Multi-Item Supply Chain and Revenue Management Problems

JOERN MEISSNER

Submitted in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy in Business
under the Executive Committee of the
Graduate School of Art and Sciences

Columbia University
2004

ABSTRACT

Multi-Item Supply Chain and

Revenue Management Problems

Joern Meissner

The areas of Revenue Management and Supply Chain Management represent two fundamental pillars for the management of industries that procure and distribute consumer products. The former is concerned with the management of the *demand* processes and the development of methodologies and systems required to support this management function. The area of Supply Chain Management is concerned with the the design of a supply process to match a given demand pattern as efficiently as possible. It may therefore be viewed as the complement of the Revenue Management area.

Operations management papers have demonstrated that the operational environment and associated cost structures may have a fundamental impact on the equilibrium behavior in the industry, in general, and the resulting price levels in particular. Little remains known, however, about how prices should be set in a competitive environment, in the *simultaneous* presence of two other major complications:

(i) time dependent demand functions and cost parameters, and

(ii) scale economies in the operational costs

Conversely, traditional inventory and procurement planning models assume that the demand processes for the finished goods are exogenously given, when, in reality, these demand processes can be managed by appropriate price choices, inter alia. It is of critical importance to understand how effective replenishment strategies are affected by pricing decisions and how replenishment strategies and pricing decisions are to be integrated effectively.

This dissertation focuses on the following four areas of complicating factors affecting the union of Supply Chain Management and Revenue Management:

(A) *Pricing Decisions.* Here we distinguish between two types of settings. In the first case, the firm operates as a monopolist or in an environment of imperfect competition, with the competitors' prices (temporarily) fixed. In the second case, prices need to be determined in an environment of imperfect, oligopolistic competition.

(B) *Time-dependent demand functions and cost structures.*

(C) *Economies of scale in the operational cost.* These arise, for example, from *fixed* cost components in the procurement processes, i.e. production and distribution setup costs.

(D) *Capacity Limitations*, i.e. limits on how many units can be produced in any given period or, in the aggregate, over the complete planning horizon. Such capacity limits often create interdependencies between different products sharing the same production or distribution equipment.

Contents

1 Introduction **1**

2 Progressive Interval Heuristics for Multi-Item Capacitated Lot-Sizing Problems **7**

2.1 Introduction 7

2.2 Literature Review 14

2.3 The multi-item model with joint setup cost (JS) 16

2.4 Progressive interval heuristics: worst case bounds for optimality gaps . . . 18

2.5 Solution methods for a single interval problem: polynomial and asymptotically optimal heuristics 27

2.6 The general (JIS) – model and numerical results 33

3 Probabilistic Analysis of Progressive Interval Heuristics for Multi-Item Capacitated Lot-Sizing Problems **45**

3.1 Introduction 45

3.2 The model and the class of progressive interval heuristics 49

3.3 Almost sure asymptotic optimality 53

3.4 Products with limited shelf life 61

4 Dynamic Pricing Strategies for Multi-Product Revenue Management Problems **68**

4.1 Introduction 68

4.2 Single Resource, Multi-Product Model 75

4.2.1 Dynamic pricing model 75

4.2.2 The Single-Resource Capacity Control Problem 79
4.3 Analysis of the pricing and capacity control problems 80
4.3.1 Dynamic programming formulation of the pricing problem 80
4.3.2 DP formulation of the capacity control problem 82
4.3.3 A unified analysis of the pricing and capacity control problems . . . 84
4.3.4 An efficient frontier for multi-product pricing and capacity controls 87
4.4 Deterministic analysis of the multi-product pricing problem 88
4.4.1 Solution to the deterministic multi-product pricing problem 88
4.4.2 Heuristic policies extracted from the deterministic analysis 91
4.4.3 Dynamic Pricing Network Revenue Management Problems 100
4.5 Numerical examples 101

5 Price Competition under Time-Varying Demands and Dynamic Lot Sizing Costs 112

5.1 Introduction 112
5.2 Model and Notation 119
5.3 The best response problem under multiplicative seasonalities and constant setup costs 124
5.4 The Best Response Problem: The General Case 140
5.5 The Equilibrium Analysis 145
5.6 Numerical Examples 150

Acknowledgements

First and foremost, I am greatly indebted to my advisors, Professors Awi Federgruen and Costis Malglaras. Their talent and drive was a fundamental component of this dissertation, and their advice and insights about academic life was invaluable. Working with them has transformed me not only as a researcher, but also as a person.

As my principal advisor, Professor Awi Federgruen provided me with constant support and tireless encouragement. He never failed to motivate me. His research style is an archetype that I will strive to emulate throughout my career.

My frequent interactions and collaborations with Professor Costis Maglaras were also very valuable learning experiences. I am grateful for his support and enthusiasm.

I would also like to thank Professors Daniel Bienstock, Guillermo Gallego, and Sergei Savin for serving on my dissertation committee and providing me with useful suggestions. Professors Sebastian Ceria and Fangruo Chen have been my advisors during the first phase of the doctoral program at Columbia Business School. I am indebted to both of them and all my teachers here at Columbia for a stimulating and rewarding experience that has broadened my horizon. Professors Linda Green and Julien Bramel have advised me during my first experiences on front of the classroom and they have taught me how to convey excitement about the field of operations management to my students.

My fellow graduate students also played an integral role in my education. In particular, I am grateful to Bin Yu, and my office-mates Jeremy Staum and Gad Allon for many fruitful discussions. I am also thankful for the collaboration with former Columbia Ph.D. Michal Tzur.

Finally, I wish to express my gratitude for all the years of encouragement and support by my parents, Johannes and Karin Meissner. This thesis is dedicated to them.

Chapter 1

Introduction

The areas of Revenue Management and Supply Chain Management represent two fundamental pillars for the management of industries that procure and distribute consumer products. The former is concerned with the management of the *demand* processes and the development of methodologies and systems required to support this management function. The area of Supply Chain Management is concerned with the the design of a supply process to match a given demand pattern as efficiently as possible. It may therefore be viewed as the complement of the Revenue Management area.

While many challenges remain in the development of these two areas, by themselves, there is an increasing recognition of the importance to integrate the two areas with the help of models which *simultaneously* manage the supply and demand processes. The integrated area is often referred to as Enterprise Profit Optimization; its state-of-the-art is still in its infancy.

Determining the 'right' price to charge for a product is a complex task. A voluminous literature in economics and marketing has been devoted to models which prescribe how prices should be set in industries in which a limited number of competing firms offer similar products, which may therefore be viewed as substitutes. These papers typically model the interaction among competitors as a noncooperative game, see VIVES (2000) and TIROLE (1988) for survey texts.

More recently, operations management papers have demonstrated that the operational environment and associated cost structures may have a fundamental impact on the equilibrium behavior in the industry, in general, and the resulting price levels in particular. See CACHON (2003) for a recent survey. Little remains known, however, about how prices should be set in a competitive environment, in the *simultaneous* presence of two other major complications:

(i) time dependent demand functions and cost parameters, and

(ii) scale economies in the operational costs

Conversely, traditional inventory and procurement planning models assume that the demand processes for the finished goods are exogenously given, when, in reality, these demand processes can be managed by appropriate price choices, inter alia. It is of critical importance to understand how effective replenishment strategies are affected by pricing decisions and how replenishment strategies and pricing decisions are to be integrated effectively.

This dissertation focuses on the following four areas of complicating factors affecting the union of Supply Chain Management and Revenue Management:

(A) *Pricing Decisions.* Here we distinguish between two types of settings. In the first case, the firm operates as a monopolist or in an environment of imperfect competition, with the competitors' prices (temporarily) fixed. In the second case, prices need to be determined in an environment of imperfect, oligopolistic competition in which the firm's demand depends not only on its own price choices but equally on those made by each of the firm's competitors.

(B) *Time-dependent demand functions and cost structures.*

(C) *Economies of scale in the operational cost.* These arise, for example, from *fixed* cost components in the procurement processes, i.e. production and distribution setup costs.

(D) *Capacity Limitations*, i.e. limits on how many units can be produced in any given period or, in the aggregate, over the complete planning horizon. Such capacity limits often create interdependencies between different products sharing the same production or distribution equipment.

This dissertation consists of an analysis *four* finite horizon planning models. In each, the planning horizon is partitioned into a finite number of periods, with decisions restricted to the the beginning of these periods. Each of the four models integrates several of the complicating factors (A)–(D), as explained below. Each corresponds with one of chapter 2–5 of this dissertation, with the following titles:

Chapter 2: Progressive Interval Heuristics for Multi-Item Capacitated Lot-Sizing Problems

Chapter 3: Probabilistic Analysis of Progressive Interval Heuristics for Multi-Item Capacitated Lot-Sizing Problems

Chapter 4: Dynamic Pricing Strategies for Multi-Product Revenue Management Problems

Chapter 5: Price Competition under Time-Varying Demands and Dynamic Lot Sizing Costs

Chapters 2 and 3 correspond with Federgruen, Meissner, and Tzur (2002) and Federgruen and Meissner (2004a), respectively. They consider a multi-item planning model in which the time series of demands for each of the items under consideration, is given exogenously. The model integrates complicating factors (B)–(D), above. While the model, itself, has been addressed since the seventies, in a voluminous literature, we develop the first class of heuristics which can be designed to be *simultaneously* polynomially bounded *and* asymptotically optimal as T, the number of periods in the planning horizon tends to infinity. The heuristics can also be designed to generate an ϵ-optimal solution for *any* $\epsilon > 0$. The properties are obtained in the first paper, as *worst case guarantees*, merely assuming that some of the model parameters are uniformly bounded from above and below. For example, we assume that the demand and capacity values are uniformly bounded from above. The second paper relaxes this assumption and establishes that performance guarantees hold, *with probability one*, if the periods' demand and capacity values are guaranteed as independent realizations from an arbitrary (common) multivariate distribution, possibly with *unbounded* support. To our knowledge, this paper represents the first probabilistic analysis of heuristics in entire lot sizing literature. The first paper also reports on an extensive numerical study demonstrating the practical importance of the proposed class of heuristics.

Chapter 4 corresponds with Maglaras and Meissner (2004) and addresses a planning model in which all demand must be satisfied from inventory that is available at the beginning of the planning horizon. In other words, the model addresses settings where inventory can *not* be replenished, so that the complicating factory (C) is irrelevant in this setting. The model succeeds in integrating the factors (A) and (D). More specifically, the model assumes that that in each period the demand rate for each of the items can be adjusted by selecting price levels for all n items. While each item's demand rate, in any given period, is given by a demand function of the complete price vector which prevails in this period, these demand functions are assumed to be time-homogenous. However, the generalization to the case of time-dependent demand functions is rather simple, thus enabling the integration of the factors (A), (B) and (D). As such, the model represents an important generalization and extension to the classical papers by GALLEGO and VAN RYZIN (1994, 1997). The model is shown to be applicable, with minor changes, to a setting where the price levels are determined, upfont, but the firm has the discretion to accept or reject individual product requests.

Finally, the fifth and last chapter corresponds with FEDERGRUEN and MEISSNER (2004b) and addresses a *competitive* pricing model for an industry in which each of N competing firms sell a distinct item or product brand. The different items or brands are (close) substitutes. This model integrates the complicating factors (A)–(C), i.e. it succeeds in establishing the *simultaneous* incorporation of the complicating factors (B) and (C) into an oligopoly model with price- or quantity competition. Prior work has addressed only *one* of the factors (B) or (C) in the context of competitive pricing (factor (A)).

For example, BERNSTEIN and FEDERGRUEN (2003) address a setting where each firm incurs fixed as well as variable procurement costs along with (linear) inventory carrying costs. However, the model assumes an infinite horizon setting with *time-invariant* demand functions and cost parameters. Here the long-run average operational costs are given by the simple closed-form Economic Order Quantity (EOQ) cost function, i.e. the costs are given by the sum of a term that is proportional with the demand value itself and one that is proportional with the *square root* of the demand value, thus reflecting scale economies. Cachon and Harker (2002) similarly consider, for an industry with two firms, a setting with a single set of *time-invariant* demand functions and with a closed form cost function given by a *concave* power function of the demand volume, (possibly in conjunction with a *linear* cost component), once again to reflect scale economies. Other than the EOQ-cost model above, the authors show that their cost structure arises in a specific service competition model.

At the same time, a stream of marketing papers address competitive pricing problems under time-dependent demand functions, however with simple linear cost functions, and under the assumption that each period's demand is procured in the same period, i.e. no inventories are carried. Perakis and Sood (2003, 2004) and Kachani, Perakis and Simon (2004) also address competitive pricing problems under time-varying demand functions. Since each firm starts the planning horizon with a known inventory and inventories can not be replenished at any time during the horizon, these models consider *no* replenishment costs.

While the demand processes in chapter 2, 3 and 5 are *deterministic*, the fourth chapter represents them as homogeneous Poisson processes.

Chapter 2

Progressive Interval Heuristics for Multi-Item Capacitated Lot-Sizing Problems

2.1 Introduction

This chapter addresses capacitated dynamic lot-sizing models. We consider a family of N items which are produced in the same facility or replenished by the same outside supplier. Demands are specified for each item and each period of a given horizon of T periods. If in a given period an order is placed for some or all of the items, set-up costs are incurred. The aggregate order size is constrained by a capacity limit. The objective is to find a lot-sizing strategy that satisfies the demands for all items over the entire horizon without backlogging, and which minimizes the sum of inventory carrying, fixed and variable order costs. All demands, cost parameters and capacity limits may be time-dependent, reflecting for example, general time-series of forecasts, customer orders, seasonal fluctuations of the cost parameters, or changes in the capacity due to new acquisitions or scheduled maintenance.

In the basic model, the setup cost for an order in any given period only depends on the period index but not on the composition of the order. This assumption is satisfied in many, if not most practical applications, e.g. when the setup cost represents the fixed cost of dispatching a truck or barge or that of initiating a production run in a batch production facility. We refer to this basic case as the Joint Setup cost (JS) model. We extend the model to allow for *item-dependent* setup costs in addition (or in lieu of) the *joint* setup costs and refer to this generalized model as the Joint and Item-dependent Setup cost (JIS)-model.

This capacitated dynamic lot-sizing model is one of the most frequently used deterministic inventory planning models. It needs to be solved repeatedly for each level of a Material Requirements Planning (MRP) or Distribution Requirements Planning (DRP) system with the orders resulting from the capacitated lot-sizing problem(s) at a given level being used as the demand input parameters for the lot-sizing problem to be solved at the next level. The model represents the fundamental challenge of capacity requirements planning while assessing tradeoffs between the costs of holding inventories and the potential of exploiting economies of scale in the procurement costs.

Based on a variety of applications for the BASF and Procter & Gamble corporations as well as a production-distribution problem for the so-called PAMIPS (1995) and MEMIPS (1997) projects, Belvaux and Wolsey (2000, 2001) have developed a prototype optimization system for a class of capacitated multi-item lot-sizing problems which include the (JIS)-model. The system called *bc-prod* uses the extended modelling and optimization library of XPRESS as its engine but allows for simplified problem specification and generates various cutting plane constraints specific to the structure of lot-sizing problems. Bixby (2001) in reviewing the progress and future challenges in CPLEX's mixed integer programming capabilities emphasizes the importance of supply chain management models and within this area, the class of capacitated multi-item lot-sizing problems as being of prime importance and awaiting algorithmic improvements.

The general model is very complex. Florian et al. (1980) have in fact shown that even the single-item case ($N = 1$) is NP-complete, as opposed to the uncapacitated version which, for a planning horizon of T periods, is solvable in $O(T \log T)$ time, see Federgruen and Tzur (1991), Wagelmans et al. (1992), Aggarwal and Park (1993), and in $O(T)$ time under some mild assumptions on the data. The difficulty arises in part because under capacity restrictions, it may no longer be optimal to place an order at the last possible time; in other words, it is not possible to confine oneself to so called zero-inventory ordering policies. Polynomial time algorithms have been developed in the single-item case, but these tend to be time-consuming and restricted to special parameter settings only, see Florian and Klein (1971), Bitran and Yanasse (1982), Chung and Lin (1988) and Van Hoesel and Wagelmans (1996). Recently, Van Hoesel and Wagelmans (2001) (and Gavish and Johnson (1990) for a more restricted version of the model) developed a fully polynomial approximation scheme for the general *single-item* model, i.e. an algorithm which generates an ϵ-optimal solution for any $\epsilon > 0$, in an amount of time which is polynomial in the problem size as well as $\frac{1}{\epsilon}$.

When several items are involved ($N \geq 2$), no efficient solution methods are known, with the exception of ANILY and TZUR (2004a)'s dynamic programming algorithm for the case of *constant* capacities, which is of polynomial complexity when the number of items N is fixed. (This paper also deals with the case where multiple capacitated batches may be ordered in each period. ANILY and TZUR (2004b) develop an exponential search algorithm for the same problem.) It is for this reason that even the more advanced Manufacturing Resource Planning systems (MRPII) start with the determination of system-wide order releases without consideration of capacity constraints, i.e. on the basis of the solution (for each stage or item) of the uncapacitated single item dynamic lot sizing model. It is only in the last phase of the planning process that the elimination of capacity conflicts is attempted by heuristic adaptations of the basic schedules.

FEDERGRUEN and TZUR (1994a) have demonstrated for single-item *uncapacitated* dynamic lot-sizing models that optimal or close-to-optimal initial decisions can be made by truncating the horizon after a relatively small number of periods. A forecast horizon is found in which at most three and usually only two orders are placed (the obligatory order in the first period included). It is reasonable to expect similarly short forecast horizons to continue to apply when multiple items are considered and in the presence of capacity constraints, as long as the utilization rate is not very close to 1. See FEDERGRUEN and TZUR (1994a) for a discussion of how these forecast horizon results relate to capacitated models. This suggests that a close to optimal solution may be generated by partitioning or truncating the horizon.

We therefore develop and analyze a new *class* of so-called *progressive interval* heuristics. A progressive interval heuristic consists of J iterations. In iteration ℓ, the problem is solved to optimality for period 1 to some period T_ℓ, but all *integer* variables for periods 1 to $T_\ell - \tau$ (for some $\tau > 0$) and all *continuous* variables for periods 1 to some $t_\ell \leq T_{\ell-1}$ are fixed at their optimal values after iteration $\ell - 1$. When solving a given interval problem, we append as boundary conditions, the necessary and sufficient conditions for a feasible extension to the remainder of the planning horizon. The horizons are chosen such that $0 = T_0 \leq T_1 \leq \cdots \leq T_J = T$ and $0 = t_1 \leq t_2 \leq \cdots \leq t_J$, while $\tau \geq T_\ell - T_{\ell-1}$ the number of periods by which the horizon in the ℓ-th iteration is expanded. The complexity of any mixed integer programming method is largely determined by the number of (unrestricted) integer variables. Choosing the parameter τ sufficiently small, therefore ensures that the complexity in each iteration grows only modestly. Thus, while the heuristic solves a sequence of progressively larger problem instances, exact solution methods remain viable with only modest increases in computational effort.

We pay special attention to two extreme subsets of this class of heuristics: (i) the *Strict Partitioning heuristics* (SP): here $t_\ell = T_{\ell-1}$ and $T_\ell - T_{\ell-1} = \tau$, with the possible exception of the last interval. The planning horizon is thus *partitioned* into non-overlapping intervals and in the ℓ-th iteration, only the total cost pertaining to the newly appended τ-period interval are minimized, given the boundary conditions (in particular ending inventories) generated in the previous $(\ell - 1)$st iteration; (ii) the *Expanding Horizon heuristics* (EH): here $t_\ell = 0$ for all $\ell = 1, \ldots, J - 1$. A hybrid implementation would e.g. set $t_\ell = [T_\ell - M]^+$ for some window M. The tradeoffs are clear: (EH) [(SP)] provides, within the class of progressive interval heuristics, maximum (minimum) flexibility at the expense of maximum (minimum) incremental computational complexity in adjusting the solution from each iteration to the next.

When applied to the (JS)-model, the (SP)-heuristic can be implemented to be, simultaneously, asymptotically optimal as $T \to \infty$ and to run in $O(NT^2 \log\log T)$ time, provided some of the model parameters are uniformly bounded from above or from below. With the same choice of τ and the same interval choices, the (EH)-heuristic continues to be asymptotically optimal and runs in $O(NT^3)$ time. Our numerical study reveals, however, that it generally results in significantly better solutions than the (SP) heuristic. Both heuristics can also be designed as polynomial time approximation schemes, i.e. to be of polynomial time complexity and to guarantee and ϵ-optimal solution for any $\epsilon > 0$. To our knowledge, these are the first heuristics for multi-item capacitated lot-sizing problems to possess these properties.

While the above theoretical results refer to the (JS)-model, a comprehensive numerical study shows how in particular the (EH)-heuristic, can be effectively used for the *general* (JIS)-model (with period- and item-dependent setup costs) as well. For the latter, it is possible to find the *optimal* solution for instances with up to 150–200 setup variables (e.g. when $N = 10$ and $T = 15$ or 20). For these problem sizes, the (EH)-heuristic, generates close-to-optimal solutions with an optimality gap of up to 2% across a large set of parameter combinations. (The (SP)-heuristic, while significantly faster, often generates solutions with optimality gaps above 10%.)

While *exact* optimality gaps can not be measured for larger problem instances, our theoretical results show that (at least for the (JS) model) optimality gaps can be expected to be even lower as T, the length of the planning horizon, increases. We systematically evaluate the performance of both the (SP) and the (EH)-heuristic for problem instances with the number of items varying from 10 to 25 and the horizon length varying from 10 to 50 in the (JIS) and up to 100 in the (JS)-model. An earlier numerical study for the *single* item problem in FEDERGRUEN and TZUR (1994b) shows that problems with up to 100 periods can be solved by a slight variant of the (SP)-heuristic with an optimality gap of less than 7% and, on average, equal to 2%.

Summarizing, the main contributions of this chapter are (i) the design of a new class of heuristics; (ii) the demonstration that, for (JS), both the (SP)- and (EH)-heuristics can be designed to be of low polynomial complexity *as well as* asymptotically optimal; (iii) the proof that for finite T, both (SP)- and (EH) can be designed to to be polynomial time approximation schemes; (iv) the demonstration that a progressive interval heuristic generates close-to-optimal solutions with modest computational effort, even for large scale problems.

While our theoretical and numerical analysis are based on the (JS) and (JIS) models, we believe that the effectiveness of the progressive interval heuristics bodes well for its use in general multi-period production and inventory problems.

The remainder of this chapter is organized as follows: Section 2.2 reviews the relevant literature. In Section 2.3 we introduce the (JS)-model and its notation. In Section 2.4 we describe the new class of heuristics and develop worst case bounds for their optimality gaps. In Section 2.5, we discuss how each interval problem, which arises in an iteration of the heuristic, can be solved effectively via a general purpose mixed integer programming method or a tailor-made branch-and-bound method. This allows us to identify implementations that are simultaneously *asymptotically optimal* as well as of very reasonable and *polynomial complexity*. Finally, Section 2.6 discusses extensions to the general (JIS)-model as well as the numerical study.

2.2 Literature Review

In this section, we provide a brief review of the existing literature, beyond the papers mentioned in the introduction.

CHEN ET AL. (1994) and SHAW and WAGELMANS (1998) developed two relatively efficient pseudo-polynomial solution methods for the general single item model. Their extensions to the multi-item model result in dynamic programs with a state space of dimension N and larger, and are therefore entirely unusable except for the smallest possible number of items N. As mentioned, even for the single-item model, this chapter's heuristics are, to our knowledge, the first to be asymptotically optimal and of polynomial complexity.

All other existing methods are based on heuristics and none has provable bounds for the associated optimality gaps. These heuristics can be divided into simple constructive heuristics and mathematical programming based heuristics. The constructive heuristics include "greedy methods" in which a specific sequence is proposed to assign the capacity of a given period to satisfy its or later demand, e.g. EISENHUT (1975), LAMBRECHT and VANDER EECKEN (1978), DIXON and SILVER (1981), and MAES and VAN WASSENHOVE (1986). Other constructive heuristics start with the solution of the uncapacitated model, and search for a feasible production schedule by simple shifting routines, e.g. VAN NUNEN and WESSELS (1978), DOGRAMACI ET AL. (1981), NAHMIAS

(1989) and KARNI and ROLL (1982).

The mathematical programming based heuristics employ linear programming, lagrangean relaxation, cutting plane methods and column generation techniques. Examples include BAKER and DIXON (1978), EPPEN and MARTIN (1987), POCHET (1988), LEUNG ET AL (1989), MARTIN (1987) and TRIGEIRO ET AL. (1989). We refer to MAES and WASSENHOVE (1988), SALOMON (1990) and KUIK ET AL (1994) for detailed surveys of these methods until 1994.

State-of-the-art solution methods include, in addition to the bc-prod system mentioned in Section 2.1 (BELVAUX and WOLSEY (2000, 2001)), STADTLER (2003) and SUERIE and STADTLER (2003). Interestingly, these methods all apply variants of the (EH)-heuristic: In the 'fix-and-relax' heuristic, each consecutive problem instance expands the horizon of the previous instance by appending the same number (τ) of periods to its tail. The 'internally rolling schedule heuristics' in STADTLER (2003) and SUERIE and STADTLER (2003) use constant interval increments $\leq \tau$. In each problem instance, instead of imposing boundary conditions that are necessary and sufficient for a feasible extension till the end of the full planning horizon, the authors include the periods beyond the end of the current interval, however with all binary variables in these periods treated either as continuous variables (bc-prod) or set equal to one (STADTLER (2003) and SUERIE and STADTLER (2003)). The heuristics in the latter two papers substitute all cost parameters for the after-the-interval periods by zero, with the possible exception of variable overtime cost rates, in case the capacity constraints may be violated by scheduling overtime. (Additional heuristic changes are applied to an interval's last set of periods.)

Federguen and Tzur (1994c) describe an effective heuristic for the so-called Joint Replenishment Problem (JRP), which is similar to the (SP)-heuristic. (This heuristic can be designed to be asymptotically optimal and of polynomial complexity, under specific parameter conditions). The (JRP)-model is the special case of the (JIS)-model, which arises when no capacity constraints prevail. Federgruen and Tzur (1999) describe a general framework for a variant of the (SP)-heuristic, with applications to other types of lot-sizing problems.

2.3 The multi-item model with joint setup cost (JS)

In this section we discuss our basic model (JS) with joint setup costs only. We use the index $i \in \{1, \ldots, N\}$ to distinguish between items and the index $t \in \{1, \ldots, T\}$ to distinguish between periods. For $i = 1, \ldots, N$ and $t = 1, \ldots, T$, we specify the following parameters:

d_{it} = demand for item i in period t; $(d_{it} \geq 0)$

D_t = aggregate demand in period $t = \sum_{i=1}^{N} d_{it}$

c_{it} = variable per unit order cost for item i in period t

h_{it} = cost of carrying a unit of inventory of item i at the end of period t

K_t = setup cost incurred when an order is placed in period t

C_t = order capacity, i.e. the maximum number of units which can be ordered in period t.

Without loss of generality, we define the units of the items such that ordering one unit of an item consumes one unit of capacity. We define the following decision variables:

$$x_{it} = \text{order size for item } i \text{ in period } t;\ i = 1,\ldots,N;\ t = 1,\ldots,T$$

$$Y_t = \begin{cases} 1 & \text{if } \sum_{i=1}^{N} x_{it} > 0 \\ 0 & \text{otherwise} \end{cases} \qquad t = 1,\ldots,T$$

$$I_{it} = \text{ending inventory of item } i \text{ in period } t;\ i = 1,\ldots,N;\ t = 1,\ldots,T$$

Let I_t^0 = the *minimum* aggregate inventory at the end of period t, such that a feasible production / inventory plan exists for periods $t+1,\ldots,T$. These minimum stock levels are easily computed from the following recursion, which can be verified by induction:

$$I_t^0 = \left(D_{t+1} - C_{t+1} + I_{t+1}^0\right)^+, \qquad t = 1,2,\ldots,T-1, \quad \text{with} \quad I_T^0 = 0 \tag{2.1}$$

The multi-item model can thus be formulated as follows:

$$\text{(P)} \qquad z^* = \min\left\{\sum_{t=1}^{T}\left[K_t Y_t + \sum_{i=1}^{N}(c_{it}x_{it} + h_{it}I_{it})\right]\right\} \tag{2.2}$$

s.t.

$$I_{it} = I_{i(t-1)} + x_{it} - d_{it}, \qquad i = 1,\ldots,N, \quad t = 1,\ldots,T \tag{2.3}$$

$$\sum_{i=1}^{N} x_{it} \le C_t Y_t, \qquad t = 1,\ldots,T \tag{2.4}$$

$$\sum_{i=1}^{N} I_{it} \ge I_t^0, \qquad t = 1,\ldots,T \tag{2.5}$$

$$x_{it} \ge 0; \quad I_{it} \ge 0; \quad Y_t \in \{0,1\} \tag{2.6}$$

The above formulation is often referred to as the network formulation. The plant location formulation is an alternative that disaggregates the production quantities $\{x_{it}\}$

into $\{x_{ist}\}$ with x_{ist} = the amount of item i, ordered in period s to satisfy demand in period t.

2.4 Progressive interval heuristics: worst case bounds for optimality gaps

A progressive interval heuristic solves a sequence of J problem instances. The first instance considers the capacitated lot sizing problem that arises when restricting oneself to the first T_1 periods, i.e., it solves (P) with T replaced by T_1. In each of the subsequent instances, a given number of periods $\leq \tau$ is appended to the tail of the previous planning horizon. In the h-th iteration, a lotsizing problem $(\tilde{JS}_h)$ is solved on the complete interval $\{1, \ldots, T_h\}$, albeit that all $Y-$variables of periods $1, \ldots, T_h - \tau$ are fixed at their optimal values in the $h-1^{st}$ iteration, i.e. when solving $(\tilde{JS}_{h-1})$. Recall $T_h - T_{h-1} \leq \tau$, i.e. $T_h - \tau \leq T_{h-1}$. Thus, the number of unrestricted binary variables in each iteration remains constant, i.e. equal to τ. Moreover, the aggregate ending inventory in period T_h is constrained from below by the I^0-value.

Different progressive interval heuristics give varying amounts of flexibility to the continuous variables in each of the J problem instances. As mentioned, we focus in particular on two *extremes*: under the Strict Partitioning heuristics (SP), all interval increments $(T_\ell - T_{\ell-1}) = \tau$ with the possible exception of the last interval. Also, among the continuous variables, only those pertaining to the last τ periods of the current planning horizon are allowed to be chosen freely, (i.e. $t_\ell = T_\ell - \tau = T_{\ell-1}$) without any restrictions beyond those implied by the constraints of (P); all other continuous variables are fixed at their optimal value in the previous problem instance.

Under the expanding horizon (EH)-heuristics, *all* of the continuous variables are allowed to be varied fully (subject, of course, to the constraints (2.3) - (2.6)), i.e. $t_\ell = 0$; moreover this class allows for $T_\ell - T_{\ell-1} < \tau, \ell = 1, \dots, J-1$. If the stepsizes $(T_\ell - T_{\ell-1}) < \tau$, even many of the setup decisions determined in one iteration of the algorithm, may be revisited in subsequent iterations, on the basis of additional demand, cost and capacity information pertaining to additional periods. (We have observed that it is often effective to append a single period as one progresses from one interval to the next, i.e. $T_\ell - T_{\ell-1} = 1$.)

Various intermediate implementations may be envisioned; for example, a (moving) window of $M > \tau$ periods may be used such that the continuous variables in (up to) the last M periods are unrestricted, as opposed to the last τ (SP) or *all* periods under (EH). Federgruen and Tzur (1999) consider a slight variant of (SP) under which the size and the composition of the last (or several of the last) order periods in the previously solved iteration may be varied, along with the production quantities of the newly appended periods; see ibid for details.

Let z^{SP} and z^{EH} denote the cost of the solutions found by the (SP)- and the (EH)-heuristics, for a given choice of $\{T_\ell, t_\ell, \tau\}$. We now derive worst case bounds for their optimality gaps, under mild conditions for the cost, demand and capacity parameters. We first derive a lower bound for z^* as an explicit function of T. It is quite simple to obtain a lower bound when assuming *all* periods' demands are uniformly bounded away from zero; however, to allow for sporadic demands, we derive an alternative bound, merely assuming that the cumulative demand over a large enough time interval is uniformly bounded away from zero. Its proof, while similar to that in FEDERGRUEN and TZUR (1994c), requires major adjustments to reflect the capacity limits.

Theorem 2.1 *Assume there exists a positive integer $\theta \geq 1$, and for all $i = 1, \ldots, N$ positive constants d_{i*} such that*

$$\begin{aligned} d_{it} + \ldots + d_{i(t+\theta-1)} &\geq \theta d_{i*} \qquad i = 1, \ldots, N, \quad t = 1, \ldots, T - \theta + 1 && (2.7) \\ \sum_{t=1}^{T} d_{it} &\geq T d_{i*} \qquad i = 1, \ldots, N && (2.8) \end{aligned}$$

In addition, assume there exist constants K_* and C^* and for each $i = 1, \ldots, N$ constants h_{i*} and c_{i*} such that $K_t \geq K_*, C_t \leq C^*$, $h_{it} \geq h_{i*}$ and $c_{it} \geq c_{i*}$ for all $t = 1, \ldots, T$. Let $d_* = \sum_{i=1}^{N} d_{i*}$, $\kappa = \sum_{i=1}^{N} c_{i*} d_{i*}$, $H_* = \frac{1}{2} \sum_{i=1}^{N} h_{i*} d_{i*}$. Then, $z^* \geq \gamma T$ where

$$\gamma \stackrel{\text{def}}{=} \kappa + \begin{cases} \frac{K_*}{2\theta} & \textit{if} \quad \sqrt{\frac{H_*}{\hat{K}}} \geq \frac{1}{2\theta} \\ \left(2\sqrt{(K_* + 2H_*\theta^2)H_*} - 3H_*\theta\right) & \textit{if} \quad \frac{d_*}{C^*} < \sqrt{\frac{H_*}{\hat{K}}} < \frac{1}{2\theta} \\ \left(\frac{(K_* + 2H_*\theta^2)d_*}{C^*} + \frac{H_* C^*}{d_*} - 3H_*\theta\right) & \textit{if} \quad \sqrt{\frac{H_*}{\hat{K}}} \leq \frac{d_*}{C^*} < \frac{1}{2\theta} \end{cases} \qquad (2.9)$$

Proof: We obtain a lower bound by replacing all fixed order costs by K_*, all capacities by C^* and for each item $i = 1, \ldots, N$ all variable order cost rates by c_{i*}, and holding cost rates by h_{i*}. We refer to the resulting problem as the *transformed problem.* Consider a solution in which $m \geq 1$ orders are placed. For $\ell = 1, \ldots, m$ let n_ℓ denote the number of periods in the ℓ^{th} order cycle, i.e. the interval which contains the ℓ^{th} order period and all subsequent periods prior to the next order interval (if any). (The m^{th} interval terminates with period T.)

We first derive a lower bound for the total *holding* costs incurred in a single order cycle of n periods in the transformed problem. Note that zero-inventory ordering may fail to be optimal in the capacitated model, i.e. the starting inventory in the first period may be positive for some or all items. However, the holding cost in the order cycle is clearly bounded from below by assuming that the starting inventory equals zero.

Renumber the periods in this cycle as $1, \dots, n$ and let $n = \psi\theta + \tau$ with $0 \le \tau < \theta$, i.e., $\psi = \left\lfloor \frac{n}{\theta} \right\rfloor$. Fix $i = 1, \dots, N$. Observe by our assumption that in each of the intervals $[(j-1)\theta + \tau + 1, j\theta + \tau]$ for $j = 1, \dots, \psi$ at least θd_{i*} units are demanded for item i. Being ordered in or after period 1, the lowest holding costs for these demands arise when θd_{i*} units are demanded in period $(j-1)\theta + \tau + 1$ (i.e. in the *first* period of this interval) and none in the remaining periods of the interval $[(j-1)\theta + \tau + 1, j\theta + \tau]$. It follows that the holding costs in a single order cycle of n periods are bounded from below by
$\sum_i h_{i*}\theta d_{i*} \sum_{j=0}^{\psi-1} (\tau + j\theta) = \sum_i h_{i*}\theta d_{i*} \left[\psi\tau + \frac{1}{2}\theta\psi(\psi - 1)\right]$
$= \sum_i h_{i*}\theta d_{i*} \left[\left\lfloor \frac{n}{\theta} \right\rfloor \tau + \frac{1}{2}\theta \left\lfloor \frac{n}{\theta} \right\rfloor \left(\left\lfloor \frac{n}{\theta} \right\rfloor - 1\right)\right] \ge \sum_i \frac{1}{2} h_{i*}\theta^2 d_{i*} \left[\left(\frac{n}{\theta} - 1\right)^+ \left(\frac{n}{\theta} - 2\right)^+\right]$
$= g(n)$ where $g(x) \overset{def}{=} H_*\theta^2 \left(\frac{x}{\theta} - 1\right)^+ \left(\frac{x}{\theta} - 2\right)^+$ is convex.

This implies the following lower bound for the *total* cost over the complete horizon:

$$\begin{aligned} z^* &\ge \kappa T + \min_m \left\{ K_* m + \min_{n_\ell} \left[\sum_{\ell=1}^m g(n_\ell) : \sum_{\ell=1}^m n_\ell = T \right] | m \ge \frac{Td_*}{C^*} \right\} \\ &= \kappa T + \min_m \left\{ K_* m + \left(H_*\theta^2\right) m \left(\frac{T}{m\theta} - 1\right)^+ \left(\frac{T}{m\theta} - 2\right)^+ \right. \\ &\qquad \left. \left| \frac{Td^*}{C^*} \le m \le \max\left(\frac{Td^*}{C^*}, \frac{T}{2\theta}\right) \right. \right\} \end{aligned} \tag{2.10}$$

The lower bound for m may be imposed because when $mC^* < Td_*$, it is infeasible to satisfy all demand. The equality in (2.10) follows, since, by the convexity of $g(\cdot)$, equal values $n_\ell = \frac{T}{m}, \ell = 1, ..., m$ achieve the minimum to its left. (The upper bound for m may be imposed since the minimand to the right of (2.10) is increasing for $m > \frac{T}{2\theta}$.) (2.9) follows from (2.10), noting that for $m \leq \frac{T}{2\theta}$, the $(\cdot)^+$ operators may be ignored. ■

We now derive an upper bound for the optimality gap of the (SP)- and (EH)-heuristic. The bound is established under the parameter conditions of the lower bound Theorem 2.1, a uniform lower (upper) bound for the capacities (holding cost rates) and a condition which specifies that a uniform slack capacity exists over any cycle of θ periods, i.e.

(S) there exists a constant $\sigma > 0$ and an integer ζ such that

$$\sum_{r=t+1}^{t+\zeta} C_r \geq \sum_{r=t+1}^{t+\zeta} D_r + \sigma \quad \text{for all} \quad t = 0, \ldots, T - \zeta. \tag{2.11}$$

We first need the following lemma which shows that under condition (S) a uniform upper bound prevails for all minimum reserve stocks $\left\{I_t^0\right\}$:

Lemma 2.1 *Let condition (S) hold and assume a constant C^* exists such that $C_t \leq C^*$. Then*

$$I_t^0 \leq U \stackrel{\text{def}}{=} \zeta C^* - \sigma \qquad t = 1 \ldots, T. \tag{2.12}$$

Proof. By repeated substitutions in (2.1), we get for all t = 1, ..., T:
$I_t^0 = \max_{t+1 \leq s \leq T} \left[\sum_{r=t+1}^{s} (D_r - C_r)\right]^+ = \max_{t+1 \leq s \leq \min(T, t+\zeta-1)} \left[\sum_{r=t+1}^{s} (D_r - C_r)\right]^+$ $\leq \max_{t+1 \leq s \leq \min(T,t+\zeta-1)} \sum_{r=t+1}^{s} D_r \leq \sum_{t=t+1}^{\min(T,t+\zeta-1)} D_r$ where, by (S), the second equality follows from $\sum_{r=t+1}^{s} (D_r - C_r) \geq \sum_{r=t+1}^{s-\zeta} (D_r - C_r)$ for $s \geq t + \zeta$. Thus, I_t^0 can be

bounded by a sum of ζ consecutive aggregate demands, hence by a sum of ζ consecutive capacity values minus σ, given (2.11). This proves (2.12). ■

Theorem 2.2 *Let (S) hold. Assume there exists an integer $\theta \geq 1$ and for each $i = 1, \ldots, N$ a constant d_{i*} such that:*

$$(d_{it} + \ldots + d_{i,t+\theta-1}) \geq \theta d_{i*}, \quad t = 1, \ldots, T - \theta + 1; \tag{2.13}$$

$$\sum_{t=1}^{T} d_{it} \geq T d_{i*}. \tag{2.14}$$

In addition, assume there exist constants K^*, K_*, C^* and C_* and for each $i = 1, \ldots, N$ constants $h_{i*}, h_i^*, c_{i*}, c_i^*$ such that for all $t \geq 1, K_* \leq K_t \leq K^*$, $C_* \leq C_t \leq C^*$, $h_{i*} \leq h_{it} \leq h_i^*$ and $c_{i*} \leq c_{it} \leq c_i^*$. Let $\Delta c^* = \max_i \left[c_i^* - c_{i*}\right]$, $\eta = \frac{K^*}{C_*} + \Delta c^*$, $h_* = \min_i h_{i*}$, $\Delta h^* = \max[h_i^* - h_{i*}]$ and $D_* = \sum_{i=1}^{N} d_{i*}$. Let γ be defined as in (2.9) and

$$\rho_1 = K^* + C^* \left(\left\lfloor \frac{\eta}{h_*} \right\rfloor \eta - \frac{1}{2} \left\lfloor \frac{\eta}{h_*} \right\rfloor \left(\left\lfloor \frac{\eta}{h_*} \right\rfloor + 1 \right) h_* \right). \tag{2.15}$$

$$\rho_2 = U \left[(\Delta c^* + K^*) + \left(\left\lfloor \frac{U}{\sigma} \right\rfloor + 1 \right) \zeta \Delta h^* + \left(\frac{\Delta c^* + K^*}{h_*} \right) \Delta h^* \right]. \tag{2.16}$$

$$\rho = \rho_1 + \rho_2. \tag{2.17}$$

Then,

$$(a) \qquad \frac{z^{SP} - z^*}{z^*} \leq \frac{(J-1)\rho}{\gamma T}, \qquad (b) \qquad \frac{z^{EH} - z^*}{z^*} \leq \frac{(J-1)\rho}{\gamma T}.$$

Proof: (a) We show that an optimal solution of the complete problem can be transformed, in two phases, into one which is achievable by the (SP)-heuristic, adding at most $(J-1)\rho$ to the total cost. In Phase I, the optimal solution is transformed into one with all intervals' ending *aggregate* inventory *equal* to their minimum I^0-level. In Phase II, the composition of the reserve stock at the end of each of the intervals is made identical to that of the solution of the (SP)-heuristic.

To describe the transformation in Phase I, renumber the periods in the first ℓ intervals from $1,\ldots,T_\ell$, starting with T_ℓ and going backwards, i.e. period t is now renumbered as $T_\ell - t + 1, t = 1,\ldots,T_\ell$. With this numbering, period t occurs t periods before the end of the ℓ^{th} interval.

In the optimal solution, let Q_{ir} denote the number of units of item i ordered in period r to satisfy demands in some future period in the $(\ell+1)^{st}$ or later intervals $(i = 1,\ldots,N, r = 1,\ldots,T_\ell)$. Also, let $Q_r = \sum_i Q_{ir}$. The starting aggregate inventory of the $(\ell+1)^{st}$ interval is $= \sum_{r=1}^{T_\ell} Q_r > I_1^0$. Since a feasible solution exists for $(JS_{\ell+1})$ with a starting inventory of I_1^0 *only*, it is *feasible* to postpone the orders for $\left(\sum_{r=1}^{T_\ell} Q_r\right) - I_1^0$ units to periods that belong to the $(\ell+1)^{st}$ interval itself. The transfer of these order quantities requires at most $\left\lceil \frac{\left(\sum_{r=1}^{T_\ell} Q_r\right) - I_1^0}{C_*} \right\rceil$ additional setups in the $(\ell+1)^{st}$ interval, and therefore at most $\left\lceil \frac{\left(\sum_{r=1}^{T_\ell} Q_r\right) - I_1^0}{C_*} \right\rceil K^* \le \left\lceil \frac{\left(\sum_{r=1}^{T_\ell} Q_r\right)}{C_*} \right\rceil K^* \le K^* + \frac{K^*}{C_*} \sum_{r=1}^{T_\ell} Q_r$ in additional setup costs. An upper bound for the *total* additional costs due to the transfer of these order quantities is therefore given by:

$$\max \left\{ \sum_{i=1}^{N} \sum_{r=1}^{T_\ell} \left[c_i^* - c_{ir} - \sum_{s=1}^{r} h_{i,s} \right] Q_{i,r} + \frac{K^*}{C_*} \sum_{r=1}^{T_\ell} \sum_{i=1}^{N} Q_{ir} + K^* \;\middle|\; 0 \le \sum_{i=1}^{N} Q_{ir} \le C_r, \forall r \right\}.$$

This linear progam decomposes into T_ℓ single constraint problems. Each is straightforwardly solved in closed form: for each $r = 1,\ldots,T_\ell$, set $Q_{ir} = C_r$ for any item i whose objective function coefficient is largest, unless all $\{Q_{ir} : i = 1, \ldots, N\}$ - variables have negative coefficients, in which case it is optimal to set $Q_{ir} = 0$ for all $i = 1, \ldots, N$. This results in the upper bound $K^* + \sum_{r=1}^{T_\ell} \max_i \left[\frac{K^*}{C_*} + c_i^* - (c_{i,r} + h_{i,r} + h_{i,r-1} + h_{i,1})\right]^+ C_r$ $\leq K^* + \sum_{r=1}^{T_\ell} [\eta - rh_*]^+ C_r \leq K^* + C^* \sum_{r=1}^{\Lambda} [\eta - rh_*] \leq K^* + C^* \left\{\Lambda\eta - \frac{1}{2}\Lambda(\Lambda+1)h_*\right\}$ $= \rho_1$ where $\Lambda = \left\lfloor \frac{\eta}{h_*} \right\rfloor$ is an upper bound on the number of periods in which inventory may be held prior to the ℓ-th interval for use during the ℓ-th or later intervals. (The first equality follows from the fact that the $\Lambda + 1$-st until the T_ℓ-th term in the sum to its left vanish). Apply the transfer process sequentially to the intervals $\ell = J-1, J-2, \ldots, 1$ to end up with a solution in which all intervals' ending aggregate inventory *equals* the minimum I^0-level and whose cost exceeds z^* by at most $(J-1)\rho_1$.

Let L_ℓ denote the longest shelf life of any unit in stock at the end of the ℓ-th interval $\ell = 1, \ldots, J-1$. In Phase II, we transform the Phase I solution by changing the item identity of at most $I^0_{T_\ell}$ units in stock at the end of period T_ℓ, without any *additional* changes in the order and inventory plan. This maintains feasibility, leaves total set-up costs unaltered and adds at most:

$$\sum_{\ell=1}^{J-1} I^0_{T_\ell} \left(\Delta c^* + L_\ell \Delta h^*\right) \tag{2.18}$$

variable order and holding costs. In view of Lemma 2.1, to show that the summand in (2.18) is bounded by ρ_2, it suffices to show that $L_\ell \leq \left(\left\lfloor \frac{U}{\sigma} \right\rfloor + 1\right)\zeta + \frac{\Delta c^* + K^*}{h^*} \stackrel{\text{def}}{=} \bar{L}$.

Assume first that at least one of the periods $t^* \in \left\{T_\ell - \left(\left\lfloor \frac{U}{\sigma} \right\rfloor + 1\right)\zeta + 1, \ldots, T_\ell\right\}$ has slack capacity (in the Phase I solution). In this case, if one of the $I^0_{T_\ell}$ units in the reserve stock has a shelf life of more than $\bar{L}$ periods, the ordering of this unit can be postponed until t^*, thereby reducing inventory costs by at least $h_*\left(\bar{L}-\left(\left\lfloor\frac{U}{\sigma}\right\rfloor+1\right)\zeta\right) = h_*\frac{\Delta c^*+K^*}{h_*} = \Delta c^* + K^*$, offsetting any increase in the variable ordering cost (and possibly one setup cost), due to the postponement. Thus, if any of the $I^0_{T_\ell}$ units has a shelf life larger than $\bar{L}$, a full capacity order is placed in each period of the interval $\left[T_\ell-\left(\left\lfloor\frac{U}{\sigma}\right\rfloor+1\right)\zeta+1,\ldots,T_\ell\right]$, resulting in an ending inventory of at least $\sum_{t=T_\ell-(\lfloor\frac{U}{\sigma}\rfloor+1)\zeta+1}^{T_\ell}(C_t-D_t) \geq \left(\left\lfloor\frac{U}{\sigma}\right\rfloor+1\right)\sigma > U$ units, which contradicts Lemma 2.1.

(b): Let $I^{(\ell)}$ denote the N-vector of ending inventories at the end of the ℓ-th interval, as determined in the ℓ-th iteration of the (EH)-heuristic, $\ell = 1,\ldots,J$, and let $\left\{Y^{EH}_t : t = 1,\ldots,T\right\}$ be the Y-vector chosen by this heuristic. Transform the optimal solution into a solution $\pi^{(II)}$ with cost value $z^{(II)}$ via Phase I and Phase II transformations as in part (a) except that in Phase II the ℓ-th interval's vector of ending inventories is now matched to $I^{(\ell)}$. With $T_{-1} = T_0 = 0$, let $\pi^{(\ell)}$ be an optimal solution of the mixed integer program (P^ℓ), where $\ell = 0,\ldots,J$.

$$(P^\ell): \quad z^{(\ell)} = \min(2.2) \tag{2.19}$$

$$\text{s.t.} \quad (2.3) - (2.6) \tag{2.20}$$

$$I_{iT_h} = I^{(h)}_i \quad i = 1,\ldots,N, \quad h = \max(\ell-1,1),\max(\ell,1),\ell+1,\ldots,J \tag{2.21}$$

$$Y_t = Y^{EH}_t \quad t = 1,\ldots,T_{\ell-1}. \tag{2.22}$$

(P^{ℓ^*+1}) is obtained from (P^{ℓ^*}), by simultaneously *adding* the constraints $Y_t = Y^{EH}_t, t = T_{\ell^*-1}+1,\ldots,T_{\ell^*}$ and *eliminating* the constraints $I_{iT_{\ell^*-1}} = I^{(\ell^*-1)}_i$, $i = 1,\ldots,N$. Since $\pi^{(\ell^*)}$ satisfies (2.21) and (2.22) for $\ell = \ell^*$, i.e. since it maintains the same ending inventories at the end of the ℓ^*-th interval as the EH-heuristic does at the end of the ℓ^*-th iteration, and since it is restricted to the same order periods in the first $(\ell^* - 1)$ intervals as the (EH)-heuristic is in its ℓ^*-th iteration, it follows that both $\pi^{(\ell^*)}$ and the solution obtained by the (EH)-heuristic in its ℓ^*-th iteration, minimize total costs over the first T_{ℓ^*} periods subject to the constraints (2.21) - (2.22) with $\ell = \ell^*$. This implies that $\pi^{(\ell^*)}$ can be chosen such that $Y_t = Y_t^{EH}, \quad t = T_{\ell^*-1}+1, \ldots, T_{\ell^*}$ and hence $Y_t = Y_t^{EH}$ for all $t = 1, \ldots, T_{\ell^*}$. Thus, $\pi^{(\ell^*)}$ is a feasible solution of (P^{ℓ^*+1}) so that

$$z^{EH} = z^{(J)} \leq z^{(J-1)} \leq \cdots \leq z^{(0)} \leq z^{(II)} \leq z^* + (J-1)\rho, \tag{2.23}$$

where the equality follows from the (EH)-solution optimizing $P^{(J)}$, the last inequality from part (a) and the one before that from $\pi^{(II)}$ being a feasible solution of $P^{(0)}$. ■

Remark: The proof of Theorem 2.2 reveals that a tighter bound, with ρ replaced by a smaller value, may be computed in any given instance, once the number of intervals and their lengths have been specified.

2.5 Solution methods for a single interval problem: polynomial and asymptotically optimal heuristics

We now discuss how a single interval problem in an iteration of the progressive interval heuristic can be solved effectively. We have found that the *general* purpose branch-and-bound method embedded in CPLEX is very effective to solve (JS) problems; see Section 2.6 for details. Alternatively, several tailor-made branch-and-bound methods can be used. Below, we discuss *three* such methods. Two of them have the distinct advantage over the CPLEX-based algorithm that their complexity, for the (SP)-heuristic, is of the order $O(2^{\tau}P(\tau))$ with $P(\cdot)$ a polynomial in τ (The complexity is $O(2^{\tau}P(T))$ for the (EH)-heuristic.). Theorem 2.2 shows that the two heuristics can be designed to be *asymptotically optimal*, e.g. by choosing every (except possibly the last) interval increment $T_{\ell} - T_{\ell-1} = \tau, \ell = 1, \ldots, J-1$ with

$$\tau = \lceil \alpha \log T \rceil \qquad \text{for some} \qquad \alpha > 0 \tag{2.24}$$

or more generally by choosing $\tau = o(T)$, as $T \to \infty$ (The last increment $T_{\ell} - T_{\ell-1} = T - \left\lfloor \frac{T}{\tau}\tau \right\rfloor$). Thus, by choosing τ as in (2.24), we obtain an algorithm which is *simultaneously* asymptotically optimal and of polynomial complexity.

Our three branch-and-bound methods are based on three bounds for the value of z^*.

z^{LB_1} = minimum cost value in the uncapacitated model, i.e. ignoring constraints (2.4)

z^{LB_2} = $\max_{\lambda \geq 0} z(\lambda)$ where

$$z(\lambda) = \min \left\{ \sum_{t=1}^{T} (K_t Y_t + \sum_{i=1}^{N} (c_{it} x_{it} + h_{it} I_i) + \lambda_t [C_t Y_t - \sum_i x_{it}]) \right.$$
$$\text{s.t. } (2.3), (2.5), (2.6) \}.$$

In other words, z^{LB_2} is the value of the Lagrangean dual associated with the relaxation of the capacity constraints (2.5). Clearly, $z^{LB_2} \geq z(0) = z^{LB_1}$.

z^{LB_3} = $z^{LB}\text{var} + z^{LB}\text{fix}$, where

z^{LB}var = minimum value of the variable costs, i.e. minimum cost value when all setup costs are reduced to zero, and

z^{LB}fix = minimum value of the fixed (setup) costs required to satisfy all demands when in each period t the best observed and yet unused setup cost and capacity value can be used (instead of only K_t and C_t being available).

Therefore, z^{LB}fix is a lower bound on the minimum value of the fixed costs, i.e.

$$\begin{aligned} z^* &\geq \min\left\{\sum_{t=1}^{T}\left[\sum_{i=1}^{N} c_{it}x_{it} + \sum_{i=1}^{N} h_{it}I_{it}\right] \text{ s.t. (2.3) - (2.6)}\right\} \\ &\quad + \min\left\{\sum_{t=1}^{T} K_t Y_t \text{ s.t. (2.3)-(2.6)}\right\} \\ &= z^{LB}\text{var} + z^{LB}\text{fix} = z^{LB_3} \end{aligned} \tag{2.25}$$

In the single item case (N=1), z^{LB_1} can clearly be evaluated via any of the solution methods for the single uncapacitated model. (This can be done in O($T \log T$) time, see the Introduction.) In the multi-item case, evaluation of z^{LB_1} reduces to the solution of the joint replenishment problem (JRP) without item-specific setup costs. In the important special case where no speculative motives for carrying inventory prevail, the complexity of this method is easily verified to be $O(NT^2)$, see FEDERGRUEN and TZUR (1994c). For general variable holding and order costs, any of the known lower bounds for the JRP can be invoked, e.g. the bound in FEDERGRUEN and TZUR (1994c) which requires $O((N + K^*)T \log T)$ time where $K^* = \max_t K_t$.

To evaluate z^{LB_2}, the above methods need to be embedded in an unconstrained optimization technique which searches for the maximizing vector λ.

z^{LB_3} is the sum of two components: z^{LB}var is the minimum cost network flow in a network of special structure. Ahuja and Hochbaum (2004, §6.3)'s algorithm solves this problem in $O(NT \log T)$ time. To compute z^{LB}fix, observe that it is optimal to sequentially postpone setups until the last feasible period, since in any given period, any *prior* (unused) capacity and setup cost value may be chosen. Thus, assume that the first j setup periods $t(1), t(2), \ldots, t(j)$ have been determined, together with their "adopted" capacities and setup cost values; the next setup period $t(j+1)$ (if any) is then obtained as the first period t after $t(j)$ for which $\sum_{s=1}^{t} D_s$ is in excess of the sum of the adopted capacities for periods $t(1), \ldots, t(j)$; it is then optimal to assign to this setup period the best observed and yet unused setup cost and capacity value. This sequence of setup periods (and associated setup costs and capacity values) can be determined in $O(T \log T)$ time, by maintaining two ordered lists of unused capacity and setup cost parameters. Thus, z^{LB_3} can be computed in $O(NT \log T)$ time.

Branch-and-bound methods

Our branch and bound algorithm bears the following similarities to that in Federgruen and Tzur (1994c): 1) it implicitly enumerates all possible subsets of the τ undetermined order periods; 2) it characterizes each node of the b & b tree by a partition of the periods into sets S^+, S^- and S^0, with S^+ the set of periods in which one is committed to place an order, S^- the set in which no order is allowed and S^0 the set of periods where no decision is fixed yet; 3) the root of the tree has all τ periods in the set S^0 and every non-terminal node has two successor nodes, one with an additional period shifted from S^0 to S^+ and one with the same period shifted to S^-. (This period is selected according to a specific branching rule.) At any of the leaf nodes, for a given set of order periods, the problem reduces to a polynomially solvable network problem.

Compared to Federgruen and Tzur (1994c), a different lower bound is used to evaluate each node of the b & b tree. For all $r = 1, 2, 3$, and a given node characterized by S^+, S^-, S^0 let, $Z^{LB_r} = \sum_{i \in S^+} K_i$ + the value of z^{LB_r} when the setup cost for periods $i \in S^+(S^-)$ is changed to $0(\infty)$ and the capacity for periods $i \in S^-$ is changed to 0. Each of the values $Z^{LB_1}, Z^{LB_2}, Z^{LB_3}$, can be used as a lower bound for any node in the tree; Z^{LB_3} gives the optimal solution value for nodes at the bottom of the tree, where $S^0 = \emptyset$.

We now conclude that both the (SP)- and (EH)-heuristic can be implemented as an asymptotically optimal and polynomially bounded heuristic, e.g., if all intervals are chosen as in (2.24).

Corollary 2.1 *Consider the (SP)-heuristic with interval lengths specified by (2.24) and with each interval problem solved by the above branch-and-bound procedure.*

(a) In the general multi-item case, the heuristic has complexity $O(NT^2 \log\log T)$ if each node is evaluated by the value Z^{LB_3} and $O((N + K^*)T^2 \log\log T)$ if evaluated by Z^{LB_1}.

(b) In the multi-item case without speculative motives, the heuristic has complexity $O(NT^2 \log T)$ if each node in the branch-and-bound tree is evaluated by Z^{LB_1}.

(c) In the single item case (N = 1) the heuristic has complexity $O(T^2 \log\log T)$ if each node in the branch-and-bound tree is evaluated by Z^{LB_1} or Z^{LB_3}.

(d) Assume the parameter conditions of theorem 2.2 are satisfied. The heuristic is asymptotically optimal as T increases to infinity; the convergence of the optimality gap

to zero is uniform in N.

Proof. Parts (a)–(c): The (SP)-heuristic requires, to compute its solution for any given interval, at most $2^{\tau-1}$ exact evaluations, one for each leaf of the branch-and-bound tree and $2^{\tau-1}$ lower bound evaluations of the other nodes of the tree. Exact evaluation of a leaf takes O($N\tau \log \tau$) time, as shown when discussing z^{LB}var. Also, $\tau = O(\log T)$ and $J = O(\frac{T}{\log T})$. The complexity bounds in parts (a)–(c) thus follow from those associated with a single evaluation of Z^{LB_1} or Z^{LB_3} in the non-leaf nodes of the branch-and-bound tree, i.e. $O(N\tau \log \tau), O((N + K^*)\tau \log \tau), O(N\tau^2)$ and $O(\tau \log \tau)$ respectively. Part (d) follows from the discussion at the start of Section 2.5. ■

Thus, the (SP)-heuristic can be designed to be asymptotically optimal with a complexity which grows only somewhat faster than quadratically in T, and linearly in the number of items N. The (EH)- heuristic has larger complexity. For example, when implemented with interval increments of size τ and τ given by (2.24), its complexity is $O(NT^3)$ when each interval problem is solved by the above branch-and-bound procedure based on the lower bound Z^{LB_3}. On the other hand, the (EH)-heuristic tends to generate significantly superior solutions, as we shall demonstrate in the next section.

The heuristics can also be designed as polynomial approximation schemes.

Corollary 2.2 *Assume the parameter conditions of Theorem 2.2 are satisfied. For any given* $\epsilon \geq 0$*, choose* $\tau = \min\left\{T, \frac{\rho}{\epsilon\gamma}\right\}$ *and all interval increments* $T_\ell - T_{\ell-1} = \tau$ *(with the possible exception of the last interval increment which is of length* $T - \left\lfloor \frac{T}{\tau} \right\rfloor \tau$*). Assume each interval problem is solved by the above branch and bound procedure, with each node evaluated by* z^{LB_3}*. The (SP)- and (EH)-heuristics result in an* ϵ*-optimal solution with* *a complexity bound which is $O(NT)$ and $O(NT^2 \log T)$, respectively.*

Proof. The optimality gap result is obvious if $\tau = T$. Otherwise, by Theorem 2, for PI = SP and PI = EH: $\frac{z^{PI}-z^*}{z^*} \le \frac{(J-1)\rho}{T\gamma} \le \frac{(\lceil \frac{T}{\tau} \rceil - 1)}{T}\frac{\rho}{\gamma} \le \frac{\frac{T}{\tau}\rho}{\gamma T} = \epsilon$. The complexity counts are immediate from the proof of Corollary 1. ■

2.6 The general (JIS) – model and numerical results

In this section, we consider a generalization in which the fixed setup cost associated with an order depends on the specific items included in that order. More specifically, we assume that, *in addition* to the period-dependent (joint) setup cost K_t, incurred for *any* order in period t, an item-specific setup cost is incurred for any item included in the order. Thus, let

κ_{it} = setup cost incurred when ordering item i in period t; $i = 1, \ldots, N; t = 1, \ldots, T$.

The mixed-integer programming formulation in Section 2.2 is easily adjusted to incorporate these item- specific setup costs. Add a new set of zero-one variables:

$$y_{it} = \begin{cases} 1 & \text{if } x_{it} > 0 \\ 0 & \text{otherwise,} \end{cases}$$

as well as constraints:

$$x_{it} \le C_t y_{it} \quad i = 1, \ldots, N, \quad t = 1, \ldots, T. \tag{2.26}$$

$$y_{it} \le Y_t \quad i = 1, \ldots, N, \quad t = 1, \ldots, T. \tag{2.27}$$

The new objective function becomes:

$$z^* = \min \left\{ \sum_{t=1}^{T} \left[K_t Y_t + \sum_{i=1}^{N} c_{it} x_{it} + h_{it} I_{it} + \kappa_{it} y_{it} \right] \right\}. \tag{2.28}$$

The mechanisms of both the (SP) and the (EH)-heuristics are easily generalized, as well. Note that it is not necessary to solve each interval problem to optimality; to accelerate the procedure one may terminate as soon as a solution is found within a given precision (δ %) of a lower bound. While it is unknown how the bounds for the heuristics' optimality gaps can be extended or how the heuristics can be designed to be asymptotically optimal and polynomially bounded, in practice we find that in particular the (EH)-heuristic generates close-to-optimal solutions in a modest amount of time. To show this, we have conducted a numerical study, coding our heuristics in C++ and running them on a Sun 4000 work station with Solaris 7 and 2 GB of RAM.

In designing our study, we have followed the design of MAES and VAN WASSENHOVE (1986, 1988), one of the most comprehensive comparisons of known heuristics, except that they confined themselves to instances with $N = T = 12$ items and periods, while we have systematically varied the number of items between 10 and 25, and the number of periods from 10 to 50. (MAES and VAN WASSENHOVE restrict themselves to the case where *only* item-specific setup costs prevail, which remain constant across the complete planning horizon). An additional difference is that, at the end of the eighties, no solution method was capable of solving the model to optimality, even for moderate size problems with $N = T = 12$. As a consequence, the quality of the proposed heuristics was gauged by their gap with respect to the best solution found after evaluation of (up to) 1000 nodes in a tailor-made branch-and-bound tree. Today, we can solve these and many larger problems to optimality, enabling us to gauge the *actual* optimality gaps.

Our base set of problems has $N = 10$ items and a horizon of $T = 15$ periods. As in MAES and VAN WASSENHOVE (1988) all demands $\{d_{it}\}$ are independently generated from a Normal distribution with mean 100 and standard deviation of 10. With constant capacity levels C, we consider *three* levels for the 'problem density', defined as the ratio $\frac{\sum_{t=1}^{T} C_t}{\sum_{t=1}^{T} D_t} = \frac{TC}{\sum_{t=1}^{T} D_t}$: *low* density where the ratio equals 2, *medium* density where it equals 4/3 and *high* where it is 10/9. We set all variable cost rates $h_{it} = c_{it} = 1$. For each item $i = 1, \ldots, N$ we determine the fixed (item-specific) setup cost *indirectly* by first choosing the EOQ-cycle time 'Time Between Orders (TBO)' $= \sqrt{\frac{2\kappa}{hd}} = \sqrt{\frac{2\kappa}{100}} = \sqrt{\frac{\kappa}{50}}$ and determine the κ value from this identity. The TBO value is generated from a *uniform* distribution on the interval $[1, 3]$, when considering *low* TBO-values, the interval $[2, 6]$, when considering *medium* TBO-values, and $[5, 10]$ for the case of *high* TBO-values. The *joint* setup cost is calculated in the same way, i.e. from the identity $TBO = \sqrt{\frac{2K}{100N}}$.

We start by evaluating the (EH)-heuristic with respect to its optimality gap and running time, compared to the Complete Horizon Method (CHM) – the solution obtained by the standard CPLEX MIP-solver when applied to the full problem. We consider all 27 combinations which arise when combining the three problem densities, three product TBO values and three period TBO values. For each of these 27 combinations, we have generated 5 distinct problem instances and we report in Table 2.1 the average running times in CPU seconds, when solving the problem with CHM and with the (EH)-heuristic, implemented with $\tau = 5$, $T_\ell = \ell$, $\ell = 1, \ldots, J = T$ and $\delta = 1\%$. We also report the optimality gap of the solution generated by this heuristic. A hyphen indicates that (one or more) problem instances could not be solved to optimality within 6 hours, in which case the reported optimality gap refers to the best found solution by CPLEX so far. (Some of the optimality gaps are negative, implying that the (EH)-heuristic terminates with a *better* solution than CHM after 6 hours of running time!) We note that, all optimality gaps are below 1.75%.

Where comparable, the CPU times appear to be of the same order of magnitude as those in state-of-the-art heuristics such as STADTLER (2003), even though differences between the problem instances and platforms make a precise comparison impossible.

Unless specified otherwise, when CHM is used, we employ the plant location formulation. Confirming prior experience with the (JIS)-model, we have noticed that this formulation usually, though not necessarily, results in faster solutions. (In contrast, we use the network formulation, unless specified otherwise, for progressive interval heuristics, as *it* typically runs faster for these heuristics.) As a further benchmark for the (EH)-heuristic, we have verified whether exact solutions (via CPLEX 7.1) could be significantly sped up if the problem formulation is strengthened by adding the cutting plane constraints (see BARANY ET AL. (1984a,b)):

$$\sum_{t\in S} x_{it} \le \sum_{t\in S}\left(\sum_{u=t}^{l} d_{iu}\right) y_{it} + I_{il}, \quad i = 1,\ldots,N \quad \text{and} \quad l = 1,\ldots T,\ \forall S \subseteq \{1,\ldots,l\} \tag{2.29}$$

to the network formulation (and the same constraints, with x_{it} replaced by $\sum_{w=t}^{T} x_{itw}$, for the plant location formulation). More specifically, we have added the violated constraints in (2.29) after solving the LP-relaxation of the complete problem and before invoking the CPLEX MIP solver. Table 2.2 revisits the nine categories (of five problem instances each) in Table 2.1 in which the item- and period TBO are of the same type, i.e. in which they are both low, medium or high. Each of the last 6 columns reports on one of six solution methods, described below and executed with a 1 hour time limit. The first reported number is the optimality gap with respect to the best among the 6 solutions, with a * denoting a 0% gap; where the CPU time is less than 1 hour, we report this measure within parentheses (in seconds). The 6 methods are: (1) CHM using the network flow formulation by itself; (2) CHM using the network flow formulation with the addition of violated cuts; (3) CHM with the plant location formulation by itself, (4) CHM using the plant location formulation with the addition of the above violated cuts; (5) the (EH)-heuristic where each interval problem is solved with the network flow formulation, and (6) the (EH)-heuristic with the plant location formulation. We conclude that the cuts in (2.29) do not result in major improvements either in terms of CPU time or in terms of the quality of the generated solutions. (Frequently, both attributes deteriorate, in fact).

Table 2.1: (JIS): Gaps and running times of the EH heuristic with $N = 10, T = 15$ (limit 6 hours per instance)

density	low			medium			high		
TBO period	low	medium	high	low	medium	high	low	medium	high
low item TBO	0.00%	0.23%	-1.10%	0.73%	0.80%	0.31%	0.29%	0.79%	0.06%
running time	30/18	75/19	-/126	452/30	56/23	60/29	6154/55	-/126	6370/98
medium item TBO	0.72%	1.73%	1.52%	0.37%	0.32%	0.20%	0.79%	0.04%	0.49%
running time	-/80	2551/54	-/110	252/34	-/126	-/117	-/145	-/150	-/167
high item TBO	-1.33%	0.96%	0.01%	0.17%	0.89%	-0.43%	1.09%	0.01%	0.49%
running time	-/119	-/98	-/52	-/238	-/210	-/232	-/448	-/170	-/170

Table 2.2: (JIS): Gaps and CPU seconds of 4 exact and 2 EH heuristic solutions (limit 1 hour per instance) with $N = 10, T = 15$

density	TBO	Net	Net w/cuts	Plant	Plant w/cuts	EH w/net	EH w/plant
low	low	* (30)	* (244)	* (67)	* (131)	1.85% (16)	2.35% (28)
low	medium	*	0.75%	1.03%	0.25%	1.15% (38)	1.42% (106)
low	high	*	1.35%	0.90%	0.02%	5.32% (53)	2.19% (145)
medium	low	0.34%	0.64%	*	0.13%	1.17% (25)	1.16 (32)
medium	medium	0.82%	1.03%	0.46%	0.48%	0.12% (67)	* (297)
medium	high	1.02%	0.67%	0.59%	0.14%	0.15% (136)	* (548)
high	low	0.11%	0.40%	*	0.34%	0.31% (51)	0.18% (98)
high	medium	0.78%	1.01%	0.62%	0.91%	0.20% (96)	* (288)
high	high	0.91%	1.15%	1.08%	1.16%	0.04% (392)	* (947)

In Table 2.3, we show that the (EH)-heuristic, again implemented with $\tau = 5$ and $T_\ell = \ell,\ \ell = 1, \ldots, J$, can be effectively used for significantly larger problem instances. Varying N from 5 to 25 and T from 10 to 50, we report the CPU running time in seconds. We specify the parameters as above, confining ourselves to the case where the problem density is medium, as is the ‘item TBO’ and ‘period TBO’ value. Three problem instances are generated for every combination of N and T.

Table 2.3: (JIS): Running times for the (EH)-heuristic

periods	10	25	50
5 item	7	42	124
10 item	29	184	524
15 item	416	2694	4310
20 item	1600	9372	16159
25 item	20335	66634	58264

As mentioned in Section 2.4, the (SP)-heuristic is considerably faster than the (EH)-heuristic but it generally generates solutions with significantly larger optimality gaps. Table 2.4 illustrates this for a set of 27 problem instances, all with $N = 10$ and $T = 15$ and parameters as specified in our basic set. Focusing on the medium problem density case, we consider all 9 combinations of ‘product TBO’ and ‘period TBO’-values, generating 3 instances for each. We report on the running times of CHM (terminated when a solution is found within 1% of the best lower bound), the (EH)-heuristic and the (SP)-heuristic. We also report both heuristics’ average optimality gaps. While the optimality gap for the (EH)-heuristic is never in excess of 3% and on average equals 1.2%, that of the (SP)-heuristic may be as high as 33% and is on average 14.7%.

Table 2.4: (JIS): Gaps and CPU seconds for the CHM (within 1% of LB), the (EH)- and (SP)-heuristic with $N = 10, T = 15$

TBO period	low	medium	high
low item TBO	0.9%/3.9%	0.1%/7.0%	0.3%/8.6%
running time	9/7/1	9/8/1	13/6/1
medium item TBO	1.3%/11.3%	0.8%/11.5%	0.5%/11.0%
running time	262/19/1	390/19/1	208/18/1
high item TBO	2.8%/33.8%	2.9%/25.9%	1.2%/19.7%
running time	7854/23/1	5235/24/1	6750/26/1

In Table 2.5, we evaluate the optimality gaps for the (JS)-problem with period-dependent setup costs only. To this end, we consider a set of 45 problems with $N = 10$ and $T = 30$ periods; we again consider all 9 combinations of 'TBO' and problem density values and generate 5 problem instances for each for these combinations. We report the CPU times of the CHM, the (EH)- and the (SP)-heuristic, along with the optimality gaps associated with both heuristics. Once again, the (EH)-heuristic generates solutions within 1% of optimality and does so within approximately 20 seconds of CPU time. The CHM often requires several thousands of CPU seconds (i.e. many hours of CPU time); its solution times depend highly on the parameters of the problem. The (SP)-heuristic is an order of magnitude faster than the (EH)-heuristic but may generate solutions with optimality gaps as large as 15%. Clearly, the (EH)-heuristic can be employed for far larger problem instances.

Table 2.5: (JS): Gaps and CPU seconds for CHM, the (EH)-heuristic and the (SP)-heuristic with $N = 10, T = 30$

density	low	medium	high
low TBO	0.2%/6.0%	0.5%/1.6%	0.2%/0.5%
running time	44/17/1	48/17/1	21/17/1
medium TBO	0.8%/12.1%	0.1%/3.4%	0.1%/5.0%
running time	742/29/3	712/19/1	85/19/1
high TBO	0.5%/15.0%	0.1%/3.9%	0%/6.0%
running time	1150/22/2	3973/21/1	127/19/1

Finally, we consider the case with *item-dependent* set-up costs *only*. Table 2.6 compares the (EH)-heuristic and CHM, for the nine relevant item-TBO and problem density values in Table 2.1 (As in Table 2.2, the CHM is terminated after 1 hour). All of our conclusions regarding the quality of the (EH)-heuristic solutions and the running times continue to apply for this special case of the (JIS)-model.

Table 2.6: (JIS): Gaps and CPU seconds for CHM and the (EH)-heuristic with item-dependent fixed costs only ($K_t = 0, N = 10, T = 15$)

density	low	medium	high
low item TBO	0.5%	0.5%	0.4%
running time	2/7	7/7	-/13
medium item TBO	1.7%	1.6%	-1.1%
running time	-/28	-/44	-/103
high item TBO	1.0%	-0.7%	-2.5%
running time	-/41	-/126	-/358

Returning to the general (JIS)-model, Belvaux and Wolsey (2000, 2001) observe that in many applications, at most one or two items may be ordered per period. The authors refer to such models as ‘small bucket models’. Once again, the mechanics of the (SP)- and (EH)-heuristic are straightforwardly adjusted to accommodate this restriction. For small bucket models even the branch-and-bound methods of Section 2.5 are easily adjusted. Choosing $\tau = \lceil \alpha \log T \rceil$, as in (2.24), this gives rise to a polynomial time implementation of the heuristics for the (JIS)-model, where the complexity bound is a factor $O(N)$ or $O(N^2)$ larger than the corresponding complexity bound for the (JS)-model.

Similarly, the mechanics of the (SP)- and (EH)-heuristic are easily adjusted to (i) add capacity limits for individual items in each period, to (ii) allow for *multiple* capacitated order batches in every period, as in Anily and Tzur’s (2004a,b) MIMV-problem, to (iii) address the hierarchical planning problems in Graves (1982) or Van Roy and Wolsey

(1987) which differ from the (JIS) model with capacity limits for each item only by allowing the (joint) capacity to be increased with overtime at a linear penalty cost, or to (iv) handle any of the other variants mentioned in Belvaux and Wolsey (2000, 2001).

Chapter 3

Probabilistic Analysis of Progressive Interval Heuristics for Multi-Item Capacitated Lot-Sizing Problems

3.1 Introduction

This paper conducts a probabilistic analysis of an important class of heuristics for multi-item capacitated lot sizing problems. More specifically, we address the following classical problem (P): a family of N items is to be procured from the same production facility or outside supplier. The planning horizon consists of T periods (not necessarily of equal length). Demands are specified for each item and each period of the planning horizon. The aggregate order size, in any given period, is bounded by a capacity limit, which may vary over the course of the planning horizon. The costs consist of inventory carrying, variable and fixed order costs. As to the latter, the fixed order cost in any given period only depends on the period index, but not on the composition of the order. The inventory and variable order costs are proportional with the end-of-period inventories and order sizes, at item- and period-dependent cost rates. The objective is to minimize total costs for the planning horizon while satisfying all demands, without backlogging.

Despite a voluminous literature devoted to the general model (P), it continues to present a major challenge to theoreticians and practitioners alike. The problem is NP complete, even in the special case of a single item (N = 1), as shown by FLORIAN ET AL (1980). Until recently, exact and heuristic solution methods have only been successfully applied to instances with a relatively low number of items and/or periods. Chapter 2 of this dissertation investigated the following class of so-called *progressive interval* heuristics. A progressive interval heuristic consists of J iterations, where, iteration by iteration, the problem is solved, to optimality, over a progressively larger time interval $[1, T_\ell]$, i.e. $T_1 \leq T_2 \leq \cdots \leq T_J = T$. When solving a given interval problem, the necessary and sufficient conditions for a feasible extension to the remainder of the planning horizon are appended as boundary conditions. To ensure that the computational complexity in each iteration remains manageable, the heuristic fixes, in iteration ℓ, all integer variables for periods 1 to $T_\ell - \tau$ (for some $\tau > 0$) and all *continuous* variables for periods 1 to some $t_\ell \leq T_{\ell-1}$ at their optimal value after iteration $\ell - 1$. The horizons are chosen such that $0 = t_1 \leq t_2 \leq \cdots \leq t_J$ while $\tau \geq T_\ell - T_{\ell-1}$, the number of periods by which the horizon is expanded in the ℓ-th iteration.

We characterize the asymptotic performance of the progressive interval heuristics as T goes to infinity, assuming the data are realizations of a stochastic process of the following type: the vector of cost parameters follows an arbitrary process with bounded support, while the sequence of aggregate demand and capacity pairs is generated as an independent sequence with a common general bivariate distribution, which may be of *unbounded* support. We show that important subclasses of the class of progressive interval heuristics can be designed to be asymptotically optimal *with probability one*, while running with a complexity bound which grows *linearly* with the number of items N and slightly faster than *quadratically* with T. Our probabilistic analyses complement the worst case analyses in Chapter 2 where asymptotic optimally is shown under conditions which require that all demands and capacities are *uniformly bounded* and that the aggregate capacity over a large enough interval of time exceed the aggregate demand by at least a minimum slack value $\sigma > 0$. *Both* of these assumptions are somewhat restrictive, in the context of an asymptotic analysis where very large planning horizons T are considered.

For many types of complex (NP-complete) logistical planning problems, probabilistic analyses have provided performance guarantees for various classes of heuristics, fostering insights into which algorithmic approaches are effective for large size problems. One such planning area is that of vehicle routing, starting with the seminal papers by KARP (1979) and HAIMOVICH and RINNOOY KAN (1985); see COFFMAN and LUEKER (1996), FEDERGRUEN and SIMCHI-LEVI (1992) and ANILY and BRAMEL (1999) for surveys. (Some of the planning models integrate vehicle routing with inventory planning but, thus far, only in a context of demand processes that occur at *constant* rates.) Other logistical planning areas supported by probabilistic analyses include (hierarchical) facility location and sourcing models (e.g. Chan and Simchi-Levi (1996), Gallego and Simchi-Levi (1997), Fisher and Hochbaum (1980), and Romeijn and Romero Morales (2001)). See Bramel and Simchi-Levi (1997) for a general overview. Rhee and Talagrand (1987, 1989) and Rhee (1993) have shown how probabilistic analyses of a variety of logistical planning problems can be based on specific large deviation results. Our analyses, as well, are in part, based on such large deviation techniques. To our knowledge, the probabilistic analyses in this paper represent the first such analyses for inventory planning models with *time-varying* parameters (, otherwise referred to as dynamic lot sizing problems).

We conclude this section with a brief review of the relevant literature beyond the papers mentioned above. The (NP-) complexity of the problem arises from the superposition of (joint) setup costs and capacity limits. Indeed, the problem is solvable in $O(NT \log T)$ time, if either the capacity constraints are relaxed or in the absence of fixed order costs. In the former case, the problem decomposes into N independent single item lotsizing problems for which one of the $O(T \log T)$ methods by Aggarwal and Park (1992), Federgruen and Tzur (1991) or Wagelmans et al. (1992) can be used. In the latter case, the problem is solvable in $O(NT \log T)$ time with Ahuja and Hochbaum (2004)'s recent method.

There is a voluminous literature describing various heuristics for the general multi-item model. We refer to Salomon (1990) and Kuik et al. (1994) for surveys of the literature until 1994. State-of-the art solution methods include Belvaux and Wolsey (2000, 2001), Stadtler (2003) and Suerie and Stadtler (2003). These methods are all based on variants of progressive interval heuristics. See Chapter 2 for details and a more detailed literature review. Other than the above mentioned worst case analyses in the latter paper, the only performance guarantees for heuristics for capacitated lot-sizing problems with general time-dependent capacity limits are due to GAVISH and JOHNSON (1990) and VAN HOESEL and WAGELMANS (2001). The latter developed a fully polynomial approximation scheme for the general *single-item* model, after the former proposed such a scheme for a more restricted version of the problem.

The remainder of this chapter is organized as follows: In Section 3.2, we specify the model, the probability model generating its data and the class of progressive interval heuristics. In Section 3.3, we establish *almost sure* asymptotic optimality for heuristics in this class as well as their polynomial complexity bound. Section 3.4 concludes the chapter with a discussion of the case where the items' shelf life is uniformly bounded e.g. because items are perishable.

3.2 The model and the class of progressive interval heuristics

The model employs the following data, where the index $i \in \{1, \dots, N\}$ is used to distinguish between items and time periods are indexed by t. (Demands are represented as multiples of the volume that consumes *one* unit of capacity):

c_{it} = variable per unit order cost for item i in period t

h_{it} = cost of carrying a unit of inventory of item i at the end of period t

K_t = setup cost incurred when an order is placed in period t

d_{it} = demand for item i in period t; ($d_{it} \geq 0$)

D_t = aggregate demand in period $t = \sum_{i=1}^{N} d_{it}$

C_t = order capacity, i.e. the maximum number of units which can be ordered in period t.

We use the following set of decisions variables:

x_{it} = order size for item i in period t; $i = 1, \ldots, N$; $t = 1, \ldots, T$

$$Y_t = \begin{cases} 1 & \text{if } \sum_{i=1}^{N} x_{it} > 0 \\ 0 & \text{otherwise} \end{cases} \qquad t = 1, \ldots, T$$

I_{it} = ending inventory of item i in period t; $i = 1, \ldots, N$; $t = 1, \ldots, T$

Let I_t^0 = the *minimum* aggregate inventory at the end of period t, such that a feasible production / inventory plan exists for periods $t+1, \ldots, T$. These minimum stock levels are easily computed from the following recursion, which can be verified by induction:

$$I_t^0 = \left(D_{t+1} - C_{t+1} + I_{t+1}^0\right)^+, \qquad t = 1, 2, \ldots, T-1, \quad \text{with} \quad I_T^0 = 0 \tag{3.1}$$

This is the well known Lindley equation, see e.g. ASMUSSEN (1987). The following is a standard formulation:

$$
\begin{aligned}
(\mathrm{P}) \qquad z^* &= \min\left\{\sum_{t=1}^{T}\left[K_tY_t + \sum_{i=1}^{N}(c_{it}x_{it} + h_{it}I_{it})\right]\right\} && (3.2)\\
& \text{s.t.} \\
I_{it} &= I_{i(t+1)} + x_{it} - d_{it}, \qquad i = 1,\ldots,N, \quad t = 1,\ldots,T && (3.3)\\
\sum_{i=1}^{N} x_{it} &\leq C_tY_t \qquad i = 1,\ldots,N, \quad t = 1,\ldots,T && (3.4)\\
\sum_{i=1}^{N} I_{it} &\geq I_t^0 \qquad t = 1,\ldots,T && (3.5)\\
x_{it} &\geq 0; \quad I_{it} \geq 0; \quad Y_t \in \{0,1\} && (3.6)
\end{aligned}
$$

We assume that the model data are generated by the following probabilistic model: the $(2N+1)T$ cost parameters $\{K_t, c_{it}, h_{it}\}$ are generated by an *arbitrary* stochastic process with support on a hypercube in the positive orthant of $\mathbb{R}^{(2N+1)T}$. As to the sequence of aggregate demand and capacity pairs $\{(D_t, C_t) : t = 1,\ldots,T\}$,we assume:

(A) $\{(D_t, C_t) : t = 1,\ldots,T\}$ is a sequence of independent pairs of random variables, all distributed like (D, C) with a general bivariate distribution, possibly with *unbounded* support, such that the marginal distribution of D has a moment generating function, i.e. $E(e^{\theta D})$ exists for some $\theta > 0$, $\delta = E(D) > 0$ and the support of the distribution of C is bounded from below by a constant C_*. Moreover, $\mu = E(C) - E(D) > 0$.

The requirement that the demand distribution has a moment generating function which is finite in the neighborhood of the origin covers most of the distributions commonly used in (stochastic) inventory models. (e.g. the Normal, Gamma, Negative Binomial or Weibull distributions). The condition merely precludes heavy-tailed demand distributions which implies heavy-tailed distributions for the steady-state distribution of the reserve-stock variables $\left\{I_t^0\right\}$. The condition $\mu = E(C) - E(D) > 0$ is necessary to ensure that the generated problem instances be *feasible* as T becomes large. Let $\psi(\theta)$ denote the *cumulative* generating function (cgf) of the random variable (D-C) which is the logarithm of its moment generating function:

$$\psi(\theta) = \log E\left[e^{\theta(D-C)}\right].$$

Since D has a finite moment generating function on some interval $[0, \bar{\theta}]$, so does $(D-C)$, so that $\psi(\theta) < \infty$ on $[0, \bar{\theta}]$. Moreover, $\psi(\cdot)$ is differentiable with $\psi(0) = 0$ and $\psi'(0) = -\mu < 0$, by (A), so that $\psi(\theta) < 0$ for all $\theta > 0$, sufficiently small.

When the items have a limited shelf life, we show in Section 3.4 that our results continue to apply under generalizations of condition (A), allowing for various forms of intertemporal demand and capacity dependencies.

In a progressive interval heuristic, employing J iterations, the ℓ-th iteration consists of solving (P), with T replaced by T_ℓ and all (integer) Y-variables for periods $1, \ldots, T_\ell - \tau$ and all (continuous) x- and I-variables for periods $1, \ldots, t_\ell \leq T_{\ell-1}$ fixed at their optimal value in the $\ell - 1$st iteration. (In the first iteration, no restrictions apply to any of the variables.) Thus, the number of unrestricted integer variables in each (except for possibly the first) iteration is kept constant at τ. Since the complexity of any mixed integer program is primarily determined by the number of (unrestricted) integer variables, the computational complexity remains manageable when choosing τ sufficiently small, and from each iteration to the next it grows only moderately.

As in Chapter 2, we pay special attention to two *extreme* subclasses: (i) the *Strict*

Partitioning heuristics (SP), with all (except for possibly the last) interval increment $T_\ell - T_{\ell-1} = \tau$ and $t_\ell = T_{\ell-1} = T_\ell - \tau$; (ii) the *Expanding Horizon* heuristics (EH) with all $t_\ell = 0$. The (SP)-heuristics are related to the Time Partitioning heuristics, see FEDERGRUEN and TZUR (1999).

Thus, the (SP)-heuristics minimize the computational complexity of each interval problem at the expense of providing minimal flexibility to the continuous variables. The (EH)-heuristics, while of larger computational complexity, provide *maximal* flexibility for the continuous variables and even for the integer variables, in case interval increments $T_\ell - T_{\ell-1} < \tau$ are chosen. Under such choices, even many of the setup decisions made in one iteration, may be revisited in subsequent iterations, on the basis of additional demand, cost and capacity information pertaining to additional periods. The numerical study in Chapter 2 indicates that (EH)-heuristics can be used effectively to solve moderate to large size problem instances and that the solutions generated come very close to being optimal. Those generated by (SP)-heuristics typically exhibit larger optimality gaps.

3.3 Almost sure asymptotic optimality

In this section, we show that both (SP)- and (EH)-heuristics can be designed to be simultaneously *almost surely asymptotically optimal* as well as of *low polynomial complexity*. As with all (SP)-heuristics, we confine ourselves to (EH)-heuristics in which (with the possible exception of the last iteration) exactly τ periods are appended to the tail of the planning horizon, as we progress from one iteration to the next $(T_\ell - T_{\ell-1} = \tau)$.

We show that both heuristics, with cost values z^{SP} and $z^{(EH)}$, respectively, are *almost surely (a.s.) asymptotically optimal* if the interval increment τ is adjusted as a function of T, where

$$\tau = \Omega(\log T), \quad \text{i.e.} \quad \lim_{T\to\infty} \frac{\tau}{\log T} = \infty, \text{e.g.} \tag{3.7}$$
$$\tau = = \eta \lceil \log T \rceil^{\zeta}, \quad \text{for some} \quad \eta > 0 \quad \text{and} \quad \zeta > 1$$

To derive a specific complexity bound, we assume that each interval problem in each iteration is solved with a tailored branch-and-bound procedure, i.e. the b&b-procedure in Section 2.5 in which each non-leaf node of the tree is evaluated with lower bound LB_3, ibid.

Theorem 3.1 *Consider a (SP)-heuristic with $\tau = \eta \lceil \log T \rceil^{\zeta}$ for some $\eta > 0$ and $\zeta > 1$. The heuristic is asymptotically optimal , a.s. and it can be designed to run in O($NT^{\zeta+1} \log\log T$) time as well.*

Proof. Let $c_i^* > 0$ denote the essential infimum of the stochastic process $\{c_{it} : t = 1, \ldots, T\}$. Observe first that by the law of large numbers, with probability one, $\liminf_{T\to\infty} \frac{z^*}{T} \geq \sum_{i=1}^{N} c_{i*} \liminf_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} d_{it} = \sum_{i=1}^{N} c_{i*}\delta_i > 0$, with $\delta_i = E(d_{it})$. (Note that $\delta_i > 0$ for at least one $i = 1, \ldots, N$ since $\sum_{i=1}^{N} \delta_i = \delta > 0$.) In other words, with probability one, the numerator in the optimality gap $\frac{z^{SP} - z^*}{z^*}$ grows at least linearly in T. It thus suffices to show,

$$\lim_{T\to\infty} \frac{1}{T} \left[z^{SP} - z^* \right] = 0, \quad \text{a.s.} \tag{3.8}$$

As in the worst-case analysis of Theorem 2.2 in Chapter 2, we transform an optimal solution for the complete problem in *two* phases into a solution which is achievable by the (SP)-heuristic. In Phase I, the optimal solution is transformed into one with all intervals' ending *aggregate* inventories equal to their I^0-values. (Note that that the solution generated by the (SP)-heuristic satisfies this property as well.) Let z^I denote its cost value. In Phase II, the composition of the reserve stock at the end of each of the intervals is made identical to that of the solution of the (SP)-heuristic, resulting in a solution with the cost value z^{II}. This solution is one which is among the ones considered by the (SP)-heuristic, i.e. $z^{II} \geq z^{SP}$. Thus,

$$\frac{z^{SP} - z^*}{T} \leq \frac{z^{II} - z^*}{T} = \frac{z^{II} - z^I}{T} + \frac{z^I - z^*}{T} \tag{3.9}$$

Following the proof of Theorem 2.2 in Chapter 2 and given the (general) assumption about the stochastic process which generates the cost parameters, ones verifies that an integer $\Lambda > 1$ and constants B_1 and B_2 exist such that $z^I - z^* \leq (J-1)B_1 + B_2 \sum_{\ell=1}^{J-1} \sum_{r=T_\ell-\Lambda+1}^{T_\ell} C_r$. If $\Lambda > \tau$, the partial sums $\left\{\sum_{r=T-\Lambda+1}^{T_\ell} C_r : \ell = 1, \ldots, J-1\right\}$ may overlap. However, $z^I - z^* \leq (J-1)B_1 + B_2 \left\lceil \frac{\Lambda}{\tau} \right\rceil \sum_{r=T-(J-1)\Lambda+1}^{T} C_r$ is a valid upper bound. Thus, $\frac{z^I - z^*}{T} \leq \frac{(J-1)}{T} \left[B_1 + B_2 \lceil \frac{\Lambda}{\tau} \rceil \Lambda \frac{1}{(J-1)\Lambda} \sum_{r=T-(J-1)\Lambda+1}^{T} C_r\right]$ and $\lim_{T\to\infty} \frac{z^I - z^*}{T} = 0$ a.s., since $\lim_{T\to\infty} \frac{J-1}{T} = \lim_{T\to\infty} \tau^{-1} = 0$ and since, with probability one, $\lim_{T\to\infty} \frac{1}{(J-1)\Lambda} \sum_{r=T-(J-1)\Lambda+1}^{T} C_r = E(C_1) < \mu$ by the law of large numbers and the fact that the sequence $\{C_t : t = 1, 2, \ldots\}$ is an i.i.d. sequence of random variables.

To bound the additional cost incurred because of the Phase II transformation, let $\Delta c^* = \max_t \max_{i\neq\ell}[c_{it} - c_{\ell t}]$, $\Delta h^* = \max_t \max_{i\neq\ell}[h_{it} - h_{\ell t}]$, $h_* = \inf_{i,t} h_{it}$ and $K^* =$

$\max_t K_t$ and note that $\Delta c^* = O(1)$, $\Delta h^* = O(1)$ and $K^* = O(1)$ as $T \to \infty$ while $h_* > 0$.

The solution obtained after Phase I and the solution generated by the (SP)-heuristic, both have $I_{T_\ell} = I^0_{T_\ell}$ for all $\ell = 1, \ldots, J$. In Phase II, we obtain the desired composition of the ending inventory at the end of the J intervals by changing (only) the item identity of at most (all of the) $I^0_{T_\ell}$ units in the ending inventory of the $\ell - th$ interval, without any additional changes in the order- and inventory plans. The transformed solution remains feasible, incurs no additional fixed order costs and adds at most $\sum_{\ell=1}^{J-1} I^0_{T_\ell}(\Delta c^* + L_\ell \Delta h^*)$ in variable costs, where L_ℓ denotes the shelf life of the oldest unit in the reserve stock at the end of period T_ℓ. Thus, to prove (3.8) it suffices to show that

$$\lim_{T\to\infty} \frac{1}{T} \sum_{l=1}^{J-1} I^0_{T_\ell} \left(\Delta c^* + L_\ell \Delta h^*\right) = 0, \quad \text{a.s.} \tag{3.10}$$

Recall that $\left\{I^0_t\right\}$ in (3.1), when traversed *backwards*, is a Lindley process. Since the pairs $\{(D_t, C_t)\}$ are i.i.d, and since $\mu > 0$, $\left\{I^0_t\right\}$ has a limiting distribution I^0 (i.e. $\lim_{t\to\infty} I^0(T) \stackrel{w}{=} I^0$, where the convergence is in distribution.) with $E(I^0) < \infty$, see Asmussen (1987, §8.1). Moreover, in view of the remaining assumption in (A), the distribution of I^0 has an exponential tail, i.e. there exist constants α and $\beta > 0$ such that $Pr[I^0 > x] \sim \alpha e^{-\beta x}$, $x \to \infty$ (i.e. $\lim_{x\to\infty} \frac{Pr[I^0>x]}{\alpha e^{-\beta x}} = 1$) (see Assmussen (1987, §12.5). Thus, for some $x^0 > 0, Pr[I^0 > x] \le 2\alpha e^{-\beta x}$, for all $x > x^0$. Finally, let $\bar{I}(T) = \max_{t=1,\ldots,T} I^0_t$ denote the *largest* minimum reserve stock required over the entire planning horizon. Since I^0 has the same distribution as $[D - C + I^0]$, and since $0 = I^0_T \le_{st} I^0$, one easily verifies by complete induction that $I^0_t \le_{st} I^0$ for *all* $t = 1\ldots, T$.

Let $\tilde{n}(T) = \sqrt{\tau \log T}$ denote the geometric mean of τ and $\log T$ and note from (3.7) that

$$\lim_{T\to\infty} \frac{\log T}{\tilde{n}(T)} = 0 \quad \text{and} \quad \lim_{T\to\infty} \frac{\tilde{n}(T)}{\tau} = 0. \tag{3.11}$$

We first show that

$$\lim_{T\to\infty} Pr\left[L_\ell \leq \tilde{n}(T) \quad \text{for all} \quad \ell = 1,\ldots,J-1\right] = 1 \tag{3.12}$$

i.e. asymptotically the maximum shelf-life of any unit in the reserve stock at the end of any of the intervals (in the Phase I solution) is almost surely bounded by $\tilde{n}(T)$. Under (3.12) we have almost surely that (3.8) holds since

$$\begin{aligned} 0 &\leq \lim_{T\to\infty} \frac{1}{T} \sum_{\ell=1}^{J-1} I^0_{T_\ell}(\Delta c^* + L_\ell \Delta h^*) \\ &\leq \lim_{T\to\infty} \left\{ \frac{(\Delta c^* + \tilde{n}(T)\Delta h^*)}{T}(J-1) \right\} \lim_{T\to\infty} \frac{1}{J-1} \sum_{l=0}^{J-1} I^0_{T_\ell} \\ &= h^* \left(\lim_{T\to\infty} \frac{\tilde{n}(T)}{\tau} \right) E(I^0) = 0, \quad \text{a.s.} \end{aligned}$$

where the first equality follows from the the fact that the $\left\{I^0_t\right\}$-process is ergodic, so that a long-run average, sampled at equidistant epochs, converges with probability one to the expected value of the limiting distribution, while the second equality follows from (3.11).

It remains to prove (3.12). Note that $Pr[L_1 > \tilde{n}(T) \quad \text{or} \quad L_2 > \tilde{n}(T) \quad \text{or} \quad \ldots L_{J-1} > \tilde{n}(T)] \leq \sum_{l=1}^{J-1} Pr[L_\ell > \tilde{n}(T)]$. Choose $0 < \theta < \beta$ such that $\psi(\theta) < 0$. To bound each of the terms in the sum, consider first the *conditional* probability $Pr[L_\ell > \tilde{n}(T)]|I_{T_\ell} = i_\ell^0]$, which is bounded by $Pr\left[\sum_{r=T_\ell-n(T)+1}^{T_\ell}(C_r - D_r) \leq i_\ell^0 | I_{T_\ell} = i_\ell^0\right]$ with $n(T) = \tilde{n}(T) - \frac{\Delta c^* + K^*}{h_*}$.

(If a unit, in stock at the end of period T_ℓ, has a shelf life larger than $\tilde{n}(T)$, this implies that a *full capacity* order is placed is in *each* of the periods in the interval $[T_\ell - n(T) + 1, \dots ..., T_\ell]$, for otherwise the procurement of this unit could be postponed till some period in this interval with slack capacity, saving at least $\tilde{n}(T) - n(T)$ periods' carrying costs, i.e. at least $(\Delta c^* + K^*)$, more than offsetting any additional order costs. However, given the condition $I_{t_\ell} = i_\ell^0$, this situation can only happen if $\sum_{r=T_\ell-n(T)+1}^{T_\ell}(C_r - D_r) \leq i_\ell^0$.) Thus,

$$
\begin{aligned}
Pr\left[L_\ell > \tilde{n}(T) | i_\ell^0\right] &\leq Pr\left[\sum_{r=T_\ell-n(T)+1}^{T_\ell}(C_r - D_r) \leq i_\ell^0 | I_{T_\ell} = i_\ell^0\right] \\
&= Pr\left[\sum_{r=T_\ell-n(T)+1}^{T_\ell}(D_r - C_r) \geq -i_\ell^0 | I_{T_\ell} = i_\ell^0\right] \\
&= Pr\left[\sum_{l=1}^{n(T)}(D_r - C_r) \geq -i_\ell^0\right] \qquad (3.13) \\
&\leq \exp\left\{-n(T)\left(\frac{-\theta i_\ell^0}{n(T)} - \psi(\theta)\right)\right\}
\end{aligned}
$$

where the second equality follows from the fact that $I_{T_\ell}^0$ only depends on the demand and capacity values in periods $T_\ell + 1, \dots, T$, see (1), so that the conditional distributions of $\{(D_r - C_r | I_{T_\ell}^0) : r = T_\ell - n(T) + 1, \dots, T_\ell\}$ coincide with the *unconditional* distributions $\{D_r - C_r : r = T_\ell - n(T) + 1, \dots, T_\ell\}$ and hence those of $\{D_1, \dots, D_{n(T)}\}$ by the i.i.d. assumption of $\{(D_t, C_t)\}_{t=1}^{\infty}$. The last inequality in (3.13) follows from Chernoff's inequality. We thus obtain the following bound on the *unconditional* probability

$$\begin{aligned} Pr[L_\ell > \tilde{n}(T)] &\leq E_{I^0_{T_\ell}} \left\{ \exp(-n(T)\psi(\theta)) \exp(\theta \bar{I}_{T_\ell}) \right\} \\ &\leq \exp(-n(T)\psi(\theta)) E_{\bar{I}_r} \exp(\theta \bar{I}_T) \end{aligned} \tag{3.14}$$

Note that

$$\begin{aligned} & E \exp\{\theta \bar{I}(T)\} \\ = \quad & -\int_0^\infty e^{\theta x} d[1 - Pr(\bar{I}(T) \leq x)] = 1 + \int_0^\infty \theta e^{\theta x} Pr[\bar{I}(T) > x] dx \\ = \quad & 1 + \int_0^\infty \theta e^{\theta x} Pr[I_1^0 > x \quad \text{or} \quad I_2^0 > x \quad \text{or} \quad \ldots I_T^0 > x] dx \\ \leq \quad & 1 + \sum_{t=1}^T \int_0^\infty \theta e^{\theta x} Pr[I_t^0 > x] dx \\ \leq \quad & 1 + \sum_{t=1}^T \int_0^\infty \theta e^{\theta x} Pr[I^0 > x] dx = 1 + T \int_0^\infty \theta e^{\theta x} Pr[I^0 > x] dx \qquad (3.15) \\ \leq \quad & 1 + T \left[\int_0^{x^0} \theta e^{\theta x} Pr[I^0 > x] dx + \int_{x^0}^\infty \theta e^{\theta x} 2\alpha e^{-\beta x} dx \right] \\ \leq \quad & 1 + T \left[\int_0^{x^0} \theta e^{\theta x} dx + \frac{2\alpha\theta}{(\beta - \theta)} e^{e^{(\theta-\beta)x^0}} \right] = 1 + Tb \end{aligned}$$

with $b = e^{\theta x^0} - 1 + \frac{2\alpha\theta}{(\beta-\theta)} e^{(\theta-\beta)x^0}$, where the second inequality follows from $I_t^0 \le_{st} I^0$, for all t. Thus, with $a = -\psi(\theta) > 0$:

$$
\begin{aligned}
0 &\le \lim_{T\to\infty} Pr[L_1 > \tilde{n}(T) \quad \text{or} \quad L_2 > \tilde{n}(T) \quad \text{or} \quad \ldots L_{I-1} > \tilde{n}(T)] \\
&\le \lim_{T\to\infty} \sum_{l=1}^{J-1} Pr\left[L_\ell > \tilde{n}(T)\right] \\
&\le \lim_{T\to\infty} \{(J-1)\exp\{-an(T)\}\,(1+Tb)\} \\
&= b \lim_{T\to\infty} \left\{ \frac{T^2}{\tau} \exp(-\log T \frac{an(T)}{\log T}) \right\} \\
&= b \lim_{T\to\infty} \frac{T^2}{\tau} \frac{1}{T^{\frac{an(T)}{\log T}}} = 0
\end{aligned}
$$

where the last inequality follows from (3.14) and (15) and the last equality from (3.11). This proves (3.12), hence (3.10) and (3.9).

It remains to he shown that when $\tau = \eta\lceil(\log T)^\zeta\rceil$, with $\zeta > 1$, the progressive interval heuristic runs in O($NT^{1+\zeta}\log T$) time, when each interval problem is solved with the above described b&b method. The discussion in Section 2.5 shows that evaluation of any node of a b&b tree requires O($N\tau\log\tau$) time. Since this needs to be done at most 2^τ times to evaluate the complete tree, and since $\tau = O(\frac{T}{\zeta})$ interval problems need to be solved, the complexity bound follows immediately. ∎

The same simultaneous (almost sure) asymptotic optimality and polynomial complexity can be obtained for the above (EH)-heuristic, under the same choice for the interval increment τ as in (3.11). The complexity of this (EH)-heuristic is O($T\log\log T/(\log T)^{\zeta-1}$) larger than that of the (SP)-heuristic. Nevertheless, complexity grows only *linearly* with N and (only) slightly faster than cubically with T:

Theorem 3.2 *Consider an (EH)-heuristic with* $\tau = \eta \left\lceil (\log T)^{\zeta} \right\rceil$ *for some* $\eta > 0$ *and* $\zeta > 1$*. The heuristic is asymptotically optimal a.s. and can be designed to run in* $O(NT^{2+\zeta}/(\log T)^{\zeta-1})$ *time.*

Proof: The proof is analogous to that of Theorem 3.1, with only the following modifications: The Phase II transformation should modify the composition of the reserve stock at the end of periods $T_1, T_2, \ldots, T_{J=1}$ to that prevailing at the end of the ℓ-th iteration of the (EH)-heuristic. (In the case of the (EH)-heuristic, this composition may change in subsequent iterations). As shown in Theorem 2.2(b) in Chapter 2, a *third* Phase transformation is necessary to obtain a solution which is is among the ones considered by the (EH)-heuristic, but this third transformation only *reduces* the cost value. The derivation of the complexity bound is again analogous, except that the evaluation of a single node in *one* of the b&b trees now requires $O(NT \log T)$ time. ∎

3.4 Products with limited shelf life

Thus far, we have assumed that items can be kept in stock for an unlimited amount of time. In this Section, we address the situation where the shelf life of each item is bounded by an (integer) constant λ, perhaps because the items are perishable. We refer to the survey paper by Nahmias (1982) for a review of inventory models with limited shelf lives. Within the context of dynamic lot sizing models, the complication of a fixed shelf life has not been addressed until Hsu (2000) who showed that the single item uncapacitated model can be solved in $O(T^2)$ time. (For this case, Hsu addresses, in addition, more general life time models and more general order and inventory cost functions then those used in (P).)

We show that, in the presence of a limited shelf life, almost sure asymptotic optimality of (SP)- and (EH)-heuristics can be established under conditions even more general than (A), for example:

(A^f) The sequence $\{(D_t, C_t)\}$ is strongly ergodic, i.e. for any Lipschitz continuous function $g : \mathbb{R} \to \mathbb{R}$, there exists a constant G such that

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} g((D_t, C_t); (D_{t+1}, C_{t+1}); (D_{t+\Lambda}, C_{t+\Lambda})) = G \qquad \text{a.s.} \tag{3.16}$$

Moreover, $0 < \mu \stackrel{def}{=} \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} C_t - \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} D_t$ (a.s.).

The condition is related to that of asymptotic mean stationarity, see e.g. Gray (1990). Beyond the case of i.i.d. aggregate capacity and demand pairs considered under (A), (A^f) encompasses a large variety of processes, for example:

(I) $\{(D_t, C_t)\}$ is stationary and ergodic

(II) $\{(D_t, C_t)\}$ is a so-called 'world driven' process. Here, the distribution of (D_t, C_t) is time invariant but it depends on the state of the world W_t, with $\{W_t\}$ a Markov process with a finite or countable state space which is ergodic (i.e., the Markov chain has a single positive recurrent set of states). Thus the conditional distributions $\{(D_t, C_t)|W_t = w\}$ are time-invariant. Moreover, $\lim_{t\to\infty} W_t \stackrel{w}{=} W$ and $\lim_{t\to\infty}(D_t, C_t) \stackrel{w}{=} ((D, C)|W)$. See Zipkin (2000) for a detailed discussion of the use of world driven demand processes in inventory models.

(III) A third type of process satisfying (A^f) and modeling different types of intertemporal correlations, is where the process $\{(D_t, C_t)\}$ is autoregressive, e.g. a stable ARMA(p,q) process, i.e.

$$D_t = \sum_{i=1}^{p} \varphi_i D_{t-i} + \sum_{j=1}^{q} \psi_j \epsilon_{t-j} + \epsilon_t \quad \forall t \tag{3.17}$$

$$C_t = \sum_{i=1}^{p} \hat{\varphi}_i C_{t-i} + \sum_{j=1}^{q} \hat{\psi}_j \hat{\epsilon}_{t-j} + \hat{\epsilon}_t \quad \forall t \tag{3.18}$$

where $\{\epsilon_t\}_{t=-\infty}^{+\infty}$ and $\{\hat{\epsilon}_t\}_{t=-\infty}^{+\infty}$ are independent sequences of i.i.d. random variables with finite second moments. A sufficient condition for the processes to be stable is that the characteristic polynomials $\Phi(z) = \sum_{i=1}^{p} \varphi_i z^i$ $[\hat{\Phi}(z) = \sum_{i=1}^{p} \hat{\varphi}_i z^i]$ and $\Psi(z) = \sum_{i=1}^{q} \psi_i z^i$ $[\hat{\Psi}(z) = \sum_{i=1}^{q} \hat{\psi}_i z^i]$ do not have common (complex) roots and that the roots of the former are outside the unit circle.

Lemma 3.1 *Assume the process $\{(D_t, C_t)\}$ is of type (I)–(III). Then $\{(D_t, C_t)\}$ is strongly ergodic.*

Proof: (I) Immediate, see e.g. Proposition 6.31 in Breiman (1992).

(II) The process $\{(W_t, W_{t+1}, \ldots, W_{t+\lambda})\}$ is a Markov process, whose Markov chain has a single positive recurrent set of states, i.e. there exists a state of the process with a finite expected recurrence time. Almost sure convergence of the limit to the left of (3.16) then follows from the renewal reward theorem.

(III) It suffices to prove strong ergodity of $\{D_t\}$ and $\{C_t\}$ separately. We prove the

former; the proof of the latter is identical. Since the ARMA process is stable, there exists a constant $0 < a < 1$ such that $D_t = \sum_{j=0}^{t} \alpha_j \epsilon_{t-j}$, with $|\alpha_j| < a^j$, see e.g. SAMORODNITSKY and TAQQU (1994). The sequence $\{D_t\}$ is non-stationary. Let $D_t^0 \overset{def}{=} \sum_{j=0}^{\infty} \alpha_j \epsilon_{t-j}$. $\{D_t^0\}$ is clearly stationary and it is well known to be ergodic. Fix a function $g : \mathbb{R}^{\lambda} \to \mathbb{R}$ that is Lipschitz continuous. By the argument for (I), there exists a constant G such that $\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} g(D_t^0, D_{t+1}^0, \ldots, D_{t+\lambda}^0) = G$ a.s..
To show that
$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} g(D_t, D_{t+1}, \ldots, D_{t+\Lambda}) = G$ a.s., as well, it suffices to show that for any $\delta > 0$,

$$|\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} [g(D_t, D_{t+1}, \ldots, D_{t+\lambda}) - g(D_t^0, \ldots, D_{t+\lambda}^0)]| < \delta \quad \text{a.s.} \tag{3.19}$$

Since for any integer $n \geq 1$,

$$\begin{aligned} & |\lim_{T\to\infty} \frac{1}{T} \sum_{t=n}^{T} [g(D_t, D_{t+1}, \ldots, D_{t+\Lambda}) - g(D_t^0, D_{t+1}^0, \ldots, D_{t+\Lambda}^0)]| \\ \leq \; & \lim_{T\to\infty} \frac{1}{T} \sum_{t=n}^{T} |g(D_t, D_{t+1}, \ldots, D_{t+\Lambda}) - g(D_t^0, D_{t+1}^0, \ldots, D_{t+\Lambda}^0)| \end{aligned}$$

it follows from the Lipschitz continuity of $g(\cdot)$ that it suffices to show, for any $\delta > 0$, that an integer $n \geq 1$ exists such that

$\lim_{T\to\infty} \frac{1}{T} \sum_{t=n}^{T} |D_t - D_t^0| \leq \lim_{T\to\infty} \frac{1}{T} \sum_{t=n}^{T} \sum_{j=t+1}^{\infty} |\alpha_j||\epsilon_{t-j}|$
$\leq \lim_{T\to\infty} \frac{1}{T} \sum_{t=n}^{T} \sum_{j=t+1}^{\infty} a^j |\epsilon_{t-j}| = \lim_{T\to\infty} \frac{1}{T} \sum_{t=n}^{T} a^{t+1} \sum_{j=0}^{\infty} a^j |\epsilon_{-j-1}|$
$\leq \lim_{T\to\infty} \sum_{t=n}^{T} a^{t+1} \left\{ \lim_{r\uparrow 1} \sum_{j=0}^{\infty} r^j |\epsilon_{-j-1}| \right\}$
$= \lim_{T\to\infty} \frac{1}{T} \sum_{t=n}^{T} a^{t+1} \left\{ \lim_{M\to\infty} \frac{1}{M+1} \sum_{j=0}^{M} |\epsilon_{-j-1}| \right\} \leq a^{n+1} E|\epsilon| < \delta \quad$ a.s.

where the equality follows from the Abel-Tauberian theorem and the next to last inequality from the law of large numbers. (Since the random variable ϵ has a finite second moment, E $|\epsilon| < \infty$.) The last inequality is satisfied for all $n \geq \lfloor \log(\rho / E|\epsilon|) / \log a \rfloor$. ■

To pursue the algorithm's performance analysis, note that I_t^0, the minimum reserve stock at the end of period t is now given by:

$$I_t^{0,f} = \max_{t+1 \leq s \leq t+\lambda} \sum_{r=t+1}^{s} (D_r - C_r) \tag{3.20}$$

instead of (3.1). (Using repeated substitutions in (3.1), note that $I_t^{0,f} = I_t^0$ when $\lambda = \infty$) In other words, assuming that a feasible solution exists, to ensure that a solution for the first t periods $[1, \dots, t]$ can be extended into a feasible solution over the complete horizon $[1, \dots, T]$, it is *necessary* and *sufficient* that $I_t \geq I_t^{0,f}$. (If $I_t < I_t^{0,f}$, aggregate demand in the periods $t+1, \dots, s$ (for some $t+1 \leq s \leq t+\lambda$) exceeds $(\sum_{r=t+1}^{s} C_r + I_t)$, so demand in $[t+1, s]$ can not be satisfied even when placing a full capacity order in each of the periods of this interval. At the same time, if $I_t \geq I_t^{0,f}$, the first period whose demand can not be met, has a period index greater than $t + \lambda$ and any additional inventory at the end of period t is of no use to meet this demand.)

Theorem 3.3 *Assume items have a fixed shelf life time $\lambda > 0$ and (A^f) holds.*

(a) Consider an (SP)-heuristic with $\tau = \eta \lceil \log T \rceil$ for some $\eta > 0$. The heuristic is asymptotically optimal, a.s., and it can be designed to run in $O(N^2 T^2 \log T (\log N +$

$\log\log T)^2)$ *time as well.*

(b) Consider an (EH)-heuristic with $\tau = \eta \lceil \log T \rceil$ *for some* $\eta > 0$*. The heuristics is asymptotically optimal, a.s., and it can be designed to run in* $O\left(N^2T^4(\log N + \log T + \log^2 N / \mathrm{l}\right.$ *time as well.*

Proof: The proof is analogous to that of Theorem 3.1 and 3.2 and is, in fact, simpler.

$\lim_{T\to\infty} \frac{z^I - z^*}{T} = 0$ a.s., is verified as in the proof of Theorem 3.1. Moreover, it was shown there that for $\lim_{T\to\infty} \frac{z^{II} - z^I}{T} = 0$ a.s., it is sufficient to verify that (3.10) holds. Since $L_\ell \le \lambda$, (3.10) reduces to showing that $\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} I_t^{0,f}$ converges to a constant a.s.. This, however, follows from (A^f), since, by (16), $I_t^{0,f}$ is a Lipschitz continuous function of $\{(D_{t+1}, C_{t+1}), (D_{t+2}, C_{t+2}), \ldots, (D_{t+\lambda}, C_{t+\lambda})\}$.

We now verify the complexity bounds. Under a fixed life time, the minimum cost network flow problem to be solved in each node of the b&b-trees, associated with the different interval instances, now needs to be solved by a standard method, rather than the Ahuja and Hochbaum (2004) method. The best strongly polynomial time algorithm to solve minimum cost network flow problems is due to Orlin (1989). The network flow model has a source, a sink and two sets of nodes; the first set has a node for every period and the second one has a node for every period / item combination. Thus, the model has $O(N\tau)$ nodes and $O(N\tau)$ arcs in the (SP)-heuristic and $O(NT)$ nodes and arcs in the (EH)-heuristic. Orlin's method solves the problem therefore in $O(N^2\tau^2\log^2 N\tau)$ and $O(N^2T^2\log^2 NT)$, respectively. Since $J = T/\tau$ interval instances are solved and since in each, in the worst case, all 2^T nodes of the b&b tree need to be evaluated, the complexity bounds follow readily.

■

Acknowledgment: We are greatly indebted to Assaf Zeevi for many helpful suggestions and comments regarding earlier versions of this chapter.

Chapter 4

Dynamic Pricing Strategies for Multi-Product Revenue Management Problems

4.1 Introduction

Consider a firm that owns a fixed capacity of a certain resource that is consumed in the process of producing or offering multiple products or services, and which must be consumed over a finite time horizon. The firm's problem is to maximize its total expected revenues by selecting the appropriate dynamic controls. We will consider two well studied problem formulations. In the first, the firm is assumed to be a monopolist or to operate in a market with imperfect competition, and thus to have power to influence the demand for each product by varying its price. In this setting, the firm's problem is to choose a dynamic pricing strategy for each of its products in order to optimize expected revenues. In the second situation, prices are assumed to be fixed either by the competition or through a higher-order optimization problem, and the firm's problem is now to choose a dynamic capacity allocation rule that controls when to accept new requests for each of these products. In the sequel, these two problems will be referred to as the 'dynamic pricing' and 'capacity control' formulations, respectively. Revenue management problems of that sort originated in the late 1970's in the context of the airline industry, and have since been successfully introduced in a variety of other areas such as hotels, cruise lines, rental cars, retail etc. For example, the first of these problems may arise in the retail industry, while the second one tends to be associated with the airline industry (although there are examples of airlines that practice revenue management through a dynamic pricing policy as well).

Both of these problems have been studied quite extensively in the revenue management literature, in each case highlighting the structure of the optimal controls, proposing near-optimal heuristics, and evaluating their performance in extensive numerical studies for both stylized examples as well as real-life applications. This paper illustrates how these two problems can be reduced to a common formulation and thus be treated in a unified manner, and explores some of the consequences of this formulation. Broadly speaking this is done as follows: Consider a firm that owns capacity of a single resource and offers multiple products, and suppose for now that the aggregate rate at which capacity is consumed is given. One can then compute the vector of product prices or capacity controls in each case in order to maximize the instantaneous expected revenue subject to the constraint that the aggregate capacity consumption equals the aforementioned rate. This is akin to the basic microeconomics problem of resource allocation subject to a budget constraint. Its solution, which in some cases can be obtained in closed form, defines an *aggregate expected revenue rate* as a function of the capacity consumption rate, and identifies the optimal way of translating the latter into a vector of product-level controls. One can now reformulate both the 'dynamic pricing' and the 'capacity control' problems as 'single resource, single product' pricing problems, where the firm controls the instantaneous resource consumption rate and where revenues are accrued according to appropriate *aggregate revenue function.* Pricing and capacity decisions are then extracted from the optimal capacity consumption rate in the manner described above.

This formulation constitutes the main modelling contribution of the Chapter, which then proceeds to explore some of its theoretical and practical implications in dynamic pricing and capacity control revenue management problems. Specifically, we show that the multi-product dynamic pricing problem introduced by GALLEGO and VAN RYZIN (1997) and the capacity control problem of LEE and HERSH (1993) can be recast within this common framework, and be treated as different instances of a single-product pricing problem for appropriate concave revenue functions (Propositions 4.1 and 4.2). This highlights the common structure of the pricing and capacity control problems and allows us to treat both in a unified framework. This, of course, recovers well-known structural results regarding the monotonicity properties of the value function and the associated controls (see Propositions 4.3 and Corollary 4.2) that were previously derived in the literature while studying each of these problems in isolation, see, e.g., GALLEGO and VAN RYZIN (1994, 1997), LEE and HERSH (1993), LAUTENBACHER and

Stidham (1999), Zhao and Zheng (2000), and the recent book by Talluri and van Ryzin (2004a). Moreover, Corollary 4.1 establishes that the optimal multi-product pricing policy has a certain monotone and nested structure. Although our analysis is done in a stationary setting, most of our results extend to allow for non-stationary demand models, in which case the associated dynamic pricing single-product problems are of the form studied by Zhao and Zheng (2000).

A similar type of demand aggregation appeared in Talluri and van Ryzin (2004) while analyzing a capacity control problem for a system where customer behavior is captured through a discrete choice model. Demand aggregation techniques have been exploited in the past in the numerical solution of the dynamic programs associated with these revenue management problems. Finally, similar ideas of demand aggregation arise in the context of 'equivalent workload formulations' in stochastic network theory; see Harrison and van Mieghem (1996) for background, and Maglaras (2003b) for a recent application of this idea in the context of a joint pricing and scheduling problem.

This new formulation leads to several qualitative and quantitative insights. The first concerns the derivation of simple pricing and capacity control heuristics. Specifically, both multi-product formulations are reduced into dynamic pricing problems for single-product models of the type studied by Gallego and van Ryzin (1994). Based on their analysis this implies that a static pricing heuristic is optimal for the deterministic and continuous (fluid) approximation of the underlying problems, and asymptotically optimal in an appropriate sense for the original problems. This was already observed by Gallego and van Ryzin (1997), but their characterization of that policy in the multi-product setting was implicit. Our reduction of the multi-product problem to an appropriate single-product one leads to a closed-form characterization of that fixed-price heuristic (see Proposition 4.4); see also the review article by BITRAN and CALDENTEY (2003) for a discussion of deterministic multi-product pricing problems. In contrast to the dynamic programming analysis discussed above, the fluid formulation of these problems allows us to consider a more general class of models where products may differ with respect to their capacity requirements. Based on the solution of the fluid formulation we propose three heuristics: (i) a static pricing heuristic; (ii) a static pricing heuristic applied in conjunction with an appropriate capacity allocation policy; and (iii) a 'resolving' heuristic that re-evaluates the fluid policy as a function of the current state and time-to-go (which is derived by expressing the fluid solution in feedback form).

The first of these heuristics was suggested by GALLEGO and VAN RYZIN (1997). Once prices are fixed, the firm's problem has been reduced to one of capacity control, of the type analyzed in LEE and HERSH (1993), which motivates the second heuristic. Policies that combine static prices with capacity controls as in (ii) have been suggested in other papers such as MCGILL and VAN RYZIN (1999), FENG and XIAO (2004), and LIN ET. AL. (2003). Finally, the 'resolving' heuristic (iii) is widely applied in practice, but to the best of our knowledge has not been analyzed theoretically thus far. The only exception was the negative result of COOPER (2002) that illustrated through an example that resolving may in fact do worse than applying the static fluid policy.

Subsection 4.4.2 establishes that all three heuristics achieve asymptotically optimal performance under fluid scaling, i.e., in the spirit of GALLEGO and VAN RYZIN (1997) and COOPER (2002) (Propositions 4.5 to 4.7). These results show that the phenomenon demonstrated in Cooper's example does not persist in problems with large capacity and demand, where in fact resolving achieves the asymptotically optimal performance. Moreover, the numerical results of Section 4.5 illustrate that the dynamic heuristics (ii) and (iii) tend to improve the firm's performance.

The second insight is structural and offers a partial characterization of good pricing and capacity allocation policies. The formulation advanced in this paper and specifically the subproblem of translating a capacity consumption rate into a set of product-level controls that jointly maximize instantaneous revenue, defines an *efficient frontier* for the firm's pricing and capacity control strategy. This captures in a tractable way the interactions between products due to cross-elasticity effects and the joint capacity constraint. The idea of an efficient frontier has also appeared in Talluri and van Ryzin (2004) in the context of a capacity control problem for a model with customer choice among products, and in Feng and Xiao (2000, 2004) while studying pricing problems with a predetermined set of price points; the latter set of papers uses the term maximum increasing concave envelope.

While the main emphasis of this paper is not computational, it is worth noting that the control dimension reduction presented in this paper may lead to computational simplifications in cases where the subproblem of inferring the optimal demand rates given an aggregate consumption rate is solvable in closed-form. While this is not generally true, it does admit a simple solution in cases such as the linear demand model and the multinomial logit model (both are reviewed in Section 4.5), and it also leads to a simple characterization of the revenue function in the capacity control formulation. Moreover, it leads to algorithmic and computational simplifications in the case where there is a fixed set of price points for each product, as in the model studied in Feng and Xiao (2004).

Finally, we extend this formulation to the network case. The same decomposition in (a) choosing the aggregate resource consumption rates and then (b) translating these decisions to a vector of product prices still holds. As expected, the complexity of each of these steps increases in the network setting. However, the structural insights gleaned by this problem formulation provide a promising direction to follow in developing efficient network controls.

The remainder of the paper is structured as follows. This section concludes with some additional comments on the related literature. Section 4.2 describes the model and the associated problem formulations. Section 4.3 demonstrates the reduction of the dynamic programming formulations to that of a single-product pricing problem, and derives some of its structural properties. Section 4.4 discusses several insights and extensions that hinge on the previous results, and Section 4.5 provides some numerical illustration of our results and offers some concluding remarks.

The papers by Elmghraby and Keskinocak (2002), Bitran and Caldentey (2003), and McGill and van Ryzin (1999), and the book by Talluri and van Ryzin (2004a) provide comprehensive overviews of the areas of dynamic pricing and revenue management. The modelling framework adopted in this paper closely matches that of Gallego and van Ryzin (1994, 1997), and, as in their work, we also partly focus on deterministic and continuous-dynamics (fluid) approximations of the underlying discrete problems to derive simple heuristics for the multi-product revenue management problems. See also Kleywegt (2001) for a fluid model approach to multi-product network revenue management problem. Standard references on revenue management with capacity controls are Lee and Hersh (1993), Brumelle and McGill (1993) and Lautenbacher and Stidhman (1999).

4.2 Single Resource, Multi-Product Model

This section poses the multi-product dynamic pricing and capacity control problems. We start with the former, which follows the model of Gallego and van Ryzin (1997), and then provide the necessary changes to formulate the latter, which is the model analyzed by Lee and Hersh (1993).

4.2.1 Dynamic pricing model

Consider a firm that is endowed with C units of capacity of a single resource that is used in producing or offering multiple products or services, indexed by $i = 1, \ldots, n$. Each unit of product i consumes one unit of capacity, and for simplicity we assume that C is integer valued. There is a finite horizon T over which the resources must be used, and capacity cannot be replenished up to that time. The firm is either a monopolist or is assumed to operate in a market with imperfect competition, and, in that, has power to influence the demand for each product by varying its menu of prices. Let $p(t) = [p_1(t), \ldots, p_n(t)]$ denote the vector of prices at time t. The demand process is assumed to be n-dimensional non-homogeneous Poisson process with rate vector λ determined through a demand function $\lambda(p(t))$, where $\lambda : \mathcal{P} \to \mathcal{L}$, $\mathcal{P} \subseteq \mathbb{R}^n$ is the set of feasible price vectors, and $\mathcal{L} = \{x \geq 0 : x = \lambda(p),\ p \in \mathcal{P}\} \subseteq \mathbb{R}^n_+$ is the set of achievable demand rate vectors. We assume that $\mathcal{L}$ is a convex set. A simple interpretation of the demand model in the single product case is as follows: if the total arrival rate of potential customers is Λ, and each customer has an independent identically distributed (IID) reservation price (i.e., maximum price they are willing to pay) for the product drawn from some distribution F, then customers that have reservation prices $\geq p$ choose to buy the product and $\lambda(p) = \Lambda(1 - F(p))$. Following GALLEGO and VAN RYZIN (1994, 1997) we consider *regular* demand functions that satisfy the following requirements. In the sequel, x' will denote the transpose of any vector or matrix x (the use of T is reserved for the time horizon), and for any real number y, $y^+ := \max(0, y)$.

Assumption 4.1 *A demand function is said to be regular if it is a continuously differentiable, bounded function, and: (a) for each product i, $\lambda_i(p)$ is strictly decreasing in p_i, (b) $\lim_{p_i \to \infty} \lambda_i(p) = 0$ (i.e., consumers have bounded wealth), and (c) the revenue rate $p'\lambda(p) = \sum_{i=1}^n p_i \lambda_i(p)$ is bounded for all $p \in \mathcal{P}$ and has a finite maximizer $\bar{p}$.*

Note that this class of demand functions incorporate product complementarity and substitutions effects. For ease of exposition, we assume that the demand function is stationary. The extension to a non-stationary demand model can be done along the lines of GALLEGO and VAN RYZIN (1994, 1997) and ZHAO and ZHENG (2000). Also, this demand model assumes that customer decisions only depend on the current price vector and not on past and/or future pricing decisions. A more general model would allow customers to learn the firm's pricing policy and adjust their actions accordingly, and thus incorporate the strategic interaction between the firm and the customers' collective behavior. While this may seem appealing, it leads to a complicated game-theoretic analysis, and is often avoided both in the revenue management literature and in practice. See VULCANO ET. AL. (2002, § 4.3) for a discussion of this issue with appropriate pointers to the literature.

We assume that there exists a continuous inverse demand function $p(\lambda)$, $p : \mathcal{L} \to \mathcal{P}$, that maps an achievable vector of demand rates λ into a corresponding vector of prices $p(\lambda)$. This allows us to view the demand rate vector as the firm's control, and once this is selected infer the appropriate prices using the inverse demand function. The expected revenue rate as a function of the vector of demand rates, λ, is defined by $R(\lambda) := \lambda' p(\lambda)$. In the sequel we will assume that the revenue rate functional $R(\cdot)$ is continuous, bounded and strictly concave.

Example: Under a linear demand model, the demand for product i is given by

$$\lambda_i(p) = \Lambda_i - b_{ii}p_i - \sum_{j \neq i} b_{ij}p_j,$$

where Λ_i is the market potential for product i and b_{ii}, b_{ij} are the price and cross-price sensitivity parameters. This can be expressed in vector form as $\lambda(p) = \Lambda - Bp$, for the obvious choice of Λ, B. The inverse demand function is then given $p(\lambda) = B^{-1}(\Lambda - \lambda)$, and the revenue function is given by $R(\lambda) = \lambda' B^{-1}(\Lambda - \lambda)$. Assumption 1 requires that $b_{ii} > 0$ for all i. To ensure that B^{-1} exists and the inverse demand function is well defined, and that the revenue function is concave we further require that either $b_{ii} > \sum_{j \neq i} |b_{ji}|$ or $b_{ii} > \sum_{j \neq i} |b_{ij}|$ for all i; both conditions guarantee that B is invertible and that its eigenvalues have positive real parts (see HORN and JOHNSON (1994, Thm. 6.1.10). If the products are substitutes ($b_{ij} \leq 0$ for all $j \neq i$), the first condition reduces to $\sum_j b_{ji} > 0$, which implies that an increase in p_i leads to a reduction in the total demand for all products. Alternatively, the second condition says that the effect of a marginal increase in p_i on the demand for product i is bigger than that of a marginal increase on the prices of all other products.

The problem that we address is roughly described as follows: given an initial capacity C for that resource, a selling horizon T, and a demand function that maps the menu of prices into a set of demand rates for each product, the firm's goal is to choose a non-anticipating dynamic pricing strategy for each product in order to maximize its total expected revenues. The restriction to non-anticipating policies implies that the decisions at time t can only depend on the current state of the system as well as the time history up to time t, and not on future events.

We assume that the salvage value of remaining capacity at time T is zero. (Otherwise, at least in the case where the salvage value per unit of capacity at time T is constant, one could modify the objective function to be total expected revenue in excess of the salvage values.) We will adopt a discrete-time formulation, which assumes that time has been discretized in small intervals of length δt, indexed by $t = 1, \ldots, T$, such that $\mathbb{P}(\text{product } i \text{ arrival in } [0, \delta t]) = \lambda_i \delta t + o(\delta t)$ for all products i, and $\mathbb{P}(\text{product } i \text{ and } j \text{ arrivals in } [0, \delta t]) = \lambda_i \lambda_j (\delta t)^2 + o((\delta t)^2)$, where $o(x)$ implies that $o(x)/x \to 0$ as $x \to 0$. With slight abuse of notation, in the sequel we will write λ_i in place of $\lambda_i \delta t$; i.e., λ_i will not refer to the rate of the underlying Poisson arrival process, but to the probability that a product i request occurs in each time interval. We will refer to λ_i as the demand or the buying probability for product i, interchangeably. Hence, the corresponding demand random vector for period t, denoted by $\xi(t; \lambda)$, is Bernoulli with probabilities λ that are controlled by the vector of posted prices, and

$\mathbb{P}(\xi_i(t) = 1) = \lambda_i(p(t))$ and $\mathbb{P}(\xi_i(t) = 0) = 1 - \lambda_i(p(t))$ for all products i. As stated above, we will treat the demand rates λ_i as the control variables and infer the appropriate prices through the inverse demand relationship. The discrete-time formulation of the dynamic pricing problem of GALLEGO and VAN RYZIN (1997) is as follows:

$$\max_{\{\lambda(t),\ t=1,\dots,T\}} \left\{ \mathbb{E}\left[\sum_{t=1}^{T} p(\lambda(t))'\xi(t;\lambda)\right] \ :\ \sum_{t=1}^{T} e'\xi(t;\lambda) \leq C \text{a.s. and } \lambda(t) \in \mathcal{L} \forall t \right\}, \tag{4.1}$$

where e is the vector of ones of appropriate dimension and 'a.s.' stands for almost surely.

4.2.2 The Single-Resource Capacity Control Problem

The second problem that we will consider is the one studied by LEE and HERSH (1993) that takes the product prices as exogenously fixed and strives to optimize over the capacity allocation decisions. In more detail, the price vector p is fixed a priori, and this also fixes the corresponding vector of demand rates $\lambda = \lambda(p)$. In the context of this problem and without any loss of generality we will assume that the products are labelled in such a way that $p_1 \geq p_2 \geq \cdots \geq p_n$. The firm has discretion as to which product requests to accept at any given time. This is modelled through the control $u_i(t)$ for each product i, which is equal to the probability of accepting such a request at time t. It is customary to assume that the firm is 'opening' or 'closing' products (or fare classes), thus leading to controls $u_i(\cdot)$ that take the values of 0 or 1, but this need not be imposed as a restriction. A more general class of controls could allow the firm to flip a coin upon arrival of a product i request with probability of success given by $u_i(t)$ and make the accept/rejection decision accordingly. The dynamic capacity control problem can be formulated as follows:

$$\max_{\{u(t),t=1,\ldots,T\}} \left\{ \mathbb{E}\left[\sum_{t=1}^{T} p'\xi(t;u\lambda) \right] : \sum_{t=1}^{T} e'\xi(t;u\lambda) \leq C \text{a.s. } 1 \text{ and } u_i(t) \in [0,1] \forall t \right\}, \tag{4.2}$$

where $u\lambda$ above denoted the vector with coordinates $u_i\lambda_i$.

4.3 Analysis of the pricing and capacity control problems

This section describes how to reduce (4.1) and (4.2) into dynamic optimization problems where the control is the (one-dimensional) aggregate capacity consumption rate. Subsequently, we derive some structural properties for these two problems through a unified analysis.

4.3.1 Dynamic programming formulation of the pricing problem

Consider the dynamic pricing problem posed in (4.1). Let x denote the number of remaining units of capacity at the beginning of period t, and $V(x,t)$ be the expected revenue-to-go starting at time t with x units of capacity left. Then, the Bellman equation associated with (4.1) is:

$$V(x,t) = \max_{\lambda \in \mathcal{L}} \left\{ \sum_{i=1}^{n} \lambda_i \left[p_i(\lambda) + V(x-1,t+1) \right] + (1 - e'\lambda)\, V(x,t+1) \right\}, \tag{4.3}$$

with the boundary conditions

$$V(x, T+1) = 0 \ \ \forall\ x \quad \text{and} \quad V(0, t) = 0 \ \ \forall\ t. \tag{4.4}$$

Letting $\Delta V(x,t) = V(x,t+1) - V(x-1,t+1)$ denote the marginal value of one unit of capacity as a function of the state (x,t), (4.3) can be rewritten as

$$V(x,t) = \max_{\lambda \in \mathcal{L}} \left\{ R(\lambda) - \sum_{i=1}^{n} \lambda_i \Delta V(x,t) \right\} + V(x,t+1). \tag{4.5}$$

The gist of the proposed solution approach is to rewrite (4.5) in terms of the one-dimensional *aggregate* rate of capacity consumption defined by $\rho := \sum_{i=1}^{n} \lambda_i$. We first consider the maximum achievable revenue rate subject to the constraint that all products jointly consume capacity at a rate ρ, which is given by the solution to the following parametric optimization problem

$$R^r(\rho) := \max_{\lambda} \ \{R(\lambda) \ : \ \textstyle\sum_{i=1}^{n} \lambda_i = \rho, \ \lambda \in \mathcal{L}\}. \tag{4.6}$$

Note that (4.6) is concave maximization problem over a convex set, and its solution is readily computable, often in closed form (examples are given in Section 4.5). Moreover, $R^r(\cdot)$ is a concave function and satisfies the conditions of Assumption 1, and $R^r(\rho)/\rho$ is the optimal 'average' price per unit of capacity, which is interpreted as an inverse demand function for this aggregate model. The unique vector of demand rates that achieves that optimum will be denoted by $\lambda^r(\rho)$. Let $\mathcal{R} := \{\rho \ : \ \sum_{i=1}^{n} \lambda_i = \rho, \ \lambda \in \mathcal{L}\}$ be the set of achievable capacity consumption rates. Then:

Proposition 4.1 *The dynamic pricing problem (4.1) can be reduced to the dynamic program*

$$V(x,t) = \max_{\rho \in \mathcal{R}} \{R^r(\rho) - \rho \Delta V(x,t)\} + V(x,t+1), \tag{4.7}$$

and the boundary condition (4.4) expressed in terms of the one-dimensional aggregate consumption rate. In particular, if $\rho^*(x,t)$ denotes the optimal solution of (4.7) and (4.4) and $\lambda^*(x,t)$ and $p^*(x,t)$ denote the optimal demand rate and price vectors associated with (4.1), then,

$$\lambda^*(x,t) = \lambda^r(\rho^*(x,t)) \quad \textit{and} \quad p^*(x,t) = p(\lambda^r(\rho^*(x,t))). \tag{4.8}$$

Next, we show how the capacity control formulation can also be reduced to a problem that resembles (4.7), and then proceed to analyze the structure of both problems in an unified manner.

4.3.2 DP formulation of the capacity control problem

Using the notation established above, the Bellman equation associated with (4.2) is

$$V(x,t) = \max_{u_i \in [0,1]} \left\{ \sum_{i=1}^{n} \lambda_i u_i [p_i + V(x-1,t+1)] + (1 - u'\lambda)\, V(x,t+1) \right\} \tag{4.9}$$

with the boundary condition (4.4), which using the marginal value of capacity ΔV becomes

$$V(x,t) = \max_{u_i \in [0,1]} \left\{ \sum_{i=1}^{n} \lambda_i u_i p_i - u'\lambda \Delta V(x,t) \right\} + V(x,t+1). \tag{4.10}$$

Suppose products are labelled such that $p_1 \geq p_2 \geq \cdots \geq p_n$. Observe that $\rho = u'\lambda$ and define

$$R^a(\rho) = \max_{u} \left\{ \sum_{i=1}^{n} u_i \lambda_i p_i \;:\; u'\lambda = \rho,\; u_i \in [0,1] \right\}$$

to be the maximum instantaneous revenue rate when the aggregate capacity consumption rate (or probability of a sale at that time) is given by ρ. Let $u^a(\rho)$ be the control that attains that maximum. Using the structure of the knapsack problem that defines $R^a(\cdot)$ we get that

$$R^a(\rho) = \min_{i} \; c_i + p_i \rho \qquad \text{and} \qquad u_k^a(\rho) = \min\left(\frac{(\rho - \sum_{i<k} \lambda_i)^+}{\lambda_k}, 1 \right), \tag{4.11}$$

where $c_1 = 0$ and $c_i = \sum_{k<i} \lambda_k (p_k - p_i)$, and for any $x \in \mathbb{R}$, $x^+ := \max(x, 0)$. That is, the optimal solution starts from offering product 1 (with the highest price) and keeps adding more products until the total probability of a sale in period t becomes equal to ρ. In practice, of course, one would always set $u_1^a(\rho) = 1$ for all $\rho \geq 0$, i.e., it is never optimal to close the highest-fare class. The 'average' expected selling price is $R^a(\rho)/\rho = \min_i c_i/\rho + p_i$, which is, of course, decreasing in ρ, and $R^a(\cdot)$ is concave. Combining these results, (4.2) can be expressed as a dynamic problem in terms of ρ as

follows:

$$V(x,t) = \max_{0 \le \rho \le \sum_{i=1}^{n} \lambda_i} \left\{ R^a(\rho) - \rho \Delta V(x,t) \right\} + V(x,t+1). \tag{4.12}$$

subject to the boundary condition (4.4). This structural result is summarized below.

Proposition 4.2 *The capacity control problem (4.2) can be reduced to the dynamic program (4.12) and (4.4) expressed in terms of the one-dimensional aggregate consumption rate* ρ*. In particular, if* $\rho^*(x,t)$ *denotes the optimal solution of (4.12) and (4.4) and* $u^*(x,t)$ *denote the optimal policy for (4.2), then* $u^*(x,t) = u^a(\rho^*(x,t))$.

This last result was also derived in TALLURI and VAN RYZIN (2004), while considering the capacity control problem for a model with customer choice. While their model and emphasis was different than ours, one of their key findings was also that the optimal policy in multi-product settings can be expressed in terms of the aggregate probability of a sale and the associated expected revenue, i.e., ρ and $R^a(\rho)$.

4.3.3 A unified analysis of the pricing and capacity control problems

The similarity of the dynamic programs for the pricing and capacity control problems highlights their common structure and allows them to be treated in a unified manner. For both (4.7) and (4.12) the optimal control $\rho^*(x,t)$ is computed from

$$\rho^*(x,t) = \operatorname*{argmax}_{\rho \in \mathcal{R}} \{R(\rho) - \rho \Delta V(x,t)\},$$

where $R(\cdot)$ is any concave increasing revenue function. Using the properties of $R(\cdot)$ one gets that $\rho^*(x,t)$ is decreasing in $\Delta V(x,t)$, which using a backwards induction argument in t gives that $\Delta V(x,t)$ is decreasing in x and t. These standard results for single-product dynamic pricing problems are summarized below; a proof can be found in Talluri and van Ryzin (2004, Prop. 5.2 Ch. 4).

Proposition 4.3 Talluri *and* van Ryzin *(2004, Prop. 5.2 Ch. 4) For both problems defined in (4.1) and (4.2) we have that:*

1. *$\rho^*(x,t)$ is decreasing in the marginal value of capacity $\Delta V(x,t)$, and*
2. *$\Delta V(x,t)$ is decreasing in x and t.*

Next, we specialize these results to the dynamic pricing and capacity allocation problems. We first consider the dynamic pricing problem and for illustrative purposes focus on the case where the products are non-substitutes. That is, the demand for product i is only a function of the price for that product p_i. In that case, the Lagrangian associated with (4.6) is given by $L(\lambda,x,y) = R(\lambda) + x(\rho - \sum_{i=1}^{n} \lambda_i) - y'\lambda$, and the associated first order conditions are given by $\partial R(\lambda)/\partial\lambda_i = x + y_i$, for some $x \geq 0$ and $y_i \leq 0$ with $y_i = 0$ if $\lambda_i > 0$. It is easy to show that x is decreasing in ρ (i.e., the shadow price for the capacity consumption constraint decreases as the associated rate ρ increases), and that $\lambda_i^r(\rho)$ is decreasing in x. This is summarized below.

Corollary 4.1 *Consider the problem specified in (4.3) and (4.4) and further assume that the products are non-substitutes, i.e., $\lambda_i(p) = \lambda_i(p_i)$ for all i. $\lambda_i^*(x,t)$ is non-decreasing in $\rho^*(x,t)$ (and non-increasing in $\Delta V(x,t)$).*

A similar result can be obtained when products are substitutable provided that the demand model satisfies certain conditions analogous to the ones given for the price sensitivity matrix B in §4.2.1 when describing the linear demand model example.

Finally, we specialize these results to the capacity control problem to recover some well known structural properties of the optimal policy, see, e.g., Lee and Hersh (1993). Our derivation based on the aggregate control offers some new intuition as to why they hold. To start with, the optimal control $\rho^*(x,t)$ is the solution to the following optimization problem

$$\max_{0\le\rho\le\sum_{i=1}^{n}\lambda_i} \quad \min_i \; c_i + (p_i - \Delta V(x,t))\rho.$$

Let $i^*(x,t) = \max\{i \ge 1 : p_i \ge \Delta V(x,t)\}$. Then, by inspecting the form of the piecewise linear objective function involved in the calculation of $\rho^*(x,t)$ we get that

$$\rho^*(x,t) = \sum_{i\le i^*(x,t)}^{n} \lambda_i.$$

That is, the solution is 'bang-bang' in the sense that the form of the optimal control is such that $u_i^*(x,t)$ is 0 if $i > i^*(x,t)$ and 1 if $i \le i^*(x,t)$. In addition, from Proposition 4.3 part 1 we see that $i^*(x,t)$ is decreasing in the marginal value of capacity $\Delta V(x,t)$. Therefore:

Corollary 4.2 *For the capacity control problem (4.2) or equivalently, (4.12) and (4.4), the optimal allocation policy is* nested *in that* $u_i^*(x,t) = 1$ *if* $i \le i^*(x,t)$, *and* $u_i^*(x,t) = 0$ *otherwise, and* $i^*(x,t)$ *is decreasing in the marginal value of capacity* $\Delta V(x,t)$.

4.3.4 An efficient frontier for multi-product pricing and capacity controls

The subproblem of computing the optimal revenue subject to a constraint on the aggregate capacity consumption rate specified in (4.6) and (4.11) defines an efficient frontier $(\rho, R^r(\rho))$ and $(\rho, R^a(\rho))$ for the dynamic pricing and capacity allocation problems, respectively. As in the context of portfolio optimization, the efficient frontier provides a systematic framework for comparing different policies and highlights the structure of the optimal controls for these two problems. It may also lead to computational improvements if this subproblem can be solved efficiently, preferably in closed form. This can be achieved for some commonly used demand models such as the linear and the multinomial logit; both are reviewed in section 4.5.

As mentioned in the introduction the idea of an efficient frontier has also appeared in Feng and Xiao (2000, 2004) and Talluri and van Ryzin (2004). The structure of the dynamic programs studied in this section has been observed in other revenue management papers, such as Lin et. al. (2003) and their study of single-resource capacity control problems where each arrival may request multiple units of capacity, and Vulcano et. al. (2002) and their analysis of optimal dynamic auctions. The second of these papers studies a discrete time model where demand arrives in batches, and in each period the firm runs an auction among the potential buyers of that period. Given an announced auction mechanism, the firm observes the bids and selects how many units to award by balancing total consumption with the extracted revenues in that period. This is the discrete-time, batch demand analog to choosing ρ and computing $R^r(\rho)$, and in that sense the dynamic programming structure in Vulcano et. al. (2002)

is closely related to the one observed here.

4.4 Deterministic analysis of the multi-product pricing problem

This section studies deterministic (fluid model) formulations of multi-product revenue management problems to provide some structural results (Section 4.4.1) and suggest simple and implementable heuristics for the underlying problems (Section 4.4.2). The latter have desirable theoretical performance guarantees and are shown to perform well in the numerical experiments of the next section. Finally, section 4.4.3 sketches how to extend these ideas to the network setting.

4.4.1 Solution to the deterministic multi-product pricing problem

The dynamic program of section 4.3.1 is generally not solvable in closed form, and one has to refer to numerical techniques for the computation of the optimal pricing decisions, which is often difficult and results in policies that are hard to implement. This motivates the use of approximate models that are analytically tractable and may lead to practical solutions. The most natural candidate is the 'fluid' model that has deterministic and continuous dynamics, and is obtained by replacing the discrete stochastic demand process by its rate, which now evolves as continuous process. It is rigorously justified as a limit under a strong-law-large-numbers type of scaling when we let the potential demand and the capacity grow proportionally large; see GALLEGO and VAN RYZIN (1994, 1997), and Section 4.4.2.

It is simplest to describe the fluid model in continuous time (this is consistent with Gallego and van Ryzin (1994, 1997)). In more detail, the realized instantaneous demand for product i at time t in the fluid model is deterministic and given by $\lambda_i(t)$. and allow product i requests to consume capacity at a rate of $a_i > 0$ units per unit of demand, and denote by a the vector $[a_1, \dots, a_n]$. This is a generalization of the model considered thus far that assumed uniform capacity requirements that with no loss of generality can be taken to be 1, which however can be addressed with no increase in complexity. With a general capacity requirement vector a the capacity consumption rate is defined by $\rho = a'\lambda$, and the definitions of R^r and λ^r can be appropriately adjusted to reflect that change. The system dynamics are given by $dx(t)/dt = -\sum_{i=1}^n a_i \lambda_i(t)$, $x(0) = C$, together with the boundary condition that $x(T) \geq 0$. i.e., this model has deterministic and continuous dynamics. The firm selects a demand rate $\lambda_i(t)$ (or a price) at each time t. The fluid control formulation of our revenue management problem is the following:

$$\max_{\{\lambda(t), t\in[0,T]\}} \left\{ \int_0^T R(\lambda(t))dt \ : \ \int_0^T a'\lambda(t)dt \leq C \text{ and } \lambda(t) \in \mathcal{L} \ \forall t \right\}. \tag{4.13}$$

Single-product problem: For the case with *one* product Gallego and van Ryzin (1994) showed that a fixed price (and thus a constant demand rate) is optimal for (4.13). This is described as follows. Let $\hat{\lambda} = \operatorname{argmax}\{R(\lambda) : \lambda \in \mathcal{L}\}$ and $\hat{p} = p(\hat{\lambda})$ be the demand rate and price that maximize the instantaneous revenue rate in the absence of any capacity considerations, respectively. Also, define $\lambda^0 = C/T$ be the *run-out* rate that depletes capacity at time T, and let $p^0 = p(\lambda^0)$ be the associated price. Then, GALLEGO and VAN RYZIN (1994) showed that

$$\bar{\lambda} = \min(\hat{\lambda},\ \lambda^0) \qquad \text{and} \qquad \bar{p} = \max(\hat{p},\ p^0)$$

are the optimal demand rate and price for the fluid formulation of the pricing problem given in (4.13). (To differentiate from the solution of the dynamic programming formulation of the previous section we have used an overbar to denote the fluid solution.) That is, the optimal demand and price are constant and do not depend on the state of the system at any given time t. Intuitively, the solution uses the revenue maximizing price $\hat{p}$ unless this will deplete the capacity too soon, in which case the firm can increase its unit price to p^0 and sell its capacity by time T, while accruing higher total revenues. GALLEGO and VAN RYZIN (1997, §4.5) extended these results to multiple products, but in that case without providing such a succinct solution.

Multi-product problem: Following the approach of section 4.3 we can reduce the multi-product problem to an appropriate single-product one, and thus solve it in closed form. Specifically, recalling the definitions of the aggregate revenue function $R^r(\rho)$ and optimal demand rate vector $\lambda^r(\rho)$ in (4.6) adjusted for the fact that $\rho = a'\lambda$, (4.13) can be rewritten as:

$$\max_{\{\rho(t), t\in[0,T]\}} \left\{ \int_0^T R^r(\rho(t))dt \ :\ \int_0^T \rho(t)dt \le C,\ \rho(t) \in \mathcal{R}\ \forall t \right\}. \tag{4.14}$$

Note that (4.14) is the same as (4.13) for a single product with the revenue function R^r, and hence, it is solvable using the approach described above. Let $\rho^0 := C/T$ and

$\hat{\rho} = \text{argmax}_{\rho} R^r(\rho)$. Then, the optimal solution to (4.14) is to consume capacity at a constant rate $\bar{\rho}$ given by

$$\bar{\rho} := \min(\hat{\rho},\ \rho^0). \tag{4.15}$$

The corresponding vector of demand rates that maximizes the instantaneous revenues subject to the constraint that capacity is consumed at a rate $\bar{\rho}$ is given by $\lambda^r(\bar{\rho})$, and the corresponding prices are $p(\lambda^r(\bar{\rho}))$. A direct verification that this solution satisfies the optimality conditions for (4.13) establishes the following result.

Proposition 4.4 *Let $\bar{\lambda}(\cdot)$ and $\bar{p}(\cdot)$ denote the optimal vectors of demand rates and prices for (4.13). Then, $\bar{\lambda}, \bar{p}$ are constant over time and are given by*

$$\bar{\lambda}(t) = \lambda^r(\bar{\rho}) \quad \textit{and} \quad \bar{p}(t) = p(\lambda^r(\bar{\rho})). \tag{4.16}$$

4.4.2 Heuristic policies extracted from the deterministic analysis

Based on the preceding analysis we propose three heuristics for the underlying revenue management problems, which we analyze in the asymptotic setting introduced in GALLEGO and VAN RYZIN (1997) and COOPER (2002). Among other things we will show that the dynamic heuristic that 'resolves' the fluid policy as t progresses is asymptotically optimal in an appropriate sense.

I. Pricing heuristics

a. A static pricing heuristic: The first and simplest of our heuristics implements a static pricing policy $\bar{p}$ specified in Proposition 4.4. This policy corresponds to the make-to-order heuristic of GALLEGO and VAN RYZIN (1997).

The static nature of policy (a) is desirable for implementation purposes (see GALLEGO and VAN RYZIN (1997)), but also removes any form of capacity control. This issue does not arise in the fluid formulation, because capacity is then drained along an optimal deterministic trajectory, but it may be relevant in the underlying stochastic problem when capacity is close to being depleted. The next heuristics provide two possible adjustments to this static policy that add such capacity control capability and seem of practical interest. We start by recognizing that the solution of the fluid pricing problem of Section 4.4.1 can also be described in feedback form as

$$\bar{\rho}(x,t) = \min\left(\hat{\rho}, \frac{x}{T-t}\right), \tag{4.17}$$

where x is the remaining capacity at time t. The deterministic trajectory of the fluid model is, of course, such that $x/(T-t) = C/T$ for all t if $\hat{\rho} \geq C/T$, and $x/(T-t) = (C-\hat{\rho}t)/(T-t) \geq C/T$ if $\hat{\rho} < C/T$. In both cases, $\bar{\rho}(x,t) = \min(\hat{\rho}, C/T)$ for all x, t along the fluid trajectory of the capacity process, and thus (4.17) is identical to the static control derived in (4.15).

b. A List Price Capacity Control (LPCC) heuristic: One way to implement (4.17) is by coupling capacity control decisions together with the static pricing policy given in (a). Specifically, our second heuristic is defined as follows:

1. price according to $\bar{p}$ and label products such that $\bar{p}_1/a_1 \geq \bar{p}_2/a_2 \geq \cdots \geq \bar{p}_n/a_n$, and

2. compute $\bar{\rho}(x,t)$ and use the capacity controls $u_1(x,t) = 1$ if $x > 0$, $u_1(0,t) = 0$,

and

$$u_i(x,t) = \begin{cases} 1 & \text{if } \bar{\rho}(x,t) - \sum_{j<i} a_j \bar{\lambda}_j \geq a_i \bar{\lambda}_i \\ 0 & \text{otherwise} \end{cases} \quad \text{for } k \geq 2. \tag{4.18}$$

Note that with prices fixed, this control can only reduce the aggregate capacity consumption rate over its nominal value of $\sum_{i=1}^{n} a_i \bar{\lambda}_i$, but can never increase it. This heuristic makes a product available only if the fluid solution starting from that state would choose to sell this product in all future time periods, and 'closes' the product if the fluid solution would dictate only partial acceptance of the associated demand. When implementing in a discrete-time setting, rounding errors can be avoided by setting $u_k(x,t) = 1$ if $\bar{\rho}(x,t) - \sum_{i<k} a_i \bar{\lambda}_i \geq (T-t+1)(a_k \bar{\lambda}_k)$.

This policy is a refinement of the static pricing policy in (a) and the make-to-order heuristic of Gallego and van Ryzin (1997). Other examples of joint pricing and capacity controls can be found in the recent papers by Vulcano et. al. (2002), by Lin et. al. (2003) and Feng and Xiao (2004).

c. A dynamic pricing heuristic: The third policy translates the aggregate control $\bar{\rho}(x,t)$ into product-level rates (and prices) through

$$\lambda(x,t) = \lambda^r(\bar{\rho}(x,t)) \quad \text{and} \quad p(x,t) = p(\lambda(x,t)), \tag{4.19}$$

where the mapping $\lambda^r(\cdot)$ was the maximizer in (4.6). This corresponds to the idea of 'resolving' the fluid problem as we step through time, which is widely applied in practice, where, however, the resolving occurs at discrete points in time, e.g., daily or weekly depending on the application setting. Despite its practical appeal and use, to the best of our knowledge policies that use this resolving idea have not been analyzed theoretically, other than the isolated example provided by Cooper (2002) that illustrated that resolving may in fact degrade the performance of the fluid heuristic in a stochastic problem of multi-product capacity control problem. Our preceding discussion illustrates that 'resolving' is nothing but implementing the fluid policy in feedback form. The analysis that follows will characterize its behavior in an appropriate asymptotic sense, and the numerical results of the next section will demonstrate that it tends to outperform the other two candidate policies.

II. Asymptotic performance analysis of the pricing heuristics

The remainder of this subsection offers a brief asymptotic characterization of the performance under these three heuristics that shows that all three are (fluid-scale) asymptotically optimal in a regime where the potential demand and capacity grow proportionally large; this is consistent with the setup and the criterion of Gallego and van Ryzin (1997) and Cooper (2002). Specifically, using k as an index, we will consider a sequence of problems with demand model $\lambda^k(\cdot) = k\lambda(\cdot)$ and capacity $C^k = kC$, and we will let k increase to infinity; a superscript k will denote quantities that scale with k. Let N_i for $i = 1, \ldots, n$ denote independent unit rate Poisson processes, and recall the functional strong-law-of-large numbers for the Poisson process that asserts that as $k \to \infty$ and for all $t \geq 0$,

$$\frac{N_i(kt)}{k} \to t \quad \text{a.s.} \tag{4.20}$$

For all of the candidate policies the capacity dynamics can be expressed as follows. Let

$$A_i^k(t) = \int_0^t k\lambda_i(s)ds \tag{4.21}$$

where $\lambda_i(t)$ is the demand rate for product i at time t. Then, the cumulative demand for that product up to time t is equal to $N_i(A_i^k(t))$ and the remaining capacity at time t is

$$X^k(t) = C^k - \sum_{i=1}^n a_i N_i(A_i^k(t)). \tag{4.22}$$

Our goal here is to analyze the 'fluid-scale' behavior of the capacity process defined as $\bar{X}^k(t) := X^k(t)/k$ under the three candidate policies. The analysis of the static policy (a) is related to that of GALLEGO and VAN RYZIN (1997), while those of the dynamic policies are new.

Analysis of the static heuristic (a): In this case, the firm uses the constant price vector $\bar{p}$, which results in the demand rates $\lambda^k(\bar{p}) = k\bar{\lambda}$. We could either assume that the demand rates become 0 when the capacity is depleted, or keep them unchanged but modify (4.22) to $X_a^k(t) = (C^k - \sum_{i=1}^n a_i N_i(A_i^k(t)))^+$. (The subscript is used to identify the policy.) For simplicity we will proceed with the latter. For this policy we have that $A_i^k(t) = \bar{\lambda}_i kt$, and thus as $k \to \infty$

$$\frac{N_i(A_i^k(t))}{k} \to \bar{\lambda}_i t \quad \text{a.s.,} \quad \text{uniformly in } t \in [0, T]. \tag{4.23}$$

It immediately follows that as $k \to \infty$ and for all $t \in [0,T]$

$$\bar{X}_a^k(t) \to \left(C - \sum_{i=1}^{n} a_i \bar{\lambda}_i t \right)^+ = C - \bar{\rho} t \quad \text{a.s.;}$$

the $(\cdot)^+$ was removed since from (4.15) $\bar{\rho} t \le C$ for all $t \in [0,T]$. Let R_a^k denote the revenues extracted under policy (a), and $\tau^k := \inf\{s \ge 0 : \sum_{i=1}^n a_i N_i(\bar{\lambda}_i k s) \ge C^k\}$ be the random time where the aggregate capacity requested reaches or exceeds the available capacity C^k. Then,

$$R_a^k := \sum_{i=1}^{n} \bar{p}_i N_i(k \bar{\lambda}_i \min(T, \tau^k)) - \delta, \tag{4.24}$$

where δ is a random variable that corrects revenues for the case where $\tau^k < T$, which is bounded above by $\max_i \bar{p}_i$. (We will not delve into an accurate description of δ, since it will turn out to be asymptotically negligible.) From (4.23) and arguing by contradiction shows that $(T - \tau^k)^+ \to 0$ a.s., as $k \to \infty$. Combining with (4.24) we get the following result.

Proposition 4.5 *Suppose that demand and capacity are scaled according to $\lambda^k(\cdot) = k\lambda(\cdot)$ and $C^k = kC$, and consider the static pricing policy $p^k(x,t) = \bar{p}$ for all x, t and all k. Then, as $k \to \infty$, $\bar{X}_a^k(t) \to C - \bar{\rho} t$ a.s., uniformly in t, and $\frac{1}{k} R_a^k \to \sum_{i=1}^n \bar{p}_i \bar{\lambda}_i T$ a.s.*

That is, the static pricing policy (a) is asymptotically optimal in that it achieves as $k \to \infty$ the revenue extracted in the fluid model; this is the criterion used in Gallego and van Ryzin (1997) and Cooper (2002). This is also referred to as fluid-scale asymptotic optimality (see Maglaras (2000)) to highlight that it pertains to optimality with respect to the highest order revenue term. We will next analyze the dynamic pricing policy (c), and then return to deal with (b).

Analysis of the dynamic heuristic (c): The dynamic nature of this policy requires a more detailed study. The cumulative demand for product i up to time t is equal to $N_i(A_i^k(t))$, where

$$A_i^k(t) = \int_0^t k\lambda_i(s)ds \quad \text{where} \quad \lambda_i(s) = \lambda_i^r\left(\min(\hat{\rho}, X_c^k(s)/(k(T-s)))\right), \tag{4.25}$$

for $\lambda_i^r(\cdot)$ defined in (4.6), and $X_c^k(t)$ denotes the remaining capacity at time t under policy (c), defined in (4.22). Now, $\bar{X}_c^k(t)$ is bounded ($\bar{X}^k(t) \in [0, C]$) and therefore the sequence $\{\bar{X}_c^k(t) : t \in [0, T]\}$ is tight (see GLYNN (1990, §3)), and therefore has a converging subsequence, say $\{k_j\}$, such that along this subsequence, $\bar{X}_c^{k_j}(t) \to \bar{x}_c(t)$ a.s., uniformly in t; at this point limit trajectory may depend on the converging subsequence, but as we will see momentarily all possible limit solutions coincide. Given the continuity of $\lambda_i^r(\cdot)$ and using Lemma 2.4 of DAI and WILLIAMS (1994b) we get that as $k \to \infty$

$$\frac{1}{k}A_i^k(t) = \int_0^t \lambda_i^r\left(\min\left(\hat{\rho}, \frac{\bar{X}_c^k(s)}{T-s}\right)\right)ds \to \int_0^t \lambda_i^r\left(\min\left(\hat{\rho}, \frac{\bar{x}_c(s)}{T-s}\right)\right)ds \tag{4.26}$$

a.s., uniformly in t. Using (4.20), (4.22) and (4.26) we get that as $k \to \infty$

$$\begin{aligned}\bar{X}_c^k(t) = C - \frac{1}{k}\sum_{i=1}^n a_i N_i(A_i^k(t)) \quad &\to \quad C - \int_0^t \min\left(\hat{\rho}, \frac{\bar{x}_c(s)}{T-s}\right)ds \\ &= \quad C - \bar{\rho}t,\end{aligned} \tag{4.27}$$

where the last equality follows from identifying that this is the unique solution to (4.27), and the convergence is almost sure, uniformly in t. Finally, the revenues extracted under policy (c) are

$$R_c^k = \sum_{i=1}^{n} \int_0^t p_i^k(t) dN_i(A_i^k(t)),$$

where $p^k(t)$ is the price vector that corresponds to $\lambda^r(\min(\hat{\rho}, \bar{X}_c^k(t)/(T-t)))$, the demand rate vector at time t, and the integrals should be interpreted in the Riemann-Stieltjes sense. From (4.26) and (4.27) we have that $\lambda_i^r(\min(\hat{\rho}, \bar{X}_c^k(t)/(T-t))) \to \bar{\lambda}_i$ a.s., uniformly in t. By the continuity of the inverse demand function we have that $p^k(t) \to \bar{p}$ a.s., uniformly in t, and therefore using again Lemma 2.4 of Dai and Williams (1994b) we get the result summarized below.

Proposition 4.6 *Suppose that demand and capacity are scaled according to* $\lambda^k(\cdot) = k\lambda(\cdot)$ *and* $C^k = kC$, *and consider the dynamic pricing heuristic defined through (4.25). Then,* $\bar{X}_c^k(t) \to C - \bar{\rho}(t)$ *a.s., uniformly in* $[0,T]$, $\frac{1}{k}R_c^k \to \sum_{i=1}^n \bar{p}_i\bar{\lambda}_i T$ *a.s.*

Analysis of the LPCC heuristic (b): This policy is defined through $A_i^k(t) = k\bar{\lambda}_i \cdot \int_0^t u_i^k(t)dt$, where $u_i^k(t)$ was defined in (4.18) and can be expressed as follows:

$$\begin{aligned} u_1^k(t) &= \mathbf{1}\{\bar{X}_b^k(t) > 0\} \quad \text{and} \\ u_i^k(t) &= \mathbf{1}\left\{\min(\hat{\rho}\frac{\bar{X}_b^k(t)}{T-t}) - \textstyle\sum_{j<i} a_j\bar{\lambda}_j \geq a_i\bar{\lambda}_i\right\} \text{for } i \geq 2, \end{aligned} \tag{4.28}$$

where $\mathbf{1}\{\cdot\}$ is the indicator function. Similarly to the analysis of policy (c), the family $\{\bar{X}_b^k(t), t \in [0,T]\}$ is tight, and thus it has a converging subsequence $\{k_j\}$ on which

$\bar{X}_b^{k_j}(t) \to \bar{x}_b(t)$ a.s., uniformly in t. Writing down $\bar{x}_b(t)$ and evaluating $u_i(t)$, reveals that in the limit model $u_i(t) = 1$ for all products i. Arguments similar arguments to the ones used above gives the following result.

Proposition 4.7 *Suppose that demand and capacity are scaled according to* $\lambda^k(\cdot) = k\lambda(\cdot)$ *and* $C^k = kC$*, and consider the LPCC heuristic defined through (4.28). Then,* $\bar{X}_b^k(t) \to C - \bar{\rho}(t)$ *a.s., uniformly in* $[0, T]$*,* $\frac{1}{k}R_b^k \to \sum_{i=1}^n \bar{p}_i \bar{\lambda}_i T$ *a.s.*

That is, resolving and using either the dynamic pricing or the LPCC heuristics is asymptotically optimal in the fluid sense, as was the static policy (a). We note that the last proposition is related in spirit to the one proved in Lin et.al. (2003) for their proposed policy. Propositions 4.6 and 4.7 established that the suboptimal behavior of the resolving idea demonstrated by the example of Cooper (2002) does not persist when one considers its performance in systems with large capacity and large demand. The same asymptotic performance would be obtained in a setting where the resolving occurs in discrete points in time, provided that this is done sufficiently frequently. If l^k is the time between resolving epochs, then the type of analysis used in studying the asymptotic behavior of discrete-review policies (see Harrison (1995b) and Maglaras (2000)) can be applied to establish that it suffices that $l^k \downarrow 0$. That is, the number of demand requests between resolving periods must be small compared to the capacity. A more refined analysis that involves a central-limit-theorem type of correction to R_a^k, R_b^k and R_c^k that is proportional to $\sqrt{k}$ would in fact show that policy (c) is better than (b), which is better than (a). These findings are illustrated numerically in the next section. Finally, we note that the feedback nature of the dynamic policies will make them more robust against model and/or parameter uncertainties, e.g., with respect to the functional form of the demand model or the size of the cross-price sensitivity parameters. While this issue is of significant practical and theoretical interest, its analysis is beyond the scope of this paper.

4.4.3 Dynamic Pricing Network Revenue Management Problems

Suppose that the firm is operating a network of resources, indexed by $j = 1, \ldots, m$, and that each product i request consumes A_{ij} units of resource j capacity. Let $A := [A_{ij}]$ denote the associated capacity consumption matrix, and assume that the initial capacity for each resource j is C_j. Then, the fluid model formulation of the network dynamic pricing problem is:

$$\max_{\{\lambda(t),\ t\in[0,T]\}} \left\{ \int_0^T R(\lambda(t))dt \ : \ \int_0^T A\lambda(t)dt \le C \text{ and } \lambda(t) \in \mathcal{L}\ \forall t \right\}. \tag{4.29}$$

As before, this problem can be expressed in terms of ρ which is defined by $\rho := A\lambda$. Specifically, let

$$R^r(\rho) := \max_{\lambda} \ \{R(\lambda) \ : \ A\lambda = \rho,\ \lambda \in \mathcal{L}\}, \tag{4.30}$$

be the maximum achievable revenue rate when resource capacity is consumed at a rate ρ, and $\lambda^r(\rho)$ denote the corresponding vector of optimal demand rates. Then, (4.29) can be reduced to

$$\max_{\{\rho(t),\ t\in[0,T]\}} \left\{ \int_0^T R^r(\rho(t))dt \ : \ \int_0^T \rho(t)dt \le C \text{ and } \rho(t) \in \mathcal{R}\ \forall t \right\}. \tag{4.31}$$

Let $\bar{\rho}$ denote the solution to (4.31). Then, $\lambda^r(\bar{\rho})$ will be the vector of optimal demand rates for (4.29). This reduction can be computationally beneficial, since as is often the case the number of products (e.g., the number of fare-class and origin-destination pairs) tends to be greater than the number of resources (e.g., number of flights in a hub-and-spoke network). One can similarly address network capacity control problems using the results of the previous subsection. Overall, this structural decomposition seems a promising direction for future work towards the development of practical network revenue management algorithms. We refer the reader to GALLEGO and VAN RYZIN (1997) and KLEYWEGT (2001) for fluid formulations to multi-product network revenue management problems.

4.5 Numerical examples

This section reports on a set of numerical examples that contrast the performance of the heuristics proposed in the previous section to that of the optimal policy obtained from the dynamic program. We review a variety of settings that explore the effects of the joint capacity constraint and of cross-price sensitivities in multi-product revenue management.

The base model that we use has two products, each consuming one unit of capacity per request, and a linear demand relationship of the form $\lambda(p) = \Lambda - Bp$ with

$\Lambda = [.3, .1]$ and $T = 200$ time periods. Towards the end of the section we will study examples with three products and non-uniform capacity requirements. The price set is defined as $\mathcal{P} = \{p \,:\, \Lambda - Bp \geq 0\}$. Recall that the inverse demand and revenue functions are given by $p(\lambda) = B^{-1}(\Lambda - \lambda)$ and $R(\lambda) = \lambda' B^{-1}(\Lambda - \lambda)$, respectively. The specific policies that we consider are the following:

- 'Revmax' corresponds to the monopoly price vector $\bar{p}$ that maximizes the aggregate instantaneous revenue rate disregarding the capacity constraints, computed as follows:

$$\begin{aligned} \hat{\lambda} &= \operatorname{argmax}\ \{\lambda' B^{-1}(\Lambda - \lambda) \,:\, \lambda \geq 0\} = (1/2)(B^{-1} + B^{-1'})^{-1}\Lambda \quad \text{and} \\ \hat{p} &= B^{-1}(\Lambda - \hat{\lambda}). \end{aligned}$$

- 'Fluid' implemented the price vector $\bar{p} = p(\lambda^r(\bar{\rho}))$ as defined in (4.15) and (4.16).

For the linear demand model one can derive closed form expressions for the aggregate revenue function $R^r(\rho)$, the corresponding revenue maximizing demand vector $\lambda^r(\rho)$, and get a simple characterization of policies (a) and (b). We will illustrate this point for the special case where B is diagonal, i.e., $b_{ij} = 0$ for all $i \neq j$, in which case there are no cross-price sensitivity effects due to product substitution and/or complementarities. In this case, $B = \mathbf{diag}(b_{11}, \ldots, b_{II})$ and $B^{-1} = \mathbf{diag}(1/b_{11}, \ldots, 1/b_{II})$. The revenue function is $R(\lambda) = \sum_{i=1}^{n} \lambda_i(\Lambda_i - \lambda_i)/b_{ii}$ and $\partial R(\lambda)/\partial \lambda_i = (\Lambda_i - 2\lambda_i)/b_{ii}$. With no loss of generality we will assume that products are labelled such that $\Lambda_1/b_{11} \geq$

$\Lambda_2/b_{22} \geq \cdots \geq \Lambda_I/b_{II}$. The 'Revmax' policy is given by

$$\hat{\lambda}_i = \Lambda_i/2 \quad \text{and} \quad \hat{p}_i = \Lambda_i/2b_{ii}.$$

The 'Fluid' policy is defined as follows. For a sequence of constants $0 = \rho^{(0)} \leq \rho^{(1)} \leq \cdots \leq \rho^{(I)}$ (specified below) we set $\hat{\imath}(\rho)$ as a function of ρ to be $\hat{\imath}(\rho) = \min\{i \geq 0 : \rho^{(i)} \geq \rho\}$, and then

$$\lambda_j^r(\rho) = \begin{cases} \hat{\lambda}_j - \mu B_{jj} & j \leq \hat{\imath}(\rho) \\ 0 & j > \hat{\imath}(\rho) \end{cases} \quad \text{where} \quad \mu = \frac{\sum_{k \leq \hat{\imath}} \hat{\lambda}_k - \rho}{\sum_{k \leq \hat{\imath}} B_{kk}},$$

and $R^r(\rho) = \lambda^r(\rho)' B^{-1}(\Lambda - \lambda^r(\rho))$ which is a piecewise concave quadratic function in ρ. The expression for μ comes from the form of the Lagrange multiplier that takes into account the 'open' products. It remains to define the constants $\rho^{(i)}$, which is done recursively as follows:

$$\begin{aligned} \rho^{(1)} &= \min\left\{\rho \geq 0 \,:\, \rho = l_1,\ \frac{\Lambda_1 - 2l_1}{b_{11}} = \frac{\Lambda_2}{b_{22}}\right\}, \\ \rho^{(2)} &= \min\left\{\rho \geq 0 \,:\, \rho = l_1 + l_2,\ \frac{\Lambda_1 - 2l_1}{b_{11}} = \frac{\Lambda_2 - 2l_2}{b_{22}} = \frac{\Lambda_3}{b_{33}}\right\} \end{aligned}$$

and so on. A similar argument can be applied when the cross-price sensitivity parameters are non-zero and the capacity requirements are not all equal to 1.

- 'Decoupled-DP' is the following heuristic: each product manager calculated upfront a dynamic pricing strategy for his product by solving a single-item DP, disregarding cross-elasticity effects with the other product. At each point t in time, the remaining capacity is split according to the nominal split prescribed by the fluid solution, $\tilde{C}_i(t) := \lambda^r(\bar{p}) \cdot (T - t)$. The product managers then implement the pricing strategy according to the remaining time and their assigned inventory. This heuristic is one of many possible 'refinements' over the deterministic fluid solution, and is used for illustrative purposes.

- 'LPCC' is the joint (list) pricing and capacity control heuristic defined through (4.18).
- 'DynPrice' is the dynamic implementation of the fluid policy defined through (4.19).
- 'DP' implemented the solution of the dynamic program outlined in Section 3.1.

The expected revenue under the first two static pricing rules were computed analytically using a binomial model, while those under the next three policies were obtained by averaging out revenues from 1,000 simulated sample paths.

I. The effect of the joint capacity constraint. Table 1 studies a series of problems with increasing capacity. The price sensitivity matrix is $B = \mathbf{diag}(1, 6)$, which implies that there are no cross-price sensitivity effects ($b_{12} = b_{21} = 0$) to isolate the effect of the joint capacity constraint on the performance of the various heuristics.[1] The resulting demand model was

$$\lambda_1(p) = .3 - 1p_1 \quad \text{and} \quad \lambda_2(p) = .1 - 6p_2.$$

[1]The parameter values in this and other examples that follow were selected from a large pool of cases tested, and are representative in terms of the optimality gaps observed for the various policies. To demonstrate the effects of the joint capacity constraint and cross-price sensitivities, the relative magnitudes of Λ and B for the two products are such that in most cases it is economically optimal to offer both products.

Table 4.1: Results for $B = \mathbf{diag}(1,6)$. For each policy we report the expected revenue ($\times 1000$) and the optimality gap relative to the performance of the optimal policy (DP).

C	Revmax	Decoupled-DP	Fluid	LPCC	DynPrice	DP
15	171.87 (45.9%)	186.00 (41.4%)	304.25 (4.2%)	305.25 (3.9%)	305.40 (3.8%)	317.52
20	229.17 (39.3%)	236.64 (37.3%)	366.29 (3.0%)	367.34 (2.7%)	372.18 (1.5%)	377.67
25	286.41 (31.4%)	285.38 (31.7%)	404.96 (3.0%)	406.98 (2.6%)	416.05 (0.4%)	417.63
30	342.94 (22.2%)	324.83 (26.3%)	420.98 (4.4%)	426.04 (3.3%)	436.05 (1.0%)	440.53
35	394.66 (12.6%)	348.15 (22.9%)	428.65 (5.1%)	443.25 (1.9%)	445.76 (1.3%)	451.63
40	432.53 (5.4%)	367.89 (19.5%)	432.53 (5.4%)	454.76 (0.7%)	454.55 (0.7%)	457.00
45	451.40 (1.8%)	396.58 (13.7%)	451.40 (1.8%)	458.73 (0.2%)	458.63 (0.2%)	459.50
50	457.18 (0.7%)	414.24 (10.0%)	457.18 (0.7%)	459.16 (0.3%)	458.93 (0.3%)	460.39

The ‘Fluid’ pricing problem becomes unconstrained, i.e., the capacity constraint is not binding at its optimum, for $C \geq 40$ units, in which cases the ‘Revmax’ and ‘Fluid’ prices coincide. We observe the following. First, the relative performance under the ‘Fluid’ and ‘Revmax’ heuristics improves as the capacity C increases; this is consistent with the results by Gallego and van Ryzin (1994, 1997) that also illustrated that the effect of the capacity constraint is more pronounced when C is scarce. Second, for the cases where the capacity is scarce, the ‘Fluid’ heuristic that incorporates the capacity constraint significantly outperforms ‘Revmax,’ but its regret over the ‘DP’ policy can still be substantial (0.7%-5.4%). Third, in comparing ‘Fluid’ with ‘LPCC’ we note that when the capacity is small $C \leq 20$ the fluid prices effectively switch off product 2 and operates as a single-product system, where the two rules are almost identical. As the capacity increases, it is optimal to offer both products and the effect of the capacity control capability of LPCC becomes more evident. Switching from an effectively single-product to a two-product solution also causes the optimality gaps to be non-monotone. Fourth, the dynamic pricing heuristic that effectively resolves the fluid policy at each time point is significantly better than all these heuristics when the caapcity is scarce. Finally, the 'Decoupled-DP' policy performs very poorly even when the firm has ample capacity, mainly because it disregards the pooling effects across products.

II. The cross-price elasticity effects. The next two tables study how cross-price sensitivity effects impact the performance of these heuristics by gradually increasing the interaction terms b_{12} and b_{21}. Table 2 used $b_{12} = -.4$ and $b_{21} = -.6$, which corresponds to the demand model $\lambda_1(p) = .3 - 1p_1 + .4p_2$ and $\lambda_2(p) = .1 - 6p_2 + .6p_1$, while Table 3 had $b_{12} = -0.8$ and $b_{21} = -1.2$. Again, the 'Decoupled-DP' heuristic performs very poorly when compared to all other candidates, and its performance deteriorates substantially as the magnitude of the cross-price interaction terms increases. The 'Fluid' and 'LPCC' policies that incorporate the interaction effects in their static prices perform consistently well, but again the magnitude of their respective optimality gaps tends to increase as the interaction coefficients increase. Based on a wide range of examples ran with random interaction terms b_{ij} while keeping b_{ii} constant, we found that the 'LPCC' heuristic outperforms the 'Fluid' policy by about .75% to 2% in cases where the capacity is sufficiently large so as to want to offer both products. The dynamic heuristic adds an other 1% of improvement.

Table 4.2: Expected revenues and optimality gaps when $B = [1 \ \ -.4; -.6 \ \ 6]$.

C	Revmax	Decoupled-DP	Fluid	LPCC	DynPrice	DP
20	235.77 (42.0%)	121.35 (70.2%)	394.27 (3.1%)	396.13 (3.6%)	399.21 (1.9%)	406.81
30	353.42 (26.1%)	199.95 (58.2%)	457.89 (4.2%)	464.66 (2.8%)	472.74 (1.1%)	478.11
40	457.84 (8.4%)	262.36 (47.5%)	475.35 (4.9%)	492.57 (1.5%)	496.94 (0.6%)	500.07
50	500.63 (1.1%)	342.00 (32.4%)	500.63 (1.1%)	505.57 (0.1%)	506.05 (0.0%)	506.17

Table 4.3: Expected revenues and optimality gaps when $B = [1 \ \ -.8; -1.2 \ \ 6]$.

C	Revmax	Decoupled-DP	Fluid	LPCC	DynPrice	DP
20	265.22 (44.8%)	75.03 (84.4%)	465.13 (3.2%)	465.19 (3.1%)	472.12 (1.7%)	480.29
30	397.77 (30.1%)	133.86 (76.5%)	545.46 (4.1%)	551.33 (3.1%)	562.73 (1.0%)	568.68
40	524.30 (12.6%)	204.75 (65.9%)	572.29 (4.6%)	589.80 (1.7%)	597.80 (0.4%)	599.90
50	597.98 (2.1%)	260.79 (57.3%)	597.98 (2.1%)	606.49 (0.7%)	606.70 (0.7%)	610.99

III. Mulitple products. Table 4 reports results for a model with three products, and provides a brief illustration of the consistently good performance of the LPCC and DynPrice policies.

Table 4.4: A three-product example: $\Lambda = [.2 \ \ .1 \ \ .1]$ and

$B = [1 \ \ -.8 \ \ -.4; -1.2 \ \ 3 \ \ -.6; -.3 \ \ -.6 \ \ 4]$.

C	Revmax	Decoupled-DP	Fluid	LPCC	DynPrice	DP
10	132.08 (56.0%)	12.81 (95.7%)	237.65 (20.8%)	251.98 (16.1%)	284.04 (5.4%)	300.16
20	264.15 (42.3%)	38.13 (91.7%)	436.29 (4.7%)	444.12 (3.0%)	451.15 (1.4%)	457.63
30	395.99 (25.4%)	92.95 (82.5%)	507.77 (4.3%)	511.24 (3.6%)	525.38 (1.0%)	530.51
40	513.68 (8.4%)	157.74 (71.9%)	534.99 (4.6%)	543.67 (3.0%)	556.64 (0.7%)	560.65
50	563.05 (1.1%)	199.91 (64.9%)	561.52 (1.4%)	563.82 (1.0%)	566.42 (0.5%)	569.35

IV. Non-uniform capacity requirements The next two tables summarize results for a model with two products that have different capacity requirements. This change complicates the associated dynamic programming formulation, in the sense that the structural result of Section 4.3.1 that allows us to consider the problem in terms of the aggregate capacity consumption rate no longer holds. Numerically, this affects the backwards induction step required to solve the problem. In contrast, the fluid analysis of Section 4.4.1 and the heuristics extracted therein are still valid. In the notation of Section 4.4, we will assume that product i consumes a_i units of capacity and $a_1 \neq a_2$. The results of the next two tables suggest that the fluid model heuristics perform quite well in cases where the capacity requirements are small compared to the capacity itself. This effect is more pronounced in Table 6 because in this example the product that requires more capacity per request (product 1) happens to be the more profitable product as well, hence making the overall capacity small compared to a_1 in the first two or three rows.

Table 4.5: Expected revenues and optimality gaps when $B = [1\ -.4; -.6\ \ 6]$ and $a = [1\ \ 2]$.

C	Revmax	Fluid	LPCC	DynPrice	DP
10	92.61 (63.0%)	234.09 (6.6%)	235.65 (5.9%)	233.64 (6.7%)	250.54
20	181.29 (55.3%)	394.35 (2.8%)	394.92 (2.7%)	399.77 (1.5%)	405.75
30	273.21 (42.7%)	456.32 (4.3%)	459.51 (3.7%)	464.22 (2.7%)	476.91
40	361.53 (27.1%)	468.52 (5.6%)	485.66 (2.1%)	488.78 (1.5%)	496.05
50	442.54 (12.0%)	473.95 (5.7%)	494.47 (1.7%)	497.77 (1.0%)	502.77
60	491.22 (2.9%)	489.40 (3.2%)	500.14 (1.1%)	504.46 (0.3%)	505.81
70	502.87 (0.8%)	503.02 (0.7%)	504.94 (0.4%)	505.07 (0.3%)	506.73
80	505.12 (0.3%)	505.74 (0.2%)	506.10 (0.2%)	506.33 (0.1%)	506.86

Table 4.6: Expected revenues and optimality gaps when

$B = [1\ \ -.4; -.6\ \ 6]$ and $a = [2\ \ 1]$.

C	Revmax	Fluid	LPCC	DynPrice	DP
10	66.90 (51.8%)	120.96 (12.9%)	121.81 (12.3%)	120.51 (13.3%)	138.92
20	136.64 (45.8%)	235.07 (6.7%)	235.57 (6.5%)	234.83 (6.8%)	252.04
30	206.30 (39.5%)	325.97 (4.4%)	327.28 (4.0%)	327.54 (3.9%)	340.95
40	276.01 (32.2%)	393.26 (3.3%)	394.49 (3.0%)	399.77 (1.7%)	406.89
50	345.91 (23.4%)	438.37 (3.0%)	438.61 (2.9%)	449.08 (0.6%)	451.85
60	411.77 (14.2%)	461.79 (3.7%)	463.35 (3.4%)	473.84 (1.2%)	479.64
70	465.83 (6.0%)	473.78 (4.4%)	477.88 (3.6%)	491.75 (0.8%)	495.78
80	494.94 (1.7%)	493.53 (2.0%)	495.15 (1.7%)	500.27 (0.7%)	503.60
90	503.42 (0.6%)	502.08 (0.8%)	505.33 (0.2%)	505.85 (0.1%)	506.26

V. An analytical example: the Multinomial Logit (MNL) model We complete this section by illustrating that for the commonly used MNL choice model the subproblems of computing the revenue function and demand rates as functions of the aggregate demand rate are solvable in closed form. In the MNL model we assume that potential customers arrive to the firm with utilities for each product i given by $u_i + \xi_i$, where u_i is the deterministic portion that is common to all customers and ξ_i is the random term (that differentiates customers). The MNL choice model hinges on the assumption that these random terms ξ_i are random variables drawn from a Gumbel distribution with mean zero and parameter one (the latter is assumed w.l.o.g.), which as IID across products and customers. The deterministic component of the utility can be written as $u_i = v_i - p_i$, where v_i denotes the 'average value' of the product and p_i is its price. Finally, we denote by $u_0 + \xi_0$ the utility of the no-purchase option, where ξ_0 is IID with the ξ_i's and w.l.o.g. we will assume that $u_0 = 0$. Finally, let Λ be the aggregate customer arrival rate or market size. Under these assumptions the demand rates for product i is given by

$$\lambda_i(p) = \Lambda \frac{e^{v_i - p_i}}{1 + \sum_j e^{v_j - p_j}}.$$

Suppose further that the cost of a unit of product i to the firm is given by $c_i > 0$. Then, the profit rate function as a function of the price vector is given by $\sum_{i=1}^n \lambda_i(p)(p_i - c_i)$. With some abuse of notation define the aggregate profit rate function $R^r(\rho)$ to incorporate the product costs c_i as follows

$$R^r(\rho) := \max \left\{ \sum_{i=1}^n \lambda_i(p)(p_i - c_i) \;:\; \sum_{i=1}^n \lambda_i(p) = \rho,\ \lambda(p) \geq 0 \right\}.$$

Then, simple algebraic manipulations give that

$$R^r(\rho) = \rho \ln(\sum_j e^{v_j - c_j}) - \rho \ln(\rho/(\Lambda - \rho)) \tag{4.32}$$

and

$$\lambda_i(\rho) = \rho \frac{e^{v_i - c_i}}{\sum_j e^{v_j - c_j}} \quad \text{and} \quad p_i(\rho) = c_i + \ln(\textstyle\sum_j e^{v_j}) - \ln(\rho/(\Lambda - \rho)). \tag{4.33}$$

This calculation offers a nice insight about the structure of the optimal pricing strategy, namely that each product is priced at its cost plus a common premium that depends on the aggregate capacity consumption rate. From a computational viewpoint, one can now use the expression for $R^r(\rho)$ in the dynamic program of the form discussed in Subsection 4.3.1 to efficiently solve for the optimal capacity consumption rate $\rho^*(x,t)$ and then use (4.33) to translate this into product-level controls.

Chapter 5

Price Competition under Time-Varying Demands and Dynamic Lot Sizing Costs

5.1 Introduction

Determining the 'right' price to charge for a product is a complex task. A voluminous literature in economics and marketing has been devoted to models which prescribe how prices should be set in industries in which a limited number of competing firms offer similar products, which may therefore be viewed as substitutes. These papers typically model the interaction among competitors as a noncooperative game, see VIVES (2000) and TIROLE (1988) for survey texts.

More recently, operations management papers have demonstrated that the operational environment and associated cost structures may have a fundamental impact on the equilibrium behavior in the industry, in general, and the resulting price levels in particular. See CACHON (2003) for a recent survey. Little remains known, however, about how prices should be set in a competitive environment, in the *simultaneous* presence of two other major complications:

(i) time dependent demand functions and cost parameters, and

(ii) scale economies in the operational costs

In contrast, progress has been made in addressing either one of these factors by itself.

For example, BERNSTEIN and FEDERGRUEN (2003) address a setting where each firm incurs fixed as well as variable procurement costs along with (linear) inventory carrying costs. However, the model assumes an infinite horizon setting with *time-invariant* demand functions and cost parameters. Here the long-run average operational costs are given by the simple closed-form Economic Order Quantity (EOQ) cost function, i.e. the costs are given by the sum of a term that is proportional with the demand value itself and one that is proportional with the *square root* of the demand value, thus reflecting scale economies. CACHON and HARKER (2002) similarly consider, for an industry with two firms, a setting with a SINGLE set of *time-invariant* demand functions and with a closed form cost function given by a *concave* power function of the demand volume, (possibly in conjunction with a *linear* cost component), once again to reflect scale economies. Other than the EOQ-cost model above, the authors show that their cost structure arises in a specific service competition model.

At the same time, a stream of marketing papers address competitive pricing problems under time-dependent demand functions, however with simple linear cost functions, and under the assumption that each period's demand is procured in the same period, i.e. no inventories are carried. Theses paper include KALISH (1983), ELIASHBERG and JEULAND (1986), CLARKE and DOLAN (1984), RAO and BASS (1985), and DOCKNER and JORGENSEN (1988). See ELMAGHRABY and KESKINOCAK (2003) for a survey. PERAKIS and SOOD (2003, 2004) and KACHANI, PERAKIS and SIMON (2004) also address competitive pricing problems under time-varying demand functions. Since each firm starts the planning horizon with a known inventory and inventories can not be replenished at any time during the horizon, these models consider *no* replenishment costs.

This paper appears to develop the first competitive pricing model which *combines* the complexity of *time-varying* demand and cost functions and that of scale economies arising from dynamic lot sizing costs. Each firm can replenish inventory in each of the T periods into which the planing horizon is partitioned. Fixed as well as variable procurement costs are incurred for each procurement order, along with inventory carrying costs. Each firm adopts, at the beginning of the planning horizon, a (single) price to be employed throughout the horizon. Based on each period's system of demand equations, these prices determine a time series of demands for each firm, which needs to service them with an optimal corresponding dynamic lot sizing plan. Scale economies always create major analytical complications in the study of price equilibria, see e.g. VIVES (2000) and CACHON and HARKER (2002). In our case, the problems are compounded by the fact that the cost structure can not be represented as a closed form analytical function of the (time series of) demand volume(s).

We model each firm's time dependent demand function as an affine transformation of a basic *time-invariant* demand function of general structure, i.e. the time invariant demand function is multiplied by a firm- and period-dependent 'seasonality' *factor* and shifted by a (firm- and period dependent) 'seasonality' *term.* This general framework includes the special cases of *multiplicative* seasonalities and *additive* seasonalities.

We establish the existence of a price equilibrium and associated optimal dynamic lotsizing plans, under mild conditions and employing a close approximation for the optimal lotsizing costs. We also design efficient procedures to compute the equilibrium prices and dynamic lotsizing plans. Finally, we characterize how the equilibrium is affected by changes in various cost and demand function parameters. Much of the analysis focuses on an individual firm's *best response* problem, i.e. the characterization of the optimal price and lot sizing strategy, in response to a given set of prices adopted by the firm's competitors. This best response problem is of interest in its own right. For the case of *multiplicative* seasonalities and *constant* setup costs, we design an efficient algorithm whose computational effort involves $O(T^2)$ elementary operations and $O(T)$ maximizations of a single variable concave function. (For many classes of demand functions, these maximizations can be performed in closed form). For the general model, we develop an alternative solution method whose efficiency is demonstrated via a numerical study. The algorithms for the best response problem are used repeatedly in the procedure which computes the overall price and lotsizing equilibrium in the industry.

The numerical study, in addition, reveals the following insights: contrary to folklore, it is not always best for a firm to operate under time-invariant demand functions or cost parameters. Even in equilibrium, all firms in the industry may be better off under certain types of seasonality patterns in the demand or cost parameters. When the above mentioned mild existence conditions for a (unique) equilibrium fail to hold, the industry may switch from having a unique equilibrium to either having none or having multiple equilibria, depending on which seasonality pattern prevails.

One of the most fundamental assumptions in our model is that each firm maintains a constant price throughout the sales season. In many industries this is the practice, either because mid-season price changes can not be implemented (e.g. catalog sales) or because they are managerially undesirable. For example, in many retail operations it is considered unacceptable to *raise* prices during the season and at most one or two markdowns in a single season are as much as contemplated. Empirical analysis has documented that prices for certain goods are extremely ‘sticky’, i.e. they change very slowly over time, if at all. Carlton (1986), for example, has investigated price data for industrial buyers over a 10 year period and has concluded that the price rigidity in many industries is striking. Kashyap (1995) analyzed the data from catalog sales and observed that the prices typically remain constant over *several* seasons, beyond the total of a year. As a third example, Cecchetti (1986) has studied newsstand prices for some 40 American magazines over a period of close to 20 years. This author concluded that prices exhibited remarkable rigidity to the extent of dropping in *real* (i.e. inflation corrected) value by as much as a quarter before a price adjustment is implemented. See Barro (1972), Sheshinski and Weiss (1977, 1983), and Caplin and Leahy (1991) for theoretical models explaining price rigidity. Finally, Blinder et al (1998) document and explain the multitude of reasons why companies maintain ‘sticky’ prices.

(As demonstrated below, at least the best response problems are considerably easier under the alternative assumption where prices *may* be chosen and varied independently in each period.)

The above not withstanding, we hope that our work will be extended to allow for a limited or an arbitrary number of price changes.

Beyond the competition models mentioned above, we now give a brief review of other literature that is relevant to our work. Even within the context of *optimization* models with a *single* decision maker, the integration of pricing and inventory strategies has only recently received the attention it deserves, even though the seminal papers on dynamic lot sizing by WAGNER and WHITIN (1958, 1959) and WAGNER (1960), 45 years ago, already addressed the need to integrate pricing and production planing decisions. We refer to ELIASHBERG and STEINBERG (1993) for a survey of early work and to ELMAGHRABY and KESKINOCAK (2003) for a more recent survey.

THOMAS (1970) addressed the dynamic lot sizing problem in which each period's demand depends (exclusively) on the price charged during this period according to a period-specific demand function. Assuming prices can be changed arbitrarily from one period to the next and that demand in a given period is independent of the prices offered in other periods, the author shows that an optimal plan can be found by a simple extension of the classical shortest path method by WAGNER and WHITIN (1958).

As in the classical dynamic lot sizing problem with exogenously specified demands, it is easily verified that replenishment orders should only be placed in periods with *zero starting inventory*. In view of this observation, the optimal procurement plan can thus again be described as the shortest path in a network in which each period is represented as a node and traversing an arc from period i to period j corresponds with the decision to satisfy all demands in period $i, i+1, \dots, j-1$ from a single order delivered in period i. The only difference with the classical WAGNER-WHITIN shortest path procedure is that evaluation of the (optimal) cost or profit value of any arc (i, j) now involves optimizing a closed form expression over the prices in periods $i, i+1, \dots, j-1$. Thus, dynamic lot sizing problems with arbitrary time-dependent prices are radically simpler than under the requirement that a single constant price be used.

The best response problem which arises within our competition model was first addressed by KUNREUTHER and SCHRAGE (1973). These authors propose a heuristic procedure which also generates an upper and lower bound for the optimal price. GILBERT (1999) considers the special case of the best response problem in which only *multiplicative* seasonalities prevail and *all* cost parameters remain constant over the entire planing horizon. One of our procedures for the best response problem is based on GILBERT's approach but reduces the computational complexity by an order of magnitude ($O(T^2)$ compared to $O(T^3)$ complexity), while addressing more general parameter settings. This complexity reduction is particularly important in our competition model where best response problems need to be solved *repeatedly* and for each of the N firms in the industry.

The remainder of the Chapter is organized as follows: In Section 5.2 we specify the general competition model and the notation. Section 5.3 is devoted to the best response problem which arises under multiplicative seasonalities and constant set-up costs. Section 5.4 addresses the fully general response problem. The equilibrium analysis for the competitive model is carried out in Section 5.5. Section 5.6 completes the paper with a numerical study.

5.2 Model and Notation

We consider an industry with N firms, each selling a distinct item or product brand. We refer to the item sold by firm i as item $i = 1, \ldots, N$. The different items are (close) substitutes of each other, e.g. different brands of 19" television sets, digital cameras, notebook computers, sport utility vehicles, tooth paste etc. We consider a planning horizon of T periods. If the firms face a natural sales season introducing a new model or variant in each season, a natural choice of T arises, e.g. $T = 52$ weeks in the automobile manufacturing industry operating with a weekly production and sales schedule. Otherwise T is chosen large enough to ensure that the firms' decisions pertaining to the initial periods of the planning horizon are not affected by this truncation of the planning process.

Each firm selects a (single) price to be used though out the planning horizon. (See Section 1 for a discussion of this assumption.) Each firm's demand in each period depends potentially on the complete vector of prices selected in the industry according to a general *time-dependent* demand function. Thus, let

$$
\begin{aligned}
p^i &= \text{the price charged by firm } i = 1, \ldots, N \\
d_t^i &= \text{the demand faced by firm } i \text{ in period } t = 1, \ldots, T \\
&= d_t^i(p^1, \ldots, p^N)
\end{aligned}
$$

The time-dependence of the demand function is characterized by *additive* seasonalities as well as *multiplicative* seasonality factors, i.e.

$$d_t^i(p) = \alpha_t^i + \beta_t^i \delta^i(p) \quad \forall \quad i = 1, \ldots, N; \; t = 1, \ldots, T \tag{5.1}$$

with $\delta^i(p)$ the *deseasonalized* demand function for firm i, $i = 1, \ldots, N$. The general specification in (5.1) contains two important special cases: (a) a purely *additive* seasonality structure arises when all $\beta_t^i = 1$. In this case, the demand function of a given firm undergoes parallel shifts as we move from one period to the next; (b) a purely *multiplicative* structure arises when all $\alpha_t^i = 0$; here, each firm's demand is scaled up ($\beta_t^i > 1$) or down ($\beta_t^i < 1$) compared to the deseasonalized norm.

The deseasonalized functions $\delta^i(p)$ are continuously differentiable, with $\frac{\partial \delta^i(p)}{\partial p^i} < 0$, i.e. a price *increase* results in in a *decrease* of the demand volume. Since the demand function $\delta^i(p)$ is strictly decreasing in p^i, it is possible to derive an *inverse* demand function $p^i = \phi(\delta^i | p^{-i})$. We only assume that firm i's revenue is *concave* in the demand volume δ^i, i.e. $R^i(\delta^i | p^{-i}) = \delta^i \phi(\delta^i | p^{-i})$ is *concave* in δ^i. An important special case arises when all deseasonalized functions are *linear*:

$$\delta^i(p) = a^i - b^i p_i + \sum_{j \neq i} \theta_j^i p^j \quad i = 1, \ldots, N, \tag{5.2}$$

Without loss of generality, we assume that

$$b^i > \sum_{j \neq i} \theta_j^i \quad , \quad i = 1, ..., N. \tag{5.3}$$

This standard assumption is often referred to as the Dominant Diagonal assumption.

It merely precludes completely unrealistic situations where an across the board price increase in the industry results in an increase of a firm's sales volume. The firms procure their goods by a production and distribution process which, in principle, allows for inventory replenishments at the beginning of each period. As in standard dynamic lot sizing problems, we assume that *fixed* as well as *variable* procurement costs are incurred as well as inventory carrying costs which are proportional to each end-of-the-period inventory. All cost parameters may fluctuate over the course of the planning horizon in arbitrary ways. Thus let

K_t^i = the *fixed* setup cost for a procurement batch delivered to firm i in period t, $i = 1, \ldots, N$; $t = 1, \ldots, T$

c_t^i = the per unit procurement cost rate for a procurement batch delivered to firm i in period t, $i = 1, \ldots, N$; $t = 1, \ldots, T$

h_t^i = the inventory carrying cost for each unit of item i carried in inventory at the end of period t, $i = 1, \ldots, N$; $t = 1, \ldots, T$

Each firm i thus selects a price p^i as well as a complete procurement schedule for the entire planning horizon to support the demand stream $\left\{d_t^i(p)\right\}$ which arises from the collective price choices. Note that his price p^i affects the profits earned by *all* firms in the industry via its impact on each firm's demand function and hence each firm's demand stream. At the same time, the procurement schedule selected by firm i impacts only his *own* profit measure. It is thus possible to conceptualize the competition model as a single stage game between N firms, in which each firm makes a single competitive choice, i.e. its price level. The competition model may thus be viewed as an example of

Bertrand price competition. (Alternatively, one may assume that each firm i selects a *basic deseasonalized* target volume δ^i. This results in a vector of prices p which satisfy the demand equations $\delta^i = \delta^i(p)$, $i = 1, \dots, N$. We discuss this Cournot competition variant at the end of Section 5.5). The game is characterized by the profit functions:

$\pi^i(p)$ = the profit earned by firm i under the price-vector p, assuming firm i adopts an optimal dynamic lot sizing schedule to respond to the demand stream $\left\{d_t^i(p) : t = 1, \dots, T\right\}$

The profit function may be written in the form:

$$\pi^i(p) = p^i \sum_{t=1}^{T} d_t^i(p) - C^i(p) \quad i = 1, \dots, N \tag{5.4}$$

where

$C^i(p)$ = the minimum total operating costs for firm i to service the demand stream $\{d_t^i(p)\}$. The difficulty in analyzing the competition model and in characterizing its equilibrium behavior, stems from the complexity of the cost functions $C^i(p)$. Clearly, the function $C^i(p)$ cannot be represented as an analytical closed form expression. The function can be *evaluated* for any given price vector p in $O(T^2)$ time using the standard Wagner-Whitin shortest path procedure and in $O(T \log T)$ time by one of the methods in Aggarwal and Park (1993), Federgruen and Tzur (1991), or Van Hoesel and Wagelmans (1992). Clearly $C^i(p)$ depends on the price vector p 'only' through the demand sequence $\left\{d_t^i(p) = d_t^i\right\}$.

Lemma 5.1 *C^i is a* piecewise linear concave *function of the demand sequence*

$\left\{d_t^i = d_t^i(p)\right\}$.

Proof: Fix a price vector $p \in R^N$. Assume firm i chooses to replenish its inventory in the set of periods $\Theta = \{t_1 = 1, t_2, \ldots, t_n\}$. It is well known that a zero-inventory ordering policy is optimal. This implies that the optimal cost under this sequence of order periods is given by:

$$\begin{aligned} C^i(p|\Theta) &= \sum_{l=1}^{n} K_{t_l}^i + \sum_{l=1}^{n} c_{t_l}^i \sum_{r=t_l}^{t_{l+1}-1} d_r^i(p) \\ &+ \sum_{l=1}^{n} \sum_{r=t_l}^{t_{l+1}-1} h_r^i(d_{r+1}^i(p) + \cdots d_{t_{l+1}-1}^i(p)) \end{aligned} \tag{5.5}$$

where $t_{n+1} = T + 1$. Finally, $C^i(p) = \min\left\{C^i(p|\Theta)|\Theta \in 2^{\{1,\ldots,T\}}\right\}$, i.e. $C^i(p)$ is the minimum of $(2^T - 1)$ linear functions and is therefore concave in the vector

$\left\{d_t^i : t = 1, \ldots, T\right\}$. ■

As a corollary of the above Lemma we obtain that, if all demand functions are *linear*, the cost functions $C^i(p)$ are piecewise linear and jointly concave in the price vector p, itself. Unfortunately, this characterization is by itself insufficient to conduct the equilibrium analyses. First, *concave* cost functions reflecting *economies* of *scale* create major analytical difficulties in equilibrium analysis, see VIVES (2000) and CACHON and HARKER (2002). Second, the fact that $C^i(p)$ is not available in closed form generates additional complexities. Finally, $C^i(p)$ fails to be jointly concave in p for more general non-linear demand functions.

Before providing a complete characterization of the industry's equilibrium behavior, we first analyze an individual firm's *best response problem*. Thus, for any $i = 1, \ldots, N$, let

$\pi^i(p^i|p^{-i}) =$ the profit for firm i when choosing price p^i, given his competitors' prices $p^{-i} = (p^1, \ldots, p^{i-1}, p^{i+1}, \ldots, p^N)$ and assuming firm i minimizes operating costs to service the resulting demand stream $\left\{d^i(p)\right\}$.

The *optimal* profit for firm i, given his competitors choose prices p^{-i} can then be written as:

$$\pi^{*i}(p^{-i}) = \max_{p^i} \pi^i(p^i|p^{-i}) \tag{5.6}$$

We refer to the optimization problem in (5.6) as firm i's *best response problem*.

5.3 The best response problem under multiplicative seasonalities and constant setup costs

In this Section, we analyze a given firm i's best response problem (5.6) in the important special case where only multiplicative seasonality factors prevail ($\alpha_t^i = 0$) and where firm i's setup cost remains constant over the course of the planning horizon $\left(K_i^i = \cdots = K_T^i \stackrel{def}{=} K^i\right)$.

Note first that under the above special structure, for any sequence of order periods $\Theta = \{t, \ldots, t_n\}$ with n setup periods (see (5.5)):

$$C^i(p|\Theta) = nK^i + \delta^i(p) \left\{ \sum_{l=1}^{n} c_{t_l}^i \sum_{r=t_l}^{t_{l+1}-1} \beta_r^i + \sum_{l=1}^{n} \sum_{r=t_l}^{t_{l+1}-1} h_r^i \left(\beta_{r+1}^i + \cdots + \beta_{t_{l+1}-1}^i\right) \right\} \tag{5.7}$$

so that

$$C^i(p) = \min_{\Theta \in 2^{\{1,\ldots,T\}}} C^i(p|\Theta) = \min_n \min_{\Theta|n} C^i(p|\Theta) = \min_n \left\{ nK^i + \delta^i(p) F_n^i(T) \right\} \qquad (5.8)$$

where

$F_n^i(t)$ = minimum total variable procurement and holding costs in periods $\{1, \ldots, t\}$ for firm i, assuming the firm's demand stream is given by the seasonality factors $\left\{\beta_1^i, \ldots, \beta_t^i\right\}$ and assuming that *exactly* n setups are performed in the first t periods $t = 1, \ldots, T; \quad n = 1, \ldots, t,\ i = 1, \ldots, N$.

In the next subsection we develop an $O(T^2)$ procedure to compute the matrix of values $\left\{F_n^i(t); t = 1, \ldots, T; n = 1, \ldots, T\right\}$. Thus, assuming these values have been determined, it follows from (2) that the best response problem (5.6) reduces to:

$$\begin{aligned}
\pi^{*i}(p^{-i}) &= \max_{p^i} \max_n \left\{ \left(\sum_{t=1}^T \beta_t^i \right) p^i \delta^i(p) - nK^i - \delta^i(p) F_n^i(T) \right\} \\
&= \max_n \max_{p^i} \left\{ \left[\left(\sum_{t=1}^T \beta_t^i \right) p^i - F_n^i(T) \right] \delta^i(p^i|p^{-i}) - nK^i \right\} \qquad (5.9) \\
&= \max_n \max_{\delta^i} \left\{ \left(\sum_{t=1}^T \beta_t^i \right) R^i(\delta^i|p^{-i}) - F_n^i(T)\delta^i - nK^i \right\}
\end{aligned}$$

In other words, the best response problem reduces to T maximization problems of a *concave* function of a *single* variable δ^i. If the revenue function $R^i(\delta^i|p^{-i})$ is *strictly* concave, its derivative $R^{i'}(\cdot|p^{-i})$ is *strictly* decreasing and has an *inverse* function $\left(R^{i'}\right)^{-1}(\cdot|p^{-i})$. The maximizing value of δ^i in (5.9) is then given by $\delta^{*i}(n) = \left(R^{i'}\right)^{-1}\left(\frac{F_n^i(T)}{\sum_{t=1}^T \beta_t^i}|p^{-i}\right)$. Substituting this demand value into (5.9), we conclude that the solution of the best response problem reduces to the determination of T closed form expressions (for $n = 1, \ldots, T$) involving the inverse derivative revenue functions $\left(R^{i'}\right)^{-1}(\cdot|p^{-i})$. Assuming this function can be evaluated in constant time, the best response problem can thus be solved in $O(T^2)$ time.

Example 5.1: Assume the deseasonalized demand functions $\delta^i(p)$ are linear, as in (5.5). This implies that at $p^i = \left[a^i + \sum_{j\neq i}\theta^i_j p^j - \delta^i\right]/b^i$, giving rise to a *quadratic* revenue function $R^i(\cdot|p^{-i})$ and an optimizing deaseasonalized demand volume

$$\delta^{*i}(n) = \left[\left(\frac{a^i + \sum_{j\neq i}\theta^i_j p^j}{2}\right) - \frac{b^i F^i_n(T)}{2\sum_{t=1}^T \beta^i_t}\right]^+ \tag{5.10}$$

and a corresponding price

$$p^{*i} = \left[a^i + \sum_{j\neq i}\theta^i_j p^j - \delta^{*i}\right]/b^i \tag{5.11}$$

Substituting this value into (5.8), we obtain:

$$\pi^{*i}(p^{-i}) = \max_{n=1,\ldots,T}\begin{cases}\left[\frac{\sqrt{\sum_t \beta^i_t}}{2\sqrt{b^i}}\left(a^i + \sum_{j\neq i}\theta^i_j p^j\right) - \frac{\sqrt{b^i}}{2\sqrt{\sum_t \beta^i_t}}F^i_n(T)\right]^2 - nK^i & \text{, if} \quad \delta^{*i}(n) > 0 \\ -nK^i & \text{, otherwise}\end{cases} \tag{5.12}$$

which provides a closed form expression for the best response profit function in terms of the values $\left\{F^i_1(T), F^i_2(T), \ldots, F^i_T(T)\right\}$. ■

An $O(T^2)$ procedure to compute the optimal lot sizing cost when restricted to a given number of order periods

In this Subsection, we develop an $O(T^2)$-procedure to compute the values $\left\{F_n^i(T) : n = 1, \ldots, T\right\}$ we need in the evaluation of (5.9). As mentioned in the introduction, this procedure bears similarities to the $O(T^3)$-procedure in GILBERT (1999) for the case where *all* cost parameters are assumed to be *constant* over the course of the planning horizon. (GILBERT addresses a *single* firm problem.) The procedure, in fact, computes *all* entries of the matrix $F \stackrel{def}{=} \left\{F_n^i(t) : n = 1, \ldots, T; t = 1, \ldots, T\right\}$, row by row, starting with the values $\left\{F_1^i(t) : t = 1, \ldots, T\right\}$ in the *first* row. To simplify the expressions we drop the superscript in this subsection. We first need the following notation:

$B(t) \quad = \quad \sum_{k=1}^{t} \beta_k$ = the cumulative demand factors over the first t periods

$c_{kl} \quad = \quad c_k + h_k + \cdots + h_{l-1}$ = the variable cost of procuring a unit in period k and maintaining it in inventory until period $l (k < l)$

$= c_{k,l-1} + h_{l-1}$

$H(t) \quad = \quad \sum_{k=1}^{t} h_k$ = the cost of carrying a unit of in stock from periods 1 until the beginning of period $t + 1$

$= H(t-1) + h_t$

$S(k,t) \quad = \quad$ the total inventory carrying cost in periods $k, k+1, \ldots, t$ when placing an order in period k to meet the cumulative demand in periods $k, \ldots, t$.

$s(t) \quad = \quad S(1,t) = s(t-1) + \beta_t H(t-1).$

$$\tilde{C}(k) \quad = \quad c_{kT} - h_{1T} = c_k - H(k-1)$$

(Note for all $k < l$, $\tilde{C}(l) - \tilde{C}(k)$ represents the *additional* cost of procuring a unit in period l as opposed to procuring it in the earlier period k and keeping it in inventory until (at least) period l.)

Note that the *first* row of the matrix F, i.e. the values $\{F_1(t) : t = 1, \ldots, T\}$ are easily obtained in $O(T)$ time from the recursion:

$$F_1(1) = c_1\beta_1; \quad F_1(t) = F_1(t-1) + c_1\beta_t + \beta_t H(t-1), \; t = 2 \ldots, T \tag{5.13}$$

Assume therefore that the first $(n-1)$ rows of F have been calculated. We now show how the $n-$th row can be determined in $O(T)$-time by a 'list-based' procedure, similar to that in FEDERGRUEN and TZUR (1991) for the *unrestricted* lot sizing problem. (The procedure is, in fact, considerably simpler than the one in FEDERGRUEN and TZUR (1991), as we take advantage of the absence of setup costs). To this end, let for all $t = 1, \ldots, T$.

$F_n(l,t)$ = the minimum cost in periods $1, \ldots, t$, under exactly n orders, when period l is the *last* order period preceding period t.

To determine whether for a given horizon $t = 1, \ldots, T$, some period l is a better 'last' order period than some earlier period $k < l$, we consider the difference function $\Delta_{n,k,l}(t) = F_n(k,t) - F_n(l,t)$. Observe that

$$F_n(l,t) \quad = \quad F_{n-1}(l-1) + S(l,t) + c_l[B(t) - B(l-1)] \tag{5.14}$$

$$F_n(k,t) \quad = \quad F_{n-1}(k-1) + S(k,t) + c_k[B(t) - B(k-1)] \tag{5.15}$$

Subtracting (5.14) from (5.15), we obtain after some algebra that

$$\Delta_{n,k,l}(t) = A_n(k,l) + [\tilde{C}(k) - \tilde{C}(l)]B(t) \tag{5.16}$$

where

$$\begin{aligned} A_n(k,l) &= F_{n-1}(k-1) - F_{n-1}(l-1) + s(l-1) - s(k) \\ &\quad + \tilde{C}(k)[B(l-1) - B(k-1)] + \beta_k H(k-1) \\ &\quad + B(k-1)(\tilde{C}(l-1) - \tilde{C}(k)) \end{aligned} \tag{5.17}$$

Thus, the difference function $\Delta_{n,k,l}(t)$ depends on the cost and demand parameters in the periods $l+1,\ldots,t$ *only via* the *single* characteristic $B(t)$, i.e. the *cumulative* demand factor in periods $1,\ldots,t$. Furthermore, it is easy to characterize for which values $B(t)$, k dominates l as a last order period. Consider the following two cases:

(I) $\tilde{C}(k) \le \tilde{C}(l)$: The condition is equivalent to $c_{k,l} \le c_k$; in this case, it is *never* strictly better to use l as the last order period rather than an earlier period. To show this, consider an optimal zero-inventory ordering plan for some planing horizon $t \ge l$, with l as the last order period. Since $c_{k,l} \le c_l$, costs do not increase when placing the order in the earlier period k. The resulting policy may fail to be a zero-inventory policy but via a series of perturbations it can be transformed into a zero-inventory policy of equal or lower cost with the same set of order periods, i.e. with a period before l as the last order period.

(II) $\tilde{C}(k) > \tilde{C}(l)$: In this case, it follows from (5.17) that l is a *strictly better* last order period than k *if and only if*

$$B(t) > R_n(k,l) \stackrel{def}{=} A_n(k,l)/[\tilde{C}(k) - \tilde{C}(l)] \tag{5.18}$$

Since each of the values in $\{F_n(t) : t = 1, \ldots, T\}$ represents the cost of a zero-inventory policy, these values are completely characterized by determining for all $t = 1, \ldots, T$:

$l_n(t)$ = the optimal last order period for the time interval $[1, t]$, when the total number of orders in this interval must equal n, i.e.

$$F_n(t) = F_n(l_n(t), t) = \min_{l \le t} F_n(l, t) \tag{5.19}$$

(If more than one period qualifies as an optimal last order period, define $l_n(t)$ as the smallest such period). Clearly $F_n(t) = \infty$, if $t < n$, so we can restrict ourselves to the case where $t \ge n$.

To construct the list $\{l_n(t) : t = 1, \ldots, T\}$, our procedure, iteratively for $j = n, \ldots, T$, constructs an (ordered) list

$\mathcal{L}_n(j) = \{i : n \le i \le j : i$ is the (lowest indexed) best *last* order period for some planning horizon $t > j$, with potential cumulative demand factor $B(t) \ge B(j)\}$, along with an ascending list of critical cumulative demand values $\{B(j) = r(1) < r(2) < \cdots < r(m)\}$, with $m = |\mathcal{L}_n(j)|$. The k-th element of the list $\mathcal{L}_n(j)$ is the best last order period (under the restriction of n orders) for any horizon $t \ge j$ with cumulative demand $r(k) \le B(t) \le r(k+1), k = 1, \ldots, m$. (Here, $r(m+1) = \infty$.) Thus, under the restriction of n orders, $\mathcal{L}_n(j)$ is a list containing all periods among the first j periods, which may arise as an optimal last order period for some horizon $t > j$. In constructing this list, we treat all *future* demand factors $\beta_{j+1}, \ldots, \beta_T$ as unknown.

Clearly, the list $\{l_n(j) : j = 1, \ldots, T\}$ of *actual* optimal last order periods is simply the list of the *first* elements of the respective lists $\{\mathcal{L}_n(1), \mathcal{L}_n(2), \ldots, \mathcal{L}_n(T)\}$.

The following lemma identifies an effective way to 'update' the lists $\{\mathcal{L}_n(j) : j = 1, \ldots, T\}$. Clearly, $\mathcal{L}_n(n) = \{n\}$ with $r(n) = B(n)$.

Lemma 5.2 *Fix $j = 1, \ldots, T$ and let $\mathcal{L}_n(j) = (i_1, \ldots, i_m)$*

(a) The periods in the list $\mathcal{L}_n(j)$ are distinct*.*

(b) $\mathcal{L}_n(j+1) \subseteq \mathcal{L}_n(j) \cup \{j+1\}$

(c) The periods appear in the list $\mathcal{L}_n(j)$ in descending *order of their $\tilde{C}$ values, i.e.*

$\tilde{C}(i_1) > \tilde{C}(i_2) > \cdots > \tilde{C}(i_m)$

(d) If $\tilde{C}(j+1) \geq \tilde{C}(i_m)$, then the list $\mathcal{L}_n(j+1) = \mathcal{L}_n(j)$ and the critical values $r(1), \ldots, r(m)$ remain unchanged as well.

(e) If $\tilde{C}(j+1) < \tilde{C}(i_m)$, period $j+1$ enters at the bottom of the list. To determine the corresponding $r(\cdot)$-value, as well as which of the existing elements of the list are to be removed, compute sequentially for $k = m, m-1, \ldots 1$ the root $r^* = R_n(i_k, j+1)$ and eliminate period i_k from the list as long as $r^* \leq r(k)$. The last computed root r^* has $r^* > r(k)$ and is the $r(\cdot)$-value for period $j+1$.

Proof. (a) Assume, to the contrary, that some period i appears more than once in the list, i.e. it is the best order period in *two* distinct intervals for the cumulative demand factor $B(t)$, but for some in between values of $B(t)$ some other period k is a better last order period. This implies that the difference function $\Delta_{n,k,l}(t)$, viewed as a function of $B(t)$, switches signs at least twice. Since the function is linear in $B(t)$ this results in a contradiction.

(b) Immediate.

(c)–(e): These parts are proven by induction with respect to $j = 1, \ldots, T$. Assume therefore that for the j-th list $\mathcal{L}_n(j)$, $\tilde{C}(i_1) > \tilde{C}(i_2) > \cdots > \tilde{C}(i_m)$. If $\tilde{C}(j+1) \geq \tilde{C}(i_m)$, we have shown under (I) above, that period i_m is a better last order period than period $j+1$ for *any* future horizon $t > j$. This proves part (d). Since the list remains unaltered, its elements continue to be ordered in decreasing order of their $\tilde{C}()$-values. If $\tilde{C}(j+1) < \tilde{C}(i_m), \tilde{C}(j+1)$ is *smaller* than the $\tilde{C}(\cdot)$-value of *all* elements of the list $\mathcal{L}_n(j)$, so that $(j+1)$ is the best last order period among all periods $1, \ldots, j+1$ for $B(t)$-values that are sufficiently large. This implies that $(j+1)$ is to be added to the list and since, by part (a), it enters in the list only once, it enters at the bottom of the list. This implies that the elements of the list $\mathcal{L}_n(j+1)$ continue to be ranked in decreasing order of their $\tilde{C}(\cdot)$-values, regardless of what elements of the list need to be eliminated, if any. This completes the induction proof for part (c). To verify the validity of the list-updating procedure in part (e), assume first that $r^* = R_n(i_m, j+1) > r(m)$. For $B(t) \geq r^*$, period $(j+1)$ is a better last order period than i_m and the list $\mathcal{L}_n(j)$ reveals that the latter dominates *all* other periods in $\{1, \ldots, j\}$ on this half line. Thus, for $B(t) > r^*$, period (j+1) is the *best* last order period among all of the first (j+1)-periods. On the other hand, for $r(m) \leq B(t) \leq r^*$, period i_m dominates period (j+1) as well as *all* periods in $\{1, \ldots, j\}$, so *it* is the best last order period among the first (j+1)-periods. It is also easily verified that each of the previous elements $i_k, k < m$ in the list continues to dominate on the interval $[r(k), r(k+1)]$ even when considering period (j+1) as an alternative. We conclude that, in this case, the list $\mathcal{L}_n(j+1)$ is obtained from the list $\mathcal{L}_n(j)$, simply by appending period $(j+1)$ to its tail with $g(m+1) = r^*$.

Consider now the remaining case where $r^* = R_n(i_m, j+1) < r(m)$. In this case, period i_m is dominated by period $(j+1)$ for $B(t) \geq r(m)$, while it is dominated by some period in $\{1, \ldots, j\}$ for $B(t) < r(m)$. This implies that period i_m is to be eliminated from the list and the updating process can now proceed with this curtailed list. ∎

Remark: Upon completion of the list $\mathcal{L}_n(j+1)$, it is advisable to consider the actual cumulative demand factor value $B(j+1)$ and eliminate from the front of the list *all* elements with an $r(\cdot)$-value below $B(j+1)$. (After all, all future horizons $t > j+1$ have $B(t) \geq B(j+1)$.)

The following summarizes he full algorithm to determine the complete matrix $F = \{F_n(t) : n = 1, \ldots, T; t = 1, \ldots, T\}$

We maintain at each iteration on ordered list

$\mathcal{L} = (\{N[FIRST], N[FIRST+1], \ldots, N[LAST]\})$ such that for $j = 1, \ldots, T$, $\mathcal{L} = \mathcal{L}_n(j)$ at the end of the j-th iteration. The records in this list are numbered FIRST, FIRST+1, . . . , LAST for appropriate values of FIRST and LAST. The k-th record $(FIRST \leq k \leq LAST)$ contains *two* numbers $\{N[k], r(k)\}$.

As explained, periods are eliminated from either the front or the tail of the list $\mathcal{L}_n(\cdot)$. We therefore distinguish between two elementary procedures:

(i) DELTOP: this procedure deletes the *first* record of the list and sets FIRST := FIRST=1;

(ii) DELBOT: the procedure deletes the *last* record of the list and sets LAST := LAST-1;

Algorithm 1 Algorithm Restricted Lot Sizing

$F_1(1) := c_1\beta_1$
for $j = 2:1:T$ **do**
    $B(j) := B(j-1) + \beta_j$
    $H(j) := H(j-1) + h_j$
    $\tilde{C}(j) := c_j - H(j-1)$
    $s(j) := s(j-1) + \beta_j H(j-1)$
    $F_1(j) := F_1(j-1) + c_1\beta_j + \beta_j H(j-1)$
end for
for $n = 2, 1, T$ **do**
    for $j = 1, 1, n-1$ **do**
        $F_n(j) := \infty$
        $FIRST := 1$
        $LAST := 1$
        $N[FIRST] := n$
        $r[FIRST] := B(n)$
        $r[FIRST+1] := \infty$
    end for
    for $j = n, 1, T$ **do**
        while $r[FIRST+1] \leq B(j)$ **do**
            DELTOP
        end while
        if $\tilde{C}(j) < \tilde{C}(LAST)$ **then**
            while $R_n(N[LAST], j) \leq r(LAST)$ **do**
                DELBOT
                $r(LAST+1) := R_n(N[LAST], j)$
                $N[LAST+1] := j$
                $LAST := LAST + 1$
            end while
            $l := l_n(j) := N[FIRST]$
            $F_n(j) := F_{n-1}(l) + c_l(B(j) - B(l-1)) + s(j) - s(l) - H(l-1)(B(j) - B(l))$
        end if
    end for
end for

It is easily verified that the computation of each row in the matrix F, i.e. the effort expended in the outer 'for do' loop in Step 1 is $O(T)$. (Note that the procedure DELTOP and DELBOT are invoked at most (T-n+1) times during this loop, since each period in $\{n, \ldots, T\}$ can be eliminated at most once, either by DELTOP or by DELBOT.

Both procedures are of complexity $O(1)$ as is the computation of the roots $R_n(1,1)$ via (5.18)).

Example 5.2: Consider an industry with $N = 3$ firms and deseasonalized demand functions:

$$\delta_1(p) = 400 - 10p_1 + p_2 + p_3 \tag{5.20}$$

$$\delta_2(p) = 250 + p_1 - 12p_2 + 10p_3 \tag{5.21}$$

$$\delta_3(p) = 250 + p_1 + 10p_2 - 12p_3 \tag{5.22}$$

The firms face a planning horizon of $T = 54$ periods. We consider six different multiplicative seasonality patterns $\{\beta_t : t = 1, \ldots, 54\}$, which are common to all three firms, as follows:

(I) (Time-invariant demand function) $\beta_t = 1 \quad ; \quad t = 1, \ldots, 54$

(II) (Linear Growth) $\beta_t = 0.25 + 1.5\frac{(t-1)}{53} \quad ; \quad t = 1, \ldots, 54$

(III) (Linear Decline) $\beta_t = 1.75 - 1.5\frac{(t-1)}{53} \quad ; \quad t = 1, \ldots, 54$

(IV) (Holiday Season at Beginning of Planning Horizon)

$$\beta_t = \begin{cases} \frac{54}{114} + \frac{540}{570}(t-1) & , t = 1, \ldots, 6 \\ \frac{594}{114} - \frac{540}{570}(t-7) & , t = 7, \ldots, 12 \\ \frac{54}{114} & , t = 13, \ldots, 54 \end{cases} \tag{5.23}$$

(V) (Holiday Season at End of Planing Horizon)

$$\beta_t = \begin{cases} \frac{54}{114} & , t = 1, \ldots, 42 \\ \frac{54}{114} + \frac{540}{570}(t-43) & , t = 43, \ldots, 48 \\ \frac{594}{114} - \frac{540}{570}(t-49) & , t = 49, \ldots, 54 \end{cases} \tag{5.24}$$

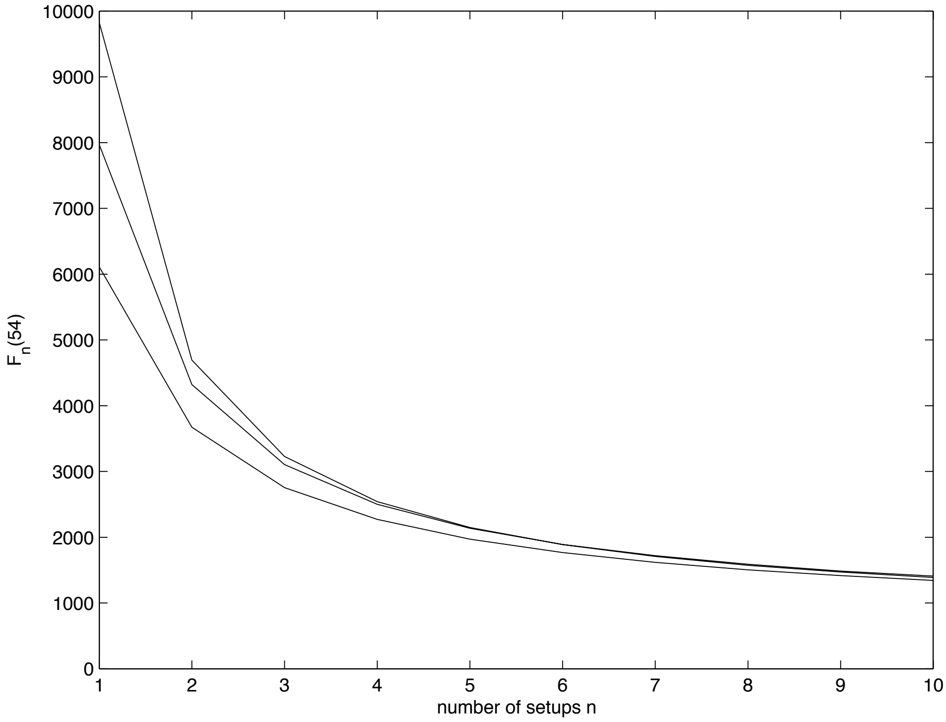
(VI) (Cyclical Pattern)

$$\beta_t = \begin{cases} 0.25 + 0.75(t-1) & , t = 1, \ldots, 3 \\ 1.75 - 0.75(t-4) & , t = 4, \ldots, 6 \\ \beta_{t \bmod 6} & , t = 7, \ldots, 54 \end{cases} \tag{5.25}$$

where t mod 6 denotes t modulo 6. Note that the *average* seasonality factor $\frac{1}{54}\sum_{t=1}^{54} \beta_t = \bar{\beta} = 1$ in all six patterns (I)–(VI). The first pattern reflects a situation where demand functions are time-invariant and the second (third) pattern one with linear growth (decline). The fourth and fifth pattern represent a planning horizon with a single season of peak demands either at the beginning or at the end of the planning horizon. Finally, the last pattern (VI) is cyclical with a cycle length of 6 periods, such that demands in the two middle periods of each cycle are 7 times their value in the first and last period, while $\beta_t = 1$ in the remaining two periods of the cycle.

All three firms share the same cost parameters which are time-invariant: $K_t^i = 1000$; $c_t^i = 15$; $h_t^i = 5$ for all $i = 1, \ldots, 3$; $t = 1, \ldots, 54$. Since the firms have identical c- and h- values and the firms have identical seasonality patterns, they also share the same values for $\left\{F_n^i(t) : t = 1, \ldots, 54\right\}$. In Figure 5.1(a) and 5.1(b) we display the values $\{F_n(54)\}$ as a function of the permitted number of setup periods $n = 1, \ldots, 54$,

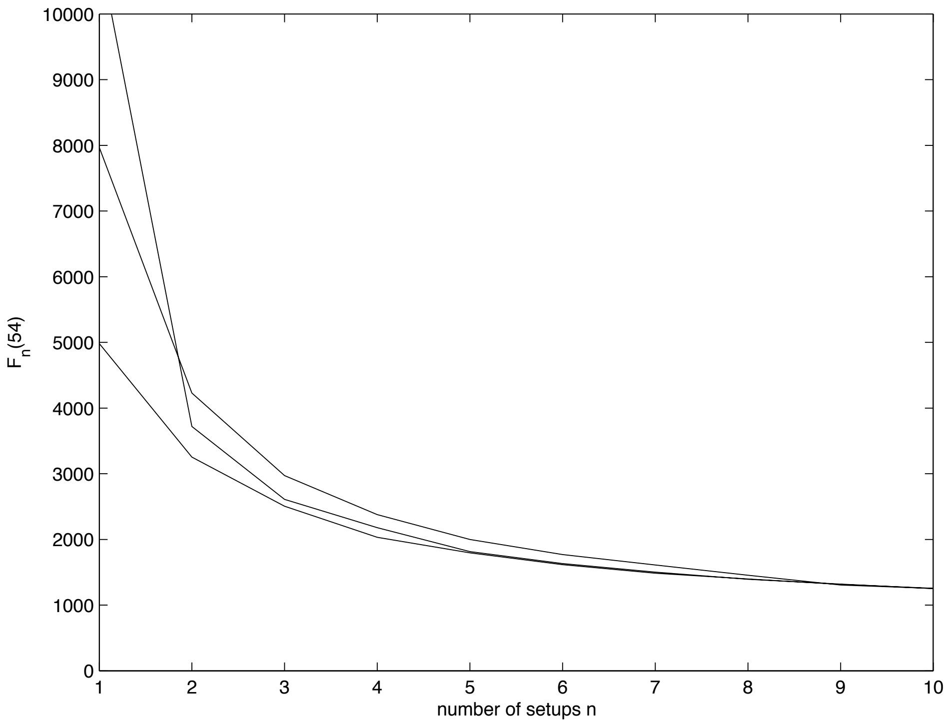
Figure 5.1(a): $F_n(54)$ as a function of the number of setups periods n, patterns (I)-(III)



for the six seasonality patterns (I)–(VI). Observe that the $F_n(54)$-values are significantly different across the six patterns, for small numbers of permitted setup periods, but the relative difference gradually decreases as n increases. Since the parameters c_t^i parameters are constant over time, the variable procurement cost component in $\{F_n(54)\}$ is identical for all $n = 1, \ldots, 54$ and for all seasonality patterns (I)–(VI). All differences in the $\{F_n(54)\}$ values are therefore attributable to differences in the holding costs. These, of course, decline to zero as n increases to 54. For all $n \geq 7$, the cyclical pattern (VI) is the *least* expensive to service and the time-invariant pattern (I) the *most* expensive. This contradicts common folklore which assumes that optimal costs are achieved when one is facing a smooth, time-invariant sales pattern.

The curves $\{F_n(54) : n = 1, \ldots, 54\}$ can be approximated very closely by curves of

Figure 5.1(b): $F_n(54)$ as a function of the number of setups periods n, patterns (IV)-(VI)



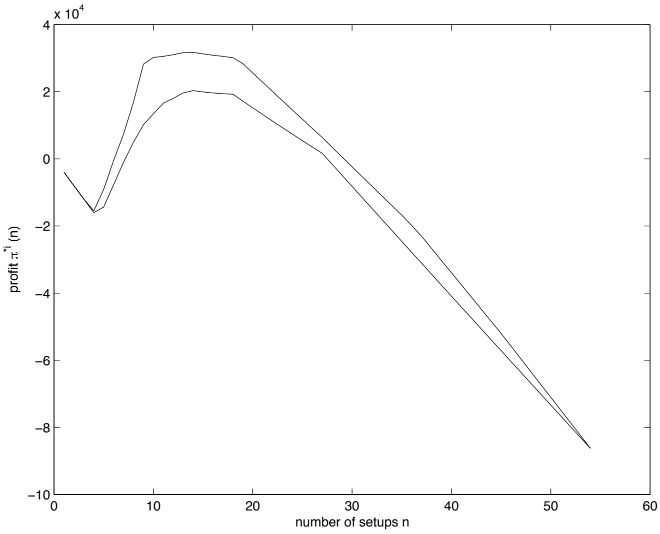
the shape

$$F_n^i(54) \sim \left(\sum \beta_t^i\right)\left[\left(c^i+\eta^i\right)+\frac{\zeta^i}{n}\right] \quad , \quad n=1,\dots,54 \tag{5.26}$$

for appropriate constants $\eta^i, \zeta^i > 0$. For example, η^i and ζ^i can be chosen to minimize the sum of squared differences between the left and the right sides of (5.26). For pattern (I) and (II), for example, we thus obtain $(\eta^1, \zeta^1) = (-2.275, 134.595)$ and $(\eta^6, \zeta^6) = (-3.378, 133.810)$ with average relative differences between the exact and the approximate curve of 0.6% and 3.3 %, respectively. We have found that approximations of the type (5.26) are, in fact, very close, across the board, for *any* combination of cost values and seasonality patterns.

Several facts explain the extreme goodness of fit: in our numerical experience, the function $F_n(T)$ is always *convexly* decreasing in n. We conjecture that this is always the case. To motivate this conjecture, note that $F_n(T)$ may be viewed as the value of an n-median problem. Once again, it is unknown whether the optimal cost-value of an n-median problem is convex in n, but it is known from Cornuejols et al. (1977), Nemhauser et al. (1978) and Nemhauser and Wolsey (1988) that the greedy heuristic solution for the n-median problem is a very close approximation (both in terms of worst case and average performance) and the cost value of this greedy heuristic solution *is* convexly decreasing. The greedy heuristic, applied to our lot sizing problem, adds in the $n+1^{\text{st}}$ iteration a new setup period to the n periods selected in the first n periods which results in the largest cost decrease. The approximation (5.26) will be used in the equilibrium analysis in Section 5.5.

Figure 5.2: Optimal profit for firm 1 as a function of permitted number of setups under pattern (I) and (VI) when $K_t^i = 1000$



We now turn to the best response problem firm 1 faces when both of his competitors choose a price value of \$30. Based on (5.11) and (5.12), Figure 5.2 exhibits the optimal profit $\pi^{*i}(n)$ for pattern (I) and (VI) as a function of the permitted number of setups n. The globally optimal prices which solve the best response problem are \$31.75 and \$30.89 and correspond with a lot sizing schedule with $n = 27$ for pattern (I) and $n = 35$ for pattern (VI).

5.4 The Best Response Problem: The General Case

In the Section we address the best response problem (5.6) for the *general* case where additive as well as multiplicative seasonalities exist and where all cost parameters vary in arbitrary ways over the course of the planning horizon. In the most general case, no solution procedure appears to exist, which, in the worst case, involves a polynomially bounded number of elementary operations or evaluations of a simple, analytical closed form function. However, a highly efficient procedure can be developed based on the following result:

For any given set of periods $\Theta \in 2^{\{1,\dots,T\}}$, let $\mathcal{P}^i(\Theta|p^{-i}) = \Big\{p^i|\Theta$ is an optimal set of order periods to service the demand stream $\{d_t^i(p) : t = 1,\dots,T\}\Big\}$.

Lemma 5.3 *Fix $i = 1,\dots,N$ and a vector p^{-i} of prices for firm i's competitors*

(a) $\mathcal{P}^i(\Theta|p^{-i})$ is a closed interval.

(b) Let $[p_0^i, p_1^i]$ denote the closed interval $\mathcal{P}^i(\Theta^0|p^{-i})$. Let $\delta_0^i = \delta^i(p_0^i|p^{-i})$ and $\delta_1^i = \delta^i(p_1^i|p^{-i})$. Firm i's profit is a concave differentiable function of δ^i on the interval $[\delta_1^i, \delta_0^i]$.

Proof: Fix $\Theta^0 \in 2^{\{1,\dots,T\}}$. We first show that $\mathcal{P}^i(\Theta^0|p^{-i})$ is convex. Thus, let $p_1^i < p_2^i \in \mathcal{P}^i(\Theta^0|p^{-i})$ and assume to the contrary that for some p_3^i with $p_1^i < p_3^i < p_2^i$, Θ^0 fails to be an optimal set of order periods. Let Θ^1 be a sequence of order periods which *is* optimal to use under price level p_3^i. Note from (5.5) that $C^i(p|\Theta)$ is an *affine* transformation of the deseasonalized demand function $\delta^i(p)$, i.e.

$C^i(p|\Theta) = A(\Theta) + B(\Theta)\delta^i(p)$, with $A(\Theta)$ and $B(\Theta)$ constants, *independent* of any price choices. Thus, $C^i(p|\Theta^0) - C^i(p|\Theta^1) = [A(\Theta^0) - A(\Theta^1)] + [B(\Theta^0) - B(\Theta^1)]\delta^i(p)$ is a *monotone* function of p^i, since $\delta^i(p)$ is a *strictly decreasing* function of p^i. Yet

$C^i((p_1^i, p^{-i})|\Theta^0) - C^i(p_1^i, p^{-i})|\Theta) \le 0$, since $p_1^i \in \mathcal{P}^i(\Theta^0|p^{-i})$

$C^i((p_3^i, p^{-i})|\Theta^0) - C^i((p_3^i, p^{-i})|\Theta^1) > 0$, since $p_3^i \notin \mathcal{P}^i(\Theta^0|p^{-i}), p_i^3 \in \mathcal{P}^i(\Theta^1|p^{-i})$

$C^i((p_2^i, p^{-i})|\Theta^0) - C^i((p_2^i, p^{-i})|\Theta^i) \le 0$, since $p_2^i \in \mathcal{P}^i(\Theta^0|p^{-i})$, indicating that this difference function *fails* to be *monotone* in p^i, thus resulting in a contradiction.

Since $\mathcal{P}^i(\Theta^0|p^{-i})$ is a convex set, it is an interval. To show that that his interval is closed, consider a sequence $\{p_k^i\}_{k=1}^{\infty}$ with $p_k^i \in \mathcal{P}^i(\Theta^0|p^{-i})$ for all $k = 1, 2, \ldots$ and $\lim_{k\to\infty} p_k^i = p_*^i$. Thus, $C^i((p_k^i, p^{-i})|\Theta^0) \leq C^i((p_k^i, p^{-i})|\Theta)$ for all order period sequences $\Theta \in 2^{\{1,\cdots,T\}}$. Taking limits on both sides of the inequality and using the continuity in p of the $C_i(p|\Theta)$ function we conclude that the limit p_*^i satisfies the inequality as well, i.e. $p_*^i \in \mathcal{P}(\Theta^0|p^{-i})$.

(b) Per definition, on the interval $\mathcal{P}^i(\Theta^0|p^{-i})$, $C^i(p) = C^i(p|\Theta^0)$ $= A(\Theta^0) + B(\Theta^0)\delta^i(p)$, so that firm i's profit $\Pi_i(\delta^i|p^{-i}) = R^i(\delta^i|p^{-i}) - A(\Theta^0) - B(\Theta^0)\delta^i$ is a concave function of δ^i. ■

Lemma 5.3 suggests the following best response algorithm. Once again, fix $i = 1, \ldots, N$, p^{-i} and a (small) precision number $\epsilon > 0$.

Algorithm: Best Response Problem

Step 0 (Initialization): Set $l := 1; \pi^* := 0, p_1^i := p_{\min}^i; \delta_1^i = \delta^i(p_1^i|p^{-i})$.

Step 1 (Iterative Step): For $p^i = p_l^i + \epsilon$, determine an optimal sequence of order periods Θ^l as well as the coefficients $A(\Theta^l), B(\Theta^l)$, i.e. $C^i(p|\Theta^l) = A(\Theta^l) + B(\Theta^l)\delta^i(p)$ (e.g. with the $O(T \log T)$ method in Federgruen and Tzur (1991)). With simple bisection, determine within ϵ, $p_u^i = \max\left\{p^i \leq p_{\max}^i|\Theta^l \text{ is an optimal sequence of order periods for the price } p^i\right\}$; $\delta_u^i = \delta^i(p_u^i|p^{-i})$. Determine $\pi_l^* = \max_{\delta_u^i \leq \delta^i \leq \delta_l^i} \left\{R^i(\delta^i|p^{-i}) - A(\Theta^l) - B(\Theta^l)\delta^i\right\}$ where the function within curled brackets is *concave* in δ^i. If $\pi_l^* > \pi^*$ then $\pi^* := \pi_l^*$.

Step 2: If $p_u^i < p_{\max}^i$ then **begin** $l := l + 1; p_l^i = p_u^i; \delta_l^i = \delta_u^i$; return to Step 1 **end**

This Algorithm thus partitions the complete price interval $[p_{\min}^i, p_{\max}^i]$ into $L \geq 1$ consecutive subintervals, such that on each subinterval a given sequence of order periods Θ remains optimal. On each interval the profit function is a simple *differentiable concave* function of δ^i whose maximum can be found by comparing the values at the end points with that in the (at most) single stationary point where the derivative of the profit function equals zero.

If the interval $[p^i_{\min}, p^i_{\max}]$ is partitioned in to L subintervals, i.e. if Step 1 is repeated L times, the total effort to identify the end points of the subintervals is $O\left(L\left\lceil \log_2\left(\frac{p^i_{\max}-p^i_{\min}}{\epsilon}\right)\right\rceil T\log T\right)$ while the effort to optimize the profit-function on each of the subintervals amounts to L maximizations of a differentiable concave (closed form) function of a single variable.

Example 5.3: Consider the previous example, however with *additive* rather than *multiplicative* seasonality terms, which are again identical to all three firms, i.e. let $\beta^i_t = 1$, and $\alpha^i_t = -100 + \frac{200}{53}(t-1), \quad t = 1, \ldots, 54.$

Thus, each period a firm's demand function is shifted upwards by a constant amount and $\frac{1}{54}\sum_{t=1}^{54}\alpha^i_t = 0$. Consider, again, firm 1's best response problem which arises when his competitors adopt a price of \$ 30, i.e. $p_2 = p_3 = 30$. In Table 1, we display the intervals identified by the 'Algorithm Best Response Problem'. along with, for each of the intervals, the optimal price for firm 1, corresponding profit level and the number of setups periods in the last column which is optimal for this interval.

Table 5.1: Intervals generated by Algorithm 'Best Response Problem' for firm 1 in Example 3

LB	UB	OptPrice	Profit	Setups
0.000	16.374	16.374	75730	54
16.384	17.141	17.141	76440	53
17.151	17.878	17.878	77110	52
17.888	18.640	18.640	77750	51
18.650	19.404	19.404	78350	50
19.414	20.151	20.151	78920	49
20.161	20.911	20.911	79450	48
20.921	21.664	21.664	79940	47
21.674	22.409	22.409	80400	46
22.419	23.168	23.168	80830	45
23.178	23.923	23.923	81210	44
23.933	24.682	24.682	81570	43
24.692	25.437	25.437	81880	42
25.447	26.184	26.184	82170	41
26.194	26.950	26.950	82410	40
26.960	27.693	27.693	82630	39
27.703	28.457	28.457	82800	38
28.467	29.210	29.210	82940	37
29.220	29.968	29.968	83050	36
29.978	30.337	30.337	83190	35
30.347	31.092	31.092	83240	34
31.102	31.475	31.475	83290	33
31.485	**32.235**	**31.611**	**83290**	**32**
32.245	32.604	32.245	83240	31
32.614	33.361	32.614	83190	30
33.371	33.736	33.371	83040	29
33.746	34.107	33.746	82760	28
34.117	34.875	34.117	82660	27
34.885	35.242	34.885	82380	26
35.252	35.620	35.252	81770	25
35.630	35.994	35.630	81280	24

5.5 The Equilibrium Analysis

In this Section, we analyze the equilibrium behavior in the industry under price—or Bertrand competition. (See however, the end of this section for a discussion of the case of Cournot competition.) Once again, we start with the case where all firms experience only multiplicative seasonalities and their fixed and variable order cost parameters remain constant throughout the planning horizon. (The best response problem of Section 5.3 thus applies; we append the assumption of constant variable order cost rates to simplify the exposition below.) Thus, let $K^i = K_1^i = \ldots = K_T^i$ and $c^i = c_1^i = \ldots = c_T^i$ for all $i = 1, \ldots, N$.

Substituting the expression (5.8) for the cost function $C^i(p)$ as well as the identity $d_t^i(p) = \beta_t^i \delta^i(p)$ in the profit function (5.4), we obtain:

$$\pi^i(p) = B^i(T)p^i\delta^i(p) - \min_{n=1,\ldots,T} \left\{ nK^i + B^i(T)c^i\delta^i(p) + \delta^i(p)\tilde{F}_n^i(T) \right\} \tag{5.27}$$

where $\tilde{F}_n^i(t)$ = minimum total holding costs in periods $\{1, \ldots, t\}$ for firm i, assuming the firm's demand stream is given by seasonality factors $\left\{\beta_1^i, \ldots, \beta_T^i\right\}$ and assuming exactly n order are placed in the first t periods, $t = 1, \ldots, T$, $n = 1, \ldots, t$, $i = 1, \ldots, N$.

$$\tilde{F}_n^i(t) = F_n^i(t) - c^i \sum_{s=1}^{t} \beta_s^i \tag{5.28}$$

It is difficult to guarantee the existence of a Nash equilibrium on the basis of the exact characterization of the cost function $C^i(p)$ as a piecewise linear function of the deseasonalized demand volume $\delta^i(p)$. Note, for example, from (5.10) – (5.12), that firm i's best response price fails to vary continuously with the competitor's prices, even tough the points of discontinuity form a set of measure zero. However, as substantiated in Section 5.3, the values $\left\{\tilde{F}_n^i(t)\right\}_{n=1}^T$ are closely approximated by a function of the from (5.26). Substituting (5.26) into (5.27) and treating the number of order periods n as a continuous variable, we obtain the (approximate) profit functions

$$\begin{aligned} \tilde{\pi}^i(p) &= B^i(T)\delta^i(p)(p^i - c^i) - \min_{n\geq 0}\left\{nK^i + B^i(T)\eta^i\delta^i(p) + \frac{B^i(T)\zeta_i\delta^i(p)}{n}\right\} \\ &= B^i(T)\delta^i(p)(p^i - c^i - \eta^i) - 2\sqrt{K^i\zeta^i}\sqrt{\delta^i(p)} \\ &= B^i(T)\left\{\delta^i(p)(p^i - \eta^i - c^i) - \left(\frac{2\sqrt{K^i\zeta^i}}{B^i(T)}\right)\sqrt{\delta^i(p)}\right\}, \; 1 = 1, \ldots, N \end{aligned} \tag{5.29}$$

where the first equality follows from simple calculus as in the well-known Economic Order Quantity (EOQ) model.

The approximation (5.26) for the values $\left\{\tilde{F}_n^i(T)\right\}$ thus permits us to represent the cost function as one consisting of an explicit linear term in $\delta^i(p)$ plus a term proportional to the square root of this deseasonalized demand value. A price competition game in which each firm selects a price from a closed interval and with profit functions of the type (5.29) has been analyzed by Cachon and Harker (2002) and Bernstein and Federgruen (2003). The former confine themselves to the case where the number of firms $N = 2$. Without guaranteeing that an equilibrium exists, they establish a sufficient condition under which the existence of *multiple* equilibria can be excluded. Bernstein and Federgruen (2003) show that the existence of an equilibrium can, indeed, not be guaranteed under completely arbitrary parameters. They identify, however, for the case of *linear* deseasonalized demand functions, see (5.2), a simple sufficient condition for the existence of an equilibrium, which relates the demand elasticity of a firm with respect to its own price to the firm's inventory-to-sales ratio. More specifically, assume the demand functions $\left\{\delta^i(p)\right\}$ satisfy (5.2).

Let $\epsilon^{ii} = \frac{\partial \delta^i(p)}{\partial p^i} \frac{p^i}{\delta^i(p)} = \frac{b^i p^i}{\delta^i(p)}$ the absolute value of the demand elasticity of firm i with respect to its own price.

$INV^i = 2\sqrt{\delta^i(p) K^i \zeta^i}$ = firm i's optimal inventory and fixed order costs over the course of the planning horizon, under the price vector p.

$REV^i = B^i(T) p^i \delta^i(p)$ = firm i's total gross revenue over the course of the planning horizon, under the price vector p.

Theorem 1 in Bernstein and Federgruen (2003) shows that a Nash equilibrium exists in the price competition model if the following condition (C) is satisfied:

$$\text{(C)} \quad \epsilon_{ii} \leq 8 \frac{REV^i}{INV^i} \quad , \quad i = 1, \dots, N \tag{5.30}$$

The ratio REV^i/INV^i is closely related to the (annual) sales-to-inventory ratio, one of Wall Street's most frequently monitored company measures. Bernstein and Federgruen (2003) argue that the ratio REV^i/INV^i is, in fact, at least 2.5 times the sales-to-inventory ratio. Moreover, analyzing data by Dun and Bradstreet (2001), the authors show for a sample of 10 consumer product lines that the average *lower* quartile of the sales-to-inventory ratio varies between 2.8 and 6.7 for the 10 product lines. As a consequence, the right hand side of the inequality in (C) varies between 56 and 134, while the absolute value of the price elasticity varies between one and five, see e.g. Tellis (1988). Thus, condition (C) is very comfortably satisfied for virtually all production lines and industries.

While condition (C) guarantees that a Nash equilibrium exists, two important questions remain: (a) Is the Nash equilibrium *unique* or can *multiple* equilibria arise, making it hard to predict which of the possible equilibria the industry will adopt and (b) How can the equilibrium (or equilibria) be computed efficiently? Theorem 2 in BERNSTEIN and FEDERGRUEN (2003) shows that under a slight tightening of condition (C), a *unique* equilibrium can indeed be guaranteed and that this unique equilibrium can be efficiently computed as the limit point of the following simple (iterative) tatônnement scheme:

Tatônnement scheme: Starting with an arbitrary price vector $p_{(0)}$, in the k-th iteration of the scheme, each firm determines the price $p^i_{(k)}$ which solves the best response problem (5.6), assuming all competing firms' prices are set according to their value in the price vector $p_{(k-1)}$. Consider condition (C2):

$$\text{(C2)} \qquad \epsilon_{ii} \leq 4\frac{REV^i}{INV^i} \tag{5.31}$$

As reviewed above, condition (C2), which is somewhat tighter than (C), is still very comfortably satisfied for virtually all product lines. We conclude:

Theorem 5.1 *Assume the deseasonalized demand functions $\{\delta^i(p)\}$ are linear, i.e. they satisfy (5.2). Assume only multiplicative seasonalities apply ($\alpha^i_t = 0$), while all fixed and variable cost parameters are constant throughout the course of the planning horizon, i.e. $K^i_1 = \ldots = K^i_T = K^i$ and $c^i_1 = \ldots = c^i_T = c^i$. Consider the price competition game under the (approximate) profit functions $\tilde{\pi}^i(p)$.*

(a) Under condition (C), the price competition game has a Nash equilibrium

(b) Under condition (C2), the price competition game has a unique *equilibrium p^*. The tatônnement scheme converges to p^* from* any *starting point p^0.*

Proof: Immediate from Theorem 1 and 2 in Bernstein and Federgruen (2003).

The tatônnement scheme consists of repeated solutions of best response problems. The availability of efficient procedures to solve the best response problems (see Section 5.3) is therefore of critical importance.

It is difficult to establish a similarly intuitive sufficient condition to guarantee the existence of an equilibrium in the general model, with nonlinear deseasonalized demand functions or time-dependent order cost parameters. However, thanks to the availability of the best response algorithms in Section 5.3 and 5.4, the tatônnement scheme *can* be applied effectively to the fully general model. When convergent, its limit point *is*, of course, a Nash equilibrium. If convergent to the *same* limit point, regardless of its starting point $p_{(0)}$, the equilibrium is in fact *unique*. Thus, the tatônnement scheme provides an algorithmic mechanism to establish the existence of a Nash equilibrium in *any* given instance of the general model. Indeed, we have applied the scheme to a large variety of problem instances of the general model and have found hat a unique equilibrium exists in the vast majority of cases.

We conclude this Section with a brief discussion of the case where the firms engage in Cournot (quantity competition) as opposed to Bertrand (price competition). Consider, again, the conditions stated in Theorem 1, and assume the approximate profit functions $\tilde{\pi}^i(\cdot)$ are used. Under (5.5), the system of (deseasonalized) demand functions can be inverted resulting in the inverse demand functions:

$$p^i = \hat{a}^i - \hat{b}^i \delta^i - \sum_{j \neq i} \hat{\theta}^i_j \delta^j \quad , \quad i = 1, \ldots, N \quad \text{where} \quad \hat{b}^i, \hat{\theta}^i_j \geq 0 \tag{5.32}$$

It follows from Theorem 3 in BERNSTEIN and FEDERGRUEN (2003) that an equilibrium exists under condition (C) and that this equilibrium is *unique* under (C2) and the condition

$$\hat{b}^i > \sum_{j \neq i} \hat{\theta}^i_j \quad , \quad i = 1, \ldots, N, \tag{5.33}$$

the direct counterpart of (5.5) for *inverse* demand functions. Thus, the same condition (C) guarantees an equilibrium both under price- and under quantity competition. Moreover, an analogous pair of conditions guarantees that the equilibrium is *unique*. However, in contrast to the case of price competition, it is no longer possible to ensure that the tatônnement scheme converges to an equilibrium, even when the existence of a *unique* equilibrium can be guaranteed.

5.6 Numerical Examples

In this Section, we report on a numerical study conducted to investigate the effect on the equilibrium behavior of seasonality patterns and economies of scale resulting from fixed setup costs.

Table 5.2 displays the equilibrium in our base example (Example 5.2) under each of the six seasonality patterns (I)–(VI) and four combinations of setup cost values: (a)

$K_t^i = 500$; (b) $K_t^i = 1000$; (c) $K_t^i = 4000$ and (d) $K_t^i = 5400$. Each of the cells in the table corresponds with one of the 24 problem instances. Whenever a *unique equilibrium* exists, we display, sequentially, the equilibrium prices (p^*), volumes(δ^*), profit levels (π^*) and the number of order periods (n^*) for each firm. (If no equilibrium exists, or in case multiple equilibria arise, the cell is left empty.)

In our example, firms 2 and 3 have *identical* characteristics. Firm 1 has a larger intercept as well as a significantly higher 'total price sensitivity' defined by $\bar{b}_i = b_i - \sum_{j \neq i} \Theta_j^i > 0$, see (5.5). (The total price sensitivity measures the marginal decline in sales volume due to a *universal* price increase in the industry; note $\bar{b}_1 = 8$ and $\bar{b}_2 = \bar{b}_3 = 1$.) The larger total price sensitivity for firm 1 induces it to adopt a lower price than his competitors, in each of the 24 problem instances considered, the lower *direct* price sensitivity $10 = \frac{\partial \delta_i^i(p)}{\partial p^1} = b^1 < b^2 = b^3 = 12$ not withstanding.

While the differences in the price equilibria are relatively small, they often have very significant impacts on equilibrium volumes and in particular equilibrium profits. For example, when $K_t^i = 4000$, firm 1's profit is nearly 25% larger under the cyclic seasonality pattern (VI) than under constant demands (I). When $K_t^1 = 5400$ firm 1 more than *doubles* its profit when moving from the constant demand pattern (I) to a pattern with a holiday season at the beginning (IV). The profit increases for firm 2 and 3, when moving from (I) to (IV), are in *absolute* terms, approximately equal that experienced by firm 1, even though in relative terms they amount to an increase of approximately 16%.

It is hard to predict which seasonality pattern results in higher equilibrium prices: for $K_t^1 = 500$, the prices under the constant pattern (I) are lower than those under the cyclic pattern (VI) but the opposite is true for $K_t^i = 4000$. Note, also, that the impact of a cost increase on the equilibrium prices may vary rather significantly depending on which of the seasonality patterns prevails. Under (I), equilibrium prices increase by \$0.30 for firm 1 and \$0.20 for firms 2 and 3, while the price increases under (VI) are \$0.99 for firm 1 and \$0.62 for firms 2 and 3. The order pattern, in general, and the *number* of order periods, in particular, vary greatly by firm and by seasonality pattern, see for example the case where $K_t^i = 1000$.

Table 5.2: Equilibria under six seasonality patterns and four setup cost values

		(I)	(II)	(III)	(IV)	(V)	(VI)
	p^*	30.79 32.91 21.91	30.94 33.04 33.04	30.94 33.03 33.03	31.28 32.95 32.95	31.28 32.95 32.95	31.10 33.11 33.11
K = 500	δ^*	157.91 214.96 214.96	156.59 214.87 214.87	156.67 214.88 214.88	153.12 215.38 215.38	153.12 215.38 215.38	155.22 214.87 214.87
	π^*	107.66 180.94 180.94	108.91 182.76 182.76	109.05 182.78 182.78	110.61 181.75 181.75	110.61 181.75 181.75	111.60 185.26 185.26
	n^*	54 54 54	47 50 50	47 50 50	32 54 54	32 54 54	37 45 45
	p^*	32.05 33.00 33.00	31.60 33.61 33.61	31.63 33.57 33.57	31.36 33.78 33.78	31.40 33.78 33.78	31.24 33.44 33.44
K = 1000	δ^*	145.50 216.04 216.04	151.20 214.39 214.39	150.88 214.48 214.48	153.96 213.80 213.80	153.52 213.84 213.84	154.51 214.35 214.35
	π^*	87.33 156.04 156.04	89.45 165.83 165.83	89.93 166.00 166.00	95.99 173.69 173.69	96.27 173.76 173.76	93.91 169.76 169.76
	n^*	27 54 54	34 41 41	33 41 41	32 32 32	31 32 32	35 37 37
	p^*	34.86 37.49 37.49	34.28 36.64 36.64	34.40 36.77 36.77		33.72 36.13 36.13	33.87 36.13 36.13
K = 4000	δ^*	126.38 209.88 209.88	130.48 211.00 211.00	129.51 210.87 210.87		135.03 211.46 211.46	133.58 211.61 211.61
	π^*	30.25 126.22 126.22	31.93 120.35 120.35	34.58 124.09 124.09		42.46 129.22 129.22	40.36 129.50 129.50
	n^*	14 18 18	15 20 20	14 19 19		14 18 18	14 18 18
	p^*	35.23 37.52 37.52	35.13 38.12 38.12	35.29 38.28 38.28	34.89 37.72 37.72	34.86 37.68 37.68	
K = 5600	δ^*	122.70 210.20 210.20	124.99 208.88 208.88	123.71 208.72 208.72	126.53 209.45 209.45	126.71 209.51 209.51	
	π^*	8.50 98.03 98.03	11.56 106.74 106.74	15.44 112.04 112.04	24.86 119.02 119.02	25.10 119.12 119.12	
	n^*	13 18 18	13 16 16	12 15 15	11 14 14	10 17 17	

When $K_i^t = 4000$, a unique equilibrium continues to exist under all seasonality patterns except for (IV). Here, there are *at least two* equilibria $p_1^* = [34.03; 36.28; 36.40]$ and $p_2^* = [34.03; 36.40; 36.28]$. (These equilibria are both interior points of the feasible region.) The tatônnement scheme may, in this case, converge to either one of the two equilibria *or* it cycles through the points $p_1^* = [34.02; 36.40; 36.40]$ and $p_2^* = [34.04; 36.28; 36.28]$.

When $K_i^t = 5400$, an equilibrium fails to exist for pattern (VI): the tatônnement scheme *never* converges, irrespective of the starting point. Instead, the scheme cycles between the two price points $p_1^* = [35.24; 36.54; 36.54]$ and $p_2^* = [35.25; 36.60; 36.60]$.

We conclude that the seasonality patterns of the demand functions (as well as the cost parameters) may result in significant differences in the equilibria. Sometimes, the seasonality pattern determines whether a unique equilibrium, no equilibrium or multiple equilibria prevail. The latter two cases arise, of course, in settings where conditions (C) and (C2) are violated.

We conclude this Section with a set of six problem instances with *additive* rather than *multiplicative* seasonality factors for the demand function, i.e. $\beta_t^i = 1$. The six instances are obtained from the base case in Example 5.2 (where all $\alpha_t^i = 0$), merely by varying the $\left\{\alpha_t^i\right\}$ terms. The seasonality patterns are the direct analog of (I)–(VI) in the case of multiplicative seasonalities.

For each pattern, we obtain the α_t^i-terms from the β_t^i-factors for the *same* pattern, employing the transformation $\alpha_t^i = 100(\beta_t^i - 1)$. (Note that $\frac{1}{54}\sum_{t=1}^{54} \alpha_t^i = 0 \ \forall i$.) Table 5.3 displays the equilibrium prices, demand values, profits and number of setups for the 6 instances. As in the case of multiplicative seasonalities, we note that the differences in the equilibrium prices are modest. At the same time, major differences in the number of setup periods may arise, translating into equilibrium profit differences of up to 8.4%. The differences become even larger when the setup cost parameters K_t^i are increased, see Table 5.2. As in the case of multiple seasonalities, all firms may be better off under seasonally varying demands or cost parameters, once again contradicting common folklore.

Table 5.3: Equilibria under six α-patterns

		(I)	(II)	(III)	(IV)	(V)	(VI)
	p^*	32.05 33.00 33.00	31.91 33.87 33.87	31.95 33.87 33.87	32.04 34.75 34.75	32.04 34.75 34.75	32.08 33.72 33.72
K = 1000	δ^*	78.57 116.66 116.66	80.29 115.65 115.65	80.04 115.67 115.67	80.51 114.77 114.77	80.51 114.77 114.77	79.21 115.90 115.90
	π^*	87.33 156.04 156.04	88.77 165.83 165.83	89.10 166.06 166.06	94.63 177.07 177.07	94.63 177.07 177.07	89.56 165.68 165.68
	n^*	27 54 54	32 43 43	31 43 43	31 32 32	31 32 32	36 45 45

References

AGGARWAL, A., J. PARK. 1993. Improved algorithms for economic lot size problems. *Operations Research* **41** 549–571.

AHUJA, R., D. HOCHBAUM. 2004. Solving Linear Cost Dynamic Lot Sizing Problems in $O(n \log n)$ time. Working paper, University of California, Berkeley.

ANILY, S., J. BRAMEL 1999. Vehicle Routing and the Supply Chain. In: S. TAYUR, R. GANESHAN, M. MAGAZINE (Edts): *Quantitative Models for Supply Chain Management*. Kluwer, Boston.

ANILY, S., M. TZUR. 2004(a). Shipping Multiple-Items by Capacitated Vehicles - An Optimal Dynamic Programming Approach. *Transportation Science*, to appear.

ANILY, S., M. TZUR. 2004(b). Algorithms for the Multi-Item Capacitated Dynamic Lot Sizing Problem. Working paper, Tel Aviv University.

ASMUSSEN, S. 1987. *Applied Probability and Queues*, Wiley, New York.

BAKER, K., P. DIXON. 1978. An algorithm for the dynamic lot size problem with time-varying production capacity constraints. *Management Science* **24** 710–720.

BARANY, I., T. VAN ROY, L. WOLSEY. 1984(a). Strong formulations for multi-item capacitated lot-sizing. *Management Science* **30** 1255–1261.

BARANY, I., T. VAN ROY, L. WOLSEY. 1984(b). Uncapacitated lot-sizing: the convex hull of solutions. *Mathematical Programming Study* **22** 32–43.

BARRO, R. 1972. A theory of monopolistic price adjustment. *Review of Economic Studies* **34** 17–26.

BERNSTEIN, F., A. FEDERGRUEN. 2003. Pricing and Replenishment Strategies in a Distribution System with competing Retailers. *Operations Research* **51** 409–426.

BELVAUX, G., L. WOLSEY. 2000. *bc-prod* A specialized branch-and-cut system for lot-sizing problems. *Management Science* **46** 724–738.

BELVAUX, G., L. WOLSEY. 2001. Modelling practical lot-sizing problems as mixed-integer programs. *Management Science* **47** 993–1007.

BIXBY, R. 2001. MIP: A progress report. Talk at *1st Columbia Optimization Day*, New York, NY.

BITRAN, G., H. YANASSE. 1982. Computational complexity of the capacitated lot size problem. *Management Science* **28** 1174–1186.

BITRAN, G., R. CALDENTEY. 2003. An overview of pricing models for revenue management. *Manufacturing & Service Operations Management* **5** 203–229.

BLINDER, A., E. CANETTI, D. LEBOW, J. RUDD. 1998. *Asking About Prices*, Russell Sage Foundation, New York.

BRAMEL, J., D. SIMCHI-LEVI. 1997. *The Logic of Logistics.* Springer, New York.

BREIMAN, L. 1992. *Probability.* SIAM's Classics in Applied Mathematics series. SIAM, Philadelphia.

BRUMELLE, S., J. MCGILL. 1993. Airline seat allocation with multiple nested fare classes. *Operations Research* **41** 127–137.

CACHON, G., P. HARKER. 2002. Competition and outsourcing with scale economies. *Management Science* **48** 1314–1333.

CACHON, G. 2003. Supply Chain Coordination with Contracts. In GRAVES, S.C., A.G. DE KOK: *Handbook of Operations Management*, Elsevier Publishing Company, Amsterdam, Netherlands.

CARLTON, D.W. 1986. The Rigidity of Prices. *American Economic Review* **76** 637–658.

CECCHETTI, S. G. 1986. The Frequency of Price Adjustment: A Study of the Newsstand Prices of Magazines, *Journal of Econometrics* **31** 255–274.

CHAN, L., D. SIMCHI-LEVI. 1998. Probabilistic Analyses and Algorithms for Three-Level Distribution Systems. *Management Science* **44** 1562–1576.

CAPLIN, A., J. LEAHY. 1991. State-dependent pricing and the dynamics of money and output. *The Quarterly Journal of Economics* 683–708.

CHEN, H., D. HEARN and C. LEE. 1994. A new dynamic programming method for the single capacitated dynamic lot size model. *Journal of Global Optimization* **4** 285–300.

CHEVALIER, J., A. KASHYAP, P. ROSSI. 2003. Why Don't Prices Rise during Periods of Peak Demand? Evidence from Scanner Data. *American Economic Review* **93** 15–37.

CHUNG, C., M. LIN. 1988. AN O(T^2) algorithm for the NI/G/NI/ND capacitated single item lot size problem. *Management Science* **34** 420–426.

CLARKE, D., R. DOLAN. 1984. A Simulation Analysis of Alternative Pricing Strategies for Dynamic Environments; I. Introduction. *The Journal of Business* **57** 1792–201.

COFFMAN, E., G. LUEKER. 1991. Probabilistic Analysis of Packingand Partitioning Heuristics. Wiley, New York.

COOPER, W.. 2002. Asymptotic behavior of an allocation policy for revenue management. *Operations Research* **50** 720–727.

CORNUEJOLS, G., M.L. FISHER, G.L. NEMHAUSER 1977. Location of Bank Accounts to Optimize Float: An Analytic Study of Exact and Approximate Algorithms. *Managament Science* **23** 789–810.

DAI, J., R. WILLIAMS. 1994. Existence and uniqueness of semimartingale reflecting brownian motion in convex polyhedrons. *Theory of Prabability and its Application (in Russian).* **40** 3–53. To appear in the SIAM translation of the journal with the same name.

DIXON, P., E. SILVER. 1981. A heuristic solution for the multi-item, single level, limited capacity lot sizing problem. *Journal of Operations Management* **2** 23–39.

DOCKNER, E., S. JORGENSEN. 1988. Optimal Pricing Strategies for New Products in Dynamic Oligopolies. *Marketing Science* **7** 315–334.

DOGRAMACI, A., J. PANAYIOTOPOULOS and N. ADAM. 1981. The dynamic lotsizing problem for multiple items under limited capacity. *IEE Transactions* **13** 294–303.

Dun and Bradstreet. 2000-2001. *Industry Norms and Key Business Ratios.* Dun and Bradstreet Credit Services, New York.

EISENHUT, P.. 1975. A dynamic lotsizing algorithm with capacity constraints. *AIEE Transactions* **7** 170–176.

ELIASHBERG, J., R. STEINBERG. 1993. Marketing-production joint decision-making. In ELIASHBERG, J., LILIEN, G. (eds.): Handbooks in Operations Research and Management Science: Vol.5: Marketing. Amsterdam: North-Holland, 827–880.

ELIASHBERG, J., A. JEULAND. 1986. The Impact of Competitive Entry in a Developing Market upon Dynamic Pricing Strategies. *Marketing Science* **5** 20–36.

ELMAGHRABY, W., P. KESKINOCAK. 2003. Dynamic Pricing in the Presence of Inventory Considerations: Research Overview, Current Practices and Future Directions, *Management Science* **49** 1287–1309.

EPPEN, G., R. MARTIN. 1987. Solving multi-item capacitated lot sizing problems using variable redefinitions. *Operations Research* **35** 832–848.

FEDERGRUEN, A., D. SIMCHI-LEVI. (1991). Analytical Analysis of Vehicle Routing and Inventory Routing problems. In: M. BALL, T. MAGNANTI, C. MONMA, G. NEMHAUSER (eds): *Handbooks in Operations Research and Management Science, Vol. Network Routing.* North-Holland, Amsterdam.

FEDERGRUEN, A., M. TZUR. 1991. A Simple Forward Algorithm to solve General Dynamic Lot-Sizing Models with n Periods in O(n log n) or O(n) time. *Management Science* **37** 909–925.

FEDERGRUEN, A., M. TZUR. 1994a. Minimal forecast horizons and a new planning procedure for the general dynamic lot sizing model: nervousness revisited. *Operations Research* **42** 456–469.

FEDERGRUEN, A., M. TZUR. 1994b. Capacitated Lot-Sizing Models. Working Paper, Graduate School of Business, Columbia University, New York.

FEDERGRUEN, A., M. TZUR. 1994c. The Joint Replenishment Problem with time-varying parameters: Efficient, optimal and ϵ-optimal solutions. *Operations Research* **42** 1067–1087.

FEDERGRUEN, A., M. TZUR. 1999. Time-Partitioning Heuristics: Application to One-Warehouse, Multiitem, Multiretailer Lot-Sizing Problems. *Naval Research Logistics* **46** 463–486.

FEDERGRUEN, A., J. MEISSNER, M. TZUR. 2002. Progressive Interval Heuristics for Multi-Item Capacitated Lot Sizing Problems. Accepted in *Operations Research*.

FEDERGRUEN, A., J. MEISSNER. 2004a. Probabilistic Analysis of Progressive Interval Heuristics for Multi-Item Capacitated Lot-Sizing Problems. Preprint, Graduate School of Business, Columbia University, New York.

FEDERGRUEN, A., J. MEISSNER. 2004b. Price Competition under Time-Varying Demands and Dynamic Lot Sizing Costs. Preprint, Graduate School of Business, Columbia University, New York.

FENG, Y., B. XIAO. 2000. A continuous-time yield management model with multiple prices and reversible price changes. *Management Science* **46** 644–657.

FENG, Y., B. XIAO. 2004. Integration of pricing and capacity allocation for perishable products. Preprint, Chinese University of Hong Kong.

FISHER, M., D. HOCHBAUM. 1980. Probabilistic Analysis of the K-median problem. *Mathematics of Operations Research* **5** 27–34.

FLORIAN, M., M. KLEIN. 1971. Deterministic production planning with concave costs and capacity constraints. *Management Science* **18** 12–20.

FLORIAN, M., J. LENSTRA, A. RINNOY KAN. 1980. Deterministic production planning: algorithms and complexity. *Management Science* **26** 669–679.

GALLEGO, G., G. VAN RYZIN. 1994. Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Management Science* **40** 999–1020.

Gallego, G., D. Simchi-Levi 1996. On the Effectiveness of Direct Shipping Strategy for the One-Warehouse Multi-Retailer R-Systems. *Management Science* **36** 240–248.

Gallego, G., G. van Ryzin. 1997. A multiproduct dynamic pricing problem and its application to network yield management. *Operations Research* **45** 24–41.

Gavish, B., R. Johnson. 1990. A fully polynomial approximation scheme for single-product scheduling in a finite capacity facility. *Operations Research* **38** 70–83.

Gilbert, S. 1999. Coordination of Pricing and Production Planning for Constant Priced Goods. *European Journal of Operational Research* **114** 330–337.

Glynn, P. 1990. Diffusion approximations. In D. Heyman, M. Sobel (Edts): Stochastic Model, Vol. 2 of Handbooks in OR & MS, 145–198, North-Holland, Amsterdam.

Graves, S.. 1982. Using Lagrangean Techniques to solve Hierarchical Production Planning Problems. *Management Science* **28** 260–275.

Gray, R.. 1990. *Entropy and Information Theory*. Springer Verlag, Berlin.

Haimovich, M., A. Rinnooy Kan.1985. Bounds and Heuristics for Capacitated Routing Problems. *Mathematics of Operations Research* **10** 527–542.

Harrison, J.. 1996. The BIGSTEP approach to flow management in stochastic processing networks. In F. Kelly, S. Zachary, I. Ziedlins (Edts): *Stochastic Networks: Theory and Applications*, 57–90, Oxford University Pres, Oxford.

Harrison, J., J. van Mieghem. 1996. Dynamic control of brownian networks: State space collapse and equivalent workload formulations. *Annals of Applied Probability* **7** 747–771.

Horn, R., C. Johnson. 1994. *Matrix Analysis*. Cambridge University Press.

Hsu, V.. 2000. Dynamic Economic Lot Size Model with Perishable Inventory. *Management Science* **46** 1159–1169.

Kachani, S., G. Perakis, C. Simon. 2004. An MPEC Approach to Dynamic Pricing and

Demand Learning. Working Paper, MIT, Cambridge, MA.

Kalish, S. 1983. Monopolist Pricing with Dynamic Demand and Production Cost. *Marketing Science* **2** 135–160.

Karp, R. 1977. Probabilistic Analysis of Partitioning Algorithms for the Traveling Salesman Problem. *Mathematics of Operations Research* **2** 209–224.

Karni, R., Y. Roll. 1982. A heuristic algorithm for the multi-item lotsizing problem with capacity constraints. *AIIE Trans.* **13** 249–256.

Kashyap, A.K. 1995. Sticky Prices: New Evidence from Retail Catalogs. *Quarterly Journal of Economics* **110** 245–274.

Kleywegt, A. 2001. An optimal control problem of dynamic pricing. Working Paper, Georgia Institute of Technology.

Kuik, R., M. Solomon, L. van Wassenhove. 1994. Batching decisions: structures and models. *European Journal Of Operational Research* **18** 213–263.

Kunreuther, H., L. Schrage 1973. Joint Pricing and Inventory Decisions for Constant Priced Items. *Management Science* **19** 732–738.

Lambrecht, M., J. Vander Eecken. 1978. A capacity constrained single-facility dynamic lot-size model. *European Journal Of Operational Research* **2** 132–136.

Lautenbacher, C., S. Stidham. 1999. The underlying Markov Decision Process in the single-leg airline yield-management problem. *Transportation Science* **27** 136–146.

Leung, J., T. Magnanti, R. Vachani. 1989. Facets and algorithms for capacitated lot sizing. *Mathematical Programming* **45** 331–359.

Lee, T., M. Hersh. 1993. A model for dynamic airline seat inventory control with multiple seat bookings. *Transportation Science* **27** 252–265.

Lin, G., Y. Lu, D. Yao. 2003. The stockastic knapsack revisited: structure, switch-over policies and fluid-scale asymptotic optimality. Preprint, Columbia University.

Maes, J., L. van Wassenhove. 1986. Multi Item Single Level Capacitated Dynamic Lot-sizing Heuristics: A Computational Comparison (Part I: Static Case; Part II: Rolling Horizon). *IIE Transactions* **18** 114–129.

MAES, J., L. VAN WASSENHOVE. 1988. Multi-Item Single-Level Capacitated Dynamic Lot-Sizing Heuristics: A General Review. *Journal of the Operational Research Society* **39** 991–1004.

MAGLARAS, C. 2000. Discrete-review policies for scheduling stochastic networks: Trajectory tracking and fluid-scale asymptotic optimality. *Annals of Applied Probability* **10** 897–929.

MAGLARAS, C. 2003. Revenue management for a multi-class single server queue. Preprint, Columbia University.

MAGLARAS, C., J. MEISSNER. 2004. Dynamic Pricing Strategies for Multi-Product Revenue Management Problems. Preprint, Graduate School of Business, Columbia University, New York.

MARTIN, R.. 1987. Generating alternative mixed-integer programming models using variable redefinitions. *Operations Research* **35** 820–831.

MCGILL, J., G. VAN RYZIN. 1999. Revenue management: research overview and prospects. *Transportation Science* **33** 233–256.

MEMIPS. 1997. Model enhanced solution methods for integer programming software. *Esprit Project 20118*, Public report reference DR1.1.10

NAHMIAS, S.. 1982. Perishable Inventory Theory: A Review. *Operations Research* **30** 680–708.

NAHMIAS, S.. 1989. *Production and Operations Analysis.* Homewood, Boston.

NEMHAUSER, G.L., L.A. WOLSEY, M.L. FISHER. 1978. An Analysis of Approximations for Maximizing Submodular Set Functions. *Mathematical Programming* **14** 265–294.

NEMHAUSER, G., L. WOLSEY 1988. *Integer Programming and Combinatorial Optimization.* John Wiley & Sons, New York, NY, USA.

ORLIN, J. 1989. A faster strongly polynomial minimum cost flow algorithm. *Journal of the ACM* **36** 873–886.

PAMIPS. 1995. Development of parallel algorithms and software for mixed-integer programming in industrial scheduling. *Esprit Project 8755*, Public report reference DR4.3.5.

PERAKIS G., A. SOOD. 2003. Competitive Multi-Period Pricing with Fixed Inventories. Working Paper, MIT, Cambridge, MA.

PERAKIS G., A. SOOD. 2004. Multiperiod Pricing for Perishable Products: A Robust Optimization Approach. Working Paper, MIT, Cambridge, MA.

POCHET, Y.. 1988. Valid inequalities and separation for capacitated economic lot sizing. *Operations Research Letters* **7** 109–116

RAO, R.C., F.M. BASS. 1985. Competition, strategy, and price dynamics: A theoretical and empirical investigation. *Journal of Marketing Research* **22** 283–296.

RHEE, W. 1993. Inequalities for Bin Packing – II. *Mathematics of Operations Research* **18** 685–693.

RHEE, W., M. TALAGRAND. 1989. Concentration Inequalities for the K-Median Problem. *Mathematics of Operations Research* **14** 189–202.

RHEE, W. and M. TALAGRAND. 1987. Martingale Inequalities and NP-Complete Problems. *Mathematics of Operations Research* **12** 177–181.

ROMEIJN, H., D. ROMERO MORALES. 2001. A probabilistic analysis of the multi-period sourcing problem. *Discrete Applied Mathematics* **42** 301–328.

SALOMON, M. 1990. Deterministic lotsizing models for production planning. Ph.D. dissertation, Erasmus University.

SAMORODNITSKY, G., M. TAQQU. 1994. *Stable Non-Gaussian Random Processes.* Chapman Hall, Boca Raton, Florida.

SHAW, D., A. WAGELMANS. 1998. An algorithm for single-item capacitated economic lot sizing with piecewise linear production costs and general holding costs. *Management Science* **44** 831–838.

SHESHINSKI, E., Y. WEISS. 1977. Inflation and the cost of price adjustment. *Review of Economic Studies* **44** 287–303.

SHESHINSKI, E., Y. WEISS. 1983. Optimum pricing policy under stochastic inflation. *Review of Economic Studies* **50** 513–529.

STADTLER, H. 2003. Multilevel Lot Sizing with Setup Times and Multiple Constrained Resources: Internally Rolling Schedules with Lot-Sizing Windows. *Operations Research* **51** 487–502.

SUERIE, C., H. STADTLER. 2003. The Capacitated Lot-Sizing Problem with Linked Lot Sizes. *Management Science* **49** 1039–1054.

TALLURI, K., G. VAN RYZIN. 2004. Revenue management under a general discrete choice model of consumer behavior. *Management Science* **50** 15–33.

TALLURI, K., G. VAN RYZIN. 2004. *The theory and practice of revenue management.* Kluwer Academic Publishers.

TELLIS, G. 1988. The price sensitivity of competitive demand: A meta analysis of sales response models. *Journal of Marketing Research* **15** 331–341.

THOMAS, J. 1970. Price production decisions with deterministic demand. *Management Science* **16** 747–750.

TIROLE, J. 1988. *The Theory of Industrial Organization.* MIT Press, Cambridge, MA.

TRIGEIRO, W., L. THOMAS, J. MCCLAIN. 1989. Capacitated lot sizing with setup times. *Management Science* **35** 353–366.

VAN HOESEL, C., A. WAGELMANS. 1992. Economic Lot Sizing: An (n log n) Algorithm that runs in linear Time in the Wagner-Whitin Case. *Operations Research* **40** 145–156.

VAN HOESEL, C., A. WAGELMANS. 1996. An O(T^3) algorithm for the economic lot-sizing problem with constant capacities. *Management Science* **42** 142–150.

Van Hoesel, C., A. Wagelmans. 2001. Fully Polynomial Approximation Schemes for Single-Item capacitated lot-Sizing Problems. *Mathematics of Operations Research* **26** 339–357.

Van Nunen, J., J. Wessels. 1978. Multi item lot size determination and scheduling under capacity constraints. *European Journal of Operations Research* **2** 36–41.

Van Roy, T., L. Wolsey. 1987. Solving Mixed Integer Programming Problems using Automatic Reformulation. *Operations Research* **35** 45–48.

Vives, X. 2000. *Oligopoly Pricing.* MIT Press, Cambridge, MA.

Vulcano, G., G. van Ryzin, C. Maglaras. 2002. Optimal dynamic auctions for revenue management. *Management Science* **48** 1388–1407.

Wagelmans, A., C. van Hoesel, A. Kolen. 1992. Economic lot-sizing: an O(n log n)-algorithm that runs in linear time in the Wagner-Whitin case. *Operations Research* **40** 145–156.

Wagner, H.M. 1960. A postscript to 'dynamic problems of the firm.' *Naval Research Logistics Quarterly* **7** 7–12.

Wagner, H.M., T.M. Whitin. 1958. Dynamic Version of the Economic Lot Size Model. *Management Science* **5** 89–96.

Wagner, H.M., T.M. Whitin. 1959. Comments on 'Economic Lot Sizes with Seasonal Demand.' *Operations Research* **7** 394.

Zhao, W., Y. Zheng. 2000. Optimal dynamic pricing for perishable assets with nonhomogeneous demand. *Management Science* **46** 375–388.

Zipkin, P. 2000. *Foundations of Inventory Management.* McGraw Hill, New York.