Annals of Operations Research manuscript No.
(will be inserted by the editor)

Resource Capacity Allocation to Stochastic Dynamic Competitors: Knapsack Problem for Perishable Items and Index-Knapsack Heuristic

Peter Jacko

Received: September 16, 2011 / Revised: February 27, 2012 / Revised: December 18, 2012 / Accepted: date

Abstract In this paper we propose an approach for solving problems of optimal resource capacity allocation to a collection of stochastic dynamic competitors. In particular, we introduce the knapsack problem for perishable items, which concerns the optimal dynamic allocation of a limited knapsack to a collection of perishable or non-perishable items. We formulate the problem in the framework of Markov decision processes, we relax and decompose it, and we design a novel index-knapsack heuristic which generalizes the index rule and it is optimal in some specific instances. Such a heuristic bridges the gap between static/deterministic optimization and dynamic/stochastic optimization by stressing the connection between the classic knapsack problem and dynamic resource allocation. The performance of the proposed heuristic is evaluated in a systematic computational study, showing an exceptional near-optimality and a significant superiority over the index rule and over the benchmark earlier-deadline-first policy. Finally we extend our results to several related revenue management problems.

Keywords resource allocation · Markov decision processes · knapsack problem · restless bandits · Whittle index · perishability · revenue management · retailing

An earlier version of this paper was presented at the BCAM Workshop on Bandit Problems (Bilbao, 2011) and at the INFORMS Applied Probability Society Conference (Stockholm, 2011). Research partially supported by grant MTM2010-17405 of the MICINN (Spain) and grant PI2010-2 of the Department of Education and Research (Basque Government).

P. Jacko
BCAM – Basque Center for Applied Mathematics
Mazarredo 14, 48009 Bilbao, Spain
Tel.: +34-94-6567 842
Fax: +34-94-6567 843
E-mail: jacko@bcamath.org

1 Introduction

The knapsack problem (Dantzig, 1957) is the fundamental and well-studied operations research model providing insights into the solution of more complex discrete resource capacity allocation problems. Recently, there has been a surge in the need of addressing resource capacity allocation problems in stochastic and dynamic environment in different fields. Remarkable examples include

- workforce management (allocation of number of employees to teams, e.g., for surgeries, machine repairs, client-based consultancy) (Glazebrook et al, 2005)
- dynamic allocation of the number of or the power used in transmission channels/frequences in wireless network base stations to competing users (Gesbert et al, 2007; Jacko, 2011b)
- resource allocation for multi-queue systems with a shared server pool (Yang et al, 2011; Dance and Gaivoronski, 2012; Glazebrook et al, 2011)
- dynamic allocation of a money budget to research and development projects (Loch and Kavadias, 2002; Qu and Gittins, 2011)
- dynamic allocation of machines to production of seasonal goods (Caro and Gallien, 2007)
- scheduling of (i.e., allocation of processing time to) stochastic simulations of design alternatives (Chick and Gans, 2009)
- service partitioning (allocation of the number of virtual machines) in data centers to competing processing requests (Speitkamp and Bichler, 2010; Anselmi and Verloop, 2011)
- shelf-space allocation in supermarkets (this paper).

In all of these problems, the inherent combinatorial considerations are further escalated due to the additional trade-off between exploration and exploitation. Moreover, obtaining an optimal solution to stochastic dynamic problems is often intractable due to the *curse of dimensionality*. This paper proposes a mathematical approach to a particular resource capacity allocation problem, where several stochastically and dynamically evolving competitors demand part of the capacity. A pragmatic aim is to design a well-grounded close-to-optimal dynamic solution that is generalizable to other similar or more complex problems, and that is optimal in some specific instances of the problem. This is in a direct opposition to proposing *ad-hoc* solutions for a given problem in hand, to deriving an optimal *static* solution, to obtaining dynamic solutions by *approximate* techniques (e.g., solving optimally a problem with truncated/reduced state space or time horizon, employing approximate dynamic programming, stochastic programming, simulation, etc.), or to obtaining dynamic solutions by *numerical* approaches (e.g., metaheuristics) without performance guarantees.

For the sake of concreteness, the model of this paper is presented in the setting motivated by optimal allocation of promotion space in a supermarket, where the manager has a possibility to select a number of products in order to maximize the expected revenue. We focus on perishable products with individual deadlines (non-perishable products are considered as a limiting case), therefore we refer to it as the *knapsack problem for perishable items* (KPPI). Perishability is a common phenomenon also in other fields, for instance due to contracts involving a Quality-of-Service clause.

A Markov decision process (MDP) model of the problem is formulated in Section 2. This is a constrained MDP with special structure, which belongs to the family of *weakly-coupled* MDPs (Meuleau et al, 1998). In Section 3 we discuss the relationship of KPPI with the knapsack problem and with the multi-armed restless bandit problem, which will be insightful for identifying the computational complexity and for indicating a natural direction to approach the problem. In Section 4 we present a relaxation and decomposition of the problem into single-item subproblems. While the Lagrangian approach is well-known and applied often in similar problems, we take a step further, in order to develop a solution based on so-called index values. Section 5 is dedicated to the study of indexability and derivation of closed-form index values for the single-item subproblem. The index-knapsack heuristic is developed in Section 6, where we identify cases in which it recovers optimal solutions, and discuss the intuition behind it. Performance of the index-knapsack heuristic is then studied in computational experiments presented in Section 7, showing an excellent nearly-optimal behavior and further outperforming conventional solutions. Application of our results to several revenue management problems, including variants of the dynamic product assortment problem, dynamic product pricing problem, and a loyalty card problem, is presented in Appendix A.

2 Knapsack Problem for Perishable Items

In KPPI, we assume that demand can be increased by dynamically allocating products to a limited *promotion space*, where they are more likely to attract customers. An example of the practical interest of such a tool is provided by the cooperation of Capgemini, Intel, Cisco, and Microsoft on a decision support system called *Extended Retail Solutions* which includes Dynamic Promotion Management as one of three key solution areas (cf. Capgemini et al, 2005). In their setting, the limited promotion space is given by the space and time available on the customer's loyalty card, which is used to inform and influence the particular customer by personalized messages. More conventional examples of such a promotion space include shelves close to the cash register, end-aisle displays, promotion kiosks, or a depot used for selling via the Internet.

A *perishable item* is a product unit with an associated lifetime ending at a *deadline*. At the deadline (e.g., the "best before" date) the product can no longer be sold, and only a *salvage value* is received. If an item is sold before the deadline, it yields a *revenue* (profit margin). The probability of selling depends only on whether the item is being promoted or not. The concern of KPPI is to dynamically select a subset of items to be included in a promotion space (knapsack), in order to maximize the expected total discounted sum of revenues and salvage values.

We formulate the model in discrete time as a Markov decision process. We assume that the decisions are made in some regular time moments (say, twice a day), and the problem parameters are adjusted to such time periods. Consider the time slotted into time epochs $s \in \mathcal{S} := \{0, 1, 2, \dots\}$ at which decisions can be made. Time epoch s corresponds to the beginning of time period s. Revenues are discounted over time with factor $0 \leq \beta \leq 1$.

In general, the KPPI defines a stochastic and dynamic variant of the knapsack problem with multiple units of items. As time evolves, items get sold accordingly to a stochastic demand or they perish deterministically at their deadlines. For trans-

parency, we assume in this paper that the demand is time-homogeneous (see Elmaghraby and Keskinocak (2003) for a justification of such an assumption) and we consider a single unit of each product. This assumption is, nevertheless, not crucial for our derivation of the solution.

Consider a retailer that has I perishable items to sell, labeled by $i \in \mathcal{I}$.[1] Suppose that the promotion space (knapsack) is available with capacity of $W \geq 1$ physical space units. We assume that this promotion space is *fully regenerative*, i.e., its full capacity is repetitively available at every time epoch. The capacity not used at a given epoch is lost, i.e., the promotion space is *nonmarketable*.

2.1 MDP Model of Perishable Item

In this subsection we focus on a single item and formalize it within the MDP framework.

Item i can only be sold during its lifetime, which consists of time periods $0, 1, \ldots, T_i - 1$, where $1 \leq T_i \leq \infty$ is the item's deadline. The item is on sale until the end of period $T_i - 1$, when it is removed as perished and cannot be sold anymore. If the item is sold, it yields a revenue (profit margin) $R_i > 0$ at that period. Otherwise, a salvage value is obtained in period T_i, whose expected value is denoted by $\alpha_i R_i$ for some (possibly negative) coefficient $\alpha_i \leq 1$.

The retailer can change the probability that item i is sold during a period, from $1 - q_i$ to $1 - p_i$ (with $0 < p_i, q_i \leq 1$), by placing it in a *promotion space* (knapsack). Formally,

$$p_i := \mathbb{P}\{\text{the item } i \text{ when promoted is not sold in a period}\},$$
$$q_i := \mathbb{P}\{\text{the item } i \text{ when not promoted is not sold in a period}\}.$$

We assume that such Bernoulli demand processes are independent across items. The difference $q_i - p_i$ will be called *promotion power*, as it captures the increase in the probability of being sold caused by promoting. Item i occupies $W_i \leq W$ units, and for non-triviality we assume that $\sum_i W_i > W$.

To formulate the perishable item as an MDP, we define its elements as the tuple

$$\left(\mathcal{X}_i, (\boldsymbol{W}_i^a)_{a\in\mathcal{A}}, (\boldsymbol{R}_i^a)_{a\in\mathcal{A}}, (\boldsymbol{P}_i^a)_{a\in\mathcal{A}}\right),$$

where

- The state space is $\mathcal{X}_i := \mathcal{T}_i \cup \{0\}$, where state $t \in \mathcal{T}_i := \{1, 2, \ldots, T_i\}$ means that *there are t remaining periods to the deadline and the item has not been sold* , while state 0 is an absorbing state representing a perished and/or sold item;
- The action space for states in $\mathcal{T}_i$ is $\mathcal{A} := \{0, 1\}$: we can either *promote* (action 1) or *not promote* (action 0) during the current period; state 0 is uncontrollable: only not promoting is available;

[1] We adopt the following notational conventions to ease the reading: every set is typeset in calligraphic font (e.g., $\mathcal{T}, \mathcal{I}$), the corresponding uppercase character denotes the number of its elements (T, I), and the corresponding lowercase character is used for an element (t, i). Vectors ($\boldsymbol{y}, \boldsymbol{z}$) as well as matrices ($\boldsymbol{P}$) are in boldface.

– The expected one-period space occupation (or work) $W^a_{i,t}$ in state t under action a is as follows. For any state $t \in \mathcal{T}_i$,

$$W^1_{i,t} := W_i, \qquad W^0_{i,t} := 0, \qquad W^0_{i,0} := 0;$$

– The expected one-period revenue $R^a_{i,t}$ in state t under action a is as follows. For any state $t \in \mathcal{T}_i \setminus \{1\}$,

$$\begin{aligned} R^1_{i,t} &:= R_i(1-p_i), \qquad R^1_{i,1} := R_i(1-p_i) + \beta\alpha_i R_i p_i, \\ R^0_{i,t} &:= R_i(1-q_i), \qquad R^0_{i,1} := R_i(1-q_i) + \beta\alpha_i R_i q_i, \qquad R^0_{i,0} := 0; \end{aligned}$$

– The one-period transition probability matrix $\boldsymbol{P}_i^{1|\mathcal{T}_i}$ under promoting is[2]

$$\boldsymbol{P}_i^{1|\mathcal{T}_i} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ \vdots \\ T_i \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & \cdots & T_i-1 & T_i \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1-p_i & p_i & 0 & 0 & 0 \\ \vdots & 0 & \ddots & 0 & 0 \\ 1-p_i & 0 & 0 & p_i & 0 \end{pmatrix} \end{array},$$

where $\boldsymbol{P}^{1|\mathcal{T}_i}_{i,t,s}$ is the probability of moving from state $t \in \mathcal{X}_i$ to state $s \in \mathcal{X}_i$ in one period if the item i is promoted at all states in $\mathcal{T}_i$. The one-period transition probability matrix $\boldsymbol{P}_i^{1|\emptyset}$ under not promoting is obtained analogously.

The dynamics of item i is thus captured by the *state process* $X_i(\cdot)$ and the *action process* $a_i(\cdot)$, which correspond to state $X_i(s) \in \mathcal{X}_i$ and action $a_i(s) \in \mathcal{A}$, respectively, at all time epochs $s \in \mathcal{S}$, with the initial state $X_i(0) = T_i$. Clearly, the state $X_i(s)$ at time epoch s is either $X_i(s) = T_i - s$ (i.e., the number of remaining periods to the deadline) if $s < T_i$ and the item has not yet been sold, or $X_i(s) = 0$ if either the item is perished ($s \geq T_i$) or it has been sold. As a result of deciding action $a_i(s)$ in state $X_i(s)$ at time epoch s, item i consumes the allocated capacity, earns the reward, and evolves its state for the time epoch $s+1$.

2.2 MDP Model of Empty Space

It will be advantageous to consider that the knapsack can always be completely filled with available items. Notice that an empty physical space unit can be seen as an item which is already perished, but always available for promotion. We model it as an MDP with a single state 0 and with static revenue 0. That is, an empty space unit i is defined by $\mathcal{X}_i := \{0\}, W^a_{i,0} := a, R^a_{i,0} := 0, p^a_{i,0,0} := 1$ for all $a \in \mathcal{A}$.

[2] Note that including the row "0" (referring to the uncontrollable state) in the transition probability matrices has no implications as long as the transition probabilities are equal under both actions.

2.3 KPPI Formulation

We next formulate the KPPI. Without loss of generality, let us assume that apart from the perishable items there are at least W items that correspond to the empty space units, so that the capacity can always be fully filled.

Let $\mathit{\Pi}_{\boldsymbol{X},\boldsymbol{a}}$ be the space of randomized and non-anticipative policies depending on the joint state-process $\boldsymbol{X}(\cdot) := (X_i(\cdot))_{i\in\mathcal{I}}$ and deciding the joint action-process $\boldsymbol{a}(\cdot) := (a_i(\cdot))_{i\in\mathcal{I}}$, i.e., $\mathit{\Pi}_{\boldsymbol{X},\boldsymbol{a}}$ is the *joint policy space*. Let $\mathbb{E}_0^{\boldsymbol{\pi}}$ denote the expectation over the state process $\boldsymbol{X}(\cdot)$ and over the action process $\boldsymbol{a}(\cdot)$, conditioned on the initial joint state $\boldsymbol{X}(0) = \boldsymbol{T} := (T_i)_{i\in\mathcal{I}}$ and on policy $\boldsymbol{\pi} \in \mathit{\Pi}_{\boldsymbol{X},\boldsymbol{a}}$.

For any discount factor β, the KPPI problem is to find a joint policy $\boldsymbol{\pi}$ maximizing the β-discounted aggregate revenue starting from the initial time epoch 0 subject to the family of *sample path* knapsack capacity allocation constraints, i.e.,

$$\max_{\boldsymbol{\pi}\in\mathit{\Pi}_{\boldsymbol{X},\boldsymbol{a}}} \mathbb{E}_0^{\boldsymbol{\pi}}\left[\sum_{i\in\mathcal{I}}\sum_{s\in\mathcal{S}} \beta^s R_{i,X_i(s)}^{a_i(s)}\right]$$
$$\text{subject to} \quad \sum_{i\in\mathcal{I}} W_{i,X_i(s)}^{a_i(s)} = W \text{ at each time period } s\in\mathcal{S} \qquad \text{(KPPI)}$$

3 Special Cases

Note that one could equivalently formulate KPPI using dynamic programming. However, (as we can observe in the experimental study in Section 7) the numerical computation of such equations quickly becomes intractable due to the curse of dimensionality. Moreover, the Bellman equation requires the solution of a knapsack subproblem for each possible combination of available items.

In fact, problem (KPPI) (with general one-period work, revenue and transition probability matrix) covers two well-studied problems as special cases: the NP-complete knapsack problem and the PSPACE-hard (even non-stochastic) multi-armed restless bandit problem (Papadimitriou and Tsitsiklis, 1999). So, we have the following theorem.

Theorem 1 *Problem* (KPPI) *(with general one-period work, revenue and transition probability matrix) is* PSPACE*-hard for* $\beta > 0$ *and* NP*-complete for* $\beta = 0$.

Nevertheless, formulation (KPPI) allows the relaxation and decomposition of the problem into tractable parametric subproblems, as seen in Section 4. We will next discuss the two special cases in more detail.

3.1 Knapsack Problem

Under the myopic criterion ($\beta = 0$), the KPPI reduces to a variant of the knapsack problem. In this case, the dynamics can be ignored and one only needs to determine the most valuable knapsack capacity allocation to a collection of competing items with knapsack capacity demands and rewards given depending on whether items are in the knapsack or not.

One could solve at every time epoch s the knapsack problem for $\beta = 0$, giving rise to the following *myopic knapsack rule*. In order to fill in the knapsack myopically optimally at time epoch s, we perform the following steps:

(i) Define $\mathcal{I}^{(s)}$ as the set of all unsold and unperished items (i.e., those that are not in state 0);

(ii) Compute knapsack problem values of including the item in the knapsack for each item $i \in \mathcal{I}^{(s)}$

$$v_i^{\text{myopic}} := R^1_{i,T_i-s} - R^0_{i,T_i-s} = R_i(q_i - p_i); \tag{1}$$

(iii) Solve the following 0-1 knapsack problem

$$\begin{aligned} \max_{\boldsymbol{z}} \quad & \sum_{i \in \mathcal{I}^{(s)}} z_i v_i^{\text{myopic}} \\ \text{subject to} \quad & \sum_{i \in \mathcal{I}^{(s)}} z_i W_i \leq W \qquad (\text{KP}^{\text{myopic}}) \\ & z_i \in \{0,1\} \text{ for all } i \in \mathcal{I}^{(s)} \end{aligned}$$

where $\boldsymbol{z} = (z_i : i \in \mathcal{I}^{(s)})$ is the vector of binary decision variables denoting whether each item i is selected for the promotion knapsack or not;

(iv) Select for the knapsack the items with $z_i = 1$.

The knapsack problem is NP-complete to solve optimally, but it is interesting to note that a simple greedy rule was proposed by Dantzig (1957): *Allocate the capacity to the items with the highest value/demand ratios*. In the particular case when these competing items have equal capacity demands, the Dantzig (1957)'s greedy rule is optimal and reduces to allocating the capacity to the items with highest values.

Nevertheless, the above myopic knapsack rule could be proposed as an approximate solution to the β-discounted problem since there exist extremely efficient exact algorithms for the knapsack problem (see Pisinger, 2005).

3.2 Multi-armed Restless Bandit Problem

Efficient exact algorithms are, however, not available for the *multi-armed restless bandit problem* except for instances with significantly reduced dynamics. The reason is the *curse of dimensionality* of dynamic programming. Before discussing the proposed solutions, let us first define the problem. Suppose that all the capacity demands are of one unit, i.e., $W_i = 1$ for all $i \in \mathcal{I}$. In this case, the knapsack constraint is significantly simpler and the set of policies may have one less dimension by omitting the dependence on values of W_i.

The state-of-the-art solutions proposed in the literature are the so-called *index rules*. These are greedy rules with dynamic nature, prescribing the following: *Allocate the capacity to the competitors with the highest index values*. Index values are assigned to all the states of a competitor by the index function, which is furthermore independent of the others competitors. Thus, index rules aim at decreasing the dimensionality of the problem by computing the index values for each competitor in isolation. The reader has probably realized that these index values play an analogous role as the Dantzig (1957)'s value/demand ratios. We will elaborate more on this issue in Section 6.

Gittins and Jones (1974) proved that if, in addition, the capacity $W = 1$ and the competitors remain frozen if not allocated (i.e., $\boldsymbol{P}^{1|\emptyset}$ is the identity matrix), then there is an index rule which is optimal. This problem is called the *multi-armed bandit problem* (non-restless) for being the generalization of the problem of optimal control of a one-armed bandit (slot) machine. The index which achieves such an optimality is now known as the *Gittins index*, and optimality also holds if (symmetric) competitors are allowed to appear randomly over time (Whittle, 1981). There is a huge amount of literature on this problem and computation of the Gittins index, see, e.g., Varaiya et al (1985); Katehakis and Veinott (1987); Katehakis and Derman (1987); Niño-Mora (2007a).

The restless variant of the problem (where *restless* refers to the possibility that competitors change their state even when not being allocated resource capacity) is significantly more complicated. Still, Whittle (1988) proposed to obtain an index rule after a Lagrangian relaxation and decomposition of the problem, taking as the index value of a state the value of the Lagrangian multiplier at which the optimal action in this state changes. We will call this policy the *Whittle index* rule. Although such a policy may not exist (so-called indexability is required), in case $W = 1$ there are instances in which the Whittle index rule is optimal (Jacko, 2011a). In general (and if exists), it only obeys a form of asymptotic optimality under the time-average criterion (Weber and Weiss, 1990), and is usually reported a close-to-optimal mean performance under the discounted criterion (Niño-Mora, 2007b).

To conclude this section, note that the Whittle index rule could be proposed as a solution to the β-discounted problem by properly adapting the Whittle index computation to non-unitary and non-uniform capacity demands, like in Niño-Mora (2002). Similarly, Glazebrook and Minty (2009) generalized the notion of Gittins index to general resource requirements for general W, losing, however, the optimality properties of the resulting index rule.

4 Relaxations and Decomposition

Whittle (1988) proposed what has become known as the *Whittle relaxation*: replace the infinite set of sample-path capacity constraints by a single constraint requiring to consume the capacity only *in expectation*. In the following we focus on the total discounted criterion ($\beta < 1$), but the undiscounted case ($\beta = 1$) can be treated in an analogous way after reformulating it under the time-average criterion. The Whittle relaxation of (KPPI) is the following:

$$\begin{aligned} &\max_{\boldsymbol{\pi}\in\Pi_{\boldsymbol{X},\boldsymbol{a}}} \mathbb{E}_0^{\boldsymbol{\pi}}\left[\sum_{i\in\mathcal{I}}\sum_{s\in\mathcal{S}}\beta^s R_{i,X_i(s)}^{a_i(s)}\right] \\ &\text{subject to} \quad \mathbb{E}_0^{\boldsymbol{\pi}}\left[\sum_{i\in\mathcal{I}}\sum_{s\in\mathcal{S}}\beta^s W_{i,X_i(s)}^{a_i(s)}\right] = \frac{W}{1-\beta}, \end{aligned} \tag{WR}$$

where we have employed the total discounted criterion on both sides of the capacity constraint. Consideration of the space utilization in expectation reflected in the Whittle relaxation is sufficient for the KPPI to be solved efficiently. Its solution is, however, not feasible for the original problem (KPPI), because it may imply utilization of more or less than the knapsack capacity in some periods. The optimal

solution to (WR) is a dynamic policy for adaptively time-varying knapsack capacity; in the original problem (KPPI) a dynamic policy for fixed-capacity knapsack is sought.

The Whittle relaxation (WR) can be approached by traditional Lagrangian methods. Let ν be a Lagrangian multiplier for the constraint, then we can dualize the constraint, obtaining thus the following Lagrangian relaxation

$$\max_{\boldsymbol{\pi}\in\Pi_{\boldsymbol{X},\boldsymbol{a}}} \mathbb{E}_0^{\boldsymbol{\pi}}\left[\sum_{i\in\mathcal{I}}\sum_{s\in\mathcal{S}}\beta^s R_{i,X_i(s)}^{a_i(s)}\right] - \nu\left(\mathbb{E}_0^{\boldsymbol{\pi}}\left[\sum_{i\in\mathcal{I}}\sum_{s\in\mathcal{S}}\beta^s W_{i,X_i(s)}^{a_i(s)}\right] - \frac{W}{1-\beta}\right)$$

which can be rewritten as

$$\max_{\boldsymbol{\pi}\in\Pi_{\boldsymbol{X},\boldsymbol{a}}} \sum_{i\in\mathcal{I}}\left(\mathbb{E}_0^{\boldsymbol{\pi}}\left[\sum_{s\in\mathcal{S}}\beta^s R_{i,X_i(s)}^{a_i(s)}\right] - \nu\,\mathbb{E}_0^{\boldsymbol{\pi}}\left[\sum_{s\in\mathcal{S}}\beta^s W_{i,X_i(s)}^{a_i(s)}\right]\right) + \nu\frac{W}{1-\beta} \qquad (\mathrm{LR}^{\nu})$$

Parameter ν can be interpreted as the competitive market cost per unit of the promotion space. Then, there is an optimal market cost ν^* which balances expected supply (selling free space) and expected demand (buying necessary space). If this price is known, then (LR^{ν^*}) solves (WR). In any case, the optimal solution to the Whittle relaxation or to Lagrangian relaxations (for any value of ν) yields a tractable bound for the original problem (KPPI).

4.1 Decomposition

We now set out to decompose the optimization problem (LR^{ν}) as it is standard for Lagrangian relaxations, considering ν as a parameter. Notice that any joint policy $\boldsymbol{\pi} \in \Pi_{\boldsymbol{X},\boldsymbol{a}}$ defines a set of single-item policies $\widetilde{\pi}_i$ for all $i \in \mathcal{I}$, where $\widetilde{\pi}_i$ is a randomized and non-anticipative policy depending on the *joint* state-process $\boldsymbol{X}(\cdot)$ and deciding the *item-i* action-process $a_i(\cdot)$. We will write $\widetilde{\pi}_i \in \Pi_{\boldsymbol{X},a_i}$. We will therefore study the item-i subproblem starting from time epoch 0,

$$\max_{\widetilde{\pi}_i\in\Pi_{\boldsymbol{X},a_i}} \mathbb{E}_0^{\widetilde{\pi}_i}\left[\sum_{s\in\mathcal{S}}\beta^s\left(R_{i,X_i(s)}^{a_i(s)} - \nu W_{i,X_i(s)}^{a_i(s)}\right)\right]. \qquad (2)$$

of maximizing the *expected total discounted net revenue* over policies $\widetilde{\pi}_i$. The optimal policy thus optimally resolves the trade-off between the expected total discounted revenues (with salvage values) and the expected total discounted promotion cost.

4.2 Indexability and Index Values

Now we examine the economics of promoting the perishable item. Under so-called *indexability* of the parameterized problem (2), one may identify its optimal control in terms of *index values*. The index captures the marginal rate of promotion and defines an index policy, which furnishes an optimal control of a perishable item by indicating when it is worth promoting. Indexability is defined as follows (this definition was introduced in Jacko (2010), and covers strictly more problems than the definitions introduced in Whittle (1988) and Niño-Mora (2001, 2002)).

Definition 1 (Indexability) We say that the ν-parameterized perishable item i is *indexable*, if there exist unique values $-\infty \leq \nu^*_{i,t} \leq \infty$ for all $t \in \mathcal{T}_i$ such that the following holds for every state $t \in \mathcal{T}_i$:

(i) if $\nu^*_{i,t} \geq \nu$, then it is optimal to promote perishable item in state t, and
(ii) if $\nu^*_{i,t} \leq \nu$, then it is optimal not to promote perishable item in state t.

The function $t \mapsto \nu^*_{i,t}$ is called the *index*, and $\nu^*_{i,t}$'s are called the *index values*.

It is interesting to note that if all the perishable items are indexable, then (LR$^\nu$) is optimally solved by promoting at every time period all the items with current index values larger (or equal) to ν. In the next section we propose to employ the index values in order to define a solution to the original problem (KPPI).

5 Optimal Dynamic Promotion of Perishable Item

In this section we focus on the question of indexability of perishable items. This would not be necessary, since indexability can be tested and index values can be computed numerically (Niño-Mora, 2007b). Nevertheless, in this section we prove indexability analytically under a mild condition and derive index values in a closed form. The analysis further gives additional structural results on optimal dynamic promotion.

The aim of this section is to identify an optimal solution to the problem of promotion of a single perishable item when one must pay for promotion. We interpret ν as a *promotion cost* which must be paid for each space unit occupied in every period in which the item is promoted. The optimal policy thus resolves the trade-off between the expected total discounted revenues (with salvage values) and the expected total discounted promotion cost.

Since we are now considering item i in isolation, in the following we drop the item's subscript i. We will impose a consistency requirement on promotion power, which rules out uninteresting items that should never be promoted. Indeed, the optimal action in all the states for an item with promotion power $q - p \leq 0$ is not promoting (as long as $\nu \geq 0$). We will therefore use the assumption that the promotion power be positive. Moreover, we will need the expected salvage value to be slightly restricted in order to achieve certain monotonicity property of the optimal policy.

Assumption 1 *It holds that*

(i) [Positive Promotion Power] $q - p > 0$, *and*
(ii) [Restricted Expected Salvage Value] $(1 - q) - \alpha(1 - \beta q) \geq 0$.

Note that under $\beta = 1$, the restricted expected salvage value reduces to $\alpha \leq 1$. On the other hand, for $\beta < 1$ the expected salvage value must be bounded away from 1. In any case, however, $\alpha \leq 0$ is valid.

5.1 Indexability and Index Values

Regarding the problems with finite horizon, results with index policies appear very sporadically, because of the complexity of the model, and therefore other methods

(such as dynamic programming) are often used, see, e.g., Burnetas and Katehakis (1998, 2003). Even then, the problem is usually computationally intractable. Nevertheless, there is a tractable instance, the so-called *deteriorating case*, first presented for an infinite-horizon bandit problem by Gittins (1979), which was also successfully applied in a problem with finite-horizon objective (Manor and Kress, 1997). In that setting, the bandits were, however, not restless. This was the case also for the index policies for the finite-horizon multi-armed bandit problem: Niño-Mora (2005) showed that finite-horizon bandits are indexable and provided a tractable algorithm.

We show in the main result of this section, Theorem 2, that indexability holds and index values can be obtained in closed form. Note that this problem could also be approached by standard dynamic programming techniques in order to prove structural properties; they are, however, not useful for obtaining index values.

Theorem 2 (Indexability and Time Monotonicity) *Under Assumption 1, the parameterized perishable item is indexable, and the index value for its state $t \in \mathcal{T}$ is*

$$\nu_t^* = \frac{R}{W}\left\{[(1-p)-\alpha(1-\beta p)] - \frac{[(1-q)-\alpha(1-\beta q)]\,(1-\beta p)}{(1-\beta q)+(\beta q-\beta p)(\beta p)^{t-1}}\right\}. \tag{3}$$

Moreover, the index of an item is nondecreasing as t diminishes (i.e., as the deadline approaches).

The proof of Theorem 2 is presented in Appendix C, after a more detailed description of the work-revenue analysis in Appendix B. Next we list the most appealing properties of the index values (with the proof in Appendix D).

Proposition 1 *Under Assumption 1, for any state $t \in \mathcal{T}$,*

(i) the index value is nonnegative and proportional to R/W;
(ii) an item with lower probability of being sold when not promoted ($(1-q)$'s), ceteris paribus, has higher index value.

The index resolves the trade-off between immediate and postponed promotion. Time monotonicity is a crucial property of the index, saying that the necessity of promotion increases as the deadline approaches. Based on this result, we can look for an *optimal promotion starting time* τ^*,

$$\tau^* := \max\{\tau \in \mathcal{T} :\ \nu_t^* > \nu \text{ for all } t \in \mathcal{T} \text{ such that } t \le \tau\}. \tag{4}$$

In other words, if τ^* is finite, then τ^* is the threshold time period, from which the index value is larger than the promotion cost ν, i.e. from which it is optimal to start to promote the item. If τ^* is not finite, i.e., the index value of state 1, ν_1^*, is lower than or equal to ν, then it is never optimal to promote the item. This intuition is formalized in the following proposition.

Proposition 2 *Under Assumption 1, the optimal promotion starting time τ^* is finite if and only if*

$$\frac{R}{W}\frac{(1-\beta\alpha)}{\nu}(q-p) > 1.$$

Further,

(i) if τ^* is finite, then promoting is optimal in all time periods from τ^* to 1 and not promoting is optimal in the remaining time periods;
(ii) if τ^* is not finite, then not promoting is optimal in all time periods.

The above result assures that promotion is to be done in a natural way: the item is selected for promotion only once and remains promoted as long as it remains unsold and not perished.

We further give the index values obtained in a straightforward manner from the discounted index (3) for the most important limit regimes.

Proposition 3

(i) [Undiscounted Index] Under positive promotion power assumption, in the case $\beta = 1$, the index value for state $t \in \mathcal{T}$ is

$$\nu_t^* = \frac{R}{W}(1-\alpha)(1-p)\left[1 - \frac{(1-q)}{(1-q)+(q-p)p^{t-1}}\right].$$

(ii) [Myopic Index] Under positive promotion power assumption, in the case $\beta = 0$, the index value for state $t \in \mathcal{T}$ is

$$\nu_t^* = \frac{R}{W}(q-p).$$

(iii) [Index for Nonperishable Item] Under positive promotion power assumption, the index value for a nonperishable item is

$$\nu_\infty^* = \frac{R}{W}\frac{(1-\beta)(q-p)}{1-\beta q}.$$

(iv) Under positive promotion power assumption, the index value for a perishable item with zero revenue (profit margin) and with expected salvage value $-c < 0$ is

$$\nu_t^* = \frac{c}{W}(1-\beta p)\left[1 - \frac{(1-\beta q)}{(1-\beta q)+(\beta q-\beta p)(\beta p)^{t-1}}\right].$$

6 Index-Knapsack Heuristic

If all the perishable items are indexable, we define a novel solution to (KPPI), which we call the *index-knapsack* (IK) heuristic. In order to fill in the knapsack at time epoch s, IK prescribes to perform the following steps:

(i) Define $\mathcal{I}^{(s)}$ as the set of all unsold and unperished items (i.e., those that are not in state 0);
(ii) Compute index value ν^*_{i,T_i-s} for each item $i \in \mathcal{I}^{(s)}$;
(iii) Compute knapsack problem values for each item $i \in \mathcal{I}^{(s)}$

$$v_i := W_i \nu^*_{i,T_i-s}; \tag{5}$$

(iv) Solve the following 0-1 knapsack problem

$$\begin{aligned} \max_{\boldsymbol{z}} \quad & \sum_{i \in \mathcal{I}^{(s)}} z_i v_i \\ \text{subject to} \quad & \sum_{i \in \mathcal{I}^{(s)}} z_i W_i \leq W \\ & z_i \in \{0,1\} \text{ for all } i \in \mathcal{I}^{(s)} \end{aligned} \tag{KP}$$

where $\boldsymbol{z} = (z_i : i \in \mathcal{I}^{(s)})$ is the vector of binary decision variables denoting whether each item i is selected for the promotion knapsack or not;

(v) Select for the knapsack the items with $z_i = 1$.

There are strong arguments for proposing this heuristic. Note that it closely resembles the myopic knapsack rule, with a difference only in steps (ii) and (iii). Recall that the index value is the value of the Lagrangian multiplier interpreted as the promotion cost per unit of capacity. Thus, the index value measures a (shadow) price per unit of demanded capacity for promoting the item. The knapsack problem value, however, must measure the price for promoting the item itself, therefore we propose as a proxy to multiply the index value by the item's volume.

The index value can thus be seen again as the Dantzig (1957)'s value/demand ratio. Notice that the Dantzig (1957)'s greedy rule used for solving the knapsack problem in (iv) of the index-knapsack heuristic reduces the index-knapsack heuristic to the Whittle index rule. It is well known that the Dantzig (1957)'s greedy rule yields an optimal solution to the knapsack problem if all the capacity demands are uniform; however, in the general case it is suboptimal. Our experimental study presented in Section 7 suggests that index rule analogously reveals an inferior performance with respect to the proposed index-knapsack heuristic.

The experimental study further reveals a nearly-optimal performance of the index-knapsack heuristic. Interestingly, the index-knapsack heuristic recovers optimal policies in some well-studied cases.

Theorem 3 (Optimality of Index-Knapsack Heuristic) *If $\beta = 0$ or if $T_i = 1$ for all $i \in \mathcal{I}$, then the index-knapsack heuristic is optimal. If $0 < \beta \leq 1$ and $W_i = 1$ for all items i, then the index-knapsack heuristic is optimal in all the problem instances in which the Gittins index rule or the Whittle index rule is optimal.*

Proof In the case $\beta = 0$, implementation of the myopic index values in the index-knapsack heuristic leads to recovering the myopic knapsack rule, which is optimal in this case. Similarly if all the deadlines $T_i = 1$, with a slightly different index values.

In the case $0 < \beta \leq 1$ and $W_i = 1$ for all items i, the optimal solution to (KP) is given by taking W items with highest value $v_i = \nu_i$, which is essentially the Gittins/Whittle index rule. □

7 Experimental Study

In this section we present results of systematic computational experiments, in which we evaluate the performance of the index-knapsack (IK) heuristic and the index

rule (IR). We further compare their performance to the *Earlier-Deadline-First* policy, a naïve benchmark policy.

Earlier-Deadline-First (EDF) heuristic: In order to fill in the knapsack at time epoch s, define $\mathcal{I}^{(s)}$ as the set of all unsold and unperished items (i.e., those that are not in state 0) and select items in a greedy manner until none of the remaining items fits after sorting the items so that product i_1 is preferred to product i_2, if:

(i) $T_{i_1} < T_{i_2}$,
(ii) $T_{i_1} = T_{i_2}$ and $R_{i_1}(1-\alpha_{i_1}) > R_{i_1}(1-\alpha_{i_1})$,
(iii) $T_{i_1} = T_{i_2}$ and $R_{i_1}(1-\alpha_{i_1}) = R_{i_1}(1-\alpha_{i_1})$ and $W_{i_1} < W_{i_2}$.

The following is the worst-case (i.e., revenue minimizing) solution of the knapsack subproblem whenever all the reward coefficients are nonnegative, which is valid according to Proposition 1(i).

Revenue Minimizing (MIN) solution: Leave the knapsack empty at every time epoch s.

For each fixed pair (I,T), denoting the number of products and the problem time horizon, respectively, such that $I \in \{2,3,4,\dots,8\}$ and $T \in \{2,4,6,\dots,20\}$, we have randomly generated 10^4 instances. Thus, we have tested 70 scenarios. We set $\alpha_i = 0.5$ for each product i and we assure that $T_1 := T$. We assume Poisson arrivals of customers for each item i, denoting by λ_i^0 and λ_i^1 the mean arrival rate for a non-promoted and for a promoted item, respectively. We restrict these values to $2/3 < \lambda_i^a T_i \leq 2$ for both $a \in \{0,1\}$, which assures that each item has a non-extreme probability of being sold before the deadline, since $\lambda_i^a T_i$ is the expected number of customer arrivals during the product's lifetime. Of course, the probabilities of not selling item i are $q_i := \exp\{-\lambda_i^0\}$ and $p_i := \exp\{-\lambda_i^1\}$, and we assure that $q_i > p_i$. Thus, we generate the following uniformly distributed parameters:

$$W_i \in [10,50]; \qquad R_i \in [10,50]; \qquad T_i \in [2,T]; \qquad \lambda_i^0, \lambda_i^1 \in \left(\frac{2}{3T_i}, \frac{2}{T_i}\right].$$

Finally, a uniformly distributed knapsack volume is generated: $W \in [\max\{W_i\}, 30\% \cdot \sum_i W_i)$.

We focus on the discount factor $\beta = 1$, as this is the case most likely to be implemented in practice. Moreover, our experiments (not reported here) suggest that this is also the hardest case, since the performance of IK and IR heuristics improves as the discount factor diminishes and the IK heuristic approaches optimality as $\beta \to 0$.

7.1 Performance Evaluation Measures

We obtain the maximizing policy solving the KPPI optimally by backwards recursion, which also yields the optimal objective value D^{MAX}. The objective values of the other policies are also obtained by backwards recursion, employing the respective policy at each step, denoted D^π for a policy π. We next introduce performance evaluation measures we use to report the experimental results.

The *relative suboptimality gap* of policy π, often used in the literature, is defined as

$$\text{rsg}(\pi) := \frac{D^{\text{MAX}} - D^\pi}{D^{\text{MAX}}}. \tag{6}$$

As long as $D^{\text{MAX}} > 0$ and $D^{\pi} \geq 0$, we have $0 \leq \text{rsg}(\pi) \leq 1$, where $\text{rsg}(\pi) = 0$ is obtained by the maximizing policy. However, $\text{rsg}(\pi) = 1$ is not necessarily achieved by any policy in a particular problem instance. Furthermore, if $\alpha_i \leq 0$ for some of the items, then we may have $D^{\text{MAX}} \leq 0$ and therefore may also be $\text{rsg}(\pi) < 0$. Thus, this measure may overestimate the quality of π by reporting small or negative values even for the worst policies.

This motivates us to introduce a new measure: the *adjusted relative suboptimality gap* of policy π, defined as

$$\text{arsg}(\pi) = \frac{D^{\text{MAX}} - D^{\pi}}{D^{\text{MAX}} - D^{\text{MIN}}}. \tag{7}$$

With this measure we always (as long as $D^{\text{MAX}} \neq D^{\text{MIN}}$) have $0 \leq \text{arsg}(\pi) \leq 1$, and both limiting values are achieved by admissible policies.

We further introduce a measure to be used to compare the mean performance of a policy π with respect to IK heuristic, as follows:

$$\text{ratio}(\pi) = \frac{\text{mean}(\text{rsg}(\pi))}{\text{mean}(\text{rsg}(\text{IK}))}. \tag{8}$$

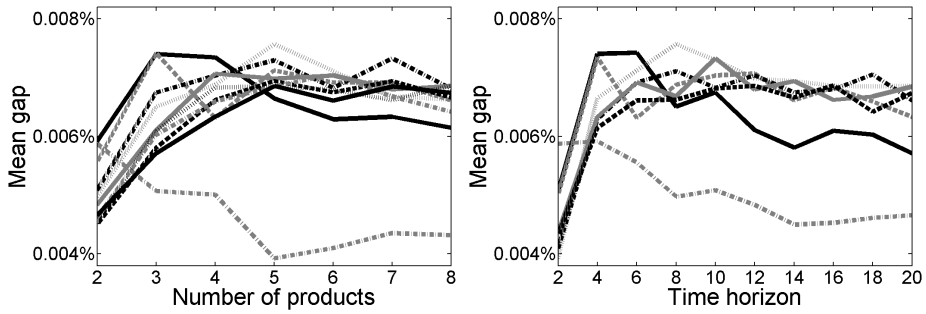
This ratio captures the extent to which the mean absolute gap (i.e., the revenue loss) created by IK heuristic may be expected to be magnified if policy π is implemented instead. Thus, we have $\text{ratio}(\pi) > 1$ if and only if policy π is on average worse than IK heuristic. An analogous ratio is used with the arsg measure.

7.2 Experimental Results

For every figure described below, subfigures (a) and (b) are projections of a 3D image. We prefer to exhibit these projections instead of corresponding 3D figures to provide a better visibility of the effects of varying a single parameter. In particular, in subfigure (a) we plot curves for several values of problem horizon T in order to observe the effect of an increase in the number of products I, while in subfigure (b) we plot curves for several values of number of products I in order to observe the effect of an increase in the problem horizon T. Note that each curve in subfigure (a) corresponds to one of the points on the horizonal axis of subfigure (b), and vice versa.

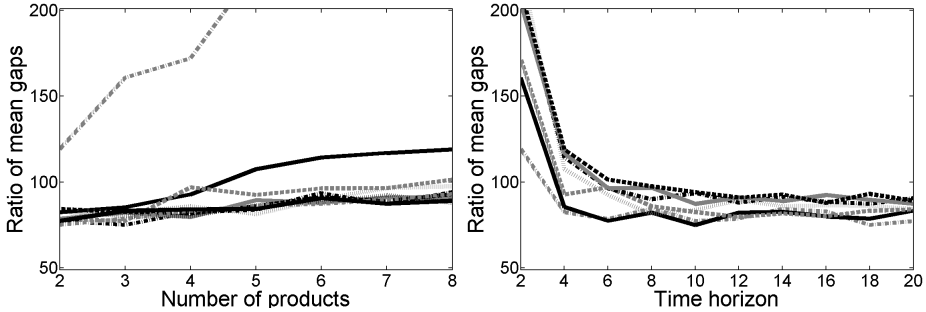
Figure 1 exhibits the projections of the mean $\text{rsg}(\text{IK})$ as function of the number of products I and the time horizon T. The figure shows an excellent mean performance of IK heuristic well below 0.01%, and further suggests that such a performance can be expected even for higher values of I and T. These strong results are further confirmed in Figure 4 considering the arsg measure, being in all cases below 0.14%.

The ratio of the benchmark EDF heuristic is presented in Figure 2 and Figure 5. The benchmark policy's mean gap is in all cases more than 50-times larger than that of IK heuristic, and the ratio grows with the number of items I. This ratio reveals that the mean performance of EDF heuristic is of the order of 0.5% in terms of rsg, which may look interesting, but it is of the order of 10% in terms of arsg, which makes EDF considerably weak.



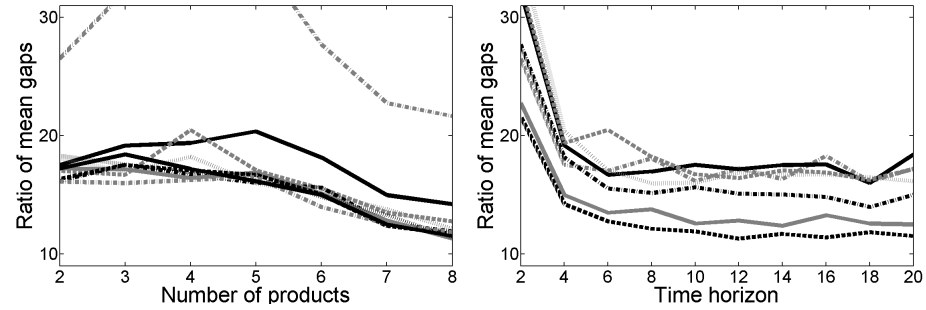
(a) curves for values $T = 2, 4, 6, \ldots, 20$ (b) curves for values $I = 2, 3, \ldots, 8$

Fig. 1 Mean relative suboptimality gap of IK heuristic.



(a) curves for values $T = 2, 4, 6, \ldots, 20$ (b) curves for values $I = 2, 3, \ldots, 8$

Fig. 2 Ratio of mean relative suboptimality gaps of EDF over IK heuristic.

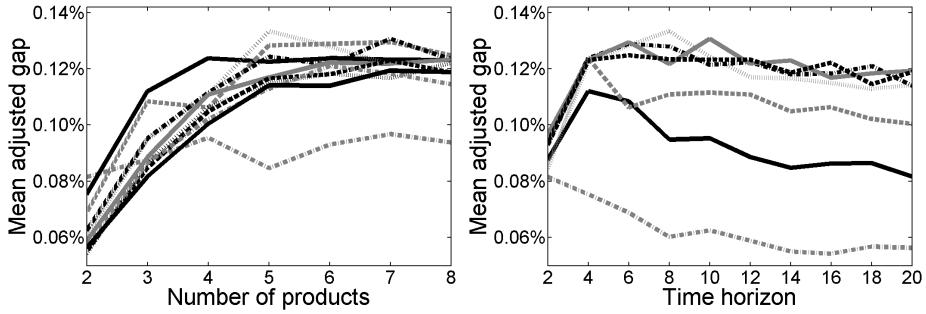


(a) curves for values $T = 2, 4, 6, \ldots, 20$ (b) curves for values $I = 2, 3, \ldots, 8$

Fig. 3 Ratio of mean relative suboptimality gaps of IR over IK heuristic.

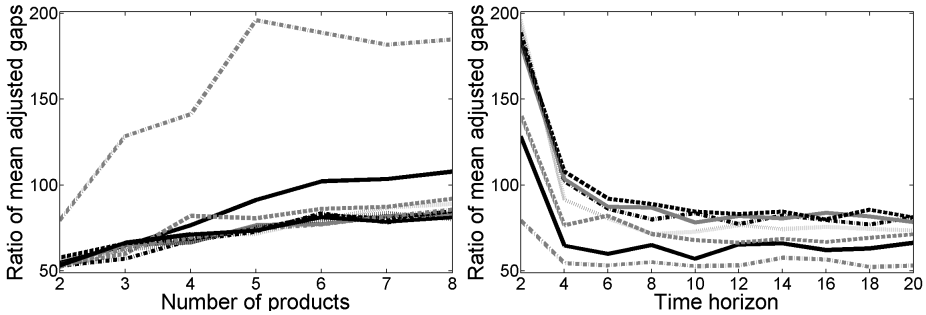
Further, in Figure 3 and Figure 6 we evaluate IR heuristic, whose mean performance is in all cases more than 10-times worse than that of IK heuristic, though improving with higher I once this passes the value 4. Therefore, the mean performance of IR heuristic is of the order of 0.1% in terms of rsg, and of the order of 1% in terms of arsg, which are still very good values of suboptimality.

Finally, we remark that the worst-case performance (out of 10^4 instances) achieved by the maximum rsg (arsg) values of IK heuristic are relatively small, ranging between 0.3% and 15% (4% and 14%) in the 70 scenarios considered. The maximum rsg (arsg) values of IR heuristic range between 1% and 8% (22% and 72%), and those



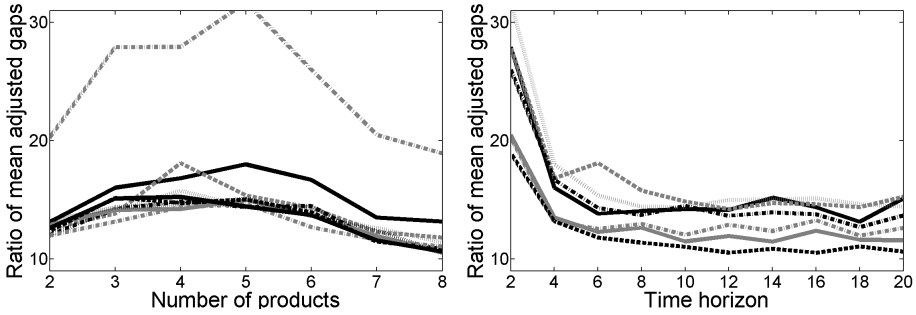
(a) curves for values $T = 2, 4, 6, \ldots, 20$ (b) curves for values $I = 2, 3, \ldots, 8$

Fig. 4 Mean adjusted relative suboptimality gap of IK heuristic.



(a) curves for values $T = 2, 4, 6, \ldots, 20$ (b) curves for values $I = 2, 3, \ldots, 8$

Fig. 5 Ratio of mean adjusted relative suboptimality gaps of EDF over IK heuristic.



(a) curves for values $T = 2, 4, 6, \ldots, 20$ (b) curves for values $I = 2, 3, \ldots, 8$

Fig. 6 Ratio of mean adjusted relative suboptimality gaps of IR over IK heuristic.

of EDF heuristic range between 3% and 8% (51% and 100%). That is, their worst-case performance is good in absolute terms, but is especially bad in the problems where promotion has small effect on the total revenue (i.e., when MAX performs close to MIN).

8 Conclusion

We have developed a dynamic and stochastic model for dynamic promotion of perishable items and proposed a tractable index-knapsack heuristic that has a natural

economic interpretation and suggests itself to be easily implementable in practice. These advantages come at the cost of possible suboptimality of such a dynamic solution, which was, however, shown to be negligible and significantly smaller than the revenue losses from implementing a naïve marketing solution of giving priority to items with earlier deadlines. Moreover, being the considered problem an extension of the notoriously difficult (PSPACE-hard) restless bandit problem, the nearly-optimal performance of IK heuristic is an excellent result. The model is extensible to a variety of ad-hoc requirements that managers or certain circumstances may impose.

We believe that we have developed a non-trivial extension of the multi-armed bandit problem with a novel (and perhaps surprising) solution. In our model, there are four additional complications: the bandits are *restless*, because the items can get sold regardless of being in the knapsack or not, the time horizon is *finite* due to perishability, and we are to select *more than one* item for the knapsack, which is allowed to be filled partially, due to the *heterogeneity* of the items' capacity requirements.

Application of our results to several revenue management problems is presented in Appendix A. An important challenge is to obtain index values for an extension of KPPI taking into account price and demand changes over time, inventories of products, new stock arrivals, and cross-dependent demands. The analysis of such more general problems, however, may require the notion of indexability to be generalized in order to tackle them, especially when more than two actions are allowed for each product.

Nevertheless, this paper offers a comprehensive and generalizable modeling framework together with the nearly-optimal index-knapsack heuristic that may be relevant in applications outside the revenue management area. The items considered in resource capacity allocation problems in stochastic and dynamic environment are often perishable, either naturally or due to contract restrictions such as the Quality-of-Service clause. In addition, our results cover also nonperishable items, which can be included in the portfolio together with the perishable ones. An interesting extension of the model would be to consider random item lifetimes.

Providing provable bounds or establishing asymptotic optimality of the proposed index-knapsack heuristic remains an important open problem.

References

Anselmi J, Verloop IM (2011) Energy-aware capacity scaling in virtualized environments with performance guarantees. Performance Evaluation in press, DOI 10.1016/j.peva.2011.07.004

Burnetas AN, Katehakis MN (1998) Dynamic allocation policies for the finite horizon one armed bandit problem. Stochastic Analysis and Applications 16(5):811–824

Burnetas AN, Katehakis MN (2003) Asymptotic Bayes analysis for the finite horizon one armed bandit problem. Probability in the Engineering and Informational Sciences 17(1):53–82

Buyukkoc C, Varaiya P, Walrand J (1985) The $c\mu$ rule revisited. Advances in Applied Probability 17(1):237–238

Capgemini, Intel, Cisco, Microsoft (2005) Retailers transform processes with ERS. In: Hann M (ed) RetailSpeak, www.retailspeak.com

Caro F, Gallien J (2007) Dynamic assortment with demand learning for seasonal consumer goods. Management Science 53(2):276–292
Chick SE, Gans N (2009) Economic analysis of simulation selection problems. Management Science 55(3):421–437
Dance C, Gaivoronski AA (2012) Stochastic optimization for real time service capacity allocation under random service demand. Annals of Operations Research 193:221–253, DOI 10.1007/s10479-011-0842-2
Dantzig GB (1957) Discrete-variable extremum problems. Operations Research 5(2):266–277
Elmaghraby W, Keskinocak P (2003) Dynamic pricing in the presence of inventory considerations: Research overview, current practices, and future directions. Management Science 49(10):1287–1309
Gesbert D, Kountouris M, Health RWJ, Chae CB, Sälzer T (2007) Shifting the MIMO paradigm. IEEE Signal Processing Magazine 36:36–46
Gittins JC (1979) Bandit processes and dynamic allocation indices. Journal of the Royal Statistical Society, Series B 41(2):148–177
Gittins JC, Jones DM (1974) A dynamic allocation index for the sequential design of experiments. In: Gani J (ed) Progress in Statistics, North-Holland, Amsterdam, pp 241–266
Glazebrook K, Minty R (2009) A generalized Gittins index for a class of multiarmed bandits with general resource requirements. Mathematics of Operations Research 34(1):26–44
Glazebrook KD, Mitchell HM, Ansell PS (2005) Index policies for the maintenance of a collection of machines by a set of repairmen. European Journal of Operational Research 165:267–284
Glazebrook KD, Hodge DJ, Kirkbride C (2011) General notions of indexability for queueing control and asset management. Annals of Applied Probability 21:876–907
Jacko P (2010) Restless bandits approach to the job scheduling problem and its extensions. In: Piunovskiy AB (ed) Modern Trends in Controlled Stochastic Processes: Theory and Applications, Luniver Press, United Kingdom, pp 248–267, invited book chapter
Jacko P (2011a) Optimal index rules for single resource allocation to stochastic dynamic competitors. In: Proceedings of ValueTools, ACM Digital Library, invited paper
Jacko P (2011b) Value of information in optimal flow-level scheduling of users with Markovian time-varying channels. Performance Evaluation 68(11):1022–1036, special Issue: Performance 2011. Acceptance rate: 20%
Katehakis MN, Derman C (1987) Computing optimal sequential allocation rules in clinical trials. In: Ryzin JV (ed) Adaptive Statistical Procedures and Related Topics, vol 8, I.M.S. Lecture Notes-Monograph Series, pp 29–39
Katehakis MN, Veinott AF (1987) The multi-armed bandit problem: Decomposition and computation. Mathematics of Operations Research 12(2):262–268
Loch CH, Kavadias S (2002) Dynamic portfolio selection of NPD programs using marginal returns. Management Science 48(10):1227–1241
Manor G, Kress M (1997) Optimality of the greedy shooting strategy in the presence of incomplete damage information. Naval Research Logistics 44:613–622
Meuleau N, Hauskrecht M, Kim KE, Peshkin L, Kaelbling LP, Dean T, Boutilier C (1998) Solving very large weakly coupled Markov decision processes. In: Pro-

ceedings of the Fifteenth National Conference on Artificial Intelligence, pp 165–172

Niño-Mora J (2001) Restless bandits, partial conservation laws and indexability. Advances in Applied Probability 33(1):76–98

Niño-Mora J (2002) Dynamic allocation indices for restless projects and queueing admission control: A polyhedral approach. Mathematical Programming, Series A 93(3):361–413

Niño-Mora J (2005) Marginal productivity index policies for the finite-horizon multiarmed bandit problem. In: Proceedings of the 44th IEEE Conference on Decision and Control and European Control Conference ECC 2005 (CDC-ECC '05), pp 1718–1722

Niño-Mora J (2007a) A $(2/3)n^3$ fast-pivoting algorithm for the Gittins index and optimal stopping of a Markov chain. INFORMS Journal on Computing 19(4):596–606

Niño-Mora J (2007b) Dynamic priority allocation via restless bandit marginal productivity indices. TOP 15(2):161–198

Papadimitriou CH, Tsitsiklis JN (1999) The complexity of optimal queueing network. Mathematics of Operations Research 24(2):293–305

Pisinger D (2005) Where are the hard knapsack problems? Computers & Operations Research 32:2271–2284

Qu S, Gittins JC (2011) A forwards induction approach to candidate drug selection. Advances in Applied Probability 43(3):649–665

Smith WE (1956) Various optimizers for single-stage production. Naval Research Logistics Quarterly 3(1-2):59–66

Speitkamp B, Bichler M (2010) A mathematical programming approach for server consolidation problems in virtualized data centers. IEEE Transactions on Services Computing 3(4):266–278

Varaiya P, Walrand J, Buyukkoc C (1985) Extensions of the multiarmed bandit problem: The discounted case. IEEE Transactions on Automatic Control AC-30(5):426–439

Weber R, Weiss G (1990) On an index policy for restless bandits. Journal of Applied Probability 27(3):637–648

Whittle P (1981) Arm-acquiring bandits. Annals of Probability 9(2):284–292

Whittle P (1988) Restless bandits: Activity allocation in a changing world. A Celebration of Applied Probability, J Gani (Ed), Journal of Applied Probability 25A:287–298

Yang R, Bhulai S, van der Mei R (2011) Optimal resource allocation for multiqueue systems with a shared server pool. Queueing Systems 68:133–163

A Related Revenue Management Problems

In this section we show in a number of related revenue management problems that the KPPI framework and the results of this paper are of non-trivially wider applicability.

A.1 Optimal Policy to Dynamic Promotion Problem with Adaptive Knapsack

Consider a retailer for which it is feasible to adjust the promotion space (knapsack) dynamically, e.g. by reserving a necessary number of promotion shelves (or by hiring a necessary number of employees), where given price ν must be paid for each reserved space unit (or for each hired employee). In other words, we assume existence of a space market, where it is permitted to "buy" from other periods some amount of space if necessary or to "sell" to other periods some amount of space if it is not used. Suppose the retailer has allocated a money budget for such a purpose, whose present value is $\widetilde{W}$. We assume that such a present value was calculated using a discount factor β arising in the *perfectly competitive market* (i.e., money can be borrowed or lent for the same inter-period interest rate equal $(1-\beta)/\beta$).

Notice that this problem is in fact formulated by (LR$^\nu$). Indeed, the term $\nu W/(1-\beta)$ in (LR$^\nu$) can be understood as the present value of a regular *money budget* (νW per period) allocated for the knapsack space expected to be spent over an infinite horizon. That is, we only need to set $\widetilde{W} = \nu W/(1-\beta)$. Due to the mutual independence of the products' demand, the single-item optimal policies from Section 5 can be coupled together and we obtain an optimal policy to this multi-product problem.

Proposition 4 *Under Assumption 1 satisfied by each product, an optimal policy to the dynamic promotion problem with adaptive knapsack is to buy in every period the promotion space units necessary for promoting all the products whose current index value is greater than ν.*

A.2 Dynamic Product Assortment Problem

Consider the dynamic product assortment problem, in which a retailer wants to choose a collection of products to be displayed for purchase out of all the products available in the retailers' warehouse. The framework of this paper covers a simple variant of the product assortment problem, in which there is a single unit of each product available from producers. Notice that now the knapsack capacity is given by the total selling space available in the store, and the possible actions for each product is whether to include it in the assortment (action $a = 1$) or not ($a = 0$). Of course, the products not included in the assortment cannot be sold, i.e., $q_i := 1$ for all products i.

Expression (2) then clearly applies to this problem, and moreover, it simplifies interestingly in the undiscounted case.

Proposition 5 *In the dynamic product assortment problem the index value for state $t \in \mathcal{T}_i$ is*

(i) if $\beta < 1$, $p < 1$, and $\alpha_i \leq 0$, then

$$\nu^*_{i,t} = \frac{R_i}{W_i}\left[1 - \frac{\alpha_i\beta(1-\beta p_i)(\beta p_i)^{t-1}}{(1-\beta)+\beta(1-p_i)(\beta p_i)^{t-1}}\right](1-p_i),$$

(ii) if $\beta = 1$, $p < 1$, and $\alpha_i \leq 1$, then

$$\nu^*_{i,t} = \frac{R_i}{W_i}(1-\alpha_i)(1-p_i).$$

Recall that only products included in the assortment can be sold, and this happens with probability $1-p_i$ in every period. Interpreting this probability as a service rate $\mu_i := 1-p_i$, the undiscounted index value reduces to $c\mu$, well-known in queueing theory (see, e.g., Smith, 1956; Buyukkoc et al, 1985), where the cost $c_i := R_i(1-\alpha_i)/W_i$ is the revenue loss per unit of space occupied if item is not sold during its lifetime. We emphasize that such an index is constant over time, i.e., it does not depend on the product lifetime.

The $c\mu$-rule prescribes that products should be ordered (highest first) accordingly to the product of their profitability c and per-period attractiveness μ, and included in the assortment following such an ordering. The experimental results from the previous section suggest that IK heuristic may further give an improved solution than the straightforward $c\mu$-index ordering in the case that space requirements W_i are not equal across all the products.

A.3 Dynamic Product Pricing Problem

Consider a single product and suppose that we are given an additional parameter called *discount* (price markdown) $D_i \geq 0$, so that the revenue (profit margin) is $R_i - D_i$ instead of R_i if item i is promoted. Thus, $1 - q_i$ can be interpreted as the probability of selling the item priced at R_i, and $1 - p_i$ can be interpreted as the probability of selling the item priced at $R_i - D_i$. Let a real-valued $\widetilde{\nu}_i$ be the per-period cost of maintaining (or informing about) the price markdown of this product; thus $\widetilde{\nu}_i = 0$ may be reasonable in many practical cases. We are then addressing a simple variant of the classic *dynamic product pricing problem*.

In particular, in the dynamic product pricing problem we have the following expected one-period revenues for $t \in \mathcal{T}_i \setminus \{1\}$:

$$\widetilde{R}^1_{i,t} := (R_i - D_i)(1 - p_i), \qquad \widetilde{R}^1_{i,1} := (R_i - D_i)(1 - p_i) + \beta\alpha_i R_i p_i,$$
$$R^0_{i,t} := R_i(1 - q_i), \qquad R^0_{i,1} := R_i(1 - q_i) + \beta\alpha_i R_i q_i, \qquad R^0_{i,0} := 0$$

Denote by $\widetilde{D}_i := D_i(1-p_i)$. In order to recover the expected one-period revenues under action 1 from the KPPI framework of Section 2, we must define the expected one-period revenue for all $t \in \mathcal{T}_i$ as $R^1_{i,t} := \widetilde{R}^1_{i,t} + \widetilde{D}_i$. Let us further define the promotion cost $\nu := \left(\widetilde{\nu}_i + \widetilde{D}_i\right)/W_i$. Then, the expected one-period net revenue under action 1 is $R^1_{i,t} - \nu W^1_{i,t} = \widetilde{R}^1_{i,t} + \widetilde{D}_i - \left(\widetilde{\nu}_i + \widetilde{D}_i\right) = \widetilde{R}^1_{i,t} - \widetilde{\nu}_i$ as desired in the dynamic product pricing model. Then by the definition of indexability we have the following result.

Proposition 6 *Suppose that the problem defined above satisfies Assumption 1 and let $\nu^*_{i,t}$ be the index values given by Theorem 2. Then in the dynamic product pricing problem the following holds for state $t \in \mathcal{T}_i$:*
(i) if $\nu^*_{i,t} W_i - \widetilde{D}_i \geq \widetilde{\nu}_i$, then it is optimal to offer the item with price markdown in state t, and
(ii) if $\nu^*_{i,t} W_i - \widetilde{D}_i \leq \widetilde{\nu}_i$, then it is optimal to offer the item without price markdown in state t.

A.4 Loyalty Card Problem

Consider now the multi-product problem like the one addressed in Capgemini et al (cf. 2005), in which retailer can use customer's loyalty card to inform and influence her by personalized messages, say at the moment of arrival to the store. Suppose that the retailer wants to offer the customer a number of products with personalized price markdowns so that the total offered markdown in not larger than a budget D associated with the customer (such a budget could be a function of the customer's historical expenditures, collected "points", or any other relevant measure). Of course, now the probabilities p_i and q_i must be the probabilities of buying product i by the customer in hand (say, estimated using the customer's historical purchasing decisions).

Suppose that every product satisfies Assumption 1. Using the results of the previous subsection, suppose that $\widetilde{\nu}_i = 0$ and define the knapsack-problem rewards $\widetilde{v}_i := \nu^*_{i,t} W_i - \widetilde{D}_i$. Then, analogously to the arguments given in Section 6, we propose to use the solution to the following 0-1 knapsack subproblem in step (iv) of the IK heuristic for the loyalty card problem:

$$\max_{\boldsymbol{z}} \sum_{i \in \mathcal{I}} z_i \widetilde{v}_i$$
$$\text{subject to} \quad \sum_{i \in \mathcal{I}} z_i D_i \leq D$$
$$z_i \in \{0, 1\} \text{ for all } i \in \mathcal{I}$$

where $\boldsymbol{z} = (z_i : i \in \mathcal{I})$ is the vector of binary decision variables denoting whether each item i is offered with price markdown or not.

B Preliminaries of Work-Revenue Analysis

In order to prove Theorem 2, we describe the key points from the restless bandit framework in more detail. For a survey on this methodology we refer to Niño-Mora (2007b). We can restrict our attention to stationary deterministic policies, since it is well-known from the MDP theory that there exists an optimal policy which is stationary, deterministic, and independent of the initial state. Notice that any set $\mathcal{S} \subseteq \mathcal{T}$ can represent a stationary policy, by employing action 1 in all the states belonging to $\mathcal{S}$ and employing action 0 in all the states belonging to $\mathcal{T} \setminus \mathcal{S}$. We will call such a policy an $\mathcal{S}$-*active policy*, and $\mathcal{S}$ an *active set*.

Let $\mathcal{S} \subseteq \mathcal{T}$ be an active set. We can reformulate (1) as

$$\mathbb{R}_t^{\mathcal{S}} - \nu \mathbb{W}_t^{\mathcal{S}} := \mathbb{E}_t^{\mathcal{S}} \left[\sum_{s=0}^{t-1} \beta^s \boldsymbol{P}_{t,t-s}^{s|\mathcal{S}} R_{t-s}^{I_{\mathcal{S}}(t-s)} \right] - \nu \, \mathbb{E}_t^{\mathcal{S}} \left[\sum_{s=0}^{t-1} \beta^s \boldsymbol{P}_{t,t-s}^{s|\mathcal{S}} W_{t-s}^{I_{\mathcal{S}}(t-s)} \right], \tag{9}$$

where $\boldsymbol{P}_{i,j}^{j-i|\mathcal{S}}$ is the probability of moving from state $i \in \mathcal{X}$ to state $j \in \mathcal{X}$ in exactly $j - i$ periods under policy $\mathcal{S}$ and $I_{\mathcal{S}}(s)$ is the indicator function $I_{\mathcal{S}}(s) = \begin{cases} 1, & \text{if } s \in \mathcal{S}, \\ 0, & \text{if } s \notin \mathcal{S}. \end{cases}$ So, $\mathbb{R}_t^{\mathcal{S}}$ is the *expected total discounted revenue* under policy $\mathcal{S}$ if starting from state t, and we will write it in a more convenient way as

$$\mathbb{R}_t^{\mathcal{S}} = \mathbb{E}_t^{\mathcal{S}} \left[\sum_{s=1}^{t} \beta^{t-s} \boldsymbol{P}_{t,s}^{t-s|\mathcal{S}} R_s^{I_{\mathcal{S}}(s)} \right]. \tag{10}$$

Similarly, $\mathbb{W}_t^{\mathcal{S}}$ is the *expected total discounted promotion work* under policy $\mathcal{S}$ if starting from state t, and we will write it in a more convenient way as

$$\mathbb{W}_t^{\mathcal{S}} = \mathbb{E}_t^{\mathcal{S}} \left[\sum_{s=1}^{t} \beta^{t-s} \boldsymbol{P}_{t,s}^{t-s|\mathcal{S}} W_s^{I_{\mathcal{S}}(s)} \right]. \tag{11}$$

Let, further, $\langle a, \mathcal{S} \rangle$ be the policy which takes action $a \in \mathcal{A}$ in the current time period and adopts an $\mathcal{S}$-active policy thereafter. For any state $t \in \mathcal{T}$ and an $\mathcal{S}$-active policy, the $(t, \mathcal{S})$-*marginal revenue* is defined as

$$r_t^{\mathcal{S}} := \mathbb{R}_t^{\langle 1, \mathcal{S} \rangle} - \mathbb{R}_t^{\langle 0, \mathcal{S} \rangle}, \tag{12}$$

and the $(t, \mathcal{S})$-*marginal promotion work* as

$$w_t^{\mathcal{S}} := \mathbb{W}_t^{\langle 1, \mathcal{S} \rangle} - \mathbb{W}_t^{\langle 0, \mathcal{S} \rangle}. \tag{13}$$

These marginal revenue and marginal promotion work capture the change in the expected total discounted revenue and promotion work, respectively, which results from being active instead of passive in the first time period and following the $\mathcal{S}$-active policy afterwards. Finally, if $w_t^{\mathcal{S}} \neq 0$, we define the $(t, \mathcal{S})$-*marginal promotion rate* as

$$\nu_t^{\mathcal{S}} := \frac{r_t^{\mathcal{S}}}{w_t^{\mathcal{S}}}. \tag{14}$$

Let us consider a family of nested sets $\mathcal{F} := \{\mathcal{S}_0, \mathcal{S}_1, \ldots, \mathcal{S}_T\}$, where $\mathcal{S}_k := \{1, 2, \ldots, k\}$. We will use the following theorem to establish indexability and obtain the index values.

Theorem 4 **(Niño-Mora (2002))** *If problem (1) satisfies the following two conditions (so-called PCL($\mathcal{F}$)-indexability):*

(i) the marginal promotion work $w_t^{\mathcal{S}} > 0$ for all $t \in \mathcal{T}$ and for all $\mathcal{S} \in \mathcal{F}$, and

(ii) the marginal promotion rate $\nu_t^{\mathcal{S}_{t-1}}$ is nonincreasing in $t \in \mathcal{T}$,

then the problem is indexable, family $\mathcal{F}$ contains an optimal active set for any value of parameter ν, and the index values are $\nu_t^* := \nu_t^{\mathcal{S}_{t-1}}$ for all $t \in \mathcal{T}$.

C Proof of Theorem 2

The ultimate goal of the proof is to apply Theorem 4 and derive a closed-form expression for the index given in Theorem 2. Plugging (10) and (11) into (12) and (13), respectively, we obtain two expressions that will be used in the following analysis:

$$r_t^{\mathcal{S}} = \left(R_t^1 - R_t^0\right) - (\beta q - \beta p) \sum_{s=1}^{t-1} \beta^{t-s-1} \boldsymbol{P}_{t-1,s}^{t-s-1|\mathcal{S}} R_s^{I_{\mathcal{S}}(s)}, \tag{15}$$

$$w_t^{\mathcal{S}} = \left(W_t^1 - W_t^0\right) - (\beta q - \beta p) \sum_{s=1}^{t-1} \beta^{t-s-1} \boldsymbol{P}_{t-1,s}^{t-s-1|\mathcal{S}} W_s^{I_{\mathcal{S}}(s)}. \tag{16}$$

It is well known in the MDP theory that the transition probability matrix for multiple periods is obtained by multiplication of transition probability matrices for subperiods. Hence, given an active set $\mathcal{S} \subseteq \mathcal{T}$, we have

$$\boldsymbol{P}^{t-s|\mathcal{S}} = \left(\boldsymbol{P}^{1|\mathcal{S}}\right)^{t-s}, \tag{17}$$

where the matrix $\boldsymbol{P}^{1|\mathcal{S}}$ is an $(T+1)\times(T+1)$-matrix constructed so that its row $x \in \mathcal{X}$ is the row x of the matrix $\boldsymbol{P}^{1|\mathcal{T}}$ if $x \in \mathcal{S}$, and is the row x of the matrix $\boldsymbol{P}^{1|\emptyset}$ otherwise. For definiteness, we remark that $\boldsymbol{P}^{0|\mathcal{S}}$ is an identity matrix.

We will use the following characterization of the marginal measures.

Lemma 1 *Let $t \in \mathcal{T}$ and consider any integer $0 \le k \le T$. Then,*

$$r_t^{\mathcal{S}_k} = \begin{cases} R(q-p)\left[(1-\beta)\dfrac{1-(\beta p)^{t-1}}{1-\beta p}\right. & \\ \left.+(1-\beta\alpha)\,(\beta p)^{t-1}\right], & \text{if } k \ge t-1 \ge 0, \\ R(q-p)\left[(1-\beta)\dfrac{1-(\beta q)^{t-k-1}}{1-\beta q}\right. & \\ +(1-\beta)\,(\beta q)^{t-k-1}\,\dfrac{1-(\beta p)^{k}}{1-\beta p} & \\ \left.+(1-\beta\alpha)\,(\beta q)^{t-k-1}\,(\beta p)^{k}\right], & \text{if } T-1 \ge t-1 \ge k. \end{cases} \tag{18}$$

$$w_t^{\mathcal{S}_k} = \begin{cases} W\left[1-(\beta q-\beta p)\dfrac{1-(\beta p)^{t-1}}{1-\beta p}\right], & \text{if } k \ge t-1 \ge 0, \\ W\left[1-(\beta q-\beta p)\,(\beta q)^{t-k-1}\,\dfrac{1-(\beta p)^{k}}{1-\beta p}\right], & \text{if } T-1 \ge t-1 \ge k. \end{cases} \tag{19}$$

Proof Under an active set $\mathcal{S}_k$, from (17) we get for $T \ge t-1 \ge s \ge 1$,

$$\boldsymbol{P}_{t-1,s}^{t-s-1|\mathcal{S}_k} = \begin{cases} p^{t-s-1}, & \text{if } k \ge t-1 \ge s \ge 0, \\ q^{t-s-1}, & \text{if } T \ge t-1 \ge s \ge k, \\ q^{t-k-1}p^{k-s}, & \text{if } T \ge t-1 \ge k \ge s \ge 0, \end{cases}$$

These expressions together with the definitions of R_t^a and W_t^a plugged into (15)–(16) after simplification conclude the proof. □

The following lemma establishes condition (i) of PCL($\mathcal{F}$)-indexability.

Lemma 2 *For any integer $k \ge 0$ we have*

(i) $w_k^{\mathcal{S}_k} > 0$;

(ii) $w_t^{\mathcal{S}_k} > 0$ *for all* $t \in \mathcal{T}$.

Proof Denote by

$$h(k) := (\beta q - \beta p)\frac{1-(\beta p)^k}{1-\beta p}, \tag{20}$$

so that, using (19), $w_k^{\mathcal{S}_k} = W\,[1-h(k)]$.

(i) For $k = 0$, we have $h(0) = 0$ by definition. For $k \geq 1$, $h(k) = (1 - (\beta p)^k)\frac{\beta q - \beta p}{1 - \beta p} < 1$. □
(ii) Implied by (i) and (19). □

Next we establish condition (ii) of PCL($\mathcal{F}$)-indexability.

Lemma 3 *We have*

$$\nu_t^{\mathcal{S}_{t-1}} = \frac{R}{W}\left\{[(1-p) - \alpha(1-\beta p)] - \frac{[(1-q) - \alpha(1-\beta q)]\,(1-\beta p)}{(1-\beta q) + (\beta q - \beta p)(\beta p)^{t-1}}\right\}.$$

Moreover, under Assumption 1, $\nu_t^{\mathcal{S}_{t-1}}$ is nonincreasing in $t \in \mathcal{T}$.

Proof Consider $\mathcal{S}_{t-1}$ for $t \in \mathcal{T}$. By Lemma 1 we have

$$r_t^{\mathcal{S}_{t-1}} = R(q-p)\left[(1-\beta)\frac{1-(\beta p)^{t-1}}{1-\beta p} + (1-\beta\alpha)\,(\beta p)^{t-1}\right],$$
$$w_t^{\mathcal{S}_{t-1}} = W\left[1 - (\beta q - \beta p)\frac{1-(\beta p)^{t-1}}{1-\beta p}\right],$$

therefore

$$\nu_t^{\mathcal{S}_{t-1}} = \frac{R}{W}\frac{(q-p)\left[(1-\beta)\frac{1-(\beta p)^{t}}{1-\beta p} + (1-\beta\alpha)\,(\beta p)^{t}\right]}{1 - (\beta q - \beta p)\frac{1-(\beta p)^{t}}{1-\beta p}}. \tag{21}$$

After multiplying both the numerator and the denominator by $1 - \beta p$, and rearranging the terms, we obtain

$$\nu_t^{\mathcal{S}_{t-1}} = \frac{R}{W}\left\{[(1-p) - \alpha(1-\beta p)] - \frac{[(1-q) - \alpha(1-\beta q)]\,(1-\beta p)}{(1-\beta q) + (\beta q - \beta p)(\beta p)^{t-1}}\right\}.$$

By Assumption 1, $(1-q) - \alpha(1-\beta q) \geq 0$ and $q - p > 0$, therefore $\nu_t^{\mathcal{S}_{t-1}}$ is nonincreasing in t. □

Finally, we can apply Theorem 4 to conclude that under Assumption 1, the problem is indexable and the index values are given by $\nu_t^{\mathcal{S}_{t-1}}$ in Lemma 3, which also establishes the time monotonicity. □

D Proof of Proposition 1

(i) Immediate from (21). □
(ii) Formally, we are to prove the following statement: If the probability q is replaced by $q' \leq q$, then $\nu_t^{*\prime} \leq \nu_t^*$ for any $t \in \mathcal{T}$. It is straightforward to see that (21) is nondecreasing in q. □