Dynamic Capacity Allocation to Customers Who Remember Past Service

Daniel Adelman, Adam J. Mersereau,

MANAGEMENT SCIENCE
Vol. 59, No. 3, March 2013, pp. 592–612
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)

http://dx.doi.org/10.1287/mnsc.1120.1643

Dynamic Capacity Allocation to Customers Who Remember Past Service

Daniel Adelman
Booth School of Business, The University of Chicago, Chicago, Illinois 60637,
dan.adelman@chicagobooth.edu

Adam J. Mersereau
Kenan-Flagler Business School, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599,
ajm@unc.edu

We study the problem faced by a supplier deciding how to dynamically allocate limited capacity among a portfolio of customers who remember the fill rates provided to them in the past. A customer's order quantity is positively correlated with past fill rates. Customers differ from one another in their contribution margins, their sensitivities to the past, and in their demand volatilities. By analyzing and comparing policies that ignore goodwill with ones that account for it, we investigate when and how customer memory effects impact supplier profits. We develop an approximate dynamic programming policy that dynamically rationalizes the fill rates the firm provides to each customer. This policy achieves higher rewards than margin-greedy and Lagrangian policies and yields insights into how a supplier can effectively manage customer memories to its advantage.

Key words: dynamic programming; approximate; behavioral operations; customer relationship management; capacity allocation
History: Received January 10, 2012; accepted July 17, 2012, by Martin Lariviere, operations management. Published online in *Articles in Advance* December 19, 2012.

1. Introduction

Whenever a firm doing business with a handful of customers (or customer segments) faces more demand than it can supply, it faces a tough choice. On the one hand, there is a short-term opportunity for profit taking, by supplying only the most profitable customers today. However, this is potentially damaging to the firm's relationships with less profitable customers, who receive poor service today. When market conditions change in the future, these neglected customers could be essential to maintaining profitability, but their goodwill toward the firm may be so diminished that their business will not materialize when it is needed most.

The purpose of this paper is to develop a model that can provide insight and decision support for firms dealing with this issue. We explore the inherent dynamics of such a system and the shortcomings of ignoring customer goodwill dynamics when they are present. The problem is dynamic and high dimensional, so we employ approximate dynamic programming (ADP) to optimize the firm's decisions. The integration of a behavioral model within a modern ADP framework is novel. We believe there are opportunities to build new and interesting models with this construct in many different contexts. This paper provides a starting point for such future research.

Our model trades off three basic customer characteristics—margin, demand volatility, and memory—when assessing the value of a customer. Taking these factors into account, the supplier must decide the quality of service to provide each customer in the current period, knowing that neglecting a customer now will adversely impact future demand from the customer. The firm therefore must rationalize trade-offs between margin, demand volatility, and memory in a dynamic fashion.

We formulate the problem as an average reward Markov decision process (MDP) whose state is a vector of customer goodwill representing exponentially smoothed summaries of past fill rates provided to individual customers. Customers are heterogeneous in their smoothing parameters, which parameterize the memories of the customers. Using stochastic analysis and ADP, we draw the following conclusions:

1. *When goodwill matters*. By analyzing and evaluating a margin-greedy policy, we identify when goodwill effects are relatively (un)important. We find that goodwill effects become more important as demand variance increases and as customer memories become shorter. Also, we find that a portfolio of a few large customers requires more careful goodwill management than does a portfolio of many small customers,

and thereby we rigorously show the managerial effectiveness of broadening a firm's customer base.

2. *How goodwill matters*. A supplier generally prefers to have customers with high margins and low demand variance. When the supplier fluctuates the fill rate provided to a customer, the length of the customer's memory drives how much the goodwill fluctuates in response, and in turn affects order sizes. Our analysis shows that the length of a customer's memory has an ambiguous effect on supplier rewards. Long-term customer memories tend to dampen order variances but can have positive or negative impacts on average order sizes.

3. *What to do about it*. (a) *Intuition*. Effective goodwill management involves maintaining customer expectations within manageable ranges (potentially holding back capacity to do so), cultivating a portfolio of active customers so that the supplier can pool demand variances across customers, and actively managing the system dynamics induced by customer memories. (b) *Goodwill-sensitive policies*. Our ADP policy effectively rationalizes these various factors dynamically to significantly outperform margin-greedy and Lagrangian policies in many instances. We interpret the policy as prioritizing each customer using an "adjusted margin" consisting of the customer's margin and an offset that values the goodwill impact from fulfilling current orders.

We review related literature in the remainder of this section. In §2 we formulate our problem rigorously. Section 3 provides some insights into the inherent system dynamics and investigates, through detailed analysis of a greedy myopic heuristic, how and when the goodwill effects we model impact supplier profits. We analyze approximate policies in §4. After a numerical study in §5, we summarize our findings in §6. Most proofs are provided in the appendix.

1.1. Literature Review

Our work is related to several streams of literature. Several researchers have recognized that poor service levels in the current period may diminish demand in subsequent periods. Olsen and Parker (2008) refer to this phenomenon as "market size dynamics," which they model by assuming that customers transition between "committed" and "latent" states based on the service provided. Customers are otherwise homogeneous. Hall and Porteus (2000) and Liu et al. (2007) assume a finite pool of homogeneous customers where customers may defect to a competitor upon experiencing poor service in the current period. In Gaur and Park (2007), customers' choices among retailers are driven by their past stockout experiences. Gans (2002) assumes that each customer switches among suppliers based on her assessment of quality from past experience. Our work is distinguished from these papers by our assumption of a countable set of heterogeneous and addressable customers. We are particularly interested in instances with a relatively small number (say, 20 or fewer) customers. Another distinction is that these papers model competitive settings. Although our paper does not explicitly model competition, our model of customer behavior—in particular, our assumption that orders increase with the quality of past service—is consistent with customer behavior that arises in this literature.

The "inverse newsvendor" problem of Carr and Lovejoy (2000) involves a supplier selecting a customer portfolio as in our work. They differentiate customers by static demand attributes (i.e., mean and variance), in contrast with our dynamic model. Cachon and Lariviere (1999) examine the allocation of limited capacity to a heterogeneous set of customers using a "turn-and-earn" policy that allocates to retailers based on past sales. Their interest is in the induced competition among customers, whereas ours is in the supplier's policy optimization. There has also been work on a problem converse to ours: the allocation of buyers' orders to a heterogeneous set of suppliers. In the paper by Federgruen and Yang (2008), suppliers are heterogeneous in the likelihood of supply disruptions. Cachon and Zhang (2007) explore capacity competition among suppliers when buyers allocate demand among suppliers based on the quality of service provided.

We borrow the notion of "goodwill stock" from Nerlove and Arrow (1962), who use it to model the persistent effects of advertising efforts in the customer base. A substantial stream of literature on the optimal control of advertising expenditures followed. Surveys on this research include those by Sethi (1977) and Feichtinger et al. (1994). The interest in that research stream is on analytically characterizing the optimal timing of advertising expenditures using continuous-time optimal control formulations. Our use of goodwill is closest to what Sethi (1977, p. 688) categorizes as "advertising capital models." Customers typically come from a single homogeneous segment in this literature, and there are no natural notions of customer demand or orders. Dube et al. (2005) use the Nerlove and Arrow (1962) model in choosing advertising and pricing to a single homogeneous customer segment. A portion of their contribution is demonstration of model estimation based on data. Contemporary to our work is that of Aflaki and Popescu (2012), who consider a single customer with no capacity constraint, deterministic demand, and the possibility of customer defection. Among their findings are that customers with short memories merit better service because of the danger of defection and that loss-averse customers merit consistent service. We make related observations, but driven by stochasticity and capacity constraints.

There is a substantial interest in the marketing literature on customer lifetime value (CLV) and customer relationship management (CRM). Surveys on these topics are those by Gupta and Lehmann (2008) and Reinartz and Venkatesan (2008). This literature mainly focuses on estimation of CLV and classification-based schemes for CRM. We are aware of little work on dynamic optimization in this area. An exception is that of Simester et al. (2006), who use dynamic programming to compute a policy for mailing catalogs.

We believe that the trade-off we study is fundamental and universal, and that one explanation for the paucity of research on it may be that methodological tools to handle it have not been investigated until recently. Textbook treatments of ADP include those by Bertsekas and Tsitsiklis (1996) and Powell (2007). We specifically make use of the approximate linear programming approach (Schweitzer and Seidmann 1985, de Farias and Van Roy 2003).

2. Problem Formulation

We introduce some notation before formulating our problem as a Markov decision problem. We then develop some mathematical programming formulations of the problem, on which we will later base our main approximation approach.

2.1. Markov Decision Process Formulation

A supplier sells a single good over time to a finite set of customers $\mathcal{I} \equiv \{1, 2, \ldots, n\}$, subject to a capacity constraint of $\bar{X} > 0$ units per period. Over time, each customer develops a sense of goodwill toward the supplier based upon the quality of past service the customer has received. We denote customer i's goodwill upon entering period t by a scalar G_t^i, which we assume the customer takes as a sufficient summary statistic. At the beginning of each period, each customer i places an order with the supplier of size $y^i(G_t^i, D_t^i)$ units. This is a function of both a random demand component D_t^i and the customer's current goodwill G_t^i. Denote customer i's order in period t by y_t^i. Given the vector $\mathbf{y}_t = (y_t^1, \ldots, y_t^n)$, the supplier must choose a vector of shipments $\mathbf{x}_t = (x_t^1, \ldots, x_t^n)$ such that $0 \le x_t^i \le y_t^i$ for all i and $\sum_i x_t^i \le \bar{X}$, where x_t^i denotes the delivery quantity to customer i in period t. The supplier makes a profit of $r^i > 0$ for each unit supplied to customer i. The objective is to maximize the supplier's long-run average profit. The discounted profit criterion could be tackled using a similar methodology to the one we present. Focusing instead on long-run profits allows us to compare customers in terms of their fundamental attributes without regard to initial states or transient effects.

We model goodwill as an exponential smoothing of utilities derived from past supplier fill rates, inspired by the classical Nerlove and Arrow (1962) model. Customer i updates her goodwill as follows:

$$G_{t+1}^i = \begin{cases} \beta^i G_t^i + u^i(x_t^i / y^i(G_t^i, D_t^i)) & \text{if } y^i(G_t^i, D_t^i) > 0, \\ \beta^i G_t^i & \text{if } y^i(G_t^i, D_t^i) = 0, \end{cases}$$

where $\beta^i \in (0, 1)$ is a customer-specific memory parameter and $u^i\colon [0, 1] \to \mathbb{R}_+$ is a nondecreasing, continuous customer-specific utility function. A higher β^i corresponds with longer-term memory of past goodwill. Of course, other models of customer behavior are possible. Fader and Hardie (2009) provide a review of probability models of customer retention, for example. Our choice of model is driven by parsimony, as it yields a one-dimensional state for each customer and is arguably the simplest way to summarize a customer's past service.

We assume that the vector of demands $\mathbf{D}_t = (D_t^1, D_t^2, \ldots, D_t^n)$ in period t follows a stationary distribution with cumulative distribution function $\mathcal{F}$ defined over a compact Borel vector space $\mathcal{D}$ of demand scenarios. We permit demands across customers to be statistically dependent, but we require $\mathbf{D}_t$ to be independent across periods. Let $\mathbf{G}_t = (G_t^1, G_t^2, \ldots, G_t^n)$ denote the vector of goodwills in period t. We drop the time index t when appropriate.

To formulate the problem as a Markov decision process, we first define state and action spaces. The lowest goodwill is achieved by a zero fill rate, i.e., solves $G^i = \beta^i G^i + u^i(0)$, which yields $G^i = u^i(0)/(1 - \beta^i)$. Likewise, the highest goodwill achievable is $G^i = u^i(1)/(1 - \beta^i)$. Hence, we define the goodwill state space as the compact set $\mathcal{G} = \{\mathbf{G} \in \mathbb{R}_+^n\colon u^i(0)/(1 - \beta^i) \le G^i \le u^i(1)/(1 - \beta^i)\; \forall\, i \in \mathcal{I}\}$. Given a Borel space Z, we denote by $\mathbb{B}(Z)$ the Banach space of all Borel measurable bounded functions on Z, i.e., functions f having finite norm $\|f\| = \sup_{z \in Z} |f(z)|$. For a given goodwill state $\mathbf{G}$, the action space $\mathcal{X}(\mathbf{G})$ represents the allocations provided under each demand scenario and is defined by

$$\mathcal{X}(\mathbf{G}) = \left\{ \mathbf{x} \in \mathbb{B}(\mathcal{D})\colon \text{ for all } \mathbf{D} \in \mathcal{D}, \right.$$
$$\left. \sum_{i \in \mathcal{I}} x^i(\mathbf{D}) \le \bar{X}, 0 \le x^i(\mathbf{D}) \le y^i(G^i, D^i)\; \forall\, i \in \mathcal{I} \right\}.$$

It is also useful to define the action space after a demand scenario $\mathbf{D} \in \mathcal{D}$ is observed:

$$\mathcal{X}(\mathbf{G}, \mathbf{D}) = \left\{ \mathbf{x} \in \mathbb{R}_+^n\colon \sum_{i \in \mathcal{I}} x^i \le \bar{X}, 0 \le x^i \le y^i(G^i, D^i)\; \forall\, i \in \mathcal{I} \right\}.$$

Throughout this paper, we will make the following assumption:

Assumption 1. *The ordering functions $\{y^i\}_{i \in \mathcal{I}}$ are continuous, nondecreasing, lower bounded away from zero, and bounded on $\mathcal{G} \times \mathcal{D}$.*

Under this assumption, the state–action space

$$\mathcal{M} = \{(\mathbf{G}, \mathbf{x}): \mathbf{G} \in \mathcal{G}, \mathbf{x} \in \mathcal{X}(\mathbf{G})\}$$

is compact. The assumption that the y^is are lower bounded away from zero avoids mathematical complications around division by zero. We require y^is to be nondecreasing for several of our analyses in §3, but our formulation and ADP methodology do not require this assumption.

Let $\boldsymbol{\beta}\mathbf{G} + \mathbf{u}(\mathbf{x}/\mathbf{y}(\mathbf{G}, \mathbf{D}))$ be a vectorized representation of the expressions $\beta^i G^i + u^i(x^i/y^i(G^i, D^i))$ for all i, which denote the next state. We can write the unichain optimality equations in two equivalent ways:

$$h(\mathbf{G}) = \mathsf{E}_{\mathbf{D}}\left[\max_{\mathbf{x}\in\mathcal{X}(\mathbf{G},\mathbf{D})}\left\{\sum_i r^i x^i - \rho + h\left(\boldsymbol{\beta}\mathbf{G} + \mathbf{u}\left(\frac{\mathbf{x}}{\mathbf{y}(\mathbf{G},\mathbf{D})}\right)\right)\right\}\right] \quad \forall \mathbf{G} \in \mathcal{G}, \tag{1}$$

$$h(\mathbf{G}) = \max_{\mathbf{x}\in\mathcal{X}(\mathbf{G})} \mathsf{E}_{\mathbf{D}}\left[\sum_i r^i x^i(\mathbf{D}) - \rho + h\left(\boldsymbol{\beta}\mathbf{G} + \mathbf{u}\left(\frac{\mathbf{x}(\mathbf{D})}{\mathbf{y}(\mathbf{G},\mathbf{D})}\right)\right)\right] \quad \forall \mathbf{G} \in \mathcal{G}, \tag{2}$$

where $h(\cdot) \in \mathbb{B}(\mathcal{G})$ is the bias function and $\rho \in \mathbb{R}$ is the gain. Our problem is not unichain, but these optimality equations suffice for our purposes because they lead to valid approximations and bounds.

Several features make this problem challenging to solve. First, we consider multiple, addressable customers in contrast to related works on single-customer problems (e.g., Sethi 1977, Aflaki and Popescu 2012) or on homogeneous pools of customers (e.g., Gaur and Park 2007). The capacity constraint in our problem does not permit our problem to be decomposed by customers without loss of optimality. Second, stochastic demands lead to stochastic problem dynamics so that a steady state in our problem is necessarily characterized by a complex joint probability distribution. Third, the use of fill rate as a measure of service quality induces nonlinear state dynamics that complicate analysis. Despite these complications, we believe that these features are essential to our problem.

2.2. Mathematical Programming Formulations

Changing the equality in (2) into an inequality leads to the infinite-dimensional linear program

$$(\mathrm{D}_0) \quad \inf_{\rho, h} \rho \tag{3a}$$

$$h(\mathbf{G}) \geq \mathsf{E}_{\mathbf{D}}\left[\sum_{i\in\mathcal{I}} r^i x^i(\mathbf{D}) - \rho + h\left(\boldsymbol{\beta}\mathbf{G} + \mathbf{u}\left(\frac{\mathbf{x}(\mathbf{D})}{\mathbf{y}(\mathbf{G},\mathbf{D})}\right)\right)\right] \quad \forall (\mathbf{G}, \mathbf{x}) \in \mathcal{M}, \tag{3b}$$

$$h \in \mathbb{B}(\mathcal{G}), \rho \in \mathbb{R}. \tag{3c}$$

Because feasible solutions (ρ, h) have a bounded bias function, it follows from Lemma 6.4.1 of Hernández-Lerma and Lasserre (1996) that the solution to (D_0) provides an upper bound on the reward of any policy starting from any state.

To formulate the dual program, let $\mathcal{B}(Z)$ denote the collection of Borel subsets of a Borel space Z. For any set $\bar{\mathcal{G}} \in \mathcal{B}(\mathcal{G})$, let $\mathcal{M}(\bar{\mathcal{G}}) = \{(\mathbf{G}, \mathbf{x}) \in \mathcal{M}: \mathbf{G} \in \bar{\mathcal{G}}, \mathbf{x} \in \mathcal{X}(\mathbf{G})\}$ be the set of associated state–action pairs leading from states $\mathbf{G} \in \bar{\mathcal{G}}$. Also, let $\mathbb{M}(Z)$ denote the Banach space of signed measures on the Borel space Z having finite total variation norm, and let $\mathbb{1}\{\cdot\}$ denote the indicator function. We then have

$$(\mathrm{P}_0) \quad \sup_{\mu} \int_{(\mathbf{G},\mathbf{x})\in\mathcal{M}} \mathsf{E}_{\mathbf{D}}\left[\sum_{i\in\mathcal{I}} r^i x^i(\mathbf{D})\right] d\mu(\mathbf{G}, \mathbf{x}), \tag{4a}$$

$$\mu(\mathcal{M}(\bar{\mathcal{G}})) - \int_{(\mathbf{G}',\mathbf{x}')\in\mathcal{M},\, \mathbf{D}\in\mathcal{D}} \mathbb{1}\left\{\boldsymbol{\beta}\mathbf{G}' + \mathbf{u}\left(\frac{\mathbf{x}'(\mathbf{D})}{\mathbf{y}(\mathbf{G}',\mathbf{D})}\right) \in \bar{\mathcal{G}}\right\} \cdot d\mu(\mathbf{G}', \mathbf{x}')\, d\mathcal{F}(\mathbf{D}) = 0 \quad \bar{\mathcal{G}} \in \mathcal{B}(\mathcal{G}), \tag{4b}$$

$$\mu(\mathcal{M}) = 1, \tag{4c}$$

$$\mu \geq 0, \mu \in \mathbb{M}(\mathcal{M}). \tag{4d}$$

The measure μ represents the average time spent in each set of state–action pairs. Constraint (4b) requires that the flow in equals the flow out for every Borel subset of $\mathcal{G}$, whereas constraint (4c) ensures that μ is a probability measure.

3. The Greedy Policy

A natural policy to study is the myopic margin-greedy policy (or simply the "greedy" policy), which allocates capacity based on margins alone and without regard to goodwill; that is, in period t the supplier solves

$$\max \sum_i r^i x_t^i \quad \text{s.t.} \quad \sum_i x_t^i \leq \bar{X}, \quad 0 \leq x_t^i \leq y_t^i \quad \forall i \in \mathcal{I}.$$

Understanding when the greedy policy performs well or poorly reveals when goodwill effects are most important for the supplier to account for. We provide results suggesting when greedy allocations perform well (§§3.2 and 3.3), but we also illustrate two reasons why a greedy policy can fall short (§§3.4 and 3.5). Along the way, we prove that the greedy policy can perform arbitrarily badly, and we gain insight into behaviors of good policies. First, we build some preliminary insight through a stochastic analysis under the greedy policy.

3.1. System Dynamics Under the Greedy Policy

Consider a single-customer problem with capacity $\bar{X}$, a linear utility function $u(f) = f$, and a linear ordering function $y(G, D) = (1-\beta)GD$.[1] To

[1] To satisfy Assumption 1, our analysis of this ordering function in this and subsequent sections carries through if we add a small positive constant to all orders. We suppress this for expositional convenience.

Figure 1 Illustration of the System of Equations Governing the Problem Dynamics in the Simplified Context of §3.1

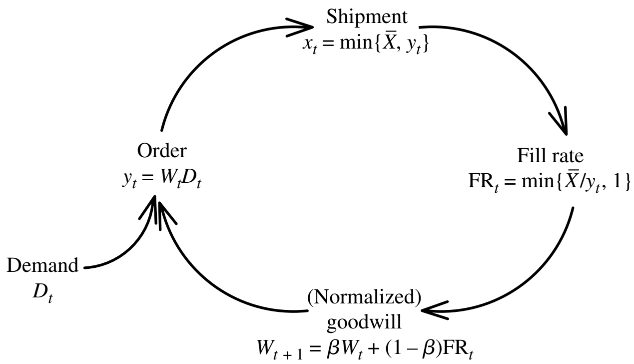
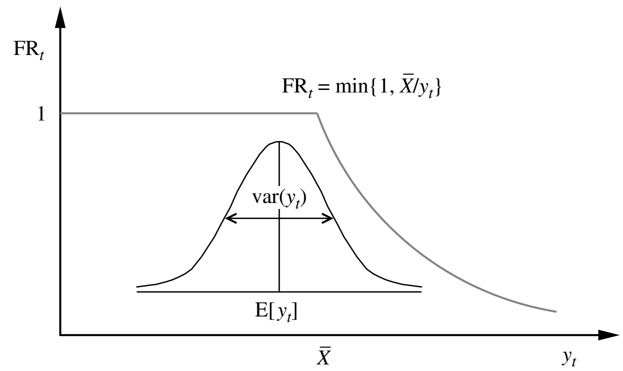


Figure 2 Illustration of Fill Rate as a Function of y_t in the Simplified Context of §5.2



streamline the exposition, we will represent goodwill using the normalized rescaling $W_t = (1-\beta)G_t$, updated according to the equation $W_{t+1} = \beta W_t + (1-\beta)u(x_t/y(W_t/(1-\beta), D_t))$. It is straightforward to show that this is equivalent to the update $G^i_{t+1} = \beta^i G^i_t + u^i(x^i_t/y^i(G^i_t, D^i_t))$. The customer order at time t is therefore $y_t = \bar{y}_t(W_t, D_t) = W_t D_t$.

If we further simplify by assuming the supplier is using a greedy policy, then we can express the period t shipment in closed form as $x_t = \min\{\bar{X}, y_t\}$ and the fill rate as $\mathrm{FR}_t = \min\{\bar{X}/y_t, 1\}$. Figure 1 illustrates the system of equations that relates the demands, orders, shipments, fill rates, and normalized goodwills in this simplified context. We assume that this process is stationary.

Viewing $y_t = W_t D_t$ as a random variable, we see that the function $x_t = \min\{\bar{X}, y_t\}$ is nondecreasing and concave in y_t. Therefore, we expect the expected shipments $\mathsf{E}[x_t]$ and hence expected rewards to be increasing in $\mathsf{E}[y_t]$ and decreasing in the variability of y_t;[2] that is, the supplier has a natural preference for customers with high expected orders and low order variability.

The variability of y_t is driven in part by variability in W_t. Taking the variance of both sides of the update equation $W_{t+1} = \beta W_t + (1-\beta)\mathrm{FR}_t$,

$$\mathrm{Var}[W_{t+1}] = \beta^2 \mathrm{Var}[W_t] + (1-\beta)^2 \mathrm{Var}[\mathrm{FR}_t] + \beta(1-\beta)\mathrm{Cov}[W_t, \mathrm{FR}_t],$$

and recognizing that $\mathrm{Var}[W_t] = \mathrm{Var}[W_{t+1}]$ because of the stationarity of the W_t process,

$$\mathrm{Var}[W_t] = \frac{1-\beta}{1+\beta}\mathrm{Var}[\mathrm{FR}_t] + \frac{2\beta}{1+\beta}\mathrm{Cov}[W_t, \mathrm{FR}_t].$$

Because $\mathrm{Var}[\mathrm{FR}_t]$ is bounded, we have that $\mathrm{Var}[W_t] \to \mathrm{Cov}[W_t, \mathrm{FR}_t]$ in the limit as $\beta \to 1$. Furthermore, the covariance will be less than or equal to zero because $\mathrm{FR}_t = \min\{1, \bar{X}/(W_t D_t)\}$ is a nonincreasing function of W_t for fixed D_t. It follows that $\mathrm{Var}[W_t] \to 0$ as $\beta \to 1$.

Thus, it seems that large β tends to dampen the variability of a customer's orders. On the other hand, the impact of β on *expected* orders is less clear. Note that W_t is simply a weighted average of past fill rates and that FR_t is stationary; thus in steady state we expect $\mathsf{E}[W_t] = \mathsf{E}[\mathrm{FR}_t]$, implying that $\mathsf{E}[y_t] = \mathsf{E}[W_t]\mathsf{E}[D_t] = \mathsf{E}[\mathrm{FR}_t]\mathsf{E}[D_t] = \mathsf{E}[\min\{\bar{X}/y_t, 1\}]\mathsf{E}[D_t]$. The function $\min\{\bar{X}/y_t, 1\}$ is illustrated in Figure 2. It is neither concave nor convex as a function of y_t. As a result, a large β's tendency to reduce the variability in y_t can have a positive or negative influence on expected orders. We will explore this further in §5.2.

To summarize, the brief analysis in this section reveals the complexity of the trade-offs in our problem. We see that the firm has a natural preference, at least under the greedy policy, for customers with high expected orders and low order variability. Long customer memories tend to dampen order variability, but long customer memories can have a positive or negative influence on expected orders.

3.2. Asymptotic Optimality

Consider a sequence of problems indexed by n, generated as follows. Problem n endows the manager with capacity $\bar{X} = n\bar{x}$ to service n customers, with $\bar{x}$ some fixed positive number. Customers are identical except for their margins and memory parameters. Define fixed infinite sequences $(r^1, r^2, \ldots,)$ and $(\beta^1, \beta^2, \ldots,)$ such that $0 < r^i < r_{\max}$ for some $r_{\max}$, $0 < \beta^i < 1$ for all i, and (for simplicity) $r^i \neq r^j$ for $i \neq j$. We assume that problem n consists of customers with margins $r^1, \ldots, r^n$ and memory parameters $\beta^1, \ldots, \beta^n$. D^is are independent (across time and customers) and identically distributed. We let $y() = y^i()$ and $u() = u^i()$ for all i, and we let D indicate a random variable

[2] We require some technical conditions to make this statement rigorous. For example, $\mathsf{E}[x_t]$ increases if we make the distribution of y_t stochastically greater, for example, by adding a positive constant to y_t. Also, $\mathsf{E}[x_t]$ is decreasing in the riskiness of y_t when we measure "riskiness" in the second-order stochastic dominance sense (Müller and Stoyan 2002, Theorem 8.1.1); this occurs, for example, if we add a zero-mean random variable to y_t. See Müller and Stoyan (2002) for definitions and further details.

with identical distribution to D_t^i for all i and t. When considering problem n, we reindex the customers such that $r^1 > r^2 > \cdots > r^n$. Per Assumption 1, $y()$ is bounded. Specifically, there exists some $0 < y_{\max} < \infty$ such that $\Pr\{y(u(1)/(1-\beta^i), D) < y_{\max}\} = 1$ for all i.

Define $\rho_n^{\mathcal{G}}$ as the average reward per customer under the greedy policy in problem n, and let ρ_n^* indicate the average reward per customer in problem n obtained under an optimal policy. The following theorem establishes the asymptotic optimality of the greedy policy. Similar results can be found in the literature on restless bandit problems.[3]

Theorem 1. $\lim_{n\to\infty}(\rho_n^* - \rho_n^{\mathcal{G}}) = 0$.

We prove this theorem in the appendix. It suggests that a portfolio of few large customers requires more careful goodwill management than does a portfolio of many small, independent customers. This is consistent with our interest, expressed in §1, in instances involving a handful of addressable customers. As the number of independent customers grows, demand uncertainties are pooled across customers and the system approaches a deterministic fluid model. As we will see in the next subsection, a greedy allocation of capacity is optimal when the system is deterministic.

3.3. The Deterministic Case

In this section, we consider a version of the problem in which the demand random variable $\mathbf{D}$ is constant and deterministic. For convenience, here we write $x^i = x^i(\mathbf{D})$ and $y^i(G^i) = y^i(G^i, D^i)$. Assume without loss of generality that customers are sorted in decreasing order of their margins (i.e., $r^1 \geq r^2 \geq \cdots \geq r^n$). Define the state–action pair $(\mathbf{G}^*, \mathbf{x}^*)$, where

$$(x^*)^i = \begin{cases} y^i\left(\dfrac{u^i(1)}{1-\beta^i}\right) & \text{if } \displaystyle\sum_{j=1}^{i} y^j\left(\dfrac{u^j(1)}{1-\beta^j}\right) \leq \bar{X}, \\[2ex] \bar{X} - \displaystyle\sum_{j=1}^{i-1} y^j\left(\dfrac{u^j(1)}{1-\beta^j}\right) & \\ \qquad \text{if } \displaystyle\sum_{j=1}^{i-1} y^j\left(\dfrac{u^j(1)}{1-\beta^j}\right) \leq \bar{X} \leq \displaystyle\sum_{j=1}^{i} y^j\left(\dfrac{u^j(1)}{1-\beta^j}\right), & \\[2ex] 0 & \text{otherwise}, \end{cases}$$

and $(G^*)^i$ solves the equation

$$\beta^i (G^*)^i + u^i\left(\frac{(x^*)^i}{y^i((G^*)^i)}\right) = (G^*)^i.$$

It is straightforward to show that this equation has a solution and that the solution satisfies $u^i(0)/(1-\beta^i) \leq (G^*)^i \leq u^i(1)/(1-\beta^i)$. Therefore, $\mathbf{x}^*$ constitutes a feasible policy starting from state $\mathbf{G}^*$, yielding reward $\rho^* = \sum_i r^i \cdot (x^*)^i$ each period.

Proposition 1. *When* $\mathbf{D}$ *is deterministic, a Dirac measure concentrated on* $(\mathbf{G}^*, \mathbf{x}^*)$ *solves* (P_0).

It follows from Proposition 1 that ρ^* is the maximum average reward achievable under any policy starting from any state. We conclude that an optimal steady-state allocation in the deterministic case is greedy with respect to customer margins. An important implication of this is that customer memories only impact a supplier's optimal steady-state rewards when there is demand variance. Proposition 1 characterizes the optimal allocations given to customers in steady state, but it does not provide guidance on how to reach that steady state. A greedy *policy* may not be optimal, as illustrated in the next subsection.

3.4. The Greedy Policy Can Be Arbitrarily Bad

In this section, we present a family of instances designed to demonstrate that the greedy policy can perform arbitrarily badly. Consider a family of deterministic, single-customer problems parameterized by the customer's discount factor β. The seller has capacity $\bar{X} = 1$, and the customer's ordering function $y()$ is illustrated in Figure 3.

It is not difficult to design a simple policy for each instance that maintains customer goodwill $W = \beta$. Suppose $W_t = \beta$ for some t; then the customer will place order $y_t = y(\beta) = 1$. The seller can ship $x_t = \beta < 1$ to the customer and maintain goodwill $W_{t+1} = \beta$.

On the other hand, for particular choices of ordering function parameters ϵ and $\bar{Y}$, the greedy policy can cycle between a high goodwill (where the customer places a very large order $\bar{Y}$ and receives a poor

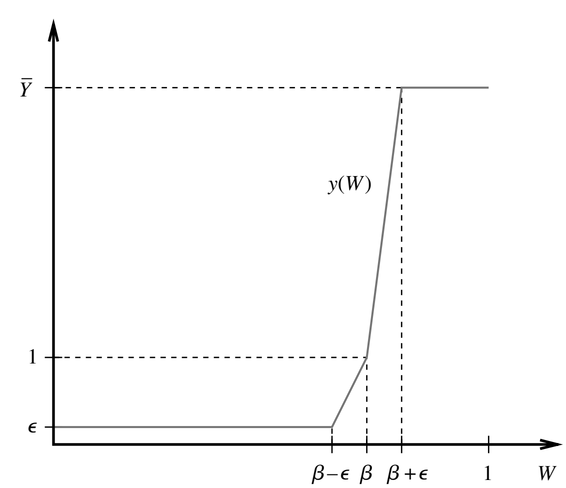


Figure 3 Customer Ordering Function in the Family of Instances Considered in §3.4

[3] Although the statement of our theorem is similar to a conjecture by Whittle (1988) analyzed by Weber and Weiss (1990), these results do not apply to our problem. Furthermore, our proof is significantly different, owing to problem-specific structure and the fact that our problem is defined on a continuous state space.

Management Science 59(3), pp. 592–612, © 2013 INFORMS

fill rate) and several periods of low goodwill (where the customer places orders of size ϵ and receives 100% fill rate). The parameter $k(\beta)$, defined below, determines the length of such a cycle. When we take $\beta \to 1$, $\epsilon \to 0$, and $\bar{Y} \to \infty$, the fraction of periods with low goodwill increases, and the average rewards achieved under the greedy policy become arbitrarily small.

This is the essence of Proposition 2. First, we define the family of problems in detail. We consider $\beta < 1$ and $\beta(1+\beta^2) > 1$. (Restricting ourselves to $\beta \in (0.69, 1)$ is sufficient for this to be true.) Define

$$k(\beta) = \left\lceil \frac{\log(1+\beta^2)}{-\log(\beta)} \right\rceil - 1.$$

The condition $\beta(1+\beta^2) > 1$ implies that $k(\beta) \geq 1$. Given $k = k(\beta)$, we define the parameter $\bar{Y} = \bar{Y}(\beta)$ as the solution to the equation

$$\frac{\beta - 1 + \beta^k}{\beta^k} = \beta(1-\beta+\beta^2) + (1-\beta)(1/\bar{Y}).$$

Equivalently, $\bar{Y} = (1+\beta^2-\beta^{-k})^{-1}$. Such a $\bar{Y} > 0$ exists because the definition of k implies $\beta^{-k} < 1+\beta^2$. Let $\epsilon = \epsilon(\beta)$ be a small positive number satisfying $0 < \epsilon < \min\{(1-\beta)^2, 1/k\}$. Finally, let $\rho^*(\beta)$ denote the average reward under an optimal policy, and let $\rho^{\mathcal{G}}(\beta)$ denote the average reward under the greedy policy.

Proposition 2. *For the family of instances considered here with $W_0 = \beta$,*
(a) *$\rho^*(\beta) \geq \beta$ for all β, and*
(b) $\lim_{\beta \to 1} \rho^{\mathcal{G}}(\beta) = 0$.

It follows from the proposition that the relative difference between $\rho^*(\beta)$ and $\rho^{\mathcal{G}}(\beta)$ approaches infinity as $\beta \to 1$. This family of instances provides an extreme example of a customer with high expectations, little tolerance for lapses in service, and a long memory of past service. The example yields insight into one reason the greedy policy is suboptimal; it can set customer expectations to unsustainably high levels. Better policies for these instances, in contrast, strategically withhold capacity to keep customer goodwills within a manageable range. We explore another reason for the suboptimality of the greedy policy in the following section.

3.5. Portfolio Effects

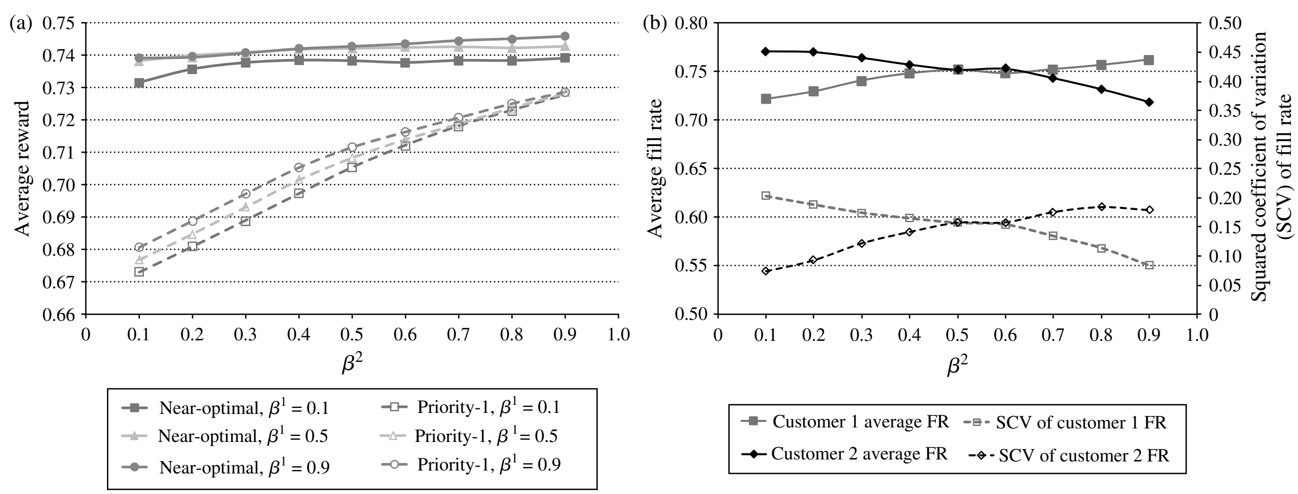
We consider a simple two-customer problem in which the two customers $i \in \{1, 2\}$ generate the same margin and face independent two-point demand random variables ($\Pr\{D^i = 0.5\} = 1/2$ and $\Pr\{D^i = 1\} = 1/2$ for $i \in \{1, 2\}$). The seller's capacity is 0.75, the utility functions are taken to be the identity function ($u^i(x) = x$ for $i \in \{1, 2\}$), and the ordering functions are taken to be linear ($y^i(G^i, D^i) = (1-\beta^i)G^iD^i$ for $i \in \{1, 2\}$). In short, the two customers are assumed to be identical except for their memory parameters. We evaluate two policies via simulation:

(i) The first policy, "priority-1," always prioritizes customer 1, allocating any remaining capacity to customer 2 until either customer 2's order is filled or the supplier runs out of capacity. This policy exhibits behavior typical of a greedy policy in that it strictly prioritizes customers without regard to memories or demand variances.

(ii) The second is a "near-optimal" policy, generated using a fine grid approximation of the bias function (see Judd 1998). This method is not suitable for high-dimensional problems, but is tractable here given this instance's two-dimensional state space and simple demand model.

Numerical simulation results in Figure 4(a) show that the near-optimal policy achieves average rewards up to approximately 9% larger than the priority-1 policy. We found that the priority-1 policy gives customers 1 and 2 average fill rates of approximately 91%

Figure 4 **(a) Average Rewards Generated by Near-Optimal and Priority-1 Policies in the Example of §3.5; (b) Decisions of the Near-Optimal Policy When $\beta^1 = 0.5$**



and 40%, respectively, over a wide range of choices of the memory parameters β^1 and β^2. In contrast, Figure 4(b) indicates that the average fill rates generated by the near-optimal policy are much more equitable. For intuition into why this approach is effective, consider the random variable $Y = y^1(G^1, D^1) + y^2(G^2, D^2)$, representing the aggregate order quantities seen by the seller. By maintaining similar goodwills among the customers, the seller effectively pools the uncertainty in the random variables D^1 and D^2, yielding a less variable Y than if the seller favored one customer over the other. We conclude that a supplier can take advantage of a "portfolio effect" by keeping multiple customers ordering actively and pooling demand variances across customers. Essentially, it is important for the supplier to keep both customers reasonably satisfied so that when one of them experiences a low demand, the other can be counted on to supply orders if its demand materializes.

Figure 4(b) reveals that, although the fill rates are more equitably distributed under the near-optimal policy than under the priority-1 policy, the near-optimal policy rewards and allocations do vary with the customer memories. The supplier prefers customers with bigger βs—its rewards are increasing with both β^1 and β^2 under the near-optimal policy—but the supplier treats the low-β customer better. Customer 2 gets lower and more variable fill rates as β^2 increases relative to β^1. The analysis of §3.1 suggests that short customer memories tend to amplify order variances relative to long customer memories. We see that the near-optimal policy here counteracts this amplification by supplying the short-memory, or less loyal, customer more stably. Thus, our results offer a mathematical argument for the adage "the squeaky wheel gets the grease." In addition, as customer 2 becomes less sensitive (i.e., β^2 increases), customer 1 benefits because it becomes profitable for the supplier to transfer some of the risk from customer 1 to customer 2. As a result, the near-optimal policy's rewards are much less sensitive to β^2 than are the priority-1 rewards.

4. Approximations

The results of §3 suggest that, for many problems of interest, the greedy policy is inadequate. In this section, we develop a new policy based on approximate dynamic programming. Our approximation is based on a polynomial approximation of the optimal bias function h of problem (D_0). We formulate the problem as primal and dual mathematical programs under this approximation, prove duality, and show how to solve it and generate policies from it. First, however, we discuss why a natural Lagrangian relaxation approach to the problem lacks strength.

4.1. A Lagrangian Approach

Consider our average reward MDP formulation (2), and observe that each customer could be managed independently if not for the coupling constraints $\sum_i x^i(\mathbf{D}) \le \bar{X}$ inherent in $\mathscr{X}(\mathbf{G})$ for all $\mathbf{D}$. We can define a natural Lagrangian relaxation in which we remove the coupling constraints and add terms $\lambda \mathsf{E}_{\mathbf{D}}[(\bar{X} - \sum_i x^i(\mathbf{D}))]$ to the objective function for each $\mathbf{G} \in \mathscr{G}$:

$$h^\lambda(\mathbf{G}) = \lambda \bar{X} + \max_{\mathbf{x}(\mathbf{D})} \mathsf{E}_{\mathbf{D}}\biggl[\sum_i (r^i - \lambda) x^i(\mathbf{D}) - \rho^\lambda + h^\lambda\biggl(\boldsymbol{\beta}\mathbf{G} + \mathbf{u}\biggl(\frac{\mathbf{x}(\mathbf{D})}{\mathbf{y}(\mathbf{G}, \mathbf{D})}\biggr)\biggr)\biggr], \tag{5a}$$

$$\text{s.t.}\quad 0 \le x^i(\mathbf{D}) \le y^i(G^i, D^i) \quad \forall\, \mathbf{D} \in \mathscr{D},\ i \in \mathscr{I}. \tag{5b}$$

We will demonstrate below that this system of equations has a (constant-gain) solution with bounded bias function h^λ, which is sufficient for optimality of ρ^λ in the relaxed problem by Theorems 5.2.2 and 5.2.4 of Hernández-Lerma and Lasserre (1996). It has been shown in related settings that such a Lagrangian relaxation is equivalent to enforcing the linking constraint in expectation rather than almost surely if λ is chosen to minimize the gain upper bound ρ^λ (Whittle 1988, Adelman and Mersereau 2008, Gittins et al. 2011).

The following proposition establishes that problem (5) decomposes by customer. Let $\mathscr{G}^i$ denote the set $\{G^i \in \mathbb{R}_+ \colon u^i(1)/(1-\beta^i) \le G^i \le u^i(1)/(1-\beta^i)\}$.

Proposition 3. *Equations* (5) *are solved by* $h^\lambda(\mathbf{G}) = \sum_i h_i^\lambda(G^i)$ *for all* $\mathbf{G} \in \mathscr{G}$ *and* $\rho^\lambda = \lambda\bar{X} + \sum_i \rho_i^\lambda$, *where* $h_i^\lambda(G^i)$ *and* ρ_i^λ *solve the following for all* $G^i \in \mathscr{G}^i$ *and* $i \in \mathscr{I}$:

$$h_i^\lambda(G^i) = -\rho_i^\lambda + \max_{x^i(D^i)} \mathsf{E}_{D^i}\biggl[(r^i - \lambda)x^i(D^i) + h_i^\lambda\biggl(\beta^i G^i + u^i\biggl(\frac{x^i(D^i)}{y^i(G^i, D^i)}\biggr)\biggr)\biggr], \tag{6a}$$

$$\text{s.t.}\quad 0 \le x^i(D^i) \le y^i(G^i, D^i) \quad \forall\, \mathbf{D} \in \mathscr{D}. \tag{6b}$$

Our problem does not fit the definition of a "weakly coupled" dynamic program (Adelman and Mersereau 2008) because we do not assume demands are independent across customers. However, the decomposition of Proposition 3 relies on a weaker condition, namely, that the distribution of the next period's state for customer i is independent of the current state and action associated with all other customers $j \ne i$.

The following proposition characterizes the decision rule arising from the relaxation.

Proposition 4. *An optimal decision rule* $\{x^i(D^i)\}_{i\in\mathscr{I}}$ *for problem* (6) *is the following*:
1. $x^i(D^i) = y^i(G^i, D^i)$ *for all* i *such that* $r^i \ge \lambda$;
2. $x^i(D^i) = 0$ *for all* i *such that* $r^i < \lambda$.

The decision rule of Proposition 4 is greedy and does not distinguish among customers based on variances or memory parameters. This decision rule may not be feasible in the original problem. One natural, feasible decision rule is an index-based approach (e.g., Whittle 1988), whereby for all i we determine the λ^i for which the supplier is indifferent between shipping zero and positive quantities in problem (6), then allocate capacity greedily to customers in decreasing order of the λ^i's. However, it is clear from Proposition 4 that we will have $\lambda^i = r^i$ for all i, and therefore this index policy will coincide with the greedy policy. An alternate Lagrangian-derived policy solves, for any given $\mathbf{G}$ and $\mathbf{D}$,

$$\max_{\mathbf{x}\in\mathscr{X}(\mathbf{G},\mathbf{D})}\left\{\sum_i r^i x^i + \sum_i h_i^\lambda\left(\beta^i G^i + u^i\left(\frac{x^i}{y^i(G^i,D^i)}\right)\right)\right\}, \quad (7)$$

where $h_i^\lambda()$ is as in the proof of Proposition 4:

$$h_i^\lambda(G^i) = \begin{cases} -(r^i-\lambda)\displaystyle\sum_{t=0}^{\infty}\mathsf{E}\Bigg[y^i\left(\frac{u^i(1)}{1-\beta^i}, D^i\right) \\ \quad - y^i\left((\beta^i)^t G^i + u^i(1)\displaystyle\sum_{s=0}^{t-1}(\beta^i)^s, D^i\right)\Bigg] & \text{if } r^i \ge \lambda, \\ 0 & \text{if } r^i < \lambda. \end{cases} \quad (8)$$

If we assume a linear ordering function $y^i(G^i, D^i) = (1-\beta^i)G^i D^i$, then it is straightforward to show

$$h_i^\lambda(G^i) = -\{r^i-\lambda\}^+\mathsf{E}[D^i]\left\{\frac{u^i(1)}{1-\beta^i} - G^i\right\}.$$

The alternate Lagrangian policy (7) then solves

$$\begin{aligned} \max\ & \left\{\sum_i r^i x^i - \sum_i \{r^i-\lambda\}^+\mathsf{E}[D^i] \right. \\ & \left. \cdot\left[\frac{u^i(1)}{1-\beta^i} - \beta^i G^i - u^i\left(\frac{x^i}{y^i(G^i,D^i)}\right)\right]\right\} \\ \text{s.t. } & \sum_i x^i \le \bar{X} \\ & 0 \le x^i \le y^i(G^i, D^i). \end{aligned}$$

Further assuming $u^i(f) = f$, this problem becomes equivalent to a linear, continuous knapsack problem that is solved by allocating capacity greedily to customers in descending order of the indices

$$r^i + \frac{\mathsf{E}[D^i]\{r^i-\lambda\}^+}{y^i(G^i,D^i)} \quad (9)$$

until capacity is exhausted. This policy prioritizes customers dynamically depending on their goodwills in each period, and we can think of it as adjusting the greedy policy by adding to the customer's margin r^i an expression that approximates the customer's future marginal value. Intuitively, this expression balances goodwills among similar customers; as the goodwill G^i decreases for customer i, she orders a smaller quantity $y^i(G^i, D^i)$, which in turn raises the priority of customer i. Hence, a customer whose goodwill is too low is propped up, at the expense possibly of a customer whose goodwill is currently high. Despite this intuitive appeal, we will see in numerical tests (see §5.3) that this policy performs only slightly better than the greedy policy. We proceed with a more sophisticated policy in the next subsection. In §4.3, we will view the new policy as amending the index policy (9).

4.2. Approximating the Bias Function

Now suppose we are given a collection $\mathscr{K} = \{1, 2, \ldots, K\}$ of basis functions $\{\phi_k(\cdot)\}_{k\in\mathscr{K}}$, where $\phi_k \in \mathbb{B}(\mathscr{G})$. For a weight vector $\mathbf{w} = (w_1, w_2, \ldots, w_K) \in \mathbb{R}^K$, we can approximate the bias function as

$$h(\mathbf{G}) \approx \sum_{k\in\mathscr{K}} w_k\phi_k(\mathbf{G}) \quad \forall\, \mathbf{G}\in\mathscr{G}.$$

Substitute the approximation into (D_0) to obtain the semi-infinite program

$$(\mathrm{D}_\phi) \quad \inf_{\rho,\mathbf{w}} \rho \quad (10a)$$

$$\sum_{k\in\mathscr{K}} w_k\phi_k(\mathbf{G}) \ge \mathsf{E}_{\mathbf{D}}\Bigg[\sum_{i\in\mathscr{I}} r^i x^i(\mathbf{D}) - \rho + \sum_{k\in\mathscr{K}} w_k\phi_k\left(\boldsymbol{\beta}\mathbf{G} + \mathbf{u}\left(\frac{\mathbf{x}(\mathbf{D})}{\mathbf{y}(\mathbf{G},\mathbf{D})}\right)\right)\Bigg] \quad \forall\,(\mathbf{G},\mathbf{x})\in\mathscr{M}, \quad (10b)$$

$$\mathbf{w}\in\mathbb{R}^K,\ \rho\in\mathbb{R}. \quad (10c)$$

Because this h approximation is in $\mathbb{B}(\mathscr{G})$, we have that (D_ϕ) upper bounds (D_0) and therefore upper bounds the average reward of any policy.

The dual of this program is

$$(\mathrm{P}_\phi) \quad \sup_{\mu} \sum_{(\mathbf{G},\mathbf{x})\in M} \mathsf{E}_{\mathbf{D}}\Bigg[\sum_{i\in\mathscr{I}} r^i x^i(\mathbf{D})\Bigg]\mu(\mathbf{G},\mathbf{x}), \quad (11a)$$

$$\sum_{(\mathbf{G},\mathbf{x})\in M}\left[\phi_k(\mathbf{G}) - \mathsf{E}_{\mathbf{D}}\left[\phi_k\left(\boldsymbol{\beta}\mathbf{G} + \mathbf{u}\left(\frac{\mathbf{x}(\mathbf{D})}{\mathbf{y}(\mathbf{G},\mathbf{D})}\right)\right)\right]\right] \cdot\mu(\mathbf{G},\mathbf{x}) = 0 \quad \forall\, k\in\mathscr{K}, \quad (11b)$$

$$\sum_{(\mathbf{G},\mathbf{x})\in M}\mu(\mathbf{G},\mathbf{x}) = 1, \quad (11c)$$

$$\mu \ge 0,\ M = \mathrm{supp}(\mu)\subset\mathscr{M},\ |M| < \infty, \quad (11d)$$

where $\mathrm{supp}(\cdot)$ denotes the support set.

Theorem 2 (Duality). *Suppose $\{\phi_k\}_{k\in\mathscr{K}}$ are continuous functions on $\mathscr{G}$. Then (P_ϕ) is solvable, and* $\sup(P_\phi) = \inf(D_\phi)$.

The solvability of (D_ϕ) is not necessarily guaranteed, but in our numerical work we will only solve (P_ϕ). We will do so using column generation. For a given dual price vector $\mathbf{w}$ on (11b) and price ρ on (11c), to find a new column we solve

$$\max_{(\mathbf{G},\mathbf{x})\in\mathcal{M}} \mathsf{E}_{\mathbf{D}}\Biggl\{\Biggl[\sum_{i\in\mathcal{I}} r^i x^i(\mathbf{D})\Biggr] - \sum_{k\in\mathcal{K}} w_k \cdot \Biggl[\phi_k(\mathbf{G}) - \mathsf{E}_{\mathbf{D}}\Biggl[\boldsymbol{\phi}_k\Biggl(\boldsymbol{\beta}\mathbf{G}+\mathbf{u}\Biggl(\frac{\mathbf{x}(\mathbf{D})}{\mathbf{y}(\mathbf{G},\mathbf{D})}\Biggr)\Biggr)\Biggr]\Biggr] - \rho\Biggr\}.$$

If the optimal objective value is positive, then we add a new variable $\mu(\mathbf{G}^*,\mathbf{x}^*)$ corresponding with an optimal solution $(\mathbf{G}^*,\mathbf{x}^*)$. If the optimal objective value is nonpositive, then the current solution to (P_ϕ) is optimal. When $\mathbf{D}$ is a continuous random variable, this is an optimization over functions $\mathbf{x}(\mathbf{D})$. In all of our numerical results, we consider discrete $\mathbf{D}$s. Even so, this subproblem is generally nonconvex, so it may be difficult to guarantee that the optimal objective value is nonpositive in practice.

A common challenge in implementing approximate dynamic programming is the choice of the basis functions $\{\phi_k\}_{k\in\mathcal{K}}$. In our numerical results in §5, we employ the following polynomial approximation of $h(\mathbf{G})$:

$$h(\mathbf{G}) = \sum_{i\in\mathcal{I}}\sum_{j=1}^{N} w_{ij}\cdot(G^i)^j, \tag{12}$$

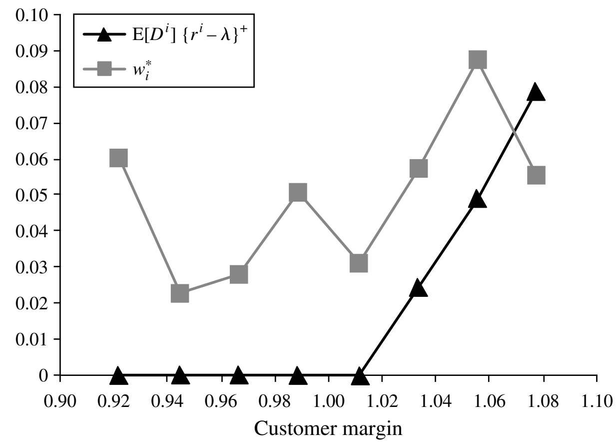
with $N=3$, and where $(G^i)^j$ indicates the goodwill component G^i raised to the jth power. This is a natural approximation architecture. One justification for this choice follows from the bias function under the Lagrangian relaxation, expressed in Equation (8). Assume, as we will in §5, that ordering functions take the form $y^i(G^i,D^i) = D^i z^i(G^i)$, and suppose that we approximate the share functions z^i using polynomials. (By the Weierstrass approximation theorem, we can approximate any share function z^i arbitrarily well using polynomials.) Equation (8) implies that the bias function is a separable (by customer) function of polynomials under the Lagrangian relaxation.

A natural control policy is based on solving the right-hand side of (2) under the bias function approximation implied by an optimal solution to (D_ϕ). Suppose in a given period the demand realization is $\mathbf{D}$ and the customers are in goodwill state $\mathbf{G}$. Given optimal dual variables $\mathbf{w}^*$ obtained from solving (P_ϕ), the supplier solves

$$\begin{aligned}\max_{\mathbf{x}} \quad & \sum_{i\in\mathcal{I}}\Biggl(r^i x^i + \sum_{k\in\mathcal{K}} w_k^*\phi_k\Biggl(\boldsymbol{\beta}\mathbf{G}+\mathbf{u}\Biggl(\frac{\mathbf{x}}{\mathbf{y}(\mathbf{G},\mathbf{D})}\Biggr)\Biggr)\Biggr)\\ \text{s.t.} \quad & \sum_{i\in\mathcal{I}} x^i \le \bar{X}, \quad 0\le x^i\le y^i(G^i,D^i) \quad \forall\, i\in\mathcal{I},\end{aligned} \tag{13}$$

or possibly a multiperiod lookahead version of this problem.

Figure 5 Comparison of Components of the Lagrangian and ADP-Based Index Policies for an Eight-Customer Instance



4.3. Properties of a Linear Approximation

In the special case of an identity utility function $u^i(f)=f$, and a linear basis function $\phi_i(G^i)=G^i$ for each $i\in\mathcal{I}$, the optimization (13) becomes a simple linear program with an index solution. Specifically, assuming $w_i^*\ge 0$ for all i (see below), an optimal solution allocates capacity to customers in decreasing order of the customer indices

$$r^i + \frac{w_i^*}{y^i(G^i,D^i)} \tag{14}$$

until capacity is exhausted.

It is instructive to compare this index policy with the index policy (9) derived from the Lagrangian relaxation in §4.1 for linear y^i. Both approximate future customer value using an expression with order quantity $y^i(G^i,D^i)$ in the denominator, which has the effect of dynamically balancing customer goodwills as discussed after Equation (9). The two indices differ only in the numerators of this expression. Figure 5 compares the two numerators, $\mathsf{E}[D^i]\{r^i-\lambda\}^+$ and w_i^* (computed with a linear bias function approximation), for an eight-customer instance considered in our numerical results in §5.3, where the customers have equally spaced margins between 0.90 and 1.10 and demand scenarios are randomly drawn from a common distribution. We see that the two sets of indices are quite distinct. For this instance, the Lagrangian-derived policy weighs customers' future values according to a monotonic function of their margins, whereas the ADP policy indicates a more nuanced customer evaluation that depends on specific customer demand profiles. We have found the ADP policy to significantly outperform the Lagrangian-derived policy even under a linear approximation.

In spite of this, the ADP policy under the linear basis function approximation is limited in its ability to capture certain problem features. The following proposition implies that this policy does not hold back capacity.

Proposition 5. *When the linear bias function approximation $h(\mathbf{G}) \approx \sum_i w_i G^i$ is used, we can impose the constraints $w_i \geq 0$ for all i on (D_ϕ) without loss of optimality.*

Hence, the objective of problem (13) is an increasing function of x_i for all i; that is, a solution to (13) will fulfill orders until either capacity is exhausted or all orders have been fulfilled. Recall from §3.4 that holding back capacity can be an important lever for the supplier in some circumstances. For this reason, in our numerical results we use adjustments to the linear approximation policy that facilitate holding back capacity, including a higher-order bias function approximation and adding lookahead to problem (13).

Furthermore, we show that for a large set of problem instances, the w_i^* values and hence the decisions produced by (13) are invariant to the β^i parameters when a linear approximation is used. Denote problem (D_ϕ) under the linear approximation as $(D_\phi\text{–}\boldsymbol{\beta})$ to signify its dependence on $\boldsymbol{\beta}$.

Proposition 6. *Assume the linear bias function approximation is used and that*

1. *$u^i()$ does not depend on β^i, and*
2. *$y^i(G^i, D^i) = D^i z^i(G^i) = D^i \hat{z}^i(G^i(1-\beta^i)/u^i(1))$ for some $\hat{z}^i()$ that does not depend on β^i.*

Then an optimal solution $(\rho^*, \mathbf{w}^*)$ to $(D_\phi\text{–}\boldsymbol{\beta})$ is optimal in $(D_\phi\text{–}\bar{\boldsymbol{\beta}})$ for any $\bar{\boldsymbol{\beta}}$.

Proof. Under the proposition assumptions, the constraints (10b) of (D_ϕ) are

$$\sum_i w_i G^i(1-\beta^i) \geq \mathsf{E}_{\mathbf{D}}\left[\sum_i r^i x^i(\mathbf{D}) - \rho + \sum_i w_i u^i\left(\frac{x^i(\mathbf{D})}{D^i \hat{z}^i(G^i(1-\beta^i)/u^i(1))}\right)\right] \quad \forall (\mathbf{G}, \mathbf{x}) \in \mathcal{M}(\boldsymbol{\beta}), \quad (15)$$

where we denote the set $\mathcal{M}$ by $\mathcal{M}(\boldsymbol{\beta})$ to emphasize its dependence on $\boldsymbol{\beta}$. We can write $\mathcal{M}(\boldsymbol{\beta})$ as

$$\mathcal{M}(\boldsymbol{\beta}) = \Bigg\{(\mathbf{G}, \mathbf{x})\colon \mathbf{G} \in \mathbb{R}_+^n, \mathbf{x} \in \mathbb{B}(\mathcal{D}), u^i(0) \leq G^i(1-\beta^i) \leq u^i(1)\ \forall i, \sum_{i\in\mathcal{I}} x^i(\mathbf{D}) \leq \bar{X} \quad \forall \mathbf{D} \in \mathcal{D}, 0 \leq x^i(\mathbf{D}) \leq D^i \hat{z}^i\left(\frac{G^i(1-\beta^i)}{u^i(1)}\right) \quad \forall i \in \mathcal{I} \text{ and } \forall \mathbf{D} \in \mathcal{D}\Bigg\}.$$

Observe that G^i and β^i only appear in the constraints (15) and in the definition of the set $\mathcal{M}(\boldsymbol{\beta})$ through the term $G^i(1-\beta^i)$. Therefore, $(\mathbf{G}, \mathbf{x}) \in \mathcal{M}(\boldsymbol{\beta})$ is equivalent to $(\bar{\mathbf{G}}, \mathbf{x}) \in \mathcal{M}(\bar{\boldsymbol{\beta}})$, where $\bar{\mathbf{G}}$ is defined by $\bar{G}^i = G^i(1-\beta^i)/(1-\bar{\beta}^i)$, and the constraint $(\mathbf{G}, \mathbf{x})$ of $(D_\phi\text{–}\boldsymbol{\beta})$ is identical to the constraint $(\bar{\mathbf{G}}, \mathbf{x})$ of $(D_\phi\text{–}\bar{\boldsymbol{\beta}})$. Therefore, $(D_\phi\text{–}\boldsymbol{\beta})$ is numerically identical to $(D_\phi\text{–}\bar{\boldsymbol{\beta}})$, implying the desired result. □

Given this result, it is straightforward to show that the policy's allocation $\mathbf{x}$ in goodwill state $\mathbf{G}$ under $\boldsymbol{\beta}$ is the same as its allocation in goodwill state $\bar{\mathbf{G}} = \mathbf{G}(1-\boldsymbol{\beta})/(1-\bar{\boldsymbol{\beta}})$ under $\bar{\boldsymbol{\beta}}$. Goodwill $\mathbf{G}$ under $\boldsymbol{\beta}$ corresponds naturally with $\bar{\mathbf{G}}$ under $\bar{\boldsymbol{\beta}}$ in that both map to the same normalized goodwill $\mathbf{W} = (1-\boldsymbol{\beta})\mathbf{G} = (1-\bar{\boldsymbol{\beta}})\bar{\mathbf{G}}$; that is, both $\mathbf{G}$ and $\bar{\mathbf{G}}$ represent the same average fill rate and the same ratio of goodwill to maximum achievable goodwill. Thus, Proposition 6 implies that, under a linear approximation, a customer's allocation in a period is determined by her normalized goodwill, without regard to $\boldsymbol{\beta}$.

On the other hand, under a particular sample path, we will not generally have $\bar{\mathbf{G}}_t = \mathbf{G}_t(1-\boldsymbol{\beta})/(1-\bar{\boldsymbol{\beta}})$. Thus, the allocations to customers will be $\boldsymbol{\beta}$-dependent over time under the linear approximation because of the differing problem dynamics induced by different $\boldsymbol{\beta}$s.

5. Numerical Study

This section explores the problem and proposed ADP policy through a numerical study. Our goals here are to demonstrate the quality of the proposed policy versus the greedy policy, to reinforce some of our earlier analytical insights, and to make some new observations.

We assume an identity utility function $u^i(f) = f$ throughout, and we present results assuming y^i takes the form $y^i(G^i, D^i) = D^i z^i(G^i) = (1-\beta^i)G^i D^i$. (We add a small constant to each order to avoid division by zero.) For the purposes of this section, we do not solve (D_ϕ) directly. Instead, we solve (P_ϕ) using column generation as described in §4.2. We refer to the solution to this problem as the "LP solution." We implement the column generation and policy simulations in AMPL, and we use KNITRO version 6.0 as our linear and nonlinear programming solver. As we pointed out in §4.2, the optimal solution to (D_ϕ) upper bounds the achievable average rewards under all policies. However, we terminate the column generation approach when the reduced profits relative to the objective function fall below a prespecified small tolerance. Therefore, our solutions are not provably feasible in (D_ϕ) and are not guaranteed to bound the optimal average reward. In contrast to the policy analyzed in §4.3, we use a third-order approximation ($N = 3$) and add one step of lookahead to the policy problem (13). For the instances we consider, we have found the third-order approximation to give significantly tighter (approximate) bounds but indistinguishable policy performances compared with a linear approximation. Lookahead stabilizes and noticeably improves the policy in our examples.

Figure 6 Simulated Policy Rewards and Customer B Allocations as r^B Varies—$r^A = 1$ and the Two Customers Are Identical Except for Their Margins

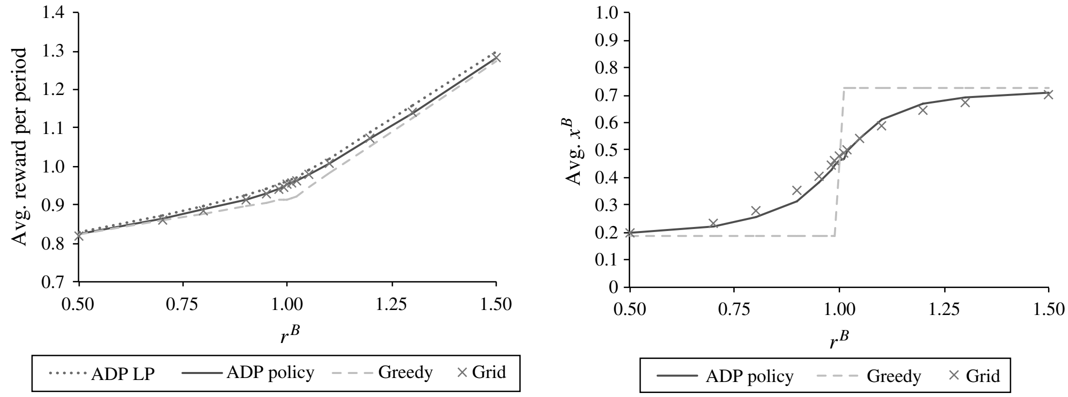
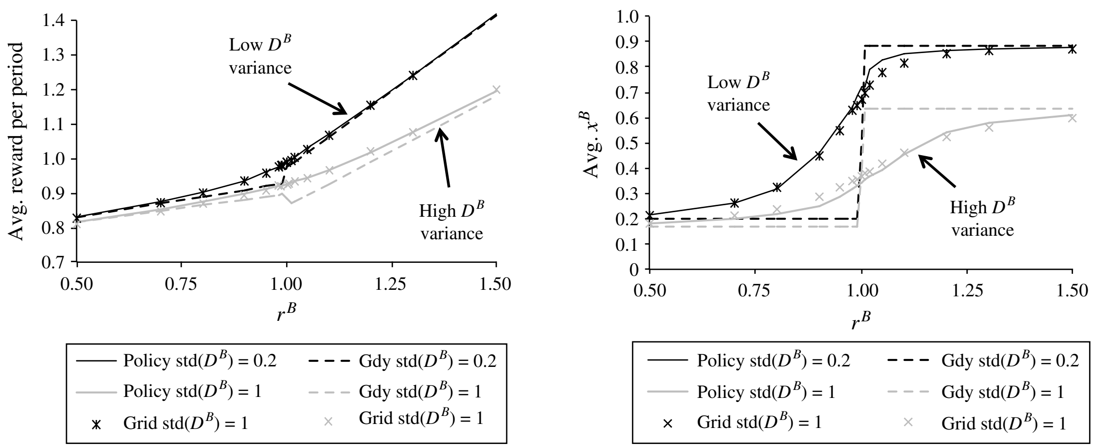


Figure 7 Simulated Policy Rewards and Customer B Allocations as r^B Varies—Cases Are Shown for High and Low Customer B Demand Variance



We start in §5.1 with results on two-customer instances. Section 5.2 builds some intuition into the role of $\boldsymbol{\beta}$. Section 5.3 presents and interprets results on larger examples.

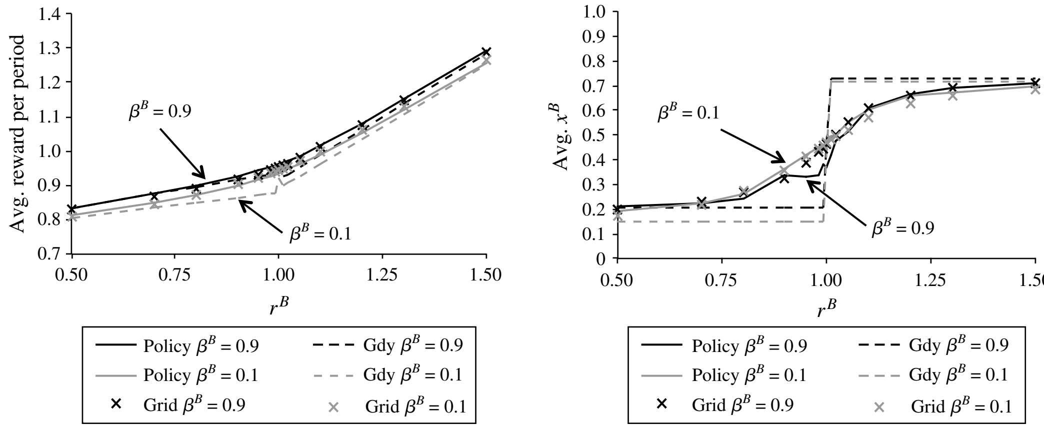
5.1. Two-Customer Examples

Consider an example with two customers, A and B, who differ only in their margins. We construct such an example by generating 20 equiprobable demand scenarios in which the D^i values are drawn independently from a lognormal distribution with mean one and standard deviation 0.6.[4] We take $\beta^A = \beta^B = 0.5$, $\bar{X} = 1.0$, z^A and z^B to be linear share functions.

Figure 6 gives, for the ADP policy, greedy policy, and grid-based policy used in §3.5, the average problem reward and the average allocation x^B to customer B while r^B is varied and $r^A = 1$. We see that the ADP policy performance is quite close to that predicted by the LP (D_ϕ) and nearly identical to that of the grid-based policy. Not surprisingly, problem rewards and customer B shipments are both increasing in r^B under both policies. Both policies have a "kink" in rewards at $r^B = 1$. To the right of this point, customer B receives a majority of the shipments, and thus her margin has a greater impact on the problem reward. Whereas the behavior of the greedy policy depends only on the relative ranking of the two margins, the more sophisticated policies more evenly allocate their capacity to the two customers, particularly when the two margins are relatively close. This is consistent with the "portfolio effect" we identified in §3.5. Accordingly, when the two margins are close, the approximate policy exhibits its greatest reward advantage over the greedy approach. In Figure 7, we show results for problem instances in which customers differ in the variances of their demand random variables: the generative lognormal distribution for customer A's demand scenarios has mean 1 and standard devation 0.6, whereas the generative lognormal distribution for customer B has mean 1 and

[4] In fact, we generate 10 demand scenarios in such a way. The remaining 10 scenarios are the same as the first 10 except that each D^i is assigned to the opposite customer. In this way, we make the two customers' demands statistically identical.

Figure 8 Simulated Policy Rewards and Customer B Allocations as r^B Varies—Cases Are Shown for High and Low Customer B Memory



standard deviation either 0.2 or 1.0. It is intuitive that low demand variance would be advantageous to the supplier; indeed we see that the supplier's rewards are highest when customer B's demand variance is lowest. We see that shipments made to customer B depend inversely on her demand variance under all three policies. Interestingly, when customer B has low variance and r^B is just below $r^A = 1$, Customer B receives larger orders on average (and experiences higher goodwill) than customer A under the ADP policy. (For $r^B = 0.99$, average $x^A = 0.31$ and average $x^B = 0.68$.) Thus, we see that lower customer volatility can substitute for higher margin under a good policy.

Finally, Figure 8 illustrates cases in which the customers have statistically identical (but independent) D^is, but the customers differ in their memory parameters β^i. We observe that β^B has its greatest impact when $r^B < r^A$. In these cases, the majority of the supplier's capacity goes to customer A, and the amount of leftover capacity available for customer B is variable. The supplier benefits here when customer B is more loyal and thus less sensitive to this variability. Interestingly, the rewards of the ADP and grid-based policies seem to be less impacted by the change in β^B than the rewards of the greedy policy. As in §3.5, we see here that differing customer loyalties can be somewhat neutralized with a sufficiently intelligent policy. We further explore the impact of customer memory parameters in the following subsection.

5.2. The Impact of Customer Memory

In §3.1 we argued that although longer memories tend to dampen customer order variances, longer memories can have a positive or negative impact on a customer's expected orders. In this section, we work in a simplified context to further understand the impact of $\boldsymbol{\beta}$. Specifically, we numerically illustrate on a single-customer instance that high β can either be good or bad for the supplier.

Take a single customer with margin $r = 1$ and ordering function $y(W, D) = WD$, and take the supplier's capacity to be 0.5. The random demand D takes values 0.5 and 1.0 with equal probability, yielding a demand standard deviation of 0.25. Figure 9 gives the simulated average shipments (which correspond with the average rewards because $r = 1$) under the ADP policy as a function of β for this problem instance along with a "low variance" instance where the two demand scenarios are 0.625 and 0.875 for a demand standard devation of 0.125. Not surprisingly, increasing demand variance leads to lower overall shipments and lower overall rewards. Interestingly, the shipments are increasing in β (albeit slightly) under low demand variance and decreasing in β under high demand variance. In this particular instance, we have empirically observed that the ADP policy yields results indistinguishable from both the greedy and grid-based policies.

Figure 9 Average Shipments to a Single Customer Under Two Variance Scenarios and for Several Choices of β

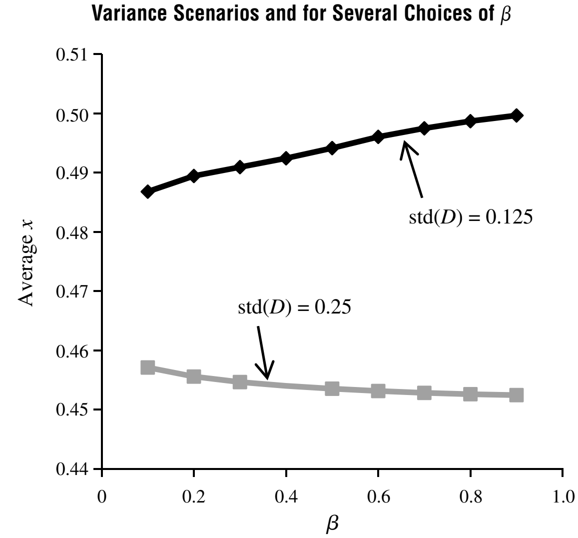
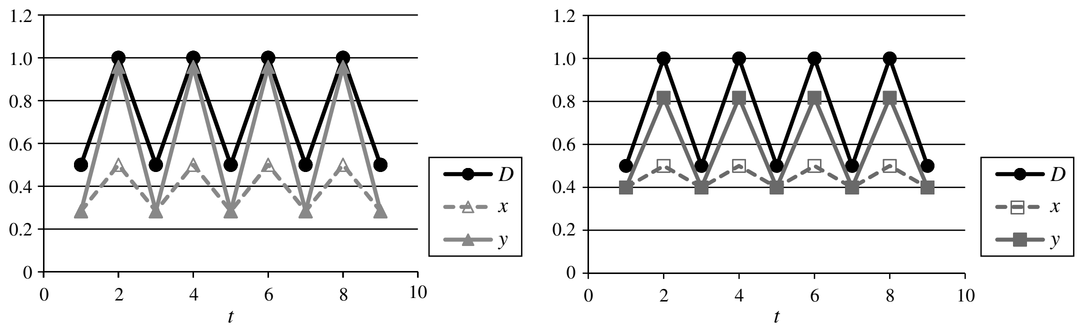


Figure 10 Orders and Shipments to a Single Customer over Time with $\beta = 0.1$ (Left) and $\beta = 0.9$ (Right) Under a Sample Path with Alternating Demands



To illuminate the interaction between β and variance, consider a particular sample path under which D simply alternates (deterministically) between 0.5 and 1.0. Figure 10 shows this sample path and the steady-state orders and shipments under it for two different choices of β. Notice that in even-numbered periods, $D = 1$ and the customer places large orders that significantly exceed the supplier's capacity. In odd-numbered periods, however, $D = 0.5$, and the supplier can fully fulfill the customer orders. The customer therefore receives fill rates that alternate between good and poor. Because of her shorter memory, the low-β customer in the left-hand plot experiences large goodwill fluctuations that lead to a more variable order stream. Because of his limited capacity, the supplier is capped in his ability to satisfy big orders. The net impact is that the *average* shipment on this sample path decreases as β decreases. We see that a high β can be beneficial to the supplier inasmuch as it can smooth the customer's order stream and therefore increase average shipment sizes.

On the other hand, consider the sample path of Figure 11, where D alternates between four-period sequences of constant demands. In both plots of Figure 11, when D is high, for example, from $t = 5$–8, the supplier sells out of his capacity but nevertheless falls well short of the customer orders. This damages the customer goodwill for periods 9–12. The low-β customer's goodwill, and hence its orders, recover quickly. However, the high-β customer still orders less than capacity through period 12, effectively "holding a grudge." In short, a small β is advantageous here because the customer recovers quickly from low goodwill states.

Our formulation in §2 assumes that demands are independent over time, in which case a realistic sample path will contain demand subsequences like the ones considered as well as others. These results show that β can have impact, but they also demonstrate that its impact is complex. Although we have observed in other numerical contexts (e.g., §3.5) that higher customer β usually leads to higher average rewards for the supplier, the example here shows that this need not always be the case. This reinforces a similar insight generated in §3.1.

5.3. Larger Examples

We report on results on larger examples to better understand the performance of the ADP policy relative to the greedy policy and, by extension, the impact of memory effects on supplier rewards.

Figure 11 Orders and Shipments to a Single Customer with $\beta = 0.1$ (Left) and $\beta = 0.9$ (Right) Under a Sample Path with Alternating Length-4 Sequences of Constant Demands

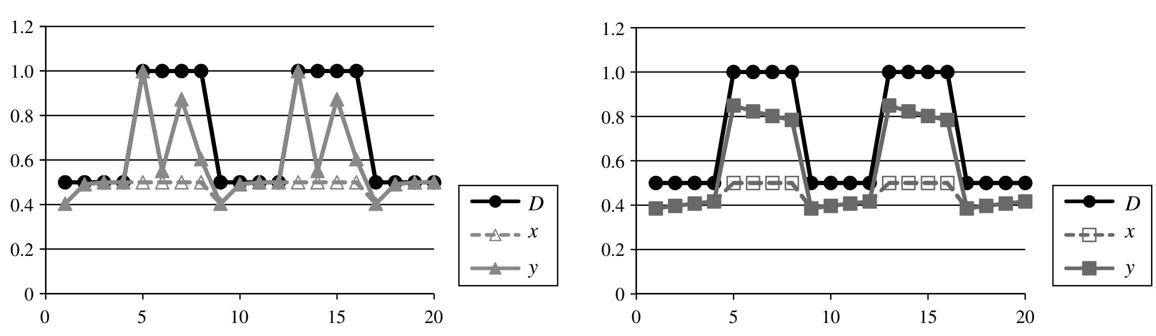
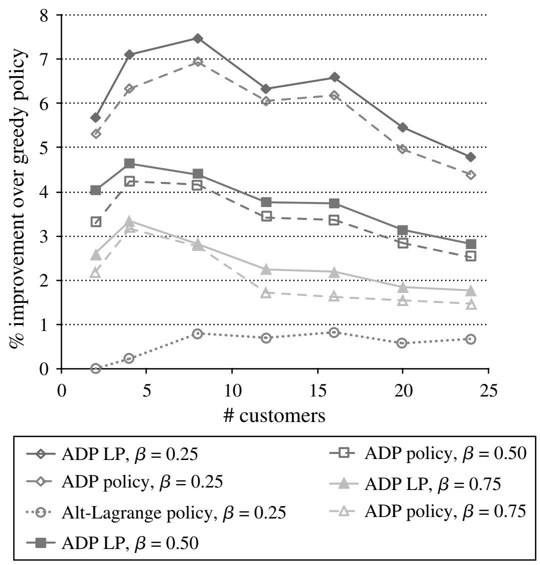


Figure 12 LP Solutions, ADP Policy Rewards, and Selected Alternate Lagrangian Policy Rewards for the Problem Instances of §5.3, Expressed in Terms of Percent Improvement over Greedy Policy Rewards



We consider families of instances indexed by the number of customers n. All customers in a problem instance share a common $\beta = \beta^i$, $i = 1, \ldots, n$, either 0.25, 0.50, or 0.75. Margins r^i, $i = 1, \ldots, n$ are chosen equally spaced on the range $[0.9, 1.1]$ according to the formula $0.9 + 0.2i/(n+1)$. (For example, for $n = 4$ the customer margins are $0.94, 0.98, 1.02$, and 1.06.) Uncertainty is modeled through 30 scenarios. On each scenario and for each customer i, the realization of D^i is an independent draw from a lognormal distribution with mean 1 and standard deviation 0.9. The supplier capacity is taken to be $\bar{X} = n/2$. We simulate both the greedy and ADP policies over 1,000 periods for five sets of randomly generated uncertainty scenarios. We present in Figure 12 the simulated average rewards resulting from the ADP policy with lookahead.[5] We measure the average rewards of the ADP policy relative to the average rewards obtained by the greedy policy. For comparison, we also include in the figure the expected rewards implied by our final solution to the LP (D_ϕ). For $\beta = 0.25$, we show the average rewards achieved by the alternate Lagrangian policy (9).

We observe that the ADP policy achieves average rewards quite close to what our solutions to (D_ϕ) predict. We also see that the ADP policy achieves significantly higher rewards than the greedy policy, especially for moderate numbers of customers. The gap nears 7% for $n = 8$ customers when $\beta = 0.25$. We observe that the gap between the two policies narrows as n gets larger, as predicted by Theorem 1. Finally, we see that the advantage of the ADP policy over the greedy policy is most pronounced when $\beta = 0.25$. We note that this is driven by sensitivity of the greedy policy rewards to β; we find the ADP policy rewards to be relatively insensitive to β in these examples. We also find that the alternate Lagrangian policy (9) slightly outperforms the greedy policy but substantially underperforms the ADP policy.

To gain insight into the customer portfolios cultivated by the ADP and greedy policies, we examine the simulated average goodwills under both policies for eight-customer instances with $\beta = 0.50$ and $\bar{X} = 4$. For this purpose, we generate D^i scenarios using three correlation matrices: the identity matrix and the two correlation matrices illustrated in Figure 13. (We generate multivariate normal variates with the specified correlation matrices, and then we transform the variates to lognormal variates by taking their exponentials. The effective underlying lognormal distributions have mean 1 and standard deviation 0.9 as before.) Figure 14 plots simulated average goodwills of each of the eight customers, averaged over five randomly generated instances for each correlation matrix.

We observe first that the ADP policy produces a more balanced customer porfolio than the greedy policy under all three correlation structures. For the independent case in particular, the smallest margin customer receives average goodwill 1.16 under the ADP policy versus 0.63 under the greedy policy. Intuitively, the ADP policy reduces aggregate demand variance by diversifying its capacity investment among customers. Whereas the different correlation structures have only a modest impact on the customer portfolios generated by the greedy policy, the impact of correlation structure on the ADP policy portfolios is pronounced. Under correlation matrix A, the ADP policy effectively diversifies its portfolio among three correlated customer groups, favoring the customers with the fourth- and sixth-highest margins in particular relative to the independent case. Under correlation matrix B, the ADP policy favors the two lowest-margin but negatively-correlated customers relative to customers with higher margins and positive demand correlations. In summary, we observe the ADP policy managing complex customer demand characteristics to trade-off margin and demand variance in a fashion akin to a financial portfolio manager trading off risk and return.

[5] We also solved some instances without lookahead. For the instances in Figure 12 with $\beta = 0.25$, we found that the performance without lookahead was on average 0.73% worse than with lookahead but still significantly better than the greedy policy.

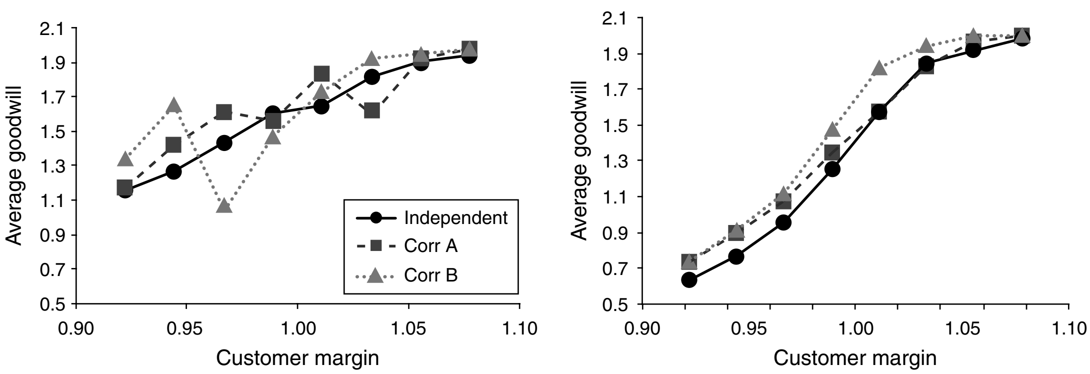
Figure 13 Correlation Matrices A (Left) and B (Right) Used to Generate the Instances Underlying Figure 14

r^i	0.92	0.94	0.97	0.99	1.01	1.03	1.06	1.08
0.92	1	3/4	3/4					
0.94	3/4	1	3/4					
0.97	3/4	3/4	1					
0.99				1	3/4			
1.01				3/4	1			
1.03						1	3/4	3/4
1.06						3/4	1	3/4
1.08						3/4	3/4	1

r^i	0.92	0.94	0.97	0.99	1.01	1.03	1.06	1.08
0.92	1	– 3/4						
0.94	–3/4	1						
0.97			1	3/4	3/4	3/4	3/4	3/4
0.99			3/4	1	3/4	3/4	3/4	3/4
1.01			3/4	3/4	1	3/4	3/4	3/4
1.03			3/4	3/4	3/4	1	3/4	3/4
1.06			3/4	3/4	3/4	3/4	1	3/4
1.08			3/4	3/4	3/4	3/4	3/4	1

Note. Blank cells indicate zero correlations.

Figure 14 Average Goodwills Observed Under the ADP (Left) and Greedy (Right) Policies for Eight-Customer Examples with Various Correlation Structures



6. Concluding Remarks

Customer memories bring a complex but consequential set of trade-offs to a firm's management of its customer portfolio, particularly when the number of customers is small to moderate, when there is significant volatility in customer demands, and when customers have relatively short-term memories of past service. Our approximate policy consistently dominates the greedy policy by managing customer expectations, pooling customer demand volatilities, and actively managing the system dynamics induced by customer memories. In all, our paper provides new insights into when goodwill matters, the mechanisms that link goodwill to firm profits, and how to manage it. We believe our work paves the way for further exploration of goodwill dynamics in other relationship contexts.

Acknowledgments

Daniel Adelman thanks the University of Chicago Booth School of Business, and Adam J. Mersereau thanks the University of North Carolina at Chapel Hill Kenan-Flagler Business School and the W. Benton Pipkin Faculty Development Fund for financial support of this research. The authors thank seminar participants at Cornell University, Harvard University, McGill University, New York University, University of Michigan, University of Southern California, University of Chicago, and Duke-UNC Operations Management Workshop for constructive feedback. The authors also thank Pradeep Chintagunta and Reid Hastie for useful discussions during the project, Rod Parker for helpful pointers to the literature, and a review team consisting of editor Gérard Cachon, department editors Preyas Desai and Martin Lariviere, one anonymous associate editor, and four anonymous reviewers for high-quality, constructive reviews. The authors accept sole responsibility for any errors or omissions.

Appendix. Proofs

Proof of Theorem 1. We model customer memory using "normalized goodwill" as described at the end of §2.1. The normalized goodwill $u(0) \le W_t^i \le u(1)$ of customer i in period t is defined using the recursion $W_{t+1}^i = \beta^i W_t^i + (1-\beta^i)u(x_t^i/y(W_t^i, D_t^i))$, where we abuse notation slightly by expressing y as a function of W_t^i rather than G_t^i.

We assume the system is started from a state sampled from the steady-state distribution under the greedy policy, and therefore that the system is stationary (i.e., the system state at time t is ex ante independent of t). Let random variables $W_{(n)}^i$ and $\mathrm{FR}_{(n)}^i$ indicate the steady-state normalized goodwill of customer i and the steady-state fill rate provided to customer i, respectively, in problem n under the greedy policy. The fraction $R = \min\{1, \bar{x}/\mathsf{E}[y(u(1), D)]\}$ is constant across problems and represents the proportion of demand that can be fully satisfied if all customers have maximal goodwill and place their expected order quantities.

The outline of our argument is as follows. First, we show that as the problem is scaled up, the top Rn customers tend to 100% fill rate. We then show that, in the limit, these customers bring revenues approaching those of a relaxed version of the problem in which the decision maker adheres to a capacity constraint that holds in expectation. We proceed with the following lemmas.

Lemma 1. *For any* $0 < \alpha < R$, $\lim_{n\to\infty} \Pr\{FR_{(n)}^{\lfloor \alpha n\rfloor} = 1\} = 1$.

Proof. Fix α.

$$\begin{aligned}
&\Pr\{FR_{(n)}^{\lfloor \alpha n\rfloor} = 1\}\\
&= \Pr\left\{n\bar{x} \geq \sum_{i=1}^{\lfloor \alpha n\rfloor} y(W_{(n)}^i, D^i)\right\} \geq \Pr\left\{n\bar{x} \geq \sum_{i=1}^{\lfloor \alpha n\rfloor} y(u(1), D^i)\right\}\\
&= 1 - \Pr\left\{\sum_{i=1}^{\lfloor \alpha n\rfloor} y(u(1), D^i) > n\bar{x}\right\}\\
&\geq 1 - \Pr\left\{\sum_{i=1}^{\lfloor \alpha n\rfloor} y(u(1), D^i) > Rn\mathsf{E}[y(u(1), D)]\right\}\\
&= 1 - \Pr\left\{\sum_{i=1}^{\lfloor \alpha n\rfloor} y(u(1), D^i) - \lfloor \alpha n\rfloor \mathsf{E}[y(u(1), D)]\right.\\
&\qquad\qquad \left. > (Rn - \lfloor \alpha n\rfloor)\mathsf{E}[y(u(1), D)]\right\}\\
&\geq 1 - \Pr\left\{\left|\sum_{i=1}^{\lfloor \alpha n\rfloor} y(u(1), D^i) - \lfloor \alpha n\rfloor \mathsf{E}[y(u(1), D)]\right|\right.\\
&\qquad\qquad \left. > (Rn - \lfloor \alpha n\rfloor)\mathsf{E}[y(u(1), D)]\right\}\\
&\geq 1 - \frac{\lfloor \alpha n\rfloor \operatorname{Var}[y(u(1), D)]}{\{(Rn - \lfloor \alpha n\rfloor)\mathsf{E}[y(u(1), D)]\}^2} \xrightarrow{n\to\infty} 1.
\end{aligned}$$

The last inequality is an application of Chebyshev's inequality. □

Corollary 1. *For any* $\alpha < R$,
(a) $p\lim_{n\to\infty} FR_{(n)}^{\lfloor \alpha n\rfloor} = 1$,
(b) $p\lim_{n\to\infty} W_{(n)}^{\lfloor \alpha n\rfloor} = u(1)$.

Proof. Part (a) follows directly from Lemma 1 because $\Pr\{1 - FR_{(n)}^{\lfloor \alpha n\rfloor} > 0\} = 1 - \Pr\{FR_{(n)}^{\lfloor \alpha n\rfloor} = 1\}$.

To prove part (b), consider the vector of past fill rates $(FR_{-1}, FR_{-2}, FR_{-3}, \ldots,)$, extending backward in time and generating the normalized goodwill $W_{(n)}^{\lfloor \alpha n\rfloor}$; that is, express $W_{(n)}^{\lfloor \alpha n\rfloor} = \sum_{s=0}^{+\infty} \beta^s(1-\beta)u(FR_{-s})$. Fix an $\epsilon > 0$, and let $S(\epsilon) = \lceil \ln(\epsilon/u(1))/(\ln \beta)\rceil - 1$, which implies that

$$\sum_{s=0}^{S(\epsilon)} \beta^s(1-\beta)u(1) \geq u(1) - \epsilon.$$

Then we can write

$$\begin{aligned}
&\Pr\{u(1) - W_{(n)}^{\lfloor \alpha n\rfloor} \geq \epsilon\}\\
&= \Pr\left\{\sum_{s=0}^{+\infty} \beta^s(1-\beta)u(FR_{-s}) \leq u(1) - \epsilon\right\}\\
&\leq \Pr\left\{\sum_{s=0}^{S(\epsilon)} \beta^s(1-\beta)u(FR_{-s}) \leq u(1) - \epsilon\right\}\\
&\leq \Pr\{FR_{-s} < 1 \text{ for at least one } s \in [0, S(\epsilon)]\}\\
&\leq [S(\epsilon)+1]\cdot \Pr\{FR_{(n)}^{\lfloor \alpha n\rfloor} < 1\} \xrightarrow{n\to\infty} 0.
\end{aligned}$$

The final inequality holds because $\Pr\{FR_{-0} < 1$ or $FR_{-1} < 1$ or, $\ldots$, or $FR_{-S(\epsilon)} < 1\} \leq \sum_{s=0}^{S(\epsilon)} \Pr\{FR_{-s} < 1\}$ by Boole's inequality and by the stationarity of the system, which implies that the FR_{-s} random variables are identically distributed. □

Define $\rho_n^{\mathcal{G}}$ as the average reward per customer under the greedy policy in problem n. Define $\rho_n^{\mathcal{R}}$ as the optimal average reward per customer in a relaxed problem in which the decision maker must make shipment decisions so that he satisfies the capacity constraint in expectation: $\mathsf{E}[\sum_{i=1}^n x_t^i] \leq n\bar{x}$. It is straightforward to see that no feasible policy in the relaxed problem can do better than fully satisfying the top $\lceil Rn\rceil$ customers in each time period, i.e., $\rho_n^{\mathcal{R}} \leq (1/n)\sum_{i=1}^{\lceil Rn\rceil} \mathsf{E}[r^i y(u(1), D^i)]$.

Lemma 2. $\lim_{n\to\infty}(\rho_n^{\mathcal{R}} - \rho_n^{\mathcal{G}}) = 0$.

Proof. Fix n and a fraction $0 < \eta < R$. When the DM has sufficient capacity to satisfy the top $\lfloor (R-\eta)n\rfloor$ customers he will receive revenue at least $\sum_{i=1}^{\lfloor (R-\eta)n\rfloor} r^i y^i(W_{(n)}^i, D^i)$ under the greedy policy. Therefore,

$$\begin{aligned}
n\rho_n^{\mathcal{G}} \geq\, &\mathsf{E}\left[\sum_{i=1}^{\lfloor (R-\eta)n\rfloor} r^i y^i(W_{(n)}^i, D^i) \,\middle|\, \sum_{i=1}^{\lfloor (R-\eta)n\rfloor} y^i(W_{(n)}^i, D^i) \leq n\bar{x}\right]\\
&\cdot \Pr\left(\sum_{i=1}^{\lfloor (R-\eta)n\rfloor} y^i(W_{(n)}^i, D^i) \leq n\bar{x}\right).
\end{aligned}$$

We can bound the conditional expectation by writing

$$\begin{aligned}
&\mathsf{E}\left[\sum_{i=1}^{\lfloor (R-\eta)n\rfloor} r^i y^i(W_{(n)}^i, D^i)\right]\\
&= \mathsf{E}\left[\sum_{i=1}^{\lfloor (R-\eta)n\rfloor} r^i y^i(W_{(n)}^i, D^i) \,\middle|\, \sum_{i=1}^{\lfloor (R-\eta)n\rfloor} y^i(W_{(n)}^i, D^i) \leq n\bar{x}\right]\\
&\quad \cdot \Pr\left(\sum_{i=1}^{\lfloor (R-\eta)n\rfloor} y^i(W_{(n)}^i, D^i) \leq n\bar{x}\right)\\
&\quad + \mathsf{E}\left[\sum_{i=1}^{\lfloor (R-\eta)n\rfloor} r^i y^i(W_{(n)}^i, D^i) \,\middle|\, \sum_{i=1}^{\lfloor (R-\eta)n\rfloor} y^i(W_{(n)}^i, D^i) > n\bar{x}\right]\\
&\quad \cdot \Pr\left(\sum_{i=1}^{\lfloor (R-\eta)n\rfloor} y^i(W_{(n)}^i, D^i) > n\bar{x}\right)\\
&\leq \mathsf{E}\left[\sum_{i=1}^{\lfloor (R-\eta)n\rfloor} r^i y^i(W_{(n)}^i, D^i) \,\middle|\, \sum_{i=1}^{\lfloor (R-\eta)n\rfloor} y^i(W_{(n)}^i, D^i) \leq n\bar{x}\right]\\
&\quad \cdot \Pr\left(\sum_{i=1}^{\lfloor (R-\eta)n\rfloor} y^i(W_{(n)}^i, D^i) \leq n\bar{x}\right)\\
&\quad + \Pr\left(\sum_{i=1}^{\lfloor (R-\eta)n\rfloor} y^i(W_{(n)}^i, D^i) > n\bar{x}\right) \sum_{i=1}^{\lfloor (R-\eta)n\rfloor} r_{\max} y_{\max}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
n\rho_n^{\mathcal{G}} \geq\, &\mathsf{E}\left[\sum_{i=1}^{\lfloor (R-\eta)n\rfloor} r^i y^i(W_{(n)}^i, D^i)\right]\\
&- \Pr\left(\sum_{i=1}^{\lfloor (R-\eta)n\rfloor} y^i(W_{(n)}^i, D^i) > n\bar{x}\right) \sum_{i=1}^{\lfloor (R-\eta)n\rfloor} r_{\max} y_{\max}
\end{aligned}$$

for any n. We know from Lemma 1 that

$$\Pr\left(\sum_{i=1}^{\lfloor (R-\eta)n\rfloor} y^i(W^i_{(n)}, D^i) > n\bar{x}\right) \xrightarrow{n\to\infty} 0;$$

therefore, for any $\epsilon > 0$ there is a $N(\eta,\epsilon)$ such that the following inequality is valid for $n > N(\eta,\epsilon)$:

$$\begin{aligned} n\rho_n^{\mathcal{G}} &\geq \mathsf{E}\left[\sum_{i=1}^{\lfloor (R-\eta)n\rfloor} r^i y^i(W^i_{(n)}, D^i)\right] - \epsilon \sum_{i=1}^{\lfloor (R-\eta)n\rfloor} r_{\max} y_{\max} \\ &\geq \mathsf{E}\left[\sum_{i=1}^{\lfloor (R-\eta)n\rfloor} r^i y^i(W^{\lfloor (R-\eta)n\rfloor}_{(n)}, D^i)\right] - \epsilon \sum_{i=1}^{\lfloor (R-\eta)n\rfloor} r_{\max} y_{\max}. \end{aligned}$$

The second inequality follows because the greedy policy always gives priority to customers with higher margin (and equivalently, lower index)—implying $W^{i+1}_{(n)} \leq_{st} W^i_{(n)}$ for all i—and because y is nondecreasing.

For $n > N(\eta,\epsilon)$, we can therefore bound the cost difference between the relaxed and greedy policies as

$$\begin{aligned} \rho_n^{\mathcal{R}} - \rho_n^{\mathcal{G}} &\leq \frac{1}{n}\sum_{i=1}^{\lceil Rn\rceil} \mathsf{E}[r^i y^i(u(1), D^i)] \\ &\quad - \frac{1}{n}\mathsf{E}\left[\sum_{i=1}^{\lfloor (R-\eta)n\rfloor} r^i y^i(W^{\lfloor (R-\eta)n\rfloor}_{(n)}, D^i)\right] \frac{\epsilon}{n} \sum_{i=1}^{\lfloor (R-\eta)n\rfloor} r_{\max} y_{\max} \\ &\leq \frac{1}{n}\mathsf{E}\left[\sum_{i=1}^{\lfloor (R-\eta)n\rfloor} r^i \{y^i(u(1), D^i) - y^i(W^{\lfloor (R-\eta)n\rfloor}_{(n)}, D^i)\}\right] \\ &\quad + \frac{1}{n}\sum_{i=\lfloor (R-\eta)n\rfloor}^{\lceil Rn\rceil} r_{\max}\mathsf{E}[y(u(1), D)] + \frac{\epsilon}{n}\sum_{i=1}^{\lfloor (R-\eta)n\rfloor} r_{\max} y_{\max} \\ &\leq \frac{r_{\max}}{n}\mathsf{E}\left[\sum_{i=1}^{\lfloor (R-\eta)n\rfloor} \{y(u(1), D) - y(W^{\lfloor (R-\eta)n\rfloor}_{(n)}, D)\}\right] \\ &\quad + \frac{1}{n}\sum_{i=\lfloor (R-\eta)n\rfloor}^{\lceil Rn\rceil} r_{\max}\mathsf{E}[y(u(1), D)] + \frac{\epsilon}{n}\sum_{i=1}^{\lfloor (R-\eta)n\rfloor} r_{\max} y_{\max}. \end{aligned}$$

In the last expression, the second term can be upper bounded by a constant times η (plus a constant times $1/n$), and the third term can be upper bounded by a constant times ϵ. We can choose both η and ϵ arbitrarily small and consider $n > N(\eta,\epsilon)$. This gives the desired result if we can show

$$\lim_{n\to\infty} \mathsf{E}[y(u(1), D) - y(W^{\lfloor (R-\eta)n\rfloor}_{(n)}, D)] = 0. \tag{16}$$

Fix a realization d of D. By Corollary 1, $W^{\lfloor (R-\eta)n\rfloor}_{(n)} \xrightarrow{p} u(1)$. This and the continuity of y imply that we can choose arbitrarily small constants ζ and δ such that $\Pr\{y(W^{\lfloor (R-\eta)n\rfloor}_{(n)}, d) \leq y(u(1), d) - \zeta\} \leq \delta$ for $n > M(\zeta,\delta)$ for some function M. For $n > M(\zeta,\delta)$, we have $\mathsf{E}[y(W^{\lfloor (R-\eta)n\rfloor}_{(n)}, d)] \geq (1-\delta)\cdot\{y(u(1), d) - \zeta\}$. By pushing δ and ζ to zero, we get that $g_n(d) \equiv \mathsf{E}[y(W^{\lfloor (R-\eta)n\rfloor}_{(n)}, d)]$ converges to $g(d) \equiv y(u(1), d)$ for every d. Because $0 \leq g_n(d) \leq g(d)$, the bounded convergence theorem implies that $\lim_{n\to\infty}\mathsf{E}[g_n(D)] = \mathsf{E}[g(D)]$, or equivalently, (16). □

Theorem 1 follows from this lemma because $\rho_n^{\mathcal{G}}$ is the average reward of a feasible policy and $\rho_n^{\mathcal{R}}$ is the optimal reward of a problem relaxation. Therefore, $\rho_n^{\mathcal{G}} \leq \rho_n^* \leq \rho_n^{\mathcal{R}}$ for all n. □

Proof of Proposition 1. We first demonstrate that $\mu = \delta(\mathbf{G}^*, \mathbf{x}^*)$ is feasible in (P_0). In the deterministic case,

$$\mathcal{X}(\mathbf{G}) = \left\{\mathbf{x}\in\mathbb{R}^n\colon \sum_{i\in\mathcal{I}} x^i \leq \bar{X}, 0 \leq x^i \leq y^i(\mathbf{G})\right\},$$

where we suppress in our notation the dependence on the constant $\mathbf{D}$. Thus, we have $\mathbf{x}^* \in \mathcal{X}(\mathbf{G}^*)$, $(\mathbf{G}^*, \mathbf{x}^*) \in \mathcal{M}$, and therefore $\mu(\mathcal{M}) = 1$. To verify the first set of constraints (4b) in (P_0) we consider two cases. When $\mathbf{G}^* \in \bar{\mathcal{G}}$, then the constraint is equivalent to $1 - \mathbb{1}\{\boldsymbol{\beta}\mathbf{G}^* + \mathbf{u}(\mathbf{x}^*/\mathbf{y}(\mathbf{G}^*)) \in \bar{\mathcal{G}}\} = 0$, which holds because $\mathbf{G}^*$ solves $\boldsymbol{\beta}\mathbf{G}^* + \mathbf{u}(\mathbf{x}^*/\mathbf{y}(\mathbf{G}^*)) = \mathbf{G}^*$. When $\mathbf{G}^* \notin \bar{\mathcal{G}}$, then the constraint is equivalent to $0 - \mathbb{1}\{\boldsymbol{\beta}\mathbf{G}^* + \mathbf{u}(\mathbf{x}^*/\mathbf{y}(\mathbf{G}^*)) \in \bar{\mathcal{G}}\} = 0$, which holds because $\boldsymbol{\beta}\mathbf{G}^* + \mathbf{u}(\mathbf{x}^*/\mathbf{y}(\mathbf{G}^*)) = \mathbf{G}^* \notin \bar{\mathcal{G}}$.

We prove optimality of $(\mathbf{G}^*, \mathbf{x}^*)$ by demonstrating a solution to (D_0) that achieves the same objective value as our proposed solution. Consider the solution (h^*, ρ^*) to (D_0), where $\rho^* = \sum_{i\in\mathcal{I}} r^i(x^*)^i$ and $h^*(\mathbf{G}) = 0$ for all $\mathbf{G}\in\mathcal{G}$. Feasibility in (D_0) is equivalent to the inequality $\rho^* = \sum_{i\in\mathcal{I}} r^i(x^*)^i \geq \sum_{i\in\mathcal{I}} r^i x^i$ holding for every $(\mathbf{G},\mathbf{x})\in\mathcal{M}$. Consider the maximization problem $\max_{(\mathbf{G},\mathbf{x})\in\mathcal{M}}\sum_i r^i x^i = \max_{\mathbf{G}\in\mathcal{G},\mathbf{x}\in\mathcal{X}(\mathcal{G})}\sum_i r^i x^i$. Because $y^i(G^i)$ is nondecreasing in G^i, a relaxation of this problem is

$$\max_{\mathbf{x}} \sum_{i\in\mathcal{I}} r^i x^i \quad \text{s.t. } 0 \leq x^i \leq y^i\left(\frac{u^i(1)}{1-\beta^i}\right) \quad \forall\, i\in\mathcal{I},\ \sum_i x^i \leq \bar{X},$$

which is a continuous linear knapsack problem with known greedy solution $\mathbf{x}^*$. Thus, we have that $\sum_{i\in\mathcal{I}} r^i(x^*)^i \geq \sum_{i\in\mathcal{I}} r^i x^i$ holds for every $(\mathbf{G},\mathbf{x})\in\mathcal{M}$.

The solution $\mu = \delta(\mathbf{G}^*, \mathbf{x}^*)$ yields an objective value in (P_0) of $\sum_{i\in\mathcal{I}} r^i(x^*)^i$, and the solution (h^*, ρ^*) to (D_0) yields the same objective. The proposition therefore follows from the weak duality of (D_0) and (P_0). □

Proof of Proposition 2. We fix β (letting $\bar{Y} = \bar{Y}(\beta)$, $k = k(\beta)$, and $\epsilon = \epsilon(\beta)$) and analyze the rewards of two policies. First, consider a policy that ships $x_t = \min\{\bar{X}, \beta y(W_t)\}$ at time t. We show via induction that the policy generates reward β each period. Suppose $W_{t-1} = \beta$, then $y(W_{t-1}) = 1$. The policy then sets $x_{t-1} = \min\{1, \beta\} = \beta$ and $W_t = \beta W_{t-1} + (1-\beta)(x_{t-1}/y(W_{t-1})) = \beta^2 + (1-\beta)\beta = \beta$. Part (a) of the proposition then follows because the optimal average reward exceeds the average reward of this feasible policy.

Next consider the greedy policy. We first show an inequality that will be useful in the analysis:

$$2 - \frac{1}{\beta} < 2 - \frac{1}{\beta} + (1-\beta)^2 - \epsilon < 2 - \frac{1}{\beta} + \frac{(1-\beta)^2}{\beta} - \epsilon = \beta - \epsilon. \tag{17}$$

Now we trace through the greedy policy, starting from $W_0 = \beta$.

(1) *Period* 0
 (a) $W_0 = \beta$, $y(W_0) = 1$, $x_0 = 1$.
(2) *Period* 1
 (a) $W_1 = \beta W_0 + (1-\beta)(x_0/y(W_0)) = \beta^2 + 1 - \beta$.
 (b) By the definition of ϵ, we have $\epsilon < (1-\beta)^2 = 1 + \beta^2 - 2\beta$, therefore $\beta^2 + 1 - \beta > \beta + \epsilon$ and $y(W_1) = \bar{Y}$.
 (c) The greedy policy allocates $x_1 = \bar{X} = 1$.

(3) *Period 2*

(a) $W_2 = \beta W_1 + (1-\beta)(x_1/y(W_1)) = \beta(1-\beta+\beta^2) + (1-\beta)(1/\bar{Y}) = (\beta-1+\beta^k)/\beta^k$, where the last equality follows from the definition of $\bar{Y}$.

(b) $W_2 = (\beta-1+\beta^k)/\beta^k = \beta^{-k}(\beta-1)+1 < \beta^{-1}(\beta-1)+1 = 2-1/\beta$. By (17), this means $W_2 < \beta-\epsilon$ so $y(W_2)=\epsilon$.

(4) *Periods* $t = 3, \ldots, k+1$

(a) For each of the periods $t = 2, \ldots, k$, we have $y(W_t)=\epsilon$, $x_t=\epsilon$, and $W_{t+1} = \beta^{t-1}W_2 + 1 - \beta^{t-1}$.

(b) For $2 \le t \le k$ we clearly have $W_{t+1} = \beta^{t-1}W_2 + 1 - \beta^{t-1} \le \beta^{k-1}W_2 + 1 - \beta^{k-1} = \beta^{k-1}((\beta-1+\beta^k)/\beta^k) + 1 - \beta^{k-1} = 1 - 1/\beta + \beta^{k-1} + 1 - \beta^{k-1} = 2 - 1/\beta < \beta - \epsilon$.

(5) *Period* $k+1$

(a) $W_{k+1} = 2-1/\beta < \beta-\epsilon$, so $y(W_{k+1})=\epsilon$, $x_{k+1}=\epsilon$.

(b) $W_{k+2} = \beta W_{k+1} + (1-\beta) = \beta[2-1/\beta] + (1-\beta) = \beta$.

It is clear that the cycle repeats after $k+1$ periods after accruing two periods of reward 1 and $k-1$ periods of reward ϵ. The average reward per period for the greedy policy is therefore $((k-1)\epsilon+2)/(k+1)$. By the definition of ϵ we have that $k\epsilon < 1$; therefore, $((k-1)\epsilon+2)/(k+1) < 3/(k+1)$.

Now we vary β. Writing $k = k(\beta)$ and $\epsilon = \epsilon(\beta)$ as functions of β, and noting that $k(\beta) \to \infty$ as $\beta \to 1$, we have that $0 \le \lim_{\beta\to 1}(k(\beta)\epsilon(\beta)+2)/(k(\beta)+2) \le \lim_{\beta\to 1} 3/(k(\beta)+1) = 0$. This proves part (b) of Proposition 2. $\square$

Proof of Proposition 3. Our solution solves (5) if we can show the following expression equals zero for all $\mathbf{G}$:

$$\begin{aligned}
&\sum_i h_i^\lambda(G^i) - \lambda\bar{X} \\
&\quad - \max_{0\le \mathbf{x}(\mathbf{D})\le \mathbf{y}(\mathbf{G},\mathbf{D})} \mathsf{E}_{\mathbf{D}}\bigg[\sum_i (r^i-\lambda)x^i(\mathbf{D}) - \lambda\bar{X} - \sum_i \rho_i^\lambda \\
&\qquad + \sum_i h_i^\lambda\bigg(\beta^i G^i + u^i\bigg(\frac{x^i(\mathbf{D})}{y^i(G^i,D^i)}\bigg)\bigg)\bigg] \\
&= \sum_i h_i^\lambda(G^i) + \sum_i \rho_i^\lambda \\
&\quad - \mathsf{E}_{\mathbf{D}}\bigg[\max_{0\le \mathbf{x}\le \mathbf{y}(\mathbf{G},\mathbf{D})} \sum_i \bigg\{(r^i-\lambda)x^i + h_i^\lambda\bigg(\beta^i G^i + u^i\bigg(\frac{x^i}{y^i(G^i,D^i)}\bigg)\bigg)\bigg\}\bigg] \\
&= \sum_i h_i^\lambda(G^i) + \sum_i \rho_i^\lambda \\
&\quad - \mathsf{E}_{\mathbf{D}}\bigg[\sum_i \max_{0\le x^i\le y^i(G^i,D^i)} \bigg\{(r^i-\lambda)x^i + h_i^\lambda\bigg(\beta^i G^i + u^i\bigg(\frac{x^i}{y^i(G^i,D^i)}\bigg)\bigg)\bigg\}\bigg] \\
&= \sum_i h_i^\lambda(G^i) + \sum_i \rho_i^\lambda \\
&\quad - \sum_i \mathsf{E}_{D^i}\bigg[\max_{0\le x^i\le y^i(G^i,D^i)} \bigg\{(r^i-\lambda)x^i + h_i^\lambda\bigg(\beta^i G^i + u^i\bigg(\frac{x^i}{y^i(G^i,D^i)}\bigg)\bigg)\bigg\}\bigg] \\
&= \sum_i h_i^\lambda(G^i) + \sum_i \rho_i^\lambda \\
&\quad - \sum_i \max_{0\le x^i(D^i)\le y^i(G^i,D^i)} \bigg\{\mathsf{E}_{D^i}\bigg[(r^i-\lambda)x^i(D^i) \\
&\qquad + h_i^\lambda\bigg(\beta^i G^i + u^i\bigg(\frac{x^i(D^i)}{y^i(G^i,D^i)}\bigg)\bigg)\bigg]\bigg\}.
\end{aligned}$$

Clearly, this will equal zero for any $\{(h_i^\lambda(G^i), \rho_i^\lambda), i \in \mathcal{I}\}$ that solves (6). $\square$

Proof of Proposition 4. For $r^i \ge \lambda$, we claim that the following satisfy (6):

$$h_i^\lambda(G^i) = -(r^i-\lambda)\sum_{t=0}^{\infty}\mathsf{E}\bigg[y^i\bigg(\frac{u^i(1)}{1-\beta^i}, D^i\bigg) - y^i\bigg((\beta^i)^t G^i + u^i(1)\sum_{s=0}^{t-1}(\beta^i)^s, D^i\bigg)\bigg], \tag{18a}$$

$$\rho_i^\lambda = (r^i-\lambda)\mathsf{E}\bigg[y^i\bigg(\frac{u^i(1)}{1-\beta^i}, D^i\bigg)\bigg]. \tag{18b}$$

The proposed gain and bias function satisfy (6) if the following quantity is zero:

$$\begin{aligned}
&h_i^\lambda(G^i) + \rho_i^\lambda - \max_{0\le x^i(D^i)\le y^i(G^i,D^i)} \mathsf{E}_{D^i}\bigg[(r^i-\lambda)x^i(D^i) \\
&\qquad + h_i^\lambda\bigg(\beta^i G^i + u^i\bigg(\frac{x^i(D^i)}{y^i(G^i,D^i)}\bigg)\bigg)\bigg] \\
&= -(r^i-\lambda)\sum_{t=0}^{+\infty}\mathsf{E}\bigg[y^i\bigg(\frac{u^i(1)}{1-\beta^i}, D^i\bigg) \\
&\qquad - y^i\bigg((\beta^i)^t G^i + u^i(1)\sum_{s=0}^{t-1}(\beta^i)^s, D^i\bigg)\bigg] \\
&\quad + (r^i-\lambda)\mathsf{E}\bigg[y^i\bigg(\frac{u^i(1)}{1-\beta^i}, D^i\bigg)\bigg] \\
&\quad - \max_{0\le x^i(D^i)\le y^i(G^i,D^i)} \bigg\{\mathsf{E}[(r^i-\lambda)x^i(D^i)] - (r^i-\lambda) \\
&\quad \cdot \sum_{t=0}^{+\infty}\mathsf{E}\bigg[y^i\bigg(\frac{u^i(1)}{1-\beta^i}, D^i\bigg) - y^i\bigg((\beta^i)^t\bigg(\beta^i G^i + u^i\bigg(\frac{x^i(D^i)}{y^i(G^i,D^i)}\bigg)\bigg) \\
&\qquad + u^i(1)\sum_{s=0}^{t-1}(\beta^i)^s, D^i\bigg)\bigg]\bigg\} \\
&= -(r^i-\lambda)\sum_{t=0}^{+\infty}\mathsf{E}\bigg[y^i\bigg(\frac{u^i(1)}{1-\beta^i}, D^i\bigg) \\
&\qquad - y^i\bigg((\beta^i)^t G^i + u^i(1)\sum_{s=0}^{t-1}(\beta^i)^s, D^i\bigg)\bigg] \\
&\quad + (r^i-\lambda)\mathsf{E}\bigg[y^i\bigg(\frac{u^i(1)}{1-\beta^i}, D^i\bigg) - y^i(G^i, D^i)\bigg] \\
&\quad + (r^i-\lambda)\sum_{t=0}^{+\infty}\mathsf{E}\bigg[y^i\bigg(\frac{u^i(1)}{1-\beta^i}, D^i\bigg) \\
&\qquad - y^i\bigg((\beta^i)^{t+1} G^i + u^i(1)\sum_{s=0}^{t}(\beta^i)^s, D^i\bigg)\bigg] \\
&= -(r^i-\lambda)\sum_{t=1}^{+\infty}\mathsf{E}\bigg[y^i\bigg(\frac{u^i(1)}{1-\beta^i}, D^i\bigg) \\
&\qquad - y^i\bigg((\beta^i)^t G^i + u^i(1)\sum_{s=0}^{t-1}(\beta^i)^s, D^i\bigg)\bigg] \\
&\quad + (r^i-\lambda)\sum_{t=0}^{+\infty}\mathsf{E}\bigg[y^i\bigg(\frac{u^i(1)}{1-\beta^i}, D^i\bigg) \\
&\qquad - y^i\bigg((\beta^i)^{t+1} G^i + u^i(1)\sum_{s=0}^{t}(\beta^i)^s, D^i\bigg)\bigg] \\
&= 0.
\end{aligned}$$

In the second equality, we recognize that the maximized term is nondecreasing in $x^i(D^i)$ for all D^i because $r^i - \lambda \ge 0$ and because y^i and u^i are both nondecreasing functions, and we therefore set $x^i(D^i) = y^i(G^i, D^i)$.

For $r^i < \lambda$, we claim that the solution $h_i^\lambda(G^i) = 0$, $\rho_i = 0$ satisfies (6). The proposed gain and bias function satisfy (6) if the following quantity is zero:

$$h_i^\lambda(G^i) + \rho_i^\lambda - \max_{0 \le x^i(D^i) \le y^i(G^i, D^i)} \mathsf{E}_{D^i}\left[(r^i - \lambda)x^i(D^i) + h_i^\lambda\left(\beta^i G^i + u^i\left(\frac{x^i(D^i)}{y^i(G^i, D^i)}\right)\right)\right]$$
$$= 0 - \max_{0 \le x^i(D^i) \le y^i(G^i, D^i)} \mathsf{E}_{D^i}[(r^i - \lambda)x^i(D^i) + 0].$$

Because $r^i - \lambda < 0$, the term inside the max is nonincreasing in $x^i(D^i)$, and so we set $x^i(D^i) = 0$. With this substitution, the quantity clearly evaluates to zero. □

Proof of Theorem 2. By Theorem 12 of Glashoff and Gustafson (1983, p. 80), the result holds provided (1) $\mathcal{M}$ is compact, (2) (P_ϕ) is consistent, and (3) (D_ϕ) meets the Slater condition, i.e., there exists a solution that satisfies (10b) with strict inequality. The compactness of $\mathcal{M}$ follows from Assumption 1 and the compactness of $\mathcal{G} \times \mathcal{D}$.

To show that (P_ϕ) is consistent, for each $i \in \mathcal{I}$ define

$$\bar{y}^i = \max_{(\mathbf{G},\mathbf{D}) \in \mathcal{G}\times\mathcal{D}} y^i(G^i, D^i) < \infty \quad \text{and}$$
$$\underline{y}^i = \min_{(\mathbf{G},\mathbf{D}) \in \mathcal{G}\times\mathcal{D}} y^i(G^i, D^i) > 0.$$

Consider the set $\tilde{\mathcal{X}} = \{\mathbf{x} \in \mathbb{R}^n\colon \sum_{i \in \mathcal{I}} x^i \le \bar{X}, 0 \le x^i \le \underline{y}^i\ \forall i \in \mathcal{I}\}$. Because $\bar{X} > 0$ and $\underline{y}^i > 0$ for all i, we can choose $\tilde{\mathbf{x}} \in \tilde{\mathcal{X}}$ that is in the interior of $\tilde{\mathcal{X}}$. Define $f^i = \tilde{x}^i/\bar{y}^i$ to be the smallest fill rate for customer i, and let $\mathbf{f}$ be the corresponding vector. Then set

$$\hat{x}^i(\mathbf{D}) = f^i y^i(G^i, D^i) \quad \forall i \in \mathcal{I}, (\mathbf{G},\mathbf{D}) \in \mathcal{G}\times\mathcal{D}.$$

By construction, $\hat{\mathbf{x}} \in \mathcal{X}(\mathbf{G})$ for all $\mathbf{G} \in \mathcal{G}$, because $0 < f^i < 1$ implies $0 < \hat{x}^i(\mathbf{D}) < y^i(G^i, D^i)$ for all i, and

$$\sum_{i\in\mathcal{I}} \hat{x}^i(\mathbf{D}) = \sum_{i\in\mathcal{I}} f^i y^i(G^i, D^i) = \sum_{i\in\mathcal{I}} (\tilde{x}^i/\bar{y}^i)\cdot y^i(G^i, D^i) \le \sum_{i\in\mathcal{I}} \tilde{x}^i < \bar{X}.$$

Next, consider a $\hat{\mathbf{G}}$ that solves $\hat{\mathbf{G}} = \boldsymbol{\beta}\hat{\mathbf{G}} + \mathbf{u}(\hat{\mathbf{x}}(\mathbf{D})/\mathbf{y}(\mathbf{G},\mathbf{D})) = \boldsymbol{\beta}\hat{\mathbf{G}} + \mathbf{u}(\mathbf{f})$. This yields $\hat{G}^i = u^i(f^i)/(1-\beta^i)$ for all i. Therefore, for every $k \in \mathcal{K}$ we have for $(\hat{\mathbf{G}}, \hat{\mathbf{x}})$

$$\phi_k(\hat{\mathbf{G}}) - \mathsf{E}_{\mathbf{D}}[\phi_k(\boldsymbol{\beta}\hat{\mathbf{G}} + \mathbf{u}(\hat{\mathbf{x}}(\mathbf{D})/\mathbf{y}(\hat{\mathbf{G}},\mathbf{D})))]$$
$$= \phi_k(\hat{\mathbf{G}}) - \mathsf{E}_{\mathbf{D}}[\phi_k(\boldsymbol{\beta}\hat{\mathbf{G}} + \mathbf{u}(\mathbf{f}))]$$
$$= \phi_k(\hat{\mathbf{G}}) - \phi_k(\boldsymbol{\beta}\hat{\mathbf{G}} + \mathbf{u}(\mathbf{f})) = \phi_k(\hat{\mathbf{G}}) - \phi_k(\hat{\mathbf{G}}) = 0.$$

Hence, we can set $M = \{(\hat{\mathbf{G}}, \hat{\mathbf{x}})\}$ and $\mu(\hat{\mathbf{G}}, \hat{\mathbf{x}}) = 1$ to obtain a feasible solution to (P_ϕ).

Finally, we need to verify that (D_ϕ) has a Slater point. Set $w_k = 0$ for all $k \in \mathcal{K}$, and then choose any $\rho \in \mathbb{R}$ that satisfies

$$\rho > \max_{(\mathbf{G},\mathbf{x}) \in \mathcal{M}} \mathsf{E}_{\mathbf{D}}\left[\sum_{i\in\mathcal{I}} r^i x^i(\mathbf{D})\right].$$

Then, this (w, ρ) satisfies (10b) with strict inequality; therefore, the Slater condition is satisfied. □

Proof of Proposition 5. Consider adding the constraints $w_i \ge 0$ for all i to (D_ϕ) to form a new problem (D'_ϕ). This affects the dual (P_ϕ) by changing the first set of constraints to inequalities, i.e.,

$$\sum_{(\mathbf{G},\mathbf{x}) \in M} \mu(\mathbf{G},\mathbf{x})\left[G^i(1-\beta^i) - \mathsf{E}_D\left[u^i\left(\frac{x^i(\mathbf{D})}{y^i(G^i, D^i)}\right)\right]\right] \le 0 \quad \forall i. \tag{19}$$

Let μ^* denote an optimal solution to the revised problem (P'_ϕ). We will construct a solution $\hat{\mu}$ to (P'_ϕ) that satisfies constraints (19) with equality without changing the objective. The desired result then follows because $\inf(\mathrm{D}_\phi) = \sup(\mathrm{P}_\phi)$ (by Theorem 2) and $\inf(\mathrm{D}'_\phi) = \sup(\mathrm{P}'_\phi)$ (by the same arguments as in the proof of Theorem 2).

Consider an i for which

$$T_i(M) \equiv \sum_{(\mathbf{G},\mathbf{x}) \in M} \mu^*(\mathbf{G},\mathbf{x})\left[G^i(1-\beta^i) - \mathsf{E}_D\left[u^i\left(\frac{x^i(\mathbf{D})}{y^i(G^i, D^i)}\right)\right]\right] < 0.$$

It is possible to partition M into two sets M_i^+ and M_i^- such that

$$\left[G^i(1-\beta^i) - \mathsf{E}_D\left[u^i\left(\frac{x^i(\mathbf{D})}{y^i(G^i, D^i)}\right)\right]\right] \ge 0 \quad \forall (\mathbf{G},\mathbf{x}) \in M_i^+,$$
$$\left[G^i(1-\beta^i) - \mathsf{E}_D\left[u^i\left(\frac{x^i(\mathbf{D})}{y^i(G^i, D^i)}\right)\right]\right] < 0 \quad \forall (\mathbf{G},\mathbf{x}) \in M_i^-,$$

and $T_i(M_i^-) < -T_i(M_i^+) \le 0$. For any $(\mathbf{G},\mathbf{x}) \in M_i^-$, consider $\hat{G}^i(\alpha_i) = (1-\alpha_i)G^i + \alpha_i(u^i(1)/(1-\beta^i))$ for $\alpha_i \in [0,1]$. Because $\hat{G}^i(\alpha_i) \ge G^i$, we have $x^i(\mathbf{D})/(y^i(\hat{G}^i(\alpha_i), D^i)) \le x^i(\mathbf{D})/(y^i(G^i, D^i)) \le 1$ for all $\alpha_i \in [0,1]$.

Define

$$T_i(M_i^-; \alpha_i) \equiv \sum_{(\mathbf{G},\mathbf{x}) \in M_i^-} \mu^*(\mathbf{G},\mathbf{x})\left[\hat{G}^i(\alpha_i)(1-\beta^i) - \mathsf{E}_D\left[u^i\left(\frac{x^i(\mathbf{D})}{y^i(\hat{G}^i(\alpha_i), D^i)}\right)\right]\right].$$

Observe that this is a continuous function of α_i (by the continuity of u^i and y^i) with $T_i(M_i^-; 0) = T_i(M_i^-)$ and $T_i(M_i^-; 1) \ge 0$. Hence, there exists an $\alpha_i^* \in [0,1]$ such that $T_i(M_i^-; \alpha_i^*) = -T_i(M_i^+)$. Let $\hat{\mu}$ equal the measure obtained from μ^* by shifting the marginal weight from G^i (for $(\mathbf{G},\mathbf{x}) \in M_i^-$) to $\hat{G}^i(\alpha_i^*)$. By the choice of α_i^*, we have

$$\sum_{(\mathbf{G},\mathbf{x}) \in M} \hat{\mu}(\mathbf{G},\mathbf{x})\left[G^i(1-\beta^i) - \mathsf{E}_D\left[u^i\left(\frac{x^i(\mathbf{D})}{y^i(G^i, D^i)}\right)\right]\right] = 0.$$

Furthermore, $\hat{\mu}$ is feasible because the marginal measure on $\mathbf{x}$ has not changed and because $y^i(\hat{G}^i(\alpha_i^*), D^i) \ge y^i(G^i, D^i)$ for all D^i and $(\mathbf{G},\mathbf{x}) \in M_i^-$. The objective to ($\mathrm{P}'_\phi$) depends only on $\mathbf{x}$ and therefore is the same under $\hat{\mu}$ as under μ^*. □

References

Adelman D, Mersereau AJ (2008) Relaxations of weakly coupled stochastic dynamic programs. *Oper. Res.* 56(3):712–727.

Aflaki S, Popescu I (2012) Managing retention in service relationships. Working paper, INSEAD, Singapore.

Bertsekas DP, Tsitsiklis JN (1996) *Neuro-Dynamic Programming* (Athena Scientific, Belmont, MA).

Cachon GP, Lariviere MA (1999) Capacity allocation using past sales: When to turn-and-earn. *Management Sci.* 45(5):685–703.

Cachon GP, Zhang F (2007) Obtaining fast service in a queueing system via performance-based allocation of demand. *Management Sci.* 53(3):408–420.

Carr S, Lovejoy W (2000) The inverse newsvendor problem: Choosing an optimal demand portfolio for capacitated resources. *Management Sci.* 46(7):912–927.

de Farias DP, Van Roy B (2003) The linear programming approach to approximate dynamic programming. *Oper. Res.* 51(6):850–865.

Dube J-P, Hitsch GJ, Manchanda P (2005) An empirical model of advertising dynamics. *Quant. Market. Econom.* 3(2):107–144.

Fader PS, Hardie BGS (2009) Probability models for customer-base analysis. *J. Interactive Marketing* 23(1):61–69.

Federgruen A, Yang N (2008) Selecting a portfolio of suppliers under demand and supply risks. *Oper. Res.* 56(4):916–936.

Feichtinger G, Hartl RF, Sethi SP (1994) Dynamic optimal control models in advertising: Recent developments. *Management Sci.* 40(2):195–226.

Gans N (2002) Customer loyalty and supplier quality competition. *Management Sci.* 48(2):207–221.

Gaur V, Park Y-H (2007) Asymmetric consumer learning and inventory competition. *Management Sci.* 53(2):227–240.

Gittins J, Glazebrook K, Weber R (2011) *Multi-Armed Bandit Allocation Indices*, 2nd ed. (John Wiley & Sons, Chichester, UK).

Glashoff K, Gustafson S-A (1983) *Linear Optimization and Approximation* (Springer-Verlag, New York).

Gupta S, Lehmann DR (2008) Models of customer value. Wierenga B, ed. *Handbook of Marketing Decision Models* (Springer Science+Business Media, New York), 255–290.

Hall J, Porteus E (2000) Customer service competition in capacitated systems. *Manufacturing Service Oper. Management* 2(2):144–165.

Hernández-Lerma O, Lasserre JB (1996) *Discrete-Time Markov Control Processes—Basic Optimality Criteria* (Springer, New York).

Judd KL (1998) *Numerical Methods in Economics* (MIT Press, Cambridge, MA).

Liu L, Shang W, Wu S (2007) Dynamic competitive newsvendors with service-sensitive demands. *Manufacturing Service Oper. Management* 9(1):84–93.

Müller A, Stoyan D (2002) *Comparison Methods for Stochastic Models and Risks* (John Wiley & Sons, Chichester, UK).

Nerlove M, Arrow KJ (1962) Optimal advertising policy under dynamic conditions. *Economica* 29(114):129–142.

Olsen TL, Parker RP (2008) Inventory management under market size dynamics. *Management Sci.* 54(10):1805–1821.

Powell W (2007) *Approximate Dynamic Programming: Solving the Curses of Dimensionality* (Wiley Interscience, Hoboken, NJ).

Reinartz WJ, Venkatesan R (2008) Decision models for customer relationship management (CRM). Wierenga B, ed. *Handbook of Marketing Decision Models* (Springer Science+Business Media, New York), 291–326.

Schweitzer PJ, Seidmann A (1985) Generalized polynomial approximations in Markovian decision processes. *J. Math. Anal. Appl.* 110(2):568–582.

Sethi SP (1977) Dynamic optimal control models in advertising: A survey. *SIAM Rev.* 19(4):685–725.

Simester DI, Sun P, Tsitsiklis JN (2006) Dynamic catalog mailing policies. *Management Sci.* 52(5):683–696.

Weber RR, Weiss G (1990) On an index policy for restless bandits. *J. Appl. Prob.* 27(3):637–648.

Whittle P (1988) Restless bandits: Activity allocation in a changing world. *J. Appl. Probab.* 25:287–298.