Revenue Management through Dynamic Cross-Selling in E-commerce Retailing*

Serguei Netessine† Sergei Savin‡, Wenqiang Xiao§,
The Wharton School Graduate School of Business
University of Pennsylvania Columbia University

April 2004, Revised November 2004

Abstract

We consider the problem of dynamically cross-selling products (e.g., books) or services (e.g., travel reservations) in the e-commerce setting. In particular, we look at a company that faces a stream of stochastic customer arrivals and may offer each customer a choice between the requested product and a package containing the requested product as well as another product, "packaging complement". Given consumer preferences and product inventories, two issues are analyzed: (1) how to select packaging complements and (2) how to price product packages to maximize profits.

We formulate the cross-selling problem as a stochastic dynamic program blended with combinatorial optimization. We demonstrate the state-dependent and dynamic nature of the optimal package selection problem and derive structural properties of the dynamic pricing problem. In particular, we focus on two practical business settings: with (the Emergency Replenishment model) and without (the Lost Sales model) the possibility of inventory replenishment in the case of a product stock-out. For the Emergency Replenishment model, we establish that the problem is separable in the initial inventory of all products and hence the dimensionality of the dynamic program can be significantly reduced. For both models several packaging/pricing heuristics are suggested and their effectiveness is tested numerically.

*We are grateful to the Departmental Editor, Associate Editor and two anonymous referees for suggestions that helped improve the paper.

†Email: netessine@wharton.upenn.edu

‡Email: svs30@columbia.edu

§Email: wx2004@columbia.edu

1 Introduction

The development of the Internet has provided significant opportunities for new business models and retailing practices partly because massive amounts of click-stream data have become available to companies and the implementation of many decisions in real time has become virtually costless. For example, Internet companies can change prices dynamically and observe consumer response. Additionally, promotional decisions (discounts, rebates, etc.) can be made dynamically. These new opportunities pose challenges to researchers in various areas of business and economics: decisions that were previously made in a static environment must now take into account the dynamic potential of Internet retailing.

One practice that has recently been widely adopted on the Internet is dynamic cross-selling. For example, an attempt to buy most books on Amazon.com will generate a suggestion to buy a package of two or more books[1]. We screened the top 100 books from Amazon.com's best-selling list for September 17, 2003 to identify configurations of cross-selling suggestions (see Table 1 for results). As is evident from this sample, only a small fraction of books (7%) was not offered in a package with any other book. Further evidence from a recent survey by The E-tailing Group Inc. [11] suggests that 62% of the top 100 Internet retailers utilize various forms of cross-selling. Other prominent examples of dynamic cross-selling are found among travel-related Internet companies that term this practice "dynamic packaging" [2]. All these observations underscore the importance of understanding the trade-offs involved in dynamic cross-selling decisions and the need to quantify the benefits from making these decisions optimally.

Cross-selling is widely acknowledged as an important part of Customer Relationship Management (CRM). However, the implementation of *dynamic* cross-selling on the Internet poses several challenges. For example, the choice of products to cross-sell must be made by the "software" on the basis of information about product inventories and customer preferences rather than by a sales associate who can ask additional questions. That is, *dynamic* cross-selling must be performed in response to every customer's purchase attempt rather than using pre-set static rules. To see challenges associated with implementing dynamic cross-selling, note that cross-selling packages are often offered at a discount so that a package of products would generate lower profit margins. Hence, if inventory for one of the products in the package is low, it might be more profitable for a company to sell products individually because there is a good chance that the product will be sold later at full price[2]. As a result, the current inventory situation must be incorporated into cross-selling decisions. The obvious complexity of the decisions involved in dynamic cross-selling on the Internet has resulted in the need for specialized software packages. Some examples of companies offering such software are Netperceptions (focusing on physical goods) and PROS (focusing on travel-related services). At the same time, academic research on dynamic cross-selling is essentially non-existent. The goal of this paper is to fill this void by providing a deeper understanding of the dynamic cross-selling problem by building the framework for its analysis, outlining the trade-offs involved, obtaining structural results and deriving efficient solution methods.

We begin by proposing a novel modeling framework for the dynamic cross-selling problem in which com-

[1] At the time of this writing, Amazon.com did not offer discounts on book packages.

[2] In this respect, the situation with selling a product as part of a package at a discount now or selling products individually later is similar to the standard yield management problem that is common for airlines.

binatorial optimization (package selection) is blended with the stochastic dynamic program (package pricing) over a finite time horizon. The time horizon is subdivided into smaller decision epochs corresponding to packaging and pricing decisions that are made more often than inventory replenishments. Customer arrivals are modeled as a stochastic, discrete-time process. Each customer attempting to buy one product (which we call *first-choice*) is offered a package of this product with another one (which we call *packaging complement*) and can (with some probability that depends on the price of the package) choose to buy a package. We demonstrate the dynamic and state-dependent nature of the optimal package selection problem and concentrate on obtaining structural properties for the dynamic pricing problem. The combinatorial problem of optimal package selection is later re-visited using heuristic approaches.

Two important business settings are further analyzed. In the first setting, the firm has an opportunity to procure an out-of-stock product at an extra cost (the Emergency Replenishment model, hereafter the ER model). For this model we show that, under any static packaging scheme, the value function is separable in the initial inventory levels of all products and hence the dimensionality of the dynamic program (hereafter, the DP) can be greatly reduced. We refer to this result as a *decomposition* property. We also show that the value function is non-decreasing concave in the initial inventory levels of all products and the optimal price of the package is a non-increasing function of time and of the inventory of the packaging complement. Interestingly, the optimal package price in this case is independent of the inventory of the first-choice product. In the second setting (the Lost Sales model, hereafter the LS model), product inventory cannot be replenished and the customer's request is simply denied if the product is out of stock. We show (through counterexamples) that few of the properties of the ER model continue to hold in this case. Most importantly, the value function is no longer separable in product inventories: in fact, for the two-product case, we show that the value function is supermodular. Moreover, the value function is generally not concave in product inventories. Under the arbitrary static product packaging rule, upper and a lower bounds for the value function are derived and used to identify settings in which the revenue function is relatively insensitive to the choice of a particular static packaging scheme. Finally, for both the ER and the LS models we suggest several heuristic solution approaches and test them numerically. We comment on the comparative advantages and robustness of different heuristics.

To summarize, this paper makes three main contributions. First, we identify dynamic cross-selling as an application of revenue management/dynamic pricing but one that involves an additional combinatorial optimization to select the packaging complement. We also propose a novel modeling framework to analyze the cross-selling (packaging/pricing) problem. We derive the structural properties of the dynamic pricing problem under any static packaging scheme, most notably the decomposition property of the objective function in the ER model. Finally, we explore the structural properties of these models to obtain efficient dynamic packaging and pricing heuristics. Two of the proposed heuristics, the so-called "two-stage" approach and the "depletion rate" heuristic, are shown numerically to have a near-optimal performance over a wide range of problem parameters. The rest of the paper is organized as follows. In the remainder of this section we survey related literature. In Section 2 we state modeling assumptions and formulate the problem. Sections 3 and 4 contain analysis of the ER and LS models, respectively. Section 5 concludes with a discussion of our results.

1.1 Related literature

In recent years the practice of cross-selling and the related practice of up-selling are receiving more and more coverage in trade publications (see, for example, Feldman [14], Peters [27]), while academic research on cross-selling is very sparse: the works of Nash and Sterna-Karwat [24] describing the application of DEA methodology to cross-sell financial services and of Kamakura et al. [19] utilizing customer databases to identify opportunities for cross-selling are the only papers addressing this issue. To the best of our knowledge, there are no papers that either model or analyze dynamic cross-selling that involves the dynamic pricing of packages and dynamic package selection.

The dynamic package pricing decision is closely related to the large stream of operations literature on dynamic pricing and revenue management (see McGill and van Ryzin [21] and Bitran and Caldentey [8] for surveys). In another recent survey, Elmaghraby and Keskinocak [9] provide a thorough discussion of the need to incorporate customized pricing considerations (of which cross-selling is one example) into dynamic pricing models; they cite many applications and notice that there are no papers in the extant literature addressing this problem. Other publications in this stream typically consider the dynamic pricing of a single product and hence do not address cross-selling issues (see, for example, Gallego and van Ryzin [16] and Aviv and Pazgal [6]). Monahan et al. [23] study the problem of pricing a single product over multiple time periods after a single inventory decision. Their setting is similar to ours in that the product is sold over multiple periods without further replenishments. However, they utilize a specific form of demand function: random shock is multiplicative and price dependence is iso-elastic, which allows them to find optimal prices and inventory in a closed form. An alternative setting is the one in which demand from several types of customers can be satisfied with the same capacity (see Maglaras and Meissner [20]). In our paper, demand from a customer may involve several types of inventory simultaneously. In this respect, we believe that the papers most relevant to our work are those analyzing the dynamic revenue management of *multiple* products, a rather sparsely populated area of literature (see Gallego and van Ryzin [17], Zhang and Cooper [34]).

On the Internet, the selection of the packaging complement can be based both on the customer information acquired during previous transactions and/or on the customer profile. To process the significant amounts of data necessary for making decisions, companies utilize data mining techniques (see Padmanabhan and Tuzhilin [26] for a survey of relevant methodologies). This approach is generally termed "personalization" (see Adomavicius and Tuzhilin [5]). Papers dealing with personalization techniques typically focus on generating a recommendation that ensures the best match between the customer and the product regardless of the potential impact on the firm's profitability. On the contrary, we assume that the company uses data mining to generate the probability distribution over customers' purchase preferences, which in turn is used to cross-sell products to maximize the firm's profit. Hence, it is quite possible that the package offered to the customer is not the best possible match from the personalization perspective.

A practice related to cross-selling is product *bundling,* which also attempts to sell a package consisting of several products rather that selling just a single product. However, bundling decisions (involving what to bundle and how to price bundles) are static and made before customer arrival; this differs from cross-selling, which is done only *after* a customer declares a desire to buy something. Nevertheless, dynamic cross-selling can be loosely interpreted as dynamic bundling, so we briefly review the related literature. Economists were the first to analyze the concept of bundling (see Stigler [31]). The cornerstone for many subsequent papers on product bundling was the seminal work of Adams and Yellen [4], who formulated a model with a firm selling two products as well as a bundle of these two products and demonstrated the benefits arising from bundling. In the operations literature Hanson and Martin [18] were, perhaps, the first to address the problem of optimally pricing bundles of products using a mathematical programming approach in a static model. Ernst and Kouvelis [10] study the issue of how much of each individual product as well as how many bundles to stock while pricing and bundling decisions are exogenous to the model. There is a wealth of marketing literature considering bundling with a particular emphasis on static pricing. Rao [29] and Stremersch and Tellis [32] review this stream. In addition, an extensive collection of papers on this topic can be found in Fuerderer et al. [15]. In all of these papers, a commitment to sell bundles is made prior to demand arrival and thereafter prices/bundles cannot be altered. Hence, the underlying models are static rather than dynamic. We are not aware of any papers analyzing dynamic product bundling.

2 Modeling dynamic cross-selling decisions

We consider an environment in which an online retail company is selling a group of m products. We assume that all products in the group target similar market segments and can therefore potentially be cross-sold with each other (see examples in the introduction). We further assume that the planning horizon is finite (representing the time between two inventory replenishments) and is separated into N decision epochs. At the beginning of each decision epoch the company observes the inventory of each product and makes cross-selling (packaging and pricing) decisions. In the online environment packaging and pricing decisions can be made as frequently as necessary, so that the length of each decision epoch can be made short compared to typical customer interarrival time. Consequently, we assume that within each epoch there is at most one customer arrival. We denote by λ_i, $i = 1, ..., m$, the probability that during any decision epoch class i customer arrives and requests one unit of i-th product ($\sum_{i=1}^{m} \lambda_i < 1$ to reflect the possibility that there is no customer arrival in a particular decision epoch). Alternatively, one can also define the probability that arriving customers do not buy a product, but the problem can be easily re-formulated to focus on customers that are willing to make a purchase.[3]

Upon requesting product i at a fixed price[4] p_i, a class i customer receives an offer of a "product i-product j" package (for some j) at a price p_{ij}. We assume that only one package of only two products is offered to the customer, which is consistent with the practice of some companies. Our experiment with Amazon.com's best-seller list (see Table 1) shows that at most one packaging complement is used for each available product. We assume that class i customers have a reservation price for "i-j" package distributed according to a cdf $F_{ij}(\cdot)$

[3] The implicit assumption here is that the probability of purchasing a product is time-invariant while in practice this probability may be affected by the firm's decisions. E.g., intensive cross-selling of product j may result in fewer customers buying this product in the future.

[4] In our analysis, we consider individual product price to be fixed but package price to be dynamic. This approach is often justified in practice, since some Internet retailers try to avoid dynamically changing prices of individual products for fear of antagonizing customers (see the example of Amazon.com's experience [1]). Alternatively, our analysis can complement the traditional single-product revenue management literature which studies the effects of dynamically changing prices for individual products (see McGill and van Ryzin [21] for a thorough survey).

with non-negative support. Consequently, a class i customer buys the package at a price p_{ij} with probability $\overline{F}_{ij}(p_{ij}) = 1 - F_{ij}(p_{ij})$, and buys only product i at price p_i with probability $F_{ij}(p_{ij})$[5]. We assume that $\bar{F}_{ij}(x)(x-y)$ is a unimodal function of x for any nonnegative constant y. This assumption holds for a wide class of distribution functions and is also a standard assumption often made in the pricing literature, see Ziya et al. [37] for discussion and references.

The issue of estimating $F_{ij}(p_{ij})$ is an important one that merits a separate study. For our purposes, we simply assume that $F_{ij}(p_{ij})$ is obtained by analyzing customer shopping behavior using data mining techniques. Two examples of approaches that can potentially be utilized in this case are found in Moe and Fader [22] and Bertsimas et al. [7]. Since we do not assume any particular functional form for these probability functions, any approach can be accommodated by our model. Note also that we classify customers based on the product they request, while in practice a company may have additional customer information that would allow for fine-tuning probability estimation further. In the extreme, each arriving customer could be treated as a separate class. Our model can be extended to incorporate such a possibility (e.g., let k be the customer index with each customer having a unique k so that $\lambda_i = \sum_k \lambda_i^k$ where λ_i^k represents customer k arriving to request book i) but at a cost of additional notation that would make exposition less tractable. To support our approach, there is some evidence that for a cross-selling recommendation, the first-choice product is more relevant than the customer profile ([3]).

In each decision epoch a company needs to decide, for each product i, which product j should be "packaged" with it and what price should be established for such a package. Suppose that at the beginning of the n-th decision epoch ($n = 1, ..., N$) the inventory levels of products are given by vector $\mathbf{I} = (I_1, ..., I_m)$ and let $V_n(\mathbf{I})$ be the optimal expected revenue accrued from that moment until the end of the planning horizon. Define $\mathbf{e}_i$, $i = 1, ..., m$ as an m-dimensional vector whose components are $(\mathbf{e}_i)_k = 1$ if $k = i$ and 0 otherwise. We are interested in selecting a sequence of packaging and pricing decisions in order to maximize $V_1(\mathbf{I})$. The functional form of the Bellman equation for $V_n(\mathbf{I})$ depends on actual inventory levels. If there is at least one unit of inventory for each product ($I_i \geq 1$ for all $i = 1, ..., m$), the Bellman equation can be expressed as

$$\begin{aligned} V_n(\mathbf{I}) &= \sum_{i=1}^{m} \lambda_i \max_{j \neq i} \left(\max_{p_{ij}} \left(F_{ij}(p_{ij}) \left(p_i + V_{n+1}(\mathbf{I} - \mathbf{e}_i) \right) + \overline{F}_{ij}(p_{ij}) \left(p_{ij} + V_{n+1}(\mathbf{I} - \mathbf{e}_i - \mathbf{e}_j) \right) \right) \right) \\ &\quad + \left(1 - \sum_{i=1}^{m} \lambda_i \right) V_{n+1}(\mathbf{I}). \end{aligned} \tag{1}$$

At the heart of the recursion (1) are two maximization operators reflecting the dynamic packaging and pricing decisions. The "outer" maximization selects the "best" product j to be packaged with product i in the case of a class i customer arrival, while the "inner" maximization selects the "best" price for each "i-j" package. Without loss of generality, we assume that the inventory left in stock at the end of the planning horizon has no salvage value and supplement (1) by the "end-of-horizon" condition $V_{N+1}(\mathbf{I}) = 0$.

[5] A more general way to model this problem is to allow the consumer to choose simultaneously among products i, j and a package ij. In this case the probability of selecting each of these options might depend on p_i, p_j and p_{ij}. However, our approach is plausible when product j has low value without product i (e.g., hotel reservation without air ticket) and it also matches practically observed applications of cross-selling relatively well.

Clearly, the dynamic packaging problem is quite complex since it involves solving a large-scale combinatorial optimization superimposed on the stochastic DP. In fact, there are very few papers that successfully derive structural results for such problems. To illustrate the complexity of this problem, we demonstrate that in the settings in which the number of products is three or more, the choice of the best packaging complement is non-trivial. In particular, the optimal packaging decisions may be state-dependent as well as dynamic. Figure 1 illustrates an example of the optimal package selection in the three-product case for the following problem parameters: $\overline{F}_{ij}(p_{ij}) = ((p_i + p_j - p_{ij})/p_j)^{\beta}$ defined over $[p_i,\ p_i + p_j]$, $i, j = 1, 2, 3$, with $N = 10, \beta = 1$, $p_1 = p_3 = 1$, $p_2 = 1.5$, and $\lambda_1 = \lambda_2 = \lambda_3 = 0.3$. This form of the reservation function is reasonable because $\overline{F}_{ij}(p_i) = 1$ and $\overline{F}_{ij}(p_i + p_j) = 0$. Figure 1a shows how the optimal packaging at the beginning of the planning horizon ($n = 1$) changes with the inventory of product 1: the packaging complement of product 2 oscillates between products 1 and 3 in a rather unusual fashion. Such "non-monotone" behavior hints at a particularly complex structural form of the optimal value function even for a relatively simple three-product case. Figure 1b illustrates the dynamic nature of the optimal packaging for the same state of the system: in this example, the shrinking of the remaining time horizon forces product 1 to change its packaging complement from product 3 to product 2.

As the above example demonstrates, packaging decisions for more than two products are quite complex. This interaction between dynamic packaging and pricing significantly complicates the analysis of the general DP formulation. Below we begin our analysis by separating packaging and pricing decisions and focusing on pricing first by assuming that packaging is static so that the packaging complement for each product is fixed and does not change with the state of the product inventory or with time. This policy may be used by a company intentionally to ensure that the packaging complement closely matches the first-choice product. For example, the packaging complement can be assigned according to the closest match based on a purchase history and/or on customer profile (of course, the packaging decision becomes trivial with only two products). Thereafter we propose several (heuristic) ways to solve the dynamic packaging problem and will test them numerically.

As mentioned above, the DP recursion (1) is applicable as long as there is at least one unit of inventory left for each product. Hence, when the company runs out of inventory for one or more products, (1) needs to be adjusted. Below we consider two alternative policies describing the company's response to the request for a product with zero inventory level. The first alternative, designated as the Emergency Replenishment model (hereafter, ER), allows the company to procure a "missing" item i at an additional cost b_i (to ensure that it is always profitable to use ER in the case of a stockout we assume that $b_i \leq p_i$). In Section 3 we prove the decomposition property of the optimal value function for the ER model under any static packaging scheme, enabling an efficient solution method for the multiproduct ER model. Under the second alternative, which we call the Lost Sales model (hereafter, LS), the customer request is simply rejected. The LS model (analyzed in Section 4) lacks such a decomposition property. Consequently, as the number of products grows, this model becomes increasingly difficult to solve. For both models we propose heuristic approaches and analyze their performance.

3 The Emergency Replenishment model

Under the ER model, the company has an opportunity to procure additional product inventory at an extra cost. This assumption may be appropriate in settings with many physical products sold over the Internet. For example, in case the retailer stocks out, products could be drop-shipped from the wholesaler directly to the customers (i.e., the order is passed on to the wholesaler/distributor who performs the fulfillment at an extra cost, see Netessine and Rudi [25] for details on drop-shipping arrangements and practical examples). In this case, b_i may represent the drop-shipping mark-up and/or additional shipping costs. Alternatively, the retailer may re-order the missing item from the wholesaler and, once it arrives, utilize a faster delivery mode to compensate for the delay (e.g., use next-day instead of the regular shipping service). In this situation, b_i may represent the extra transportation cost.

It is convenient to introduce sets of indices A_n to denote products that have at least one unit of inventory at the beginning of the n-th decision epoch. Under the ER model the appropriate generalization of (1) is given by

$$V_n\left(\mathbf{I}\right) = \sum_{i \in A_n} \lambda_i \max\left(H_i^n, J_i^n\right) + \sum_{i \notin A_n} \lambda_i \max\left(\overline{H}_i^n, \overline{J}_i^n\right) + \left(1 - \sum_{i=1}^{m} \lambda_i\right) V_{n+1}\left(\mathbf{I}\right), \tag{2}$$

with

$$H_i^n = \max_{j \neq i,\, j \in A_n} \left(\max_{p_{ij}} \left(F_{ij}\left(p_{ij}\right)\left(p_i + V_{n+1}\left(\mathbf{I} - \mathbf{e}_i\right)\right) + \overline{F}_{ij}\left(p_{ij}\right)\left(p_{ij} + V_{n+1}\left(\mathbf{I} - \mathbf{e}_i - \mathbf{e}_j\right)\right)\right)\right), \tag{3}$$

$$J_i^n = \max_{j \neq i,\, j \notin A_n} \left(\max_{p_{ij}} \left(F_{ij}\left(p_{ij}\right)\left(p_i + V_{n+1}\left(\mathbf{I} - \mathbf{e}_i\right)\right) + \overline{F}_{ij}\left(p_{ij}\right)\left(p_{ij} - b_j + V_{n+1}\left(\mathbf{I} - \mathbf{e}_i\right)\right)\right)\right), \tag{4}$$

$$\overline{H}_i^n = -b_i + \max_{j \neq i,\, j \in A_n} \left(\max_{p_{ij}} \left(F_{ij}\left(p_{ij}\right)\left(p_i + V_{n+1}\left(\mathbf{I}\right)\right) + \overline{F}_{ij}\left(p_{ij}\right)\left(p_{ij} + V_{n+1}\left(\mathbf{I} - \mathbf{e}_j\right)\right)\right)\right), \tag{5}$$

$$\overline{J}_i^n = -b_i + \max_{j \neq i,\, j \notin A_n} \left(\max_{p_{ij}} \left(F_{ij}\left(p_{ij}\right)\left(p_i + V_{n+1}\left(\mathbf{I}\right)\right) + \overline{F}_{ij}\left(p_{ij}\right)\left(p_{ij} - b_j + V_{n+1}\left(\mathbf{I}\right)\right)\right)\right). \tag{6}$$

Under the ER model all product indices remain "active" due to the possibility of outsourcing the "missing" product. Equations (3)-(6) reflect four distinct packaging possibilities that potentially exist under the ER model. For example, if an "in-stock" product is requested, it can be matched with another "in-stock" product (3), or with an "out-of-stock" product (4), with corresponding penalty. Similarly, if an "out-of-stock" product is requested, it will be procured at an extra cost, and it can be packaged with an "in-stock" product (5), or with another "out-of-stock" product (6).

3.1 Dynamic pricing under static packaging

Let $j(i)$ be the index of the product offered in a package with product i when a class i customer arrives. Under the static packaging utilized in this section, $j(i)$ is fixed for each $i = 1, 2, ..., m$. Denote by $E(i)$ the set of products for each of which product i is offered as a packaging complement, $E(i) = \{k | j(k) = i\}$. For the ER model with static packaging, (2) can be simplified to:

$$V_n(\mathbf{I}) = \sum_{i=1}^{m} \lambda_i \max_{p_{i,j(i)}} Y_{i,j(i)}^n(p_{i,j(i)}) + \left(1 - \sum_{i=1}^{m} \lambda_i\right) V_{n+1}(\mathbf{I}), \tag{7}$$

with

$$Y_{ij}^n = \left\{ \begin{array}{ll} F_{ij}(p_{ij})\left(p_i + V_{n+1}(\mathbf{I} - \mathbf{e}_i)\right) + \bar{F}_{ij}(p_{ij})\left(p_{ij} + V_{n+1}(\mathbf{I} - \mathbf{e}_i - \mathbf{e}_j)\right) & \text{if} \quad i \in A_n, j \in A_n \\ F_{ij}(p_{ij})\left(p_i + V_{n+1}(\mathbf{I} - \mathbf{e}_i)\right) + \bar{F}_{ij}(p_{ij})\left(p_{ij} - b_j + V_{n+1}(\mathbf{I} - \mathbf{e}_i)\right) & \text{if} \quad i \in A_n, j \notin A_n \\ -b_i + F_{ij}(p_{ij})\left(p_i + V_{n+1}(\mathbf{I})\right) + \bar{F}_{ij}(p_{ij})\left(p_{ij} + V_{n+1}(\mathbf{I} - \mathbf{e}_j)\right) & \text{if} \quad i \notin A_n, j \in A_n \\ -b_i + F_{ij}(p_{ij})\left(p_i + V_{n+1}(\mathbf{I})\right) + \bar{F}_{ij}(p_{ij})\left(p_{ij} - b_j + V_{n+1}(\mathbf{I})\right) & \text{if} \quad i \notin A_n, j \notin A_n \end{array} \right\}. \tag{8}$$

It turns out that under the assumption of static packaging in the ER model, the m-dimensional DP (7) can be decomposed into m one-dimensional DPs. Such decomposition greatly reduces the computational effort necessary for solving (7). Below we present the decomposition property of the ER model under static packaging. We start by introducing the following definition:

Definition 1

For $i = 1, ..., m$, let $G_n^i(I_i)$ be the function satisfying the following recursive formulae:

$$\begin{aligned} G_n^i(I_i) &= \lambda_i \left(p_i + G_{n+1}^i(I_i - 1)\right) + \left(1 - \lambda_i - \sum_{j \in E(i)} \lambda_j\right) G_{n+1}^i(I_i) \\ &\quad + \sum_{j \in E(i)} \lambda_j \max_{p_{ji}} \left(F_{ji}(p_{ji}) G_{n+1}^i(I_i) + \bar{F}_{ji}(p_{ji})\left(p_{ji} - p_j + G_{n+1}^i(I_i - 1)\right)\right), \end{aligned} \tag{9}$$

for $I_i \geq 1$ and $n = 1, ..., N$, and

$$\begin{aligned} G_n^i(0) &= \lambda_i \left(p_i + G_{n+1}^i(0) - b_i\right) + \left(1 - \lambda_i - \sum_{j \in E(i)} \lambda_j\right) G_{n+1}^i(0) \\ &\quad + \sum_{j \in E(i)} \lambda_j \max_{p_{ji}} \left(F_{ji}(p_{ji}) G_{n+1}^i(0) + \bar{F}_{ji}(p_{ji})\left(p_{ji} - p_j + G_{n+1}^i(0) - b_i\right)\right), \end{aligned} \tag{10}$$

while $G_{N+1}^i(I_i) = 0$.

Note that (10) can be expressed in closed form as follows:

$$G_n^i(0) = (N + 1 - n)\left(\lambda_i (p_i - b_i) + \sum_{j \in E(i)} \lambda_j \max_{p_{ji}} \left(\bar{F}_{ji}(p_{ji})(p_{ji} - p_j - b_i)\right)\right). \tag{11}$$

The following result states the decomposition property of the optimal revenue function:

Proposition 1

In each decision epoch n, *the optimal expected revenue function described by* (7) *can be separated into* m *parts, with each depending only on the inventory level of a single product*:

$$V_n(\mathbf{I}) = \sum\nolimits_{i=1}^{m} G_n^i(I_i),\ \ n = 1, ..., N+1,$$

where $G_n^i(I_i)$ *is defined by* (9)-(10).

We note that $G_n^i(I_i)$ can be interpreted as the expected revenue generated by the inventory of product i alone. Indeed, as (9) suggests, product i's inventory can be changed either directly through sales to class i customers or indirectly through sales to class $j \neq i$ customers as part of a "j-i" package. In the first case, the revenue generated per unit of product i sold is p_i, while in the second case it is $p_{ji} - p_j$.

Proposition 1 identifies an important property of the ER model: if a company can procure "out-of-stock" items, an m-dimensional DP problem expressed by (7)-(8) is replaced by m one-dimensional DP problems (9)-(10). This result greatly simplifies the computational effort needed to determine the optimal pricing policy. The decomposition result is somewhat surprising: one would expect that the opportunity to procure missing items would complicate, not simplify, the problem[6]. We note that the result of Proposition 1 can be rationalized as follows: it can be shown that under static packaging, the value function of the "main" dynamic program (1) is decomposable into the sum of single-product functions. The same is true for the "boundary" conditions in (7) relating to the terms in (8) with i or j outside of A_n - and the decomposition "pieces" (single-product value functions) are identical in both cases. As it turns out, such matching of decomposition "pieces" does not hold for the LS model, and neither does the decomposition result of Proposition 1. At the same time, as we will show later, the decomposition of (1) can still be applied heuristically in the LS case yielding good performance and hence has applicability beyond the ER model.

Next, we use the decomposition property of the revenue function to derive other structural properties. Let $p^*_{ij}(\mathbf{I}, n)$ be the optimal package price in decision epoch n given that the inventory levels of products are $\mathbf{I}$.

Proposition 2

a) $G_n^i(I_i)$ *is a non-decreasing concave function of* I_i *for* $i = 1, ..., m$, *and* $n = 1, ..., N$.

b) The optimal package price $p^*_{ij}(\mathbf{I}, n)$ *is non-increasing in* n, I_j, *and is independent of* I_k *for* $k \neq j$, *for* $i, j = 1, ..., m$ and $i \neq j$.

The results of Proposition 2 are established using the induction over the time index n. We note that concavity of revenue functions $G_n^i(I_i)$ facilitates the connection between the cross-selling problem we consider and the inventory ordering problem the online retailer may face. In particular, concave revenue functions can be included, along with convex inventory holding costs, in the generalized inventory ordering problem. Thus, in the absence of joint inventory ordering costs, the optimal inventory policy for each item remains to be an (s, S) policy (see Scarf [30], Veinott [33], and Zheng [35]). For situations in which joint ordering costs are significant, the (s, S) policy is no longer optimal. For such cases, however, Zheng [36] proves that a modified

[6] Interestingly, a similar observation is made by Plambeck and Ward [28] in the context of inventory procurement for assemble-to-order systems, a problem that is entirely different from ours (e.g., there are no pricing decisions involved in their model).

(s, S) policy, so-called (s, c, S) policy, is optimal in the decentralized system. Federgruen et al. [12] provide an iterative procedure to repeatedly update the (s, c, S) values for each item until the optimum is reached.

The second part of Proposition 2 indicates that the optimal package price is non-decreasing in "time-to-go" and non-increasing in the inventory of the packaged product. The first effect is similar to the one found in single-product dynamic pricing problems (see Gallego and van Ryzin [16]). The second effect is quite intuitive and shows the linkage between product availability and price that we hypothesized in the introduction. These findings are in some ways unsurprising since they also hold in simpler settings when dynamically pricing a single product. Nevertheless, it is reassuring that under a detailed and explicit model of cross-selling, this property is still achieved. Finally, the inventory of the first-choice product does not affect package price due to the decomposition property of the value function.

3.2 Heuristic approaches to dynamic packaging and pricing

The optimal solution to the cross-selling problem (1) may be hard to obtain in real time in cases involving a large number of products. Even though the decomposition property can be applied to simplify the pricing problem, the combinatorial optimization aimed at finding the best packaging complements still poses a significant challenge. In this section we propose and test numerically several heuristic packaging-pricing approaches. First, we present the myopic heuristic $\mathbf{H}^M$, which ignores product inventories and time-to-go. To account for these factors, we consider two more sophisticated heuristic approaches. Heuristic $\mathbf{H}^D$ characterizes the optimal solution in the ER model with one-shot static and deterministic demand, whose value depends on packaging and pricing decisions. This solution can then be used dynamically at each point in time in response to changing inventory levels. The last heuristic $\mathbf{H}^T$ simplifies the solution to DP by assuming that there are no packaging decisions in any periods other than the current one.

3.2.1 Myopic packaging and pricing heuristic

Perhaps the simplest approach to pricing and packaging is to ignore the impact of product inventories. We denote the resulting myopic static packaging and pricing heuristic as $\mathbf{H}^M$. From (9) assuming that $G^i_{n+1}(I_i) = G^i_{n+1}(I_i - 1)$, we get

$$j^M(i) = \arg\max_{j \neq i} \overline{F}_{ij}(p^M_{ij})(p^M_{ij} - p_i),\ i = 1, ..., m, \tag{12}$$

where p^M_{ij} is defined by

$$p^M_{ij} = \arg\max_{p_{ij}} \left(\bar{F}_{ij}(p_{ij})(p_{ij} - p_i)\right). \tag{13}$$

The static packaging scheme defined by (12) myopically chooses the packaging complement to maximize the expected profit from selling a package. Note that (12) serves as the fundamental characteristic of class i customers' "propensity" to buy product j and represents the optimal price to be charged for the "$i - j$" package in the case of the infinite product j inventory. The clear advantage of the myopic approach is its implementational and computational simplicity: its application only requires $\mathrm{O}(m^2)$ comparisons. On the other hand, the myopic heuristic assumes that the marginal value of the inventory of the potential packaging complement is negligible – an assumption that can be especially inadequate in cases in which product inventories are constrained.

3.2.2 Static deterministic approximation

Consider a static deterministic approximation to the dynamic and stochastic ER model. We assume that the "one-shot" demand for each package and each single product is deterministic but influenced by packaging and pricing decisions. Specifically, we introduce the time parameter $\tau = N - n$, which plays the role of the effective time horizon in our static deterministic analysis. In particular, the total demand from type i customers is $\lambda_i \tau$. Packaging decisions are defined by the fraction of type i demand, $0 \leq q_{ij} \leq 1$, which receives an offer of the (i-j) package (recall that this is different from the (j-i) package). Note that these continuous variables represent a generalization of the packaging decisions introduced in Section 2. Such a generalization simplifies the analysis of the deterministic model we consider. Given any packaging decisions $\{q_{ij}\}$ and pricing decisions $\{p_{ij}\}$, the total demand for the "i-j" package is given by $\lambda_i \tau q_{ij} \overline{F}_{ij}(p_{ij})$, and the total demand for single product i is given by $\sum_{j \neq i} \lambda_i \tau q_{ij} F_{ij}(p_{ij})$, for $i, j = 1, 2, ..., m$ and $i \neq j$.

Our static deterministic approximation assumes that the demand for each package and each single product is the expected value of the corresponding demand arising in the stochastic setting, given that the same static packaging and pricing decisions are used. The objective is to maximize the total revenue minus the total emergency replenishment cost by selecting the optimal values of $\{q_{ij}\}$ and $\{p_{ij}\}$. This static problem, denoted as (**P1**)**,** can be formulated as follows:

$$\max_{q_{ij}, p_{ij}} \sum_{i=1}^{m} \sum_{j \neq i} \lambda_i \tau q_{ij} \overline{F}_{ij}(p_{ij}) p_{ij} + \sum_{i=1}^{m} \sum_{j \neq i} \lambda_i \tau q_{ij} F_{ij}(p_{ij}) p_i - \sum_{i=1}^{m} b_i \left(\lambda_i \tau + \sum\nolimits_{j \neq i} \lambda_j \tau q_{ji} \overline{F}_{ji}(p_{ji}) - I_i \right)^+$$
$$\text{s.t.} \sum\nolimits_{j \neq i} q_{ij} = 1, \text{ for } i = 1, 2, ..., m,$$
$$q_{ij} \geqslant 0, \text{ for } i \neq j,$$

where $x^+ = \max\{x, 0\}$. The first part in this objective function represents the total revenue collected from selling all "i-j" packages, the second part is the total revenue from selling individual products, and the third part is the purchase cost for ERs derived from the fact that the total demand for product i consists of the demand $\sum_{j \neq i} \lambda_i \tau q_{ij} F_{ij}(p_{ij})$ for individual product i, the demand $\sum_{j \neq i} \lambda_i \tau q_{ij} \overline{F}_{ij}(p_{ij})$ for all "i-j" packages, and the demand $\sum_{j \neq i} \lambda_j \tau q_{ji} \overline{F}_{ji}(p_{ji})$ for all "j-i" packages.

Note that (**P1**) belongs to the class of constrained optimization problems with non-differentiable objective functions, which, in general, are hard to cope with. Next we consider a modified problem (**P2**) that is equivalent to (**P1**) but is easier to solve. We then formulate the dual problem (**D2**) of the primal problem (**P2**), and show that the optimal solution to (**P2**) can be derived easily from the optimal solution to (**D2**), whose objective function is linear and whose constraint set is convex. Thus, many existing algorithms (e.g., the gradient method) can be applied to solve (**D2**), which is equivalent to solving (**P1**) .

By introducing auxiliary variables $\{y_i\}$ and performing some algebraic manipulations on the objective function, problem (**P1**) can be formulated as the following equivalent problem (with the objective value differing by a constant), denoted by (**P2**),

$$\max_{y_i, q_{ij}, p_{ij}} \sum_{i=1}^{m} \sum_{j \neq i} \lambda_j \tau q_{ji} \overline{F}_{ji}(p_{ji})(p_{ji} - p_j) - \sum_{i=1}^{m} b_i y_i$$

$$\text{s.t. } y_i \geq \lambda_i \tau + \textstyle\sum_{j \neq i} \lambda_j \tau q_{ji} \overline{F}_{ji}(p_{ji}) - I_i \text{ for } i = 1, 2, ..., m,$$

$$\textstyle\sum_{j \neq i} q_{ij} = 1, \text{ for } i = 1, 2, ..., m,$$

$$y_i \geq 0 \text{ for } i = 1, 2, ..., m, \; q_{ij} \geqslant 0, \text{ for } i \neq j.$$

Define $H_{ji}(x) = \max_{p_{ji}} \left(\bar{F}_{ji}(p_{ji})(p_{ji} - p_j - x) \right)$. The main results of this subsection are summarized in the following proposition:

Proposition 3

a) The dual problem (**D2**) *of the primal problem* (**P2**) *can be formulated as the following optimization problem with a convex set of constraints:*

$$\min_{\mu_i, v_i} \sum_{i=1}^{m} \left(\mu_i (I_i - \lambda_i \tau) + v_i \right)$$

$$\text{s.t. } \lambda_j \tau H_{ji}(\mu_i) \leq v_j \text{ for } i \neq j,$$

$$0 \leq \mu_i \leq b_i \text{ for } i = 1, 2, ..., m.$$

b) Given any optimal solution $\{\mu_i^*, v_i^*\}$ *to* (**D2**), *there exists a corresponding optimal solution* $\{\{p_{ij}^*\}, \{q_{ij}^*\}, \{y_i^*\}\}$ *to* (**P2**), *determined as follows:*

i) $p_{ji}^* = \arg\max_{p_{ji}} \left(\bar{F}_{ji}(p_{ji})(p_{ji} - p_j - \mu_i^*) \right)$ *for* $i \neq j$;

ii) $\{q_{ij}^*\}$ *and* $\{y_i^*\}$ *are the solutions to* (**P2**) *with* p_{ij} *replaced by* p_{ij}^*.

Part b i) of Proposition 3 indicates that the dual solution μ_i^* can be interpreted as the marginal value of an extra unit of product i. That is, if there is ample inventory, there is no need to make an emergency replenishment of product i and hence $\mu_i^* = 0$. On the other hand, if the inventory is scarce, then μ_i^* is equal to the ER cost b_i. In other words, μ_i^* can also serve as an availability indicator for product i. By the equivalence of (**P1**) and (**P2**), the optimal solution $\{p_{ij}^*, q_{ij}^*\}$ derived from part b) is also optimal for (**P1**). This optimal solution yields another dynamic cross-selling heuristic in which decisions depend on the current product inventory and time-to-go. In the dynamic stochastic environment of (1) this heuristic (which we denote as $\mathbf{H}^D$) is implemented as follows: at the beginning of each decision epoch n, (**P2**) is solved for $\tau = N - n$ using current product inventory levels. The obtained packaging ($\{q_{ij}^*\}$) and pricing ($\{p_{ij}^*\}$) solutions are used as follows: when a customer of class i arrives, a firm conducts a randomized "coin-flip" trial (with the probability of j-th outcome being q_{ij}^*); the realized outcome j is selected and the "$i-j$" package is offered at the price p_{ij}^*.

3.2.3 A two-stage heuristic

Suppose that at the beginning of the n-th decision epoch the current inventory vector is $\mathbf{I}$. Under the "two-stage" approach, we simplify the packaging/pricing problem by assuming that *there is no packaging* in periods $n+1$ through N. Note that the number of type i customers arriving in each epoch is a Bernoulli random variable with parameter λ_i and the number of type i customers arriving in all remaining $N-n$ epochs is a binomial random variable Λ_i with parameters $(\lambda_i,\ N-n)$.Then, the marginal value of an extra unit of product $j \neq i$ can be expressed as

$$\begin{aligned} V_{n+1}(\mathbf{I}-\mathbf{e}_i) - V_{n+1}(\mathbf{I}-\mathbf{e}_i-\mathbf{e}_j) &= b_j E\left(\min(\Lambda_j, I_j) - \min(\Lambda_j, I_j - 1)\right) \\ &= b_j \Pr(\Lambda_j \geqslant I_j) = b_j \sum_{k=I_j}^{N-n} \binom{N-n}{k} \lambda_j^k (1-\lambda_j)^{N-n-k}. \end{aligned} \tag{14}$$

The package price p_{ij}^T for any (i,j) combination can be established from

$$\begin{aligned} \partial V_n(\mathbf{I})/\partial p_{ij} &= f_{ij}(p_{ij})(p_i + V_{n+1}(\mathbf{I}-\mathbf{e}_i)) - f_{ij}(p_{ij})(p_{ij} + V_{n+1}(\mathbf{I}-\mathbf{e}_i-\mathbf{e}_j)) + \overline{F}_{ij}(p_{ij}) \\ &= -f_{ij}(p_{ij})(p_{ij} - p_i - (V_{n+1}(\mathbf{I}-\mathbf{e}_i) - V_{n+1}(\mathbf{I}-\mathbf{e}_i-\mathbf{e}_j))) + \overline{F}_{ij}(p_{ij}) = 0, \end{aligned} \tag{15}$$

so that after substituting (14) to (15) and rearranging we obtain

$$p_{ij}^T = p_i + \overline{F}_{ij}\left(p_{ij}^T\right)\Big/ f_{ij}\left(p_{ij}^T\right) + b_j \sum\nolimits_{k=I_j}^{N-n} \binom{N-n}{k} \lambda_j^k (1-\lambda_j)^{N-n-k}. \tag{16}$$

Note that the expression for the myopic price p_{ij}^M introduced earlier does not contain the last term appearing in (16): p_{ij}^T (T stands for the heuristically calculated "T"erminal value of inventory) is an upper bound on p_{ij}^M. At the same time, p_{ij}^T is a lower bound on the optimal price since packaging in subsequent periods is not accounted for. Also note that the value of p_{ij}^T explicitly depends on the inventory of product j.

The packaging complement for product i is then determined as follows. Suppose that customer i arrives and we decide not to offer any package. Then we earn $p_i + V_{n+1}(\mathbf{I}-\mathbf{e}_i)$. The incremental value (denoted by Δ_{ij}) of offering the package "$i-j$" can be calculated as follows:

$$\begin{aligned} \Delta_{ij} &= \left(F_{ij}\left(p_{ij}^T\right)(p_i + V_{n+1}(\mathbf{I}-\mathbf{e}_i)) + \overline{F}_{ij}\left(p_{ij}^T\right)\left(p_{ij}^T + V_{n+1}(\mathbf{I}-\mathbf{e}_i-\mathbf{e}_j)\right)\right) - (p_i + V_{n+1}(\mathbf{I}-\mathbf{e}_i)) \\ &= \overline{F}_{ij}\left(p_{ij}^T\right)\left(p_{ij}^T - p_i - (V_{n+1}(\mathbf{I}-\mathbf{e}_i) - V_{n+1}(\mathbf{I}-\mathbf{e}_i-\mathbf{e}_j))\right) \\ &= \overline{F}_{ij}\left(p_{ij}^T\right)\left(p_{ij}^T - p_i - b_j \sum\nolimits_{k=I_j}^{N-n} \binom{N-n}{k} \lambda_j^k (1-\lambda_j)^{N-n-k}\right) \\ &= \left(\overline{F}_{ij}\left(p_{ij}^T\right)\right)^2 \Big/ f_{ij}\left(p_{ij}^T\right) \ \text{(from (16))}. \end{aligned}$$

Hence, we propose the following dynamic packaging/pricing heuristic $\mathbf{H}^T$: when a customer of type i arrives, we calculate p_{ij}^T for all j using (16). Then, the values of Δ_{ij} are computed for all j and the index of the highest value is selected:

$$j^T(i) = \arg\max_{j\neq i} \left(\left(\overline{F}_{ij}\left(p_{ij}^T\right)\right)^2 \Big/ f_{ij}\left(p_{ij}^T\right)\right). \tag{17}$$

We observe that under the two-stage heuristic both the packaging and the pricing decisions are truly dynamic: both depend on current product inventories as well as on "time-to-go". In order to illustrate packaging decisions under $\mathbf{H}^T$ we consider an example with the exponential price reservation function ($\overline{F}_{ij}(p_{ij}) = \exp(-\beta_{ij}(p_{ij} - p_i))$), symmetric price sensitivity factors β_{ij}, which depend only on the index of the first-choice product rather than on the index of the packaging complement ($\beta_{ij} = \beta_i$), and symmetric penalty costs ($b_j = b$). In this case, we obtain $p_{ij}^T = p_i + \frac{1}{\beta_i} + b\sum_{k=I_j}^{N-n}\binom{N-n}{k}\lambda_j^k(1-\lambda_j)^{N-n-k}$ and, after some algebra, $j^T(i) = \arg\max_{j\neq i}\left(\sum_{k=0}^{I_j}\binom{N-n}{k}\lambda_j^k(1-\lambda_j)^{N-n-k}\right)$. Note that the expression $S(n, I_j, \lambda_j) = \sum_{k=0}^{I_j}\binom{N-n}{k}\lambda_j^k(1-\lambda_j)^{N-n-k}$ is an increasing function of the inventory I_j and the decision epoch n, is a decreasing function of demand intensity λ_j and does not depend on the index of the first-choice product i. In other words, $S(n, I_j, \lambda_j)$ can be interpreted as a proxy of how well product j has been selling up to the decision epoch n: the lower the value of this expression, the closer the product to the status of best-seller. Thus, the price of "$i-j$" package (when it is offered) is, as expected, a decreasing function of I_j and of the decision epoch n, and an increasing function of the demand intensity λ_j. The set of packaging complements can be established as follows: at each decision epoch n, rank all products according to the current value of the parameter $S(n, I_j, \lambda_j)$ and let $\widehat{j}_s^T = \arg\max_j (S(n, I_j, \lambda_j))$ and $\widehat{j}_n^T = \arg\max_{j\neq\widehat{j}_s^T}(S(n, I_j, \lambda_j))$ be the indices of products with the highest and second-highest parameter values (note that if more than one product currently has the same parameter value, such "degeneracy" can be resolved in an arbitrary way). Then, (17) is equivalent to

$$j^T(i) = \begin{cases} \widehat{j}_s^T, & i \neq \widehat{j}_s^T, \\ \widehat{j}_n^T, & i = \widehat{j}_s^T. \end{cases} \tag{18}$$

In the view of our interpretation of $S(n, I_j, \lambda_j)$ as the best-selling index, the packaging in (18) is intuitively appealing: all products are offered in a package with the current slowest seller, and the slowest seller itself is "packaged" with the current second-slowest seller.

3.2.4 The depletion ratio heuristic

Our discussion of the "two-stage" heuristic indicates that dynamic cross-selling can be viewed as an effective tool for re-directing the demand for best-selling products to products with slower-than-desired sales. Below we propose another easy-to-implement dynamic packaging approach we call the "depletion ratio" (DR) heuristic, which emphasizes the role of dynamic packaging in the cross-selling process. Under the DR heuristic $\left(\mathbf{H}^{DR}\right)$, in each decision epoch n, each product i is assigned a "depletion ratio" index equal to the ratio of the current product inventory I_i to the rate λ_i at which product inventory would be depleted in the absence of cross-selling. Defined in this way, the DR index plays a role similar to $S(n, I_j, \lambda_j)$ defined for the "two-stage" heuristic under the exponential pricing reservation function: it indicates the current "best-selling" status of each product. For the current set of inventory values, let $j^s = \arg\max_i (I_i/\lambda_i)$ and $j^n = \arg\max_{i\neq j^s}(I_i/\lambda_i)$

be the indices of the slowest-selling and the second slowest-selling products, respectively. Then, the dynamic packaging decisions under the DR approach can be described as follows:

$$j^{DR}(i) = \begin{cases} j^s, & i \neq j^s, \\ j^n, & i = j^s. \end{cases} \tag{19}$$

The intuition behind the choice of the cross-selling complement in (19) is similar to the intuition behind (18): every product is "packaged" with the current slowest seller, while the slowest seller itself is packaged with the current second-slowest-selling product. Note that the DR approach does not restrict the choice of the package pricing policy and can be combined with simple static pricing or sophisticated best-choice dynamic pricing. In particular, in our analysis below we consider two cross-selling policies based on DR packaging: the "myopic" DR policy ($\mathbf{H}^{DRM}$), which complements DR packaging with the myopic pricing defined by (13), and the "optimal" DR policy ($\mathbf{H}^{DRO}$), which selects the best pricing under DR packaging by solving the variant of (1) with packaging complements determined by (19).

Despite the intuitive nature of DR packaging, an analytical characterization of sufficient conditions for its optimality is hard to obtain, except in rather restrictive settings. Below we provide an example of such conditions under static pricing in a symmetric product environment:

Proposition 4

Let $\lambda_i = \lambda$, $p_i = p$, $b_i = b$, $p_{ij} = q$, $p < q < p + b$, and $\overline{F}_{ij}(p_{ij}) = \gamma$ for any $j \neq i$ and $i, j = 1, 2, ..., m$. Then, DR packaging is optimal for any decision epoch $n = 1, ..., N$.

Proposition 4 considers the setting in which product packages are being sold at a fixed price and the products differ only in their inventory values. While Proposition 4 indicates the potential effectiveness of the DR packaging approach in nearly symmetric environments with static package pricing, its performance in more typical settings needs to be evaluated numerically.

3.2.5 Effectiveness of packaging and pricing heuristics: numerical study

In this section we use an extensive numerical study to test the effectiveness of pricing/packaging heuristics for the ER model. Let us denote by R_{OPT} the optimal expected profits obtained by solving (2) (note that computing the optimal dynamic packaging and pricing policy requires solving an m-dimensional DP). For each cross-selling heuristic π, we compute the value of its expected profits as follows. In each decision epoch n, let $j^\pi(n,i)$ be the packaging complement that the heuristic π selects for product i, and also let $p^\pi(n,i)$ be the price charged for the "$i - j^\pi(n,i)$" package. Then, for any initial inventory vector, the expected profits under π are evaluated using the iteration

$$\begin{aligned} V_n^\pi(\mathbf{I}) &= \sum_{i=1}^{m} \lambda_i (F_{ij^\pi(n,i)}(p^\pi(n,i)) \left(p_i + V_{n+1}^\pi(\mathbf{I} - \mathbf{e}_i)\right) + \overline{F}_{ij^\pi(n,i)}(p^\pi(n,i)) \left(p^\pi(n,i) + V_{n+1}^\pi(\mathbf{I} - \mathbf{e}_i - \mathbf{e}_j)\right) \\ &\quad + \left(1 - \sum_{i=1}^{m} \lambda_i\right) V_{n+1}^\pi(\mathbf{I}) \end{aligned} \tag{20}$$

for $\mathbf{I} > \mathbf{0}$, $n = 1, ..., N$ with appropriate "boundary" conditions. We probe the effectiveness of the $\mathbf{H}^M$, $\mathbf{H}^D$, $\mathbf{H}^T$, $\mathbf{H}^{DRM}$, and $\mathbf{H}^{DRO}$ heuristics in the ER model by comparing the relative performance gaps between R_M, R_D, R_T, R_{DRM}, R_{DRO} and R_{OPT}: $\varepsilon_k = 100\% \times (R_{OPT} - R_k)/R_{OPT}$, $k = M,\ D,\ T, DRM, DRO$.

The case of $m = 3$ products and exponential pricing reservation functions

Our search at Amazon.com reveals that in most instances the number of products connected through packaging is three or less (see Table 1). In order to reflect this observation in our numerical study, we fixed the number of products that can be potentially connected through packaging at $m = 3$. Recall that with three products optimal packaging decisions are, generally speaking, state-dependent and dynamic.

Our test suite was designed as follows. In order to isolate the effects of package pricing, we set the prices for individual products at the same level: $p_1 = p_2 = p_3 = 1$. For package reservation prices, we use the "exponential" distribution functions $\overline{F}_{ij}(p_{ij}) = \exp(-\beta_{ij}(p_{ij} - p_i))$, and assume for simplicity that the price sensitivity factors β_{ij} only depend on the index of the first-choice product rather than on the index of the packaging complement: $\beta_1 = \beta_{11} = \beta_{12} = \beta_{13}$, $\beta_2 = \beta_{21} = \beta_{22} = \beta_{23}$, $\beta_3 = \beta_{31} = \beta_{32} = \beta_{33}$. In our numerical study, each of the price sensitivity factors takes a value of 1 ("low sensitivity"), 2 ("medium sensitivity"), 5 ("high sensitivity"), or 20 ("very high sensitivity"). We set the number of time periods $N = 20$ and fix the total consumer arrival rate $\lambda = \lambda_1 + \lambda_2 + \lambda_3 = 0.8$. The "individual product" arrival rates are varied as follows: $\lambda_1 = 0.1 + 0.5\sigma_1$, $\lambda_2 = 0.1 + (0.6 - \lambda_1)\sigma_2$, and $\lambda_3 = 0.8 - \lambda_1 - \lambda_2$, where both σ_1 and σ_2 take values of 0, 0.5, and 1. Hence, we allow 7 possible combinations of arrival rates: $\{0.1, 0.1, 0.6\}$, $\{0.1, 0.6, 0.1\}$, $\{0.6, 0.1, 0.1\}$, $\{0.1, 0.35, 0.35\}$, $\{0.35, 0.1, 0.35\}$, $\{0.35, 0.35, 0.1\}$, $\{0.35, 0.225, 0.225\}$. These combinations cover three essential scenarios for a three-way splitting of total demand: "mostly single-product arrivals" (first three), "mostly two-product arrivals" (next three) and "closely-valued demands" (last one).

In order to test the sensitivity of the heuristics with respect to the initial levels of product inventory, we investigate the cases in which the initial inventories are set at $I_i = (1 + \gamma)\lambda_i N$, $i = 1, 2, 3$, where the "inventory availability" coefficient γ takes values of -0.8("severely constrained inventory"), -0.3 ("moderately constrained inventory"), 0 ("inventory matching demand"), 0.3 ("moderately slack inventory"), and 0.8 ("virtually unconstrained inventory"). Note that in the last three cases our terminology is somewhat arbitrary since, as we expect, under the optimal packaging/pricing policy the cumulative expected demand for product i will exceed $\lambda_i N$ due to cross-selling. Finally, we set the values of the ER prices b_i equal to ηp_i for $i = 1, 2, 3$, where η takes values of $0.2, 0.5$, and 0.8. Thus, in total, we test $4^3 \times 7 \times 5 \times 3 = 6,720$ problem instances. Over these instances, the averages of ε_M, ε_D, ε_T, ε_{DRM} and ε_{DRO} turn out to be 10.63%, 3.16%, 0.12%, 7.75%, 0.14%, respectively. We also observe that the two-stage and the DRO heuristics perform extremely well in almost all test instances. The relative performance gaps in the *worst* cases are merely 0.69% ("two-stage") and 1.04% (DRO) over all the tested instances. It is also interesting to observe that the deterministic approximation performs better overall than the myopic policy with about 7.5% improvement on average, which is due to the fact that the heuristic $\mathbf{H}^D$ dynamically incorporates the effect of the inventory into packaging and pricing decisions. As expected, in many cases revenue losses when using the myopic heuristic are remarkably large. We note that the DRM heuristic occupies an intermediate place between the myopic and the deterministic heuristics in terms of its performance. This observation underscores the importance of a good match between packaging and pricing decisions: "depletion ratio" packaging, being near-optimal when coupled with best-match pricing, loses its effectiveness when applied together with myopic pricing.

$\eta =$	0.2	0.5	0.8
ε_T	0.05	0.13	0.18
ε_D	0.91	2.74	5.83
ε_M	2.63	9.38	19.87
ε_{DRM}	1.61	6.72	14.92
ε_{DRO}	0.08	0.15	0.20

Table 2. Average performance gaps (in %) as functions of ER price.

$\gamma =$	-0.8	-0.3	0	0.3	0.8
ε_T	0.03	0.14	0.20	0.15	0.08
ε_D	0.16	2.40	7.44	3.97	1.84
ε_M	18.47	11.50	10.18	7.50	5.48
ε_{DRM}	17.44	9.87	6.56	3.44	1.44
ε_{DRO}	0.00	0.01	0.08	0.35	0.26

Table 3. Average performance gaps (in %) as functions of inventory availability.

Next, we report the results of the sensitivity analysis with respect to the magnitude of the ER premium relative to the product price η (Table 2), the inventory availability coefficient γ (Table 3), and the composition of total customer demand (Table 4). The values of ε_i reported in these tables refer to the averages over problem instances in which the relevant problem parameter is fixed. For example, in Table 2, the upper left value 0.05 refers to the average of the relative performance gap for the two-stage heuristic over 2,240 problem instances with $\eta = 0.2$.

$\lambda_2 =$	0.1	0.225	.35	0.6
ε_T	0.14	0.07	0.08	0.16
ε_D	3.52	2.47	2.27	3.82
ε_M	11.45	9.67	9.12	10.48
ε_{DRM}	7.97	7.28	7.22	8.15
ε_{DRO}	0.15	0.03	0.16	0.15

Table 4. Average performance gaps (in %) as functions of arrival rate λ_2.

The first observation that can be drawn from Tables 2, 3 and 4 is that $\mathbf{H}^T$ and $\mathbf{H}^{DRO}$ heuristics are consistently producing near-optimal performance for a wide range of problem parameters, while the performance of heuristics $\mathbf{H}^M$, $\mathbf{H}^D$, and $\mathbf{H}^{DRM}$ changes in a predictable manner as problem parameters are varied. For example, when the magnitude of the ER purchase cost relative to the selling price of the single product η is small, getting additional inventory is nearly cost-free. Not surprisingly, as Table 2 indicates, $\mathbf{H}^M$, $\mathbf{H}^D$, and $\mathbf{H}^{DRM}$ perform well for small values of η. Further, as Table 3 shows, when product inventory levels are high, the marginal value of an extra unit of inventory for any product is small, and all five heuristics perform well. In fact, as (16) and the result of Proposition 3b indicate, for high inventory levels both the two-stage and the deterministic heuristics coincide with the myopic heuristic, which, in turn, becomes optimal. On the other hand, when product inventory is low, the marginal value of an extra unit of product i inventory approaches b_i. Thus, the performance of the myopic heuristic, which assumes that this marginal value is zero, deteriorates. At the same time, both the two-stage and the deterministic heuristics perform very well, since, as (16) and the result of Proposition 3b indicate, in the low-inventory limit the pricing and the packaging decisions prescribed by these heuristics become optimal. Finally, Table 4 demonstrates that the performance of five heuristics is rather insensitive to changes in the relative composition of the total demand flow: the two-stage and DRO heuristics remain the best choice, followed by deterministic approximation, with myopic policy a distant fourth. We obtain similar results for fixed values of λ_1 and λ_3.

Testing two-stage and depletion ratio heuristics: the impact of the number of products and the shape of reservation functions

In the numerical study described above we have identified the "two-stage" and DRO heuristics as having the best performance across a wide range of problem parameters for the case of 3 products and exponential pricing reservation functions. Below we focus on investigating the robustness of these two heuristics with respect to changes in the number of products and in the shape of reservation functions. In particular, we introduce the following changes to our numerical suite. In addition to the exponential reservation functions $\overline{F}_{ij}(p_{ij}) = \exp(-\beta_{ij}(p_{ij} - p_i))$, we also consider the "power" reservation functions $\overline{F}_{ij}(p_{ij}) = \left(\frac{p_i+p_j-p_{ij}}{p_j}\right)^{\beta_{ij}}$. The choice of the "power" form follows from our intent to investigate the impact of the change in the shape of the pricing function (from "convex" for exponential functions to "concave" for power functions with $\beta_{ij} \leq 1$) on the performance of our two best heuristics. For both types of function, we use the symmetry assumption $\beta_{ij} = \beta_i$ and consider $\beta_i = 0.5, 1.5$. Note that the case of $\beta_i = 0.5$ can be of particular interest since for such a low value of price sensitivity parameter the cross-selling may be expected to be intense and the "two-stage" approach can really be challenged. In order to control the number of investigated problem instances, we limit the values of the "inventory availability" coefficient γ to $-0.3, 0, 0.3$. For the case of $m = 3$ products, we use the same demand patterns as described above. For the case of $m = 4$ products, we use a similar approach and fix the total consumer arrival rate $\lambda = \lambda_1+\lambda_2+\lambda_3+\lambda_4 = 0.8$. The "individual product" arrival rates are varied as follows: $\lambda_1 = 0.1 + 0.4\sigma_1$, $\lambda_2 = 0.1 + (0.5 - \lambda_1)\sigma_2$, $\lambda_3 = 0.1 + (0.6 - \lambda_1 - \lambda_2)\sigma_2$, and $\lambda_4 = 0.8 - \lambda_1 - \lambda_2 - \lambda_3$, where σ_1, σ_2, and σ_3 take values of 0, 0.5, and 1. Hence, in this case we allow 27 possible combinations of arrival rates, which, similarly to the numerical suite used in the previous section, cover all essential scenarios for a four-way splitting of the total demand: "mostly single-product arrivals", "mostly two-product arrivals", "mostly three-product arrivals" and "closely-valued demands". The remaining problem parameters were set in the same way as in numerical study above. In total, we test $2^3 \times 7 \times 3 \times 3 = 504$ problem instances for $m = 3$ products and $2^3 \times 27 \times 3 \times 3 = 1,944$ problem instances for $m = 4$ products. The results of the numerical runs are presented in Table 5.

"Two-stage"	Exponential RF	Power RF
3 products	0.13 (0.69)	0.24 (6.98)
4 products	0.86 (9.05)	1.09 (7.08)

DRO	Exponential RF	Power RF
3 products	0.16 (1.08)	0.16 (1.94)
4 products	0.46 (3.01)	0.25 (2.29)

Table 5. Average (and maximum) performance gaps (in %) for "two-stage" and DRO heuristics.

We note that, due to changes in our test suite, the deviations for the case of the exponential reservation function and $m = 3$ products differ somewhat, but not by much, from those in the previous section. We observe that both heuristics, on average, show quite a high degree of robustness with respect to changes in the reservation function and the number of products. This conclusion is especially valid for the DRO heuristic, which retains near-optimal performance even in the worst-case scenarios. On the other hand, the "two-stage" approach appears to lose its worst-case effectiveness when the number of products is increased or the shape of the reservation function is changed, or both. It is interesting to note that the worst-case performance of the two-stage heuristic is observed in problem instances with $\beta_i = 0.5$ and $\eta = 0.8$, i.e., the instances in which 1) consumers are willing to accept high package prices (for both types of reservation functions) and 2) the out-of-stock products are costly to replenish. It is intuitive that the heuristic which neglects "future" cross-selling opportunities does not perform well in cases in which cross-selling is readily accepted by customers. This performance gap opens up dramatically when the number of cross-selling choices for each product is increased from 2 ($m = 3$) to 3 ($m = 4$). The high cost of replenishment further accentuates the profit loss resulting from the use of an ineffective cross-selling approach.

The ultimate choice between the DRO and the "two-stage" heuristics may depend on the specific features of the business environment in which the firm operates. On the one hand, if the nature of the firm's inventory allows it to limit the number of likely complements for each product to 2, the firm may be advised to use the "two stage" approach as long as its customers are relatively sensitive to package prices: the pricing policies under "two-stage" approach are much easier to compute than those under the DRO heuristic (which requires solving a number of optimization problems at each decision epoch). On the other hand, the DRO may be a much more effective policy in cases in which the number of products is high or the estimated customer price sensitivity for product packages is low.

4 The Lost Sales model

Although in some cases it is possible to procure an out-of-stock product, in a variety of situations this might prove impossible or too costly. For example, when cross-selling travel services, there might not be seats available on the requested route. Hence, it will be necessary to deny the customer's request. To address this issue, we consider an alternative setting in which there is no opportunity to replenish inventory and a customer request is simply denied in the case of a stock-out. Similarly to the previous section, it is convenient to introduce the sets of indices A_n to denote products that have at least one unit of inventory at the beginning of the n-th decision epoch. Then, under the LS model, the generalization of (1) is given by

$$\begin{aligned} V_n(\mathbf{I}) &= \sum_{i\in A_n} \lambda_i \max_{j\neq i,\, j\in A_n} \left(\max_{p_{ij}} \left(F_{ij}(p_{ij}) \left(p_i + V_{n+1}(\mathbf{I}-\mathbf{e}_i)\right) + \overline{F}_{ij}(p_{ij}) \left(p_{ij} + V_{n+1}(\mathbf{I}-\mathbf{e}_i-\mathbf{e}_j)\right)\right)\right) \\ &\quad + \left(1 - \sum_{i\in A_n} \lambda_i\right) V_{n+1}(\mathbf{I}). \end{aligned} \tag{21}$$

As (21) indicates, the "lost sales" feature of the inventory dynamics is reflected in the fact that the set of product indices that actively participate in the DP transformation on the right-hand side of (1) is "shrinking" as time passes.

4.1 Dynamic pricing under the static packaging

As indicated earlier, in the general case the LS model possesses few structural properties. However, the two-product case is somewhat more amenable to analysis. Hence, we begin our analysis with this simpler case.

Under the LS model, when a company runs out of inventory for one of the products, we obtain

$$V_n(I_1,0) = \lambda_1 \left(p_1 + V_{n+1}(I_1-1,0)\right) + (1-\lambda_1)V_{n+1}(I_1,0), \tag{22}$$

for $I_1 \geq 1$, $n = 1,...,N$ and

$$V_n(0,I_2) = \lambda_2 \left(p_2 + V_{n+1}(0,I_2-1)\right) + (1-\lambda_2)V_{n+1}(0,I_2), \tag{23}$$

for $I_2 \geq 1$, $n = 1,...,N$. In addition, when the inventories of both products are depleted, we have

$$V_n(0,0) = 0,\ n = 1,...,N. \tag{24}$$

Using the induction over the time index n, we can formalize the structural properties of the LS model in a two-product case:

Proposition 5

a) The optimal expected revenue $V_n(I_1,I_2)$ is a non-decreasing function of I_1 and I_2, respectively, for any $n = 1,...,N$:

$$V_n(I_1+1,I_2) - V_n(I_1,I_2) \geq 0, V_n(I_1,I_2+1) - V_n(I_1,I_2) \geq 0. \tag{25}$$

b) The optimal expected revenue $V_n(I_1,I_2)$ is a supermodular function of (I_1,I_2), for any $n = 1,...,N$:

$$V_n(I_1+1,I_2+1) - V_n(I_1,I_2+1) \geq V_n(I_1+1,I_2) - V_n(I_1,I_2). \tag{26}$$

c) The optimal package price $p^*_{ij}(I_1,I_2,n)$ is non-decreasing in I_i, for $(i,j) = (1,2)$ and $(2,1)$:

$$\begin{aligned} p^*_{12}(I_1+1,I_2,n) &\geq p^*_{12}(I_1,I_2,n), n = 1,...,N, \\ p^*_{21}(I_1,I_2+1,n) &\geq p^*_{21}(I_1,I_2,n), n = 1,...,N. \end{aligned} \tag{27}$$

The statement of part a) of Proposition 5 is intuitively appealing and is similar to those established in the literature on single-product pricing (see, e.g., Gallego and van Ryzin [16], and Federgruen and Heching [13]). The statement of part b) can be rationalized as follows: the higher the inventory level for product j, the more opportunities there are for selling product i. Thus, the marginal profit from adding one unit of product i increases as the inventory level of product j increases. Note that part b) is in sharp contrast to the ER model in which the optimal value function was separable in inventories of all products and hence supermodularity held trivially as an equality in (26).The intuition behind Proposition 5c can be explained as follows: the larger the inventory of product i, the more opportunities there will be to sell this product in the future. Therefore, the package offered to a customer requesting product i can be priced high as there will be other opportunities in the future to sell the same package. This result should be contrasted with related results in single-product revenue management where larger product inventory typically leads to lower price (see, e.g., Gallego and van Ryzin [16]).

The statements of Proposition 5 are best illustrated with an example that also reveals additional insights into the problem. The two pricing curves shown in Figure 2 depict the time trajectories of the optimal price for the "1-2" package with the initial states ($I_1 = 1$, $I_2 = 5$) and ($I_1 = 25$, $I_2 = 5$). We use the following reservation functions and problem parameters in this example: $\overline{F}_{12}(p_{12}) = ((p_1 + p_2 - p_{12})/p_2)^{\beta_{12}}$ defined over $[p_1, p_1 + p_2]$, $\overline{F}_{21}(p_{21}) = ((p_1 + p_2 - p_{21})/p_1)^{\beta_{21}}$ defined over $[p_2, p_1 + p_2]$ with $\beta_{12} = \beta_{21} = 1$, $p_1 = 1$, $p_2 = 2$, and $\lambda_1 = \lambda_2 = 0.4$. Both curves are monotone, in accordance with (25), and, as (27) indicates, a higher value of I_1 diminishes the incentive to offer a discount on the package. In addition, we observe that both pricing trajectories "reduce" to the same myopic price once the remaining planning horizon $N - n$ becomes less than the inventory of product 2, $I_2 = 5$. In this case, the existing inventory of product 2 cannot possibly be depleted over the time remaining until the end of the horizon, and the system is prompted to behave as if the inventory of product 2 is infinite. In this case,

$$p_{12}^*(I_1, I_2, n) = \arg\max_{p_{12}} \left(\bar{F}_{12}(p_{12})\,(p_{12} - p_1)\right), \tag{28}$$

which depends neither on the products' inventories nor on the time index. In other words, the heuristic price p_{ij}^M becomes optimal.

It seems reasonable to believe that, similar to the ER model, the concavity property of $V_n(I_1, I_2)$ and the monotonicity of $p_{ij}^*(I_1, I_2, n)$ with respect to I_j should hold. Unfortunately, these intuitive properties do not hold for the LS model. The following counterexample demonstrates this. Let $p_1 = 100, p_2 = 200, \lambda_1 = 0.8, \lambda_2 = 0.2$, and $N = 9$. Then, for the reservation function $F_{ij}(x) = (x - p_i)/p_j$, for $x \in [p_i, p_i + p_j]$, it is easy to verify that $V_3(2,1) + V_3(0,1) - 2V_3(1,1) = 0.219 > 0$ and $p_{21}^*(2,2,2) > p_{21}^*(1,2,2)$.

We now turn to the multi-product case. Consider any static product packaging rule $B = \{j(i), i = 1, ..., m\}$. The multi-product LS model with static packaging B and dynamic pricing can be formulated as follows:

$$\begin{aligned} V_n^{LS}(\mathbf{I}) \quad = \quad & \sum_{i \in A_n, j(i) \in A_n} \lambda_i \max_{p_{i,j(i)}} F_{i,j(i)}\left(p_{i,j(i)}\right)\left(p_i + V_{n+1}^{LS}(\mathbf{I} - \mathbf{e}_i)\right) \\ & + \overline{F}_{i,j(i)}\left(p_{i,j(i)}\right)\left(p_{i,j(i)} + V_{n+1}^{LS}\left(\mathbf{I} - \mathbf{e}_i - \mathbf{e}_{j(i)}\right)\right) \\ & + \sum_{i \in A_n, j(i) \notin A_n} \lambda_i \left(p_i + V_{n+1}(\mathbf{I} - \mathbf{e}_i)\right) + \left(1 - \sum_{i \in A_n} \lambda_i\right) V_{n+1}^{LS}(\mathbf{I}). \end{aligned} \tag{29}$$

In the multi-product case, the few properties possessed by the two-product LS model vanish. For example, the supermodularity of the optimal value function no longer holds, and the LS model retains only the monotonicity of optimal revenues with the product inventory levels, which can be established using the induction over time index n:

Lemma 1

The optimal expected revenue is non-decreasing in the inventory level, i.e., $V_n^{LS}(\mathbf{I} + \mathbf{e}_i) \geqslant V_n^{LS}(\mathbf{I})$, for $i = 1, ..., m$ and $n = 1, ..., N$.

4.2 Heuristic approaches

Since even the stand-alone dynamic pricing problem under the LS model is rather intractable, there is little hope of efficiently solving the dynamic pricing/packaging problem without the use of heuristics. Hence, it is desirable to know how much we might lose by implementing a simple static packaging policy. Below we derive upper and lower bounds for the expected profit under the LS model with static packaging and dynamic pricing. Using these two bounds, we identify situations in which expected revenues are relatively insensitive to the choice of a particular static packaging scheme.

Definition 2

Let $L_n^i(I)$ be the function satisfying the following recursive relation:

$$L_n^i(I) = \lambda_i \left(p_i + L_{n+1}^i\left(I-1\right)\right) + \left(1-\lambda_i\right) L_{n+1}^i\left(I\right), \tag{30}$$

for $I \geqslant 1$, and $L_n^i(0) = 0$, for $i = 1, ..., m$, and $n = 1, ..., N$, while $L_{N+1}^i(I) = 0$.

Definition 3

Let $U_n^i(I)$ be the function satisfying the following recursive relation:

$$U_n^i(I) = \lambda_i \left(p_i + U_{n+1}^i\left(I-1\right) + \overline{F}_{i,j(i)}(p_{i,j(i)}^M)(p_{i,j(i)}^M - p_i)\right) + \left(1-\lambda_i\right) U_{n+1}^i\left(I\right), \tag{31}$$

for $I \geqslant 1$, and $U_n^i(0) = 0$, for $i = 1, ..., m$, and $n = 1, ..., N$, while $U_{N+1}^i(I) = 0$.

Define $L_n(\mathbf{I}) = \sum_{i=1}^m L_n^i(I_i)$ and $U_n(\mathbf{I}) = \sum_{i=1}^m U_n^i(I_i)$. Note that $L_n(\mathbf{I})$ is the expected revenue under the LS model with no packaging. The following proposition uses the induction over n to show that $L_n(\mathbf{I})$ and $U_n(\mathbf{I})$ are a lower bound and an upper bound, respectively, for the optimal expected revenue $V_n^{LS}(\mathbf{I})$ in the LS model.

Proposition 6

a) $L_n(\mathbf{I}) \leq V_n^{LS}(\mathbf{I}) \leq U_n(\mathbf{I})$*, for* $n = 1, ..., N$*.*

b) For $n = 1, ..., N$*,*

$$\frac{V_n^{LS}\left(\mathbf{I}\right) - L_n\left(\mathbf{I}\right)}{L_n\left(\mathbf{I}\right)} \leq \max\left\{ \frac{p_{j(i)}}{p_i} \,\middle|\, i = 1, ..., m \right\}. \tag{32}$$

Note that the upper bound given in part b) of Proposition 6 is independent of customer reservation prices and arrival processes. If each product is significantly more expensive than its packaging complement, then the bound in (32) is tight, implying that the impact of the packaging decision on expected profit is negligible. Given the price of each product, we can now easily compute the upper bound for each packaging topology. Thus, without the demand information, we are able to identify the packaging configurations with the tight upper bound. Consider an example in which $(p_1, p_2, p_3) = (100, 10, 1)$ and product 1 is packaged with product 2 and product 2 with product 3 (product 3 is not packaged with any other product). According to (32), under this packaging configuration, the highest possible revenue is at most 10% above the revenue achieved in the absence of cross-selling.

Even with static packaging, the computation of the optimal price under the LS assumption requires solving an m-dimensional DP problem (29) that is computationally challenging for large m. Therefore, there exists a need to develop a simple, easy-to-implement and effective pricing heuristic. As (29) indicates, the optimal package price can be expressed as the maximizer of the following expression:

$$p_{ij}^{*}(\mathbf{I}, n) = \arg\max_{p_{ij}} \left(\bar{F}_{ij}(p_{ij})(p_{ij} - p_i + (V_{n+1}(\mathbf{I} - \mathbf{e}_i - \mathbf{e}_j) - V_{n+1}(\mathbf{I} - \mathbf{e}_i))\right). \tag{33}$$

A natural approximation under the above modification is to use the package prices defined by p_{ij}^{M} in (13) which essentially ignores the last two terms of (33). This heuristic package price is constant over the time horizon and is independent of the inventory level of either product. We reuse the notation $\mathbf{H}^M$ to denote this heuristic (it is exactly the same heuristic applied to the two different models, ER and LS). The following result states that p_{ij}^{M} is the lower bound on the optimal price $p_{ij}^{*}(\mathbf{I}, n)$, and characterizes the sufficient conditions under which the heuristic $\mathbf{H}^M$ is optimal:

Lemma 2

a) The optimal price is always no less than the price determined by the heuristic $\mathbf{H}^M$: $p_{ij}^{*}(\mathbf{I}, n) \geq p_{ij}^{M}$, for any $\mathbf{I}$, and n.

b) Heuristic $\mathbf{H}^M$ is optimal in period n when the available inventory is $I_k \geq N+2-n$ for all $k = 1, 2, ..., m$.

The proof of part a of Lemma 2 utilizes the monotonicity of the value function established in Lemma 1. Part b of Lemma 2 indicates that the static pricing heuristic is optimal in cases in which the inventory levels of every product are high, thus formalizing the observation in Figure 2. It is to be expected that the performance of the myopic pricing heuristic worsens as the inventory becomes more "constrained".

All heuristics used for the ER model can be easily modified for the LS model. For the $\mathbf{H}^D$ heuristic the only change is to set the ER prices b_i ($i = 1, 2, ..., m$) to a "big" number in (**P2**), and then the same procedure can be applied to derive the heuristic solution. For the $\mathbf{H}^T$ heuristic it suffices to replace b_i with p_i whenever it appears. Additionally, after observing that the DP formulations of the ER and LS models differ on the "boundary" of state space, it is natural to use optimal pricing that employs the separability property for the ER model while making myopic packaging decisions as a heuristic solution for the LS model. We call such an approximation a decomposition heuristic, denoted by $\mathbf{H}^S$.

$\gamma =$	-0.8	-0.3	0	0.3	0.8
ε_T	0.03	0.20	0.43	0.44	0.24
ε_S	0.86	1.52	3.01	2.94	2.56
ε_D	2.93	4.75	8.75	5.62	2.59
ε_M	10.64	10.26	10.16	8.60	6.58
ε_{DRM}	12.70	10.78	8.42	5.42	2.43
ε_{DRO}	0.01	0.01	0.13	0.63	0.50

Table 6. Average performance gaps (in %) as functions of inventory availability.

$\lambda_2 =$	0.1	0.225	.35	0.6
ε_T	0.33	0.11	0.13	0.38
ε_S	2.53	1.57	1.64	2.08
ε_D	5.04	4.89	4.52	5.20
ε_M	9.22	9.35	9.43	8.93
ε_{DRM}	7.99	7.90	7.88	7.96
ε_{DRO}	0.29	0.05	0.30	0.25

Table 7. Average performance gaps (in %) as functions of arrival rate λ_2.

Below we report the results of the numerical study testing the performance of heuristics $\mathbf{H}^M$, $\mathbf{H}^D$, $\mathbf{H}^T$, $\mathbf{H}^S$, $\mathbf{H}^{DRM}$ and $\mathbf{H}^{DRO}$ applied to the LS model. Our test suite is set similarly to the one used for the ER numerical study described above, except, naturally, for the absence of the ER prices b_i. Thus, for the LS model we test in total $4^3 \times 7 \times 5 = 2,240$ problem instances. Over these problem instances, the average relative performance gaps for the four heuristics (ε_M, ε_D, ε_T, ε_S, ε_{DRM} and ε_{DRO}) are 9.25%, 4.93%, 0.27%, 2.18%, 7.95%, and 0.26%, respectively. Tables 6 and 7 indicate that, just as in the ER case, the two-stage and DRO heuristics perform extremely well, with worst-case relative performance gaps of 1.8% and 1.55%. Comparing Tables 3 and 6, we observe that the performance of the deterministic heuristic worsens in the LS model, and the decomposition heuristic becomes the third-best performer.

5 Summary

In this paper we study the problem of dynamically cross-selling products on the Internet. While there are related studies of static bundling and static cross-selling decisions, to the best of our knowledge there are no studies that, like ours, consider cross-selling in the dynamic setting and identify it as an opportunity complementary to single-product revenue management. The dynamic aspect of the cross-selling problem is important since the Internet provides a truly dynamic environment that differs from the static setting of the brick-and-mortar store in many aspects. These aspects are also found in the dynamic pricing/revenue management literature, which we draw upon in our analysis. However, this literature has predominantly focused on dynamic pricing for a single product and hence does not address dynamic pricing of product packages under cross-selling.

Our findings can be concisely summarized in the following two tables. Table 8 characterizes the optimal cross-selling approaches under different inventory levels of individual products. When inventory is ample (last column) there is little need to account for product inventories when making cross-selling decisions: the main driving force behind the optimal package price is consumer preferences. This is when myopic cross-selling policies work well. If inventory is severely constrained (first column), revenue management through dynamic cross-selling is beneficial in the ER model but much less so in the LS model. Understandably, the simple myopic heuristic performs very poorly in this case. Finally, most benefits from dynamic cross-selling arise when inventory is approximately equal to expected demand. In this case it is particularly important to have a good heuristic solution. Indeed, from Tables 3 and 6 we can see that the performance of all heuristics worsens for intermediate levels of inventory because of the impact of incorrect cross-selling decisions. However, the two-stage and DRO heuristics are still good approximations of the optimal dynamic cross-selling policy.

	Inventory$\ll$Exp. Demand	Inventory=Exp. Demand	Inventory$\gg$Exp. Demand
ER model	Dynamic cross-selling		Myopic cross-selling
LS model	No cross-selling	Dynamic cross-selling	Myopic cross-selling

Table 8. The optimal cross-selling approaches for different inventory levels.

Table 9 further illustrates our earlier observation that dynamic cross-selling is complementary to more traditional single-product revenue management. While single-product revenue management is most effective when product inventory is highly constrained, most benefits from dynamic cross-selling arise when product inventory is approximately equal to expected demand. Therefore, a potentially fruitful direction for future research would be to explore simultaneous dynamic pricing of individual products as well as cross-selling (we analyze only the latter). Although in some cases (e.g., book retailing) companies might be reluctant to change the prices of individual products dynamically, in other applications (especially travel services) this extension might be plausible. We hope that future research will consider this option.

	Inventory$\ll$Exp. Demand	Inventory=Exp. Demand
RM for individual products	Large benefits	Some benefits
RM through cross-selling	Some benefits	Large benefits

Table 9. Comparison of revenue management (RM) for individual products and using cross-selling.

We believe that research in the area of dynamic cross-selling has the potential to make an important practical impact. According to several experts, dynamic cross-selling of non-air elements of the travel package with air tickets has the potential to generate large incremental revenues in an industry plagued by low margins (see Feldman [14]). Web merchants who analyze revenue increases from cross-selling report a 5% increase on average (see Peters [27]). Although early experiments with dynamic pricing of individual products on the Internet have failed (see the example of Amazon.com [1]) due to poor perception by customers, dynamic cross-selling seems to be flourishing. The reason might be that customers perceive a package as "fairly" priced as long as it is not priced above the sum of its components, and moreover, cross-selling offers may provide valuable information to customers by introducing them to new products. While Internet companies are in need of easy-to-implement algorithms for dynamic cross-selling, operations research currently offers little help in this respect. In particular, efficient algorithms are needed to deal with multi-dimensional cross-selling problems that otherwise are too computationally intensive. Some obvious extensions of our work may include allowing individual product prices to be adjusted dynamically, allowing multiple packages to be offered to the same customer and letting customers self-select a package. A study of the related practice of up-selling customers to products with higher revenues seems to be another fruitful research direction.

References

[1] "Is dynamic pricing really so bad?", *CIO magazine*, October 25 (2000), available at http://comment.cio.com/soundoff/102500.html.

[2] "Design your own discount gateway", *Business Week*, March 8 (2004).

[3] Presentation of A. Weigend, former Chief-Scientist of Amazon.com, at the Wharton School, University of Pennsylvania, March 17 (2004).

[4] Adams, W.J., and J.L. Yellen, "Commodity packaging and the burden of monopoly", *Quarterly Journal of Economics*, **90** (1976), 475-498.

[5] Adomavicius, G., and A. Tuzhilin, "An architecture of e-butler: a consumer-centric online personalization system", *International Journal of Computational Intelligence and Applications*, **2** (2002), 313-327.

[6] Aviv, Y., and A. Pazgal. "A partially observed Markov decision process for dynamic pricing", Working Paper (2003), Washington University at St. Louis.

[7] Bertsimas, D.J., Mersereau, A.J., and N.R. Patel. Dynamic classification of online customers. *Proceedings of SIAM International Conference on Data Mining* (2003).

[8] Bitran, G., and R. Caldentey, "Commissioned paper: an overview of pricing models for revenue management", *Manufacturing and Service Operations Management*, **5** (2003), 203-229.

[9] Elmaghraby, W., and P. Keskinocak, "Dynamic pricing in the presence of inventory considerations: research overview, current practices, and future directions", *Management Science*, **49** (2003), 1287-1309.

[10] Ernst, R., and P. Kouvelis, "The effect of selling packaged goods on inventory decisions", *Management Science*, **45** (1999), 1142-1155.

[11] E-tailing Group Inc. web page, http://www.e-tailing.com/.

[12] Federgruen, A., Groenevelt, H., and H.C. Tijms, "Coordinated replenishment in a multi-item inventory system with compound Poisson demands", *Management Science*, **30** (1984), 344-357.

[13] Federgruen, A., and A. Heching, "Combined pricing and inventory control under uncertainty", *Operations Research*, **47** (1999), 454-475.

[14] Feldman, J.M. "Package deal", *Air Transport World*, **40**, No.3 (2003), 40-42.

[15] Fuerderer, R., Herrmann, A., and G. Wuebker, editors, *Optimal packaging: marketing strategies for improving economic performance*, Springer (1999).

[16] Gallego, G., and G. van Ryzin, "Optimal dynamic pricing of inventories with stochastic demand over finite horizons", *Management Science*, **40** (1994), 999-1020.

[17] Gallego, G., and G. van Ryzin. "A multiproduct dynamic pricing problem and its applications to network yield management", *Operations Research*, **45** (1997), 24-41.

[18] Hanson, W., and R.K. Martin, "Optimal package pricing", *Management Science*, **36** (1990), 155-174.

[19] Kamakura, W.A., Wedel, M., de Rosa, F., and J.A. Mazzon. "Cross-selling through database marketing: a mixed data factor analyzer for data augmentation and prediction", *International Journal of Research in Marketing,* **20** (2003), 45-65.

[20] Maglaras, C., and J. Meissner, "Dynamic pricing strategies for multi-product revenue management problems", Working Paper (2003), Columbia University.

[21] McGill, J.I., and G. van Ryzin, "Revenue management: research overview and prospects", *Transportation Science*, **33**, (1999), 233-256.

[22] Moe, W.W., and P.S. Fader. "Dynamic conversion behavior at e-commerce sites." *Management Science,* **50**, (2004), 326-335.

[23] Monahan, G.E., N.C. Petruzzi and W. Zhao, "The dynamic pricing problem from a newsvendor perspective", *Manufacturing & Service Operations Management*, 6 (2004), 73-91.

[24] Nash, D., and A. Sterna-Karwat, "An application of DEA to measure branch cross selling efficiency." *Computers and Operations Research*, **23**, No.4 (1996), 385-392.

[25] Netessine, S., and N. Rudi, "Supply chain choice on the Internet", Working Paper (2003), University of Pennsylvania, available at http://www.netessine.com.

[26] Padmanabhan, B., and A. Tuzhilin, "On the use of optimization for data mining: theoretical interactions and eCRM opportunities", *Management Science,* **49** (2003), 1327-1343.

[27] Peters, K. "May we suggest". *Internet Retailer*, January (2004).

[28] Plambeck, E., and A. Ward, "A separation principle for assemble-to-order systems with expediting", Working Paper (2003), Stanford University.

[29] Rao, V.R., "Pricing models in marketing", in *Handbooks in operations research and management science: Marketing* by J. Eliashberg and G.L. Lilien, editors, North-Holland (1993).

[30] Scarf, H., "The optimality of (s, S) policies in the dynamic inventory problem", *Mathematical Models in Social Sciences*, K. Arrow, S. Karlin and P. Suppes, editors, Stanford University Press (1960).

[31] Stigler, G., "A note on block booking", Reprinted in *Organization of Industry*, Chicago University Press (1963).

[32] Stremersch, S., and G.J. Tellis, "Strategic packaging of products and prices: a new synthesis for marketing", *Journal of Marketing*, **66** (2002), 55-72.

[33] Veinott, A.F., "On the optimality of (s, S) inventory policies: new conditions and a new proof", *SIAM Journal of Applied Mathematics*, **14** (1966), 1067-1083.

[34] Zhang, D., and W.L. Cooper. "Revenue management for parallel flights with customer choice behavior", Working Paper (2003), University of Minnesota.

[35] Zheng, Y.-S., "A simple proof for optimality of (s, S) policies in infinite-horizon inventory systems", *Journal of Applied Probability*, **28** (1991), 802-810.

[36] Zheng, Y.-S., "Optimal control policy for stochastic inventory systems with Markovian discount opportunities", *Operations Research*, **42** (1994), 721-738.

[37] Ziya, S., H. Ayhan and R.D. Foley, "Relationship among three assumptions in revenue management", *Operations Research*, **52**, (2004), 804-809.

Appendix

Proof of Proposition 1

For the proof of Proposition we need an additional result:

Lemma A1

Let $g_{ij}(x) = \max_y \left(\bar{F}_{ij}(y)(y-x)\right)$ where $1 \le i \ne j \le m$.

a) For any $x_1 \ge 0$ and $x_2 \ge 0$,

$$\max(0, x_2 - x_1) \ge g_{ij}(x_1) - g_{ij}(x_2) \ge \min(0, x_2 - x_1). \tag{34}$$

b) For any $x_1 \ge x_2 \ge 0$, let y_k be the maximizer of $g_{ij}(x_k)$ $(k = 1, 2)$. Then $y_1 \ge y_2$.

Proof of Lemma A1

Let y_k be the maximizer of $g_{ij}(x_k)$ $(k = 1, 2)$. We consider two cases: $x_1 \ge x_2$ and $x_1 \le x_2$. If $x_1 \ge x_2$, then by definition, $g_{ij}(x_2) = \bar{F}_{ij}(y_2)(y_2 - x_2) \ge \bar{F}_{ij}(y_1)(y_1 - x_2) \ge \bar{F}_{ij}(y_1)(y_1 - x_1) = g_{ij}(x_1) \ge \bar{F}_{ij}(y_2)(y_2 - x_1)$. This yields $0 \le g_{ij}(x_2) - g_{ij}(x_1) \le \bar{F}_{ij}(y_2)(x_1 - x_2) \le x_1 - x_2$ and $\bar{F}_{ij}(y_1) \le \bar{F}_{ij}(y_2)$, i.e., $y_1 \ge y_2$. If $x_1 \le x_2$, by symmetry, we have $0 \le g_{ij}(x_1) - g_{ij}(x_2) \le x_2 - x_1$. Combining these two cases, we get (34). ■

The statement trivially holds for $n = N + 1$. Using induction, suppose that for some $n = 1, ..., N$

$$V_{n+1}(\mathbf{I}) = \textstyle\sum_{i=1}^m G_{n+1}^i(I_i). \tag{35}$$

We need to prove that $V_n(\mathbf{I}) = \sum_{i=1}^m G_n^i(I_i)$.

$$\begin{aligned}
& V_n(\mathbf{I}) \\
= & \sum_{i \in A_n, j(i) \in A_n} \lambda_i \left(p_i + V_{n+1}(\mathbf{I} - \mathbf{e}_i) + \max_{p_{i,j(i)}} \left(\bar{F}_{i,j(i)}(p_{i,j(i)}) \left(p_{i,j(i)} - p_i + V_{n+1}(\mathbf{I} - \mathbf{e}_i - \mathbf{e}_{j(i)}) - V_{n+1}(\mathbf{I} - \mathbf{e}_i) \right) \right) \right) \\
& + \sum_{i \in A_n, j(i) \notin A_n} \lambda_i \left(p_i + V_{n+1}(\mathbf{I} - \mathbf{e}_i) + \max_{p_{i,j(i)}} \left(\bar{F}_{i,j(i)}(p_{i,j(i)}) \left(p_{i,j(i)} - p_i - b_{j(i)} \right) \right) \right) \\
& + \sum_{i \notin A_n, j(i) \in A_n} \lambda_i \left(p_i - b_i + V_{n+1}(\mathbf{I}) + \max_{p_{i,j(i)}} \bar{F}_{i,j(i)}(p_{i,j(i)}) \left(p_{i,j(i)} - p_i + V_{n+1}(\mathbf{I} - \mathbf{e}_{j(i)}) - V_{n+1}(\mathbf{I}) \right) \right) \\
& + \sum_{i \notin A_n, j(i) \notin A_n} \lambda_i \left(p_i - b_i + V_{n+1}(\mathbf{I}) + \max_{p_{i,j(i)}} \bar{F}_{i,j(i)}(p_{i,j(i)}) \left(p_{i,j(i)} - p_i - b_{j(i)} \right) \right) \\
& + \left(1 - \sum_{i=1}^m \lambda_i \right) V_{n+1}(\mathbf{I}) \\
= & \sum_{i \in A_n, j(i) \in A_n} \lambda_i \left(p_i + G_{n+1}^i(I_i - 1) - G_{n+1}^i(I_i) + \max_{p_{i,j(i)}} \bar{F}_{i,j(i)}(p_{i,j(i)}) \left(p_{i,j(i)} - p_i + G_{n+1}^{j(i)}(I_{j(i)} - 1) - G_{n+1}^{j(i)}(I_{j(i)}) \right) \right) \\
& + \sum_{i \in A_n, j(i) \notin A_n} \lambda_i \left(p_i + G_{n+1}^i(I_i - 1) - G_{n+1}^i(I_i) + \max_{p_{i,j(i)}} \bar{F}_{i,j(i)}(p_{i,j(i)}) \left(p_{i,j(i)} - p_i - b_{j(i)} \right) \right) \\
& + \sum_{i \notin A_n, j(i) \in A_n} \lambda_i \left(p_i - b_i + \max_{p_{i,j(i)}} \bar{F}_{i,j(i)}(p_{i,j(i)}) \left(p_{i,j(i)} - p_i + G_{n+1}^{j(i)}(I_{j(i)} - 1) - G_{n+1}^{j(i)}(I_{j(i)}) \right) \right)
\end{aligned}$$

$$+\sum_{i\notin A_n, j(i)\notin A_n} \lambda_i \left(p_i - b_i + \max_{p_{i,j(i)}} \bar{F}_{i,j(i)}(p_{i,j(i)}) \left(p_{i,j(i)} - p_i - b_{j(i)}\right)\right) + \sum_{i=1}^{m} G^i_{n+1}(I_i) \quad \text{(from (35))}$$

$$= \sum_{i=1}^{m} G^i_n(I_i). \quad \text{(from (9))} \blacksquare$$

Proof of Proposition 3

a) Following the standard Lagrangian procedure to derive the dual function $q(\mu,\nu)$, we have

$$\begin{aligned}
q(\mu,\nu) &= \max_{q\geqslant 0, y\geqslant 0, p_{ij}\geqslant 0} L(p,q,y,\mu,\nu) \\
&= \max_{q\geqslant 0, y\geqslant 0, p_{ij}\geqslant 0} \left\{ \begin{array}{c} \sum_{i=1}^m \sum_{j\neq i} \lambda_j \tau q_{ji} \overline{F}_{ji}(p_{ji})(p_{ji}-p_j) - \sum_{i=1}^m b_i y_i + \\ \sum_{i=1}^m \mu_i[y_i + I_i - \lambda_i\tau - \sum_{j\neq i}\lambda_j \tau q_{ji}\overline{F}_{ji}(p_{ji})] + \sum_{i=1}^m \nu_i(1-\sum_{j\neq i} q_{ij}) \end{array} \right\} \\
&= \max_{q\geqslant 0, y\geqslant 0, p_{ij}\geqslant 0} \left\{ \begin{array}{c} \sum_{i=1}^m \sum_{j\neq i} [\lambda_j \tau \overline{F}_{ji}(p_{ji})(p_{ji}-p_j-\mu_i) - \nu_j] q_{ji} + \sum_{i=1}^m (\mu_i - b_i) y_i + \\ \sum_{i=1}^m [\mu_i(I_i - \lambda_i\tau) + \nu_i] \end{array} \right\} \\
&= \max_{q\geqslant 0, y\geqslant 0} \left\{ \sum_{i=1}^m \sum_{j\neq i} [\lambda_j \tau H_{ji}(\mu_i) - \nu_j] q_{ji} + \sum_{i=1}^m (\mu_i - b_i) y_i + \sum_{i=1}^m [\mu_i(I_i-\lambda_i\tau)+\nu_i] \right\} \qquad (36)
\end{aligned}$$

where the last equation follows from the definition of $H_{ji}(\mu_i)$. Define $D_q = \{(\mu,\nu)|q(\mu,v) < +\infty\}$. It is easy to see from (36) that $D_q = \{(\mu,\nu)|\lambda_j\tau H_{ji}(\mu_i) - v_j \leq 0, \text{for } i\neq j, \text{and } \mu_i \leq b_i\}$, and $q(\mu,\nu) = \sum_{i=1}^m [\mu_i(I_i - \lambda_i\tau) + \nu_i]$ for $(\mu,\nu)\in D_q$. Hence the dual problem (**D2**) can be formulated as follows:

$$\begin{aligned}
&\min_{\mu_i, v_j} : q(\mu,v) = \sum\nolimits_{i=1}^m [\mu_i(I_i - \lambda_i\tau) + v_i] \\
&\text{s.t. } \lambda_j\tau H_{ji}(\mu_i) \leq v_j \text{ for } i\neq j, \\
&0 \leq \mu_i \leq b_i \text{ for } i = 1,2,...,m.
\end{aligned}$$

In order to show that the above constraint set is convex, it suffices to show that $H_{ji}(x)$ is convex for every pair $i\neq j$. Recall that $H_{ji}(x) = \max_{p_{ji}} \left(\bar{F}_{ji}(p_{ji})(p_{ji}-p_j-x)\right)$. Denote the maximizer of $\bar{F}_{ji}(p_{ji})(p_{ji}-p_j-x)$ by $p_{ji}(x)$. Then from the first-order condition, we have

$$\frac{d\bar{F}_{ji}(p_{ji})(p_{ji}-p_j-x)}{dp_{ji}}|_{p_{ji}=p_{ji}(x)} = 0. \qquad (37)$$

Taking the derivative of $H_{ji}(x)$, we have

$$\frac{dH_{ji}(x)}{dx} = \frac{d\bar{F}_{ji}(p_{ji})(p_{ji}-p_j-x)}{dp_{ji}}|_{p_{ji}=p_{ji}(x)} \frac{dp_{ji}(x)}{dx} - \bar{F}_{ji}(p_{ji}(x)) = -\bar{F}_{ji}(p_{ji}(x)), \qquad (38)$$

where the last equality follows from (37). Hence $d^2H_{ji}(x)/dx^2 = (dF(x)/dx)\,(dp_{ji}(x)/dx) \geqslant 0$, which follows from part b) of Lemma A1 and the fact that $F(x)$ is a cumulative distribution function. This proves part a).

b) It is easy to ensure that the constraint set of (**D2**) is compact by imposing a finite bound on v. Hence, the Weierstrass Theorem ensures the existence of the optimal solution to (**D2**). Denote the optimal solution to (**D2**) by (μ^*, ν^*). It is easy to verify that any solution to (**D2**) is regular. It follows from the KKT necessary conditions that there exist $\{q^*_{ij}\}$ and $\{y^*_i\}$ satisfying the following KKT conditions of (**D2**) (note that (37) is used in deriving the KKT conditions):

$$\left(y^*_i + I_i - \lambda_i\tau - \sum\nolimits_{j\neq i} \lambda_j\tau q^*_{ji}\overline{F}_{ji}(p_{ji}(\mu^*_i))\right)\mu^*_i = 0, \text{ for } i = 1, 2, ..., m, \tag{39}$$

$$\sum\nolimits_{j\neq i} q^*_{ij} = 1, \text{ for } i = 1, 2, ..., m, \tag{40}$$

$$(\mu^*_i - b_i)y^*_i = 0, \text{ for } i = 1, 2, ..., m, \tag{41}$$

$$(\lambda_j\tau H_{ji}(\mu^*_i) - v^*_j)q^*_{ji} = 0 \tag{42}$$

$$y^*_i \geqslant 0, \text{ for } i = 1, 2, ..., m, \tag{43}$$

$$y^*_i \geqslant -I_i + \lambda_i\tau + \sum\nolimits_{j\neq i} \lambda_j\tau q^*_{ji}\overline{F}_{ji}(p_{ji}(\mu^*_i)), \text{ for } i = 1, 2, ..., m, \tag{44}$$

$$q^*_{ij} \geqslant 0, \text{ for } i \neq j. \tag{45}$$

Define $p^*_{ji} = p_{ji}(\mu^*_i)$. Note that (40), (43), (44), and (45) are the four constraints of (**P2**). Hence, $\{\{q^*_{ij}\}, \{y^*_i\}, \{p^*_{ij}\}\}$ is a feasible solution to (**P2**), and it yields the following objective value:

$$\begin{aligned}
& \sum\nolimits_{i=1}^{m} \sum\nolimits_{j\neq i} \lambda_j\tau q^*_{ji}\overline{F}_{ji}(p^*_{ji})(p^*_{ji} - p_j) - \sum\nolimits_{i=1}^{m} b_i y^*_i \\
= & \sum\nolimits_{i=1}^{m} \sum\nolimits_{j\neq i} \lambda_j\tau q^*_{ji}\overline{F}_{ji}(p^*_{ji})(p^*_{ji} - p_j) - \sum\nolimits_{i=1}^{m} \mu^*_i y^*_i \quad \text{(by (41))} \\
= & \sum\nolimits_{i=1}^{m} \sum\nolimits_{j\neq i} \lambda_j\tau q^*_{ji}\overline{F}_{ji}(p^*_{ji})(p^*_{ji} - p_j - \mu^*_i) + \sum\nolimits_{i=1}^{m} \mu^*_i(I_i - \lambda_i\tau) \quad \text{(by (39))} \\
= & \sum\nolimits_{i=1}^{m} \sum\nolimits_{j\neq i} \lambda_j\tau q^*_{ji}H_{ji}(\mu^*_i) + \sum\nolimits_{i=1}^{m} \mu^*_i(I_i - \lambda_i\tau) \text{ (by definition of } H_{ji}(x)) \\
= & \sum\nolimits_{i=1}^{m} \sum\nolimits_{j\neq i} q^*_{ji}v^*_j + \sum\nolimits_{i=1}^{m} \mu^*_i(I_i - \lambda_i\tau) \quad \text{(by (42))} \\
= & \sum\nolimits_{i=1}^{m} v^*_i + \sum\nolimits_{i=1}^{m} \mu^*_i(I_i - \lambda_i\tau), \quad \text{(by (40))}
\end{aligned}$$

which is equal to the optimal value of the dual. By weak duality, $\{\{q^*_{ij}\}, \{y^*_i\}, \{p^*_{ij}\}\}$ is an optimal solution to (**P2**). This proves i) of part b. One can obtain the optimal solution to (**P2**) using the above KKT conditions. An alternative way is to solve a linear programming problem (**P2**) with p_{ij} being replaced by p^*_{ij}. This proves ii) of part b. ■

Proof of Proposition 4

Since $q < p + b$, it is suboptimal to choose a product with zero inventory as a packaging complement. Therefore, the dynamic packaging problem reduces to:

$$\begin{aligned}
V_n(\mathbf{I}) = & \sum\nolimits_{i\in A(\mathbf{I})} \lambda \left((1-\gamma)(p + V_{n+1}(\mathbf{I} - \mathbf{e}_i)) + \max_{j\neq i,\, j\in A(\mathbf{I})} \gamma\left(q + V_{n+1}(\mathbf{I} - \mathbf{e}_i - \mathbf{e}_j)\right)\right) \\
& + \sum\nolimits_{i\notin A(\mathbf{I})} \lambda \left((1-\gamma)(p - b + V_{n+1}(\mathbf{I})) + \max_{j\neq i,\, j\in A(\mathbf{I})} \gamma\left(q - b + V_{n+1}(\mathbf{I} - \mathbf{e}_j)\right)\right) \\
& + (1 - m\lambda)V_{n+1}(\mathbf{I}),
\end{aligned} \tag{46}$$

where $A(\mathbf{I})$ denotes products that have at least one unit of inventory at the beginning of the n-th decision epoch when the inventory is $\mathbf{I}$. The boundary condition is $V_{N+1}(\mathbf{I})=0$. In order to prove the optimality of the DR packaging, we need to show that $V_n(\mathbf{I}-\mathbf{e}_k)\geq V_n(\mathbf{I}-\mathbf{e}_l)$ for any $\mathbf{I}=(I_1,I_2,...,I_m)$ with $I_k\geq I_l\geq 1$ and any $n=1,...,N+1$. First, from the boundary condition, this inequality trivially holds when $n=N+1$. By induction, suppose it also holds up to some $n+1$, i.e.,

$$V_s(\mathbf{I}-\mathbf{e}_k)\geq V_s(\mathbf{I}-\mathbf{e}_l)\,,\ \text{for any } \mathbf{I}=(I_1,I_2,...,I_m) \text{ with } I_k\geq I_l\geq 1 \text{ and } s=n+1,...,N+1. \tag{47}$$

We prove the desired result for two separate cases: $I_k=I_l=1$ and $I_k>I_l=1$ (the proof for the case of $I_k,I_l>1$ follows similar steps).

Case 1: $I_k=I_l=1$.

Define $B=\{i|i\neq k,l,\text{and } I_i>0\}$ and $C=\{i|i\neq k,l,\text{and } I_i=0\}$. Then $A(\mathbf{I}-\mathbf{e}_k)=B\cup\{l\}$ and $A(\mathbf{I}-\mathbf{e}_l)=B\cup\{k\}$. Note that for any $i\in B$, by (47), we have

$$\max_{j\neq i,\,j\in B}\gamma\left(q+V_{n+1}(\mathbf{I}-\mathbf{e}_i-\mathbf{e}_j-\mathbf{e}_k)\right)\geq \max_{j\neq i,\,j\in B}\gamma\left(q+V_{n+1}(\mathbf{I}-\mathbf{e}_i-\mathbf{e}_j-\mathbf{e}_l)\right),$$

which implies that

$$\max_{j\neq i,\,j\in B\cup\{l\}}\gamma\left(q+V_{n+1}(\mathbf{I}-\mathbf{e}_i-\mathbf{e}_j-\mathbf{e}_k)\right)\geq \max_{j\neq i,\,j\in B\cup\{k\}}\gamma\left(q+V_{n+1}(\mathbf{I}-\mathbf{e}_i-\mathbf{e}_j-\mathbf{e}_l)\right). \tag{48}$$

Similarly, for any $i\in C$, by (47), we have

$$\max_{j\neq i,\,j\in B\cup\{l\}}\gamma\left(q-b+V_{n+1}(\mathbf{I}-\mathbf{e}_j-\mathbf{e}_k)\right)\geq \max_{j\neq i,\,j\in B\cup\{k\}}\gamma\left(q-b+V_{n+1}(\mathbf{I}-\mathbf{e}_j-\mathbf{e}_l)\right). \tag{49}$$

By(46), we have

$$\begin{aligned}
& V_n(\mathbf{I}-\mathbf{e}_k) \\
= \; & \textstyle\sum_{i\in B}\lambda\left((1-\gamma)(p+V_{n+1}(\mathbf{I}-\mathbf{e}_i-\mathbf{e}_k))+\max\limits_{j\neq i,\,j\in B\cup\{l\}}\gamma\left(q+V_{n+1}(\mathbf{I}-\mathbf{e}_i-\mathbf{e}_j-\mathbf{e}_k)\right)\right) \\
& +\lambda\left((1-\gamma)(p+V_{n+1}(\mathbf{I}-\mathbf{e}_l-\mathbf{e}_k))+\max\limits_{j\neq l,\,j\in B\cup\{l\}}\gamma\left(q+V_{n+1}(\mathbf{I}-\mathbf{e}_l-\mathbf{e}_j-\mathbf{e}_k)\right)\right) \\
& +\sum_{i\in C}\lambda\left((1-\gamma)(p-b+V_{n+1}(\mathbf{I}-\mathbf{e}_k))+\max\limits_{j\neq i,\,j\in B\cup\{l\}}\gamma\left(q-b+V_{n+1}(\mathbf{I}-\mathbf{e}_j-\mathbf{e}_k)\right)\right) \\
& +\lambda\left((1-\gamma)(p-b+V_{n+1}(\mathbf{I}-\mathbf{e}_k))+\max\limits_{j\neq k,\,j\in B\cup\{l\}}\gamma\left(q-b+V_{n+1}(\mathbf{I}-\mathbf{e}_j-\mathbf{e}_k)\right)\right) \\
& +(1-m\lambda)V_{n+1}(\mathbf{I}-\mathbf{e}_k)\,, \\
\geq \; & \textstyle\sum_{i\in B}\lambda\left((1-\gamma)(p+V_{n+1}(\mathbf{I}-\mathbf{e}_i-\mathbf{e}_l))+\max\limits_{j\neq i,\,j\in B\cup\{k\}}\gamma\left(q+V_{n+1}(\mathbf{I}-\mathbf{e}_i-\mathbf{e}_j-\mathbf{e}_l)\right)\right) \\
& +\lambda\left((1-\gamma)(p+V_{n+1}(\mathbf{I}-\mathbf{e}_k-\mathbf{e}_l))+\max\limits_{j\neq l,\,j\in B\cup\{k\}}\gamma\left(q+V_{n+1}(\mathbf{I}-\mathbf{e}_k-\mathbf{e}_j-\mathbf{e}_l)\right)\right)
\end{aligned}$$

$$
\begin{aligned}
&+\textstyle\sum_{i\in C}\lambda\left((1-\gamma)(p-b+V_{n+1}(\mathbf{I}-\mathbf{e}_l))+\max_{j\neq i,\,j\in B\cup\{k\}}\gamma\left(q-b+V_{n+1}(\mathbf{I}-\mathbf{e}_j-\mathbf{e}_l)\right)\right)\\
&+\lambda\left((1-\gamma)(p-b+V_{n+1}(\mathbf{I}-\mathbf{e}_l))+\max_{j\neq l,\,j\in B\cup\{k\}}\gamma\left(q-b+V_{n+1}(\mathbf{I}-\mathbf{e}_j-\mathbf{e}_l)\right)\right)\\
&+(1-m\lambda)V_{n+1}(\mathbf{I}-\mathbf{e}_l)\ \text{ (by (47), (48), and (49))}\\
=\ & V_n(\mathbf{I}-\mathbf{e}_l).
\end{aligned}
$$

Case 2: $I_k > I_l = 1$.

Note that for any $n = 1, 2, ..., N$, $i = 1, 2, ..., m$, and inventory vector $\mathbf{I}$, we have $V_n(\mathbf{I} - \mathbf{e}_i) + b \geq V_n(\mathbf{I})$, which together with (47), implies that

$$
V_{n+1}(\mathbf{I}-\mathbf{e}_k-\mathbf{e}_k) \geq -b + V_{n+1}(\mathbf{I}-\mathbf{e}_l), \tag{50}
$$

and for any $j \in B$,

$$
V_{n+1}(\mathbf{I}-\mathbf{e}_k-\mathbf{e}_j-\mathbf{e}_k) \geq -b + V_{n+1}(\mathbf{I}-\mathbf{e}_j-\mathbf{e}_l). \tag{51}
$$

$$
\begin{aligned}
& V_n(\mathbf{I}-\mathbf{e}_k)\\
=\ & \sum_{i\in B}\lambda\left((1-\gamma)(p+V_{n+1}(\mathbf{I}-\mathbf{e}_i-\mathbf{e}_k))+\max_{j\neq i,\,j\in B\cup\{k,l\}}\gamma\left(q+V_{n+1}(\mathbf{I}-\mathbf{e}_i-\mathbf{e}_j-\mathbf{e}_k)\right)\right)\\
&+\lambda\left((1-\gamma)(p+V_{n+1}(\mathbf{I}-\mathbf{e}_l-\mathbf{e}_k))+\max_{j\neq l,\,j\in B\cup\{k,l\}}\gamma\left(q+V_{n+1}(\mathbf{I}-\mathbf{e}_l-\mathbf{e}_j-\mathbf{e}_k)\right)\right)\\
&+\lambda\left((1-\gamma)(p+V_{n+1}(\mathbf{I}-\mathbf{e}_k-\mathbf{e}_k))+\max_{j\neq k,\,j\in B\cup\{k,l\}}\gamma\left(q+V_{n+1}(\mathbf{I}-\mathbf{e}_k-\mathbf{e}_j-\mathbf{e}_k)\right)\right)\\
&+\sum_{i\in C}\lambda\left((1-\gamma)(p-b+V_{n+1}(\mathbf{I}-\mathbf{e}_k))+\max_{j\neq i,\,j\in B\cup\{k,l\}}\gamma\left(q-b+V_{n+1}(\mathbf{I}-\mathbf{e}_j-\mathbf{e}_k)\right)\right)\\
&+(1-m\lambda)V_{n+1}(\mathbf{I}-\mathbf{e}_k)\\
\geq\ & \sum_{i\in B}\lambda\left((1-\gamma)(p+V_{n+1}(\mathbf{I}-\mathbf{e}_i-\mathbf{e}_l))+\max_{j\neq i,\,j\in B\cup\{k\}}\gamma\left(q+V_{n+1}(\mathbf{I}-\mathbf{e}_i-\mathbf{e}_j-\mathbf{e}_l)\right)\right)\\
&+\lambda\left((1-\gamma)(p+V_{n+1}(\mathbf{I}-\mathbf{e}_k-\mathbf{e}_l))+\max_{j\neq l,\,j\in B\cup\{k\}}\gamma\left(q+V_{n+1}(\mathbf{I}-\mathbf{e}_k-\mathbf{e}_j-\mathbf{e}_l)\right)\right)\\
&+\sum_{i\in C}\lambda\left((1-\gamma)(p-b+V_{n+1}(\mathbf{I}-\mathbf{e}_l))+\max_{j\neq i,\,j\in B\cup\{k\}}\gamma\left(q-b+V_{n+1}(\mathbf{I}-\mathbf{e}_j-\mathbf{e}_l)\right)\right)\\
&+\lambda\left((1-\gamma)(p-b+V_{n+1}(\mathbf{I}-\mathbf{e}_l))+\max_{j\neq l,\,j\in B\cup\{k\}}\gamma\left(q-b+V_{n+1}(\mathbf{I}-\mathbf{e}_j-\mathbf{e}_l)\right)\right)\\
&+(1-m\lambda)V_{n+1}(\mathbf{I}-\mathbf{e}_l)\ \text{ (by (47), (50), and (51))}\\
=\ & V_n(\mathbf{I}-\mathbf{e}_l).\ \blacksquare
\end{aligned}
$$

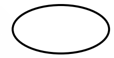
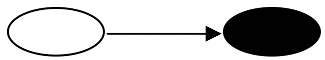
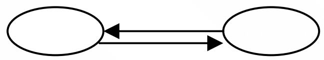
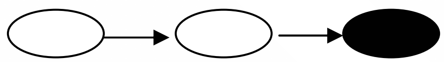
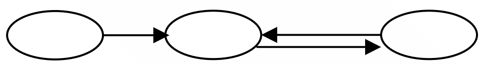
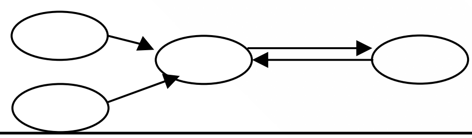
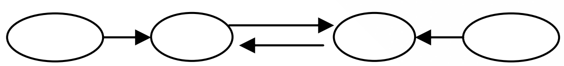
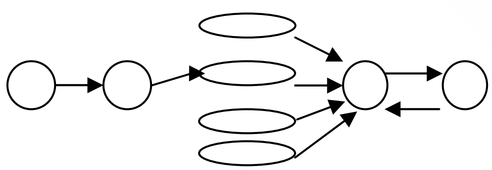
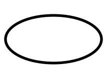
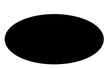
Packaging Configuration	Number of Observations
	7
	51
	6
	14
	6
	4
	4
	8

Table 1: Packaging configurations for the top 100 best selling books at Amazon.com:
 - book on the top100 list,  - book not on the top 100 list.

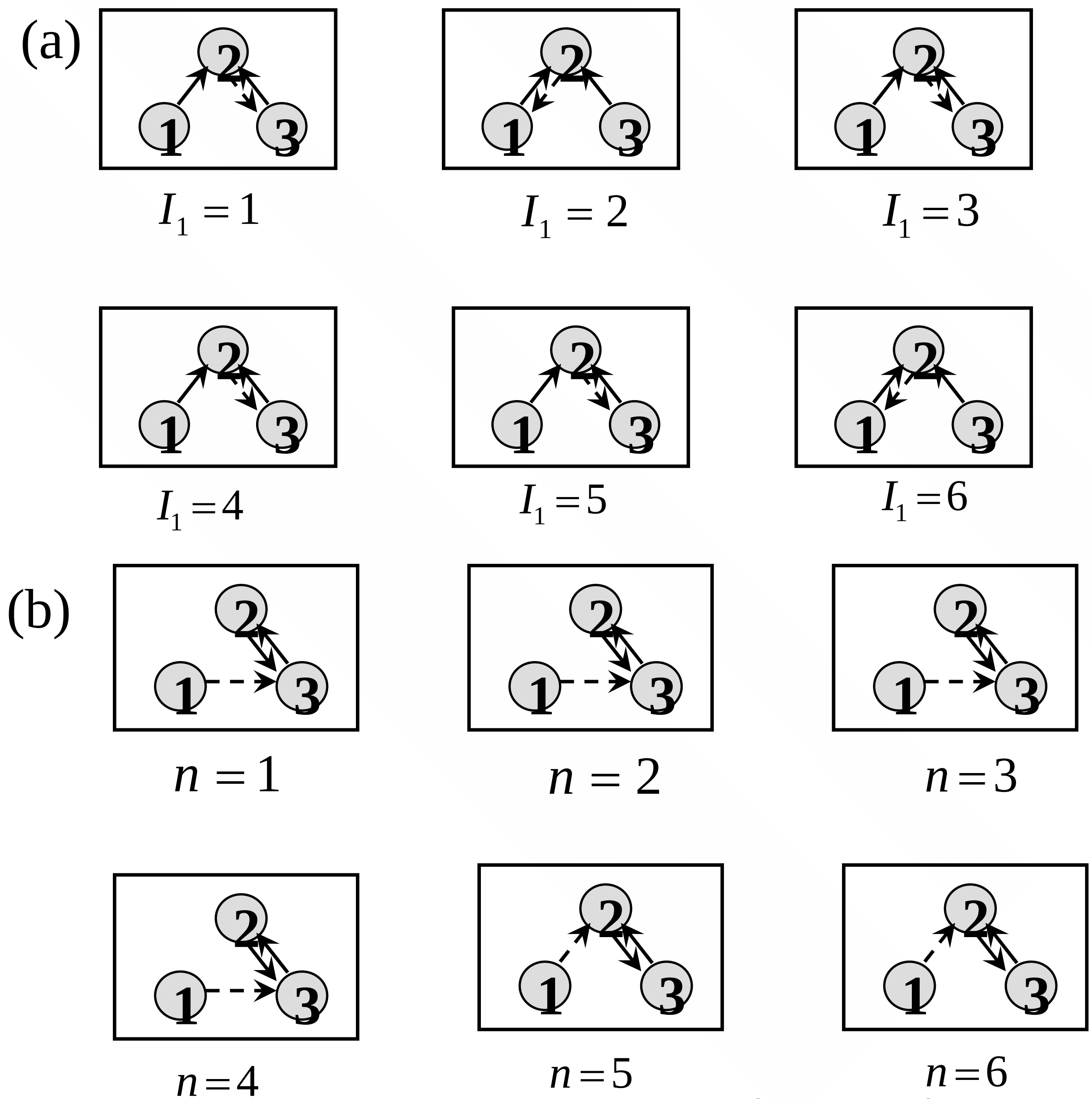


Fig.1. 3-product case: optimal dynamic packaging as a function of the state of the system (a – for I_2=2, I_3=4, n=1) and time (b – for for I_1=1, I_2=2, I_3=2). N=10, $\overline{F}_{ij}\left(p_{ij}\right) = \left(\left(p_i + p_j - p_{ij}\right)/p_j\right)^{\beta}$, p_1=β= p_3=1, p_2=1.5.

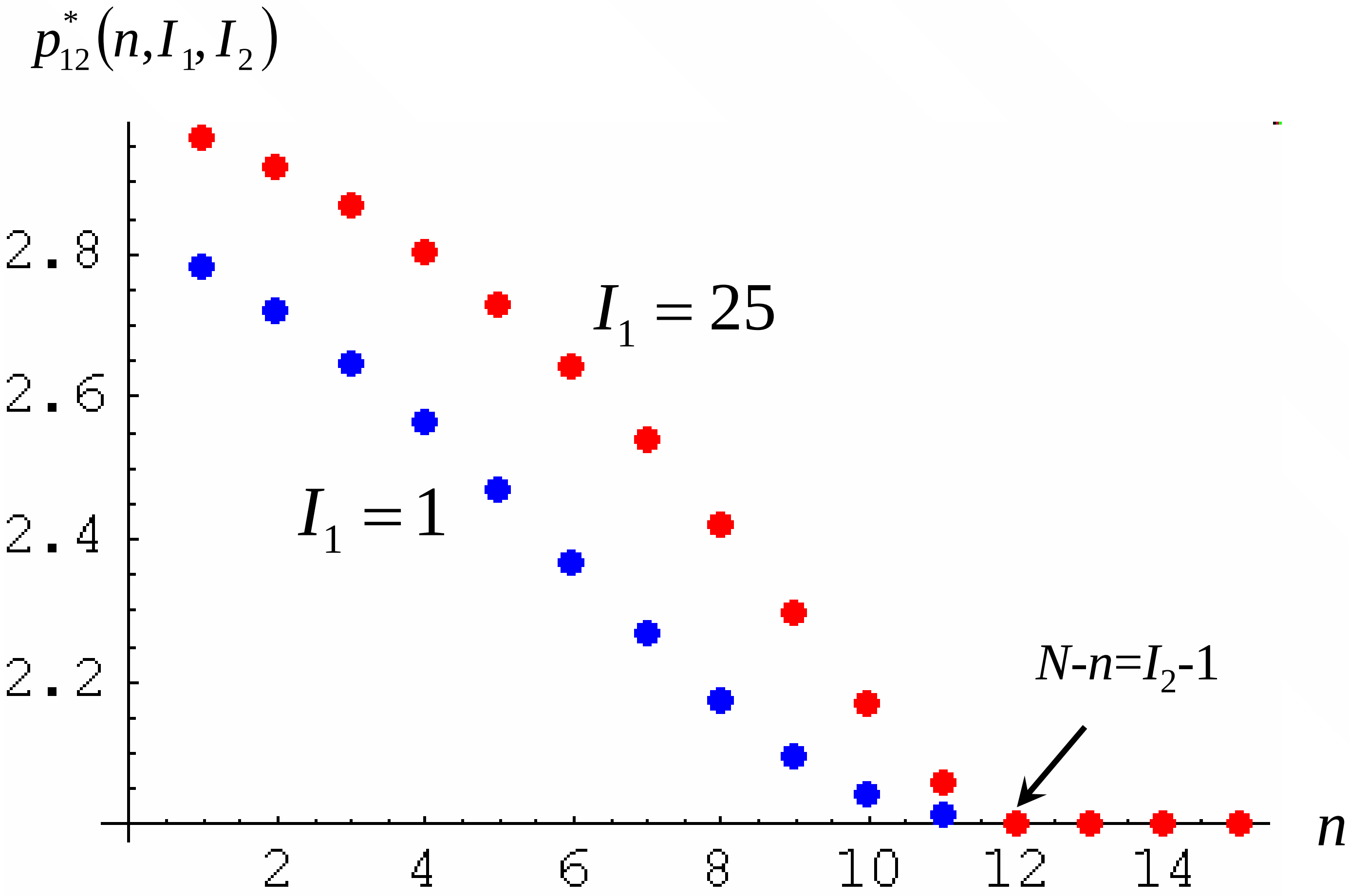


Fig. 2. Optimal price for the "1-2" package as a function of the time index n for the Lost Sales model. Pricing functions and model parameters: $\overline{F}_{ij}(p_{ij})=((p_i+p_j-p_{ij})/p_i)^{\beta}$, $\underline{p}_{ij} = p_i$, $\overline{p}_{ij} = p_i + p_j$, $\beta = 1$, $p_1 = 1$, $p_2 = 2$, $\lambda_1 = \lambda_2 = 0.4$, $I_2 = 5$.