *(Generation Decomposition Property) An optimal production plan $Q^*_{u,v}$ for problem $P_{u,v}$ consists of production sequences which can be independently solved.*

Proof By assumption, $I_{u-1} = 0$. Clearly, in an optimal production plan, $I^*_v = 0$. If $I^*_t \neq 0$ for $t \in \{u, \ldots, v-1\}$, then there is a single production sequence. Otherwise, $I_k = 0$ for some $k \in \{u, \ldots, v-1\}$. In this case, there are $k+1$ generations by definition. From Theorem 1, each generation can be solved as a sub-problem. Hence, the result. □

In the remainder of this section, we provide results on the characteristics of generations and optimal production sequences.

Corollary 2 *(Generation Characteristics) For a generation $\langle u, v \rangle$,*

(i) $q_u = d_u \geq 0$ *if* $u = v$;

(ii) $\sum_{s=u}^{t} q_s > \sum_{s=u}^{t} d_s$ *for* $t \in \{u, u+1, \ldots, v-1\}$ *if* $u < v$;

(iii) $q_u > 0$ *if* $u < v$;

(iv) $d_v > 0$ *if* $u < v$.

Proof *(i)* Follows from (3.10b). *(ii)* By definition. That is, if $\sum_{s=u}^{t} q_s = \sum_{s=u}^{t} d_s$, then the generation would have ended at $v = t$, which contradicts the definition. *(iii)* Immediately follows from the previous two results. *(iv)* We prove the result by contradiction. Suppose that $d_v = 0$. Then, the inventory balance equation of period v, $I_v = q_v + I_{v-1} - d_v$, implies $0 = q_v + I_{v-1}$, which is possible only if $q_v = I_{v-1} = 0$ due to the non-negativity of these variables. But this contradicts the definition of a generation, hence the result.□

The above lemma implies that a generation whose total demand is zero consists of a single no-production period, and that a generation with at least two periods can neither end with a zero-demand period nor start with a no-production period. Next, we present our results on the structure of the optimal production plan. In any production plan, there may be production and no-production periods. Given a production plan $Q_{u,v}$, let $S(Q_{u,v})$ denote the set of production periods. A special class of production plans forms the basis of the characterization of the optimal solution. We introduce this class below.

Definition 2 *A production plan $Q_{u,v} = (q_u, \ldots, q_v)$ is of class G if*

$$r_i w_i {q_i}^{r_i - 1} = r_j w_j {q_j}^{r_j - 1} - \sum_{s=i}^{j-1} h_s \tag{3.12}$$

for any $i, j \in S(Q_{u,v})$ and $u \leq i < j \leq v$.

Now, we can give the fundamental results on the optimal production plan structure.

Theorem 2 *(Optimal Production Plan Structure I) In an optimal production plan $Q^*_{1,T}$, for any generation $\langle u, v \rangle$,*
(i) $Q^*_{uv} = (d_u)$ if $1 \leq u = v \leq T$,
(ii) $Q^*_{u,v} = (D_{uv}, 0, \ldots, 0)$ if $1 \leq u < v \leq T$ and $r_t \leq 1$ for $t \in [u, v]$,
(iii) $Q^*_{u,v} = (q^*_u, \ldots, q^*_v)$ is of class G if $1 \leq u < v \leq T$ and $r_t > 1$ for $t \in [u, v]$,

Proof *(i)* Directly follows from the definition of a generation. *(ii)* If $r_t \leq 1$ for $t \in [u, v]$, then the production costs are concave. The problem reduces to the classical lot-sizing problem and the result follows. *(iii)* From Theorem 1, $Q^*_{u,v}$ can be found by solving $P_{u,v}$ independently. The proof rests on obtaining the optimal solution for the sub-problem $P_{u,v}$. Similar to Eqn (3.10a), we can write the total cost over the generation $\langle u, v \rangle$, g_{uv} as follows.

$$g_{uv} = \sum_{t \in S(Q^*_{uv})} K_t + \min \sum_{t=u}^{v} \Big[\Big(\sum_{i=t}^{v} h_i \Big) q_t + w_t q_t^{r_t} \Big] - \sum_{t=u}^{v} h_t D_{u,t}$$

Note that g_{uv} is convex in the production quantities, and that the feasible region defined by Eqns (3.10b) and (3.10c) is regular. Constructing the Lagrangean $\mathcal{L}_{uv}$ for $P_{u,v}$, we have

$$\mathcal{L}_{uv} = g_{uv} - \sum_{t=u}^{v} \Big[\lambda_t \Big(\sum_{i=u}^{t} q_i - \sum_{i=u}^{t} d_i \Big) + \mu_t q_t \Big]$$

where λ_t and μ_t above denote the shadow prices of the constraints. From the first order optimality (Karush-Kuhn-Tucker) conditions, we have, for $t \in S(Q^*_{uv})$,

$$\mu_t q^*_t = 0, \tag{3.13}$$

$$\lambda_t \left(\sum_{i=1}^{t} q^*_i - \sum_{i=1}^{t} d_i \right) = 0, \tag{3.14}$$

$$\frac{\partial \mathcal{L}}{\partial q_t}|_{q_t = q^*_t} = r_t w_t q_t^{*r_t - 1} + \sum_{i=t}^{v} (h_i - \lambda_i) - \mu_t = 0. \tag{3.15}$$

Since in a generation, $I_t = \big(\sum_{i=u}^{t} q_i - \sum_{i=u}^{t} d_i \big) > 0$ for $t \in \{u, \ldots, v-1\}$, (3.14) implies that,

$$\lambda^*_t = 0, \qquad t \in \{u, \ldots, v-1\}. \tag{3.16}$$

Substituting (3.16) in (3.15) we find

$$r_t w_t q_t^{*r_t - 1} = \lambda_v - \sum_{i=t}^{v} h_i + \mu_t \quad \forall t \in S(Q^*_{uv}) \tag{3.17}$$

Now, $q^*_t > 0$ for $\forall t \in S(Q^*_{uv})$ together with (3.13) implies that $\mu^*_t = 0$ for $t \in S(Q^*_{uv})$. Hence, writing (3.17) for $i < j$, $i, j \in S(Q^*_{uv})$ gives $r_i w_i q_i^{*r_i - 1} = \lambda_v - \sum_{s=i}^{v} h_s$ and $r_j w_j q_j^{*r_j - 1} = \lambda_v - \sum_{s=j}^{v} h_s$. Equating the two expressions via λ_v, $r_i w_i q_i^{*r_i - 1} = r_j w_j q_j^{*r_j - 1} - \sum_{s=i}^{j-1} h_s$. Clearly, this is a production plan of class

G. Hence, the result.□

The above result implies that it suffices to consider only those feasible production plans that are of class G in order to optimize the problem $P_{u,v}$ for any horizon $[u, v]$. We shall exploit this property when we develop our forward dynamic programming solution approach. Theorem 2 characterizes the relationship among the production quantities within a generation. Next, we establish the relationship between the production quantities of two consecutive generations in an optimal production plan.

Theorem 3 *(Optimal Production Plan Structure II) If $r_t \geq 1$ for all t, in an optimal production plan, for generations $\langle u, v\rangle$ and $\langle v+1, v'\rangle$,*

$$r_{v+1}w_{v+1}(q^*_{v+1})^{r_{v+1}-1} \leq r_l w_l q_l^{*r_l-1} + \sum_{i=l}^{v} h_i, \tag{3.18}$$

where, l is the last production period in $\langle u, v\rangle$.

Proof Let $Q^*_{u,v'}$ be the optimal production plan for $[u, v']$. If $q^*_{v+1} = 0$, the result follows immediately. Otherwise, consider the modified feasible production plan $Q'_{u,v'} = (q'_u, \ldots, q'_{v'})$ such that $q'_l = q^*_l + \epsilon$, $q'_{v+1} = q^*_{v+1} - \epsilon$ and $q'_t = q^*_t$ for $t \in \{u, u+1, \ldots, l-1, l+1, \ldots, v, v+2, \ldots, v'\}$ where $\epsilon > 0$. Due to the optimality of $Q^*_{u,v'}$ we have

$$\begin{aligned}\sum_{i=u}^{v'}[w_i q_i^{*r_i} + h_i I_i] \leq & \sum_{i=u}^{l-1}[w_i q_i^{*r_i} + h_i I_i] \\ & + w_l(q^*_l + \epsilon)^{r^n_l} + \sum_{i=l}^{v} h_i(I_i + \epsilon) + w_{v+1}(q^*_{v+1} - \epsilon)^{r_{v+1}} \\ & + h_{v+1}I_{v+1} + \sum_{i=v+2}^{v'}[w_i q_i^{*r_i} + h_i I_i]\end{aligned}$$

which implies

$$\left[w_{v+1}{q^*_{v+1}}^{r_{v+1}} - w_{v+1}(q^*_{v+1} - \epsilon)^{r_{v+1}}\right] \leq [w_l(q^*_l + \epsilon)^{r_l} - w^n_l q_l^{*r_l}] + (h_l + ... + h_v)\epsilon.$$

Dividing both sides by ϵ and taking the limit $\epsilon \to 0$, we get

$$w_{v+1}\frac{d}{dq_{v+1}}q_{v+1}^{r_{v+1}}|_{q_{v+1}=q^*_{v+1}} \leq w_l\frac{d}{dq_l}q_l^{r_l}|_{q_l=q^*_l} + (h_l + ... + h_v)$$

which becomes

$$r^n_{v+1}w_{v+1}{q^*_{v+1}}^{r_{v+1}-1} \leq r_l w_l {q^*_l}^{r_l-1} + (h_l + ... + h_v).\square$$

The above theorem, enables us to check whether a proposed bisecting of the sub-horizon $[u, v']$ can be optimal. So far, we have provided structural results of the optimal production plans for the general case that allows for non-zero fixed production (setup) costs. Next, we focus on the special case of $K_t = 0 \quad \forall t$, which enables us to obtain further results on the optimal production plans.

3.4 A special case: zero setup costs ($K_t = 0$)

Recall that, in the classical lot-sizing problem with the non-speculative cost structure ($c_t + h_t > c_{t+1} \quad \forall t$), the optimal production plan consists of lot-for-lot productions in the absence of setup costs. This has two implications: *(i)* each period is one generation, and *(ii)* production is done only in periods of non-zero demand. In the presence of production functions, these results no longer hold. In particular, it is optimal to produce in every period within a generation $\langle u, v\rangle$ if $D_{uv} > 0$. This result follows from the property below.

Lemma 1 *If $r_t \geq 1$ and $K_t = 0 \quad \forall t$, in an optimal production plan, for generation $\langle u, v\rangle$, $q^*_j > 0$ if $q^*_t > 0$ for $u \leq t < j \leq v$.*

Proof Proof by contradiction. We first establish that $q^*_{t+1} > 0$. Suppose that in the subplan $Q_{u,v}$, $q_t > 0$ and $q_{t+1} = 0$. We will show that this subplan can be improved. To do so, consider the feasible subplan $Q'_{u,v} = (q'_u, \ldots, q'_v)$ with $q'_t = q_t - \epsilon$, $q'_{t+1} = \epsilon$ and $q'_i = q_i$ for $i \in \{u, \ldots, v\}\backslash\{t, t+1\}$ such that

$0 < \epsilon < \min\{I_t, q_t, 1\}$. By definition of a generation, I_t is positive and q_t is positive by assumption. Therefore, such a positive ϵ which guarantees the feasibility of $Q'_{u,v}$ always can be found. We denote the corresponding costs of these two subplans by π and π'.

$$\begin{aligned}\tilde{\pi} - \pi' &= [w_t q_t^{r_t} + h_t I_t] - [w_t(q_t - \epsilon)^{r_t} + h_t(I_t - \epsilon) + w_{t+1}\epsilon^{r_{t+1}}] \\ &= w_t\Big({q_t}^{r_t} - (q_t - \epsilon)^{r_t}\Big) + h_t\epsilon - w_{t+1}\epsilon^{r_{t+1}} \\ &\geq h_t\epsilon - w_{t+1}\epsilon^{r_{t+1}} = \epsilon(h_t - w_{t+1}\epsilon^{r_{t+1}-1}).\end{aligned}$$

The inequality above follows from nonnegativity of the parenthetical term in the former expression. The last expression is positive for any $\epsilon < \left(\frac{h_t}{\bar{w}_{t+1}}\right)^{1/(\bar{r}_{t+1}-1)}$ if h_t is positive. If $h_t = 0$ then,

$$\begin{aligned}\pi - \pi' &= [w_t {q_t}^{r_t}] - \left[w_t^j(q_t - \epsilon)^{r_t} + w_{t+1}\epsilon^{r_{t+1}}\right] \\ &= w_t\left[{q_t}^{r_t} - (q_t - \epsilon)^{r_t}\right] - w_{t+1}\epsilon^{r_{t+1}}.\end{aligned}$$

Consider the function $f(x) = q^x - (q-\epsilon)^x$. Derivative of this function with respect to x is $f'(x) = \ln(q)q^x - \ln(q-\epsilon)(q-\epsilon)^x$ which is always positive for $(q > \epsilon)$. Therefore, $f(x)$ is an increasing function of x for $(q > \epsilon)$. Let $\lfloor x \rfloor$ be the greatest integer equal or less than x. Then, $w_t\left[{q_t}^{r_t} - (q_t - \epsilon)^{r_t}\right] \geq w_t\left[{q_t}^{\lfloor r_t \rfloor} - (q_t - \epsilon)^{\lfloor r_t \rfloor}\right]$. Suppose $\epsilon < 1$. Then, $w_{t+1}\epsilon^{r_{t+1}} < w_{t+1}\epsilon^{\lfloor r_{t+1} \rfloor}$ for all j, as well. Therefore,

$$\begin{aligned}\pi - \pi' &\geq w_t\left[{q_t}^{\lfloor r_t \rfloor} - (q_t - \epsilon)^{\lfloor r_t \rfloor}\right] - w_{t+1}\epsilon^{\lfloor r_{t+1} \rfloor} \\ &= w_t\Bigg[{q_t}^{\lfloor r_t \rfloor} - {q_t}^{\lfloor r_t \rfloor} + \binom{\lfloor r_t \rfloor}{1}{q_t}^{\lfloor r_t \rfloor - 1}\epsilon + {q_t}^{\lfloor r_t \rfloor - 2} - \binom{\lfloor r_t \rfloor}{2}{q_t}^{\lfloor r_t \rfloor - 2}\epsilon^2 + \\ &\qquad \ldots \pm \epsilon^{\lfloor r_t \rfloor}\Bigg] - w_{t+1}\epsilon^{\lfloor r_{t+1} \rfloor} \\ &\geq \left[M_1\epsilon - M_2(\epsilon^2 + \epsilon^3 + \ldots + \epsilon^{\max(\lfloor r_t \rfloor, \lfloor r_{t+1} \rfloor)})\right] \\ &\geq \left[M_1\epsilon - M_2\frac{\epsilon^2}{1-\epsilon}\right]\end{aligned}$$

where $M_1 = w_t\lfloor r_t \rfloor {q_t}^{\lfloor r_t \rfloor - 1}(> 0)$ and $M_2 = \max_{2 \leq i \leq \lfloor r_t \rfloor}\left\{w_t\binom{\lfloor r_t \rfloor}{i}{q_t}^{\lfloor r_t \rfloor - i} + w_{t+1}\right\}(> 0)$. The last expression is positive for $\epsilon < M_1/(M_1 + M_2)$.

Hence, by choosing any positive ϵ less than $\min\left\{I_t, q_t, 1, \left(\frac{h_t}{w_{t+1}}\right)^{1/(r_{t+1}-1)}, \frac{M_1}{M_1+M_2}\right\}$ the subplan $Q_{u,v}$ can always be improved and, hence, it is not optimal. Having established the result for t and $t+1$, it can be extended to the remaining periods similarly by induction over periods $t+2$ to v. □

It follows from the lemma above that all periods within a generation are production periods provided that the total demand is positive and setup costs are negligible.

For convex production and zero setup costs, the optimal solution behaves in a particular way with respect to demand increases and horizon extensions. If the last period's demand is increased (all else being the same), then in the optimal production plan for the modified problem, (1) the number of generations cannot increase, and (2) the optimal solution to the original problem is retained up to a regeneration point obtained in the original problem. That is, only the last generation in the original solution may merge with previous ones to form a longer last generation in the modified problem's solution. If the problem horizon is extended, then, in the optimal solution, either the new period constitutes the (new) last generation in addition to those obtained in the original problem or the effect of extending the problem horizon is similar to a demand increase in the last period of the original problem. We formally state these properties in the following theorem.

Theorem 4 *(Planning Horizon Theorem) Given a problem $P_{1,t}$ with demands $\boldsymbol{d}_t = (d_1, \ldots, d_t)$ and $r_i > 1$ and $K_i = 0$ for $i = 1, ..., t$, suppose the optimal production plan is $Q^*_{1,t} = Q^*_{t_1,t_2-1} \cup Q^*_{t_2,t_3-1} \cup \ldots Q^*_{t_k,t}$ where k denotes the number of generations in the plan and t_j denotes the regeneration points with $t_1 = 1$.*

- **(i)** *For a modified problem $\bar{P}_{1,t}$ with modified demands $\bar{\boldsymbol{d}}_{1,t} = (d_1, \ldots, d_{t-1}, d_t + x)$ where $x > 0$, the optimal production plan, $\bar{Q}^*_{1,t}$ is given as $Q^*_{t_1,t_2-1} \cup \ldots \cup Q^*_{t_{i-1},t_i-1} \cup \bar{Q}^*_{t_i,t}$ where $\bar{Q}^*_{t_i,t}$ denotes the (new) production sequence for the (new) last generation and $i \in \{1, \ldots, k\}$.*
- **(ii)** *For problem $P_{1,t+1}$ with demands $\boldsymbol{d}_{t+1} = (d_1, \ldots, d_t, d_{t+1})$, the optimal production plan is $Q^*_{1,t+1} = Q^*_{t_1,t_2-1} \cup \ldots \cup Q^*_{t_{i-1},t_i-1} \cup \bar{Q}^*_{t_i,t+1}$ where $\bar{Q}^*_{t_i,t+1}$ denotes the (new) production sequence for the (new) last generation $i \in \{1, \ldots, k+1\}$ with $t_{k+1} = t + 1$ if $r_{t+1} > 1$ and $K_{t+1} = 0$.*

Proof **(i)** We construct the optimal production plan of $\bar{P}_{1,t}$ by changing $Q^*_{1,t}$.

First we construct the new G-class production subplan for $[t_k, t]$, $\bar{Q}_{t_k,t}$. If $\sum_{n=1}^{m} r_{t_k-1} w_{t_k-1} q^{*}_{t_k-1}{}^{(r_{t_k-1}-1)} + h_{t_k-1} \geq r_{t_k} w_{t_k} \bar{q}_{t_k}^{(r_{t_k}-1)}$ then $\bar{Q}^{*}_{1,t}$ is optimal with $i = k$. Otherwise, we can improve it by transferring some portion of the total demand of $[t_k, t]$, say ϵ, to the period t_k. However, this results in $r_{t_k-2} w_{t_k-2} q^{*}_{t_{k-2}}{}^{r_{t_k-2}-1} + h_{t_k-2} \leq r_{t_k-1} w_{t_k-1} (q^{*}_{t_k-1} + \epsilon)^{r_{t_k-1}-1}$ which implies that it can be improved again. By similar argument, transferring some positive portion of x to all periods within $[t_{k-1}, t_k - 1]$ gives a better objective cost. We continue this procedure in backward way until we reach to period t_i such that after augmenting $q^{*}_{t_i}$ to $\bar{q}_{t_i} = q^{*}_{t_i} + \epsilon_{t_i}$ we still have the optimality inequality of Theorem 3, that is, $r_{t_i-1} w_{t_i-1} q^{*}_{t_i}{}^{r_{t_i-1}-1} + h_{t_i-1} \geq r_{t_i} w_{t_i} \bar{q}_{t_i}^{r_{t_i}-1}$ and no further improvement can be made. Hence optimal augmentation of the old production quantities gives a new G-class production subplan and it stops in one of the t_i, $i \in \{1, ..., k\}$.

(ii) If d_{t+1} is such that $r_t w_t q^{*}_t{}^{r_t-1} + h_t \geq r_{t+1} w_{t+1} d_{t+1}^{r_{t+1}-1}$, the given plan $Q^{*}_{1,t+1}$ is optimal with $i = k+1$ and the new period is itself a generation. Otherwise, the rest of the proof follows from part **(i)** by considering $x = d_{t+1} - \tilde{d}$ where $\tilde{d}$ solves $r_t w_t q^{*}_t{}^{r_t-1} + h_t = r_{t+1} w_{t+1} (\tilde{d})^{r_{t+1}-1}$. $\square$

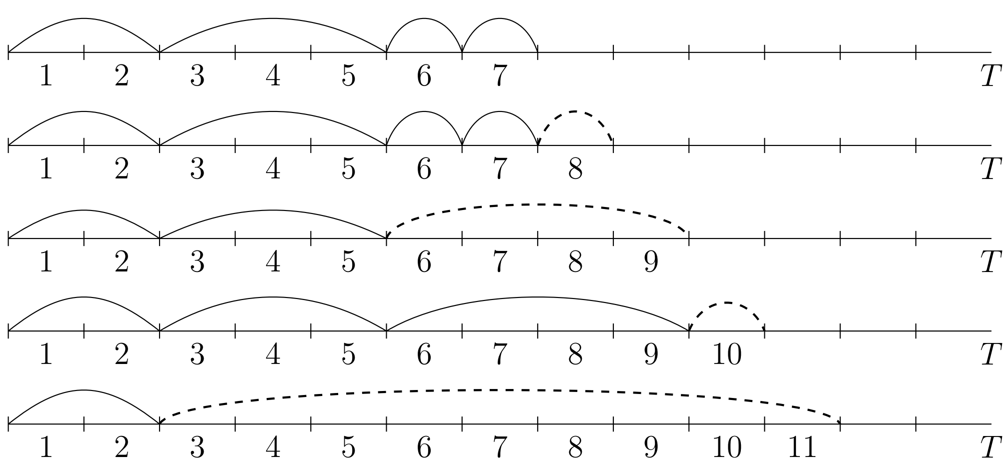


Figure 3.2: Evolution of generations in the optimal solution for $P_{1,T}$ for $T = 7, 8, 9, 10, 11$. ($h_t = h = 0.1$, $m = 1$, $w^1_t = w = 0.01$, $r^1_t = r = 2$, $K_t = 0$ for all $t \in \{1, \ldots, T\}$ and demands $\mathbf{d} = (50, 100, 0, 70, 80, 40, 45, 30, 80, 35, 250)$.)

An illustration of this property is given with an example in Table 3.2 and Figure 3.2 as evolution of the optimal solution is depicted for successively longer problem horizons. As horizon extends from $T = 7$ to $T = 8$, the former set of generations is retained and the new period comprises the last generation, whereas, the last generation merges with three former ones as horizon further extends to $T = 9$. Thus, the last generation in an optimal solution can only extend and its regeneration point can only shift toward the time origin. (See also $T = 10, 11$.) This theorem is of interest for settings where production plans may be done on a rolling horizon basis. In certain cases, the merging of the last generation with the previous ones may continue up to the first period. Unlike the classical lot-sizing problem, there exists no guaranteed partitioning of the problem horizon even for zero setup costs.

3.5 Solution algorithms and heuristics

The dynamic lot sizing problem with convex economic production functions can be solved in a number of ways: Direct application of the available generic optimizers on the given mixed integer nonlinear programming (MINLP) formulation; a backward dynamic programming (DP) formulation with inventory levels as states and time periods as stages; a forward DP formulation with exhaustive and heuristic search subroutines; and, heuristics specially developed for the problem at hand. We considered all of these approaches. Below, we discuss the particulars of each approach with its merits and disadvantages.

Problem $P_{u,v}$ is already formulated as an MINLP problem. Therefore, one option is to employ the commercially available solvers which have been developed for generic MINLP problems. In the next chapter we provide the results and discussion about the performance these optimization packages. A direct application of the given MINLP formulation resulted in poor performance of the available solvers; sometimes no solution could be found at all. To overcome this, a possibility is to consider reformulations of the MINLP problem similar to those in [1] making the problem more amenable to the available solvers. A small numerical study indicated that there is indeed room for improvement in the performance of the generic solvers with different reformulations. But, for large scale problems, we still encountered the difficulties of computational time and iteration limits.

Another option is to obtain the optimal solution to problem $P_{1,T}$ by a general backward dynamic programming (DP) algorithm. To this end, define $J_t^T(I_t)$ as the minimum total cost under an optimal production plan for periods $t+1$ through T, where I_{t-1} is ending inventory as defined before and follows the recursion $I_t = I_{t-1} + q_t - d_t$ for all t. (We retain all other notation introduced previously.) Then

$$J_{t-1}^T(I_{t-1}) = \min_{q_t \geq \max(0, d_t - I_{t-1})} \left\{ K_t \mathbb{1}_{q_t>0} + h_t I_t + w_t q_t^{r_t} + J_t^T(I_t) \right\} \quad t \in \{1, \ldots, T\} \tag{3.19}$$

with $\mathbb{1}_{q_t>0}$ denoting the indicator for a setup and the boundary condition in period T being $J_T^T(I_T) = 0$ for all I_T. The optimal solution is found using the above recursion and $J_0^T(0)$ denotes the minimum cost over the problem horizon. The main difficulty with this backward DP algorithm is the curse of dimensionality. For real valued demands, implementing the above formulation requires discretization of ending inventories (and production quantities) with a suitable step-size, say, δ. Then the total memory requirement for the cost-to-go array is of size $[\sum_{t=1}^T \sum_{i=t}^T d_i]/\delta$. As the problem horizon extends, it becomes prohibitively high preventing its usage for large problems. However, it is possible to use the structural properties of optimal solutions and formulate the problem as a forward DP problem which we discuss next.

Generation decomposition property in Corollary 1 implies that an optimal plan for $P_{u,v}$ can be found by considering generations over $[u, v]$ which can be independently solved. This property forms the basis of the forward dynamic programming recursion which uses only the period information. The logic of the forward DP rests on partitioning any given problem. For any problem horizon t, we construct the feasible production plans by considering the last generation in the plan, $\langle i, t \rangle$, for some $i \in [1, t]$ and the best solution obtained for $[0, i-1]$ where period 0 denotes the time origin for convenience. Formally, we can state the forward DP algorithm as follows. Let f_t^* be the cost under an optimal production plan for $[1,t]$ given that $I_0 = 0$. Then, for $t = 1, \ldots, T$, we have

$$f_t^* = \min_{1 \leq i \leq t-1} \{f_{i-1}^* + g_{i,t}\}, \tag{3.20}$$

where $g_{i,t}$ is the cost associated with generation $\langle i, t\rangle$, $f_0^* \equiv 0$ and f_T^* is the optimal cost for problem $P_{1,T}$. To find the optimal production sequence for generation $\langle i, t\rangle$, we search over the feasible production plans of class G as implied by Theorem 2. Specifically, we start with some production period set, S for the given generation $\langle i, t\rangle$ and solve for the positive production quantities that satisfy the condition for class G plans. (If the obtained production plan is not feasible, it is discarded as having infinite cost.) If necessary, we update the set S and find new production sequences until no further cost improvements are achieved. Recall that, if $K_t = 0 \quad \forall t$, production is done in all periods within a generation except for one-period generations with zero demands. For this case, it suffices to choose the initial S as containing all of the periods $\{i, i+1, \ldots, t\}$ and no updating is necessary. Furthermore, as the algorithm progresses (as t is increased to $t+1$), from Theorem 4, instead of minimizing over $i \in [1,t]$, it is sufficient to consider only the regeneration points $\{t_1, ..., t_k\} \cup \{t\}$, where t_j's, $j = 1, ..., k$, denote the regeneration points obtained for problem horizon t. The above algorithm is guaranteed to give the optimal solution for problem $P_{1,T}$, (*i.e.*, $f_t^* = F_{1,T}^*$). We provide the pseudo-code for the forward DP algorithm in the Algorithm 1. For zero fixed setup costs, it has a computational complexity of $O(T^2)$; in practice, this translates to the algorithm being able to solve large scale problems with 300 periods within a millisecond on a personal computer. For positive setup costs, however, production may not be done in all periods in a generation $\langle i, t\rangle$, and all 2^{t-i+1} possible sets must be considered for S as candidates for new production sequences. The forward DP algorithm that considers all these sets provides the optimal solution and has $O(T^2 2^T)$ run time complexity. But such an exhaustive search is prohibitively time-consuming rendering the exact solution by the given forward DP formulation impractical for $K > 0$ and long problem horizons.

In the absence of reasonably fast exact solution methodologies, one may resort to approximate solutions. We develop a version of the above forward DP algorithm in two forms which are labeled with *inc.ex* and *PsdWW*. The former is used as benchmark. Additionally we propose eight heuristics for problem P_{1T} and we name them with the abbreviate name of the classical lot sizing heuristics which they are based on. The first four heuristics,WW, SM, EOQ and LUC are single step heuristics which are based on stopping rules and variants of the Silver-Meal, EOQ and Least-Unit-Cost based heuristics for the classical lot sizing problem. Heuristic *WW* is a variant of the Wagner-Whitin solution that employs the forward DP algorithm while imposing demand integrality on the production quantities. The first four heuristics are single step heuristics.

The remaining four heuristics, which we call the G-heuristics, are two-step hybrids that use the set of production periods of the solutions obtained by the first three heuristics and improve them via G-class production subplans .

For all heuristics, we adopt the following notation. The solution for problem $P_{1,T}$ obtained under heuristic j consists of the set of production quantities denoted by $Q_T^{(j)} = \{q_1^{(j)}, q_2^{(j)}, \ldots, q_T^{(j)}\}$ and the index set of production periods for the problem horizon denoted by $\Omega_T^{(j)}$ in which period t is a production period if $q_t^{(j)} > 0$ for $t = 1, \ldots, T$, and results in the cost, $f_T^{(j)} = \sum_{t \in \Omega_T^{(j)}} K_t + \sum_{t=1}^{T} \left[h_t I_t + w_t (q_t^{(j)})^{r_t} \right]$ with I_t as defined before. Below, we explain the construction and particulars of each heuristic in detail.

Heuristic *SM* is similar in construction to the heuristic in Silver and Meal (1973) developed for the classical dynamic lot sizing problem. Under this heuristic, the beginning period of each generation is its sole production period. That is, for any generation $\langle u, v \rangle$, we set $q_i = D_{u,v}$ for $i = u$ and $q_i = 0$ for $u < i \leq v$. The generations themselves are obtained in a forward manner along the problem horizon by means of a stopping rule. A generation starting in period u terminates in period $u + \hat{l}(u)$ where $\hat{l}(u) = \max\{l : \frac{g_{u,u+l}^{(SM)}}{l} \geq \frac{g_{u,u+l+1}^{(SM)}}{l+1}, u \leq l \leq T\}$ with $g_{u,v}^{(SM)} := K_u + \left[\sum_{s=u}^{v-1} h_s D_{s+1,v} \right] + \left[w_u D_{u,v}^{r_u} \right]$ being the cost associated with the periods $[u, v]$. The generation terminates at $\hat{l}(u)$ because the cost per period starts to increase after that. The solution algorithm starts with the initial period of the problem horizon. Once the stopping rule is satisfied and $\hat{l}(1)$ is found, the production plan over $[1, \hat{l}(1) + 1]$ is retained and the procedure is repeated for the remaining periods starting with period $\hat{l}(1)+1$ until the entire horizon is covered. The pseudo-code is provided at the end of the chapter and has $O(T)$ computational complexity. Under this heuristic, the quantity produced in period t is given as $q_t^{(SM)} = D_{t,\hat{l}(t)}$ if $t \in \Omega_T^{(SM)}$ and zero, otherwise. Then, we have $f_T^{(SM)} = \sum_{i\in\Omega_T^{(SM)}} g_{i,i+\hat{l}(i)}^{(SM)}$. By design, with this heuristic, demand integrality is preserved in production quantities and each production period constitutes a generation start in the solution. The stopping rule computation differs from the classical Silver-Meal heuristic in order to incorporate the nonlinear production costs in our setting.

Heuristic *LUC* is similar in construction to the Least-Unit-Cost developed for the classical dynamic lot sizing problem. Similar to *SM*, the beginning period of each generation is its sole production period and the generations are obtained in a forward manner by means of a stopping rule. A generation starting in period u terminates in period $u+\hat{l}(u)$ where $\hat{l}(u) = \max\{l : \frac{g_{u,u+l}^{(LUC)}}{D_{u,u+l}} \geq \frac{g_{u,u+l+1}^{(LUC)}}{D_{u,u+l+1}}, u \leq l \leq T\}$ with $g_{u,v}^{(LUC)} := K_u + \left[\sum_{s=u}^{v-1} h_s D_{s+1,v}\right] + \left[w_u D_{u,v}^{r_u}\right]$ being the cost associated with the periods $[u,v]$. The generation terminates at $\hat{l}(u)$ because the cost per unit of product starts to increase after that. Except the stopping rule, the structure of this heuristic is similar to SM and the quantity produced in period t is given as $q_t^{(LUC)} = D_{t,\hat{l}(t)}$ if $t \in \Omega_T^{(LUC)}$ and zero, otherwise. Then, we have $f_T^{(LUC)} = \sum_{i\in\Omega_T^{(LUC)}} g_{i,i+\hat{l}(i)}^{(LUC)}$.

Heuristic *EOQ* is based on a variant of the economic order quantity (EOQ) model which was developed by Harris [71] for linear acquisition costs. To develop the heuristic, consider the following stylized continuous time counterpart for our production setting. Demand for the item is deterministic with a constant rate, D over an infinite problem horizon with stationary cost parameters. Production is done in lots of constant size $\tilde{Q}$ (because of infinite horizon) incurring nonlinear costs in the production quantity. The objective is to minimize the total cost rate $TC(\tilde{Q}) = KD/\tilde{Q} + h\tilde{Q}/2 + \left[\sum_{n=1}^{m} w^n \tilde{Q}^{r^n}\right] D/\tilde{Q}$ where K stands for the fixed setup cost and h for the unit holding cost rate. Let the minimum total cost rate be denoted by TC^* and the corresponding optimal lot size by $\tilde{Q}^*$. We have the following result.

Lemma 2 *The total cost rate $TC(\tilde{Q})$ is quasi-convex for $r \geq 1$ and has a unique minimizer $\tilde{Q}^*$ which solves*

$$K - h(\tilde{Q}^*)^2/2D + (1-r)w(\tilde{Q}^*)^{r-1} = 0.$$

Proof The proof rests on standard optimization techniques. We first solve for the extrema that satisfy the first order condition. We have $\frac{d}{dQ}TC(Q) = TC'(Q) = -KD/Q^2 + h/2 - [wQ^r]\, D/Q^2 + [rwQ^{r-1}]\, D/Q = 0$
$\Rightarrow K + (1-r)w^n Q^{*r} = \frac{h}{2D}Q^{*2}$

$$\Rightarrow KD = \frac{h}{2}Q^{*2} - D(1-r)wQ^{*r} \tag{3.21}$$

where Q^* denote(s) the extremum(a). Considering the second derivative evaluated at the extrema, we have
$\frac{d}{dQ}TC'(Q) = TC''(Q) = 2KD/Q^3 + \left[(r-2)(r-1)wQ^{r-3}\right] D$
where at the critical point(s) of TC since (3.21) holds, we have
$TC''(Q^*) = hQ^{*2} - 2D\left[(1-r)wQ^{*r}\right] + D\left[(r-2)(r-1)wQ^{r-3}\right]$
$= hQ^{*2} + D\left[(-2+2r+r^2-3r+2)wQ^{*r}\right]$
$= hQ^{*2} + D\left[r(r-1)wQ^{*r}\right] > 0.$
So the function is convex in its all critical points and therefore, is a quasi-convex. Therefore, it has a unique extremum point.□

Note that the above result reduces to the classical EOQ result for $r = 1$. For the general case, it does not yield a closed-form solution for $\tilde{Q}^*$ but the uniqueness of the solution allows for an efficient linear search for it. Under heuristic *EOQ* the production quantity in period t is found as $q_t^{(EOQ)} = \min\Big([D_{t,T} - I_{t-1}]^+, \max([d_t - I_{t-1}]^+, \tilde{Q}^*)\Big)$ for $1 \leq t \leq T$ starting with $I_0 = 0$. The solution algorithm starts with the initial period of the problem horizon, and production quantities are obtained as one proceeds over the entire problem horizon. The pseudo-code for the algorithm is provided in the Algorithm 1 and has $O(T)$ computational complexity. We have $f_T^{(2)} = \sum_{t\in\Omega_T^{(EOQ)}} K_t + \sum_{t=1}^{T}\left[h_t I_t + w_t^n (q_t^{(2)})^{r_t}\right]$. The condition on the net remaining total demands

($[D_{t,T} - I_{t-1}]^+$) ensures ending inventory to be zero. Unlike the above heuristic, demand integrality is not preserved under this heuristic.

The rest of the heuristics and the inclusion/exclusion DP based benchmark heuristic employ the forward DP algorithm introduced above and obtain solutions by means of simple rules to construct the set S in a generation resulting in a possibly suboptimal solution. The benchmark algorithm differs from the exact one only in its construction of S. It has the advantage of providing solutions within reasonable times and the goodness of the solutions can be improved by developing efficient set-construction heuristics. Below, we explain the details of these heuristics.

Heuristic WW is obtained by employing the forward recursive procedure in Eqn(3.20) while imposing the condition that demand integrality is preserved. Hence, for any generation $\langle u, v\rangle$ in the solution, we set the quantity produced in period i, $q_i^{(WW)} = D_{u,v}$ for $i = u$ and $q_i^{(WW)} = 0$ for $u < i \leq v$ and search over all possible generations over the problem horizon. Let $g_{i,t}^{(WW)}$ be the total cost of the subproblem $[i, t]$ which constitutes a single generation $\langle i, t\rangle$, $f_0^{WW} \equiv 0$. Then, for $t = 1, \ldots, T$, $f_t^{(WW)} = \min_{1\leq i\leq t}\{f_{i-1}^{(WW)} + g_{i,t}^{(WW)}\}$ where $g_{i,t}^{(WW)} = K_i + \left[\sum_{s=i}^{t-1} h_s D_{s+1,t}\right] + \left[w_t D_{i,t}^{r_i}\right]$. Due to the imposition of demand integrality, this heuristic may be viewed as a version of the classical Wagner-Whitin solution methodology. It has the same computational complexity as the classical Wagner-Whitin algorithm [5] and it reduces to the solution in the classical setting, for $r_t = 1$. In its implementation, the forward DP algorithm is employed wherein the production period set S for a generation $\langle u, v\rangle$ is constructed as consisting of only period u. Aside from being a viable fast solution technique, heuristic WW is important in that its performance illustrates the significance of demand splitting in the case of nonlinear production costs and the importance of class G production subplans.

Next, we introduce heuristics *GSM*, *GLUC*, *GEOQ* and *GWW* which exploit the G-class property of the production subplans. They work as follows. First, we obtain an initial solution to the problem $P_{1,T}$ by one of the above four heuristics.

Of this initial solution obtained via heuristic j, we take only the set of production periods $\Omega_T^{(j)}$, and use it as the given global set of production periods. That is, as we implement the forward DP algorithm, we construct the set S for the generation $\langle u, v\rangle$ using the subset of $\Omega_T^{(j)}$ corresponding to the problem subhorizon $[u, v]$. In practice, this amounts to simply reading off the indexes of the production periods, if any, in the set-construction subroutine. The rest of the algorithm is applied as before. Hence, these heuristics are two-step improvement extensions of first four ones. That is, *GSM* takes $\Omega_T^{(SM)}$ obtained by heuristic *SM* and improves on it via class G subplans in accordance with Theorem 2, heuristic *GEOQ* takes $\Omega_T^{(EOQ)}$ obtained by heuristic *EOQ* and improves on it, and so forth. By construct, the use of the initial solutions implies that we construct the set S *a priori* and, hence, need only to consider a smaller fraction of class G subplans. This greatly reduces the computational effort. The benchmark forward DP algorithm has $O(T^2)$ computational time complexity, given that $\Omega_T^{(i)}$ is provided as pre-processed data. We denote the usage of these S-construction heuristics in the pseudo-codes as instructions denoted by $\Omega_T^{(i)} \to S$. The performance of this group of heuristics depends, to some extent, on the performance of the initial solution which gives $\Omega_T^{(i)}$. But, the significant improvements over the initial solutions indicate that developing the G-class subplans for generations is the main factor for obtaining good solutions.

Now, we consider another method of constructing the set S for a generation in the solution: *inc.ex*. In this method, we create the set S for each generation under consideration according to three set-construction rules used conjunctively as the algorithm proceeds over the problem horizon. *(i)* The first rule is a greedy exclusion rule. Initially, S contains all periods within the generation. One by one, each period (other than the first) is excluded in the updated S. The best is retained and the greedy improvement is repeated with the remaining periods until no further improvement. To avoid possible local optima, we also implemented a scatter search by means of updating S randomly as follows. *(ii)* The second rule is a randomized exclusion rule. This is the randomized version of the greedy exclusion rule. Initially, S is full. A period is randomly selected to be excluded from the updated S. This is repeated for n times. The best is retained and the greedy improvement is repeated with the remaining periods until no further improvement. *(iii)* Finally, a randomized inclusion rule. Initially, S contains only the first period of the generation. This corresponds to the solution in the classical dynamic lot-sizing problem. A period is randomly selected to be included in the updated S. This is repeated for n times. The best is retained and the greedy improvement is repeated with the remaining periods until no further improvement. The conjunctive use of these rules implies that, for a generation considered in the solution, set S that gives the minimum cost among all those constructed by the three rules is taken as the production period set for that generation. With the implementation of the S-construction subroutine using the above rules, the forward DP algorithm has a computational complexity of $O(T^4)$ in the presence of positive setup costs. Clearly, this algorithm cannot guarantee optimality for positive setup costs; however, our preliminary numerical tests (with problem horizon length of 100 periods) indicate that the suboptimality decreases significantly for long problem horizons with average deviations from the optimal (obtained by backward DP algorithm) of approximately 0.1%. Therefore, we adopted this solution algorithm as our *benchmark* solution methodology.

Heuristic PsdWW is in fact another variant of *inc.ex* which is based on a forward DP. Similarly, the production periods are not predetermined, rather, the set S is determined as the recursion proceeds. In psdWW, to form a subplan within $[u, v]$, the first period is selected as a production period, i.e. $S = \{u\}$, and the cost of the subplan is calculated as $g_1^{psdWW} = K_u + \sum_{t=u}^{v-1} h_t I_t + w_u (D_{uv})^{r_u}$. Then the second one, $u{+}1$, is also selected to be a production period, i.e. $S = \{u, u{+}1\}$, and its cost is calculated as $g_2^{psdWW} = K_u + K_{u+1} + \sum_{t=u}^{v-1} h_t I_t + w_u(\tilde{q}_u)^{r_u} + w_u(\tilde{q}_{u+1})^{r_{u+1}}$ where $r_u w_u (\tilde{q}_u)^{r_u} + h = w_{u+1}(\tilde{q}_{u+1})^{r_{u+1}}$. If $g_1^{psdWW} > g_2^{psdWW}$ then we make also the third period a production period and continue this procedure until $g_n^{psdWW} < g_{n+1}^{psdWW}$ and we keep the subplan corresponding to g_n^{psdWW}. This procedure results in a production plan consisting of production subplans -separated with regeneration points- in which the first or more consecutive periods are production periods. In fact, if only a single production period is allowed in each subplan, then the Wagner-Whitin type solution will be obtained. That is why we have termed this heuristic as pesudo-Wagner-Whitin.

Algorithm 1 Forward DP and Heuristic Algorithms

Require: Problem instance, Algorithm name
Ensure: Computes solution for $P_{1,T}$

function FORWARD_ZEROSETUP()
  for $i = 1$ to T **do**
    $f_i = f_{i-1} + g_{i,i}$
    $RegP(i) \leftarrow i$ ▷ $RegP(i)$ =first period of the generation including i
    $clrp \leftarrow i$ ▷ clrp=candidate last regeneration point
    while $clrp > 1$ and feasible_class-G_exists **do**
      if class-G sequence for the segment $(clrp, i)$ is feasible **then**
        $f_i = f_{clrp-1} + g_{clrp,i}$
        $RegP(i) = clrp$
      else
        feasible_class-G_exists← False
      end if
      $clrp - 1 \leftarrow RegP(clrp - 1)$
    end while
  end for
end function

function $SM()$
  while $i \leq T$ **do**
    $l \leftarrow -1$
    repeat
      $l \leftarrow l + 1$
    until $\left(\frac{g^{SM}_{i,i+l}}{l} \leq \frac{g^{SM}_{i,i+l+1}}{l+1}\right)$ or $(i + l \geq T)$
    $q_i \leftarrow D_{i,i+l}$
    $i \leftarrow i + l + 1$
  end while
end function

function $LUC()$
  while $i \leq T$ **do**
    $l \leftarrow -1$
    repeat
      $l \leftarrow l + 1$
    until $\left(\frac{g^{LUC}_{i,i+l}}{D_{i,i+l}} \leq \frac{g^{LUC}_{i,i+l+1}}{D_{i,i+l+1}}\right)$ or $(i + l \geq T)$
    $q_i \leftarrow D_{i,i+l}$
    $i \leftarrow i + l + 1$
  end while
end function

function $EOQ()$
  $\tilde{Q}^* \leftarrow \arg\min_{\tilde{Q}>0} TC(\tilde{Q})$
  $Ds \leftarrow D_{1,T}$: Total remaining demand
  for $i = 1$ to T **do**
    update net demand: $\hat{d}$
    $q_i = \min(Ds, \max(\hat{d}_i, \tilde{Q}^*))$
    $Ds \leftarrow Ds - q_i$
  end for
end function

Algorithm 1 (Continued)

function $WW()$
  $f_0 \leftarrow 0$
  for $i = 1$ to T **do**
    $f_i = f_{i-1} + g^{WW}_{i,i}, \quad q_i \leftarrow D_{i,i}$
    for j=i-1 down to 1 **do**
      if $f_{j-1} + g^{WW}_{j,i} < f_i$ **then** $f_i \leftarrow f_{j-1} + g^{WW}_{j,i}, \quad q_i \leftarrow D_{i,j}$
    end for
  end for
end function

function InclusionExclusion(j, i)
  for $n = 1$ to 3 **do** ▷ method selection : n
    for $nn = 1$ to $Nrand$ **do** ▷ number of randomizations : Nrand
      if $n = 1, 2$ **then**
        $S = \{j, \ldots, i\}$
      else
        $S \leftarrow \{j\}$
      end if
      $Q_{j,i} \leftarrow$ class-G sequence induced by S
      if $Q_{j,i}$ is infeasible **then**
        $Q_{j,i} \leftarrow (d_j, d_{j+1}, \ldots, d_i)$
      end if
      $g_{ji} \leftarrow$ cost of $Q_{j,i}$
      repeat
        Improvement_Observation $\leftarrow$ False
        switch n
          case 1 $klow = j$, $kup = i$ ▷ greedy exclusion heuristic
          case 2 $klow =$ Random index from $\{j, \ldots, i\} \backslash S$, $kup = klow$ ▷ random exclusion heuristic
          case 3 $klow =$ Random index from S, $kup = klow$ ▷ random inclusion heuristic
        for $k = klow$ to kup **do**
          if $k \notin S$ and $n = 1, 2$ **then** $S' = S \cup \{k\}$
          else $S' = S \backslash \{k\}$
          end if
          $Q'_{j,i} \leftarrow$ class-G sequence induced by S'
          $g'_{j,i} \leftarrow$ cost of $Q'_{j,i}$
          if $Q'_{j,i}$ is feasible and $g'_{j,i} < g_{ji}$ **then**
            $candidate \leftarrow j$, Improvement_Observation$\leftarrow$ True
            $g_{ji} \leftarrow g'_{j,i}$, $Q_{j,i} \leftarrow Q'_{j,i}$
          end if
        end for
        $S \leftarrow S'$
      until Improvement_Observation=False
    end for
  end for
end function

Algorithm 1 (Continued)

function PSDWW_SUBPLAN(i, j)
  $k \leftarrow i, \quad n \leftarrow 1$
  $Q_{ij}^1 \leftarrow (D_{ij}, 0, \ldots, 0)$
  $g_1^{psdWW} \leftarrow$ cost of Q_{ij}^1
  repeat
    $k \leftarrow k + 1$, $n \leftarrow n + 1$,
    $S \leftarrow S \cup k$
    $Q_{ij}^k \leftarrow$ type-G production plan induced by S
    $g_n^{psdWW} \leftarrow$ the cost of Q_{ij}
  until $(g_{n-1}^{psdWW} < g_n^{psdWW})$
end function

function G-HEURISTICS_AND_FORWARDDPS$(algName)$
  $f_0 \leftarrow 0$ ▷ $algName \in \{GSM, GLUC, GEOQ, GWW, inc.ex, psdWW\}$
  if algName$= H_k$ **then** ▷ $H_k \in \{GSM, GLUC, GEOQ, GWW\}$
    Call H_{k-4} ▷ $H_{k-4} \in \{SM, LUC, EOQ, WW\}$
    $\Omega \leftarrow$ production period indices
  end if
  for $i = 1$ to T **do**
    $f_i = f_{i-1} + g_{i,i}$
    for $j = i - 1$ down to 1 **do**
      if algName$\neq ForwardDP$ **then**
        $S \leftarrow$ indices in $[j, i] \cap \Omega$
        $Q_{j,i} \leftarrow$ class-G sequence induced by S
      else Call $InclusionExclusion(j, i)$
      end if
      if $Q_{j,i}$ is infeasible **then** $g_{j,i} \leftarrow \infty$ **else** $g_{j,i} \leftarrow$ cost of $Q_{j,i}$
      if $f_{j-1} + g_{j,i} < f_i$ **then** $f_i \leftarrow f_{j-1} + g_{j,i}$
    end for
  end for
end function

Before we proceed with our detailed numerical study, we illustrate the implementation of the proposed solution algorithms through a small example. We have $h_t = h = 0.1$, $w_t^1 = w = 0.01$, $r_t^1 = r = 2$, $K_t = K$ for all $t \in \{1, \ldots, T\}$, $T = 12$, $K = \{0, 100\}$ and the demand vector, $\mathbf{d} = (50, 100, 0, 70, 80, 40, 45, 30, 80, 35, 250, 75)$. We assume that production quantities can be real numbers. In Table 3.1 , we present the optimal production plans $Q^*_{1,i}$, the corresponding total cost $f^*_{1,i}$, the regeneration points in the optimal solution and the candidate solutions developed for problem $P_{1,i}$ as the DP progresses over the horizon length $i = 1, \ldots, T$ for $K = 0$. Note that for zero setup costs, the forward DP is guaranteed to find the optimal. But, for $K > 0$, the forward algorithm does not guarantee the optimal solution. In Table 3.2, for different sub-problem horizon lengths i, we present the optimal production plan $Q^*_{1,i}$ and the corresponding total cost $F^*_{1,i}$ as obtained by the backward DP algorithm and the counterparts $\tilde{Q}_{1,i}$ and $\tilde{F}_{1,i}$ obtained by the forward DP employing with a discretization increment of $\delta = 0.01$ units. As the forward algorithm partitions the problem into the last generation $\langle k+1, i \rangle$ and the sub-horizon $[1, k]$, it results in some (globally suboptimal) local optima. Although the resulting production plans may differ significantly, the resulting maximum cost deviation from the optimal is about 0.95% for $i = 5$ and less than 0.014% for $i = 12$. As the problem horizon increases, the performance of the forward algorithm improves, as expected. For the illustrative example, only one of the set-construction rules (the greedy inclusion updating routine) has been used to find the best production sequence for the last generation. Based on similar preliminary studies, the other two rules (randomized search routines discussed above) have been developed and implemented which result in significant improvements within generations. Hence, they have been embedded to be used conjunctively in the benchmark solution algorithm for the numerical study. For the case of $K = 100$, we provide the solutions obtained with the proposed eight additional heuristics in Tables 3.3–3.4.

i	$Q^*_{1,i}$	f^*_i	Regeneration points	Minimization search
1	$\{50\}$	25	$\{1\}$	$\{\mathbf{g_{1,1}}\}$
2	$\{72.5, 77.5\}$	114.88	$\{1\}$	$\{\mathbf{g_{1,2}}, f^*_1 + g_{2,2}\}$
3	$\{72.5, 77.5\}, \{0\}$	114.88	$\{1, 3\}$	$\{g_{1,3}, \mathbf{f^*_2} + \mathbf{g_{3,3}}\}$
4	$\{72.5, 77.5\}, \{32.5, 37.5\}$	142.75	$\{1, 3\}$	$\{g_{1,4}, \mathbf{f^*_2} + \mathbf{g_{3,4}}, f^*_3 + g_{4,4}\}$
5	$\{72.5, 77.5\}, \{45, 50, 55\}$	197.38	$\{1, 3\}$	$\{g_{1,5}, \mathbf{f^*_2} + \mathbf{g_{3,5}}, f^*_4 + g_{5,5}\}$
6	$\{72.5, 77.5\}, \{45, 50, 55\}, \{40\}$	213.38	$\{1, 3, 6\}$	$\{g_{1,6}, f^*_2 + g_{3,6}, \mathbf{f^*_5} + \mathbf{g_{6,6}}\}$
7	$\{72.5, 77.5\}, \{45, 50, 55\}, \{40\}, \{45\}$	233.63	$\{1, 3, 6, 7\}$	$\{g_{1,7}, f^*_2 + g_{3,7}, f^*_5 + g_{6,7}, \mathbf{f^*_6} + \mathbf{g_{7,7}}\}$
8	$\{72.5, 77.5\}, \{45, 50, 55\}, \{40\}, \{45\}, \{30\}$	242.63	$\{1, 3, 6, 7, 8\}$	$\{g_{1,8}, f^*_2 + g_{3,8}, f^*_5 + g_{6,8}, f^*_6 + g_{7,8}, \mathbf{f^*_7} + \mathbf{g_{8,8}}\}$
9	$\{72.5, 77.5\}, \{45, 50, 55\}, \{41.25, 46.25, 51.25, 56.25\}$	296.44	$\{1, 3, 6\}$	$\{g_{1,9}, f^*_2 + g_{3,9}, \mathbf{f^*_5} + \mathbf{g_{6,9}}, f^*_7 + g_{8,9}, f^*_6 + g_{7,9}, f^*_8 + g_{9,9}\}$
10	$\{72.5, 77.5\}, \{45, 50, 55\},\{41.25, 46.25, 51.25, 56.25\},\{35\}$	308.69	$\{1, 3, 6, 10\}$	$\{g_{1,10}, f^*_2 + g_{3,10}, f^*_5 + g_{6,10}, \mathbf{f^*_9} + \mathbf{g_{10,10}}\}$
11	$\{72.5, 77.5\}, \{50, 55, 60, 65, 70, 75, 80, 85, 90\}$	629.38	$\{1, 3\}$	$\{g_{1,11}, \mathbf{f^*_2} + \mathbf{g_{3,11}}, f^*_5 + g_{6,11}, f^*_{10} + g_{11,11}\}$
12	$\{72.5, 77.5\}, \{50, 55, 60, 65, 70, 75, 80, 85, 90\}, \{75\}$	685.63	$\{1, 3, 11, 12\}$	$\{g_{1,12}, f^*_2 + g_{3,12}, \mathbf{f^*_{11}} + \mathbf{g_{12,12}}\}$

Table 3.1: Forward dynamic programming algorithm solution. ($w^1_t = w = 0.01$, $h_t = h = 0.1$, $r^1_t = r = 2$, $K_t = 0$ for all $t \in \{1, \ldots, T\}$.)

i	$Q^*_{1,i}$	$f^*_{1,i}$	$\tilde{f}_{1,i}$	$\tilde{Q}_{1,i}$
1	{50}	125	125	{50}
2	{72.5,77.5}	314.88	314.88	{72.5,78.5}
3	{72.5,77.5},{0}	314.88	314.88	{72.5,78.5},{0}
4	{107,113,0,0}	461.88	463.88	{72.5,78.5},{0},{70}
5	{93.33,98.33,0,108,34,0}	621.83	627.75	{72.5,78.5},{0},{72.5,77.5}
6	{75,80,0,90,95,0}	701.5	701.75	{72.5,78.5},{0},{92.5,97.5,0}
7	{93.33,98.33,0,108,34,0},{85,0}	798.59	804.46	{72.5,78.5},{0},{73.3,78.3,83.3,0}
8	{75,80,0,90,95,0},{75,0}	860.75	861	{72.5,78.5},{0},{92.5,97.5,0},{75,0}
9	{75,80,0,90,95,0},{75,0},{80}	1024.75	1025	{72.5,78.5},{0},{92.5,97.5,0},{75,0},{80}
10	{75,80,0,90,95,0},{90,0,100,0}	1092	1092.25	{72.5,78.5},{0},{92.5,97.5,0},{90,0,100,0}
11	{75,80,0,90,95,0},{98.75,0,108.75,113.75,118.75}	1613.82	1614.06	{72.5,78.5},{0},{92.5,97.5,0},{98.75,0,108.75,113.75,118.75}
12	{75,80,0,90,95,0},{98.75,0,108.75,113.75,118.75},{75}	1770.06	1770.31	{72.5,78.5},{0},{92.5,97.5,0},{98.75,0,108.75,113.75,118.75},{75}

Table 3.2: Comparison of solutions for $P_{1,i}$ obtained by backward and forward dynamic programming algorithms, $Q^*_{1,i}$, $f^*_{1,i}$ and $\tilde{Q}^*_{1,i}$, $\tilde{f}^*_{1,i}$ ($w^1_t = w = 0.01$, $h_t = h = 0.1$, $r^1_t = r = 2$, $K_t = 100$ for all $t \in \{1, \ldots, T\}$, $T = 12$.)

T	Q_T Heuristic *SM*	f_T	f_T	Q_T Heuristic *GSM*
1	{50}	125	125	{50}
2	{50},{100}	325	314.88	{72.5,77.5}
3	{50},{100},{0}	325	314.88	{72.5,77.5},{0}
4	{50},{100},{0},{70}	474	463.88	{72.5,77.5},{0},{70}
5	{50},{100},{0},{70},{80}	638	627.75	{72.5,77.5},{0},{72.5,77.5}
6	{50},{100},{0},{70},{120,0}	722	701.5	{75,80,0,90,95,0}
7	{50},{100},{0},{70},{120,0},{45}	842.25	821.75	{75,80,0,90,95,0},{45}
8	{50},{100},{0},{70},{120,0},{75,0}	881.25	860.75	{75,80,0,90,95,0},{75,0}
9	{50},{100},{0},{70},{120,0},{75,0} ,{80}	1045.25	1024.75	{75,80,0,90,95,0},{75,0},{80}
10	{50},{100},{0},{70},{120,0},{75,0} ,{115,0}	1117	1092	{75,80,0,90,95,0},{90,,0,100,0}
11	{50},{100},{0},{70},{120,0},{75,0} ,{115,0},{250}	1842	1669.14	{88.57,93.57,0,103.57,108.57,0,118.57,0,128.57,0,138.57}
12	{50},{100},{0},{70},{120,0},{75,0},{115,0},{325,0}	2280.75	1892.53	{99.29,104.29,0,114.29,119.29,0,129,29,0,139.29,0,149.29,0}
T	Q_T Heuristic *LUC*	f_T	f_T	Q_T Heuristic *GLUC*
T	Q_T	f_T	f_T	Q_T
1	{50}	125	125	{50}
2	{150,0}	335	335	{150,0}
3	{150,0,0}	335	335	{150,0,0}
4	{150,0,0},{70}	484	484	{150,0,0},{70}
5	{150,0,0},{70},{80}	648	647.88	{150,0,0},{72.5,77.5},{40}
6	{150,0,0},{70},{80},{40}	764	763.88	{150,0,0},{73.33, 78.33, 83.33,0}
7	{150,0,0},{70},{80},{85,0}	824.75	824.58	{150,0,0},{73.33, 78.33, 83.33,0}
8	{150,0,0},{70},{80},{85,0},{30}	933.75	933.58	{150,0,0},{73.33, 78.33, 83.33,0}, {30}
9	{150,0,0},{70},{80},{85,0},{110,0}	1053.75	1051.5	{150,0,0},{77.5,82.5,87.5,0,97.5,0}
10	{150,0,0},{70},{80},{85,0},{110,0}, {35}	1166	1163.8	{150,0,0},{77.5,82.5,87.5,0,97.5,0},{35}
11	{150,0,0},{70},{80},{85,0},{110,0}, {35},{235}	1891	1654.2	{150,0,0},{88.33,93.33,98.33,0},{108.33,0,118.33,123.33}
12	{150,0,0},{70},{80},{85,0},{110,0}, {35},{235},{75}	2047.25	1810.4	{150,0,0},{88.33,93.33,98.33,0},{108.33,0,118.33,123.33},{75}

Table 3.3: Illustrative example showing solutions of heuristics *SM, GSM, LUC, GLUC* ($w_t^1 = w = 0.01$, $h_t = h = 0.1$, $r_t^1 = r = 2$, $K_t = 100$ for all $t \in \{1, \ldots, T\}$.)

T	Q_T Heuristic EOQ	f_T	f_T	Q_T Heuristic $GEOQ$
1	{50}	125	125	{50}
2	{96.82,53.18}	326.71	314.88	{72.5,77.5}
3	{95.35,54.65},{0}	325.31	314.88	{72.5,77.5},{0}
4	{95.74,95.74,0,28.51}	504.34	463.88	{72.5,77.5},{0},{70}
5	{96.08,96.08,0,96.08,11.77}	698.17	627.75	{72.5,77.5},{0},{72.5,77.5}
6	{95.86,95.86,0,95.86,52.42,0}	726.84	701.50	{75,80,0,90,95,0}
7	{95.74,95.74,0,95.74,95.74,0,2.03}	898.90	821.75	{75,80,0,90,95,0},{45}
8	{95.50,95.50,0,95.50,95.50,0,32.99,0}	1010.52	960.75	{75,80,0,90,95,0},{75,0}
9	{95.74,95.74,0 95.74,95.74,0,95.74,0,16.29}	1108.92	1024.75	{75,80,0,90,95,0},{75,0},{80}
10	{95.59,95.59,0 95.59,95.59,0,95.59,0,52.04,0}	1135.02	1092.00	{75,80,0,90,95,0},{90,0,100,0}
11	{96.65,96.65,0 96.65,96.65,0,0 ,96.65 ,96.65,0,200.10}	1714.97	1657.07	{87.86,92.86,0,102.86,107.86,0,0,122.86 ,127.86,0,137.86}
12	{96.67,96.67,0 96.67,96.67,0,0,96.67 ,96.67,0,200.01},{75}	1871.09	1813.32	{87.86,92.86,0,102.86,107.86,0,0,122.86 ,127.86,0,137.86},{75}

T	Q_T Heuristic WW	f_T	f_T	Q_T Heuristic GWW
1	{50}	125	125	{50}
2	{50},{100}	325	314.88	{72.5,77.5}
3	{50},{100},{0}	325	314.88	{72.5,77.5},{0}
4	{50},{100},{0},{70}	474	463.88	{72.5,77.5},{0},{70}
5	{50},{100},{0},{70},{80}	638	627.75	{72.5,77.5},{0},{70},{72.5,77.5}
6	{50},{100},{0},{70},{120,0}	722	701.50	{75,80,0,90,95,0}
7	{50},{100},{0},{80},{85,0}	814.75	804.46	{72.5,77.5},{0},{73.33,78.33,83.33,0}
8	{50},{100},{0},{80},{115,0,0}	880.75	863.46	{72.5,77.5},{0},{83.33,88.33,93.33,0,0}
9	{50},{100},{0},{80},{85,0},{110,0}	1043.75	1031.38	{72.5,77.5},{0},{77.5,82.5,87.5,0,97.5,0}
10	{50},{100},{0},{80},{115,0,0},{115,0}	1116.5	1098.88	{72.5,77.5},{0},{85,90,95,0,0,110,0}
11	{50},{100},{0},{80},{115,0,0},{115,0} ,{250}	1841.5	1680.79	{89.29,94.29,0,104.29,109.29,114.29,0,0 ,129.29,0,139.29}
12	{50},{100},{0},{80},{115,0,0},{115,0},{250},{75}	1997.75	1837.04	{89.29,94.29,0,104.29,109.29,114.29,0,0,129.29,0,139.29},{75}

Table 3.4: Illustrative example showing solutions of heuristics EOQ, $GEOQ$, WW, GWW ($w_t^1 = w = 0.01$, $h_t = h = 0.1$, $r_t^1 = r = 2$, $K_t = 100$ for all $t \in \{1, \ldots, T\}$.)

3.6 Numerical study

In this section, we present and discuss our findings in a numerical study.

For our numerical study, we considered a problem horizon of $T = 300$ periods. Period demands are generated randomly from normal distribution with mean $\mu \in \{50, 100, 200\}$ and standard deviation $\sigma(= 40)$; negative demand values have been replaced by zero demands. We denote the three demand patterns by $d1, d2$ and $d3$ respectively. All other system parameters are stationary. We set unit holding cost rate, $h_t = h = 1$ and setup cost is selected as a function of the mean demand rate, $K_t = K = [J^2/2]\mu$ where J may be viewed as a proxy for the average length of a production lot *if* production costs were linear as in the classical lot-sizing problem. We have $J \in \{0, 2, 3, 4, 5\}$ with $J = 0$ corresponding to zero setup cost. The production cost structure was chosen with $r_t^1 = r \geq 1$. This corresponds to the Cobb-Douglas type economic production function with convex costs. We selected $r \in \{1, 1.1, 1.5, 2.0, 2.2\}$; note that $r = 1$ corresponds to the classical lot sizing setting used as benchmark. To select the cost coefficient $w_t^1 = w$, we considered the variable cost of production per unit when a production quantity equals the average demand per period, $\bar{c}$ where $\bar{c} = [w\mu^r]/\mu = w\mu^{r-1}$. Then, letting $a = h/\bar{c}$, we have $w = h\mu/(a\mu^r)$ with $a \in \{0.02, 0.05, 0.1, 0.2\}$. Note that the resulting variable cost for a production quantity of q units is given by $[h\mu/a](q/\mu)^r$, and that, as a increases $\bar{c}$ decreases since we hold h equal to unity. Overall, our experimental set contains $120(= 5 \times 4 \times 6)$ *parameter instances* for each of the three levels of demand mean. For visual displays, we used a short-hand notation to denote an experiment instance, $d_iK_ja_nr_s$, where, for example, $d3K2a1r4$ corresponds to the parameter values, $\mu = 200$, $J = 2$, $a = 0.02$ and $r = 2.0$. For each particular experiment instance we generated 10 demand stream *replications*. The average fraction of *zero* demand values in the generated replication streams for each demand pattern is approximately 11%, 0.67% and 0%, respectively with resulting averages of 51.9, 100.4 and 200.7.

Next, we discuss our findings from our numerical study. We first report our findings on the relative performance of the proposed heuristics in comparison with the benchmark forward DP solution. Then, we provide our sensitivity analysis on the basis of the solutions obtained by the benchmark DP algorithm.

3.6.1 Comparison of heuristics

We conducted our numerical study to investigate the following: (i) percentage deviation from the benchmark minimum cost for each heuristic; (ii) dominance of heuristics among themselves; (iii) impact of the cost parameter values and demand patterns on performances of the heuristics. All heuristic comparisons have been performed on a static basis (when demands for the entire problem horizon are known at the beginning of the problem horizon.) For our numerical study, we considered the same experimental set described above. The rationale for this set has been the earlier performance studies for the classical problem setting; in particular, Simpson [32].

We use as the benchmark the best solution to problem (P) by means of the forward DP solution algorithm discussed above. The total cost over the problem horizon under a particular heuristic is denoted by $f_T^{(j)}$ and the best solution with the forward DP is denoted by $\tilde{f}_T$. For computing the total cost under a particular heuristic for a problem instance, we used the corresponding algorithm provided in the Algorithm 1. For each experiment instance, we measure the performance of heuristic j in terms of percentage deviations from $\tilde{f}_T$ as follows.

$$\Delta_j\% = \frac{f_T^{(j)} - \tilde{f}_T}{\tilde{f}_T} \times 100$$

We discuss the performance of heuristics *SM, LUC, EOQ, WW* and heuristics *GSM, GLUC, GEOQ, GWW, psdWW* separately. For each heuristic, we report (i) the minimum, maximum, median and average percentage deviations, and (ii) the number of instances for which zero or negative deviations have been obtained for three different demand variance levels across all 1000 experiment instances. A negative deviation implies that a *better* solution has been found by the heuristic than the forward DP algorithm. We begin our discussion of the heuristic performances with their overall behavior. In Table 3.5, we report the performance statistics for heuristics SM through WW for $1000(= 10 \times 100)$ experiment instances for each of the three different demand patterns. We see that as the demand variance decreases (from $d1$ to $d3$), percentage deviations also decrease for all heuristics. All heuristics have left-skewed performance distributions for all demand variance levels. Heuristic WW (that is, solving the problem in a forward DP algorithm while imposing demand integrality) turns out to be the best performer except for high variance levels on average. It is followed very closely by Heuristic EOQ and LUC. The high performance of heuristic WW is due to the fact that the problem is solved optimally albeit under the restriction of demand integrality.

Note that 611 $(= 124+(161+8+8)+(235+36+39))$ out of $3000(= 3 \times 1000)$ cases have resulted in zero deviations from the best solution with the forward DP (N(0) column in Table 3.5), implying that demand integrality was preserved in the best forward DP solution for such instances. For the remaining instances, the deviations obtained under heuristic WW may be viewed as the impact of not smoothing the production across successive periods within a generation. The second-best performance of heuristic EOQ points to the fundamental trade-offs captured by Harris's formula; and, being a single step heuristic, its performance is excellent. The number of instances for which this heuristic resulted in the same solution as the forward DP also increases as demand variance decreases, as expected. The Silver-Meal heuristic typically performs well in the classical setting. It is surprising that heuristic SM did not do as well following the others with a relatively large gap. In contrast, LUC performs very well since its structure considers average unit costs and prevents high production costs due to the convexity of the production costs.

As demand variance decreases, allowing for production smoothing seems to be counter-productive. This may explain the performance of heuristics EOQ and WW for d3; after all, both preserve demand integrality by construct. None of the heuristics in this group resulted in a solution better than the benchmark forward DP algorithm. (See N(-) in Table 3.5.)

Next, we look at sensitivity of heuristic performance with respect to the fixed setup cost K and the production cost nonlinearity measured through r. In Table 3.7, we present the minimum, maximum, median and average percentage deviations for the heuristics for different demand patterns and cost parameters. (See also Figures 3.3 and 3.4 for pictorial depictions.) The performances of the heuristics are roughly similar for different measures (minimum, maximum percentage deviations etc.); hence, we focus on the average percentage deviations. On average, the performances of all the heuristics deteriorate as demand variance increases. The performance also worsens as the nonlinearity in production cost (r) increases for moderate and high demand variances (d1 and d2). For low variance (d3), the performances of all heuristics lie within the 4% band with those of heuristics *SM* and the performance gap among the heuristics gets smaller for large r values. With respect to the fixed setup cost, the performances of heuristics *SM* and *WW* improve as K increases for all demand patterns. The performance of heuristic *EOQ* improves for moderate and high variance levels but worsens slightly for $d3$. Although, heuristic *WW* was deemed to be the best performer in general, it does not do so well in comparison with *EOQ* for large r values. As the production cost becomes more and more nonlinear, this heuristic starts to underperform especially as demand variance increases. This is due to the fact that heuristic *EOQ* allows for demand splitting while heuristic *WW* cannot smooth the production over successive periods. Heuristics *EOQ* and *WW* have similar performances for low and moderate values of fixed setup cost, but the former performs slightly better for large K values. Once again, this indicates that the variant of Harris's formula captures the fundamental trade-offs. Worst case performance is also of theoretical and practical interest. In terms of maximum percentage deviations, heuristic *SM* always results in the maximum percentage deviations. Heuristic *WW* clearly dominates *EOQ* and *LUC* for small nonlinearity in production cost but their performance gaps decrease as r and K get large. Although *WW* is the best performer in this group and may result in zero deviation from the benchmark solution in a large portion of the solutions, it is still important to point out that imposing demand integrality may cause very large deviations (up to 65%) in certain cases. The performance deteriorates rapidly with large production cost nonlinearity and low setup costs. As we discuss below for the G-class heuristics, exploiting the optimal generation structures for given production period decisions does improve the solutions significantly.

The last group of heuristics (G-class) results in significant improvements over the first group. In some cases, the average improvements are about twenty-fold. Actually, the large performance difference between the heuristic groups directly implies the importance of class G production plans in a solution and points to the impact of production cost non-linearity. In Table 3.6, we report the performance statistics for this group for $1000(= 10 \times 100)$ experiment instances for each of the three different demand patterns. Note that in some experiment instances, heuristics *GEOQ*, *GLUC* *GWW* obtained better solutions than the benchmark forward DP rendering minimum deviations negative. Also, the number of instances for which the forward DP solution was obtained increases with this group; $1188(= 157 + 223 + 808)$ instances for heuristic *GWW*, $701(= 65{+}636)$ instances for *GEOQ*, 114 instances for *GSM* and 604(=570+34) instances for *GLUC*, and 1321(=148+419+754) instances for *psdWW*.

Demand	Algorithm	min	max	median	average	N(0)	N(-)
d1	*SM*	1.23	89.23	17.97	24.43	0	0
	LUC	0.08	65.03	5.79	13.58	0	0
	EOQ	0.08	65.03	6.24	12.47	0	0
	WW	0	65.03	5.01	12.76	124	0
d2	*SM*	0.79	28.2	10.72	11	0	0
	LUC	0	19.90	2.30	4.58	8	0
	EOQ	0	19.9	3.24	4.71	8	0
	WW	0	19.9	1.58	4.16	161	0
d3	*SM*	0.5	7.59	3.18	3.33	0	0
	LUC	0	5.37	0.70	1.17	39	0
	EOQ	0	7.4	1.69	1.99	36	0
	WW	0	5.37	0.44	1.04	235	0

Table 3.5: Percentage deviation statistics for heuristics which preserve demand integrality.

Demand	Algorithm	min	max	median	average	N(0)	N(-)
d1	*GSM*	1.09	44.17	13.06	15.39	0	0
	GLUC	-0.41	18.87	3.13	4.53	0	31
	GEOQ	0.06	18.87	1.8	3.36	0	0
	GWW	-0.01	18.87	1.73	3.68	157	2
	psdWW	-0.35	8.38	0.20	0.68	148	36
d2	*GSM*	0.79	15.84	6.78	6.63	0	0
	GLUC	-.02	3.17	0.69	0.78	34	1
	GEOQ	-0.01	2.28	0.29	0.4	65	5
	GWW	-0.01	2.37	0.19	0.34	223	3
	psdWW	-0.01	3.01	0.02	0.26	419	30
d3	*GSM*	0	7.04	1.37	1.94	114	0
	GLUC	0	1.41	0	0.14	570	0
	GEOQ	0	1.29	0	0.17	636	0
	GWW	0	0.65	0	0.02	808	0
	psdWW	0	3.0	0.0	0.19	754	0

Table 3.6: Percentage deviation statistics for heuristics with G-class type of production subplans.

The number of zero percentage deviation solutions increases as the demand variance gets smaller. Heuristics *GLUC*, *GEOQ* and *GWW* perform almost equally well; the magnitude of deviations is small when they differ most for low demand variance. In Table 3.8, we present the performance statistics with respect to r and K. (See also Figures 3.3 and 3.4 for pictorial depictions.) With respect to nonlinearity in production cost, the behavior of the heuristics are not monotone. However, overall, there is a tendency for the performances to deteriorate as r gets large. Likewise, demand variance impacts the performances negatively. An interesting observation is that, with large demand variance, r impacts performances negatively whereas with low demand variance, all heuristics tend to converge to the benchmark. (See Figure 3.3 (f).) The effect of the fixed setup cost K is similar to that observed for the first group but with smaller deviations. For Heuristics *GEOQ* and *GWW*, the improvement of performance for smaller demand variances becomes more pronounced than those of their counterparts *EOQ* and *WW*. Lastly, we should mention the computational efficiency achieved by the proposed heuristics. As discussed above, the backward DP formulation which guarantees optimality is prohibitively slow and memory-inefficient for practical use. The computational times statistics measured in seconds for each heuristic and the benchmark forward DP algorithm for different demand patterns are presented in Table 3.9. As expected, the first four heuristics have similar and the smallest run times, and their G-class versions in the second four heuristics have relatively larger and similar times but are still reasonably fast. The benchmark forward DP algorithm solves on average in roughly four to six minutes. The statistics provided for the heuristics have been obtained for a 2.3GHz processor whereas those for the benchmark have been obtained for a 3.3GHz processor. To conclude, our numerical comparison of the heuristics reveals that (i) imposing demand integrality (using *WW*) may result in large deviations, especially in the presence of large production nonlinearities and low setup costs, (ii) a variant of Harris's formula that captures the fundamental trade-offs among setup costs, inventory holding costs and production costs provides a quick and reasonably good heuristic (*EOQ*), (iii) construction of G-class subplans is essential in developing heuristic solutions for the dynamic lot sizing problem in the presence of production cost nonlinearities, (iv) the computational time improvements through the proposed heuristics are justifiably significant, and finally (v) among all the proposed heuristics, *psdWW* almost always catches the benchmark with less computational time and among the others, *GWW* and, then, *GEOQ* and *GLUC* are, by far, the best ones.

		Production cost exponent levels					Setup cost levels				
Heuristic	statistic	r1	r2	r3	r4	r5	k1	k2	k3	k4	k5
SM	min	1.23	1.48	2.92	8.77	9.65	8.59	2.23	1.23	1.25	1.24
(d1)	max	12.99	25.76	40.07	74.28	89.23	89.23	83.29	79.88	74.21	68.06
	median	3.09	9.65	20.38	42.87	49.27	30.19	21.94	18.83	14.14	9.65
	average	*4.67*	*10.5*	*19.14*	*40.31*	*47.47*	*36.36*	*27.95*	*23.06*	*18.96*	*15.75*
LUC	min	0.08	0.64	1.47	3.99	5.24	0.08	0.49	0.57	0.45	0.31
	max	7.54	5.31	17.54	48.4	65.03	65.03	58.48	55.18	50.97	46.21
	median	1.16	2.05	6.41	25.78	33.46	11.39	7.15	5.47	5.19	4.51
	average	*1.8*	*2.38*	*7.01*	*24.48*	*32.16*	*19.11*	*15.16*	*13.08*	*11.08*	*9.4*
EOQ	min	0.08	0.51	1	1.6	1.64	0.08	0.3	0.25	0.4	0.65
	max	11.81	9.41	17.54	48.4	65.03	65.03	57.72	53.51	48.26	41.64
	median	2.12	2.87	4.98	23.07	29.91	11.39	7.66	6.28	4.94	4.46
	average	*3.61*	*3.3*	*6.14*	*21.21*	*28.03*	*19.19*	*13.98*	*11.7*	*9.6*	*7.83*
WW	min	0	0	0.03	2.26	3.49	0.08	0	0	0	0
	max	1.94	6.86	17.54	48.4	65.03	65.03	58.47	55.12	50.89	46.09
	median	0.01	1.08	6.03	25.47	33.35	11.39	7.01	5.29	3.75	2.15
	average	*0.25*	*1.68*	*6.28*	*23.91*	*31.6*	*19.19*	*14.64*	*12.21*	*9.85*	*7.83*
SM	min	0.79	1.12	2.05	4.74	5.15	8.06	2.17	0.93	0.79	0.97
(d2)	max	11.15	17.48	21.35	25.67	28.2	28.2	28.03	27.56	25.63	24.95
	median	2.42	7.94	11.72	15.56	16.99	14.8	12.58	10.14	7.67	6.19
	average	*4.14*	*7.74*	*11.29*	*14.9*	*16.91*	*15.49*	*12.73*	*10.56*	*8.77*	*7.43*
LUC	min	0	0.14	0.46	1.84	2.5	0	0.17	0.21	0.21	0.15
	max	3.17	2.33	5.06	14.76	19.9	19.9	19.14	18.31	17.28	15.99
	median	0.48	0.91	2.08	8.08	10.92	3.05	2.57	2.42	2.19	1.94
	average	*0.83*	*0.97*	*2.37*	*7.88*	*10.83*	*5.73*	*5.15*	*4.58*	*3.98*	*3.44*
EOQ	min	0	0.14	0.52	0.75	0.94	0	0.19	0.12	0.51	0.18
	max	8.86	6.66	5.06	14.76	19.9	19.9	18.91	17.78	16.09	14.19
	median	2.09	1.3	2.51	7.1	10.08	3.05	3.65	3.77	3.12	2.46
	average	*2.78*	*2.12*	*2.5*	*6.79*	*9.34*	*5.73*	*5.26*	*4.79*	*4.16*	*3.57*
WW	min	0	0	0	0.77	1.41	0	0	0	0	0
	max	0.44	1.82	5.06	14.76	19.9	19.9	19.14	18.29	17.2	15.9
	median	0	0.27	1.92	7.99	10.85	3.05	2.48	1.85	1.24	0.8
	average	*0.06*	*0.5*	*2.04*	*7.6*	*10.56*	*5.73*	*4.85*	*4.11*	*3.36*	*2.71*
SM	min	0.5	0.69	0.88	0.79	1.06	0.79	0.83	0.71	0.58	0.5
(d3)	max	7.32	7.59	6.2	5.05	6.36	7.32	7.02	7.59	6.83	5.58
	median	1.92	4.33	3.81	2.81	3.19	3.89	3.59	3.01	2.8	2.76
	average	*2.92*	*4.07*	*3.57*	*2.83*	*3.27*	*4.01*	*3.74*	*3.25*	*2.94*	*2.72*
LUC	min	0	0	0.08	0.51	0.93	0	0	0.03	0.06	0.02
	max	1.41	0.88	1.26	3.94	5.37	5.37	5.17	4.94	4.64	4.31
	median	0.18	0.22	0.57	1.85	2.65	0.53	0.59	0.72	0.8	0.8
	average	*0.3*	*0.25*	*0.58*	*1.94*	*2.79*	*1.28*	*1.2*	*1.15*	*1.14*	*1.1*
EOQ	min	0	0	0.08	0.46	0.41	0	0.02	0.03	0.17	0.09
	max	7.4	6.43	4.29	3.94	5.37	5.37	7.4	7.07	6.58	6.7
	median	1.84	0.3	0.77	2.18	2.55	0.53	1.27	2.05	1.91	1.69
	average	*2.38*	*1.72*	*1.18*	*2.04*	*2.61*	*1.28*	*1.94*	*2.25*	*2.31*	*2.15*
WW	min	0	0	0	0.36	0.81	0	0	0	0	0
	max	0.09	0.39	1.26	3.94	5.37	5.37	5.17	4.94	4.64	4.31
	median	0	0.06	0.45	1.79	2.6	0.53	0.47	0.36	0.47	0.34
	average	*0.01*	*0.11*	*0.48*	*1.87*	*2.74*	*1.28*	*1.13*	*1.01*	*0.93*	*0.84*

Table 3.7: Percentage deviation statistics of single-step heuristics.

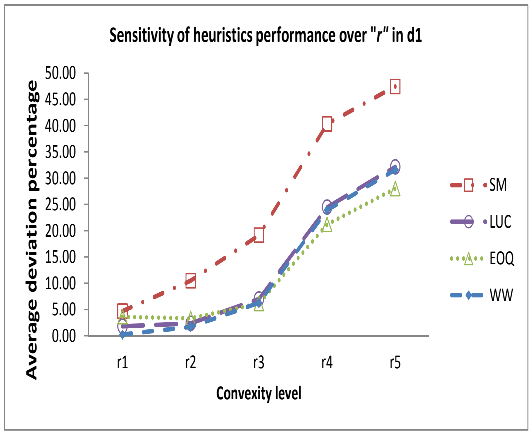
		Production cost exponent levels					Setup cost levels				
Heuristic	statistic	r1	r2	r3	r4	r5	k1	k2	k3	k4	k5
GSM	min	1.09	1.32	1.93	3.5	4.04	7.98	2.22	1.23	1.13	1.09
(d1)	max	12.94	22.12	31.05	41.25	44.17	44.17	40.22	38.75	35.25	32.28
	median	2.88	8.35	15.64	26.45	26.33	24.84	17.02	12.16	7.59	4.88
	average	*4.47*	*9.23*	*14.87*	*23.83*	*24.52*	*24.59*	*17.92*	*14.23*	*11.23*	*8.95*
GLUC	min	0.07	-0.41	0.96	1.97	2.12	0.07	-0.41	-0.26	-0.23	-0.41
	max	7.54	5.11	7.15	15.29	18.87	18.87	15.53	15.07	14.43	13.34
	median	1.04	1.18	2.67	8.15	9.16	4.65	2.74	3.13	3.03	2.83
	average	*1.67*	*1.38*	*2.99*	*7.82*	*8.79*	*5.68*	*4.54*	*4.45*	*4.15*	*3.83*
GEOQ	min	0.06	0.08	0.24	0.3	0.45	0.07	0.06	0.06	0.07	0.13
	max	5.22	3.3	7.15	15.29	18.87	18.87	15.08	13.76	12.78	10.82
	median	0.64	0.84	1.41	6.06	6.66	4.65	1.44	1.61	1.5	1.33
	average	*1.02*	*1.05*	*2.07*	*5.88*	*6.77*	*5.75*	*3.37*	*2.92*	*2.56*	*2.18*
GWW	min	-0.01	0	0	0.47	0.75	0.07	0	-0.01	-0.01	0
	max	1.11	3.3	7.15	15.29	18.87	18.87	15.47	15.05	13.77	13.45
	median	0	0.36	1.89	7.8	8.77	4.65	1.79	1.78	1.22	0.73
	average	*0.13*	*0.68*	*2.24*	*7.21*	*8.12*	*5.75*	*3.78*	*3.44*	*2.95*	*2.43*
GSM	min	0.79	1.09	1.28	1.89	1.95	4.43	2.17	0.93	0.79	0.95
(d2)	max	11.15	15.84	15.3	12.88	11.46	15.84	14.35	14.58	13.46	11.5
	median	2.29	6.43	9.19	7.53	6.68	10.02	7.89	6.18	4.17	2.76
	average	*3.98*	*6.76*	*8.38*	*7.29*	*6.73*	*10.15*	*7.96*	*6.25*	*4.87*	*3.91*
GLUC	min	0	0	0	0	-0.02	0	-0.02	0.09	0.16	0.14
	max	3.17	2.33	1.71	2.39	2.87	1.42	1.85	2.54	3.14	3.17
	median	0.48	0.38	0.53	1	1.09	0.15	0.59	0.81	0.96	0.99
	average	*0.78*	*0.54*	*0.54*	*0.96*	*1.07*	*0.27*	*0.64*	*0.88*	*1.03*	*1.08*
GEOQ	min	0	0	0	0	-0.01	0	-0.01	-0.01	-0.01	-0.01
	max	2.28	1.68	1.21	1.17	1.56	1.56	1.24	1.96	2.16	2.28
	median	0.38	0.13	0.21	0.32	0.35	0.21	0.17	0.25	0.35	0.38
	average	*0.55*	*0.36*	*0.31*	*0.37*	*0.42*	*0.33*	*0.27*	*0.4*	*0.48*	*0.54*
GWW	min	0	0	-0.01	0	-0.01	0	-0.01	0	0	-0.01
	max	0.08	0.29	0.89	2	2.37	1.56	1.32	1.61	2.24	2.37
	median	0	0.05	0.2	0.56	0.69	0.21	0.17	0.19	0.21	0.17
	average	*0.01*	*0.07*	*0.24*	*0.63*	*0.75*	*0.33*	*0.27*	*0.35*	*0.37*	*0.37*
GSM	min	0.5	0.69	0.7	0	0	0	0	0	0	0
(d3)	max	7.04	6.19	4.32	2.1	1.54	7.04	6.13	6.19	5.35	4.04
	median	1.83	3.89	2.47	0.54	0.25	1.86	2.57	1.6	1.24	1.13
	average	*2.79*	*3.47*	*2.43*	*0.65*	*0.36*	*2.58*	*2.38*	*1.9*	*1.54*	*1.32*
GLUC	min	0	0	0	0	0	0	0	0	0	0
	max	1.41	0.88	0.74	1.14	0.96	0	0.79	1.17	1.27	1.41
	median	0.17	0.03	0	0	0	0	0	0	0.13	0.17
	average	*0.29*	*0.15*	*0.11*	*0.11*	*0.06*	*0*	*0.06*	*0.14*	*0.22*	*0.29*
GEOQ	min	0	0	0	0	0	0	0	0	0	0
	max	1.29	1.24	1.03	0.76	0.51	0.04	0.99	1.16	1.09	1.29
	median	0.26	0	0	0	0	0	0	0	0.16	0.27
	average	*0.32*	*0.26*	*0.16*	*0.08*	*0.04*	*0*	*0.09*	*0.2*	*0.26*	*0.31*
GWW	min	0	0	0	0	0	0	0	0	0	0
	max	0.04	0.09	0.19	0.5	0.65	0.04	0.02	0.12	0.5	0.65
	median	0	0	0	0	0	0	0	0	0	0
	average	*0*	*0.01*	*0.02*	*0.05*	*0.04*	*0*	*0*	*0.01*	*0.04*	*0.07*

Table 3.8: Percentage deviation statistics of two-step heuristics.

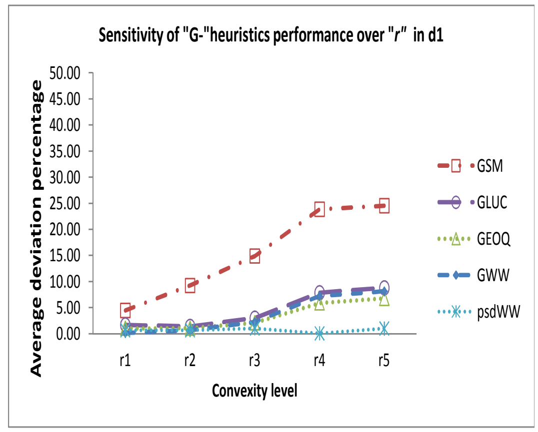
		SM	*LUC*	*EOQ*	*WW*	*GSM*	*GLUC*	*GEOQ*	*GWW*	*psdWW*	DP
	min	0	0	0	0.07	0.16	0.6	0.36	0.21	0.27	0
d1	max	0.02	0	0.02	0.34	13.48	4.28	17.68	13.54	101.48	359.64
	median	0	0	0	0.08	0.47	1.66	1.77	0.64	6.87	254.26
	average	0	0	0	0.1	1.03	1.56	2.39	1.27	25.43	207.87
	min	0	0	0	0.08	0.16	0.47	0.38	0.23	0.23	0
d2	max	0.02	0	0.02	0.36	13.7	4.38	8.69	7.43	151.55	329.89
	median	0	0	0	0.09	0.56	1.97	1.62	0.69	6.05	284.87
	average	0	0	0	0.11	0.92	1.8	1.79	1.01	22.91	230.56
	min	0	0	0	0.08	0.16	0.38	0.39	0.24	0.19	0
d3	max	0.03	0.02	0.02	0.26	3.32	3.7	3.86	4.85	92.59	373.74
	median	0	0	0	0.08	0.53	1.64	1.78	0.7	4.73	254.31
	average	0	0	0	0.09	0.61	1.51	1.67	0.79	22.25	240.08

Table 3.9: Execution time statistics for the entire experiment set measured in seconds.

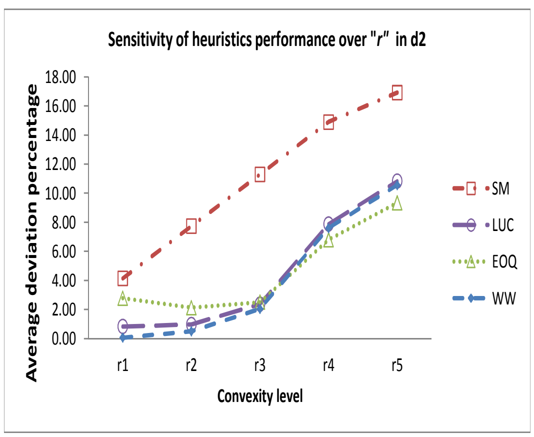
('0' indicates run time less than 1 millisecond.)



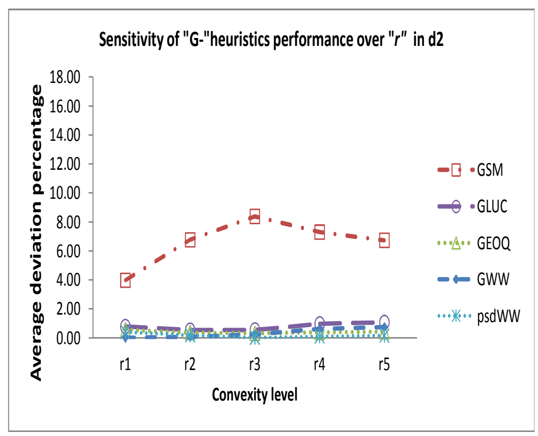
(a)



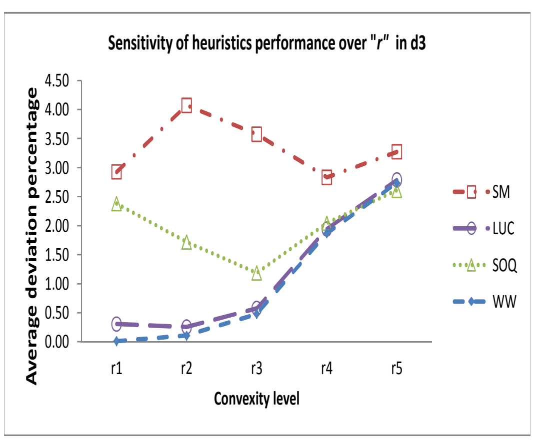
(b)



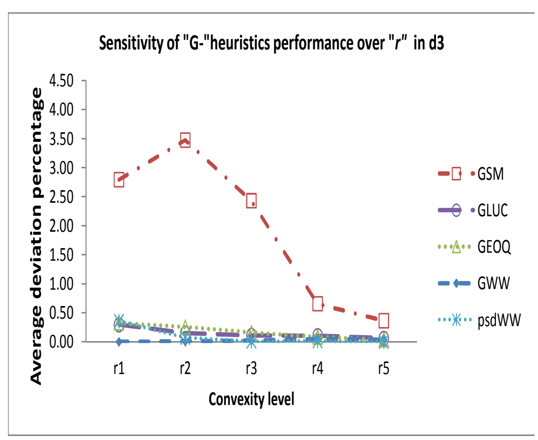
(c)



(d)

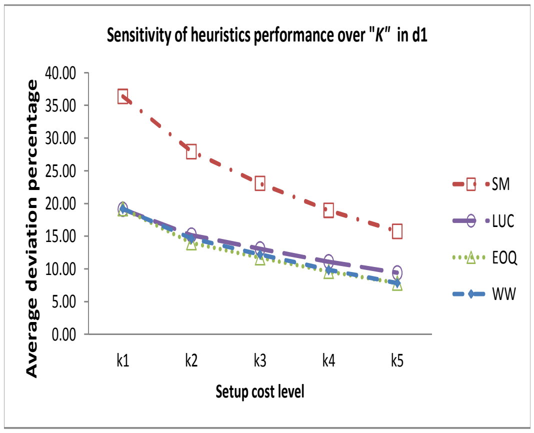


(e)

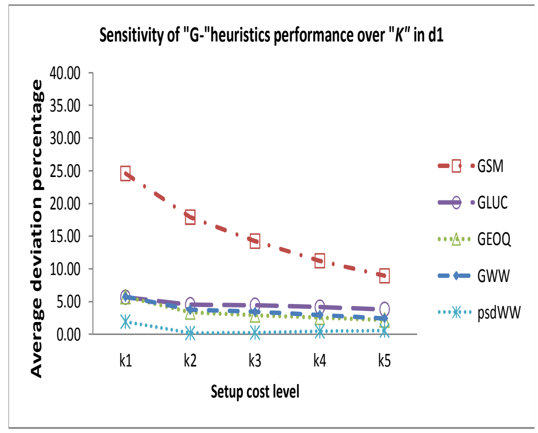


(f)

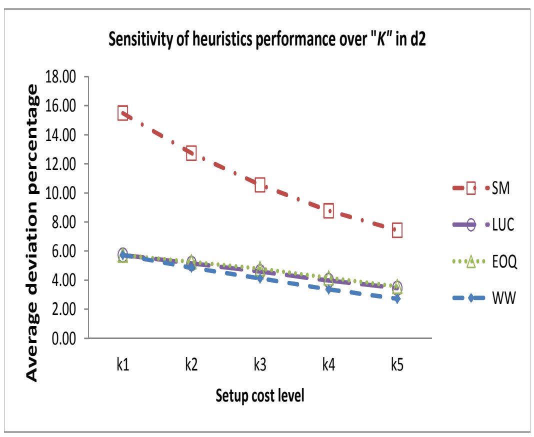
Figure 3.3: Average percentage deviation of heuristics versus production cost convexity levels for d1, d2 and d3.



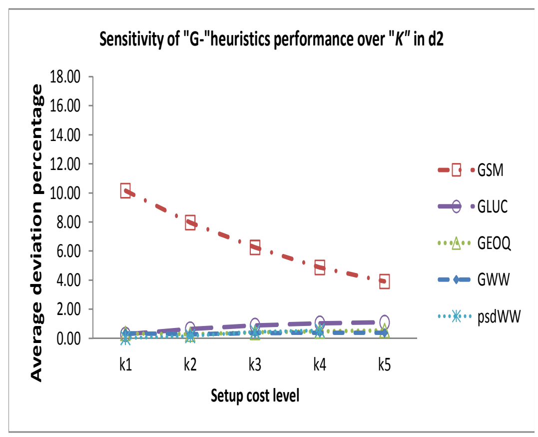
(a)



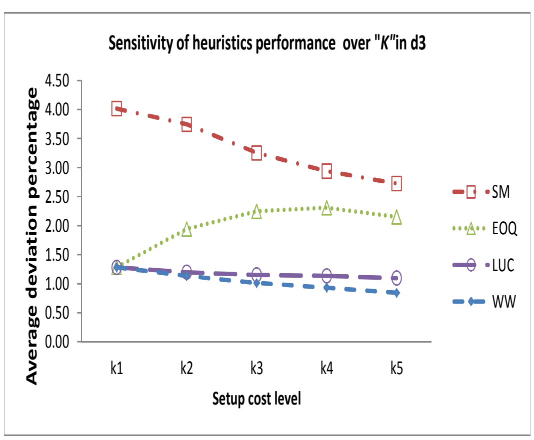
(b)



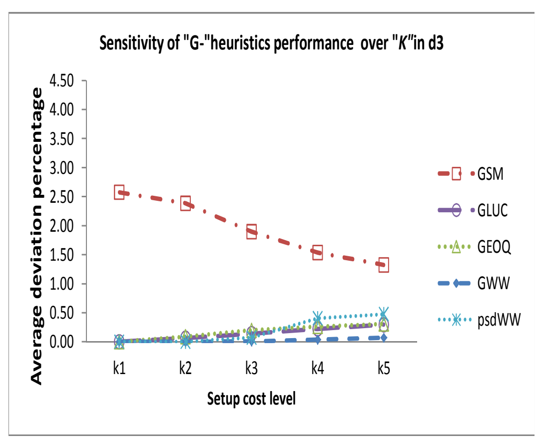
(c)



(d)



(e)



(f)

Figure 3.4: Average percentage deviation of heuristics versus setup cost levels for $d1$, $d2$ and $d3$.

3.6.2 Sensitivity analysis

We study the impact of experiment parameters through *(i)* the average minimum total cost over the horizon, *(ii)* the average number of generations over the problem horizon, *(iii)* the average percentage of production periods, *(iv)* the average percentage of generations of certain types. All findings are based on the benchmark solutions.

Average minimum total cost

The minimum total costs behave as expected. (See Figure 3.5 (a)-(b) for an illustration of costs averaged over replications and all r and a values and selected two K, values.) As K decreases and a increases, the minimum total cost decreases. The impact of non-linearity in variable production costs (measured through r) becomes more pronounced with larger K. The parameters a and r interact in the variable production cost component. That is, the same cost can be obtained by distinct pairs of a and r; smaller a and r in one, and larger a and r in the other. Furthermore, the average total cost is not monotone in r.

Average number of generations

In Figure 3.5 (c)-(d), we illustrate the number of generations in the optimal solution for $P_{1,T}$ averaged over the ten replications for all r and a values and $K = \{K1, K5\}$. We begin our discussion with zero setup costs. For $K1$ and linear variable production cost ($r0$), the Lot-for-Lot (LFL) solution is obtained. In this case, the solution is insensitive to a values as total production cost over the horizon is constant, $(h/a)D_{1,T}$. As r increases, the solution deviates from LFL, splitting some production quantities over a number of periods in order to exploit the marginal cost structure. This smoothes out the production plan resulting in longer generations on average. As demand's coefficient of variation increases, this effect increases. Increasing a, which decreases cost for a production quantity allowing for larger production quantities at the same cost, counters the advantages of smoothing out. As setup costs increase, batching effects dominate extending the generations lengths resulting in smaller average number of generations in the solution. (Compare overall tendencies for $K1$ and $K5$.) However, the interaction between the setup cost and the variable production cost component is not simple. As K increases, on one hand, it is desirable to increase the quantity q_t in a production period to compensate for the fixed cost; on the other hand, increasing quantities result in a tendency to split/smooth out production as observed before.

The percentage of production periods

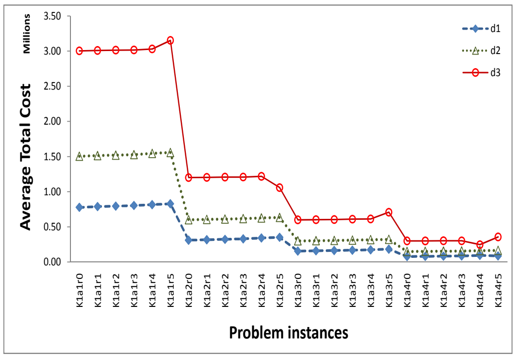
The percentage of production periods is of interest from a number of perspectives: It stands for the number of setups needed over a problem horizon; it can be used as a proxy for the average quantity produced over a problem horizon; and finally, it is a measure for utilization of production assets which is of managerial concern. For zero setup costs and $r0$, there is production only in periods with non-zero demands. As $r(> 1)$ increases, it becomes more advantageous to do a new setup due to increasing marginal production costs. Thus, smoothing out of production occurs resulting in larger percentages of production periods. For positive setup costs, batching effects emerge reducing the percentage but the effect of r is as for zero setup costs. For a given $r(> 1)$, the percentage of production periods decreases as a increases. This arises from the fact that lower a values allow for producing in larger quantities at the same cost levels. The effect of a becomes more pronounced with increasing fixed setup costs, as K exacerbates batching effects. As the demand patterns change from $d1$ to $d3$, the percentage of production periods increases. This is due to two reasons. First, the average demand per period μ increases, rendering a new setup more economical to do than to incur increasing marginal production costs. Second, the variability of demand (measured either as fraction of zeros or coefficient of variation) decreases that allows for smoother production plans.

Percentage of different types of production sequences

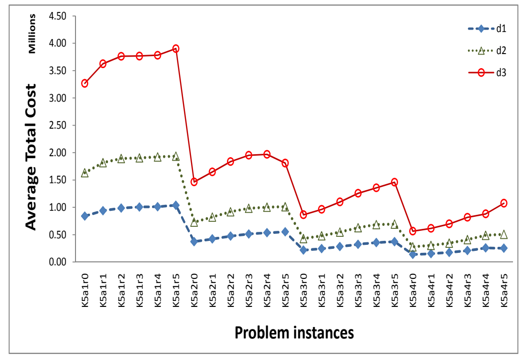
In our problem, a production sequence may have four forms: *(i)* A single period in which no production is done (a generation with zero demands). *(ii)* A single period in which the production equals that period's demand (LFL). *(iii)* A single production period followed by a number of no-production periods; the production quantity covers the total demand in the generation (Wagner-Whitin type). Finally, *(iv)* a combination of production and no-production periods, to which we refer as non-Wagner-Whitin type. In regards to the occurrence of these types, the overall findings can be summarized as follows. As fixed setup cost K increases, production quantities tend to increase to compensate for the fixed cost. This results in longer generations as more demands of future periods are covered by production in a particular period. As r increases, the tendency to split production over periods (production smoothing) increases. This also results in longer generations on average. Finally, as the ratio of holding cost to mean production cost per unit a increases, production quantities tend to increase since larger lots can be produced at the same cost. The result is again longer generations on average. The two fundamental tendencies- batching and production smoothing- manifest themselves in the production sequence types. When batching effect is more dominant, more LFL or Wagner-Whitin type sequences occur in the optimal production plan. When production smoothing effect is more dominant, the optimal production plan consists of more non-Wagner-Whitin type production sequences of class G.

We report the percentage of LFL production sequences over all generations in the optimal solution for $P_{1,T}$ averaged over the ten replications for all r and a values and selected two K values in Figure 3.5(e)-(f). For zero setup costs and $r0$, all production sequences are LFL, as expected. As $r(> 1)$ increases, production smoothing tendency increases and generations with class G production sequences emerge resulting in decreasing LFL percentages. As a increases, variable production cost decreases allowing for larger production quantities for a given $r(> 1)$ value. Hence, we see non-decreasing LFL percentages for a given $r(> 1)$ as a increases. As K increases, it is desirable to increase the quantity q_t in a production period to compensate for the fixed cost; hence, LFL structure becomes less desirable resulting in lower percentages of LFL generations. Likewise, increasing quantities result in a tendency to split/smooth out production as observed before. The confounded effect can easily be seen throughout the Figure 3.5. Also, we see a striking increase in the cases with zero LFL generations. For zero setup costs, percentage of LFL generations increases as demand pattern changes from

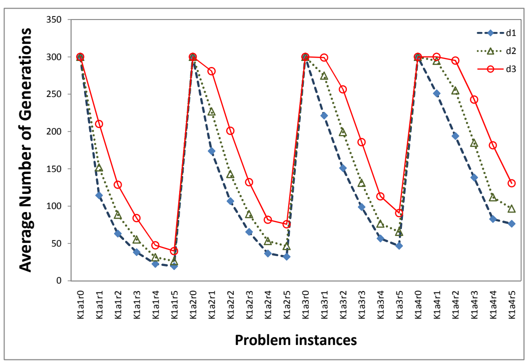
$d1$ to $d2$ to $d3$; but, the same does not hold for positive K. We observe wider ranges with respect to $r(> 1)$ as demand pattern changes from $d1$ to $d3$. The same behavior appears also as a increases, especially for large setup costs ($K5$).



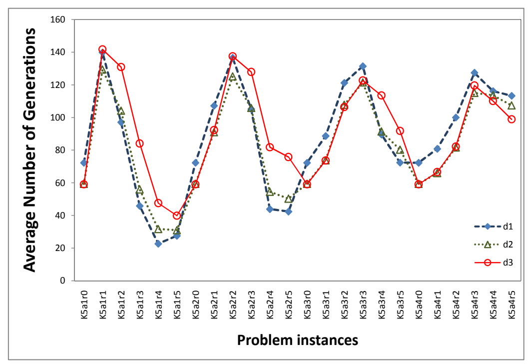
(a)



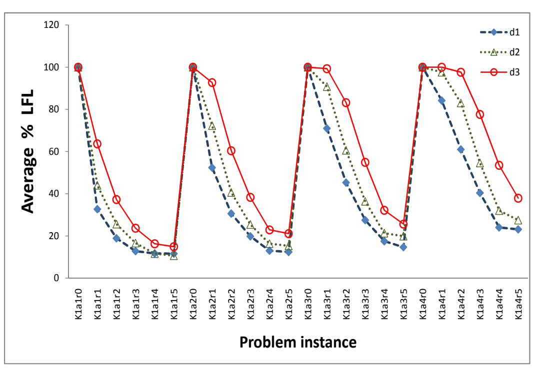
(b)



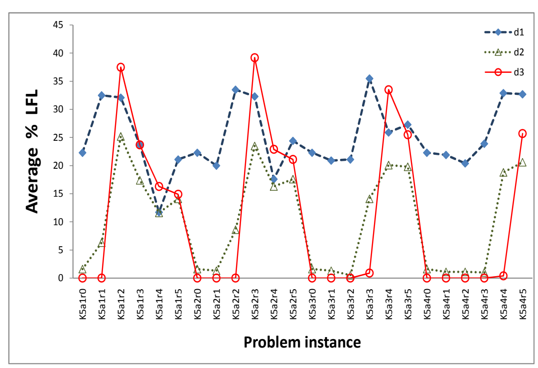
(c)



(d)



(e)



(f)

Figure 3.5: Impact of system parameters on average total cost ((a),(b)), average number of generations ((c),(d)), and average percentage of LFL generations ((e),(f)) for $K1$ and $K5$ over ten replications.

3.7 Conclusion

In this chapter, we considered the dynamic lot-sizing problem for a single item with deterministic demands where the relation between inputs and production quantities follows a Cobb-Douglas production function. We reformulate the problem into the form of the lot sizing model with nonlinear production cost function. If the total output elasticities is greater than one then the economies of scale in the production costs will hold and the existing classical exact algorithm will solve the problem optimally. We concentrated on the case in which total output elasticities are less than one and consequently we confront with a convex cost power-like production cost function in our lot sizing model for which economies of scale does not hold anymore.

The optimal solution with convex production functions exhibits behavior dissimilar to that in the classical lot-sizing problem. In particular, it is now possible to produce in a period even if its net demand is zero, and to produce part of the demand of a production period at an earlier time. We characterized the structure of the optimal policy for zero and positive setup costs. We provided further results that enabled us to develop a forward DP algorithm which guarantees optimality with $O(T^2 2^T)$ run time complexity for the problem with horizon length T. For the case of zero setup costs, it reduces to $O(T^2)$ complexity. The fundamental property of the optimal solutions is that production subplans exhibit a specific structure - herein referred to as G-class subplans. This property is retained for positive setup cost, as well. Based on this property, a version of the forward DP algorithm for positive setup costs was developed that employs three simple rules conjunctively to generate production sequences has computational complexity of $O(T^4)$ and it performed well in numerical tests. Another simpler version is also employed with less complicated rule and complexity of $O(T^3)$. In the latter one, the production periods take place in the first few beginning periods of the subplans rather than a single period in Wagner-Whitin algorithm and due to this dissimilarity we called this forward recursion pseudo-Wagner-Whitin.

We also proposed eight heuristics to solve the problem. The first four of them are single step heuristics. The remainder are two-step heuristics that improve on an initial solution obtained by the first group. They exploit the G-class production subplans and outperform the first group significantly. The best single step heuristic is a variant of the Wagner-Whitin solution algorithm in which demand integrality is imposed on the production quantities. The best G-class heuristic is its improved version. The computational time improvements through the proposed heuristics are justifiably significant. In our numerical study, we also investigated the sensitivity of the optimal production plans to the non-linearity in production functions, the average unit cost of production and setup costs. Our findings revealed the fundamental trade-offs between the batching and production smoothing tendencies. Production smoothing that is observed in our problem has important managerial implications. In an ERP (enterprise resources planning) setting, same level resource requirements such as labor are also smoothed out. Also higher level requirements become less lumpy (more uniform) allowing for simpler and steady delivery schedules and possible cost advantages. Capacity utilization rates also increase. Although we have considered an uncapacitated problem, production smoothing also has a positive impact on investment needs as capacity requirements per period are reduced. In the next chapter we discuss this problem from mathematical programming respects and discuss the performance of formulations and different optimization packages.

Chapter 4

Formulation comparison of the lot sizing problem with nonlinear convex costs

In this chapter we extend the lot sizing problem obtained in the previous chapter from two aspects. We assume that in addition to the production cost, the holding cost may follow a nonlinear convex form. Also it is assumed that the product is deteriorating. The motivation behind this assumption arises from application of the Cobb-Douglas production function in agriculture sector as discussed in Chapter 2. Following the classification of linear models, we provide the AGG, NIF and FAL reformulations of the MINLP model and their conic quadratic counterparts. We employ a previously proposed perspective cut and introduce two new cuts based on the structure of the optimal solution to strengthen the reformulations. We investigate the solution performance of the resulting 18 reformulations in a numerical study on commercial solvers.

The lot sizing problem with linear production and holding costs has been investigated extensively from the aspects of both problem construction and solution methodology. The existing mathematical programming models of the classical (uncapacitated) lot sizing problem have been classified into four formulation types bt Brahimi et al. [1]: *(i)* Formulation with aggregated production quantity variables, (the AGG model), *(ii)* formulation without the inventory variables (the *NIF* model), *(iii)* uncapacitated facility location type formulation (the FAL model), and *(iv)* the shortest path formulation (the *SHP* model). Reformulations of an optimization problem are desirable as they may exploit certain structural properties of the optimal solutions or may be more amenable to available optimization techniques. The FAL formulation is known to be strong and the LP-relaxation gives integer valued solutions for the setup variables. Günlük and Linderoth[72] point out the research gap in this regard. They state that such reformulations are not available for the nonlinear lot sizing problem and would be of great use if their particular efficacy were established. They even envision the possibility of commercial solvers automatically converting a given problem with non-linear costs into a stronger reformulation as in the case of the FAL reformulation of linear lot sizing problems.

Here, we attempt to further fill this gap. In particular, we consider production and holding costs that may involve power function terms of the production quantities and/or inventories, and decaying inventories. Following the classification in [1], we provide the AGG, *NIF* and FAL reformulations of the MINLP model and their conic quadratic counterparts. Furthermore, we introduce the perspective cut proposed by Aktürk et al. [48] and two new optimality cuts based on the structure of the optimal solution to strengthen the reformulations. We investigate the solution performance of the resulting 18 reformulations for the non-linear dynamic lot sizing problem.

The remainder of this chapter is organized as follows: in Section 4.1 we introduce the lot sizing problem and provide our base MINLP formulation. We discuss and provide the alternative MINLP reformulations in Section 4.2, and their conic quadratic counterparts in Section 4.3. In Section 4.4 we provide some analytical results and propose ways to strengthen the models. Section 4.5 contains the discussion of our numerical study and findings.

4.1 Basic model and formulation

We consider a lot sizing problem with discrete time periods and a planning horizon of T units, with deterministic demands $d_t, t = 1, \ldots, T$. A fixed setup cost K_t, is incurred at period t if production takes place in that period. Holding and production costs are incurred according to non-decreasing, continuous and convex power functions $H_T(I_t) \propto I_t^{\eta_t/\theta_t}$ and $P_t(X_t) \propto X_t^{\alpha_t/\beta_t}$, respectively where I_t denotes the inventory level at the end of period t and X_t is the production amount in period t; $I_t, X_t \in \mathbb{R}_+ \quad \forall t$. The parameters $\eta_t, \theta_t, \alpha_t, \beta_t \in \mathbb{N} \quad \forall t$ and determine the degree of nonlinearity in the holding and production cost functions. Holding costs are assumed to be independent of the age of the items in stock. All demands are satisfied without backlogging. Items in stock decay at a rate of δ_t for $t, t \geq 1$. That is, δ_t fraction of the ending inventory of the period t becomes unusable before it is transferred to period $t+1$ and a decay cost of π_t is incurred for each unit decaying in period t. The objective is to minimize the total cost over the problem horizon. The optimization problem is formally stated as follows.

$$\text{(AGG0)} \quad \min \sum_{t=1}^{T} \Big[K_t y_t + H_t(I_t) + P_t(X_t) + \pi_t \delta_t I_t \Big] \tag{4.1a}$$

s.t.

$$I_0 = 0 \tag{4.1b}$$

$$(1 - \delta_{t-1}) I_{t-1} + X_t - d_t = I_t, \qquad t = 1, ..., T \tag{4.1c}$$

$$X_t \leq M y_t, \qquad t = 1, ..., T \tag{4.1d}$$

$$I_t \geq 0, \qquad t = 1, ..., T \tag{4.1e}$$

$$y_t \in \{0, 1\}, \qquad t = 1, ..., T \tag{4.1f}$$

$$X_t \geq 0, \qquad t = 1, ..., T \tag{4.1g}$$

where y_t denotes the binary setup variable which takes on the value of one if $X_t > 0$ and zero, otherwise, and M is a sufficiently big number. Constraint (4.1b) gives the initial inventory and (4.1c) gives the inventory balance equation between consecutive periods after losing (δ_{t-1}) fraction of the inventory in transition from period $t-1$ to t. Constraint (4.1d) is for setup detection, and the rest indicate the variable types. Note that the model above is a mixed integer non-linear programming (MINLP) model. The formulation above has $2T$ continuous and T binary variables with $2T+1$ constraints.

Following the classification of Brahimi et al. [1], the original problem statement above corresponds to the AGG formulation of the nonlinear problem at hand. In the following sections, we provide and discuss alternative reformulations of the original problem.

4.2 MINLP reformulations

The first reformulation rests on eliminating the inventory variables: The recursive relationship for inventory balance (4.1c) and the initial inventory level in (4.1b) combine to yield $I_t = \sum_{i=1}^{t} a_{i,t}\big(X_i - d_i\big)$ for $t = 1, 2, \ldots, T$ while I_0 drops from the formulation, where $a_{i,t} = \prod_{j=i}^{t-1}(1-\delta_j)$ denotes the proportion of the product remaining from period i that is usable in t with $a_{i,t} = 1$ for $i = t$. Substituting this expression in place of inventory in the rest of the problem statement eliminates all inventory variables and replaces the non-negativity constraint on inventory levels with $\sum_{i=1}^{t} a_{i,t}X_i \geq \sum_{i=1}^{t} a_{i,t}d_i$ for $t = 1, 2, \ldots, T$. Unless stated otherwise, we retain all other notations in the reformulations. The resulting model is the *NIF0* reformulation of the lot sizing problem with non-linear costs. It has fewer decision variables and is beneficial when the nonlinearity in the model arises only from the production costs. However, when the holding cost function is non-linear, the *NIF0* reformulation results in an objective function with a more complicated form so that the commercially available generic solvers typically fail. In our preliminary investigations, all of the commercial solvers that we tested failed to solve this reformulation for a subset of our numerical study set with holding costs represented via power functions. This difficulty was overcome when the inventory variables were retained as auxiliary variables in the formulation in the arguments of the cost functions. For consistency in comparing reformulations with and without non-linearity in holding costs, we opted to use in our study this modified reformulation labeled as *MNIF0*.

It is formally stated as follows.

$$
\begin{aligned}
\text{(MNIF0)} \quad & \min \sum_{t=1}^{T} \left[K_t y_t + H_t(I_t) + P_t(x_t) + \pi_t \delta_t I_t\right] && && \text{(4.2a)} \\
& \text{s.t.} \\
& I_t = \sum_{i=1}^{t} a_{i,t}\big(x_i - d_i\big), && t = 1, ..., T && \text{(4.2b)} \\
& \sum_{i=1}^{t} a_{i,t} x_i \geq \sum_{i=1}^{t} a_{i,t} d_i, && t = 1, ..., T && \text{(4.2c)} \\
& x_t \leq M_t y_t, && t = 1, ..., T && \text{(4.2d)} \\
& x_t \geq 0, \quad y_t \in \{0, 1\}, && t = 1, ..., T. && \text{(4.2e)}
\end{aligned}
$$

4.2.1 Facility Location problem based formulation (FAL)

Krarup et al. [73] propose a reformulation of the lot sizing problem with linear costs and non-decaying items as an uncapacitated facility location problem - the *FAL* reformulation. It is known to be stronger than others, and solving the LP relaxation of the problem results in integer values for the binary setup variables. The reformulation rests on strengthening the *NIF0* model further by disaggregating the production quantity in a period into the assigned amounts to be used to satisfy demand in a particular future period. We adopt the same approach for our second reformulation of the non-linear lot sizing problem. We define x_{tj} as a new non-negative decision variable denoting the amount produced in period t to be used to satisfy the demand in period j,$(j \geq t)$ so that X_t in the *AGG0*, *NIF0* and *MNIF0* reformulations above is given by $\sum_{j=t}^{T} x_{tj}$. The new variables x_{tj}

are more informative than their aggregates, X_t. Thus, demand satisfaction constraints can be strict equalities in this reformulation. When all x_t are replaced by the new variables, we obtain the *FAL0* reformulation of the original non-linear problem. However, due to similar difficulties in solver performance as in the *NIF0* reformulation, we opted to retain x_t as an auxiliary variable in the arguments of non-linear production cost functions. This reformulation has $2T + \frac{T(T+1)}{2}$ continuous variables, T binary variables, and $3T + \frac{T(T+1)}{2}$ constraints. The resulting modified problem statement is the *MFAL0* reformulation and formally stated as follows.

$$\text{(MFAL0)} \quad \min \sum_{t=1}^{T} \left[K_t y_t + H_t(I_t) + P_t(X_t) + \pi_t \delta_t I_t \right] \tag{4.3a}$$

s.t.

$$\sum_{i=1}^{t} a_{i,t} x_{it} = d_t, \qquad t = 1, ..., T \tag{4.3b}$$

$$I_t = \sum_{j=t+1}^{T} \sum_{i=1}^{t} a_{i,t} x_{ij}, \qquad t = 1, ..., T \tag{4.3c}$$

$$X_t = \sum_{j=t}^{T} x_{tj}, \qquad t = 1, ..., T \tag{4.3d}$$

$$x_{tj} \leq M y_t, \qquad t = 1, ..., T, j = t, ..., T \tag{4.3e}$$

$$y_t \in \{0, 1\}, \quad x_{tj} \geq 0, \qquad t = 1, ..., T, j = t, ..., T. \tag{4.3f}$$

4.2.2 Shortest Path reformulation (SHP)

The optimal solution for the classical lot sizing problem has the so-called zero inventory order property (ZIO) established by Wagner and Whitin [5], which states that, in an optimal production plan, the starting inventory and production quantity of a period can not be positive simultaneously (*i.e,* $I^*_{t-1} \times x^*_t = 0$). Waterer [22] demonstrates that this property holds for decaying inventories with linear costs, as well. By exploiting this result, Evans [74] has shown that the classical lot sizing problem can be modeled as a shortest path problem (SHP).

Eppen et al. [75] has extended this reformulation to the case of the multi-item capacitated lot sizing problem. However, as shown in Chapter 3, ZIO does not necessarily hold when the production costs are convex. In particular, convex functions may force the model to attain its optimal in the interior point of the feasible region. As an illustration, consider a simple 2-period problem with decay rate $\delta = 10\%$, fixed setup cost $K_t = 200$, production cost function $P_t(x_t) = 0.01x_t^2$, holding cost function $H(I_t) = I_t$ and zero unit decay cost. Suppose the demands are $d_1 = 50$ and $d_2 = 100$. For this setting, it is easy to establish that the optimal solution is $q_1^* = 66.575$ and $q_2^* = 85.083$, which implies that $I_1^* = 16.575$ and $I_1^* \times x_2^* \neq 0$. Therefore, no SHP reformulation can be proposed due to the lack of ZIO property.

4.3 Conic Quadratic Reformulations

We consider second order cone programming (SOCP) in the way that Lobo [76] defines. In a SOCP, a linear function is minimized over the intersection of an affine set and the product of Cartesian product cones.

$$\min \quad \mathbf{f}^T x \tag{4.4}$$

$$s.t. \quad \|\mathbf{c}_i x + \mathbf{d}_i\| \leq \mathbf{a}_i^T x + \mathbf{b}_i, \quad i = 1, \dots, m \tag{4.5}$$

where the decision vector is $x \in \mathbb{R}^n$, and parameter vectors are $\mathbf{f} \in \mathbb{R}^n$, $\mathbf{c}_i \in \mathbb{R}^{(n_i-1)\times n}$, $d_i \in \mathbb{R}^{n_i-1}$, $\mathbf{a}_i \in \mathbb{R}^n$, $\mathbf{b}_i \in \mathbb{R}$, and $\mathbf{f}^T$ and $\mathbf{a}_i^T$ denote the transpose of vector $\mathbf{f}$ and $\mathbf{a}_i$, respectively Lobbo [76]. The constraint (4.5) is defined as a standard cone. Any cone constraint of the form $w^2 \leq uv$ $(u, w \in \mathbb{R}_+)$ defined as a rotated cone can be expressed as a standard cone; $\sqrt{(2w)^2 + (u-v)^2} \leq u + v$. Furthermore, an inequality of the form $z^{\alpha/\beta} \leq r$ with $\alpha > \beta > 0$, $r \in \mathbb{R}$ and $z \in \mathbb{R}_+$ is equivalent to the inequality $z^{2^m} \leq r^\beta z^{2^m-\alpha} 1^{\alpha-\beta}$ where $m = \lceil \log_2 \alpha \rceil$.

Following the binary tree approach of combining variables illustrated in Alizadeh [42] it is easy to verify that an inequality of the form $t^{2^m} \leq$

$s_1 s_2 \dots s_{2^m}$ $(t \in \mathbb{R}, s_i \in \mathbb{R}_+)$ can be expressed by *at most* $2^m - 1$ cone constraints (*e.g.,* Akturk [48]) To the best of our knowledge, the exact number of the needed cone constraints and a construction algorithm are not provided in the literature. In the following, we develop this result. First, we introduce the following definition.

Definition 3 *An inequality of the form*

$$z^{2^m} \leq s_1 s_2 \dots s_{2^m} \quad (z \in \mathbb{R}, s_i \in \mathbb{R}_+) \tag{4.6}$$

is an ***n-block power-m*** *inequality if the variables* $s_1, s_2, \dots, s_{2^m}$ *form* n $(1 \leq n \leq 2^m)$ *distinct blocks such that* $s_1 s_2 \dots s_{2^m} = t_1^{r_1} t_2^{r_2} \cdots t_n^{r_n}$ *where* $r_1 + r_2 + \cdots + r_n = 2^m$ *(i.e.,* $s_j = t_i$ *for some* $j \in \{1, \dots, 2^m\}$ *and* $i = \{1, \dots, n\}$*.)*

For an *n-block power-m* inequality, let r_i be expressed in binary terms as $r_i = \sum_{j=0}^{m} b_j^i 2^j$ with $b_j^i \in \{0, 1\}$ for $i = 1, \dots, n$. Then, $t_i^{r_i} = \prod_{j=0}^{m} t_i^{b_j^i 2^j}$ for $i = 1, \dots, n$. Define the sets $B_{t_i} = \{k \in \{1, \dots, m\} | b_j^i = 1\}$ for $i = 1, \dots, n$ and $T_j = \{t_i^{2^j} | j \in B_{t_i}\}$ for $j = 0, 1, \dots, m$. Let $|\Omega|$ denote the cardinality of set Ω.

Lemma 3 *For an n-block power-m inequality,* $\sum_{k=0}^{j} \frac{|T_k|}{2^{j-k}}$ *is even for j=0,1,...,n.*

Proof The proof rests on induction. Define $\sigma(k) = \sum_{i=0}^{n} \left[\sum_{j=k}^{m} b_j^i 2^j \right]$ for $0 \leq k \leq m$. By construct, $\sum_{i=0}^{n} r_i$ is even. Hence, $\sigma(0)$ is also even. Collecting terms, we have

$$\sigma(0) = \sigma(1) + \sum_{i=0}^{n} b_0^i 2^0 = \sigma(1) + 2^0 |T_0|.$$

As $\sigma(1)$ is even (because it involves power of 2 terms all of which are larger than 1), we have the result for $j = 0$. For $j = 1$, we have

$$\sigma(0) = \sigma(2) + \sum_{i=0}^{n} b_1^i 2^1 + 2^0 |T_0| = \sigma(2) + 2^1 |T_1| + 2^0 |T_0|.$$

Dividing both sides by 2, we get

$$\sigma(0)/2 = \sigma(2)/2 + |T_1| + |T_0|/2.$$

The LHS is by definition 2^{m-1} which is even, and the term $\sigma(2)/2$, which is $\sum_{i=0}^{n}\sum_{j=2}^{m} b_j^i 2^{j-1}$, is also even. Therefore, $|T_1| + |T_0|/2$ must be even, as well. For $j \geq 2$, we proceed in a similar fashion collecting the terms b_j^i and obtain the result.□

One can write the exponents of the RHS variables in an *n-block power-m* inequality in the form of the summation of the complete power of two by simply using the fact that $r_i = \sum_{j=0}^{m-1} b_j 2^j$, and assign a leaf node corresponding to each positive b_j^i with the node degree 2^j while the leaf nodes are sorted on the basis of their degrees. Using the binary tree approach to group the variables (called nodes), every pair of two nodes with the same degree d is combined and a new parent node is created with the degree of $2d$, which corresponds to adding a constraint of the form $w^2 \leq uv$ where w takes place in the parent node but u, v in the child nodes. This procedure continues until only a single orphan node remains with the degree of 2^m. Lemma 3 ensures that following this procedure, each single node can always mate with another node of the same degree. Each such combination reduces the number of orphan nodes by one. Hence, the total number of combinations, N_m is equal to total number of leaf nodes minus one. On the other hand, the number of leaf nodes is equal to the number of positive b_j^i's where the non-redundancy follows from the fact that none of the thus constructed constraints at the leaf nodes can be further eliminated. Thus, we have the following fundamental result.

Proposition 1 *An n-block power-m inequality can be equivalently expressed by exactly* $N_m = \sum_{j=0}^{m-1}\sum_{i=1}^{n} b_j^i - 1$ *inequalities of the form* $w^2 \leq uv$ $(u, v, w \in \mathbb{R}_+)$ *as* *follows.*

$$z^{2^m} \le s_1 s_2 \dots s_{2^m} \Leftrightarrow \begin{cases} u_{j,i}^2 \le v_{j-1,2i-1} v_{j-1,2i} & \textit{for } N'_{j-1} \neq 0;\ i = 1, 2, \dots, N'_{j-1}; \\ & \qquad j = 0, \dots, m-1 \\ z^2 \le v_{m-1,1} v_{m-1,2} & \end{cases} \tag{4.7}$$

where $v_{j,i}^{2^j}$ *is the* i*th element of the set* T'_j *with* $T'_0 = T_0$, $T'_j = T_j \cup \{u_{j,1}^{2^j}, \dots, u_{j,N'_{j-1}}^{2^j}\}$ *and* $N'_{j-1} = |T'_{j-1}|/2$ *for* $j = 1, \dots, m-1$.

Proof We provide a constructive proof using the binary tree approach to group the variables (see Fig.4.1). Build the leaf nodes as follows. Write exponents of the RHS variables in the form of the summation of the complete power of two by simply using the fact that $r_i = \sum_{j=0}^{m-1} b_j 2^j$. Assign a leaf node corresponding to each positive b_j^i with the node degree 2^j and sort the leaf nodes based on their degrees. In the binary tree approach of combining variables (called nodes), each two nodes with the same degree d are combined and create a new parent node with the degree of $2d$. This corresponds to adding a constraint of the form $w^2 \le uv$ in which w takes place in the parent node but u, v in the child nodes. This procedure continues until only a single orphan node remains with the degree of 2^m. Lemma 3 ensures that following this procedure, each single node can always mate with another node of the same degree. Each such combination reduces the number of orphan nodes by 1. Hence, the total number of combinations, N_m is equal to total number of leaf nodes minus one. On the other hand, the number of leaf nodes is equal to the number of positive b_j^i's. The non-redundancy follows from the fact that none of the thus constructed constraints at the leaf nodes can be further eliminated. $\square$

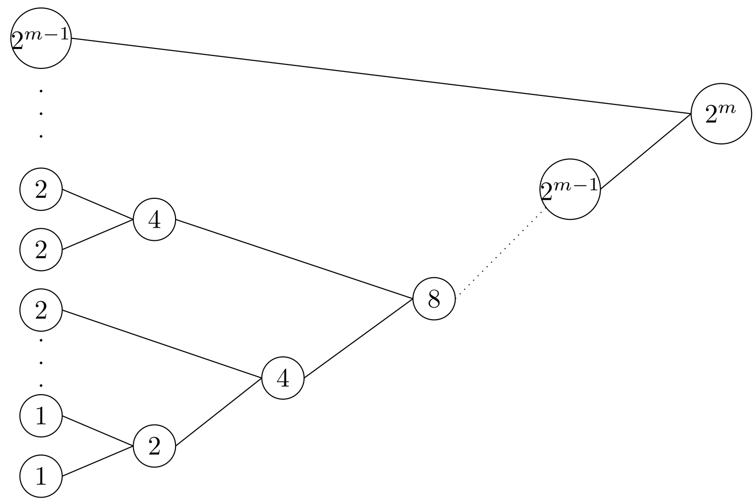


Figure 4.1: Illustration of the binary tree approach of combining the variables.

Algorithm 2 Conic representation of $x^{\alpha/\beta} \leq z$

$l \leftarrow 0, \quad CC \leftarrow \emptyset, \quad m \leftarrow \lceil \log_2 \alpha \rceil$ $\qquad \triangleright$ CC =Cone-Constraints

$T_j \leftarrow \left\{ w \in \{x^{2^{j_x}}\} \cup \{z^{2^{j_z}}\} \cup \{1^{2^{j_1}}\} \mid j_x \in B_x, j_z \in B_z, j_1 \in B_1 \right\}, \quad j = 1, \dots, m-1$

for $k = 0$ to $m-2$ **do**

  while $T_k \neq \emptyset$ **do**

    select two arbitrary $u^{2^k}, v^{2^k} \in T_k$

    $l \leftarrow l + 1$

    $CC \leftarrow CC \cup \{w_l^2 \leq uv\}$

    $T_k \leftarrow T_k \backslash \{u^{2^k}, v^{2^k}\}$

    $T_{k+1} \leftarrow T_{k+1} \cup \{w_l^{2^{k+1}}\}$

  end while

end for

$CC \leftarrow CC \cup \{x^2 \leq uv\}$ where $u^{2^{m-1}}, v^{2^{m-1}} \in T_{m-1}$

Note that the construction is not unique due to the numbering of the elements within the sets T_j' even though the number of cone constraints, N_m is fixed for an *n-block power-m* inequality.
This proposition implies that the number of required cone constraint to represent a *n-block power-m* is equal to the total number of '1's in the binary representation of the exponents in that n-block power-m.

In the *AGG0*, *MNIF0* and *MFAL0* formulations with cost expressions constructed via power functions, we only encounter terms of the form $f(x) = x^{\alpha/\beta}$

with $\alpha > \beta > 0$. Therefore, we provide a streamlined procedure in Algorithm 2 which constructs the equivalent cone constraint set for its epigraph ($epi(f) = \{(x,z)|x^{\alpha/\beta} \leq z\}$) on the basis of Proposition 1.

To illustrate how the proposed algorithm operates, we provide following examples.

Example 1. Consider the inequality $t^{16} \leq x^3y^6z^7$. This is a *3-block power-4* inequality with $t_1 = x$, $t_2 = y$ and $t_3 = z$. In binary notation r_i, we have $b_1^0 = b_1^1 = 1$ and $b_1^j = 0$ for $j = 2,3$, $b_2^1 = b_2^2 = 1$ and $b_2^j = 0$ for $j = 0,3$ and $b_3^0 = b_3^1 = b_3^2 = 1$ and $b_3^j = 0$ for $j = 3$ (i.e, $3 = (0011)_2$, $6 = (0110)_2$ and $7 = (0111)_2$). Thus, $N_m = 7 - 1 = 6$. Applying the construction in Proposition 1, we get $t^2 \leq u_1v_1$, $u_1^2 \leq zy$, $v_1^2 \leq u_2v_2$, $u_2^2 \leq xy$, $v_2 \leq zv_3$, $v_3^2 \leq xz$.

Example 2. The equivalent conic representation of the function $f_2(x) = x^{5/4}$ is $\left\| \begin{matrix} 2x \\ z - w_2 \end{matrix} \right\| \leq z + w_2$, $\left\| \begin{matrix} 2w_2 \\ x - w_1 \end{matrix} \right\| \leq x + w_1$, $\left\| \begin{matrix} 2w_1 \\ x - 1 \end{matrix} \right\| \leq x + 1.$
Since,
$epi(f_1) = \{(x,z)|f_2(x) \leq z\} = \{(x,z)|x^{5/4} \leq z\} \Rightarrow x^5 \leq z^4 \overset{m=3}{\Rightarrow} x^{2^3} \leq x^3 \times z^4 \times 1 = x^{2^1}x^{2^0}z^{2^2}1^{2^0}$
$B_x = \{1,0\}$, $B_z = \{2\}$ and $B_1 = \{0\}$. Thus, Algorithm 2 starts with $T_0 = \{1^{2^0}, x^{2^0}\}$, $T_1 = \{x^{2^1}\}$ and $T_2 = \{z^{2^2}\}$. In the first pass (k=0) we get $CC = \{w_1^2 \leq x \times 1\}$ and $T_0 = \emptyset, T_1 = \{x^{2^1}, w_1^{2^1}\}$. In the second pass (k=1), $CC = \{w_1^2 \leq x \times 1, \quad w_2^2 \leq w_1 \times x\}$ and $T_0 = T_1 = \emptyset, T_2 = \{z^{2^2}, w_2^{2^2}\}$. Finally, line 12 results in $CC = \{w_1^2 \leq x \times 1, \quad w_2^2 \leq w_1 \times x, \quad x^2 \leq w_2 \times z\}$ and the algorithm stops.

The conic quadratic reformulation of the *AGG0*, *MNIF0* and *MFAL0* models rests on *(i)* retaining all of the original linear terms in the objective function and the constraints, *(ii)* replacing the power function terms of the production quantities and/or the ending inventories in the objective function with a single linear variable, say z (and/or r) and *(iii)* adding the equivalent set of cone constraints on z (and/or r) to the optimization problem statement. The resulting models are referred to as the *C-AGG0*, *C-MNIF0* and *C-MFAL0* reformulations.

4.4 Strengthening the models

Optimization models can be strengthened by introducing cuts which exploit the structure of the optimal solution or some other property of the objective function. In this section, we discuss how this can be achieved for the problem at hand. In particular, we introduce three types of cuts that will result in stronger re-statements of the models. The first two of these cuts are based on a structural property of the optimal solution and hold for both MINLP and SOCP formulations. The third cut is developed for the SOCP formulation only. For convenience of exposition, we present the cuts for the MINLP formulation. Their counterparts for the other reformulations are obtained by direct substitution of the intermediate variables defined above. Moreover, we assume that production quantities and inventories are continuous variables, and use the notation f' to denote the first derivative of f with respect to its argument.

The fundamental structural property of an optimal solution on which the proposed cuts will be based is provided below.

Proposition 2 *In an optimal solution, the following inequality holds between any pair $i < j$ where $X_j^* X_i^* > 0$.*

$$\frac{P_i'(X_i^*)}{a_{i,j}} + \sum_{s=i}^{j-1} \Big[\frac{H_s'(I_s^*)}{a_{s,j}} + \frac{\pi_s \delta_s}{a_{s,j}} \Big] \geq P_j'(X_j^*) \tag{4.8}$$

Proof Suppose we modify the optimal solution in such a way that X_j^* is reduced by an amount of ϵ and ϵ / a_{ij} is added to X_i^*, all others are kept unchanged. The modified solution is still feasible and we have

$P_i(X_i^* + \frac{\epsilon}{a_{i,j}}) + P_j(X_j^* - \epsilon) + \sum_{s=i}^{j-1} \Big[H_s(I_s^* + \frac{\epsilon}{a_{s,j}}) + \pi_s \delta_s (I_s^* + \frac{\epsilon}{a_{s,j}}) \Big]$
$\geq P_i(X_i^*) + P_j(X_j^*) + \sum_{s=i}^{j-1} [H_s(I_s^*) + \pi_s \delta_s (I_s^*)]$

Hence,

$P_i(X_i^* + \frac{\epsilon}{a_{i,j}}) - P_i(X_i^*) + \sum_{s=i}^{j-1} \Big[H_s(I_s^* + \frac{\epsilon}{a_{s,j}}) - H_s(I_s^*) + \pi_s \delta_s (I_s^* + \frac{\epsilon}{a_{s,j}}) - \pi_s \delta_s (I_s^*) \Big]$
$\geq +P_j(X_j^*) - P_j(X_j^* - \epsilon)$.

Dividing both sides by ϵ and taking the limits as ϵ it tends to zero provides the desired result.□

The condition in Proposition 2 implies that the marginal production cost of an additional unit in period j cannot exceed that of producing it in period i and carrying it over while incurring some loss due to decay. It is the non-linear counterpart of the Wagner-Whitin condition in the classical lot sizing problem which is typically imposed a priori for a non-trivial solution.

Corollary 3 *Suppose that $P_t'(0) = 0$, $K_t > 0$, $\forall t$. For any pair of periods $i < j$, there exists $M > 0$ such that*

$$\frac{P_i'(X_i^*)}{a_{i,j}} + \sum_{s=i}^{j-1} \left[\frac{H_s'(I_s^*)}{a_{s,j}} + \frac{\pi_s \delta_s}{a_{s,j}}\right] + (1 - y_i)M \geq P_j'(X_j^*) \tag{4.9}$$

Corollary 4 *Suppose that $P_t'(0) = 0$, $K_t > 0$, $\forall t$. For any pair of periods $i < j$,*

$$\frac{P_i'(X_i^*)}{a_{i,j}} + \sum_{s=i}^{j-1} \left[\frac{H_s'(I_s^*)}{a_{s,j}} + \frac{\pi_s \delta_s}{a_{s,j}}\right] \geq P_j'(X_j^*) \times y_i \tag{4.10}$$

Proof In an optimal solution, if $X_i^* > 0$ then $y_i^* = 1$ and both (4.9) and (4.10) reduce to (4.8) which holds for $X_j^* > 0$ due to the assumption above and non-negativity of the left hand side. Otherwise, $y_i^* = 0$ and, therefore, (4.9) will always hold by, for example, setting $M = P_i'(d_j + \ldots + d_T)$, and (4.10) will hold due to the non-negativity of its left hand side. □

Our first cut, Cut 1, is the inequality (4.9) incorporated into the reformulations as a constraint set for the consecutive period pairs (i, j) for $i = t$, $j = i + 1$ and $t \in \{1, 2, \ldots, T - 1\}$. Similarly, our second cut, Cut 2 is the inequality (4.10) incorporated into the reformulations as a constraint set for the consecutive period pairs (i, j) for $i = t$, $j = i + 1$ and $t \in \{1, 2, \ldots, T - 1\}$. Although the inequalities in Corollary (3) and (4) are valid for all periods $i < j$, we observed numerically that introducing the cuts as such deteriorate the model performance considerably.

Therefore, we develop the proposed cuts only for the adjacent periods. Both cuts can be employed in the MINLP and conic formulations. For the SOCP models, we express the nonlinear terms by their equivalent cone constraints. When Cut 1 is included in the formulation, the MINLIP reformulations introduced above are denoted by *AGG1*, *MNIF1* and *MFAL1*. With Cut 2, we get the reformulations *AGG2*, *MNIF2* and *MFAL2*. Their conic counterparts are denoted by the *C-* prefix, such as *C-MFAL1*. The main advantage of Cuts 1 and 2 is that they cut off sub-optimal feasible solutions. In this sense, they are non-redundant. Consequently, in the branch-and-bound procedure, the proposed cuts provide coat bounds at least as tight as one gets in their absence.

The third type of strengthening is based on the cut already introduced in [48] by modifying the cone constraints. It was shown in [72] to be a perspective cut as defined in [77]. The Cut 3 is obtained by replacing any polynomial term in a formulation of the form $z^{2^m} \leq r^{\beta} \times z^{2^m-\alpha} \times 1^{\alpha-\beta}$ to $z^{2^m} \leq r^{\beta} \times z^{2^m-\alpha} \times y_t^{\alpha-\beta}$ where y_t denotes the binary variable indicating a setup when equal to one.

When the perspective cut is employed in the conic formulations *C-AGG0*, *C-MNIF0* and *C-MFAL0*, we obtain *C-AGG3*, *C-MNIF3* and *C-MFAL3*. Likewise, the reformulations *C-AGG1*, *C-MNIF1* and *C-MFAL1* become *C-AGG4*, *C-MNIF4* and *C-MFAL4*; the reformulations *C-AGG2*, *C-MNIF2* and *C-MFAL2* become *C-AGG5*, *C-MNIF5* and *C-MFAL5*.

4.5 Computational Results and Conclusion

In our numerical study, we investigated the solution performance of the different reformulations of the lot sizing problem at hand introduced in the previous section. The MINLP formulations were solved with seven different solvers on the noes-server (http://www.neos-server.org/neos): AlphaECP, BARON, Bonmin, Couenne, DICOPT, LINDOGlobal and KNITRO. Execution time limits of the solvers were set to 1200 seconds and all other options were operated at default. (We refer the reader to [78], [79] and [80] for detailed operational information on this online optimization server.) The conic quadratic formulations were coded in C++ language with gcc compiler calling ILOG CPLEX 12.5.1 with ILOG Concert Technology and solved via MIQCP on a microcomputer with 4GB RAM and Intel(R) Core(TM)2 Quad CPU Q8300 2.5GHz with execution time limit of 300 seconds.

The test bed has been constructed with $T = 100$-period problems. The demands per period were generated from a normal distribution with mean $\mu = 50$ units with negative valued demands rounded off to zero. For each parameter instance, six realizations (R) were generated, one of them with zero variance (constant demands) and the rest with standard deviation $\sigma = 40$. We set unit holding cost rate, $h_t = h = 1$ and select the setup cost as a function of the average demand per period as $K_t = K = [J^2/2]\mu$, in which J may be viewed as a proxy for an average replenishment interval if production and holding costs were linear and was assigned the values of 2 and 5 giving $K \in \{110, 650\}$. Production costs were taken as $\gamma_t x^{\alpha_t/\beta_t}$ with $\alpha_t/\beta_t = \alpha/\beta \in \{3/2, 5/4\}$. To set the γ_t values, we selected the ratio of the unit holding in linear case to production cost in stationary form (i.e., $\frac{h}{\gamma d^{\alpha/\beta}/d}$ with $h = 1$) from $\{0.05, 0.25\}$. Thus, $\gamma_t = \gamma \in \{0.55, 2.32\}$ when $\alpha/\beta = 3/2$ and $\{1.46, 6.24\}$ when $\alpha/\beta = 5/4$. Holding costs were also selected in two forms: $H(I) = hI^{\eta_t/\theta_t}$ with $\eta_t/\theta_t = \eta/\theta \in \{1, 5/4\}$.. Two constant levels of decay rate $\delta_t = \delta \in \{0.05, 0.10\}$ were used and unit decay cost was set to $\pi_t = 0$ and $\pi_t = 2$. Hence, totally 384 ($= \eta/\theta \times \alpha/\beta \times \pi \times \delta \times K \times \gamma \times R = 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 6$) instances were used in our numerical experiment which we group in four sets of 96 based on parameters η/θ and α/β.

All solvers reported as to whether or not the obtained solution was deemed optimal. However, this is based on the intrinsic computation of lower bounds for each solver and may fail to recognize optimality in some cases. Conic quadratic reformulations obtained a feasible solution in all experiment instances, but in 12 instances, none of the reformulations reported optimality. Thus, out of 384 ($= 96 \times 4$) instances, the best available solution corresponds to the reported optimal in 372 instances and, in the remainder, it is simply the best among those obtained by all reformulations.

We begin our discussion with the overall performance of the MINLP reformulations. We provide the summary results in Tables 4.1–4.2, where each production and holding cost pair corresponds to 96 parameter instances. The notation nfs is the number of instances in which the solver gave a feasible solution, and Δ % is the average percentage of deviation from the best available solution for those feasible solutions. The overall poor solver performance is to be expected with MINLP formulations. However, the facility location reformulation is known to be strictly stronger than the two equivalent *AGG* and *NIF* reformulations in the linear cost case, and it is expected that it would be so in the presence of non-linear costs. Therefore, it is interesting that, in our numerical study, the *MFAL* reformulations did not perform as well as the *AGG* and *MNIF*, with or without Cuts 1 and 2. In particular, LINDOGlobal failed to solve any of *MFAL* reformulations in all experiment instances. Among the solvers, Couenne gave higher quality solutions but, at the same time, failed more frequently to find a feasible solution. BARON and KNITRO failed less frequently than the others in finding feasible solutions while the average solution quality in BARON was the worst but a more consistent pattern was observed for KNITRO. DICOPT provided a feasible solution in a larger number of instances but with low quality solutions. As the commercial solvers were used at their default settings, this empirical finding further points to the existing need for developing problem- and solver- specific techniques for MINLP problems.

Overall, the Cuts 1 and 2 result in a significant decrease in the the number of instances with feasible solutions for all MINLP formulations except for the MFAL in a few cases. Among the 6048($= 9 \times 7 \times 96$) outputs of 9 different formulations and 7 different solvers for 96 parameter instances, in only 116 cases –less than 2% of the instances– the feasible solution was reported to be optimal by the solver and all of them detected only by LINDOGlobal.

Next, we consider the performance of the conic counterparts of the MINLP reformulations above (without the perspective cuts) for the same experiment set. Conic quadratic reformulations obtained a feasible solution in all experiment instances. The corresponding average percentage deviations from the best available solution, Δ % are provided in Table 4.3. We see that the conic formulations perform much better than their MINLP counterparts, as expected. The *C-MNIF* reformulation is dominated by either *C-AGG* or *C-MFAL* reformulations with the latter providing a better solution in three of the four groups of instances. In the absence of the perspective cuts, the effects of the proposed Cuts 1 and 2 are mixed. Overall, for the *C-AGG* reformulation, Cut 1 performs slightly better than Cut 2, and they improve the objective for the case when inventory holding cost is also nonlinear although the nominal difference is quite small. For both *C-MNIF* and *C-MFAL* reformulations, Cuts 1 and 2 deteriorate the solution quality slightly.

In Tables 4.4 – 4.5, we present the performance summaries for all of the conic quadratic reformulations. The tabulated results are the minimum, maximum and average of total cpu times (cpu), the average percentage gap between the continuous relaxation at the root node and the best known feasible solution ($rgap$), the number of instances when the solution was reported as optimal (nopt), and the number of instances with unreported optimal solutions (unropt). The last entity corresponds to the instances where the computed lower bound is too loose to report as optimal the solution obtained for the particular reformulation but the obtained solution is reported as optimal by some other reformulation. Also in Figure 4.2, we depict the instances (indexed in the order $\pi \times \delta \times K \times \gamma \times R$) for which a particular conic reformulation reported an optimal solution. The 12 instances for which none of the eighteen conic reformulations obtained a reported optimal are also especially marked by $\times$ on the horizontal axis. All these instances correspond to constant demands. It is surprising to see that there are some instances with no reported optimal solutions for our set, because our test bed was designed to capture the cost values encountered in realistic operational practice and is not challenging at first glance. The difficulty in detecting optimality may result from the constant demand pattern, although problem instances with constant demands were optimized in other groups and cost parameter values.

The strength of the reformulations with the perspective cuts, especially that of *C-MFAL3* is clearly observed. The inclusion of Cuts 1 and 2 along with the perspective cut increases the number of reported optimality instances and decreases the average computational times significantly for *C-AGG* and *C-MNIF* reformulations. Their positive impact is less pronounced for the already strong *C-MFAL* reformulation. However, they may be helpful in identifying some of the optimal solutions that are not reported by *C-MFAL3*. For instance, with $\alpha\beta = 1.5$ and $\gamma = 1.25$, *MFAL4* and *MFAL5* were able to report one and two more optimal solutions, respectively. But, for the same four instances with $\alpha\beta = 1.5$ and linear holding costs and for the same eight instances with $\alpha\beta = 1.25$ and and $\gamma = 1.25$, none of the *C-MFAL* reformulations with the perspective cuts obtained reported optimality. Interestingly, all of these instances correspond to the constant demand pattern. We see that as non-linearity in holding cost increases, the solution performances worsen. On the other hand, the models with $\alpha/\beta = 1.5$ perform better than the others with $\alpha/\beta = 1.25$ although the latter is closer to a linear function. This may be because the number of auxiliary variables and required cone constraints is larger in the latter.

With linear holding costs, inclusion of perishability increases the average solution time, and consequently decreases the number of (detected or undetected) optimal solutions. However, with non-linear holding costs, we observed the opposite. These behaviors were similar for different conic reformulations. Thus, one can conclude that perishability in the form of decay plays an important role on the solvability of a problem. For setups, we observed that solution performances deteriorate as the setup cost increases. This is to be expected. As the contribution of setups in overall costs increases, the solution becomes more sensitive to selection of y_t, burdening the search. The rest of the cost parameters did not exhibit a discernible pattern.

Our experiments demonstrate the following for lot sizing problems with non-linear cost functions. Alternative reformulations of MINLP models do not enhance solvability of the lot sizing problem. Especially the MINLP uncapacitated facility location reformulation performs poorly on commercial solvers with their default settings despite expectations otherwise. Cuts based on optimal solution structures do not result in improvements, may even deteriorate performance with the added non-linear constraint. The performance of a particular MINLP reformulation depends highly on the commercial solver. Conic quadratic reformulations result in significant improvements as expected. In terms of solution speed and solution quality, the facility location reformulation is the strongest conic model, followed by the *NIF* and *AGG* reformulations. Perspective cuts greatly increases the strength of the models whereas cuts based on optimal structure increase the ability to detect optimality. Overall, our study provides empirical evidence to support the conjecture that facility location reformulation for non-linear problems is the stronger model.

		AlphaECP		BARON		Bonmin		Couenne		DICOPT		LINDOGlobal		KNITRO	
	Model	Δ%	nfs	Δ%	nfs	Δ%	nfs	Δ%	nfs	Δ%	nfs	Δ%	nfs	Δ%	nfs
$\eta = 1, \theta = 1, \alpha = 3, \beta = 2$	AGG0	4.44	85	16.915	84	0.705	26	1.11	32	41.065	76	10.35	80	13.99	76
	AGG1	30.85	4	243.57	42	1.16	9	2.71	31	60.51	67	154.93	59	13.36	42
	AGG2	30.85	4	166.54	68	0.98	2	2.64	31	73.67	79	68.02	90	13.36	42
	MNIF0	4.86	74	20.79	84	0.695	26	1.24	28	51.6	13	13.32	80	13.99	76
	MNIF1	10.195	33	144.19	39	1.25	19	1.14	34	44.18	11	108.53	52	13.57	84
	MNIF2	13.9	32	175.34	53	1.43	14	1.88	40	48.85	20	37.44	70	13.57	84
	MFAL0	4.73	86	15.15	90	-	0	-	0	14.38	61	-	0	12.75	63
	MFAL1	7.8	22	18.06	85	3.29	1	0.11	5	13.29	31	-	0	11.79	23
	MFAL2	7.8	22	17.175	88	-	0	-	0	68.82	66	-	0	11.145	44
$\eta = 5, \theta = 4, \alpha = 3, \beta = 2$	AGG0	7	64	17.64	61	-	0	0.27	25	71.42	24	6.56	82	11.18	82
	AGG1	-	0	55.31	42	19.85	1	4.87	32	97.51	18	9.84	64	11.39	40
	AGG2	-	0	81.95	35	-	0	3.82	28	98.06	60	30.37	88	11.39	40
	MNIF0	19.74	85	18.27	54	-	0	0.36	28	88.78	11	11.14	82	12.84	44
	MNIF1	-	0	14.94	5	-	0	0.3	13	63.9	18	11.08	82	12.31	73
	MNIF2	-	0	23.62	2	0.47	1	1.03	36	66.85	79	35.85	45	11.93	76
	MFAL0	4.64	87	13.64	92	-	0	1.08	25	10.13	9	-	0	7.32	4
	MFAL1	-	0	140.36	25	-	0	0.63	72	26.25	11	-	0	-	0
	MFAL2	-	0	169.52	51	-	0	0.91	30	7.51	4	-	0	10.47	41

Table 4.1: Average deviation percentage from the best available solution over 96 instances and number of instances which resulted in feasible solution by the solvers

	Model	AlphaECP Δ%	AlphaECP nfs	BARON Δ%	BARON nfs	Bonmin Δ%	Bonmin nfs	Couenne Δ%	Couenne nfs	DICOPT Δ%	DICOPT nfs	LINDOGlobal Δ%	LINDOGlobal nfs	KNITRO Δ%	KNITRO nfs
$\eta = 1, \theta = 1, \alpha = 5, \beta = 4$	AGG0	3.81	88	37.83	85	0.91	35	0.99	40	14.73	90	9.73	88	20.33	88
	AGG1	28.63	8	121.73	45	0.21	6	4.59	42	32.26	70	13.03	78	21.1	44
	AGG2	28.63	8	137.38	44	0.07	5	4.52	43	47.8	70	17.06	91	21.1	44
	MNIF0	4.01	88	67.2	88	0.87	37	1.06	37	12.9	65	18.89	88	20.59	86
	MNIF1	9.83	14	158.17	92	0.23	5	1.09	37	16.25	70	15.9	88	20.33	88
	MNIF2	20.43	6	227.77	76	0.34	13	1.23	40	19.64	82	80.67	83	20.55	88
	MFAL0	3.3	89	4.51	93	4.34	5	1.21	31	5.06	70	-	0	19.73	83
	MFAL1	5.86	12	20.12	82	2.78	36	1.4	51	32.87	22	-	0	4.33	18
	MFAL2	5.82	11	10.63	77	-	0	66.62	9	43.84	56	-	0	17.29	36
$\eta = 5, \theta = 4, \alpha = 5, \beta = 4$	AGG0	5.37	44	25.52	88	-	0	0.38	24	42.81	42	5.95	88	24.12	45
	AGG1	-	0	152.36	41	-	0	5.25	36	99.35	33	11.28	69	24.07	23
	AGG2	-	0	166.97	39	0.43	1	2.69	26	79.01	17	15.19	91	24.07	23
	MNIF0	18.84	85	20.03	63	-	0	0.39	23	62.8	15	15.43	88	27.76	22
	MNIF1	-	0	66.58	7	4.37	1	0.37	25	72.16	32	25.02	83	25.36	38
	MNIF2	-	0	56.52	6	6.39	1	0.65	17	49.81	86	128.43	34	24.79	41
	MFAL0	4.30	90	9.85	94	-	0	0.89	33	4.67	6	-	0	5.03	11
	MFAL1	-	0	190.48	56	-	0	0.47	73	18.85	23	-	0	-	0
	MFAL2	-	0	94.56	47	-	0	0.74	35	7.21	23	-	0	40	22

Table 4.2: Average deviation percentage from the best available solution over 96 instances and number of instances which resulted in feasible solution by the solvers

$P(.), H(.)$	$x^{3/2}, I$	$x^{3/2}, I^{5/4}$	$x^{5/4}, I$	$x^{5/4}, I^{5/4}$
Model	Δ %	Δ %	Δ %	Δ %
C-AGG0	0.16	0.41	0.17	0.47
C-AGG1	0.14	0.37	0.17	0.25
C-AGG2	0.16	0.36	0.19	0.26
C-NIF0	0.28	0.56	0.39	0.68
C-NIF1	0.31	0.69	0.42	0.70
C-NIF2	0.34	0.81	0.48	0.81
C-MFAL0	0.12	0.08	0.28	0.11
C-MFAL1	0.18	0.14	0.46	0.11
C-MFAL2	0.18	0.15	0.52	0.11

Table 4.3: Average deviation percentage from the best available solution over 96 instances for conic formulations.

		cpu			rgap				
	model	min	max	average	min	max	average	nopt	unropt
$(\eta = 1, \theta = 1), (\alpha = 3, \beta = 2)$	C-AGG0	0.23	300.13	284.73	0	52.77	20.86	5	51
	C-AGG1	0.32	300.19	284.93	0	50.94	13.84	5	50
	C-AGG2	0.41	303.17	277.11	0	29.77	10.84	9	48
	C-AGG3	0.10	300.13	84.95	0	12.9	1.12	70	23
	C-AGG4	0.16	300.08	27.42	0	2.79	0.11	89	6
	C-AGG5	0.18	300.06	28.40	0	2.86	0.11	89	6
	C-NIF0	0.68	300.12	287.62	0	53.2	24.24	4	42
	C-NIF1	0.83	300.14	286.66	0	51.27	15.51	5	46
	C-NIF2	0.98	300.13	285.28	0	31.54	11.8	6	44
	C-NIF3	0.23	300.08	142.5	0	13.08	2.12	56	35
	C-NIF4	0.33	300.07	76.07	0	3.4	0.19	75	17
	C-NIF5	0.38	300.07	76.1	0	3.44	0.21	75	17
	C-MFAL0	0.95	305.83	273.99	0	22.94	8.35	14	48
	C-MFAL1	1.03	307.56	275.78	0	23.07	8.41	13	41
	C-MFAL2	1.01	306.54	278.16	0	23.24	8.47	11	41
	C-MFAL3	0.72	300.11	19.76	0	0.65	0.02	91	4
	C-MFAL4	0.81	300.09	19.95	0	0.65	0.02	91	4
	C-MFAL5	0.84	300.1	20.37	0	0.66	0.02	91	5
$(\eta = 5, \theta = 4)$ $(\alpha = 3, \beta = 2)$	C-AGG0	0.67	300.94	284.43	0	54.22	20.74	7	24
	C-AGG1	0.79	300.24	281.53	0	28.35	10.56	8	27
	C-AGG2	0.88	300.29	278.77	0	28.64	8.49	14	19
	C-AGG3	0.30	300.16	101.84	0	12.43	2.29	67	9
	C-AGG4	0.34	300.16	91.71	0	3.68	0.29	70	16
	C-AGG5	0.41	300.13	92.96	0	3.32	0.29	69	15
	C-NIF0	1.63	300.35	291.71	0	54.56	24.62	4	18
	C-NIF1	2.48	303.05	292.98	0	32.31	13.35	4	20
	C-NIF2	2.62	303.18	293.98	0	31.81	10.64	3	21
	C-NIF3	0.63	300.15	175.44	0	13.46	2.72	46	17
	C-NIF4	0.76	300.18	108.77	0	3.84	0.46	65	15
	C-NIF5	1.25	300.15	101.3	0	3.9	0.44	68	9
	C-MFAL0	15.66	302.45	252.96	0	13.56	4.89	26	31
	C-MFAL1	5.74	302.93	243.97	0	14.06	4.77	27	18
	C-MFAL2	5.48	308.67	244.85	0	13.55	4.79	28	20
	C-MFAL3	4.07	300.08	66.22	0	0.08	0.01	93	3
	C-MFAL4	4.01	300.12	63.76	0	0.02	0.01	94	2
	C-MFAL5	4.33	300.08	67.27	0	0.02	0.01	95	1

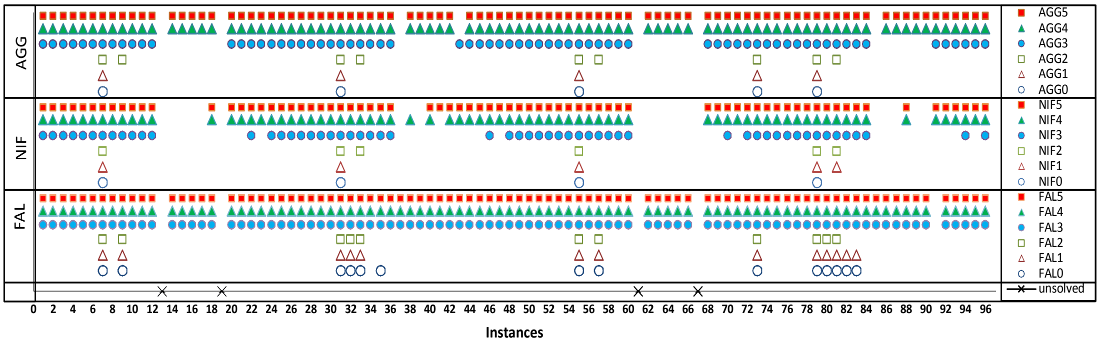
Table 4.4: Performance comparisons of the conic models

.

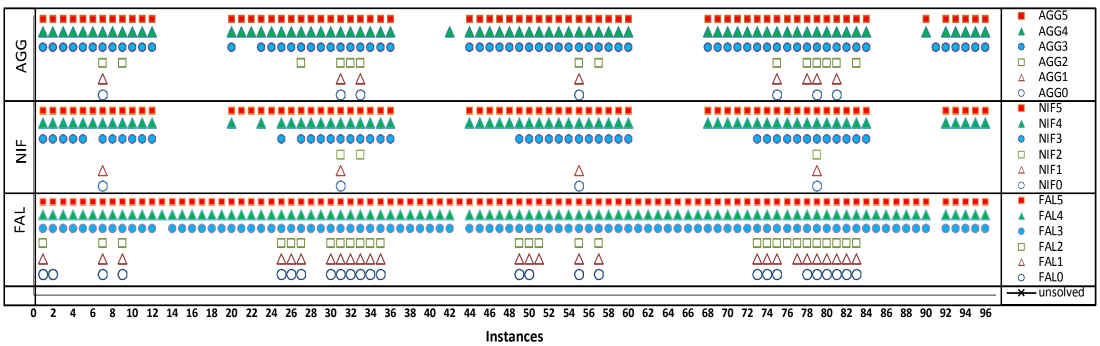
	model	cpu min	cpu max	cpu average	rgap min	rgap max	rgap average	nopt	unropt
$(\eta = 1, \theta = 1), (\alpha = 5, \beta = 4)$	C-AGG0	0.33	302.58	290.72	0	55.82	23.18	3	48
	C-AGG1	0.12	301.00	287.58	0	55.95	14.19	4	46
	C-AGG2	0.13	300.26	286.61	0	56.17	11.17	5	50
	C-AGG3	0.06	300.10	151.25	0	24.43	3.77	49	39
	C-AGG4	0.07	300.75	51.57	0	4.57	0.23	83	10
	C-AGG5	0.08	300.26	53.60	0	4.49	0.23	82	10
	C-NIF0	300	300.13	300.05	6.21	56.98	27.3	0	42
	C-NIF1	0.94	300.12	287.6	0	49.58	15.34	4	38
	C-NIF2	0.98	300.13	287.58	0	30.07	11.51	5	34
	C-NIF3	0.13	300.22	225.55	0	24.45	5.67	24	54
	C-NIF4	0.15	300.12	90.77	0	9.15	0.95	72	18
	C-NIF5	0.16	300.14	92.12	0	9.12	0.98	72	18
	C-MFAL0	0.4	307.03	278.27	0	15.67	5.09	9	53
	C-MFAL1	0.4	308.59	278.94	0	20.28	5.3	8	49
	C-MFAL2	0.42	322.53	278.96	0	18.61	5.36	9	48
	C-MFAL3	0.34	240.23	7.5	0	0.01	0.01	96	0
	C-MFAL4	0.37	254.31	7.85	0	0.01	0.01	96	0
	C-MFAL5	0.42	300.02	8.58	0	0.01	0.01	95	1
$(\eta = 5, \theta = 4), (\alpha = 5, \beta = 4)$	C-AGG0	0.58	300.37	290.80	0	61.67	22.91	3	38
	C-AGG1	0.78	302.91	251.97	0	61.84	9.77	20	29
	C-AGG2	0.90	301.95	254.71	0	62.21	7.92	19	26
	C-AGG3	0.10	307.96	220.15	0	32.75	7.04	28	33
	C-AGG4	0.11	301.14	119.09	0	12.38	0.83	65	20
	C-AGG5	0.11	301.50	120.86	0	12.24	0.84	66	17
	C-NIF0	0.93	300.69	290.85	0	55.38	25.08	3	21
	C-NIF1	1.16	300.59	291	0	30.39	11.27	3	18
	C-NIF2	1.33	300.66	290.14	0	29.81	8.97	4	18
	C-NIF3	0.17	330.01	227.64	0	32.76	8.04	24	23
	C-NIF4	0.21	300.26	147.36	0	12.07	1.28	62	10
	C-NIF5	0.21	300.44	154.38	0	11.89	1.36	58	8
	C-MFAL0	5.05	301.46	234.87	0	10.87	2.7	29	17
	C-MFAL1	5.46	302.24	235.34	0	10.69	2.73	28	13
	C-MFAL2	5.52	301.96	236.55	0	10.79	2.73	28	13
	C-MFAL3	4.3	307.06	71.37	0	1.44	0.08	87	8
	C-MFAL4	4.58	300.79	66.67	0	1.48	0.08	87	5
	C-MFAL5	4.75	300.89	75.71	0	1.46	0.09	86	7

Table 4.5: Performance comparisons of the conic models

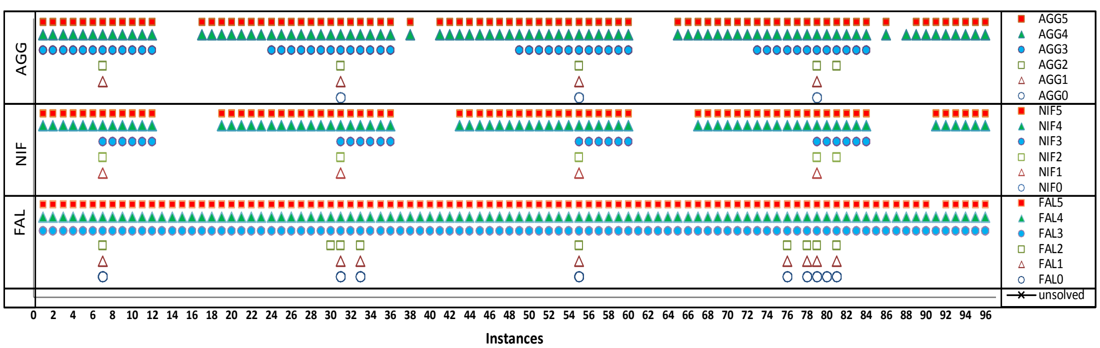
.



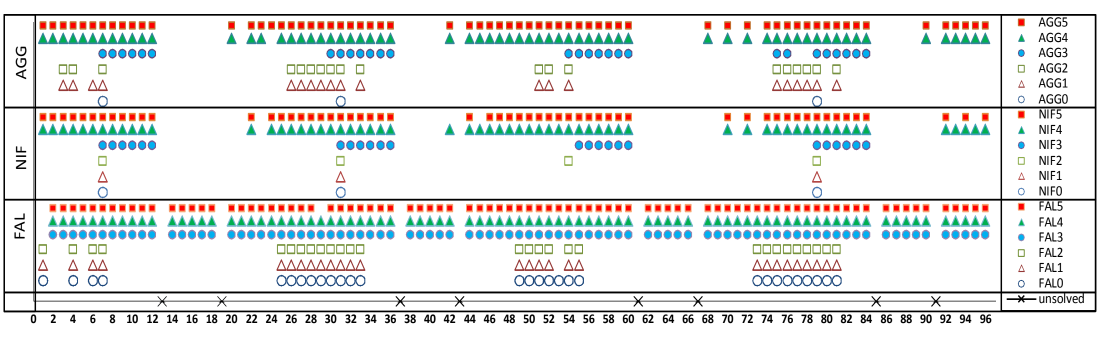
(a):($\eta = 1,\ \theta = 1$), ($\alpha = 3,\ \beta = 2$)



(b):($\eta = 5,\ \theta = 4$), ($\alpha = 3,\ \beta = 2$)



(c):($\eta = 1,\ \theta = 1$), ($\alpha = 5,\ \beta = 4$)



(d):($\eta = 5,\ \theta = 4$), ($\alpha = 5,\ \beta = 4$)

Figure 4.2: Detected and undetected optimal solutions for each formulation. Notation ($\times$) denotes instances for which no optimal was reported in study.

Chapter 5

Rolling horizon performance of the lot sizing heuristics

In this chapter we investigate the problem defined in Chapter 3 in the rolling horizon framework. In this framework the demand data is known only for a limited number of periods and the rest is revealed as time proceeds. Heuristic algorithms have three main advantages over the exact ones. *(i)* In the rolling horizon the optimality of an exact algorithm is not guaranteed and even heuristic algorithms may perform better [26]. *(ii)* The implementation of the heuristic algorithms is simpler. For instance, Chand at al.[34] in their extensive review paper quote from Hopp [81] that "*no commercial MRP package actually uses WW algorithm*". and they justify it with another quotation from the same authors stating that "*people would rather live with a problem they cannot solve than accept a solution that they do not understand*". *(iii)* The exact solution algorithms provide production plans with higher nervousness level. That is, the scheduled production plan is prone to change (reschedule) which is not an appealing phenomenon from a managerial perspective as discussed by Carlson [82].

The remainder of this chapter is organized as follows: in Section 5.1 we briefly introduce the mathematical model of our problem, then in Section 5.2 and Section 5.3, we provide numerical results and the conclusion.

5.1 The model

We consider a single item, with deterministic demands over a horizon of T periods. Demands are known for certain number of periods ahead called the forecast horizon length (FHL).

In our model x_t and I_t indicate the production amount and the inventory level at period t, respectively. We denote the demand quantity of period t by d_t. Setup and unit holding costs are expressed by K_t and h_t, respectively. The production cost of x_t amount of product is given by $P_t(x_t)$ which is a convex function. The problem over the periods $u, ..., v$ is denoted as P_{uv} and is written as follows.

$$(P_{u,v}) \quad \min_{x_u,\dots,x_v} \sum_{t=u}^{v} \Big[(K_t \mathbb{1}_{\{x_t>0\}} + h_t I_t + P_t(x_t)\Big] \tag{5.1a}$$

$$\text{s.t.}$$

$$I_{u-1} = 0,$$

$$I_t = I_{t-1} + x_t - d_t, \qquad t = u, ..., v \tag{5.1b}$$

$$I_t, x_t \geq 0, \qquad t = u, ..., v \tag{5.1c}$$

where $\mathbb{1}$ is the indicator function of the setup occurrence.
Note that if we set $u = 1$ and $v = T$ and all the demands are known, then the resulting model is the static fixed horizon version of the above problem. However, the demand quantities are unknown beyond the FHL number of periods after the current decision period. That is, at the first period we have only the demand information of periods $\{1, \dots, FHL\}$ and decide on the production quantity of the periods $1, \dots, FHL$. Then in the second period we are left with the surplus items produced in period one while the demand of the period $FHL+1$ is revealed. In our setting, we assume that only the first production quantity of each sub horizon is kept and the remaining ones may change. Thus, retaining the production quantity of the first period we can update the planned production quantities for periods $2, \dots, FHL + 1$ and the procedure continues for all t until period $T - FHL + 1$ and at step t we solve $P_{t,t+FHL-1}$. This rolling planning is outlined in Algorithm 3.

Algorithm 3 Rolling horizon setting

Require: forecast horizon length, Algorithm name
Ensure: Provide a production plan
function ROLLING_IMPLEMENTATION($AlgName$, FHL)
  for $i = 1$ to $T - FHL + 1$ **do**
    implement $AlgName$ over to the problem $P_{i,i+FHL-1}$
    keep x_i
    Update net demand using x_i
  end for
end function

5.2 Numerical study

5.2.1 Parameters setting

In this section we present our findings from the numerical study. We have conducted our experimental study to investigate the problem from the following aspects *(i)* percentage deviation of the total cost under heuristic from the optimal cost, *(ii)* impact of the demand pattern, *(iii)* impact of the forecast horizon, *(iv)* impact of system parameters. The optimal cost values are obtained by the CPLEX solver. We have considered problem horizon length of $T = 100$ periods. The demands are generated from three distributions *(a)* Gamma, *(b)* Normal and *(c)* constant demands each with two levels for the mean demand, $\mu_1 = 50$ and $\mu_2 = 200$. We refer to these demand settings as G50, G200, N50, N200, C50 and C200, respectively. The standard deviation, σ_1 , of G50, N50 and N200 is set to 40 while G200 is constructed with $\sigma_2 = 100$ in order to compare the effect of variance (σ) versus the coefficient of variation (σ/μ). We have simulated 10 realizations from the random demands. The resulting mean and variance are given in Table 5.1. The negative values of demand which were generated only in N50, were truncated to zero.

The parameter set we used in our experiments is taken similar to that in Chapter 4. The unit holding cost per period was set as $h_t = h = 1$ and the setup cost was adjusted as $K_t = K = J^2/2\mu$ where $J \in \{2, 3, 4, 5\}$. A power function of the form $P_t(x_t) = a_t x_t^r$ was used for production cost where $r \in \{1.1, 1.25, 1.5, 2, 2.2\}$

												Total
G50	μ	41.3	49.5	45.3	53.9	54.0	51.3	41.2	54.3	41.4	43.0	47.51
	σ	30.4	42.6	35.6	41.7	48.8	39.8	35.8	45.0	30.8	39.9	39.64
G200	μ	203.4	200.9	190.6	179.3	191.1	206.5	194.4	200.1	204.1	190.5	196.09
	σ	103.7	92.2	93.2	84.8	91.5	104.8	88.4	109.7	98.9	97.1	96.62
N50	μ	52.7	54.8	54.0	53.0	46.2	53.9	49.4	51.3	55.2	56.2	52.66
	σ	39.0	33.7	40.3	38.3	37.9	32.6	40.6	36.3	37.4	37.0	37.34
	0	11	9	14	10	20	8	15	14	8	12	121
N200	μ	204.0	201.7	199.9	198.0	197.1	207.4	200.0	204.8	197.8	196.2	200.70
	σ	37.3	43.1	44.7	38.4	38.4	40.1	36.5	40.7	46.7	36.6	40.37
C50	μ	50.0	-	-	-	-	-	-	-	-	-	50.00
	σ	0.0	-	-	-	-	-	-	-	-	-	0.00
C200	μ	200.0	-	-	-	-	-	-	-	-	-	200.00
	σ	0.0	-	-	-	-	-	-	-	-	-	0.00

Table 5.1: Average and standard deviation of the demand realizations

and $w_t = w = h/w\mu^{r-1}$ where $a_t = a \in \{0.02, 0.05, 0.1, 0.2\}$. All the instances have been solved both with the fixed forecast horizon length of 100 and with rolling forecast horizon lengths of 2 through 20 by step size 2. Levels of the parameters (i.e., *demand* $\times$ *Replication* $\times K \times a \times r$) lead to overall $3200(= 4 \times 10 \times 4 \times 4 \times 5)$ problem instances from normal and gamma demands and also $160(= 2 \times 1 \times 4 \times 4 \times 5)$ instances from constant demands. We have used seven solution methods including six heuristic algorithms and the CPLEX solver which is used as our benchmark corresponding to the optimal solution. They are tested against 11 different forecast horizon lengths. Therefore, our complete experiment includes 258,720 data points.

5.2.2 Performance comparisons

We define two performance measures. The first one denoted by Δ_1, is the percentage deviation of the heuristic cost from the objective value of the CPLEX solution with full demand information of the entire horizon. The second one denoted by Δ_2, is the percentage deviation of the heuristic cost from the objective value with the same forecast horizon used in optimal solution. Let TC_i^{FHL} denote the total cost obtained by implementing algorithm i (i=0 corresponds to the CPLEX) to a problem instance with forecast horizon length of FHL. Then Δ_1 and Δ_2 are defined as

$$\Delta_1 = \frac{TC_i^T - TC_0^T}{TC_0^T} \times 100$$

$$\Delta_2 = \frac{TC_i^{FHL} - TC_0^{FHL}}{TC_0^{FHL}} \times 100.$$

For the benchmark solution we fed a second order cone programming formulation of the problem into the CPLEX 12.5 solver and we limited the solution time to 15 minutes for each horizon length. Our preliminary tests showed that the small instances of the problem (in the size of horizon lengths we used) are optimally solvable within this time limit, however, in some relatively large instances ($T = 100$) the solver hits the time limit and the optimal solution is not found. Therefore, we observe also some negative Δ_1 values for static basis ($FHL = 100$).

We first present the performance of each heuristic separately for different demand patterns. In Tables 5.2–5.7 we report *(i)* the average cost deviation percentage from the optimal static solution (Δ_1), *(ii)* the average cost deviation percentage from the optimal rolling solution (Δ_2), *(iii)* the average(Δ_2^-), *(iv)* median ($\hat{\Delta}_2^-$) and also *(v)* the statistical frequency (n^-) of negative cost deviation percentages in the rolling horizon setting. In general, we observe that deviation optimality from both static and rolling horizon are increasing in the coefficient of deviation (μ/σ) of demand pattern. For each heuristic, both Δ_1 and Δ_2 take greater values for G50 and N50 compared to the G200 and N200. The demand variance in G200 is greater than that in G50, however, the coefficient of variation in G200 is smaller. Comparing G200 and N200 in Tables 5.3 – 5.5 demonstrates the same findings. The demand coefficient of variation in G200 is 0.5 while it is 0.2 in N200. On the other hand, no significant change is observed between C50 and C200 in Table 5.7 where the demand variation coefficients are equal (=0) which reveals that the coefficient of variation in demand pattern is a more important factor than the nominal value of demand variance. The order of algorithms, dominance with respect to average of Δ_1 or Δ_2 is very similar but not exactly the same. For the Gamma and Normal cases in Tables 5.2 – 5.5 the general dominance pattern is found to be $inc.ex, psdWW, GWW, GEOQ, GLUC, GSM$.

This pattern is also visually noticeable in Figures 5.3–5.8. However, for less variable demand patterns N200, C50, and C200, GLUC dominates GEOQ and is close to GWW in performance. Comparing G50 and N50 reveals that optimality deviations are slightly smaller for Normal demands with G-type heuristics while psdWW and inc.ex heuristics resulted in somewhat smaller deviations for Gamma demands.

The pictorial representation of algorithms performance, their sensitivity and behavior on the forecast horizon lengths is provided in Figure 5.5 – 5.8 in the form of box plots. The average values in these graphs are marked by $\triangledown$ and the median value by $\oplus$. In G50, G200 and N50 the median and mean of Δ_1 for the exact method, psdWW and inc.ex seem to have a decreasing pattern as the forecast horizon length increases while the two-step heuristics (except GSM) come out to have a uniform pattern with a slight increase after $FHL = 4$. In N200 these trends become decreasing likewise for the two-step heuristics (except for GSM). For constant demands, the median of the deviations converges to zero for all of the algorithms and the averages converge faster starting from $FHL = 4$ and the deviation ranges are significantly more compact than from that in the other demands. (See Figures 5.3–5.5).
Regarding Δ_2, we clearly observe that deviations converge to zero for inc.ex and psdWW in all demand patterns while the range and variability of the deviation values in two-step heuristics are preserved similar to the Δ_1 but their mean and median do not have a local minimum and appear in a increasing form in the rolling forecast horizon lengths. (See Figures 5.6–5.8).
Note that the range, mean and median of both measures (Δ_1 and Δ_2) are smaller in G200 and GN200 compared to the G50 and N50 due to the demand variability. Also focusing on constant demands in Figures 5.5 and 5.8 reveals a stationary robust performance of the algorithms.

5.2.3 Dominance of the heuristics over the exact solution

In this part we discuss on the dominance of the heuristics over the exact solution in a rolling horizon setting. In particular, we elaborate the obtained statistics about the problem instances come out with negative Δ_2 which the half part of Tables 5.2–5.3, Table 5.8 and Figure 5.2 are dedicated to. For all the heuristics results for the Gamma and Normal demand (Tables 5.2– 5.5), the number of instances which resulted in negative Δ_2 appears in relatively short forecast horizon length with 2–6 periods (see column n^-). The number of outperformed myopic exact solutions, n^-, is generally higher in G50 and G200 than in N50 and N200 and their averages are also slightly less (better) for Gamma demands (See columns Δ_2^-). Also, the value of the n^- shows a negative correlation with the coefficient of variation in demand especially for the two-step heuristics.

The magnitudes of the negative optimality deviations are also greater around the short horizon lengths. The average of negative deviations (Δ_2^-), their median ($\hat{\Delta}_2^-$) and the most negative values (Δ_2^*) mostly take place in the first three rows of each heuristic block in these tables. However, for the constant demand in Tables 5.6–5.7 we see that n^- and (Δ_2^*) are distributed uniformly across the forecast horizons without accumulation over short forecast horizon lengths.

We observe that although GSM, GLUC and GEOQ are the last three heuristics in the ranking discussed above based on the average Δ_1 or Δ_2, they do extremely better than exact myopic solutions in some instances. For instance, Δ_2^* is found as $(-1.78, -1.27, -1.42)$ in G50 , $(-3.05, -1.49, -2.45)$ in G200, $(-2.45, -0.74, -2.45)$ in N50, and $(-3.78, -2.61, -2.82)$ in N200 by means of these three heuristics, (See Tables 5.2–5.5).

G50	GWW						GSM						GEOQ					
FHL	Δ_1	Δ_2	Δ_2^-	$\hat{\Delta}_2^-$	Δ_2^*	n^-	Δ_1	Δ_2	Δ_2^-	$\hat{\Delta}_2^-$	Δ_2^*	n^-	Δ_1	Δ_2	Δ_2^-	$\hat{\Delta}_2^-$	Δ_2^*	n^-
2	9.49	1.91	-0.23	-0.17	-0.76	20	20.83	12.01	-0.57	-0.57	-0.72	2	8.59	1.10	-0.28	-0.17	-1.42	58
4	9.30	7.04	-0.14	-0.07	-0.54	29	18.92	16.36	-0.52	-0.36	-1.56	11	8.96	6.72	-0.35	-0.21	-1.27	20
6	9.58	8.51	-0.04	-0.04	-0.06	5	22.68	21.37	-0.72	-0.61	-1.78	9	9.36	8.29	-0.84	-0.86	-1.27	4
8	9.23	8.64	-0.07	-0.06	-0.21	8	23.67	22.92	-0.77	-0.77	-0.77	1	8.96	8.39	-	-	-	0
10	9.19	8.84	-0.11	-0.11	-0.11	1	24.04	23.58	-	-	-	0	9.21	8.86	-	-	-	0
12	9.47	9.25	-	-	-0.01	0	25.20	24.92	-	-	-	0	9.26	9.05	-	-	-	0
14	9.22	9.09	-0.03	-0.03	-0.03	3	25.43	25.25	-	-	-	0	8.93	8.80	-	-	-	0
16	9.21	9.13	-0.07	-0.07	-0.07	1	25.05	24.94	-	-	-	0	8.85	8.77	-	-	-	0
18	9.10	9.05	-	-	-	0	24.83	24.76	-	-	-	0	8.80	8.74	-	-	-	0
20	8.93	8.90	-0.28	-0.28	-0.28	1	24.76	24.72	-	-	-	0	8.60	8.57	-	-	-	0
100	2.20	2.20	-0.02	-0.02	-0.02	1	17.88	17.88	-	-	-	0	1.40	1.40	-	-	-	0
G50	GLUC						psdWW						inc.ex					
FHL	Δ_1	Δ_2	Δ_2^-	$\hat{\Delta}_2^-$	Δ_2^*	n^-	Δ_1	Δ_2	Δ_2^-	$\hat{\Delta}_2^-$	Δ_2^*	n^-	Δ_1	Δ_2	Δ_2^-	$\hat{\Delta}_2^-$	Δ_2^*	n^-
2	10.93	3.24	-0.36	-0.27	-1.27	27	7.39	0.00	-	-	-	0	7.39	0.00	-	-	-	0
4	10.89	8.61	-0.34	-0.27	-0.88	9	1.97	0.03	-0.09	-0.05	-0.41	60	1.98	0.04	-0.13	-0.07	-0.82	132
6	11.55	10.47	-0.77	-0.77	-0.77	1	0.90	0.02	-0.06	-0.05	-0.26	34	0.91	0.04	-0.06	-0.04	-0.23	41
8	11.54	10.94	-	-	-	0	0.49	0.02	-0.07	-0.05	-0.19	17	0.51	0.04	-0.04	-0.03	-0.18	27
10	11.49	11.13	-	-	-	0	0.30	0.02	-0.06	-0.05	-0.12	6	0.32	0.05	-0.04	-0.03	-0.11	10
12	11.76	11.54	-	-	-	0	0.19	0.02	-0.05	-0.05	-0.12	5	0.21	0.05	-0.05	-0.05	-0.05	1
14	11.46	11.33	-	-	-	0	0.12	0.02	-0.04	-0.03	-0.05	3	0.15	0.05	-0.04	-0.03	-0.06	4
16	11.48	11.40	-	-	-	0	0.08	0.02	-0.04	-0.05	-0.06	4	0.11	0.05	-0.05	-0.05	-0.05	1
18	11.40	11.34	-	-	-	0	0.06	0.02	-0.08	-0.08	-0.11	2	0.09	0.05	-	-	-0.01	0
20	11.21	11.17	-	-	-	0	0.05	0.02	-0.15	-0.15	-0.28	2	0.08	0.05	-0.28	-0.28	-0.28	1
100	3.93	3.93	-	-	-	0	0.02	0.02	-0.05	-0.03	-0.13	12	0.06	0.06	-0.02	-0.02	-0.02	1

Table 5.2: Heuristic performance in demand pattern G50

G200			GWW						GSM						GEOQ			
FHL	Δ_1	Δ_2	Δ_2^-	$\hat{\Delta}_2^-$	Δ_2^*	n^-	Δ_1	Δ_2	Δ_2^-	$\hat{\Delta}_2^-$	Δ_2^*	n^-	Δ_1	Δ_2	Δ_2^-	$\hat{\Delta}_2^-$	Δ_2^*	n^-
2	3.20	0.29	-0.09	-0.10	-0.15	5	9.34	6.24	-0.07	-0.07	-0.07	1	3.36	0.45	-0.11	-0.07	-0.76	117
4	2.21	1.49	-0.13	-0.10	-0.43	40	7.86	7.09	-0.48	-0.37	-1.16	17	2.13	1.42	-0.40	-0.23	-1.20	27
6	2.66	2.36	-0.07	-0.05	-0.22	18	8.95	8.63	-0.67	-0.33	-3.05	15	2.75	2.45	-0.69	-0.46	-2.45	8
8	2.77	2.63	-0.09	-0.07	-0.35	13	9.51	9.37	-0.25	-0.14	-0.49	3	2.88	2.75	-	-	-	0
10	2.81	2.74	-0.08	-0.06	-0.23	11	9.88	9.80	-0.10	-0.10	-0.10	1	2.98	2.91	-	-	-	0
12	2.79	2.76	-0.22	-0.25	-0.33	6	10.06	10.02	-	-	-	0	2.96	2.92	-	-	-	0
14	2.82	2.80	-0.22	-0.21	-0.32	4	9.99	9.97	-	-	-	0	2.98	2.96	-	-	-	0
16	2.78	2.78	-	-	-	0	10.29	10.28	-	-	-	0	2.95	2.94	-	-	-	0
18	2.72	2.71	-	-	-	0	10.09	10.09	-	-	-	0	2.87	2.86	-	-	-	0
20	2.65	2.65	-0.09	-0.09	-0.09	1	10.20	10.20	-	-	-	0	2.79	2.79	-	-	-	0
100	0.40	0.40	-0.04	-0.04	-0.04	1	8.74	8.74	-	-	-	0	0.41	0.41	-0.04	-0.04	-0.04	2
G200			GLUC						psdWW						inc.ex			
FHL	Δ_1	Δ_2	Δ_2^-	$\hat{\Delta}_2^-$	Δ_2^*	n^-	Δ_1	Δ_2	Δ_2^-	$\hat{\Delta}_2^-$	Δ_2^*	n^-	Δ_1	Δ_2	Δ_2^-	$\hat{\Delta}_2^-$	Δ_2^*	n^-
2	3.66	0.74	-0.23	-0.16	-0.79	23	2.91	0.00	-	-	-	0	2.91	0.00	-	-	-	0
4	3.37	2.64	-0.37	-0.34	-1.00	38	0.70	0.00	-0.07	-0.04	-0.41	67	0.71	0.00	-0.08	-0.06	-0.47	113
6	3.92	3.62	-0.62	-0.41	-1.49	6	0.29	0.01	-0.05	-0.04	-0.18	29	0.31	0.03	-0.06	-0.04	-0.37	29
8	4.16	4.02	-	-	-	0	0.13	0.01	-0.07	-0.04	-0.35	19	0.16	0.03	-0.11	-0.07	-0.35	9
10	4.16	4.09	-	-	-	0	0.07	0.01	-0.09	-0.09	-0.23	10	0.10	0.04	-0.11	-0.10	-0.23	10
12	4.14	4.10	-	-	-	0	0.04	0.01	-0.26	-0.25	-0.33	5	0.07	0.04	-0.19	-0.24	-0.33	7
14	4.16	4.14	-	-	-	0	0.02	0.01	-0.19	-0.14	-0.32	5	0.06	0.04	-0.22	-0.21	-0.32	4
16	4.14	4.13	-	-	-	0	0.02	0.01	-0.03	-0.03	-0.03	2	0.05	0.04	-	-	-	0
18	4.08	4.07	-	-	-	0	0.01	0.01	-0.08	-0.08	-0.12	2	0.04	0.04	-	-	-	0
20	3.96	3.96	-	-	-	0	0.01	0.01	-0.10	-0.09	-0.18	3	0.04	0.04	-0.09	-0.09	-0.09	1
100	1.16	1.16	-0.04	-0.04	-0.04	1	0.00	0.00	-0.06	-0.04	-0.13	16	0.04	0.04	-0.05	-0.04	-0.08	3

Table 5.3: Heuristic performance in demand pattern G200

N50			GWW						GSM						GEOQ			
FHL	Δ_1	Δ_2	Δ_2^-	$\hat{\Delta}_2^-$	Δ_2^*	n^-	Δ_1	Δ_2	Δ_2^-	$\hat{\Delta}_2^-$	Δ_2^*	n^-	Δ_1	Δ_2	Δ_2^-	$\hat{\Delta}_2^-$	Δ_2^*	n^-
2	8.42	3.26	-0.09	-0.08	-0.17	8	15.07	9.51	-0.14	-0.09	-0.34	4	8.11	2.99	-0.13	-0.08	-0.35	21
4	6.21	4.89	-0.09	-0.07	-0.31	23	12.04	10.63	-0.75	-0.66	-2.45	8	6.01	4.69	-0.70	-0.40	-2.45	15
6	6.65	6.06	-0.06	-0.04	-0.15	14	12.71	12.07	-0.82	-0.82	-0.82	1	6.65	6.06	-0.24	-0.24	-0.35	2
8	6.70	6.39	-0.05	-0.02	-0.13	7	13.85	13.51	-	-	-	0	6.81	6.50	-	-	-	0
10	6.70	6.53	-0.15	-0.15	-0.15	1	14.20	14.01	-	-	-	0	6.78	6.61	-	-	-	0
12	6.69	6.59	-0.12	-0.12	-0.22	2	14.53	14.42	-	-	-	0	6.73	6.63	-	-	-	0
14	6.71	6.65	-	-	-	0	14.72	14.65	-	-	-	0	6.73	6.67	-	-	-	0
16	6.66	6.61	-	-	-	0	14.84	14.79	-	-	-	0	6.65	6.61	-	-	-	0
18	6.55	6.53	-	-	-	0	14.65	14.62	-	-	-	0	6.55	6.53	-	-	-	0
20	6.48	6.46	-0.05	-0.05	-0.05	1	14.59	14.57	-	-	-	0	6.50	6.48	-	-	-	0
100	3.80	3.80	-	-	-	0	13.76	13.76	-	-	-	0	3.31	3.31	-	-	-	0

N50			GLUC						psdWW						inc.ex			
FHL	Δ_1	Δ_2	Δ_2^-	$\hat{\Delta}_2^-$	Δ_2^*	n^-	Δ_1	Δ_2	Δ_2^-	$\hat{\Delta}_2^-$	Δ_2^*	n^-	Δ_1	Δ_2	Δ_2^-	$\hat{\Delta}_2^-$	Δ_2^*	n^-
2	9.15	3.97	-0.28	-0.21	-0.74	23	5.94	1.00	-	-	-	0	5.94	1.00	-	-	-	0
4	7.55	6.20	-0.10	-0.10	-0.12	2	2.55	1.33	-0.08	-0.05	-0.27	29	2.61	1.39	-0.09	-0.07	-0.35	41
6	7.99	7.40	-	-	-	0	2.10	1.57	-0.06	-0.04	-0.20	30	2.12	1.60	-0.07	-0.05	-0.27	30
8	8.03	7.71	-	-	-	0	1.94	1.68	-0.09	-0.06	-0.22	11	1.94	1.67	-0.03	-0.03	-0.06	6
10	8.08	7.91	-	-	-	0	1.77	1.63	-0.12	-0.12	-0.12	1	1.79	1.65	-0.04	-0.04	-0.04	2
12	8.11	8.01	-	-	-	0	1.70	1.62	-0.10	-0.05	-0.22	3	1.73	1.65	-0.09	-0.02	-0.22	3
14	8.12	8.06	-	-	-	0	1.64	1.59	-0.03	-0.03	-0.03	4	1.66	1.61	-0.03	-0.03	-0.03	1
16	8.02	7.98	-	-	-	0	1.56	1.53	-0.03	-0.03	-0.03	1	1.59	1.55	-	-	-	0
18	7.91	7.88	-	-	-	0	1.50	1.48	-	-	-	0	1.53	1.51	-	-	-	0
20	7.82	7.80	-	-	-	0	1.47	1.46	-0.05	-0.05	-0.05	1	1.50	1.49	-0.05	-0.05	-0.05	1
100	4.84	4.84	-	-	-	0	0.22	0.22	-0.02	-0.02	-0.02	1	0.25	0.25	-	-	-	0

Table 5.4: Heuristic performance in demand pattern N50

N200			GWW						GSM						GEOQ			
FHL	Δ_1	Δ_2	Δ_2^-	$\hat{\Delta}_2^-$	Δ_2^*	n^-	Δ_1	Δ_2	Δ_2^-	$\hat{\Delta}_2^-$	Δ_2^*	n^-	Δ_1	Δ_2	Δ_2^-	$\hat{\Delta}_2^-$	Δ_2^*	n^-
2	1.02	0.00	-0.03	-0.03	-0.04	2	2.30	1.29	-0.03	-0.03	-0.03	1	1.57	0.55	-0.09	-0.09	-0.25	20
4	0.22	0.01	-0.08	-0.05	-0.33	23	1.15	0.94	-0.52	-0.47	-1.37	43	0.54	0.33	-0.16	-0.13	-0.46	39
6	0.15	0.03	-0.05	-0.04	-0.17	8	1.19	1.07	-1.03	-0.61	-3.78	25	0.50	0.38	-0.82	-0.74	-2.82	25
8	0.08	0.04	-0.06	-0.04	-0.17	5	1.23	1.19	-0.37	-0.37	-0.37	1	0.48	0.44	-	-	-	0
10	0.07	0.04	-0.09	-0.08	-0.17	3	1.28	1.25	-0.50	-0.50	-0.50	1	0.47	0.44	-	-	-	0
12	0.05	0.04	-0.07	-0.06	-0.12	3	1.30	1.29	-0.08	-0.08	-0.08	1	0.46	0.45	-	-	-	0
14	0.05	0.04	-0.18	-0.18	-0.23	2	1.31	1.30	-	-	-	0	0.45	0.45	-	-	-	0
16	0.05	0.04	-0.07	-0.07	-0.12	2	1.32	1.32	-	-	-	0	0.45	0.44	-	-	-	0
18	0.04	0.05	-0.06	-0.02	-0.14	3	1.33	1.33	-	-	-	0	0.45	0.45	-	-	-	0
20	0.04	0.05	-0.04	-0.04	-0.04	1	1.34	1.34	-	-	-	0	0.44	0.45	-	-	-	0
100	0.02	0.02	-0.05	-0.07	-0.09	7	1.75	1.75	-	-	-	0	0.20	0.20	-0.20	-0.21	-0.37	5
N200			GLUC						psdWW						inc.ex			
FHL	Δ_1	Δ_2	Δ_2^-	$\hat{\Delta}_2^-$	Δ_2^*	n^-	Δ_1	Δ_2	Δ_2^-	$\hat{\Delta}_2^-$	Δ_2^*	n^-	Δ_1	Δ_2	Δ_2^-	$\hat{\Delta}_2^-$	Δ_2^*	n^-
2	1.07	0.05	-0.16	-0.13	-0.52	29	1.01	0.00	-	-	-	0	1.01	0.00	-	-	-	0
4	0.22	0.02	-0.56	-0.47	-1.65	94	0.20	0.00	-0.11	-0.08	-0.53	37	0.24	0.03	-0.06	-0.04	-0.22	18
6	0.23	0.11	-0.67	-0.33	-2.61	48	0.12	0.00	-0.07	-0.06	-0.16	9	0.17	0.05	-0.05	-0.03	-0.15	9
8	0.23	0.18	-0.11	-0.08	-0.33	14	0.04	0.00	-0.09	-0.06	-0.19	9	0.10	0.06	-0.06	-0.04	-0.12	5
10	0.23	0.20	-0.11	-0.10	-0.33	13	0.03	0.00	-0.08	-0.08	-0.10	6	0.09	0.06	-0.09	-0.08	-0.15	3
12	0.23	0.22	-0.02	-0.02	-0.02	3	0.01	0.00	-0.05	-0.06	-0.09	9	0.08	0.06	-0.07	-0.08	-0.10	4
14	0.24	0.23	-0.02	-0.02	-0.02	1	0.01	0.00	-0.10	-0.06	-0.23	4	0.07	0.06	-0.17	-0.17	-0.23	2
16	0.24	0.24	-	-	-	0	0.00	0.00	-0.04	-0.04	-0.06	4	0.06	0.06	-0.06	-0.06	-0.10	2
18	0.24	0.24	-0.03	-0.03	-0.03	1	0.00	0.00	-0.06	-0.05	-0.14	4	0.06	0.06	-0.08	-0.08	-0.14	2
20	0.24	0.25	-	-	-0.01	0	0.00	0.00	-0.06	-0.06	-0.07	3	0.06	0.06	-0.04	-0.04	-0.04	1
100	0.18	0.18	-	-	-0.01	0	-	-	-0.15	-0.04	-0.68	34	0.06	0.06	-0.03	-0.03	-0.04	4
							0.01	0.01										

Table 5.5: Heuristic performance in demand pattern N200

C50			GWW						GSM						GEOQ			
FHL	Δ_1	Δ_2	Δ_2^-	$\hat{\Delta}_2^-$	Δ_2^*	n^-	Δ_1	Δ_2	Δ_2^-	$\hat{\Delta}_2^-$	Δ_2^*	n^-	Δ_1	Δ_2	Δ_2^-	$\hat{\Delta}_2^-$	Δ_2^*	n^-
2	0.84	0.00	-	-	-	0	0.84	0.00	-	-	-	0	1.55	0.71	-	-	-	0
4	0.20	0.07	-	-	-	0	0.13	-0.01	-1.21	-0.75	-2.35	5	0.40	0.27	-0.73	-0.73	-0.89	2
6	0.16	0.11	-	-	-	0	0.12	0.07	-1.12	-1.23	-1.60	4	0.42	0.37	-0.77	-0.77	-1.13	2
8	0.11	0.00	-2.02	-2.25	-2.81	4	0.12	0.01	-1.53	-1.95	-2.62	5	0.38	0.27	-2.72	-2.72	-2.72	1
10	0.15	0.03	-0.82	-0.60	-2.38	6	0.12	0.00	-0.92	-0.61	-2.19	7	0.37	0.25	-0.25	-0.17	-0.45	3
12	0.11	0.11	-	-	-	0	0.12	0.12	-0.29	-0.29	-0.29	1	0.38	0.38	-	-	-	0
14	0.12	0.09	-0.42	-0.42	-0.42	1	0.12	0.10	-0.51	-0.51	-0.60	2	0.36	0.34	-0.13	-0.13	-0.13	1
16	0.13	0.08	-1.26	-1.26	-2.22	2	0.12	0.07	-0.97	-0.90	-2.04	4	0.36	0.30	-	-	-	0
18	0.11	0.10	-0.29	-0.29	-0.29	1	0.12	0.11	-0.43	-0.43	-0.56	2	0.37	0.36	-0.09	-0.09	-0.09	1
20	0.11	0.07	-2.46	-2.46	-2.46	1	0.12	0.09	-2.26	-2.26	-2.26	1	0.36	0.32	-2.26	-2.26	-2.26	1
100	0.11	0.11	-	-	-	0	0.12	0.12	-	-	-	0	0.14	0.14	-	-	-0.01	0
C50			GLUC						psdWW						inc.ex			
FHL	Δ_1	Δ_2	Δ_2^-	$\hat{\Delta}_2^-$	Δ_2^*	n^-	Δ_1	Δ_2	Δ_2^-	$\hat{\Delta}_2^-$	Δ_2^*	n^-	Δ_1	Δ_2	Δ_2^-	$\hat{\Delta}_2^-$	Δ_2^*	n^-
2	0.84	0.00	-	-	-	0	0.84	0.00	-	-	-	0	0.84	0.00	-	-	-	0
4	0.13	-0.01	-1.21	-0.75	-2.35	5	0.12	-0.01	-0.55	-0.55	-0.89	2	0.21	0.08	-	-	-	0
6	0.12	0.07	-1.12	-1.23	-1.60	4	0.06	0.00	-	-	-	0	0.17	0.12	-	-	-	0
8	0.12	0.01	-1.53	-1.95	-2.62	5	0.02	-0.09	-2.02	-2.25	-2.81	4	0.12	0.01	-2.02	-2.25	-2.81	4
10	0.12	0.00	-0.92	-0.61	-2.19	7	0.07	-0.06	-0.72	-0.45	-2.38	8	0.16	0.04	-0.82	-0.60	-2.38	6
12	0.12	0.12	-0.29	-0.29	-0.29	1	0.00	0.00	-	-	-	0	0.12	0.12	-	-	-	0
14	0.12	0.10	-0.51	-0.51	-0.60	2	0.02	0.00	-0.19	-0.15	-0.42	4	0.13	0.10	-0.42	-0.42	-0.42	1
16	0.12	0.07	-0.97	-0.90	-2.04	4	0.04	-0.02	-1.26	-1.26	-2.22	2	0.14	0.09	-1.26	-1.26	-2.22	2
18	0.12	0.11	-0.43	-0.43	-0.56	2	0.01	0.00	-0.29	-0.29	-0.29	1	0.12	0.11	-0.29	-0.29	-0.29	1
20	0.12	0.09	-2.26	-2.26	-2.26	1	0.02	-0.02	-1.28	-1.28	-2.46	2	0.12	0.08	-2.46	-2.46	-2.46	1
100	0.12	0.12	-	-	-	0	0.00	0.00	-0.04	-0.04	-0.05	2	0.11	0.11	-	-	-	0

Table 5.6: Heuristic performance in demand pattern C50

C200			GWW						GSM						GEOQ			
FHL	Δ_1	Δ_2	Δ_2^-	$\hat{\Delta}_2^-$	Δ_2^*	n^-	Δ_1	Δ_2	Δ_2^-	$\hat{\Delta}_2^-$	Δ_2^*	n^-	Δ_1	Δ_2	Δ_2^-	$\hat{\Delta}_2^-$	Δ_2^*	n^-
2	0.66	0.00	-	-	-	0	0.66	0.00	-	-	-	0	1.35	0.69	-	-	-	0
4	0.11	0.04	-	-	-	0	0.06	-0.01	-0.87	-0.84	-1.65	5	0.36	0.29	-0.53	-0.53	-0.84	2
6	0.09	0.03	-	-	-	0	0.06	0.00	-0.88	-0.86	-1.77	4	0.37	0.32	-0.50	-0.50	-0.98	2
8	0.05	0.00	-1.20	-1.53	-1.73	3	0.06	0.01	-0.82	-0.79	-1.54	4	0.35	0.30	-0.21	-0.21	-0.21	1
10	0.09	0.07	-0.67	-0.67	-1.28	2	0.06	0.04	-0.46	-0.05	-1.28	3	0.33	0.31	-0.04	-0.04	-0.04	1
12	0.04	0.05	-	-	-	0	0.06	0.06	-	-	-	0	0.33	0.33	-	-	-	0
14	0.05	0.00	-0.77	-0.43	-1.62	5	0.06	0.01	-0.66	-0.43	-1.43	5	0.33	0.28	-	-	-	0
16	0.04	0.02	-1.07	-1.07	-1.62	2	0.06	0.03	-0.90	-0.90	-1.43	2	0.32	0.29	-	-	-	0
18	0.04	0.03	-0.58	-0.58	-1.12	2	0.06	0.04	-0.58	-0.58	-1.12	2	0.32	0.30	-	-	-	0
20	0.04	0.05	-	-	-	0	0.06	0.06	-	-	-	0	0.31	0.31	-	-	-	0
100	0.04	0.04	-0.02	-0.02	-0.02	2	0.06	0.06	-0.02	-0.02	-0.02	2	0.13	0.13	-0.04	-0.04	-0.06	2
C200			GLUC						psdWW						inc.ex			
FHL	Δ_1	Δ_2	Δ_2^-	$\hat{\Delta}_2^-$	Δ_2^*	n^-	Δ_1	Δ_2	Δ_2^-	$\hat{\Delta}_2^-$	Δ_2^*	n^-	Δ_1	Δ_2	Δ_2^-	$\hat{\Delta}_2^-$	Δ_2^*	n^-
2	0.66	0.00	-	-	-	0	0.66	0.00	-	-	-	0	0.66	0.00	-	-	-	0
4	0.06	-0.01	-0.87	-0.84	-1.65	5	0.07	0.00	-	-	-	0	0.12	0.05	-	-	-	0
6	0.06	0.00	-0.88	-0.86	-1.77	4	0.05	0.00	-	-	-	0	0.10	0.04	-	-	-	0
8	0.06	0.01	-0.82	-0.79	-1.54	4	0.00	-0.05	-1.20	-1.53	-1.73	3	0.05	0.00	-1.20	-1.53	-1.73	3
10	0.06	0.04	-0.46	-0.05	-1.28	3	0.05	0.03	-0.67	-0.67	-1.28	2	0.09	0.07	-0.67	-0.67	-1.28	2
12	0.06	0.06	-	-	-	0	0.00	0.00	-	-	-	0	0.06	0.06	-	-	-	0
14	0.06	0.01	-0.66	-0.43	-1.43	5	0.00	-0.05	-0.57	-0.31	-1.62	7	0.06	0.01	-0.77	-0.43	-1.62	5
16	0.06	0.03	-0.90	-0.90	-1.43	2	0.00	-0.03	-1.07	-1.07	-1.62	2	0.06	0.03	-1.07	-1.07	-1.62	2
18	0.06	0.04	-0.58	-0.58	-1.12	2	0.00	-0.01	-0.39	-0.04	-1.12	3	0.05	0.04	-0.58	-0.58	-1.12	2
20	0.06	0.06	-	-	-	0	0.00	0.00	-	-	-	0	0.06	0.06	-	-	-	0
100	0.06	0.06	-0.02	-0.02	-0.02	2	0.00	0.00	-0.03	-0.02	-0.06	4	0.04	0.04	-0.02	-0.02	-0.02	2

Table 5.7: Heuristic performance in demand pattern C200

Now we focus on the cases in which the exact solution was dominated by a heuristic solution in a rolling horizon. The related data are summarized in Table 5.8 where columns indicate demand distributions and row blocks represent the effective factors which are setup cost, production cost convexity, forecast horizon and heuristics. The number of problem instances in each block of the first four columns is 8800 whereas there are 880 in the last two columns. Figure 5.2 provides the same data in graphics by means of doughnut charts with multiple rings. Each ring corresponds to a different demand pattern with the same order as in the column titles in Table 5.8 (i.e., the most interior one corresponds to C50). In the first two blocks, we see that the frequency is high in K4 (largest setup cost) over all demand patterns and it decreases monotonically as setup cost gets smaller. Also the general trend is decreasing in $r_i, i = 1, ...4$, however, this pattern is not monotone and we have almost one exception in group of demands. For instance, from $r4$ to $r5$ in G50, from $r2$ to $r3$ in G200, from $r1$ to $r2$ in N50 and from $r3$ to $r4$ in C200. The third block in Table 5.8 depicts that among the forecast horizon lengths, as we previously observed in Tables 5.2 –5.5, for the random demands, the large values of n^- are distributed mostly in short forecast horizon lengths (See also Figure 5.2:(c)). The last block shows the magnitude of n^- across the algorithms: inc.ex and psdWW are the leading algorithms in G50, G200 and N50 while GLUC becomes the leader heuristic. For the constant demand patterns GSM and GLUC are the leading algorithms followed by psdWW. Afterwards, GWW and inc.ex perform similarly.

At the bottom of Table 5.8 the total number n^- values and their percentage among the total instances of each demand distribution are summarized. Overall, in 7.13% of all rolling horizon experiment instances we have encountered a negative value for Δ_2.

5.2.3.1 Analysis of the heuristic dominance over the forecast horizon

Let J be the economic production interval, defined as the number of periods whose demand is covered by the lot size minimizing the total cost per period with stationary demand pattern. Also define $\hat{J}$ as the effective length of production lot in an optimal production plan of the static solution which is the practical average length of production lot after solving the problem instance. J has been set during the parameters adjustment while $\hat{J}$ is calculated from the outcomes for each problem instance. The optimal effective production length is calculated as $\hat{J} = T/pn$ where pn is the number of production periods in the static solution. After calculating these $\hat{J}$ values for all instances in static basis, ($FHL = T$), they were offset to their corresponding instances in all forecast horizon length in order to create scatter plot of Δ_2^- among FHL and $\hat{J}$ which is depicted in Figure 5.1. The followings are observed in this figure: *(i)* There is no significant dominance of heuristics over the exact solution in $J = 2$ for all forecast horizon lengths except for G50. *(ii)* Most of the dot points are distributed in the region where $J > \hat{J}$ and where $J < J_{max} = 5$.

5.2.3.2 Effect of multiple solution in constant demand pattern

We observed that in some instances for constant demand pattern the exact solution showed unexpectedly poor performance in rolling horizon setting. The reason for such cases is the possibility of multiple solutions which is illustrated here by an example. For instance, consider the following parameters: $d_t = 50$, $K_t = 422.08$, $w_t = 6.74$ and $r = 1.1$. A 20-period solution of a this instance results in the following optimal solution (100 , 0 , 150 , 0 , 0 , 150 , 0 , 0 , 150 , 0 , 0 , 150 , 0 , 0 , 150 , 0 , 0 , 150 , 0 , 0). Note that it is one of the several possible optimal solutions because the first production lot (100,0) can be shifted to other places (at the end of horizon for example). But when it takes place at the beginning of the horizon it affects the solution as we update the net demands in rolling basis. Consequently, it may leads to a very poor solution. Interestingly, a little noise in demand data prevents the pattern discussed above. For instance, for the demands like (49 , 50 , 51 , 50 , 50 , 48 , 50 , 50 , 50 , 50 , 50 , 50 , 50 , 52 , 50 , 50 , 50 , 50 , 50 , 50) we obtain the production plan (150 , 0 , 0 , 148 , 0 , 0 , 150 , 0 , 0 , 150 , 0 , 0 , 102 , 0 , 150 , 0 , 0 , 150 , 0 , 0) which leads to the repetition of suitable batch size in the rolling basis. Therefore, a random data pattern is better than a pure constant one for the comparison of the exact and heuristic solutions.

5.2.4 Sensitivity analysis

Now we describe the general effect of the parameters on Δ_1 and Δ_2 in different forecast horizon lengths and also between each pair of parameters. To this aim, we have conducted a general linear model analysis of variance (ANOVA) considering the demand distribution, the algorithm, the FHL and problem parameters: the convexity level, the setup cost, and the production cost coefficient as the main effective factors. We have also considered the pairwise interaction of the main factors with the FHL to investigate the impression of the forecast horizon length on the other factors. In our analysis, Δ_1 and Δ_2 were chosen as the response variables in ANOVA as illustrated in Tables 5.9 and 5.10, respectively. We have used SPSS software to obtain ANOVA tables. The reader may refer to [83] for the details of ANOVA process. The observed small p-values enables us to conclude that statistically all the factors have significant effect on the magnitude of the both deviation types. Therefore, we compare the values of F-statistics among the factors to identify which one is more effective. Demand distribution, algorithm and production cost parameters (r and w) are the first most significant factors in both tables whereas FHL is the least significant factor among them. Among the paired interactions with forecast horizon lengths, the convexity of the production cost function is the most effective interaction while the setup cost level has the least effective interaction with FHL for both of the deviation types which is similar to as in the main factors. Nevertheless, the significance order of other paired factors does not follow the same order as the main ones but in general we can conclude the robustness of the heuristics performance with respect to FHL levels.

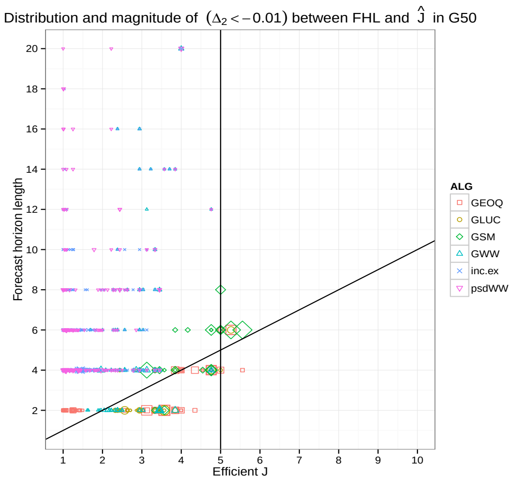
Visual depiction of the individual factor effects and the average sensitivity of the heuristics performances in the rolling horizon setting, (Δ_2), are also provided in Figures 5.9 – 5.11 for each demand, separately. These graphs are based on the mean and they disclose, that Δ_2 is increasing in r, decreasing in K, and decreasing in w levels in variable demands. However, for constant demands and N200 (which is very close to stationary demand) the monotone behavior among these factor levels vanishes. The second column in these figures brings out the robustness of the algorithms, dominance over different forecast horizon lengths as observed also in the ANOVA tables. On the other hand, Figure 5.11 illustrates that the dominance order alternates over different forecast horizons for stationary demand. That is, the heuristics can be more competitive in rolling horizon setting and the length of the forecast horizon matters in constant cases rather than in highly variable demand patterns.

5.3 Conclusion

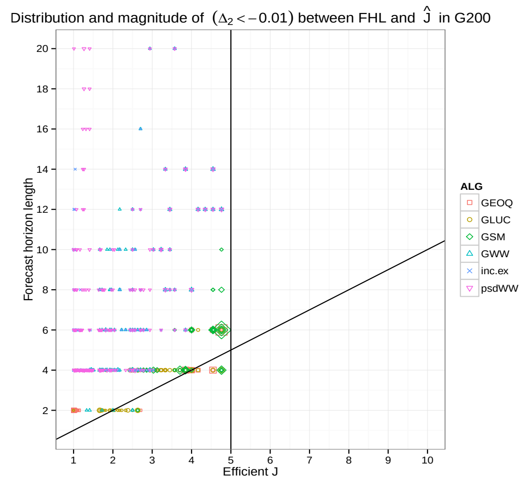
In this chapter, the behavior of the heuristics designed for the lot sizing problems with convex production cost was studied in the rolling horizon setting. Our simulations demonstrate that the performance of the heuristics is sensitive to the demand variability and the problem parameters. Even though the length of the forecast horizon does not mainly change the dominance order of the heuristics for highly variable demand. It appears to have more effect for constant or close to constant demand distributions. On the other hand, most of the heuristics perform better on a myopic rolling horizon especially with FHL being 4 periods. This observation suggests that more information about the future demands does not necessarily result in a better production plan. Our analysis for the heuristics dominance showed that most of such cases happens when the forecast horizon length is larger than the optimal economic production interval for the static solution. Another interesting finding within this study is that the heuristics which are on average worst performer with respect to other heuristics, may result in a production plan with significantly lower cost than a production plan provided by the exact method or by the other best performer heuristics. This emphasizes the importance of a heuristic algorithm in this class of the problem alike to their classical counterparts.

Table 5.8: Statistical frequency of n^- among factors

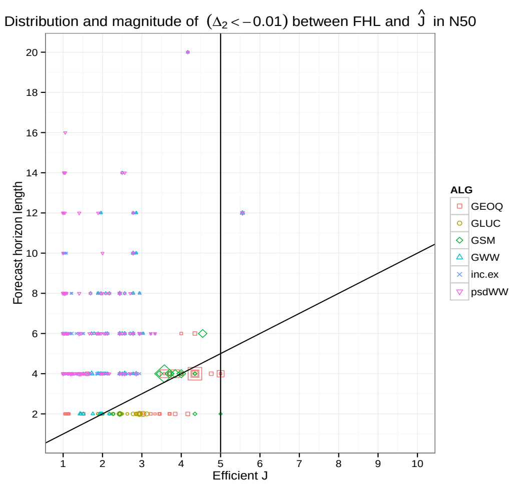
	G50	G200	N50	N200	C50	C200
K1	30	33	9	4	0	0
K2	87	78	24	38	14	12
K3	202	192	75	166	50	43
K4	255	389	189	384	64	60
r1	109	173	56	213	70	56
r2	128	116	79	128	38	44
r3	128	151	94	100	13	0
r4	98	139	61	87	3	14
r5	111	113	7	64	4	1
2	107	146	56	52	0	12
4	261	302	118	254	14	10
6	94	105	77	124	10	18
8	53	44	24	34	23	13
10	17	32	4	26	37	0
12	6	18	8	20	2	27
14	10	13	5	9	11	10
16	6	2	1	8	14	11
18	2	2	0	10	8	0
20	4	5	3	5	7	14
GWW	69	99	56	59	15	16
GSM	23	37	13	72	31	27
GEOQ	82	154	38	89	11	8
GLUC	37	68	25	203	31	27
psdWW	145	158	81	119	25	21
inc.ex	218	176	84	50	15	16
Total #	574	692	297	592	128	115
Percentage	6.5%	7.9%	3.4%	6.7%	14.5%	13.1%



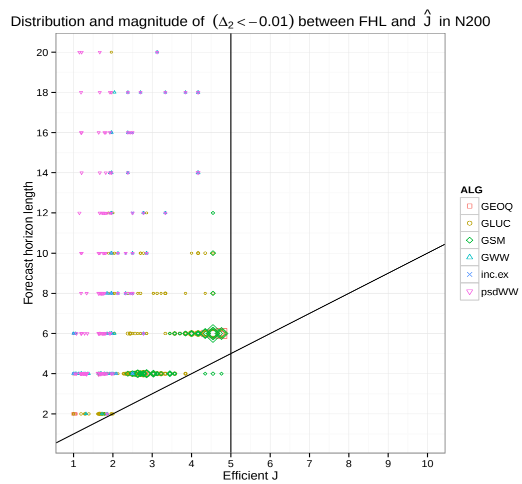
(a)



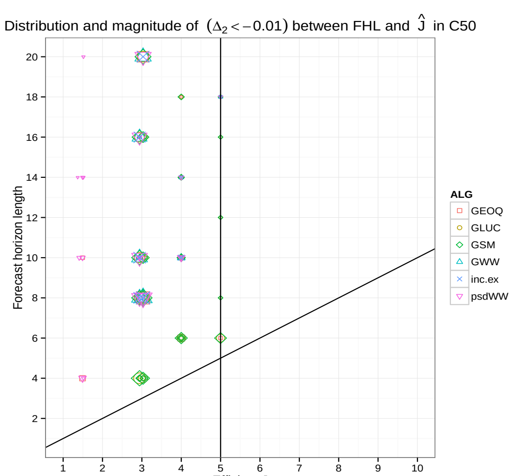
(b)



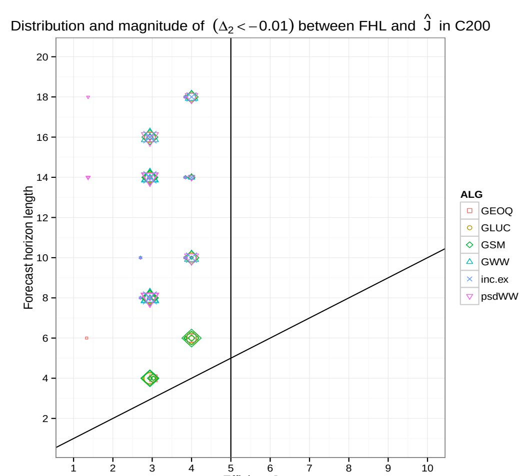
(c)



(d)

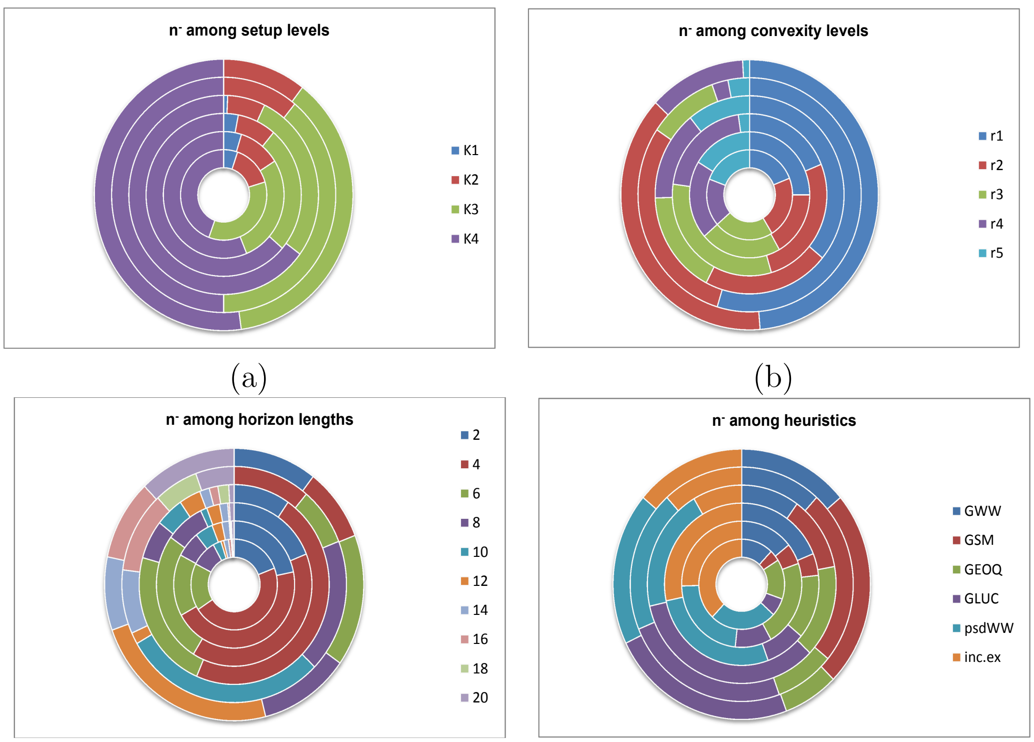


(e)



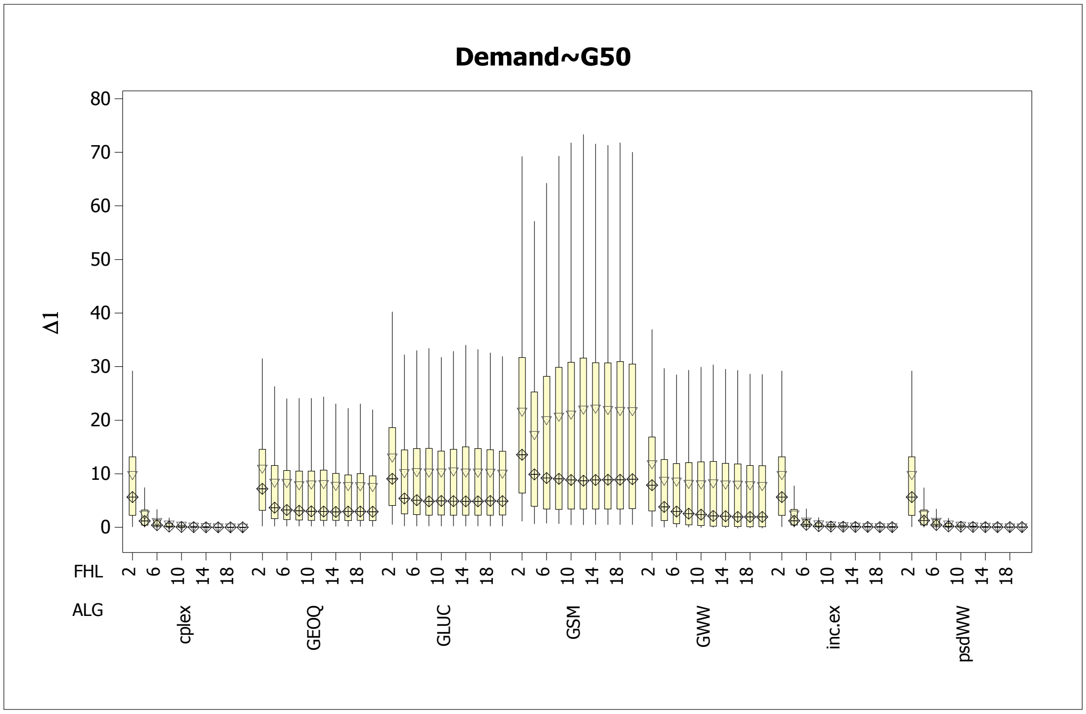
(f)

Figure 5.1: $\hat{J}$ v.s. FHL. $J \in \{2, 3, 4, 5\}$ Filter$=-0.01$

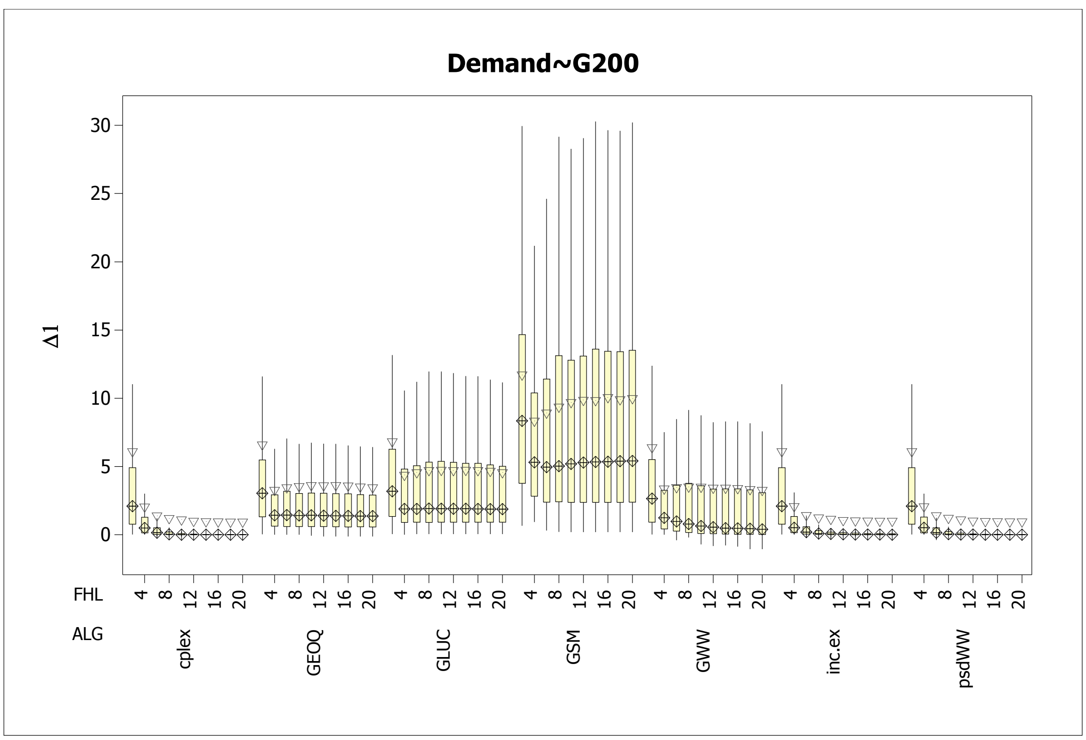


(a) (b) (c) (d)

Figure 5.2: Statistical frequency of n^{-} among the factors

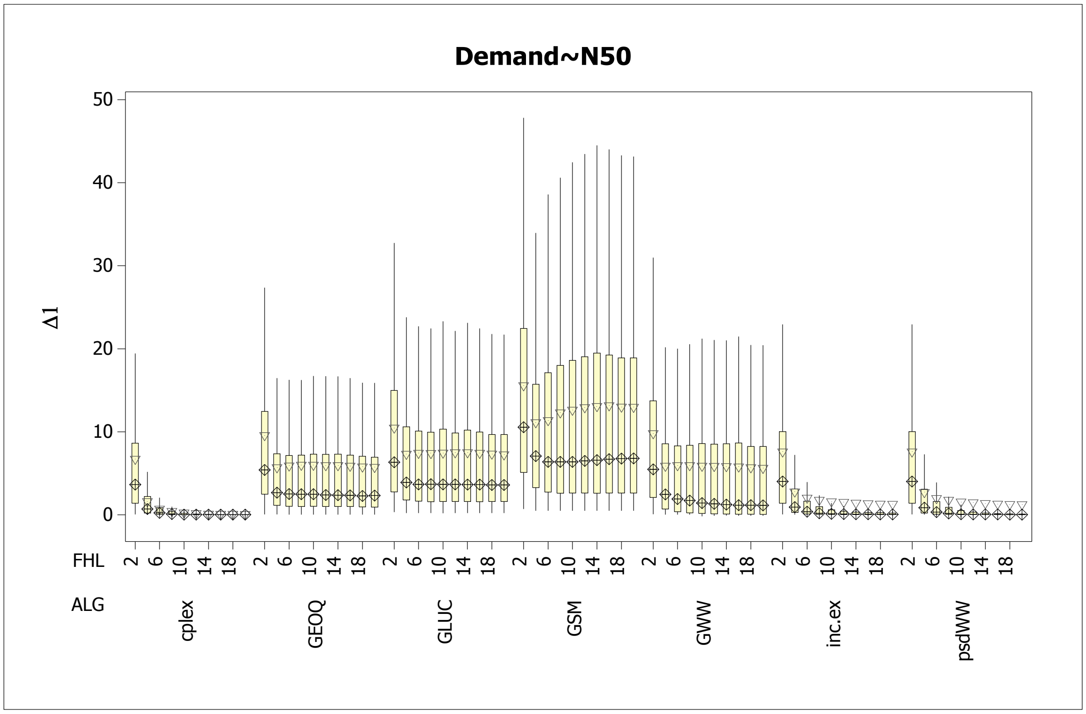


(a)

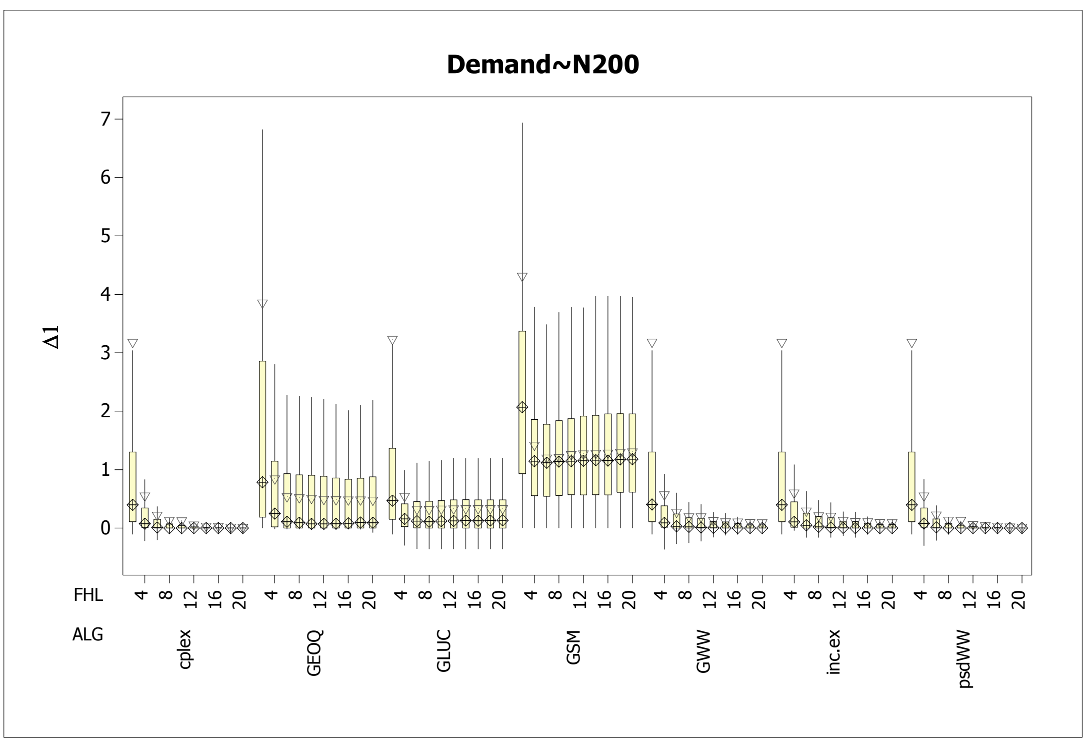


(b)

Figure 5.3: Box plots of Δ_1 values over all the instances for each algorithm in Gamma demands (a): $\mu = 50$, (b) $\mu = 200$

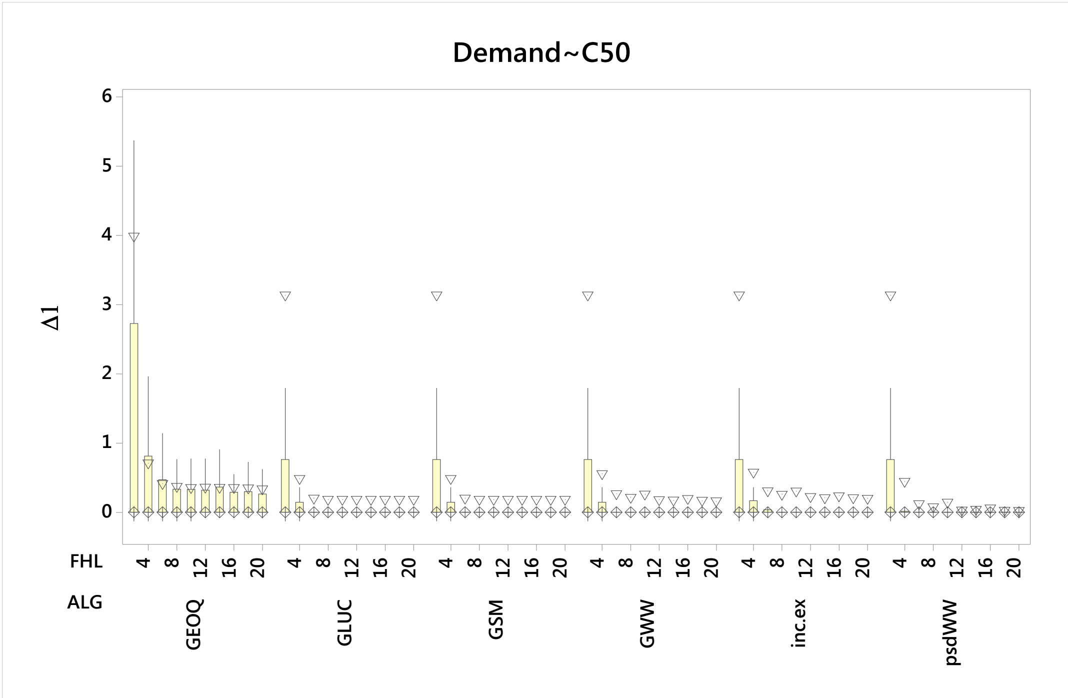


(a)

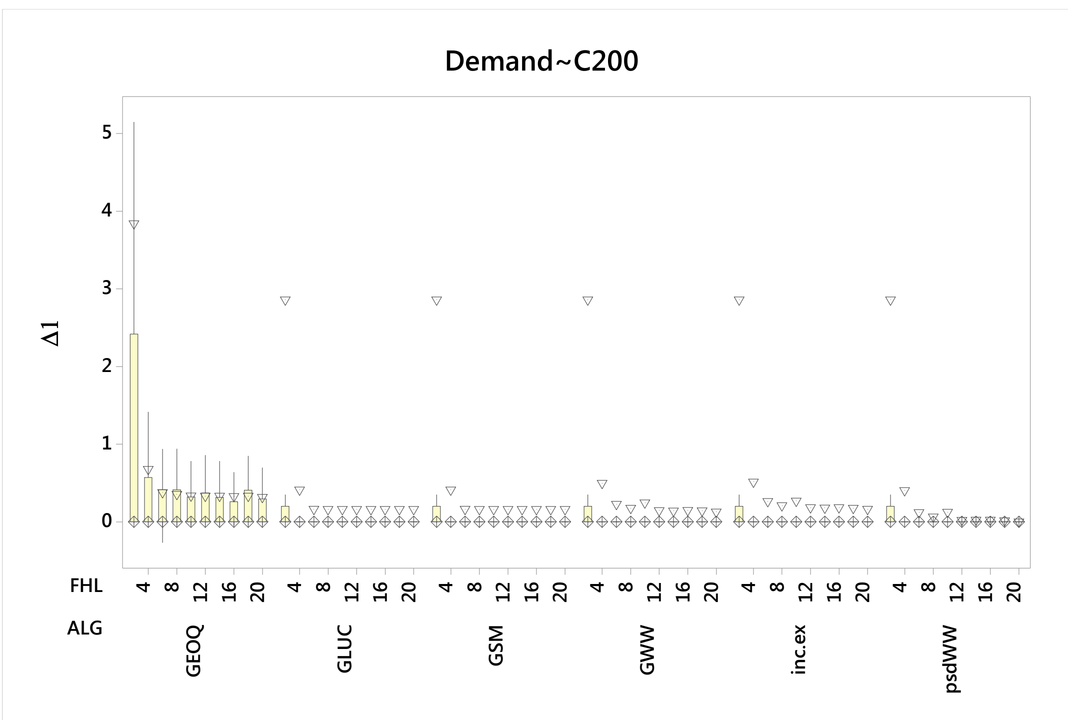


(b)

Figure 5.4: Box plots of Δ_1 values over all the instances for each algorithm in Normal demands (a): $\mu = 50$, (b) $\mu = 200$

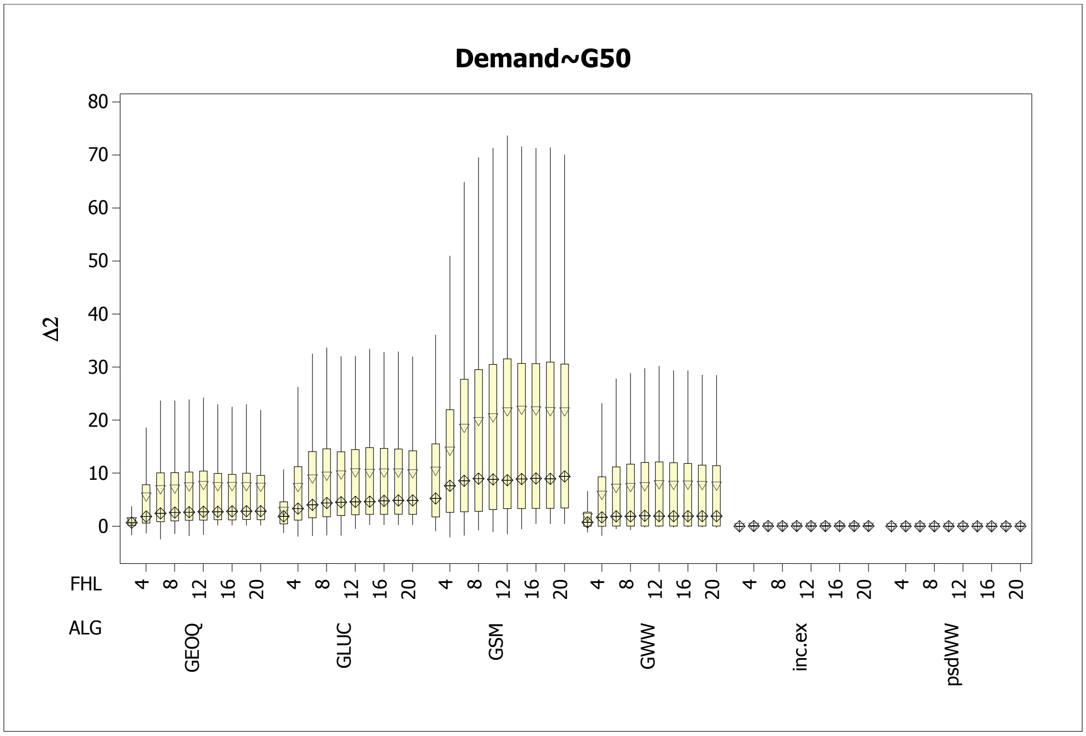


(a)

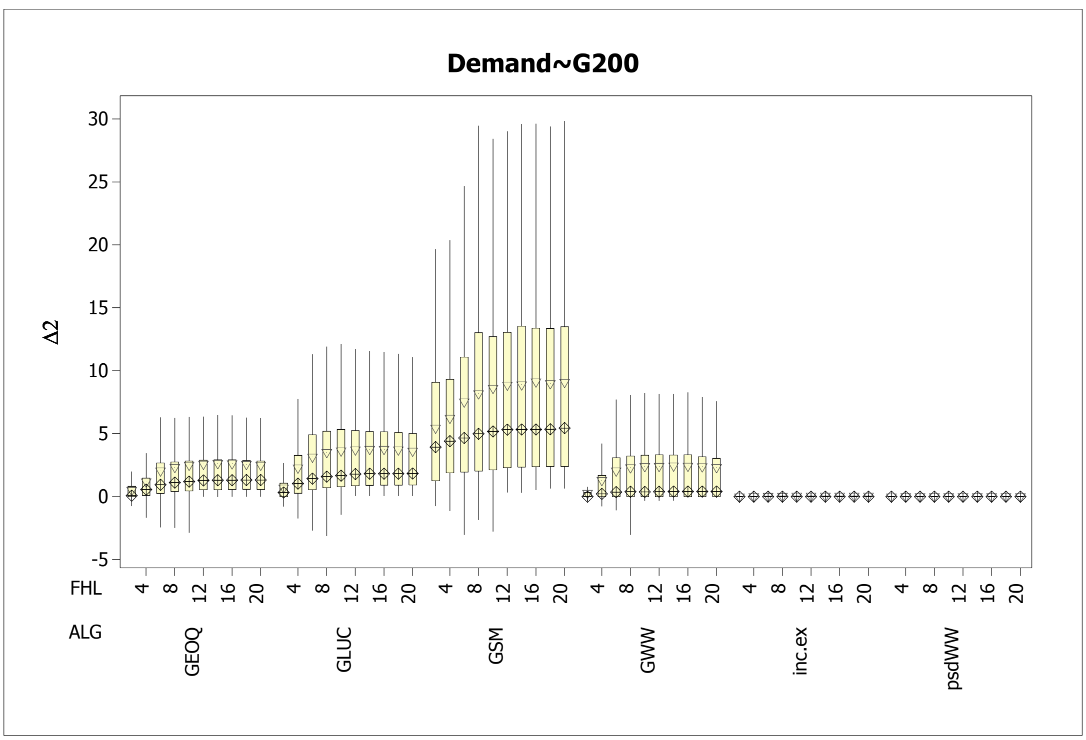


(b)

Figure 5.5: Box plots of Δ_1 values over all the instances for each algorithm in Constant demands (a): $\mu = 50$, (b) $\mu = 200$

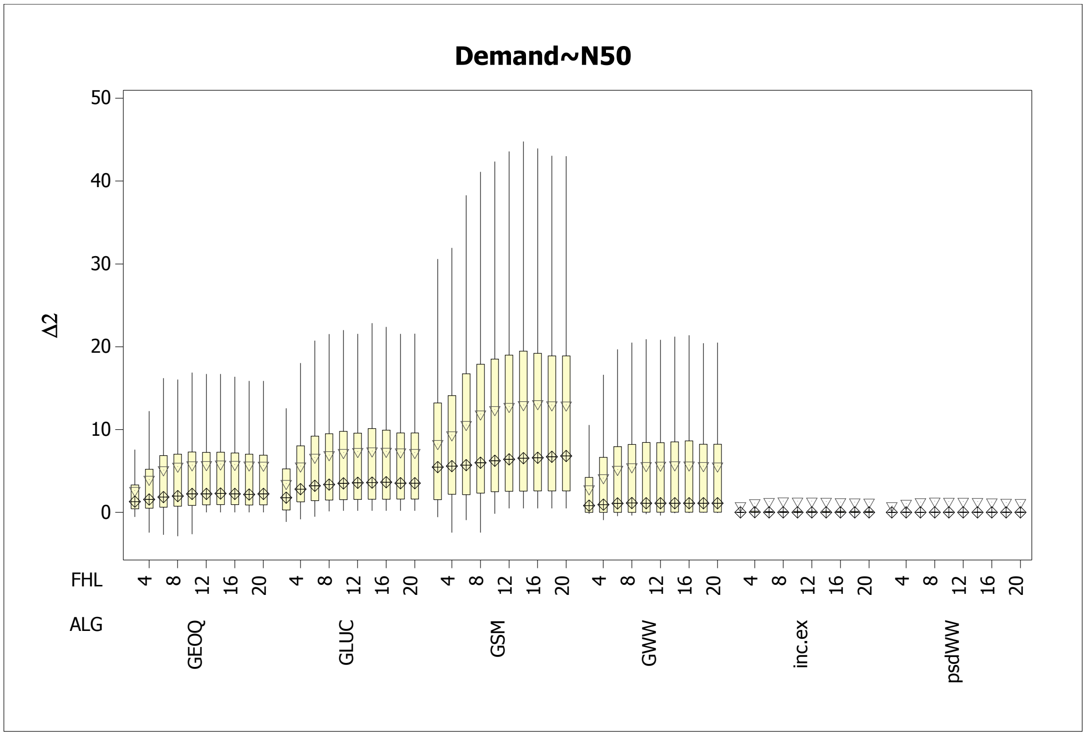


(a)

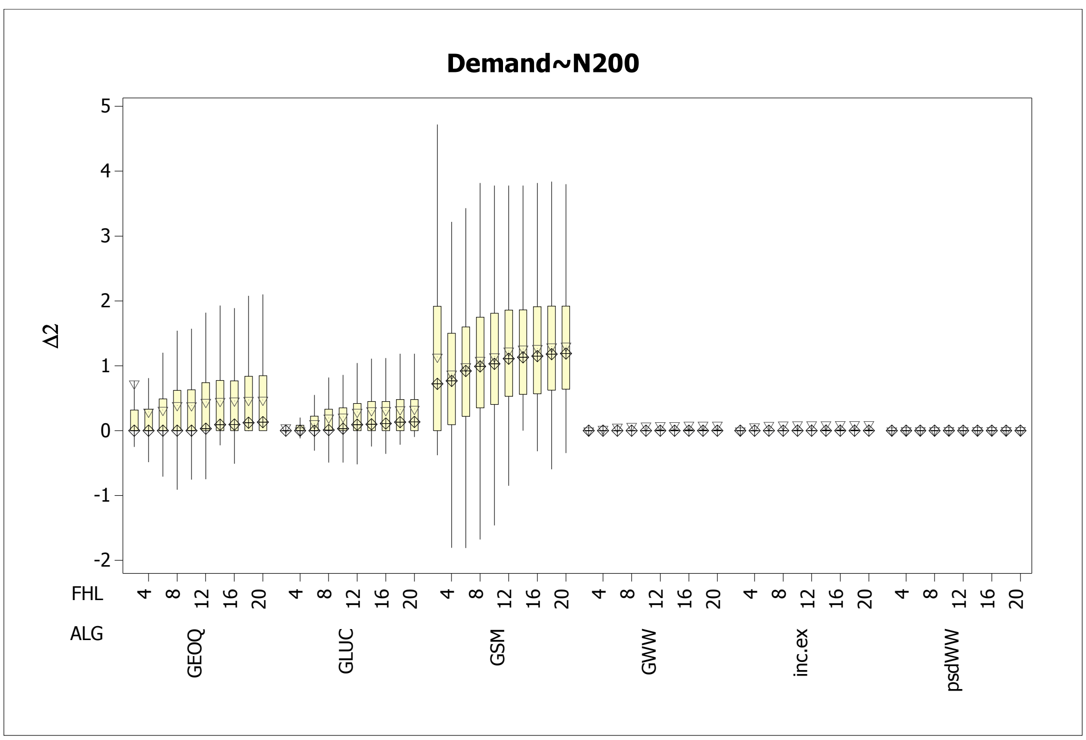


(b)

Figure 5.6: Box plots of Δ_2 values over all the instances for each algorithm in Gamma demands (a): $\mu = 50$, (b) $\mu = 200$

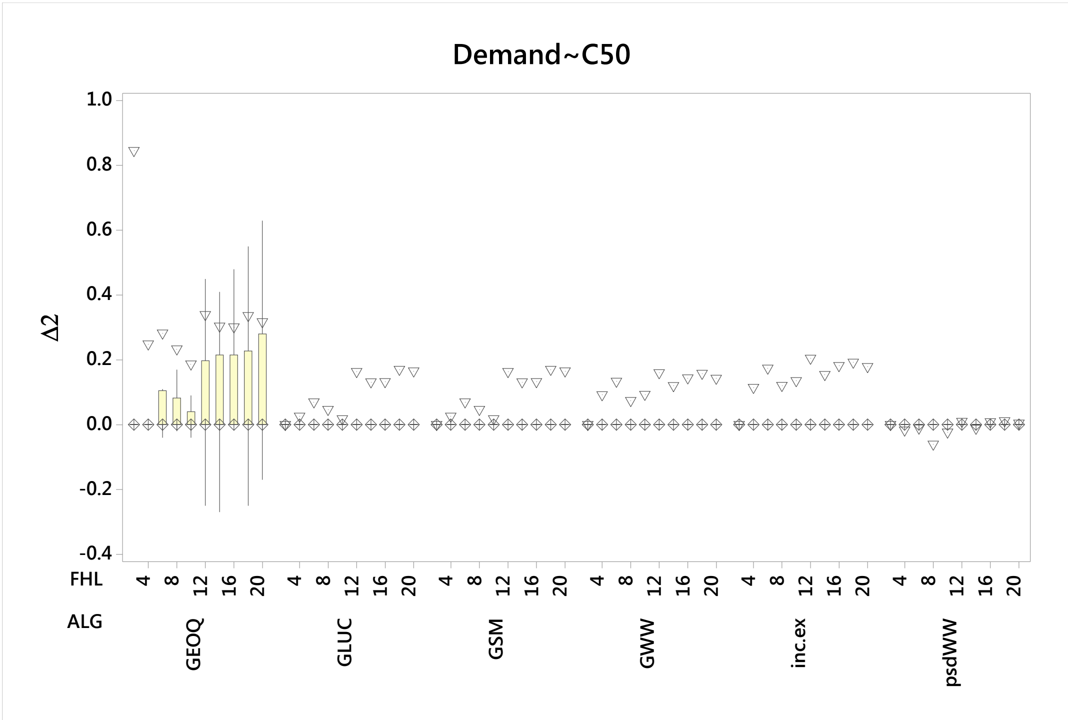


(a)

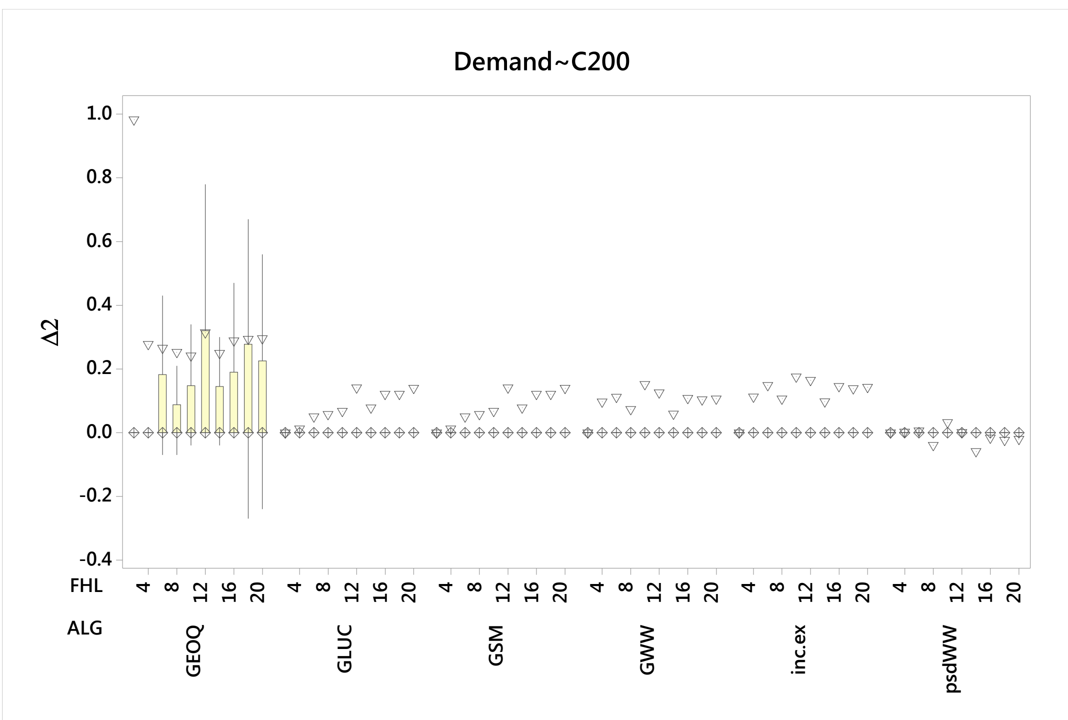


(b)

Figure 5.7: Box plots of Δ_2 values over all the instances for each algorithm in Normal demands (a): $\mu = 50$, (b) $\mu = 200$



(a)



(b)

Figure 5.8: Box plots of Δ_2 values over all the instances for each algorithm in Constant demands (a): $\mu = 50$, (b) $\mu = 200$

Table 5.9: ANOVA for effect of the factors on Δ_1

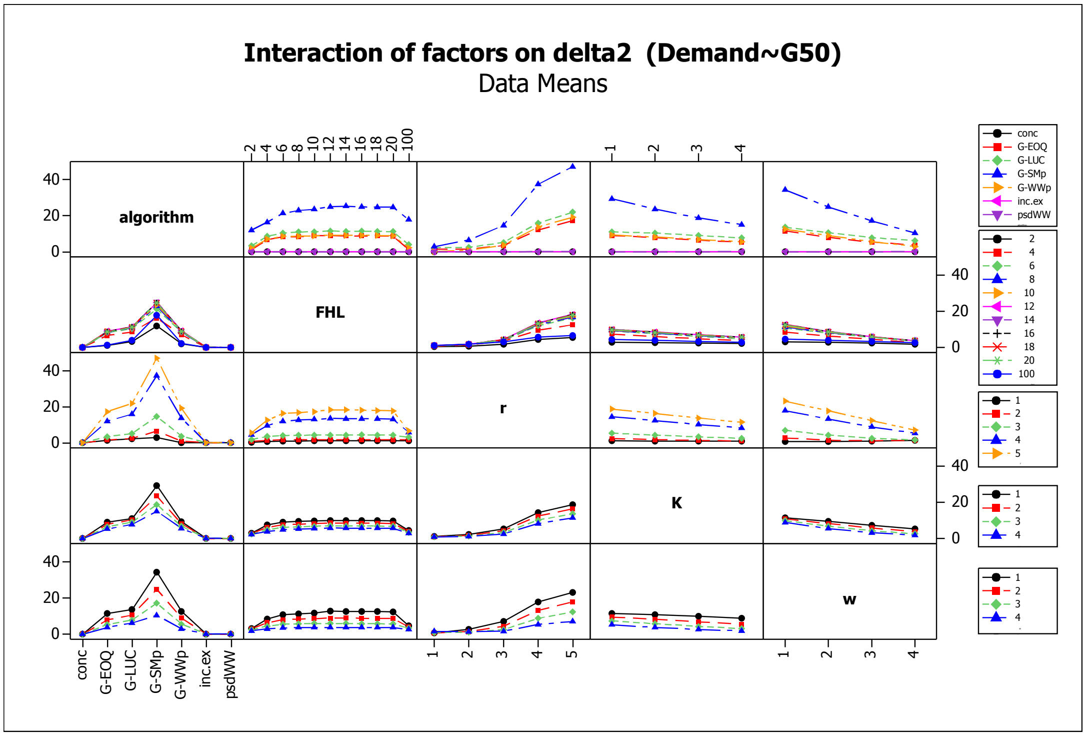
Source	Sum of Squares	df	Mean Square	F	P-value
dis	2065029.052	5	413005.8	7,940.28	0
algorithm	3291837.339	6	548639.6	10,547.92	0
FHL	30765.334	10	3076.533	59.148	0
K	98398.778	3	32799.59	630.591	0
r	2369772.721	4	592443.2	11,390.07	0
w	530110.887	3	176703.6	3,397.23	0
dis * FHL	66926.838	50	1338.537	25.734	0
algorithm * FHL	112719.953	60	1878.666	36.118	0
K * FHL	33469.164	30	1115.639	21.449	0
r * FHL	109283.366	40	2732.084	52.526	0
w * FHL	46797.551	30	1559.918	29.99	0
Error	13,444,481.05	258478	52.014		
Total	26,345,155.61	258720			

R Squared = .490 (Adjusted R Squared = .489)

Table 5.10: ANOVA for effect of the factors on Δ_2

Source	Sum of Squares	df	Mean Square	F	P-value
dis	1,563,350.75	5	312670.2	6,761.89	0
algorithm	3,228,939.27	6	538156.5	11,638.33	0
FHL	26,059.90	10	2605.99	56.358	0
K	124,329.43	3	41443.14	896.261	0
r	1,917,015.78	4	479253.9	10,364.48	0
w	441,724.08	3	147241.4	3,184.28	0
dis * FHL	105,382.59	50	2107.652	45.581	0
algorithm * FHL	123,633.27	60	2060.554	44.562	0
K * FHL	8,541.00	30	284.7	6.157	0
r * FHL	155,588.30	40	3889.707	84.12	0
w * FHL	53,388.93	30	1779.631	38.487	0
Error	11,952,031.95	258478	46.24		
Total	22,794,182.60	258720			

R Squared = .476 (Adjusted R Squared = .475)



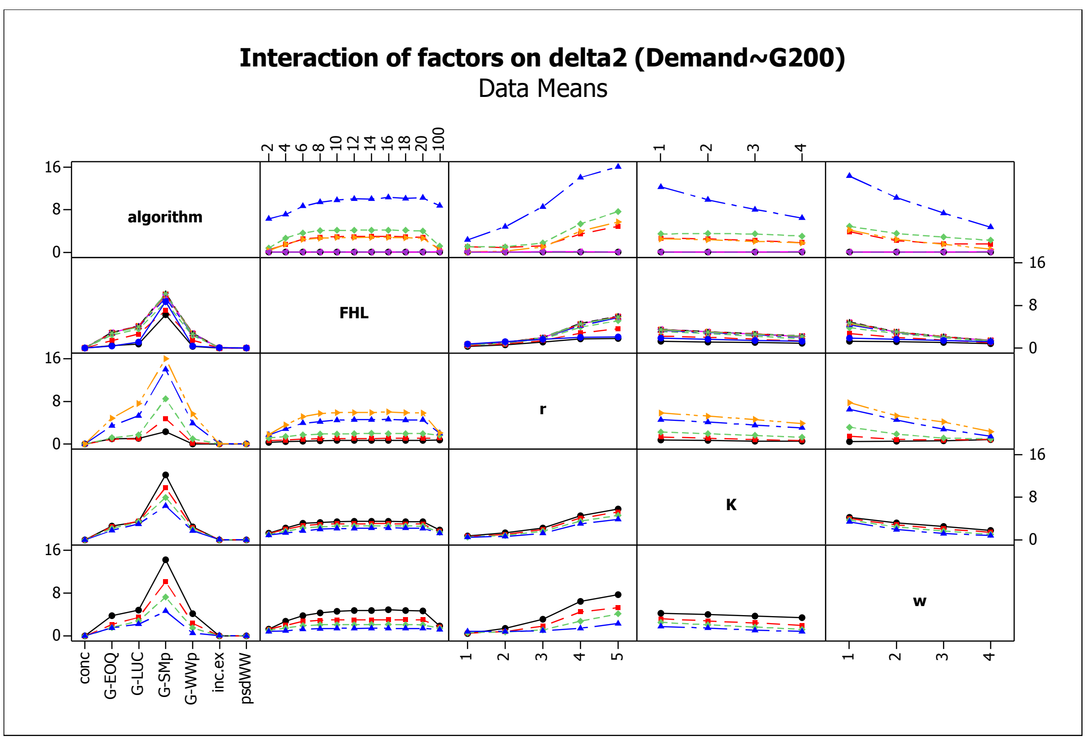


Figure 5.9: Pairwise interaction effect of the factors on Δ_2: Gamma demand

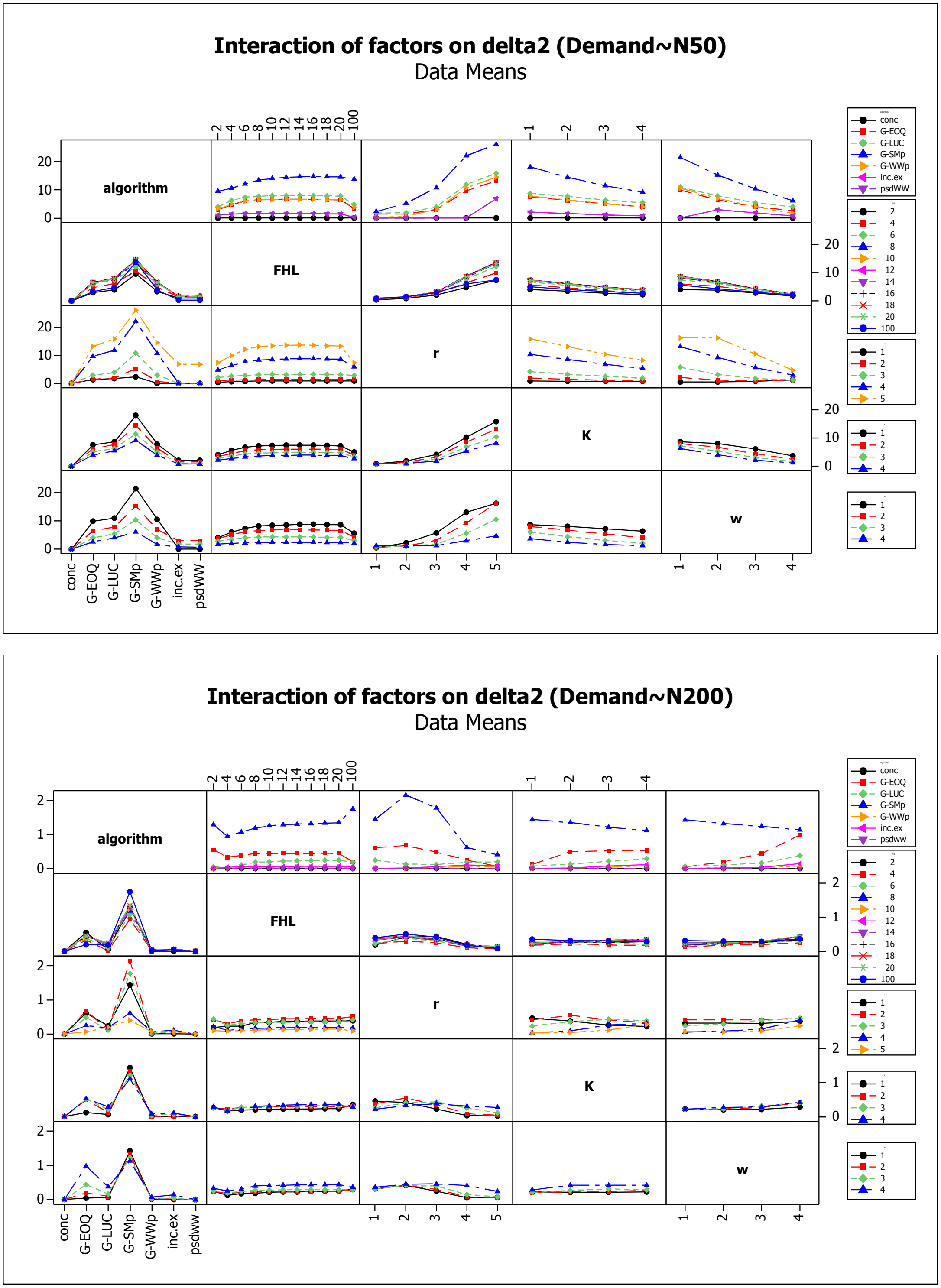
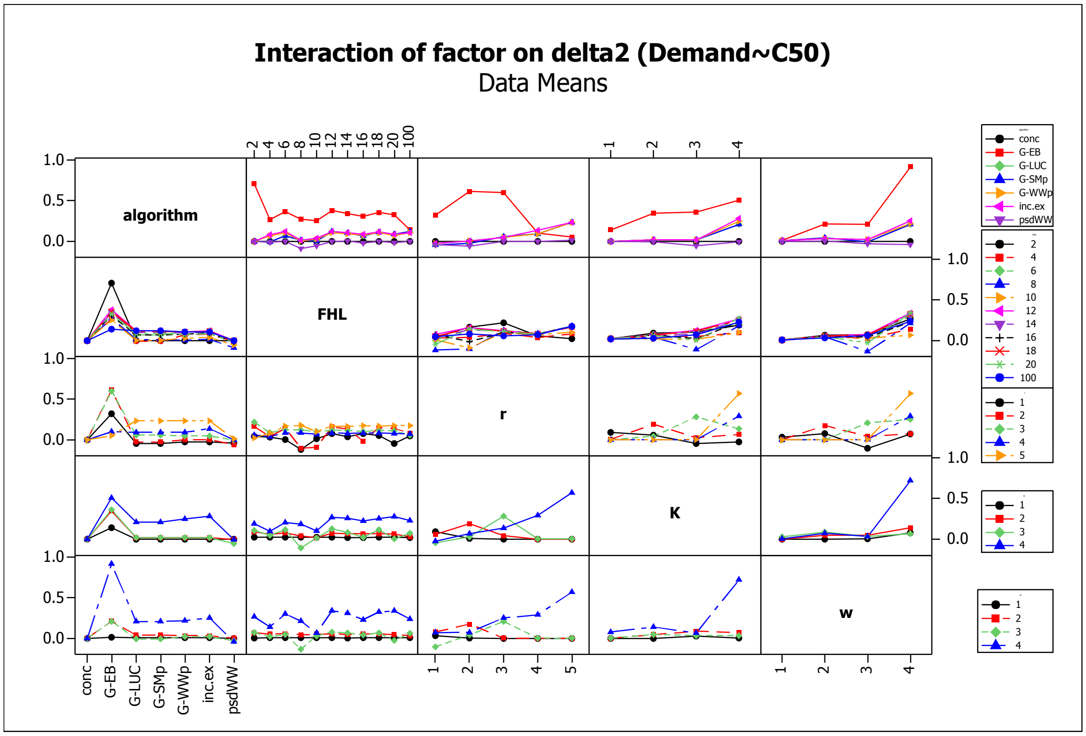


Figure 5.10: Pairwise interaction effect of the factors on Δ_2: Normal demand



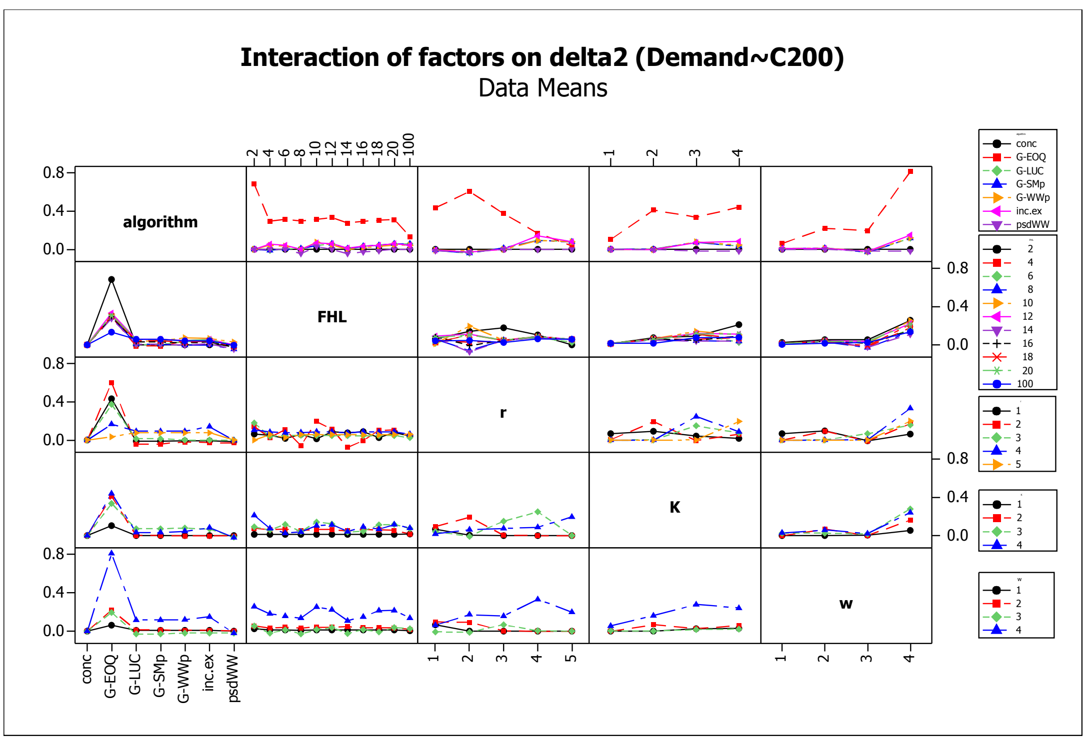


Figure 5.11: Pairwise interaction effect of the factors on Δ_2: Constant demand

Chapter 6

Lot sizing under carbon emission constraints and lost sales

In this chapter, we first extend the production planning model introduced in Chapter 3 with cumulative carbon emission constraints, and then extend it with possibility of lost sales in demand.

For the former one, we discuss two exact and two heuristic solution methods. The exact solution methods include quadratic reformulation and Benders decomposition. The heuristic approaches are Lagrangian relaxations of the carbon emission constraint, and iteratively use previously proposed algorithms in Chapter 3, with the multiplier updated at each iteration. Their difference lies in the way the multiplier is updated at each iteration.

From carbon emissions perspective, our study in this chapter is related to the work of Absi et al. [67] and Retel et al. [65]. Absi et al. [67] consider uncapacitated lot-sizing with multi-mode (different production modes with different unit cost and emission) under periodic, cumulative and rolling carbon emission constraints. All the cost functions are linear in their model and carbon emission is also a linear function of production quantity. They establish that the periodic constrained problem can be solved via existing DP algorithms while in the cumulative constrained problem is an NP-hard problem.

In the model of Retel et al. [65] carbon emission is a linear function of setup, inventory, and production quantity at each period. They assume concave production cost. They provide a fully polynomial time approximate algorithm and a Lagrangian heuristic which rests on Lagrangian relaxation of the carbon emission constraint and using Wagner-Whitin algorithm for the relaxed problem. Our work differs from theirs due to the Cobb-Douglas relation between inputs and the output quantity which results in a convex cost production cost function due to the assumption on the output elasticities. The nonlinearity and convexity of the production function leads to a more difficult problem in our setting. We propose two Lagrangian algorithms with different way of obtaining multiplier value. Also the relaxed problem in our setting is solved with our proposed G-heuristics for nonlinear cost lot sizing rather than direct use of Wagner-Whitin algorithm.

The remainder of this chapter starts with the carbon emission constrained model, followed by the solution approaches and a computational example. Then, in Section 6.2 the model with lost sales is studied which is followed by a numerical comparison of cap policies in 6.2.2.

6.1 Carbon constrained model with a total cap

We first start with introducing notations and assumptions of the model.
The production quantity is assumed to have a Cobb-Douglas form where $q_t = A_t \prod_{i=1}^{m} x_i^{\alpha_i}$ with A_t denoting the productivity factor or yield, x_i denoting the amount of resource i used in production process and α_i denoting the output elasticity of the resource i for $i = 1, ..., m$.
Each of the setup, productive resource usage and storage operations has a linear contribution in carbon emission possibly with different slopes. In our model, emission induced from production operation is a linear function of consumed resources. There are T periods of production, $t = 1, ..., T$; and m different resources, $i = 1, ..., m$. The firm manufactures only one product where its demand amount

in period t is known and equal to deterministic value of d_t. The fixed unit setup cost for any positive amount of production in period t is K_t. Carrying a unit product from period t to the next period costs h_t and the unit acquisition cost of resource i in period t is equal to c_{it}.

It is assumed that any setup for the production at period t causes to ζ_t amount of carbon emission. Storage of each unit of product at period t has γ_t amount of carbon emission and using each unit of resource i in period t results in β_{it} amount of carbon emission.

The aforementioned parameters and variables are summarized in the Table 6.1 and we refer to the mathematical formulation of our problem (PC) which is stated formally as follows.

Symbol	Definition
Parameters	
A_t	Production factor in period t
d_t	Demand in period t
K_t	Fixed setup cost in period t
h_t	Holding cost per unit of product in period t
c_{it}	Unit usage cost of resource i in period t
ζ_t	Emission amount due to a setup in period t
γ_t	Emission amount due to unit product holding in period t
β_{it}	Emission amount per unit usage of resource i in period t
α_i	Output elasticity of resource i
κ	Overall emission permit
κ_t	Emission permit in period t
Variables	
q_t	Production amount in period t
I_t	Inventory level at the end of period t
x_{it}	Quantity of resource i used in period t
y_t	0-1 variable which takes 1 if production occurs in period t

Table 6.1: Summary of the notations used in the mathematical model

$$
\begin{aligned}
\text{(PC)} \quad & \min \sum_{t=1}^{T} \Big(K_t y_t + h_t I_t + \sum_{i=1}^{m} c_{it} x_{it} \Big) && && \text{(6.1a)} \\
& s.t. \\
& I_0 = 0, && && \text{(6.1b)} \\
& I_t = I_{t-1} + q_t - d_t, && t = 1, ..., T && \text{(6.1c)} \\
& q_t = A_t \prod_{i=1}^{m} x_{it}^{\alpha_i}, && t = 1, ..., T && \text{(6.1d)} \\
& \sum_{t=1}^{T} \Big(\zeta_t y_t + \gamma_t I_t + \sum_{i=1}^{m} \beta_{it} x_{it} \Big) \leq \kappa && && \text{(6.1e)} \\
& q_t \leq \Big(\sum_{j=1}^{T} d_j \Big) y_t, && t = 1, ..., T && \text{(6.1f)} \\
& I_t, q_t \geq 0, y_t \in \{0, 1\} && t = 1, ..., T && \text{(6.1g)} \\
& x_{it} \geq 0, && t = 1, ..., T; i = 1, ..., m && \text{(6.1h)}
\end{aligned}
$$

This model is very similar to (3.2) in Chapter 3. The difference is that here we have an additional constraint, (6.1e), which corresponds to total carbon emission limit. This problem is NP-hard because even its uncapacitated counterpart is NP-hard. In the following section we discuss solution methods for this problem.

6.1.1 Exact solution approaches

6.1.1.1 Conic quadratic reformulation

As we have already observed in Chapter 4 and also discussed in Chapter 13 of [84], the performance of MINLP solvers for this mathematical programming method is poor and unreliable. Hence, we will convert the problem again into the form of the second order conic programming. We need to express constraint (6.1d) in cone representation form. To this aim, first we modify it to an inequality form $q_t \leq A_t \prod_{i=1}^{m} x_{it}^{\alpha_i}$. Output elasticity exponents can be written with rational numbers, $\alpha_i = \frac{a_i}{b_i}$ where a_i and b_i are prime with respect to each other. Let C_t be the least common multiple of the $b_1, \ldots b_m$ and $n = \lceil \log_2 C_t \rceil$. Then (6.1d) can be replaced with a m-block power-C_t defined in Chapter 4. Then, the SOCP reformulation of the problem can be obtained and be solved via CPLEX (or some other optimization packages).

6.1.1.2 Benders decomposition

The Benders decomposition algorithm is a partitioning approach for solving mixed integer linear and nonlinear programming problems. We refer the reader to [85] and [86] for the details and applications of Benders decomposition in mixed integer nonlinear problems. In general, the algorithm rests on partitioning the mathematical model into a master problem including complicating (usually integer) variables and a subproblem which is used to produce feasibility and optimality cuts to the master problem.

In problem (PC), taking y_t, q_t and consequently I_t as the first-stage variables, and x_{it} as the second-stage variables, then the projection of the problem (6.1) to the space of the main variables results in,

$$\min_{y,I,q} \sum_{t=1}^{T} \Big(K_t y_t + h_t I_t \Big) + \phi(y, I, q)$$

where constraints (6.1c),(6.1f) and (6.1g) hold and $\phi(y, I, q)$ is the following subproblem: $\phi(y, I, q) = \min_{x_{it}} \{ \sum_{t=1}^{T} \sum_{i=1}^{m} c_{it} x_{it} \}$ subject to (6.1e) and (6.1d).
The feasible region of this subproblem is convex and non-singular (i.e., interior points exists). Therefore, the strong duality holds and there is no duality gap.

Thus, $\phi(y, I, q)$ can be written as

$$\phi(y,I,q) = \max_{\mu_t,\lambda\geq 0}\Bigg\{\min_{x_{it}\geq 0}\Bigg\{\sum_{t=1}^{T}\Bigg(\sum_{i=1}^{m} c_{it}x_{it}\Bigg) - \mu_t A_t \prod_{i=1}^{m} x_{it}^{\alpha_i} + \sum_{t=1}^{T}\Bigg(\mu_t q_t + \lambda\Big(\zeta_t y_t + \gamma_t I_t + \sum_{i=1}^{m}\beta_{it}x_{it}\Big)\Bigg) - \lambda\kappa\Bigg\}\Bigg\} \tag{6.2}$$

Hence, the problem (6.1) is equivalent to

$$\text{(MP)} \quad \min_{q,y,I}\sum_{t=1}^{T}\Big(K_t y_t + h_t I_t\Big) + \eta \tag{6.3a}$$

$$s.t.$$

$$(6.1c), (6.1f), (6.1g)$$

$$\eta \geq \min_{x_{it}\geq 0}\Bigg\{\sum_{t=1}^{T}\Bigg(\sum_{i=1}^{m} c_{it}x_{it} + \lambda\beta_{it}x_{it}\Bigg) - \mu_t A_t \prod_{i=1}^{m} x_{it}^{\alpha_i}\Bigg\} + \sum_{t=1}^{T}\Big(\mu_t q_t + \lambda\Big(\zeta_t y_t + \gamma_t I_t\Big)\Big) - \lambda\kappa, \quad \forall \mu_t, \lambda \tag{6.3b}$$

Since a large number of constraints (6.3b) will not be binding at the optimality, the Benders decomposition algorithm solves the MP by ignoring all but a few of them. Hence, for a given $(y, q, I) = (y^n, q^n, I^n)$ at iteration n the optimal objective value of after solving (SP) and obtaining dual variables μ_t^n and λ^n, the dual objective can be found as,

$$\phi(y^n,q^n,I^n) = \min_{x_{it}\geq 0}\Bigg\{\sum_{t=1}^{T}\Bigg(\sum_{i=1}^{m} c_{it}x_{it} + \lambda^n\beta_{it}x_{it}\Bigg) - \mu_t^n A_t \prod_{i=1}^{m} x_{it}^{\alpha_i}\Bigg\} + \sum_{t=1}^{T}\Big(\mu_t^n q_t^n + \lambda^n\Big(\zeta_t y_t^n + \gamma_t I_t^n\Big)\Big) - \lambda^n\kappa \tag{6.4}$$

Also, constraint (6.3b) can be rewritten for iteration n in the form of

$$\eta \geq \min_{x_{it} \geq 0} \left\{ \sum_{t=1}^{T} \left(\sum_{i=1}^{m} c_{it} x_{it} + \lambda^n \beta_{it} x_{it} \right) - \mu_t^n A_t \prod_{i=1}^{m} x_{it}^{\alpha_i} \right\}$$
$$+ \sum_{t=1}^{T} \left(\mu_t^n q_t + \lambda^n \left(\zeta_t y_t + \gamma_t I_t \right) \right) - \lambda^n \kappa \tag{6.5}$$

Hence, by plugging (6.4) in (6.5), the optimality cut for the Benders master problem is obtained as

$$\eta \geq \sum_{t=1}^{T} \mu_t^n (q_t - q_t^n) + \lambda \zeta_t (y_t - y_t^n) + \lambda \gamma_t (I_t - I_t^n) \tag{6.6}$$

Solving the subproblem

A nice property of this approach is that we can solve the subproblem (SP) analytically.

$$\text{(SP)} \quad \phi(\hat{y}, \hat{I}, \hat{q}) = \min_{x_{it}} \sum_{t=1}^{T} \sum_{i=1}^{m} c_{it} x_{it} \tag{6.7a}$$

$$s.t.$$

$$A_t \prod_{i=1}^{m} x_{it}^{\alpha_i} = \hat{q}_t, \quad t = 1, ..., T \tag{6.7b}$$

$$\sum_{t=1}^{T} \sum_{i=1}^{m} \beta_{it} x_{it} \leq \kappa - \sum_{t=1}^{T} \left(\zeta_t \hat{y}_t - \gamma_t \hat{I}_t \right) \tag{6.7c}$$

$$(6.1g)$$

Constructing the Lagrangiang function and applying the KKT conditions, we find:

$$x_{it} = \frac{\alpha_i \mu_t}{c_{it} + \lambda \beta_{it}} \hat{q}_t = \frac{\alpha_i}{c_{it} + \lambda \beta_{it}} \left[\frac{\hat{q}_t}{A_t} \prod_{j=1}^{m} \left(\frac{c_{it} + \lambda \beta_{it}}{\alpha_i} \right)^{\alpha_i} \right]^r \tag{6.8}$$

where $r = 1/\sum_{k=1}^{m} \alpha_k$. Let define the function $f(\lambda)$ as

$$f(\lambda) := \sum_{t=1}^{T} \sum_{i=1}^{m} \left(\frac{\beta_{it}\alpha_i}{c_{it} + \lambda\beta_{it}} \left[\frac{\hat{q}_t}{A_t} \prod_{j=1}^{m} \left(\frac{c_{it} + \lambda\beta_{it}}{\alpha_i} \right)^{\alpha_i} \right]^r \right) \tag{6.9}$$

Then, λ is either zero –if constraint (6.7c) is loose– or it solves the following equation provided that (SP) is feasible for the given vector of $(\hat{y}, \hat{q}, \hat{I})$.

$$f(\lambda) - \kappa - \zeta_t \hat{y}_t - \gamma_t \hat{I}_t = 0 \tag{6.10}$$

The function $f(\lambda)$ can be interpreted as the emission amount in terms of the shadow price of the carbon cap and it is a decreasing function with the upper bound $f_U = f(0)$ and the lower bound $f_L = f(\infty)$ for the corresponding $\hat{q}$ vector. Hence, if in an iteration (SP) is infeasible, then we add the feasibility constraint

$$\sum_{t \in \mathbb{P}} \zeta_t y_t + \gamma_t I_t \leq \kappa - f_L^n \tag{6.11}$$

where f_L^n denotes the lower bound of the subproblem at iteration n and $\mathbb{P}$ is the set of production periods at the most recent solution.

6.1.2 Heuristic solution approaches

By Lagrangian-relaxation of the carbon emission constraint in (PC) we obtain the following model (M2)

$$\min \sum_{t=1}^{T} \Big((K_t + \lambda\zeta_t) y_t + (h_t + \lambda\gamma_t) I_t + \sum_{i=1}^{m} (c_{it} + \lambda\beta_{it}) x_{it} \Big) - \lambda\kappa \tag{6.12a}$$

$s.t.$

$(6.1c), (6.1d), (6.1f), (6.1g), (6.1h)$

For a given optimal vector of q_t^* and λ^* the optimal value of the resource variables can be obtained by the following model (SP):

$$\min \sum_{t=1}^{T} \sum_{i=1}^{m} (c_{it} + \lambda \beta_{it}) x_{it} \tag{6.13a}$$

$$q_t = A_t \prod_{i=1}^{m} x_{it}^{\alpha_i}, \quad t = 1, ..., T \tag{6.13b}$$

$$x_{it} \geq 0 \quad t = 1, ..., T; i = 1, ..., m \tag{6.13c}$$

By checking KKT necessary conditions for optimality we obtain:

$$x_{it} = q_t^r \underbrace{\left(\frac{\alpha_i}{c_{it} + \lambda \beta_{it}}\right) \left[\frac{1}{A_t} \prod_{j=1}^{n} \left(\frac{\alpha_j}{c_{jt} + \lambda \beta_{jt}}\right)^{-\alpha_j}\right]^r}_{p_{it}(\lambda)}. \tag{6.14}$$

Replacing x_{it} variables in the Lagrangian-relaxed model with (6.14), we obtain a single item lot sizing model with nonlinear power function as discussed in Chapter 3. Therefore, it can be solved with the previously proposed algorithms for a given λ. Hence, we just need to find a way to obtain the Lagrangian multiplier. To this aim, we propose two iterative procedures in the following sections. The procedure rests on updating lower and upper bounds of the Lagrangian multiplier with a bisection method. The second procedure benefits from the implicit relation between production quantities and the Lagrangian multiplier.

6.1.2.1 Bisection algorithm for finding the Lagrangian multiplier (λ)

The procedure depicted in Algorithm 4 stops either by convergence of λ or by reaching the iteration limits which occurs usually when the problem is not feasible. This algorithm uses a bisection subroutine to find the Lagrangian multiplier λ by updating its upper and lower bounds at each iteration. For a given value of λ, the complexity order of the production planning procedure is $O(T^2)$ for two-step heuristics, $O(T^3)$ for psdWW and $O(T^4)$ for inc.ex DP based algorithms. Hence the total complexity order of the procedure above is $O(\log(\lambda_U)T^2)$, $O(\log(\lambda_U)T^3)$ or $O(\log(\lambda_U)T^2)$ based on the production planning algorithm. In our numerical tests we have set the relative convergence threshold of λ as $\epsilon_2 = 0.0001$. Also we have set the maximum 1000 number of iterations as a default for our algorithm.

Algorithm 4 Bisection algorithm to find λ

$\lambda_L, \lambda \leftarrow 0$, $\lambda_U \leftarrow \infty$, $iter \leftarrow 0$
Maxiter $\leftarrow 1000, \epsilon_2 \leftarrow 0.0001$
$Solution \leftarrow$ Solve(uncapacitated) ▷ by any of the heuristic or exact approaches
$\epsilon_1 \leftarrow TE/TC$ ▷ TE and TC are total emission and cost in the current solution
if $Solution$ is NOT feasible **then**
  repeat
    $iter \leftarrow iter + 1$
    if $Solution$ is feasible **then**
      $\lambda_U \leftarrow \min(\lambda, \lambda_U)$
    else
      $\lambda_L \leftarrow \max(\lambda, \lambda_L)$
    end if
    if $\lambda_U = \infty$ **then**
      $\lambda \leftarrow \lambda + \epsilon_1$
    else
      $\lambda \leftarrow (\lambda_U + \lambda_L)/2$
    end if
    update parameters with new λ
    $Solution \leftarrow$ Solve(uncapacitated)
  until $(\frac{\lambda_U - \lambda_L}{\lambda_L} < \epsilon_2)$ OR (iter=Maxiter)
end if

6.1.2.2 Adjustive algorithm for finding the Lagrangian multiplier (λ)

An alternative way to evaluate the Lagrangian multiplier, λ, is to use the function defined in (6.9). For a given production plan $\vec{q}$ –obtained by any of the heuristics– the root of the equation $f_{\vec{q}}(\lambda) = \hat{\kappa}$, $(\hat{\kappa} := \kappa - \sum_{t=1}^{T} \zeta_t \hat{y}_t - \gamma_t \hat{I}_t)$, evaluates the corresponding shadow price of the cap constraints and consequently the Lagrangian relaxed parameters in the objective can be updated with the corresponding λ.

Then, a new vector of production quantities, $\vec{q'}$, is obtained and used for evaluating λ. This procedure is continued until λ converges to a value. There is no guarantee for the convergence of λ in this method and we have observed in our preliminary numerical tests that in some instances after a number of iterations, λ value sticks and alternates among two or more numbers preventing the procedure to stop. To avoid such situations we modify the stopping rule of the adjustive Lagrangian iteration algorithm to be the observation of a previously visited λ. That is, the procedure stops once it obtains a λ value which has already obtained in any previous iteration. The corresponding pseudo-code of this algorithm is given in Algorithm 5

Algorithm 5 Adjustive algorithm to find λ

$iter \leftarrow 0$, $S \leftarrow \{0\}$, $\lambda \leftarrow 0$, $Stop \leftarrow$False
repeat
    $iter \leftarrow iter + 1$
    $\vec{q} \leftarrow$ Solve(uncapacitated)
    $\lambda \leftarrow$ Solution of $f_{\vec{q}}(\lambda) = \hat{\kappa}$
    if $\lambda \in S$ **then**
        $Stop \leftarrow$ True
    else
        $S \leftarrow \{\lambda\} \cup S$
    end if
until $Stop$

As it will be shown in numerical results, this procedure is about 10 times faster than the previous one and it stops in almost three or four iterations for the examples we have tested.

6.1.3 A real life illustrative example

In this section, we examine the performance of our heuristic algorithms against the exact solution obtained by the CPLEX solver. We use a real life problem related to tomato production discussed in [51]. We consider four of resources considered in their model: fertilizer, Chemical, Machinery and Water. The production quantity of tomato follows a Cobb-Douglas function with these resource

Table 6.2: Parameters setting of the numerical case

Parameter name	cost factor	emission factor	output elasticity
Setup	$\{400, 200\}$	60	-
storage	1	0.1	-
Fertilizer	0.1278	0.0197	0.24
Chemical	0.5929	0.0504	0.19
Machinery	1.7172	0.0398	0.15
Water	0.2689	0.2951	0.23

quantities. The output elasticities are given in [51]. We consider a 100-period production planning and assume that an additive carbon constraint is imposed. The unit emission amount of each of the resource are taken from [87]. We have generated 10 replications of a normally distributed demand with mean $\mu = 50$ and standard deviation $\sigma = 40$. Two levels of setup cost are considered. All the parameters are summarized in Table 6.2.

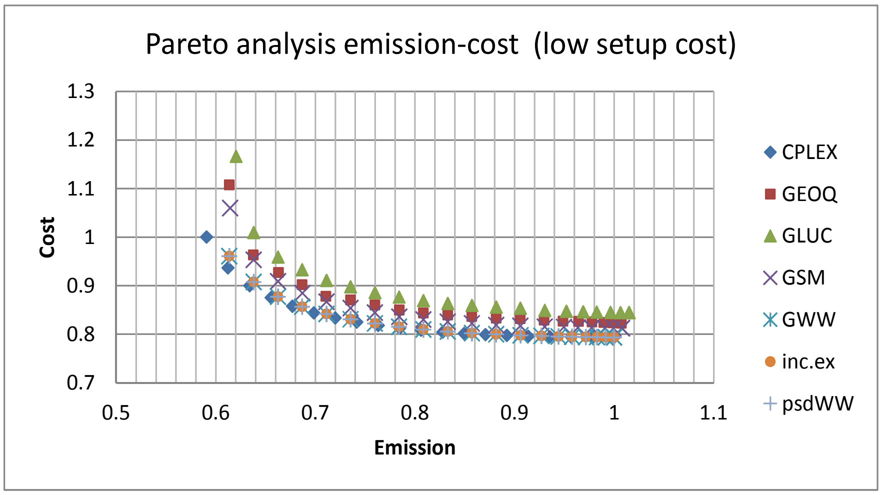
The problem is first solved by CPLEX without emission cap and then based on the resulted total carbon emission amount, a cap constraint with tightness factors $\delta \in \{5\%, 10\%, 15\%, 20\%\}$ is imposed. In such cases $\kappa = (1 - \delta) \times TE_0$ where TE_0 is the total emission in lack of the cap constraint. Computation times, deviation percentage from the optimality and carbon emission deviation percentages are provided in Table 6.3 for each Lagrangian approach and each embedded heuristic, separately. Comparing the top half of the table–which corresponds to the bisection Lagrangian– with the bottom half of the table –which corresponds to the adjustive Lagrangian– we observe that the second approach catches the same solutions with less number of iterations and consequently less CPU time.

Next, we compare these algorithms in the bi-objective fashion. To this aim, the emission function regardless of the costs, and the cost objective regardless of the emission are minimized with CPLEX optimizer. Then the emission cap levels were set into 20 points between the emission amounts of the two extreme cases. Afterwards, the minimum cost corresponding to each of the cap values are obtained by each of the heuristics and the results are pictured in Figure 6.1 as the *Pareto frontier* of the heuristics in this case study. In this figure both emission and cost amounts are normalized by dividing them to the corresponding

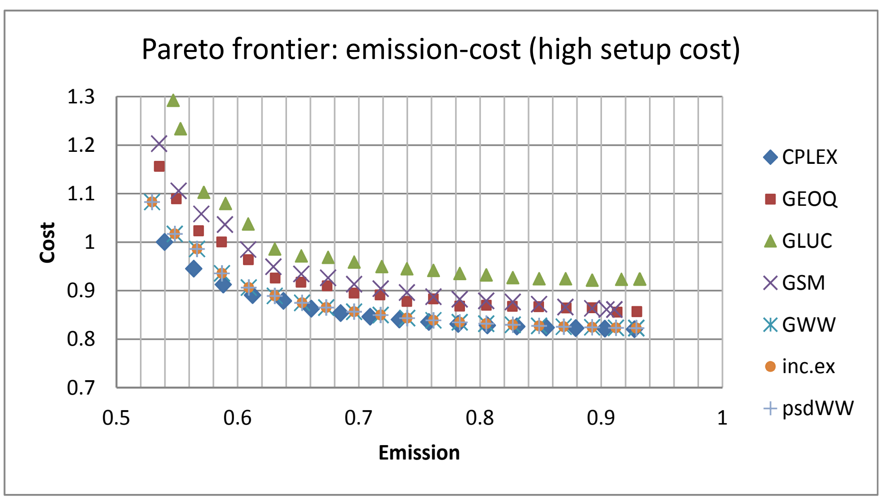
Table 6.3: Average performance of the Lagrangian heuristics

	cpu (s)			obj(%)			emission(%)			Iter.
	min	max	ave.	min	max	ave.	min	max	ave.	ave.
CPLEX	21.03	224.28	86.97	-	-	-	-	-	-	-
Bisection										
GEOQ	0.01	0.31	0.20	1.69	8.21	4.39	-1.55	0.16	-0.09	22
GLUC	0.01	0.68	0.35	3.73	17.09	9.35	-3.60	0.02	-0.14	22
GSM	0.01	0.34	0.22	1.86	7.01	3.87	-6.38	0.00	-0.43	22
GWW	0.01	7.01	0.22	0.00	0.00	0.00	0.00	0.00	0.00	19
incex	0.97	28.49	19.83	0.00	0.00	0.00	0.00	0.00	0.00	19
psdWW	0.17	4.52	3.35	0.00	0.00	0.00	0.00	0.00	0.00	19
Adjustive										
GEOQ	0.00	0.26	0.07	1.69	8.21	4.39	-1.55	0.42	-0.09	4
GLUC	0.00	0.32	0.06	3.73	17.09	9.35	-3.60	0.00	-0.14	3
GSM	0.00	0.07	0.04	1.86	7.01	3.87	-6.38	0.00	-0.43	3
GWW	0.00	7.01	0.05	0.00	0.00	0.00	0.00	0.00	0.00	2
incex	0.00	7.58	4.68	0.00	0.00	0.00	0.00	0.00	0.00	2
psdWW	0.00	1.64	0.82	0.00	0.00	0.00	0.00	0.00	0.00	2

maximum cost and emission is obtained by CPLEX. Then they are averaged over the demand realizations. Both objectives are desired to be minimized, thus the closer to the origin, the better frontier we have. We observe in Figure 6.1 that the lowest frontier belongs to the CPLEX output and inc.ex, psdWW, GWW are very close to it while GEOQ, GSM and GLUC follow them despite with a small gaps. Note that the cost parameter in this example has a low convexity level which mostly resulted in the solutions in which demand integrality holds and consequently, GWW appeared with a good performance.



(a)



(b)

Figure 6.1: Pareto frontier of the solution algorithms for the bi-objective model. (a) low setup cost (b) high setup cost

6.2 Carbon constrained model with lost sales and different cap policies

In this section, we extend the problem (PC) in the previous section from two aspects: *(i)* we allow lost sales and *(ii)* we consider different types of carbon constants, as well. The exact solution methods discussed in the previous section is applicable here. However, our purpose here is to investigate the effect of different carbon cap policies on the optimal response variables for a cost minimizer manufacturer. In particular, we have considered three main response variables including total cost, total emission and unmet lost demand. For the carbon emission constraint we have considered four main policies. The difference between the policies arises from the way that periods are selected to adhere to a carbon emission cap. Each carbon emission cap is imposed to l number of consecutive periods which we refer to as *block*. These blocks may have periods in common (rolling), or they may be disjoint without any carryover of emission allowance (cluster or seasonal). Let J denote the set of periods which blocks start over the planning horizon. The mathematical formulation of our problem which we refer to as (PLC) is given here:

$$\text{(PLC)} \quad \min \sum_{t=1}^{T} \Big(K_t y_t + h_t I_t + p_t L_t + \sum_{i=1}^{n} c_{it} x_{it} \Big) \tag{6.15a}$$

s.t.

$$I_0 = 0, \tag{6.15b}$$

$$I_t = I_{t-1} + q_t - d_t - L_t, \qquad t = 1, ..., T \tag{6.15c}$$

$$L_t \leq d_t, \qquad t = 1, ..., T \tag{6.15d}$$

$$q_t = A_t \prod_{i=1}^{n} x_{it}^{\alpha_i}, \qquad t = 1, ..., T \tag{6.15e}$$

$$\sum_{t=j}^{j+l-1} \Big(\zeta_t y_t + \gamma_t I_t + \sum_{i=1}^{m} \beta_{it} x_{it} \Big) \leq \kappa_j, \qquad \forall j \in J \tag{6.15f}$$

$$q_t \leq M y_t, \qquad t = 1, ..., T \tag{6.15g}$$

$$q_t, I_t, L_t \geq 0, y_t \in \{0, 1\}, \qquad t = 1, ..., T \tag{6.15h}$$

$$x_{it} \geq 0, \qquad i = 1, ..., m, t = 1, ..., T \tag{6.15i}$$

In the formulation above L_t denotes the amount of demand loss in period t where each unit costs p_t. Constraint (6.15f) represents carbon cap policies which is characterized by parameter l and indices set J as shown in Table 6.4. The differences between (PLC) model here and (PC) in the previous section, are in the existence of L_t here and the structure of the carbon emission constraint in 6.15f.

Policy	block length	J
cumulative	T	$\{1\}$
Rolling	l	$\{1, 2, \ldots, T-l+1\}$
Cluster (seasonal)	l	$\{1, l, 2l+1, \ldots, T-l+1\}$
periodic	1	$\{1, \ldots, T\}$

Table 6.4: Cap policy characterization

6.2.1 Some structural results regarding to lost sales

In this section we investigate the structure of lost sales in an optimal solution regardless of the carbon constraint. Following the same procedure as in Section 3.2.1, (PLC) can be reformulated in the following (PL) model which is a lot sizing problem with a nonlinear power-like convex production cost and possible lost sales.

$$\text{(PL)} \quad \min \sum_{t=1}^{T} \Big(K_t \mathbb{1}_{\{q_t>0\}} + h_t I_t + w_t q_t^r + p_t L_t \Big) \tag{6.16a}$$

s.t.

$$\sum_{i=1}^{t} q_s + L_s \geq \sum_{i=1}^{t} d_s, \qquad t = 1, ..., T \tag{6.16b}$$

$$0 \leq L_t \leq d_t, \qquad t = 1, ..., T \tag{6.16c}$$

$$q_t \geq 0, \qquad t = 1, ..., T. \tag{6.16d}$$

For the linear counterpart of the problem it is shown by Aksen et al. [88] that

Table 6.5: Parameter of the counter example

(d_1, d_2)	$(\alpha_1,\alpha_2,\alpha_3,\alpha_4)$	(c_1, c_2, c_3, c_4)	(p_1, p_2)	κ
$(10, 100)$	$(0.04, 0.04, 0.12, 0.2)$	$(0.1, 0.2, 0.4, 0.5)$	$(0.25, 1)$	∞

in an optimal solution the following results hold:
(i) $I_t^* \times q_t^* = 0, \forall t.$
(ii) $q_t^* \times L_t^* = 0, \forall t.$
(iii) $L_t^* \times (d_t - L_t^*) = 0, \forall t.$
Here we show by a counter example that these results do not necessarily hold in the presence of convex cost.

Example. Consider a 2-period problem in (PLC) form with the parameters given in Table 6.5 The optimal production and loss amounts are $(q_1^*, q_2^*, L_1^*, L_2^*) = (8, 47, 9.5, 10, 45)$. In both periods $q_t \times L_t \neq 0$; In period 2, $L_t(d_t - L_t) \neq 0$ and $I_{t-1} \times q_t \neq 0$ and therefore, none of the optimality properties of linear cost problems holds here.

Lemma 4 *If* $p_i + \sum_{s=i}^{j-1} h_s > p_j$ *for all* i, j, $1 \leq i < j \leq T$, *then in an optimal solution of (PL) the following holds within each generation* $\langle u, v \rangle$*:*
If $L_t > 0$ *for an arbitrary* t, $u \leq t < v$ *then* $L_s > 0$ *for* $\forall s, u \leq t < s \leq v$.

Proof Suppose to the contrary that, in an optimal solution, at least one of the L_s, $t < s \leq v$ is zero. Without loss of generality, suppose that $L_{t+1} = 0$ and $L_s > 0, \quad t+1 < s \leq v$. Consider the new solution: $L_t' = L_t - 1$, $L_{t+1}' = L_{t+1} + 1 = 1$, $I_t' = I_t - 1$. Then the cost change in the new solution is $p_{t+1} - p_t - h_t < 0$. The negativity of the last expression is due to the assumption and it implies that the solution is improved which contradicts with the optimality assumption. □

Proposition 3 *Suppose that* $p_i + \sum_{s=i}^{j-1} h_s > p_j$ *for all* i, j, $1 \leq i < j \leq T$. *Then in an optimal solution of (PL) the following equality holds:*
$I_t \times L_t \times (d_t - L_t) = 0, \forall t$

Proof Invoking inventory decomposition property it suffices to show that the result holds within each arbitrary generation $\langle u, v\rangle$. Let S be the set of production periods within this generation. Using Lagrangian relaxation of the constraints we come up with the following Lagrangian function:

$$\mathcal{L}_2 := \sum_{t\in S}\left(K_t + w_t q_t^r\right) + \sum_{t=u}^{v}\Big[p_t L_t + (h_t - \lambda_t)(\sum_{i=u}^{t} q_i - d_i + L_i) \\ + \mu_t(L_t - d_t) - \gamma_t L_t - \theta_t q_t\Big]. \quad (6.17)$$

Consequently, from the first order optimality rules, we must have

$$\frac{\partial \mathcal{L}_2}{\partial L_i} = \sum_{s=i}^{v}(h_s - \lambda_s) + \mu_i + p_i - \gamma_i = 0, \quad u \le i \le v, \quad (6.18)$$

$$\frac{\partial \mathcal{L}_2}{\partial q_i} = \sum_{s=i}^{v}(h_s - \lambda_s) + r w_i q_i^{r-1} = 0, \quad u \le i \le v. \quad (6.19)$$

Since $I_t > 0$ for $u \le t < v$ the complementary slackness condition (C-S) necessitates $\lambda_t = 0, \quad u \le t < v$. Suppose to the contrary that $0 < L_t < d_i$ for $u \le t < v$. Then due to the (C-S) condition $\mu_t = \gamma_t = 0$ for $u \le t < v$ and therefore, (6.18) reduces to,

$$p_i + \sum_{s=i}^{v} h_s - \lambda_v = 0, \quad u \le i \le v. \quad (6.20)$$

Similarly after deriving an equation similar to (6.20) for $j > i$ and taking the difference we obtain,

$$p_i + \sum_{s=i}^{j-1} h_s = p_j, \quad u \le i < j < v \quad (6.21)$$

which is impossible due to the assumptions. Hence, in an optimal solution, we cannot have two distinct periods in which $L_t(d_t - L_t) > 0$. Therefore, we have at most one period within the generation with partially lost demand.

Now we prove that such a period can only be the last period, namely v. Suppose

to the contrary that in an optimal solution, $0 < L_t < d_t$ for a single t, $u \leq t < v$. Lemma 4 implies that $L_i > 0 \ \forall i,\ t < i$ and on the other hand, from the argument above, they cannot be partial. That is, $L_i = d_i$ for $t < i \leq v$ which is a contradiction because each of the periods i, $(t < i)$, starts with a positive inventory (due to the generation definition) while they lose all their demands. Hence, partial demand loss is possible only in the last period.
Meanwhile, in the last period of the generation we have $I_{v-1}+q_t+L_t-d_t = I_v = 0$ which implies that $L_v = (d_v - q_v - I_{v-1}) < d_v$. This indicates that we cannot have a complete demand loss in the last period of the generation.
Thus, L_t either belongs to $\{0, d_t\}$ for $u \leq t < v$: $I_t > 0$ and $L_t(d_t - L_t) = 0$; or $0 \leq L_t < d_v$ for $t = v$ and therefore, $I_t L_t(d_t - L_t) = 0$ holds within generation $\langle u, v \rangle$ and consequently in the entire horizon. □

Remark If $0 < L_t^* < d_t$ then $q_t^* = (p_t/rw_t)^{1/(r-1)}$.

It can easily be derived by combining (6.19) and (6.20).

6.2.2 Cap policy comparison: a numerical study

In this section, we conduct a sensitivity analysis to examine how the total cost, the total emission and the total lost sales will change in an optimal production plan under different carbon cap configurations. We have considered both rolling and seasonal cap policies over the planning horizon. To set the cap amount in the parameters, first we have solved the uncapacitated model without permission to demand loss. i.e., $L_t = 0, \forall t$ and $\kappa_j = \infty, \forall j$ in (PLC); then the unit penalty for demand loss has been set as $p = \eta \frac{TC}{D}$ where TC and D denote the total cost and total demand of the corresponding instance, respectively. Using the obtained penalty for the lost demand, then the uncapacitated model resolved with permission to demand loss and the obtained emission amount are used as a base for the carbon emission cap for each of the corresponding instances under carbon emission policies. The level and value of the parameters used in our study

Table 6.6: List of parameters and their values

Name	Description	values
Horizon length		$T = 24$
Number of resources		$N = 4$
Demand	$R = 6$ number of realizations	$d \sim U(0, 100)$
Setup Cost	$K = J^2/2\bar{d} \quad J \in \{2, 6\}$	$\in \{100, 900\}$
Unit holding cost		$h = 1$
marginal cost ratio	$a := \frac{h}{rW\bar{d}^{r-1}}$ ratio of holding to production cost	$\{0.05, 0.20\}$
Technology factor	$A = \prod_{i=1}^{n} (ci/\alpha_i)^{\alpha_i} (ad^{r-1})^{1/r}$	$\{2.3, 0.9\}$
Output elasticity	$(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ sorted increasing for $r = 1.5$	$(\frac{1}{15}, \frac{1}{15}, \frac{3}{15}, \frac{5}{15})$
Resource cost	$(c_1, c_2, c_3, c_4) \sim U(0, 1)$ sorted increasing	$(0.1, 0.3, 0.5, 0.8)$
Resource emission	$(\beta_1, \beta_2, \beta_3, \beta_4) \sim U(0, 1)$ sorted decreasing	$(0.9, 0.6, 0.4, 0.2)$
Block length	$l \in \{1, 2, T/8, T/6, T/3, T\}$	$\{1, 2, 3, 4, 8, 24\}$
Cap	$\kappa_0(1-\delta)$ set for each (K, A, R, p) instance separately	
Cap tightness		$\delta \in \{5\%, 15\%, 25\%\}$
Policy:	rolling, block	
	decreasing, constant, increasing	$s := \frac{\text{cap in the first block}}{\text{cap in the last block}}$
	rolling	$s \in \{1\}$
	block: $l \neq T$	$s \in \{\frac{1}{8}, 1, 8\}$
lost sale penalty:	p=$\eta(TC_0/D)$	$\eta \in \{1, 1.5, 3\}$

are summarized in Table 6.6 and the demand realizations are depicted in Figure 6.2.

Note that we have considered both rolling and seasonal carbon emission policies. For both of them we have considered 6 different lengths for carbon emission blocks which is defined before. For the rolling emission cap policy emission caps are equally assigned while for seasonal emission policy, except for $l = T$, we have considered 3 forms of emission cap: (i) decreasing (ii) constant, and (iii) increasing. The slope of the emission cap denoted by s in Table 6.6 and defined based on the ratio of the emission cap on the first block to the last block. In total based on the parameters, $(K \times A \times R \times \delta \times \eta \times l \times s)$, $2 \times 2 \times 6 \times 3 \times 2 \times 6 \times 1$= 864 problem instances for the rolling policy, and $2 \times 2 \times 6 \times 3 \times 2 \times 5 \times 3$=2160 problem instances for the seasonal policy are tested.

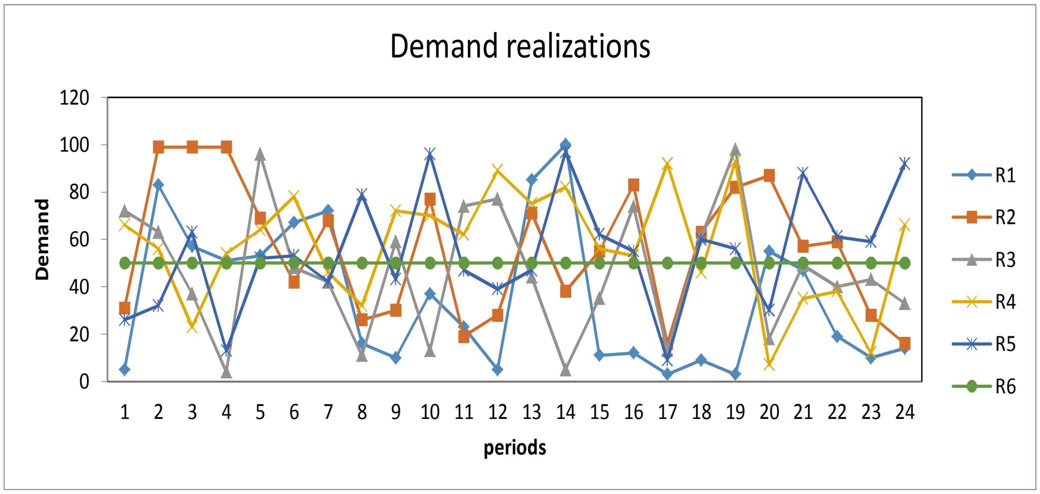


Figure 6.2: Demand realizations drawn from $U(0, 100)$ distribution

We have calculated the ratio between the optimal objective value and the emission amount of the carbon capacitated model with their corresponding values in the uncapacitated model. Then they are averaged over the instances and summarized in Tables 6.7–6.9 with labels TC and TE, respectively. In these tables LS corresponds to the average proportion of lost demands. Label l_i in the rows indicates the block length and the columns δ_i nominate the cap tightness levels in Table 6.7. Furthermore, columns s1, s2, s3 correspond to cap slope among blocks and p1, p2 indicate the lost sales penalty levels. In Table 6.7, as expected in each region TC is increasing in δ, p and decreasing in l. TE is increasing in l and p but decreasing in δ. LS is increasing in δ but decreasing in p and l. The same pattern appears in Tables 6.8–6.9 for s2.

The effect of cap slope is counter-intuitive. That is, constant uniform cap (s2) leads to the lowest cost, the highest emission and the lowest lost demand. The decreasing cap (s1) results in the second rank for all response variables while the increasing cap (s3) results in the highest cost, the lowest emission and the highest lost sales among the policies. Additionally, we see some violation of the expected trend in columns s1 and s3. For example TC increases from l_4 to l_5.

Observation 1 *Response variables are not monotone in l_i when the cap is not* *constant.*

Observation 2 *From a policy maker perspective (maybe the government), with a certain amount of total emission allowance, it seems that a seasonal emission cap with an increasing pattern is a more environment friendly policy.*

To compare the rolling with seasonal cap policy, the left (right) side of Table 6.7 should be equated to the middle part of Table 6.8 (Table 6.9).

Observation 3 *The rolling cap block policy is very close but slightly more environment-friendly policy than the seasonal cap policy with a constant cap.*

In Figures 6.3–6.4, six response variables including cost, emission, lost sales, percentage of production periods, percentage of periods with lost demand and percentage of periods with complete lost demand are depicted separately for p1 and p2. The half upper part of each group is in fact the pictorial representation of Tables 6.7–6.8. Comparing part (a) and (b) of these figures reveals that the costs are significantly affected while emissions are almost at the same level.

Observation 4 *From a policy maker perspective, it seems that cap policies has the same effect on both high and low value-added manufacturer.*

Figure 6.5 depicts the Pareto analysis for the average of the aforementioned emission and cost response variables in rolling and seasonal cap policies, respectively. Graph (a) and (b) belong to rolling cap policy and the others nominate seasonal policy. For each cap policy and lost penalty level the response variables were averaged and plotted separately. The lost sales penalty levels (p) are grouped in columns while each row corresponds to the cap slopes (s). The block lengths are plotted with different shape and the tightness levels have different colors. Therefore the cap policies are uniquely distinguishable at each figure.

For both of the figure, the trend of shapes, the colors and the vertical positions of points in the left side (p1) and the right side (p2) figures are the same which indicates that similar policies have the same effect. (See Observation 4).

Note that, regardless of the color the plotted points corresponding to relatively longer block length took place in the left top region of the graphs.

Observation 5 *Emission cap policies with short block lengths are dominated policies for the reasonable emission reduction targets.*

In Figure 6.5:(g)-(h) the emission amounts have been significantly reduced with almost similar block length and tightness levels whereas it costs more, especially when lost sales penalty is high.

Observation 6 *For a certain amount of emission allowance, the seasonal cap policy with increasing cap seems to be the most environment friendly policy from the legislator perspectives. However, it may drastically affect highly value-added industries (i.e, when p is large).*

As a general statement, we can conclude that emission cap policies can be more effective than limiting the emission allowance of the manufacturer for a certain horizon without putting controls on seasonal or other short terms interference policies. It appears that it would worthwhile for the legislators to apply dissimilar cap policies in different sectors.

Table 6.7: Average sensitivity of response variables to cap policies: rolling

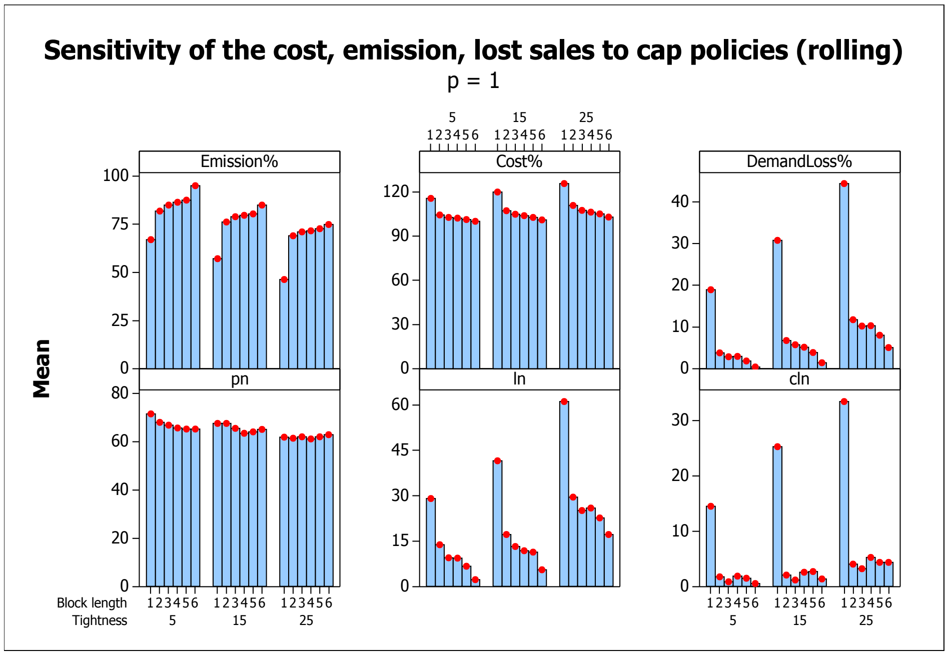
		p1 δ_1	p1 δ_2	p1 δ_3	p2 δ_1	p2 δ_2	p2 δ_3
TC	l_1	1.156	1.200	1.257	1.319	1.450	1.758
	l_2	1.044	1.072	1.110	1.056	1.112	1.190
	l_3	1.027	1.049	1.075	1.038	1.082	1.150
	l_4	1.022	1.038	1.064	1.032	1.066	1.130
	l_5	1.012	1.028	1.051	1.016	1.042	1.098
	l_6	1.001	1.010	1.030	1.001	1.012	1.043
TE	l_1	0.670	0.572	0.464	0.742	0.701	0.587
	l_2	0.819	0.762	0.690	0.830	0.772	0.713
	l_3	0.851	0.789	0.710	0.851	0.790	0.716
	l_4	0.864	0.797	0.717	0.865	0.799	0.717
	l_5	0.876	0.805	0.726	0.877	0.807	0.723
	l_6	0.950	0.850	0.750	0.950	0.850	0.750
LS	l_1	0.190	0.307	0.444	0.093	0.132	0.281
	l_2	0.038	0.068	0.118	0.002	0.014	0.034
	l_3	0.029	0.057	0.102	0.003	0.008	0.030
	l_4	0.029	0.052	0.103	0.001	0.009	0.026
	l_5	0.019	0.039	0.080	0.001	0.003	0.017
	l_6	0.004	0.015	0.050	0.000	0.000	0.000

Table 6.8: Average sensitivity of response variables to cap policies: seasonal, p1

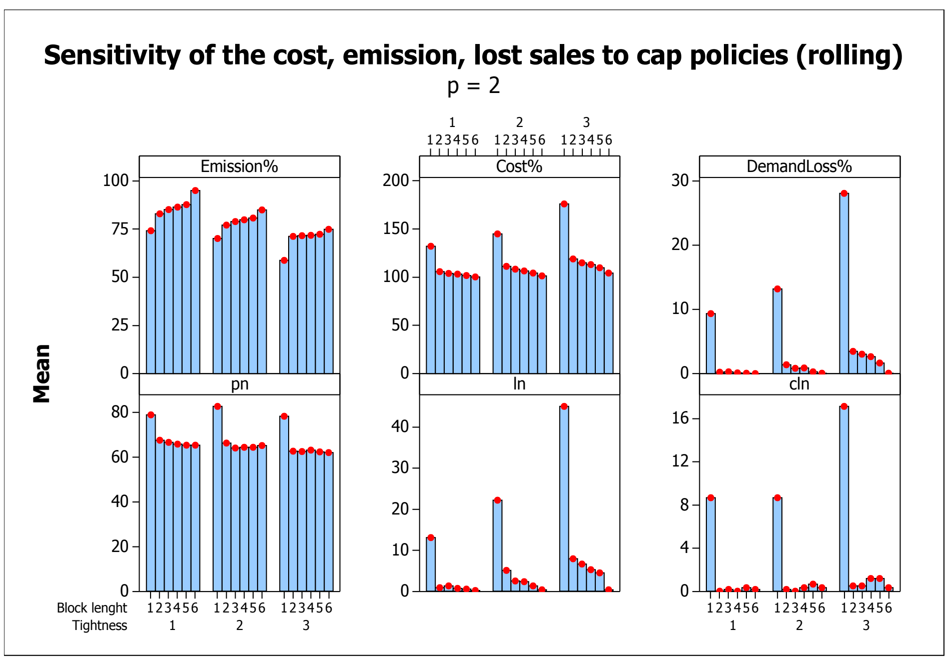
		s1			s2			s3		
		δ_1	δ_2	δ_3	δ_1	δ_2	δ_3	δ_1	δ_2	δ_3
TC	l_1	1.193	1.223	1.258	1.156	1.200	1.257	1.224	1.250	1.282
	l_2	1.100	1.120	1.146	1.038	1.064	1.100	1.139	1.159	1.184
	l_3	1.076	1.091	1.112	1.018	1.036	1.061	1.115	1.132	1.155
	l_4	1.071	1.086	1.108	1.014	1.029	1.052	1.115	1.131	1.151
	l_5	1.086	1.098	1.112	1.008	1.021	1.043	1.135	1.146	1.161
	l_6	-	-	-	1.001	1.010	1.030	-	-	-
TE	l_1	0.598	0.543	0.476	0.667	0.573	0.464	0.534	0.491	0.426
	l_2	0.751	0.698	0.644	0.833	0.767	0.699	0.671	0.625	0.574
	l_3	0.797	0.754	0.698	0.871	0.806	0.726	0.704	0.667	0.619
	l_4	0.799	0.750	0.696	0.898	0.823	0.734	0.707	0.671	0.623
	l_5	0.789	0.761	0.709	0.912	0.832	0.739	0.676	0.638	0.601
	l_6	-	-	-	0.950	0.850	0.750	-	-	-
LS	l_1	0.301	0.360	0.436	0.194	0.305	0.444	0.381	0.425	0.493
	l_2	0.149	0.191	0.237	0.034	0.063	0.106	0.232	0.274	0.321
	l_3	0.118	0.147	0.186	0.021	0.044	0.085	0.202	0.236	0.272
	l_4	0.118	0.149	0.185	0.020	0.038	0.083	0.212	0.243	0.280
	l_5	0.150	0.166	0.194	0.014	0.031	0.066	0.257	0.276	0.301
	l_6	-	-	-	0.004	0.015	0.050	-	-	-

Table 6.9: Average sensitivity of response variables to cap policies: seasonal, p2

		s1			s2			s3		
		δ_1	δ_2	δ_3	δ_1	δ_2	δ_3	δ_1	δ_2	δ_3
TC	l_1	1.567	1.675	1.812	1.319	1.451	1.758	1.725	1.823	1.940
	l_2	1.259	1.324	1.410	1.049	1.095	1.165	1.440	1.508	1.601
	l_3	1.197	1.246	1.315	1.023	1.057	1.112	1.384	1.437	1.514
	l_4	1.183	1.238	1.308	1.020	1.044	1.100	1.382	1.443	1.519
	l_5	1.243	1.282	1.335	1.011	1.031	1.077	1.497	1.528	1.569
	l_6	-	-	-	1.001	1.012	1.043	-	-	-
TE	l_1	0.661	0.613	0.548	0.742	0.700	0.587	0.582	0.536	0.480
	l_2	0.785	0.744	0.681	0.845	0.786	0.722	0.684	0.641	0.597
	l_3	0.823	0.782	0.718	0.871	0.807	0.727	0.703	0.672	0.627
	l_4	0.838	0.794	0.717	0.899	0.823	0.738	0.712	0.678	0.630
	l_5	0.841	0.796	0.732	0.911	0.831	0.741	0.676	0.640	0.605
	l_6	-	-	-	0.950	0.850	0.750	-	-	-
LS	l_1	0.228	0.277	0.345	0.093	0.133	0.281	0.319	0.366	0.421
	l_2	0.087	0.113	0.151	0.002	0.008	0.025	0.190	0.220	0.262
	l_3	0.062	0.082	0.111	0.001	0.007	0.018	0.171	0.191	0.224
	l_4	0.056	0.076	0.109	0.001	0.004	0.017	0.167	0.196	0.231
	l_5	0.080	0.098	0.123	0.000	0.002	0.008	0.236	0.248	0.263
	l_6	-	-	-	0.000	0.000	0.000	-	-	-

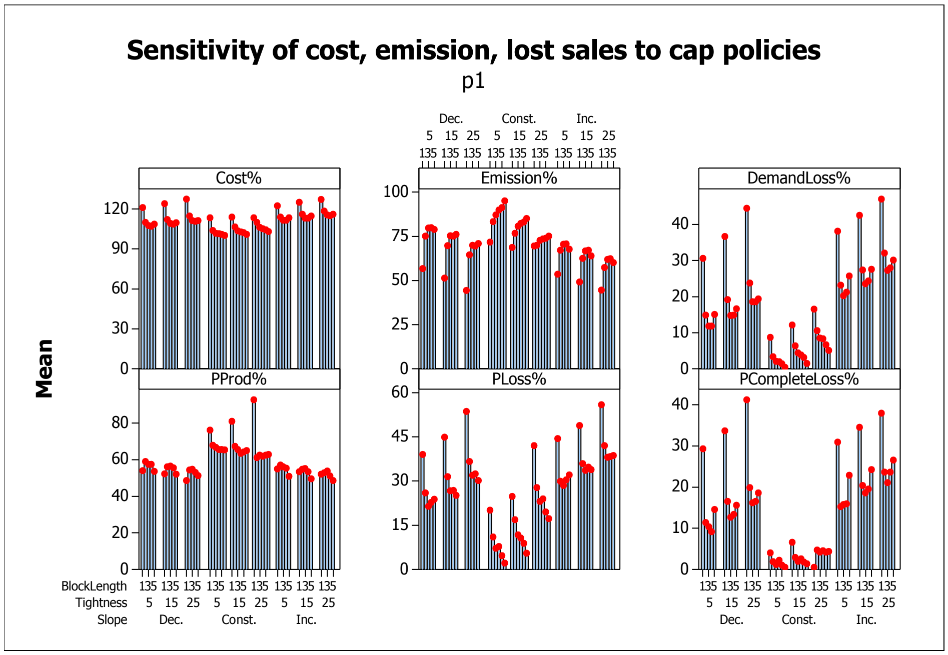


(a)

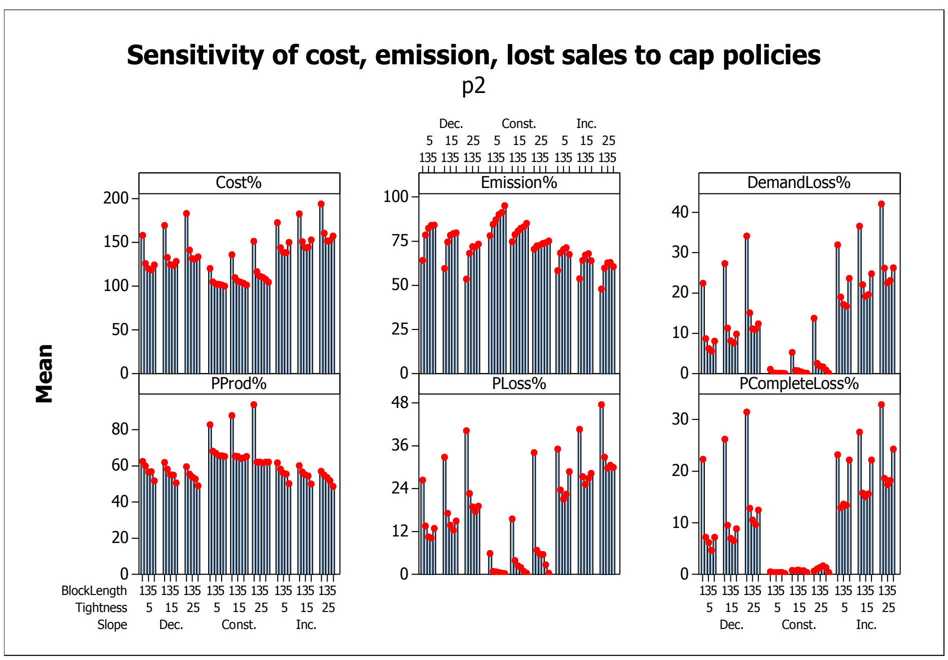


(b)

Figure 6.3: Sensitivity of response variables to cap configurations: rolling



(a)



(b)

Figure 6.4: Sensitivity of response variables to cap configurations: seasonal

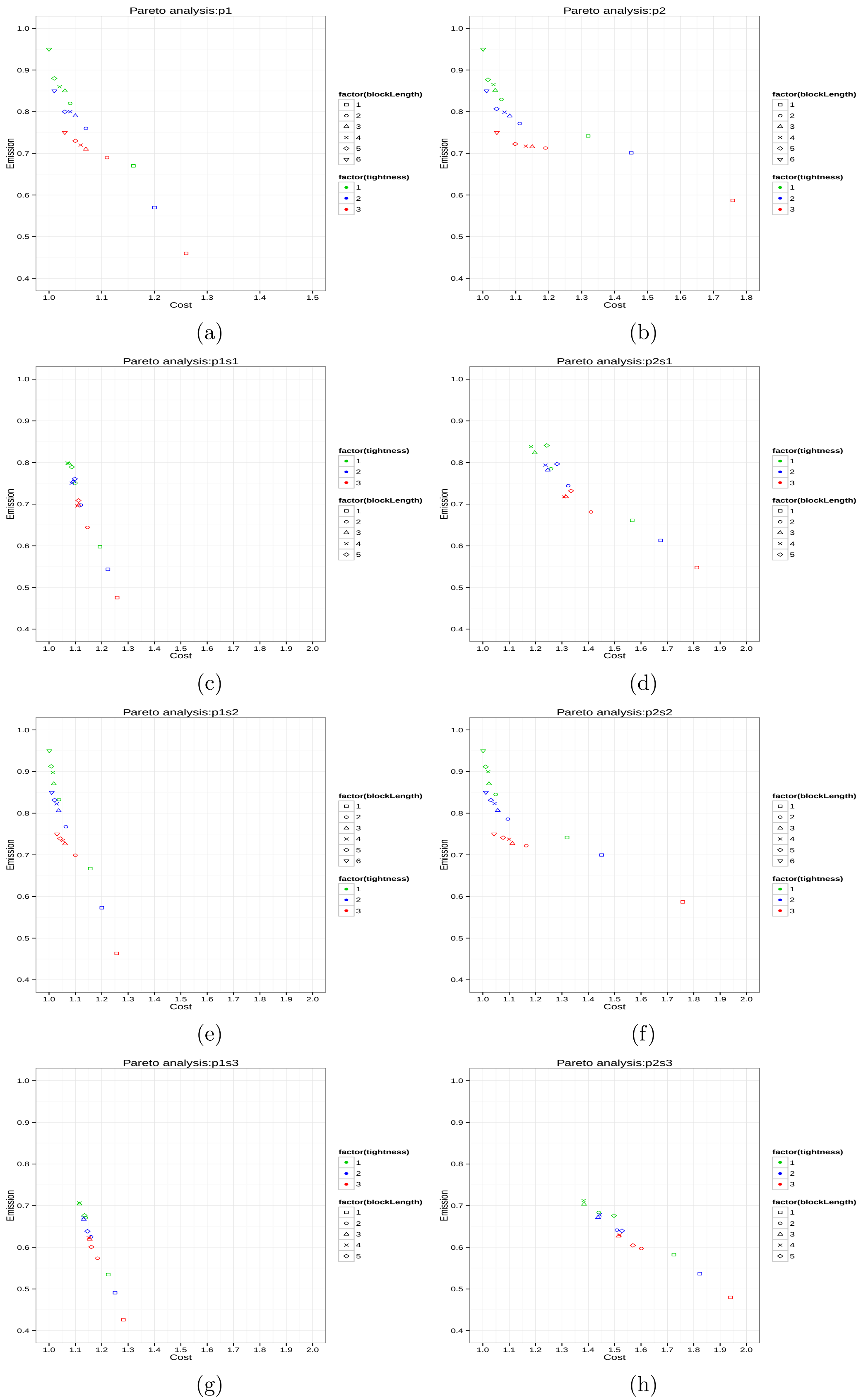


Figure 6.5: Pareto analysis of cost-emission: (a)-(b):Rolling cap, (c)-(h): Seasonal cap

Chapter 7

Conclusion and directions for future research

In this dissertation, we considered on the special variant of the production planning problem. First we focused on the unpacatitated deterministic lot sizing problem with power-like convex production cost function. This problem was motivated and originated from two aspects: in some production systems especially when the yield factor effects are taken into account the output follows a Cobb-Douglas production function which leads to a polynomial power-like cost function after some reformulation steps. On the other hand, under sustainability considerations due to the fines and charges for the overcapacity usage of the ecological resources the production cost function is directly addressed with convex power-like function in the literature. Despite the very rich literature of the lot sizing models, there was a gap in the literature about the models with convex cost and positive fixed setup cost. We designed a DP based heuristic to obtain good quality solutions and we modified some of the well known existing production planning heuristics in order to solve this problem and we conducted a very extensive numerical study to compare and measure the sensitivity of the algorithms with respect to the problem parameters.

Next, we considered the same problem with possibly deteriorating products

and also the nonlinear cost function for the holding cost, as well. We focused on this problem from the mathematical programming aspect and compared the performance of several mixed integer nonlinear optimization packages and also different well known formulations. We then applied second order cone programming and proposed a couple of optimality cuts and strengthening valid inequality to speed up the solvers execution time. We also contributed in SOCP reformulation technique with determining the exact number of cone representation constraints required for a given exponent rather than an upper bound appeared in the literature.
In our computational experiments we concluded that facility location based formulation strengthened with perspective cut is very tight which results in an optimal solution in a reasonable time or it stops with very small gap from the optimality for the moderate sized instances. Meanwhile, the optimality cuts we proposed showed positive effect only in the other formulations (AGG and NIF).

Then, we investigated the convex cost lot sizing problem in the rolling horizon basis. The motivation behind this configuration was also twofold. The first is that real or accurate demand estimates are usually possible for a limited number of periods in MRP systems which forces the planning to by myopic. The second is that in such frameworks heuristic approaches may perform better than an exact solution. We concentrated to identify such cases besides the general assessment of the heuristics. In our experiments we observed that the dominance of the heuristics to the exact method in rolling horizon setting mainly occurs when the planning horizon is longer than the effective lot size obtained from the static counterpart of the problem. On the other hand, we observed that despite the relatively poor average performance of some heuristics like Silver-Meal, they outperform the exact method with extreme deviations which the other better heuristics are unable to catch.

As a final problem, we studied the carbon emission capacitated version of the problem with Cobb-Douglas production function between the inputs and the output quantity. For this problem we proposed two mathematical programming based exact solution methods. The first one rested on the reformulation to the second order cone programming which we had already experience its reliable and robust performance while the second method was based on the generalized benders decomposition method. The advantages of the second approach is that the second-stage problem can be analytically handled with KKT optimality conditions without using any solver and the rest of the problem in master is completely mixed integer linear problem. Hence, we could solve an extremely nonlinear problem with a linear optimization package. To construct fast solution, we proposed two iterative Lagrangian methods: bisection and adjustive Lagrangian algorithms. They were integrated with our previously designed heuristics to solve the capacitated problem by relaxing the cap constraint. The adjustive Lagrangian approach is so fast and in our numerical test it converges almost in less than four iterations while it is able to catch the very high quality solutions. We also considered the possibility of lost sales demand and examined different carbon emission cap policies in forms of rolling and seasonal. Our numerical study revealed that the emission amount, the lost sales amount and the cost incurred by a manufacturer is impacted more by carbon emission policies than just considering the total emission allowance over the entire planning horizon.

Several extensions of this problem can be investigated as future research problems. For example, by considering a stochastic demand, our first problem can be extended to a stochastic lot sizing problem with convex production cost. The literature lacks such convex cost stochastic lot sizing models. The recent work of Koca et al.[47] which includes a convex term, is also approximated with a deterministic model. Therefore, there are some rooms in stochastic extension of this problem. In particular, Tunc et al. [89] has recently developed a lot sizing model with variable unit production cost which is piecewise linear and concave. Their work can be extended with convex production costs.

The rolling planning horizon study can be extended by considering the ending inventory effect as proposed by Fisher et al. [90] for the classical linear cost lot sizing. That is, each sub-horizon in the rolling setting is solved with a non-zero ending inventory and the aim is to determine an optimal value for the ending inventory amount. Also a worst case analysis of the exact solution method for the convex cost lot sizing in the rolling horizon planning, similar to the work of Qian et al. [91] for the linear case, might be a good extension for future research.

The carbon emission in a production process may vary periodically due to malfunctioning of the production facilities. Therefore, a random emission for a certain level of production is probable. Regarding to this idea, our last problem can be extended with random emission similar to sizing models with random yields which are extensively reviewed in [92].

Also, in this study we considered strict carbon cap policies. This assumption can be extended to other policies such as cap-and-trade or carbon tax frameworks. For instance, He et al. [93] compare trade and tax policies with an EOQ based production plan. Therefore, our last problem can be extended with their policies as a future research.

Bibliography

[1] N. Brahimi, S. Dauzere-Peres, N. M. Najid, and A. Nordli, "Single item lot sizing problems," *European Journal of Operational Research*, vol. 168, no. 1, pp. 1–16, 2006.

[2] D. F. Heathfield and S. Wibe, *An introduction to cost and production functions*. Humanities Pr, 1987.

[3] W. Leontief, "Introduction to a theory of the internal structure of functional relationships," *Econometrica, Journal of the Econometric Society*, pp. 361–373, 1947.

[4] Z. M. Teksan and J. Geunes, "A polynomial time algorithm for convex cost lot-sizing problems," *Operations Research Letters*, vol. 43, no. 4, pp. 359–364, 2015.

[5] H. M. Wagner and T. M. Whitin, "Dynamic version of the economic lot size model," *Management science*, vol. 5, no. 1, pp. 89–96, 1958.

[6] A. Federgruen and M. Tzur, "A simple forward algorithm to solve general dynamic lot sizing models with n periods in 0 (n log n) or 0 (n) time," *Management Science*, vol. 37, no. 8, pp. 909–925, 1991.

[7] E. A. Silver and H. A. Meal, "A heuristic for selecting lot size requirements for the case of a deterministic time varying demand rate and discrete opportunities for replenishment," *Production and Inventory Management Journal*, vol. 14, no. 2, pp. 64–74, 1973.

[8] M. Florian and M. Klein, "Deterministic production planning with concave costs and capacity constraints," *Management Science*, vol. 18, no. 1, pp. 12–20, 1971.

[9] M. Florian, J. K. Lenstra, and A. Rinnooy Kan, "Deterministic production planning: Algorithms and complexity," *Management science*, vol. 26, no. 7, pp. 669–679, 1980.

[10] G. R. Bitran and H. H. Yanasse, "Computational complexity of the capacitated lot size problem," *Management Science*, vol. 28, no. 10, pp. 1174–1186, 1982.

[11] B. Karimi, S. Fatemi Ghomi, and J. Wilson, "The capacitated lot sizing problem: a review of models and algorithms," *Omega*, vol. 31, no. 5, pp. 365–378, 2003.

[12] R. Jans and Z. Degraeve, "Meta-heuristics for dynamic lot sizing: A review and comparison of solution approaches," *European Journal of Operational Research*, vol. 177, no. 3, pp. 1855–1875, 2007.

[13] D. Quadt and H. Kuhn, "Capacitated lot-sizing with extensions: a review," *4OR*, vol. 6, no. 1, pp. 61–83, 2008.

[14] R. Jans and Z. Degraeve, "Modeling industrial lot sizing problems: a review," *International Journal of Production Research*, vol. 46, no. 6, pp. 1619–1643, 2008.

[15] L. Buschkühl, F. Sahling, S. Helber, and H. Tempelmeier, "Dynamic capacitated lot-sizing problems: a classification and review of solution approaches," *Or Spectrum*, vol. 32, no. 2, pp. 231–261, 2010.

[16] H. Ullah and S. Parveen, "A literature review on inventory lot sizing problems," *Global Journal of Researches in Engineering*, vol. 10, no. 5, 2010.

[17] M. A. Bushuev, A. Guiffrida, M. Jaber, and M. Khan, "A review of inventory lot sizing review papers," *Management Research Review*, vol. 38, no. 3, pp. 283–298, 2015.

[18] Y. Friedman and Y. Hoch, "A dynamic lot-size model with inventory deterioration," *INFOR*, vol. 16, pp. 183–188, 1978.

[19] V. N. Hsu, "Dynamic economic lot size model with perishable inventory," *Management Science*, vol. 46, no. 8, pp. 1159–1169, 2000.

[20] V. N. Hsu, "An economic lot size model for perishable products with age-dependent inventory and backorder costs," *IIE Transactions*, vol. 35, no. 8, pp. 775–780, 2003.

[21] L. Y. Chu, V. N. Hsu, and Z.-J. M. Shen, "An economic lot-sizing problem with perishable inventory and economies of scale costs: Approximation solutions and worst case analysis," *Naval Research Logistics (NRL)*, vol. 52, no. 6, pp. 536–548, 2005.

[22] H. Waterer, "Lot sizing with inventory gains," *Operations Research Letters*, vol. 35, no. 6, pp. 759–766, 2007.

[23] S. Nahmias, "Perishable inventory theory: A review," *Operations Research*, vol. 30, no. 4, pp. 680–708, 1982.

[24] F. Raafat, "Survey of literature on continuously deteriorating inventory models," *Journal of the Operational Research society*, pp. 27–37, 1991.

[25] S. Goyal and B. C. Giri, "Recent trends in modeling of deteriorating inventory," *European Journal of operational research*, vol. 134, no. 1, pp. 1–16, 2001.

[26] J. D. Blackburn and R. A. Millen, "Heuristic lot-sizing performance in a rolling-schedule environment*," *Decision Sciences*, vol. 11, no. 4, pp. 691–701, 1980.

[27] U. Wemmerlöv and D. C. Whybark, "Lot-sizing under uncertainty in a rolling schedule environment," *The International Journal of Production Research*, vol. 22, no. 3, pp. 467–484, 1984.

[28] J. H. Bookbinder and B.-T. H'ng, "Rolling horizon production planning for probabilistic time-varying demands," *International Journal of Production Research*, vol. 24, no. 6, pp. 1439–1458, 1986.

[29] R. Venkataraman and S. Smith, "Disaggregation to a rolling horizon master production schedule with minimum batch-size production restrictions," *International Journal of Production Research*, vol. 34, no. 6, pp. 1517–1537, 1996.

[30] A. R. Clark and S. J. Clark, "Rolling-horizon lot-sizing when set-up times are sequence-dependent," *International Journal of Production Research*, vol. 38, no. 10, pp. 2287–2307, 2000.

[31] H. Stadtler, "Improved rolling schedules for the dynamic single-level lot-sizing problem," *Management Science*, vol. 46, no. 2, pp. 318–326, 2000.

[32] N. Simpson, "Questioning the relative virtues of dynamic lot sizing rules," *Computers & Operations Research*, vol. 28, no. 9, pp. 899–914, 2001.

[33] W. Van Den Heuvel and A. P. Wagelmans, "A comparison of methods for lot-sizing in a rolling horizon environment," *Operations Research Letters*, vol. 33, no. 5, pp. 486–496, 2005.

[34] S. Chand, V. N. Hsu, and S. Sethi, "Forecast, solution, and rolling horizons in operations management problems: a classified bibliography," *Manufacturing & Service Operations Management*, vol. 4, no. 1, pp. 25–43, 2002.

[35] F. Sahin, A. Narayanan, and E. P. Robinson, "Rolling horizon planning in supply chains: review, implications and directions for future research," *International Journal of Production Research*, vol. 51, no. 18, pp. 5413–5436, 2013.

[36] A. Ö. Toy and E. Berk, "Dynamic lot sizing for a warm/cold process: heuristics and insights," *International Journal of Production Economics*, vol. 145, no. 1, pp. 53–66, 2013.

[37] L. Baciarello, M. DAvino, R. Onori, and M. M. Schiraldi, "Lot sizing heuristics performance," *International Journal of Engineering Business Management*, vol. 5, no. 6, pp. 1–10, 2013.

[38] M. Heck and G. Schmidt, "Lot-size planning with non-linear cost functions supporting environmental sustainability," in *Networked Digital Technologies*, pp. 1–6, Springer, 2010.

[39] R. Kian, Ü. Gürler, and E. Berk, "The dynamic lot-sizing problem with convex economic production costs and setups," *International Journal of Production Economics*, 2014.

[40] W. Karush, "On a class of minimum-cost problems," *Management Science*, vol. 4, no. 2, pp. 136–153, 1958.

[41] A. Nemirovski and K. Scheinberg, "Extension of karmarkar's algorithm onto convex quadratically constrained quadratic problems," *Mathematical Programming*, vol. 72, no. 3, pp. 273–289, 1996.

[42] F. Alizadeh and D. Goldfarb, "Second-order cone programming," *Mathematical programming*, vol. 95, no. 1, pp. 3–51, 2003.

[43] Y. Nesterov and A. Nemirovsky, "Conic formulation of a convex programming problem and duality," *Optimization Methods and Software*, vol. 1, no. 2, pp. 95–115, 1992.

[44] Y. Nesterov, A. Nemirovski, and Y. Ye, *Interior-point polynomial algorithms in convex programming*, vol. 13. SIAM, 1994.

[45] A. Nemirovski, "What can be expressed via conic quadratic and semidefinite programming," *Lecture notes*, 1998.

[46] Y. E. Nesterov and M. J. Todd, "Primal-dual interior-point methods for self-scaled cones," *SIAM Journal on optimization*, vol. 8, no. 2, pp. 324–364, 1998.

[47] E. Koca, H. Yaman, and M. S. Aktürk, "Stochastic lot sizing problem with controllable processing times," *Omega*, vol. 53, pp. 1–10, 2015.

[48] M. S. Aktürk, A. Atamtürk, and S. Gürel, "A strong conic quadratic reformulation for machine-job assignment with controllable processing times," *Operations Research Letters*, vol. 37, no. 3, pp. 187–191, 2009.

[49] A. Atamtürk, G. Berenguer, and Z.-J. Shen, "A conic integer programming approach to stochastic joint location-inventory problems," *Operations Research*, vol. 60, no. 2, pp. 366–381, 2012.

[50] R. J. Shadbegian and W. B. Gray, "Pollution abatement expenditures and plant-level productivity: a production function approach," *Ecological Economics*, vol. 54, no. 2, pp. 196–208, 2005.

[51] S. A. Hatirli, B. Ozkan, and C. Fert, "An econometric analysis of energy input–output in turkish agriculture," *Renewable and Sustainable Energy Reviews*, vol. 9, no. 6, pp. 608–623, 2005.

[52] G. Gupta, "Production function and factor productivity in the indian cement industry," *Indian Journal of Industrial Relations*, pp. 359–385, 1972.

[53] R. Kümmel, W. Strassl, A. Gossner, and W. Eichhorn, "Technical progress and energy dependent production functions," *Journal of Economics*, vol. 45, no. 3, pp. 285–311, 1985.

[54] P. Söderholm and G. Klaassen, "Wind power in europe: a simultaneous innovation–diffusion model," *Environmental and resource economics*, vol. 36, no. 2, pp. 163–190, 2007.

[55] T. Wei, "Impact of energy efficiency gains on output and energy use with cobb–douglas production function," *Energy Policy*, vol. 35, no. 4, pp. 2023–2030, 2007.

[56] N. Khanna, "Analyzing the economic cost of the kyoto protocol," *Ecological Economics*, vol. 38, no. 1, pp. 59–69, 2001.

[57] N. Banaeian, M. Omid, and H. Ahmadi, "Energy and economic analysis of greenhouse strawberry production in tehran province of Iran," *Energy Conversion and Management*, vol. 52, no. 2, pp. 1020–1025, 2011.

[58] J. Haldi and D. Whitcomb, "Economies of scale in industrial plants," *The Journal of Political Economy*, vol. 75, no. 4, pp. 373–385, 1967.

[59] I. Ozaki, "Industrial structure and employment," *Dev. Econ*, vol. 14, no. 4, pp. 341–365, 1976.

[60] L. J. Lau and S. Tamura, "Economies of scale, technical progress, and the nonhomothetic leontief production function: an application to the japanese petrochemical processing industry," *Journal of Political Economy*, vol. 80, no. 6, pp. 1167–1187, 1972.

[61] S. Nakamura, "A nonhomothetic generalized leontief cost function based on pooled data," *The Review of Economics and Statistics*, pp. 649–656, 1990.

[62] R. Hammami, I. Nouira, and Y. Frein, "Carbon emissions in a multi-echelon production-inventory model with lead time constraints," *International Journal of Production Economics*, vol. 164, pp. 292–307, 2015.

[63] S. Benjaafar, Y. Li, and M. Daskin, "Carbon footprint and the management of supply chains: Insights from simple models," *Automation Science and Engineering, IEEE Transactions on*, vol. 10, no. 1, pp. 99–116, 2013.

[64] G. Hua, T. Cheng, and S. Wang, "Managing carbon footprints in inventory management," *International Journal of Production Economics*, vol. 132, no. 2, pp. 178–185, 2011.

[65] M. Retel Helmrich, R. Jans, W. van den Heuvel, and A. P. Wagelmans, "The economic lot-sizing problem with an emission constraint," tech. rep., Econometric Institute Research Papers, 2012.

[66] S. Anutariya, C. Khompatraporn, and C. Jaturanonda, "Optimization of manufacturing system with rework under carbon emission allowance," in *The 11th Asia Pacific Industrial Engineering and Management Systems Conference, The Asia Pacific Regional Meeting of International Foundation for Production Research, Melaka*, pp. 7–10, 2010.

[67] N. Absi, S. Dauzère-Pérès, S. Kedad-Sidhoum, B. Penz, and C. Rapine, "Lot sizing with carbon emission constraints," *European Journal of Operational Research*, vol. 227, no. 1, pp. 55–61, 2013.

[68] Y. Yu, Z. Hong, C. Chu, and L. Liang, "A dynamic lot sizing model with carbon emission constraint and multi-mode production: A general property

and a polynomially solvable case," in *Networking, Sensing and Control (ICNSC), 2013 10th IEEE International Conference on*, pp. 413–417, IEEE, 2013.

[69] M. Retel Helmrich, *Green Lot-Sizing*. PhD thesis, Erasmus Research Institute of Management (ERIM), 2013.

[70] A. F. Veinott Jr, "Production planning with convex costs: A parametric study," *Management Science*, vol. 10, no. 3, pp. 441–460, 1964.

[71] F. W. Harris, "How many parts to make at once," 1913.

[72] O. Günlük and J. Linderoth, "Perspective relaxation of mixed integer nonlinear programs with indicator variables," in *Integer Programming and Combinatorial Optimization*, pp. 1–16, Springer, 2008.

[73] J. Krarup and O. Bilde, *Plant Location, Set Covering and Economic Lot Size: An 0 (mn)-Algorithm for Structured Problems*. Springer, 1977.

[74] J. R. Evans, "An efficient implementation of the Wagner-Whitin algorithm for dynamic lot-sizing," *Journal of Operations Management*, vol. 5, no. 2, pp. 229–235, 1985.

[75] G. D. Eppen and R. K. Martin, "Solving multi-item capacitated lot-sizing problems using variable redefinition," *Operations Research*, vol. 35, no. 6, pp. 832–848, 1987.

[76] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, "Applications of second-order cone programming," *Linear algebra and its applications*, vol. 284, no. 1, pp. 193–228, 1998.

[77] A. Frangioni and C. Gentile, "Perspective cuts for a class of convex 0–1 mixed integer programs," *Mathematical Programming*, vol. 106, no. 2, pp. 225–236, 2006.

[78] J. Czyzyk, M. P. Mesnier, and J. J. Moré, "The neos server," *Computing in Science and Engineering*, vol. 5, no. 3, pp. 68–75, 1998.

[79] W. Gropp and J. Moré, "Optimization environments and the neos server," *Approximation theory and optimization*, pp. 167–182, 1997.

[80] E. D. Dolan, "Neos server 4.0 administrative guide," *arXiv preprint cs/0107034*, 2001.

[81] W. J. Hopp and M. L. Spearman, *Factory physics.* Waveland Press, 2011.

[82] R. C. Carlson, J. V. Jucker, and D. H. Kropp, "Less nervous mrp systems: a dynamic economic lot-sizing approach," *Management Science*, vol. 25, no. 8, pp. 754–761, 1979.

[83] D. C. Montgomery, *Design and analysis of experiments.*

[84] B. Y. Kara, I. Sabuncuoglu, and B. Bidanda, *Global Logistics Management.* CRC Press, 2014.

[85] A. M. Geoffrion, "Generalized benders decomposition," *Journal of optimization theory and applications*, vol. 10, no. 4, pp. 237–260, 1972.

[86] C. A. Floudas, *Nonlinear and mixed-integer optimization: fundamentals and applications.* Oxford University Press on Demand, 1995.

[87] S. Sözüer, "On the newsvendor problem with multiple inputs under a carbon emission constraint," Master's thesis, Bilkent University, 2012.

[88] D. Aksen, K. Altınkemer, and S. Chand, "The single-item lot-sizing problem with immediate lost sales," *European Journal of Operational Research*, vol. 147, no. 3, pp. 558–566, 2003.

[89] H. Tunc, O. A. Kilic, S. A. Tarim, and B. Eksioglu, "The stochastic lot sizing problem with piecewise linear concave ordering costs," *Computers & Operations Research*, vol. 65, pp. 104–110, 2016.

[90] M. Fisher, K. Ramdas, and Y.-S. Zheng, "Ending inventory valuation in multiperiod production scheduling," *Management Science*, vol. 47, no. 5, pp. 679–692, 2001.

[91] T. Qian, P. C. Jones, and Y. Ye, "Worst case analysis of forward wagner–whitin algorithm with rolling horizons," *Dpt. Of Management Sciences College of Bussiness Administration. The University of Iowa*, 1996.

[92] C. A. Yano and H. L. Lee, "Lot sizing with random yields: A review," *Operations Research*, vol. 43, no. 2, pp. 311–334, 1995.

[93] P. He, W. Zhang, X. Xu, and Y. Bian, "Production lot-sizing and carbon emissions under cap-and-trade and carbon tax regulations," *Journal of Cleaner Production*, vol. 103, pp. 241–248, 2015.