Dynamic Assortment Optimization for Reusable Products with Random Usage Durations

Paat Rusmevichientong
Marshall School of Business, University of Southern California, Los Angeles, CA 90089, rusmevic@marshall.usc.edu

Mika Sumida and Huseyin Topaloglu
School of Operations Research and Information Engineering, Cornell Tech, NY, NY 10044, {ms3268,ht88}@cornell.edu

We consider dynamic assortment problems with reusable products, in which each arriving customer chooses a product within an offered assortment, uses the product for a random duration of time, and returns the product back to the firm to be used by other customers. The goal is to find a policy for deciding on the assortment to offer to each customer so that the total expected revenue over a finite selling horizon is maximized. The dynamic programming formulation of this problem requires a high-dimensional state variable that keeps track of the on-hand product inventories, as well as the products that are currently in use. We present a tractable approach to compute a policy that is guaranteed to obtain at least 50% of the optimal total expected revenue. This policy is based on constructing linear approximations to the optimal value functions. When the usage duration is infinite or follows a negative binomial distribution, we also show how to efficiently perform rollout on a simple static policy. Performing rollout corresponds to using separable and nonlinear value function approximations. The resulting policy is also guaranteed to obtain at least 50% of the optimal total expected revenue. The special case of our model with infinite usage durations captures the case where the customers purchase the products outright without returning them at all. Under infinite usage durations, we give a variant of our rollout approach that is guaranteed to obtain at least $\max\left\{\frac{1}{2}, 1 - \frac{R}{2\sqrt[3]{C_{\min}}}\right\}$ fraction of the optimal total expected revenue, where $C_{\min}$ is the smallest inventory of a product and R is the largest relative deviation of the price of a product over the selling horizon. We provide computational experiments based on simulated data for dynamic assortment management, as well as real parking transaction data for the city of Seattle. Our computational experiments demonstrate that the practical performance of our policies is substantially better than their performance guarantees and that performing rollout yields noticeable improvements.

Key words: dynamic assortment optimization, reusable products, and choice modeling

1. Introduction

Revenue management problems focus on making capacity allocation decisions for limited inventories of products over a finite selling horizon. These problems have applications in areas as diverse as airline, hotel, electric power, health care, consumer credit, cruise line, and advertising capacity management (Ozer and Phillips 2012). The dynamic programming formulations of revenue management problems are generally intractable because they require high-dimensional state variables that keep track of the remaining inventory of each product. Thus, computing the optimal policy is computationally difficult, so researchers have focused on approximate policies. In traditional application areas of revenue management problems, the customers purchase the products for final consumption. Some emerging industries, however, focus on renting out products such as computing capacity, fashion items and storage units. In these industries, each customer requests a product, uses the product for a possibly random duration of time and returns the product back to the firm, at which point the product can be used by other customers. For example, firms such as Amazon and Google offer cloud computing services, where users utilize computing capacity for a certain duration of time before returning it. The firm needs to decide what type of computing capacity to offer to each user and what prices to charge. Firms such as Rent the Runway and Glam Corner rent fashion items to shoppers through their online platforms. The firm needs to decide which assortment of fashion items to offer to each shopper and at what prices. Firms such as Cube Smart and Make Space lease storage units, where customers return the leased storage units back after using them for a certain duration of time. The firm needs to decide what prices to charge for the storage units as a function of the current occupancy. Using real-time information on the availability of street parking spaces, city governments have the opportunity to dynamically adjust the price for parking spaces, where each driver uses a parking space for a certain duration of time before leaving and making it available for other drivers. When making capacity management decisions in such environments, the firm must consider the on-hand product inventories, as well as the products that are currently in use.

In this paper, we consider dynamic assortment problems with reusable products. In our problem setting, we have access to a set of products with limited inventories. Customers randomly arrive into the system. Among the products for which we currently have available units on-hand, we offer an assortment of products to the arriving customer. The customer either chooses a product from the offered assortment or decides to leave the system. If the customer chooses a product, then she uses the product for a random duration of time. After a usage duration, the customer returns the product. The returned product can be used to offer an assortment to another customer in the future. The goal is to find a policy for deciding on the assortment to offer to each customer so that the total expected revenue over a finite selling horizon is maximized.

Our dynamic programming formulation of the problem allows for a broad class of choice models for describing the choice process of the customers, non-stationarities in the revenue structure, and arbitrary distributions for the random usage durations. In our formulation, the randomness in the usage duration is not resolved until the customer returns the rented product back, but we can also modify our formulation to address the case where the usage duration is revealed before the firm makes its assortment offering decision. To our knowledge, our model is the first revenue management model with limited inventories of reusable resources, where the customers can choose among the offered products according to a broad class of choice models, there can be non-stationarities in the revenue structure, and the distributions of the usage durations can be arbitrary. The dynamic programming formulation of the problem requires a high-dimensional state variable that keeps track of the inventories of the products that are available on-hand, as well as the products that are currently in use. Therefore, finding the optimal policy is computationally difficult. We propose tractable policies that provide performance guarantees.

Main Contributions: In Section 3, we provide a method to construct linear approximations to the optimal value functions in the dynamic programming formulation of the problem. Our method uses an efficient backward recursion over the time periods in the selling horizon. At each time period, we solve a static assortment problem, where we adjust the product revenues by time-dependent constants computed from the recursion and find an assortment of products that maximizes the expected adjusted revenue from a customer (Section 3.1). We show that the greedy policy with respect to our linear value function approximations is guaranteed to obtain at least 50% of the optimal total expected revenue (Section 3.2), but this policy turns out to be agnostic to inventory levels. Specifically, whether this policy offers a particular product at a particular time period does not depend on the exact on-hand inventory level of this product, as long as on-hand inventory is available. To remedy this shortcoming, we construct separable and nonlinear value function approximations, as discussed next.

In Section 4, we perform rollout on a static policy to construct separable and nonlinear value function approximations. We start with a static policy that offers the same assortment at a particular time period, irrespective of the state of the system (Section 4.1). We compute the total expected revenue obtained by the static policy starting at each state and at each time period, which can be done by focusing on each product separately. We use these total expected revenues as the value function approximations at different states and at different time periods. In this way, we obtain separable and nonlinear value function approximations. In rollout, we use the greedy policy with respect to these value function approximations (Bertsekas and Tsitsiklis 1996). We show that the policy obtained through the rollout approach is also guaranteed to yield at least 50% of the optimal total expected revenue (Section 4.2). This policy is not agnostic to inventory levels, unlike the greedy policy with respect to linear value function approximations. We demonstrate that we can efficiently perform rollout when the usage duration follows a negative binomial distribution (Section 4.3) or when the usage duration is infinite (Section 4.4).

The case with infinite usage durations corresponds to the situation where the customers purchase the product outright, rather than renting. Under infinite usage durations, we also tailor our rollout approach to strengthen the performance guarantee. In particular, we use $C_{\min}$ to denote the smallest inventory of a product and R to denote the largest relative deviation of the price of a product over the selling horizon. If the prices of the products are stationary, then we have $R = 1$. We show how to construct separable and nonlinear value function approximations under infinite usage durations such that the greedy policy with respect to these value function approximations is guaranteed to obtain at least $\max\left\{\frac{1}{2}, 1 - \frac{R}{2\sqrt[3]{C_{\min}}}\right\}$ fraction of the optimal total expected revenue. Therefore, the tailored policy provides at least a half-approximate performance guarantee, but as the inventories of the products become large, the tailored policy becomes near-optimal. Our dynamic assortment problem with infinite usage durations corresponds to the choice-based revenue management problem over parallel flight legs operating between the same origin-destination pair, which is an important problem class that has been studied in the literature (Zhang and Cooper 2005, Liu and van Ryzin 2008, Dai et al. 2014).

In Section 5, we provide extensions of our results. We extend our approach to the case with multiple types of customers, each of whom makes choices according to a different choice model and rents the products according to different usage distributions (Section 5.1). In our setup, we know the type of a customer before offering an assortment, so that we can personalize the assortment according to known customer features. Since different customer types can have different usage distributions and we know the type of a customer before offering an assortment, this extension allows us to capture the case where the usage duration is revealed before we offer an assortment to a customer. We also extend our approach to the cases where we set the prices of the products rather than choosing an assortment to offer (Section 5.2) and when we only approximately solve the assortment optimization problems used in the construction of our value function approximations (Section 5.3).

In Section 6, we provide computational experiments. We formulate a linear program that yields an upper bound on the optimal total expected revenue (Section 6.1). In our first set of computational experiments, we consider dynamic assortment management, where we offer an assortment of products to each arriving customer. (Section 6.2). Our policies perform remarkably well when compared with the upper bound on the optimal total expected revenue and yield average improvements of 1-10% when compared with other benchmarks. In our second set of computational experiments, we consider the problem of dynamically adjusting the prices for street parking spaces (Section 6.3). We treat each parking space as a reusable product with a random usage duration. Using data from the city of Seattle to estimate the model parameters, dynamically adjusting the prices improves the total expected revenues by 2-7% when compared to static pricing.

Literature Review: We review the recent work on revenue management models with reusable products. Levi and Radovanovic (2010) study a model that assumes independent demands across products, without any choice behavior for the customers. Focusing on the infinite selling horizon setting with stationary demand, the authors establish a performance guarantee for a static policy that does not consider the real-time state of the system. Owen and Simchi-Levi (2017) extend this work to incorporate customer choice behavior and a finite selling horizon. The authors assume that the usage durations are exponentially distributed and note that this assumption can be relaxed under a stationary customer choice process. They develop a policy that is guaranteed to obtain $1/e$ fraction of the optimal total expected revenue. This policy may offer products for which there is no on-hand inventory. If the customer chooses a product for which there is no on-hand inventory, then she must leave without making a purchase. The policy in their paper is also static because it offers each assortment of products with a fixed probability that does not depend on the real-time state of the system. By contrast, our model can accommodate arbitrary distributions for the random usage durations and arbitrary non-stationarities in the choice process of the customers. The policy that we construct takes the real-time state of the system into consideration. As long as the choice process of the customers is governed by a random utility maximization principle, our policy never offers products for which there is no on-hand inventory.

Lei and Jasin (2016) develop a model with reusable resources, deterministic usage durations, and advance reservations. Their model includes multiple resources, and each product uses a different combination of resources. The authors give a data-dependent performance guarantee and show that their model is asymptotically optimal when the resource inventories and the number of time periods in the selling horizon scale up linearly at the same rate. Chen et al. (2017) study a model with multiple units of a single reusable product, random usage durations and advance reservations. Their model allows random usage durations, but the usage duration is revealed at the time of the reservation. The authors also provide a data-dependent performance guarantee and show that their model is asymptotically optimal when the product inventory and the customer arrival rate scale up linearly at the same rate. In our work, we do not allow advance reservations, but we provide constant-factor performance guarantees that are not dependent on the problem data, work with arbitrary usage duration distributions and allow the randomness in the usage durations to be resolved when the customer returns the product.

Golrezaei et al. (2014) study dynamic assortment problems with non-reusable products. In essence, their model is a special case of ours with infinite usage durations. Considering the case with multiple customer types, the authors construct a policy that is guaranteed to obtain at least 50% of the optimal total expected revenue, even when the sequence of customer type arrivals is chosen by an adversary. As the product inventories become arbitrarily large, the performance guarantee of their policy improves from 50% to $1-1/e$. When the type of a customer is drawn from a stationary distribution over the time periods, the performance guarantee further improves to 75%. The key idea in this work is to adjust revenue from the sale of each product by a revenue modifier, which is an increasing function of the current inventory of the product. The policy offers the assortment that maximizes the expected adjusted revenue from each customer. As the inventory of a product is depleted, its adjusted revenue decreases and the policy is more likely not to offer this product. The proof of the performance guarantee in Golrezaei et al. (2014) is based on formulating a deterministic linear programming approximation and constructing a dual feasible solution to the approximation by using the revenue modifier. We note that it is possible to formulate a similar linear program under reusable products, but this linear program has a capacity constraint for each product and at each time period, thus ensuring that the expected number of products that are on-hand and in use at any time period will not exceed the product inventory. The dual of this linear program is substantially more complicated and it is not clear how to construct a feasible dual solution by using the revenue modifier.

Considering the problem setting in Golrezaei et al. (2014), we can tailor our results in this paper to non-reusable products to obtain stronger performance guarantees. Under non-reusable resources, we can give a variant of our rollout approach that is guaranteed to obtain $\max\left\{\frac{1}{2}, 1-\frac{R}{2\sqrt[3]{C_{\min}}}\right\}$ fraction of the optimal total expected revenue, where $C_{\min}$ and R are as discussed earlier in this section. If, for example, the prices of the products are stationary and each product has at least 100 units of inventory, then this performance guarantee computes to be 89%. Similar to the policy in Golrezaei et al. (2014), all of our policies are based on adjusting the revenue from the sale of each product by a revenue modifier. The revenue modifiers in Golrezaei et al. (2014) are multiplicative, whereas our revenue modifiers are additive. The construction of the revenue modifiers in Golrezaei et al. (2014) only uses the current and initial inventory levels, whereas the construction of our revenue modifiers uses all of the problem data. Thus, the revenue modifier in Golrezaei et al. (2014) is robust as it is insensitive to a large part of the problem data. Our computational experiments, however, indicate that using all of the problem data pays off and our policies can perform noticeably better than the one in Golrezaei et al. (2014). Motivated by the online resource allocation setting, Stein et al. (2016), Wang et al. (2016) and Gallego et al. (2016) also consider problems that involve allocating products to customers arriving over time and provide policies with performance guarantees, but this stream of work does not deal with reusable products either.

In the work discussed so far, the initial inventories of the products are exogenously given. There is work on computing stocking quantities at the beginning of the selling horizon when the customers arriving over time choose among the products according to a certain choice model. The paper by van Ryzin and Mahajan (1999) gives a model to compute the optimal stocking quantities under the assumption that a customer can choose a product for which there is no on-hand inventory, in which case, she leaves without a purchase, possibly resulting in a penalty or emergency procurement cost. Other work considers the case where a customer chooses only among the products for which there is on-hand inventory. Mahajan and van Ryzin (2001) use stochastic descent to compute stocking quantities without a performance guarantee. Honhon et al. (2010) use a choice model based on ranked preference lists and compute the optimal stocking quantities through a dynamic program, whose running time is exponential in the number of products. Under ranked preference lists, Aouad et al. (2015) give approximation algorithms, whereas Goyal et al. (2016) give polynomial-time approximation schemes. Under the multinomial logit model, Aouad et al. (2017) provide approximation algorithms. These papers do not consider reusable products.

Finally, our work is related to revenue management problems under customer choice. Zhang and Cooper (2005) compute upper bounds on the optimal value functions for the choice-based parallel flights problem. Gallego et al. (2004) focus on network revenue management problems and study static policies extracted from a deterministic linear program. Adelman (2007) constructs linear value function approximations when customers arrive into the system to purchase fixed products without a choice process. His approach yields upper bounds on the optimal value functions, but without a performance guarantee. Tong and Topaloglu (2013) show that the approach in Adelman (2007) boils down to solving a linear program whose dimensions increase only linearly with the numbers of itineraries, flights and time periods. Liu and van Ryzin (2008) develop dynamic programming decomposition methods for decomposing the dynamic programming formulation of the network revenue management problem by the flight legs. The authors obtain separable and nonlinear value function approximations, also without a performance guarantee. Zhang and Adelman (2009) and Vossen and Zhang (2015) extend the work of Adelman (2007) to include a customer choice process. Their approach requires solving a linear program whose number of constraints increases exponentially with the number of itineraries. Therefore, the linear program is solved using column generation. We can solve the column generation subproblem efficiently under some choice models, but there is no a priori bound on the number of columns that need to be generated to obtain the optimal solution. Overall, the work discussed in this paragraph constructs linear and nonlinear value function approximations, but without performance guarantees.

Organization: In Section 2, we provide a dynamic programming formulation for our dynamic assortment problem with reusable products. In Section 3, we design a policy that is guaranteed to obtain at least 50% of the optimal total expected revenue. This policy uses linear value function approximations. In Section 4, we use rollout on a static policy to obtain separable and nonlinear value function approximations. The resulting policy is also guaranteed to obtain at least 50% of the optimal total expected revenue. In Section 5, we describe extensions to the cases where we have multiple customer types, we make pricing decisions and we can solve the assortment problems with errors. In Section 6, we give our computational experiments. In Section 7, we conclude.

2. Problem Formulation

We have a set of products with limited inventories. At each time period in the selling horizon, we decide on the set of products to offer. A customer arriving into the system either chooses to rent one of the offered products or decides to leave the system without renting anything. We capture the choice process of the customers through a discrete choice model. If the customer chooses to rent one of the offered products, then she uses the product for a random duration of time by paying an upfront fee and a per-period rental fee for each time period that she uses the product. After using the product for a random duration of time, the customer returns the product, at which point, we can rent the product to another customer. Our goal is to find a policy for maximizing the total expected revenue over the selling horizon. We describe the problem data, state and transition dynamics, followed by a dynamic programming formulation of the problem.

Problem Data: We have n products indexed by $\mathcal{N} = \{1, \ldots, n\}$. For each product $i \in \mathcal{N}$, let $C_i \in \mathbb{Z}_+$ denote its initial inventory level. There are T time periods in the selling horizon indexed by $\mathcal{T} = \{1, \ldots, T\}$. Each time period corresponds to a small interval of time and there is exactly one customer arrival at each time period. It is not difficult to extend our model to the case with at most one customer arrival at each time period. If we offer the subset of products S, then a customer arriving at time period t chooses product i with probability $\phi_i^t(S)$. Naturally, we have $\phi_i^t(S) = 0$ for all $i \notin S$. If a customer chooses to rent product i at time period t, then she immediately pays a one-time upfront fee of r_i^t. If a customer is renting product i during time period t, then she also pays a per-period rental fee of π_i^t. Depending on the specific application at hand, one of the fees can be zero; in addition, one or both of the fees can be stationary. The per-period rental fee can also depend on how long the product has been in use.

We use the generic random variable $\mathsf{Duration}_i$ to represent the random rental duration of product i. The random variable $\mathsf{Duration}_i$ has a probability mass function $f_i : \mathbb{Z}_{++} \mapsto [0, 1]$, where $\sum_{\ell=1}^{\infty} f_i(\ell) = 1$. We describe the rental duration in terms of its hazard rate $\rho_{i,\ell}$ associated with the probability mass function f_i, where for each $\ell \in \mathbb{Z}_+$, we have

$$\rho_{i,\ell} \;=\; \Pr\{\mathsf{Duration}_i = \ell + 1 \mid \mathsf{Duration}_i > \ell\} \;=\; \frac{f_i(\ell+1)}{\sum_{s=\ell+1}^{\infty} f_i(s)}.$$

The hazard rate $\rho_{i,\ell}$ is the probability that each unit of product i is returned after $\ell+1$ periods, given that the product has been used for more than ℓ periods. Since $\sum_{s=1}^{\infty} f_i(s) = 1$, we have $\rho_{i,0} = f_i(1)$, so that $\rho_{i,0}$ is the probability that a unit of product i is used for exactly one time period. The usage durations of different units are assumed to be independent of each other.

At each time period t, the following sequence of events happen. We observe the state of the system, which consists of the current on-hand units and the outstanding units that are currently being rented by the customers. Based on the state, we decide which subset of products to offer. The customer arriving at time period t chooses a unit to rent or leaves the system without renting. We collect the upfront fee for the rented unit and the rent from all customers still using their rented units. Finally, we observe whether each customer with a rented unit of product decides to return the unit, including the customer who rented a unit at the current time period.

State and Transition Dynamics: To capture the state of the system at a generic time period, let $q_{i,0}$ denote the number of units of product i available as on-hand inventory at the beginning of the time period. Also, for $\ell \geq 1$, let $q_{i,\ell}$ denote the number of units of product i that have been used for exactly ℓ time periods at the beginning of the time period. Thus, we describe the state of the system by the vector $\boldsymbol{q} = (q_{i,\ell} : i \in \mathcal{N},\ \ell = 0, 1, \ldots)$. For example, if $\boldsymbol{q}$ represents the state of the system at the beginning of time period t, then $q_{i,1}$ is the number of units of product i rented at time period $t-1$ and not returned by the beginning of time period t. Since $\sum_{\ell=0}^{\infty} q_{i,\ell} = C_i$, let $\mathcal{Q} = \{(q_{i,\ell} : i \in \mathcal{N},\ \ell = 0, 1, \ldots) : \sum_{\ell=0}^{\infty} q_{i,\ell} = C_i\ \forall\, i \in \mathcal{N}\}$ denote the set of all possible states. We assume that the system starts with no units in use. Thus, there will never be a unit in use for more than T time periods, which makes the effective set of possible states finite.

Consider the state $\boldsymbol{q}$ at the beginning of time period t. There are $q_{i,\ell}$ units of product i that have been used for exactly ℓ periods. By definition of the hazard rate, with probability $\rho_{i,\ell}$, each of the $q_{i,\ell}$ units will be returned by the beginning of time period $t+1$. Thus, if no purchase is made at time period t, then the number of units that will be available as on-hand inventory at the beginning of time period $t+1$ is $q_{i,0} + \sum_{\ell=1}^{\infty} \mathsf{Bin}(q_{i,\ell}, \rho_{i,\ell})$, where $\mathsf{Bin}(k,p)$ denotes a binomial random variable with parameters $k \in \mathbb{Z}_+$ and $p \in [0,1]$. Also, at the beginning of time period $t+1$, the number of units of product i that will have been rented out for $\ell+1$ periods will be $q_{i,\ell} - \mathsf{Bin}(q_{i,\ell}, \rho_{i,\ell})$, where the second term reflects the units that will be returned at time period t. Therefore, given the state $\boldsymbol{q}$ at the beginning of time period t, if there is no purchase by a customer, then the state $\boldsymbol{X}(\boldsymbol{q}) = (X_{i,\ell}(\boldsymbol{q}) : i \in \mathcal{N},\ \ell = 0, 1, \ldots)$ at the beginning of period $t+1$ is given by

$$X_{i,\ell}(\boldsymbol{q}) = \begin{cases} q_{i,0} + \sum_{s=1}^{\infty} \mathsf{Bin}(q_{i,s}, \rho_{i,s}) & \text{if } \ell = 0, \\ 0 & \text{if } \ell = 1, \\ q_{i,\ell-1} - \mathsf{Bin}(q_{i,\ell-1}, \rho_{i,\ell-1}) & \text{if } \ell \geq 2. \end{cases} \tag{1}$$

Note that if there is no purchase at time period t, then a unit that was on-hand will stay on-hand at time period $t+1$. Also, a unit that was in use will either be returned or it will stay in use. In the latter case, its usage duration will be at least two time periods. So, we have $X_{i,1}(\boldsymbol{q})=0$.

Dynamic Programming Formulation: We use $\mathcal{F}$ to denote the collection of feasible subsets of products that we can offer to the customers at each time period, which captures the constraints that we may impose on the offered subset of products. To formulate the problem as a dynamic program, we denote a Bernoulli random variable with parameter $\rho \in [0,1]$ by $\mathsf{Z}(\rho)$. Also, viewing the state $\boldsymbol{q}=(q_{i,\ell}: i \in \mathcal{N},\ \ell = 0,1,\ldots)$ as a vector, we let $\boldsymbol{e}_{i,k}$ be a unit vector with one in the (i,k)-th coordinate and zero everywhere else. Let $J^t(\boldsymbol{q})$ denote the maximum total expected revenue over the time periods $t,\ldots,T$, given that the system is in state $\boldsymbol{q}$ at the beginning of time period t. Then, using $\mathbb{1}_{\{\cdot\}}$ to denote the indicator function, we can compute the optimal value functions $\{J^t : t \in \mathcal{T}\}$ by solving the dynamic program

$$\begin{aligned}
J^t(\boldsymbol{q}) &= \sum_{i\in\mathcal{N}} \pi_i^t \sum_{\ell=1}^{\infty} q_{i,\ell} \\
&+ \max_{S\in\mathcal{F}} \Bigg\{ \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\geq 1\}}\, \phi_i^t(S) \Bigg(r_i^t + \pi_i^t + \mathbb{E}\Big\{ \mathsf{Z}(\rho_{i,0})\, J^{t+1}(\boldsymbol{X}(\boldsymbol{q})) + (1-\mathsf{Z}(\rho_{i,0}))\, J^{t+1}(\boldsymbol{X}(\boldsymbol{q}) - \boldsymbol{e}_{i,0} + \boldsymbol{e}_{i,1}) \Big\} \Bigg) \\
&\qquad\qquad + \Big(1 - \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\geq 1\}} \phi_i^t(S) \Big)\, \mathbb{E}\Big\{ J^{t+1}(\boldsymbol{X}(\boldsymbol{q})) \Big\} \Bigg\}, \qquad (2)
\end{aligned}$$

with the boundary condition that $J^{T+1}=0$. In the dynamic programming formulation above, we implicitly assume that even if $q_{i,0}=0$, meaning that we do not have any on-hand inventory for product i, we can offer an assortment that includes product i. Note the indicator function; if a customer chooses a product with zero on-hand inventory, then she leaves the system without renting any products. The possibility of offering products with zero on-hand inventory may be unrealistic in certain settings. Shortly in this section, in Assumption 2.1, we impose rather mild assumptions on the discrete choice model $\{\phi_i^t(S): i\in\mathcal{N},\ S\subseteq\mathcal{N}\}$ and the set of feasible decisions $\mathcal{F}$ to ensure that the optimal policy never offers a product with zero on-hand inventory, even if we are allowed to do so. So, under Assumption 2.1, it follows that the dynamic programming formulation above is equivalent to a dynamic programming formulation that explicitly imposes a constraint to ensure that we must have non-zero on-hand inventory for each product that we offer.

In the dynamic program in (2), the term $\sum_{i\in\mathcal{N}} \pi_i^t \sum_{\ell=1}^{\infty} q_{i,\ell}$ captures the rent payments from customers with already rented units at the beginning of time period t. On the other hand, the term $r_i^t + \pi_i^t + \mathbb{E}\Big\{ \mathsf{Z}(\rho_{i,0})\, J^{t+1}(\boldsymbol{X}(\boldsymbol{q})) + (1-\mathsf{Z}(\rho_{i,0}))\, J^{t+1}(\boldsymbol{X}(\boldsymbol{q}) - \boldsymbol{e}_{i,0} + \boldsymbol{e}_{i,1}) \Big\}$ corresponds to the expected revenue from a customer who selects product i at time period t. Here, $r_i^t + \pi_i^t$ reflects the one-time upfront payment and the per-period rent for the first rental period. Noting the definition of the hazard rate, we have $\rho_{i,0} = f_i(1)$. Therefore, the Bernoulli random variable $\mathsf{Z}(\rho_{i,0})$ takes a value of 1 if and only if the customer renting a unit of product i at time period t uses the product for exactly one time period. If $\mathsf{Z}(\rho_{i,0}) = 1$, then the unit is returned to the firm at the end of time period t, in which case, the state at the beginning of time period $t+1$ is $\boldsymbol{X}(\boldsymbol{q})$, which is identical to the state that we would have obtained when no rentals were made at time period t. On the other hand, if $\mathsf{Z}(\rho_{i,0}) = 0$, then the selected unit of product i will not be returned at the end of time period t. In this case, the selected unit of product i will not be on-hand at the beginning of time period $t+1$; instead, this unit will be in use for exactly one time period. Therefore, the state of the system at the beginning of time period $t+1$ will be $\boldsymbol{X}(\boldsymbol{q}) - \boldsymbol{e}_{i,0} + \boldsymbol{e}_{i,1}$. To simplify our dynamic programming formulation, note that since the rental durations of different units are independent of each other, $\boldsymbol{X}(\boldsymbol{q})$ and $\mathsf{Z}(\rho_{i,0})$ are independent of each other as well. Therefore, we obtain

$$\begin{aligned}
&\mathbb{E}\Big\{\mathsf{Z}(\rho_{i,0})\, J^{t+1}(\boldsymbol{X}(\boldsymbol{q})) + (1-\mathsf{Z}(\rho_{i,0}))\, J^{t+1}\left(\boldsymbol{X}(\boldsymbol{q}) - \boldsymbol{e}_{i,0} + \boldsymbol{e}_{i,1}\right)\Big\} - \mathbb{E}\Big\{J^{t+1}(\boldsymbol{X}(\boldsymbol{q}))\Big\} \\
&\quad = \rho_{i,0}\, \mathbb{E}\Big\{J^{t+1}\left(\boldsymbol{X}(\boldsymbol{q})\right)\Big\} + (1-\rho_{i,0})\, \mathbb{E}\Big\{J^{t+1}\left(\boldsymbol{X}(\boldsymbol{q}) - \boldsymbol{e}_{i,0} + \boldsymbol{e}_{i,1}\right)\Big\} - \mathbb{E}\Big\{J^{t+1}(\boldsymbol{X}(\boldsymbol{q}))\Big\} \\
&\quad = -(1-\rho_{i,0})\, \mathbb{E}\Big\{J^{t+1}\left(\boldsymbol{X}(\boldsymbol{q})\right) - J^{t+1}\left(\boldsymbol{X}(\boldsymbol{q}) - \boldsymbol{e}_{i,0} + \boldsymbol{e}_{i,1}\right)\Big\},
\end{aligned}$$

in which case, simply by arranging the terms, we can write the dynamic programming formulation in (2) equivalently as

$$\begin{aligned}
J^t(\boldsymbol{q}) = &\sum_{i \in \mathcal{N}} \pi_i^t \sum_{\ell=1}^{\infty} q_{i,\ell} \;+\; \mathbb{E}\Big\{J^{t+1}(\boldsymbol{X}(\boldsymbol{q}))\Big\} \\
&+ \max_{S \in \mathcal{F}} \left\{ \sum_{i \in \mathcal{N}} \mathbb{1}_{\{q_{i,0} \ge 1\}}\, \phi_i^t(S) \left(r_i^t + \pi_i^t - (1-\rho_{i,0})\, \mathbb{E}\Big\{J^{t+1}\left(\boldsymbol{X}(\boldsymbol{q})\right) - J^{t+1}\left(\boldsymbol{X}(\boldsymbol{q}) - \boldsymbol{e}_{i,0} + \boldsymbol{e}_{i,1}\right)\Big\} \right) \right\}. \quad (3)
\end{aligned}$$

Note that $J^{t+1}\left(\boldsymbol{X}(\boldsymbol{q})\right) - J^{t+1}\left(\boldsymbol{X}(\boldsymbol{q}) - \boldsymbol{e}_{i,0} + \boldsymbol{e}_{i,1}\right)$ captures the marginal value of renting one unit of product i to the customer at time period t.

Throughout the paper, we impose a mild assumption on the discrete choice model $\{\phi_i^t(S) : i \in \mathcal{N},\ S \subseteq \mathcal{N}\}$ and the set of feasible decisions $\mathcal{F}$ to ensure that the optimal policy never offers a product with zero on-hand inventory. This assumption is given below.

Assumption 2.1 (Substitutability and Feasibility) *Adding more products to an assortment does not increase the selection probability; that is, for all $S \subseteq \mathcal{N}$ and $k \in \mathcal{N}$, $\phi_i^t(S \cup \{k\}) \le \phi_i^t(S)$ for all $i \in S$. In addition, if a set of products is feasible to offer, then so are all of its subsets; that is, if $A \in \mathcal{F}$, then $S \in \mathcal{F}$ for all $S \subseteq A$.*

The first assumption ensures that products are substitutable, and thus, the probability of choosing any product never increases if more options become available. This assumption is rather mild and it holds for all choice models that satisfy the random utility maximization principle, including the multinomial logit, nested logit, d-level logit, paired combinatorial logit, and many others. In addition, the second assumption on the collection of feasible subsets $\mathcal{F}$ also holds for a broad class of assortment constraints, such as a shelf-space constraint $\mathcal{F} = \big\{S \subseteq \mathcal{N} : \sum_{i \in S} c_i \leq B\big\}$, where c_i is the space consumed by product i and B is the total shelf-space available. Under the assumption above, it is not difficult to see that the optimal policy never offers a product with zero on-hand inventory. In the maximization problem in (3), the profit contribution of product i is $\mathbb{1}_{\{q_{i,0} \geq 1\}} \times (r_i^t + \pi_i^t - (1-\rho_{i,0})\, \mathbb{E}\{J^{t+1}(\boldsymbol{X}(\boldsymbol{q})) - J^{t+1}(\boldsymbol{X}(\boldsymbol{q}) - \boldsymbol{e}_{i,0} + \boldsymbol{e}_{i,1})\})$. Let S^* be an optimal solution to this maximization problem. In this case, observe that we can drop all products with non-positive profit contributions from S^* without degrading the objective value of the maximization problem in (3) because if we drop such products, then by the substitutability assumption, the selection probabilities of all other products increase, whereas by the feasibility assumption, the subset we obtain remains feasible. Thus, the new subset that we obtain in this fashion provides an objective value to the maximization problem in (3) that is at least as large as that provided by S^*. Because the profit contribution of product i is zero when $\mathbb{1}_{\{q_{i,0} \geq 1\}} = 0$, there exists an optimal policy that never offers a product with zero on-hand inventory.

Because all products are available at the beginning of the selling horizon, the optimal total expected revenue is $J^1\left(\sum_{i \in \mathcal{N}} C_i\, \boldsymbol{e}_{i,0}\right)$. One source of difficulty in computing the optimal value functions $\{J^t : t \in \mathcal{T}\}$ is that the maximization problem in (3) is a combinatorial optimization problem that chooses the set of products to offer. However, there exist efficient algorithms to solve this problem under many discrete choice models, including the multinomial logit (Talluri and van Ryzin 2004, Rusmevichientong et al. 2014), nested logit (Davis et al. 2014, Gallego and Topaloglu 2014), d-level logit (Li et al. 2015), and paired combinatorial logit (Zhang et al. 2017), and under many different types of feasible sets $\mathcal{F}$ (Davis et al. 2013, Feldman and Topaloglu 2015, Desire et al. 2016). Later in the paper, we will also discuss how our results can be extended when we can only approximately solve the maximization problem in (3).

Although we can solve the maximization problem in (3) efficiently, to find the optimal policy, we need to compute the optimal value function $J^t(\boldsymbol{q})$ for each $\boldsymbol{q} \in \mathcal{Q}$ and $t \in \mathcal{T}$. The number of possible states $|\mathcal{Q}|$ grows exponentially with n and T, which makes it difficult to find the optimal policy. Thus, throughout the rest of the paper, we focus on developing approximate policies that are efficient to compute and have provable performance guarantees.

3. Linear Value Function Approximations

We develop an approach to construct linear approximations to the optimal value functions and analyze the performance of a policy that uses these approximations. In particular, we give a tractable recursion to come up with linear value function approximations. We show that if we use the greedy policy with respect to these linear value function approximations, then we obtain a policy that is guaranteed to obtain at least 50% of the optimal total expected revenue.

3.1 Specification of Linear Value Function Approximations

We consider an approximation $\hat{J}^t$ to the optimal value function J^t given by

$$\hat{J}^t(\boldsymbol{q}) \;=\; \sum_{i \in \mathcal{N}} \sum_{\ell=0}^{\infty} \hat{\nu}_{i,\ell}^t \, q_{i,\ell}, \tag{4}$$

where, for $\ell \geq 1$, the parameter $\hat{\nu}_{i,\ell}^t$ represents the marginal value at time period t of each unit of product i that has been in use for ℓ periods, whereas the parameter $\hat{\nu}_{i,0}^t$ represents the marginal value of each unit of product i that is currently available as on-hand inventory at time period t. We propose computing $\hat{\nu}_{i,\ell}^t$ recursively as follows.

- **Initialization:** Set $\hat{\nu}_{i,\ell}^{T+1} = 0$ for all $i \in \mathcal{N}$, $\ell \geq 0$.
- **Backward Recursion:** For $t = T, T-1, \ldots, 1$, we compute $\hat{\nu}_{i,\ell}^t$ by using $\{\hat{\nu}_{i,\ell}^{t+1} : i \in \mathcal{N},\ \ell \geq 0\}$ as follows. Let $\hat{A}^t \in \mathcal{F}$ be an assortment such that

$$\hat{A}^t \;=\; \arg\max_{S \in \mathcal{F}} \sum_{i \in \mathcal{N}} \phi_i^t(S) \Big[r_i^t + \pi_i^t - (1 - \rho_{i,0}) \left(\hat{\nu}_{i,0}^{t+1} - \hat{\nu}_{i,1}^{t+1} \right) \Big]. \tag{5}$$

Once $\hat{A}^t$ is computed, for each $i \in \mathcal{N}$, let

$$\begin{aligned} \hat{\nu}_{i,0}^t \;&=\; \hat{\nu}_{i,0}^{t+1} + \frac{1}{C_i} \phi_i^t(\hat{A}^t) \Big[r_i^t + \pi_i^t - (1 - \rho_{i,0}) \left(\hat{\nu}_{i,0}^{t+1} - \hat{\nu}_{i,1}^{t+1} \right) \Big] \\ \hat{\nu}_{i,\ell}^t \;&=\; \pi_i^t \;+\; \rho_{i,\ell} \, \hat{\nu}_{i,0}^{t+1} + (1 - \rho_{i,\ell}) \, \hat{\nu}_{i,\ell+1}^{t+1} \qquad \forall \, \ell = 1, 2, \ldots. \end{aligned} \tag{6}$$

The above description completes the specification of the approximate value function $\hat{J}^t$. We shortly give the intuition behind our approach. Because we start the system with all units available as on-hand inventory, no unit will be in use for more than T time periods. Thus, we only need to compute $\hat{\nu}_{i,\ell}^t$ for $\ell = 0, 1, \ldots, T$, so we can execute the above recursion in finite time.

We provide some intuition into the computation of $\hat{A}^t$. Intuitively speaking, we can interpret $\hat{A}^t$ as an ideal assortment to offer at time period t under the linear value function approximations when we ignore inventory availability. In particular, if we replace the value function J^{t+1} in the maximization problem on the right side of (3) with the linear approximation $\hat{J}^{t+1}(\boldsymbol{q}) = \sum_{i \in \mathcal{N}} \sum_{\ell=0}^{\infty} \hat{\nu}_{i,\ell}^{t+1} \, q_{i,\ell}$ and drop the indicator function $\mathbb{1}_{\{q_{i,0} \geq 1\}}$ to ignore inventory availability, then the objective function of this maximization problem takes the form $\sum_{i \in \mathcal{N}} \phi_i^t(S) \, [r_i^t + \pi_i^t - (1 - \rho_{i,0}) \left(\hat{\nu}_{i,0}^{t+1} - \hat{\nu}_{i,1}^{t+1} \right)]$, which is the same as the objective function of the maximization problem in (5). Next, we provide some intuition into the computation of $\hat{\nu}^t_{i,0}$, which measures the value of a unit of on-hand inventory for product i at time period t. Roughly speaking, assume that we offer the ideal assortment $\hat{A}^t$ at time period t, and if a customer selects product i at time period t, then we "direct" the customer to one of the C_i copies of product i with equal probability of $1/C_i$. In this case, the probability that a unit of product i "sees" a demand at time period t is $\phi^t_i(\hat{A}^t)\,\frac{1}{C_i}$. We write the recursion that we use to compute $\hat{\nu}^t_{i,0}$ in (6) equivalently as

$$\hat{\nu}^t_{i,0} \;=\; \frac{1}{C_i}\phi^t_i(\hat{A}^t)\Big[r^t_i + \pi^t_i + \rho_{i,0}\,\hat{\nu}^{t+1}_{i,0} + (1-\rho_{i,0})\,\hat{\nu}^{t+1}_{i,1}\Big] + \left(1 - \frac{1}{C_i}\,\phi^t_i(\hat{A}^t)\right)\hat{\nu}^{t+1}_{i,0}.$$

On the left side above, $\hat{\nu}^t_{i,0}$ is the value of a unit of product i on-hand at time period t. If we offer the ideal assortment $\hat{A}^t$ at time period t, then a unit of product i "sees" a demand with probability $\frac{1}{C_i}\phi^t_i(\hat{A}^t)$. In this case, we collect the upfront fee r^t_i and the rent π^t_i for the first time period. As discussed earlier, with probability $\rho_{i,0} = f_i(1)$, the customer rents product i for exactly one time period, in which case, she returns the product by the beginning of time period $t+1$. The value of a unit of on-hand inventory of product i at time period $t+1$ is $\hat{\nu}^{t+1}_{i,0}$. With probability $1-\rho_{i,0}$, the customer rents product i for more than one time period, in which case, the product will have been rented out at the beginning of time period $t+1$ for exactly one period. The value of a unit of product i at time period $t+1$ that has been in use for one period is $\hat{\nu}^{t+1}_{i,1}$. This discussion provides the intuition for the term $r^t_i + \pi^t_i + \rho_{i,0}\,\hat{\nu}^{t+1}_{i,0} + (1-\rho_{i,0})\,\hat{\nu}^{t+1}_{i,1}$ on the right side above. With probability $1 - \frac{1}{C_i}\phi^t_i(\hat{A}^t)$, the unit of product i does not "see" a demand, in which case this unit is still available at time period $t+1$ and the value of this unit is given by $\hat{\nu}^{t+1}_{i,0}$.

We can give a similar intuition for the recursion that is used to compute $\hat{\nu}^t_{i,\ell}$ for all $\ell = 1, 2, \ldots$. Noting the recursion $\hat{\nu}^t_{i,\ell} = \pi^t_i + \rho_{i,\ell}\,\hat{\nu}^{t+1}_{i,0} + (1-\rho_{i,\ell})\,\hat{\nu}^{t+1}_{i,\ell+1}$, recall that $\hat{\nu}^t_{i,\ell}$ on the left side is the value of a unit of product i that has been in use for ℓ periods at time period t. This unit of product i will certainly be used until the end of time period t and we will obtain the rental fee of π^t_i. Furthermore, by the definition of the hazard rate $\rho_{i,\ell}$, a unit of product i that has been in use for ℓ periods at time period t will be returned by the beginning of the next time period with probability $\rho_{i,\ell}$, in which case, the value of this on-hand unit at time period $t+1$ is $\hat{\nu}^{t+1}_{i,0}$, yielding the term $\rho_{i,\ell}\,\hat{\nu}^{t+1}_{i,0}$ on the right side. On the other hand, once again, by the definition of the hazard rate $\rho_{i,\ell}$, a unit of product i that has been in use for ℓ periods at time period t will not be returned by the beginning of time period $t+1$ with probability $1-\rho_{i,\ell}$. Therefore, this unit of product i will have been used for $\ell+1$ periods at the next time period and the value of this unit at time period $t+1$ is $\hat{\nu}^{t+1}_{i,\ell+1}$, yielding the term $(1-\rho_{i,\ell})\,\hat{\nu}^{t+1}_{i,\ell+1}$ on the right side.

The discussion in the previous two paragraphs also provides a natural interpretation for our value function approximations. In particular, our value function approximations correspond to the total expected revenue that we obtain when we use a policy that manages each unit of product i independently. Focus on one particular unit of product i. At time period t, we always offer the assortment $\hat{A}^t$, in which case, an arriving customer selects product i with probability $\phi_i^t(\hat{A}^t)$. If the customer selects product i, then we "direct" the customer to each unit of product i with equal probability of $1/C_i$. In this case, the unit of product i that we focus on "sees" a demand at time period t with probability $\phi_i^t(\hat{A}^t)\frac{1}{C_i}$. If the unit of product i that we focus on "sees" a demand, then it is rented, so we collect the upfront fee of r_i^t and the rent of π_i^t for the first time period. A unit that is rented stays in use for a random duration of time that is governed by the hazard rates $\{\rho_{i,\ell} : \ell \geq 0\}$, during which we collect the rent of π_i^t at each time period t that the unit is in use. After the usage duration has expired, the unit is returned. By the discussion in the previous two paragraphs, if we use a policy that manages each unit of product i independently in the way we just described, then $\hat{\nu}_{i,\ell}^t$ corresponds to the total expected revenue from a unit of product i that has been in use for exactly ℓ time periods at the beginning of time period t. Therefore, our value function approximations correspond to the value functions of a policy that manages each unit independently. Clearly, this policy does not pool the units of the same product together, so we certainly do not advocate using such a policy in practice. We will only use the value functions of this policy to construct value function approximations. It turns out that the greedy policy with respect to the value function approximations will have a performance guarantee.

Considering the effort to compute the parameters $\{\hat{\nu}_{i,\ell}^t : i \in \mathcal{N},\ \ell = 0, \dots, T,\ t \in \mathcal{T}\}$, we need to solve problem (5) for each time period $t \in \mathcal{T}$. The number of operations to solve this problem depends on the underlying choice model. We use Opt to denote the number of operations to solve one instance of problem (5). Next, we need to compute $\{\phi_i^t(\hat{A}^t) : i \in \mathcal{N},\ t \in \mathcal{T}\}$. The number of operations to compute these choice probabilities also depends on the underlying choice model. We use Prob to denote the number of operations to compute $\{\phi_i^t(S) : i \in \mathcal{N}\}$ for a fixed subset S and time period t. Once we compute $\{\phi_i^t(\hat{A}^t) : i \in \mathcal{N},\ t \in \mathcal{T}\}$, we can use (6) to compute each one of the parameters $\{\hat{\nu}_{i,\ell}^t : i \in \mathcal{N},\ \ell = 0, \dots, T,\ t \in \mathcal{T}\}$ in $O(1)$ operations. Thus, noting that there are $O(T^2 n)$ such parameters, we can compute all of the parameters $\{\hat{\nu}_{i,\ell}^t : i \in \mathcal{N},\ \ell = 0, \dots, T,\ t \in \mathcal{T}\}$ in $O(T \times \mathsf{Opt} + T \times \mathsf{Prob} + T^2 n)$ operations. For example, if the customers choose according to the multinomial logit model, then we can solve one instance of problem (5) in $O(n \log n)$ operations (Talluri and van Ryzin 2004). Also, for a fixed subset S and time period t, we can compute $\{\phi_i^t(S) : i \in \mathcal{N}\}$ in $O(n)$ operations. In this case, we can compute all of the parameters $\{\hat{\nu}_{i,\ell}^t : i \in \mathcal{N},\ \ell = 0, \dots, T,\ t \in \mathcal{T}\}$ in $O(Tn \log n + T^2 n)$ operations.

Lastly, although we use linear value function approximations, it is not difficult to see that the optimal value functions are not even separable by the products. In Appendix A, we give a problem instance with only one time period in the selling horizon, in which the optimal value functions are not separable by the products. We close this section with the next lemma, where we show that the marginal value of a unit of on-hand inventory becomes smaller as the end of the selling horizon approaches. We will use this property several times throughout the paper.

Lemma 3.1 (Properties of the Marginal Values) *The marginal value of on-hand inventory decreases over time; that is,* $\hat{\nu}^t_{i,0} \geq \hat{\nu}^{t+1}_{i,0}$ *for all* $t \in \mathcal{T}$ *and* $i \in \mathcal{N}$.

Proof: For notational brevity, let $\Delta^t_i = r^t_i + \pi^t_i - (1-\rho_{i,0})\left(\hat{\nu}^{t+1}_{i,0} - \hat{\nu}^{t+1}_{i,1}\right)$. We will shortly show the claim that $\phi^t_i(\hat{A}^t)\,\Delta^t_i \geq 0$ for all $i \in \mathcal{N}$. In this case, by the recursion in (6) that we use to compute $\hat{\nu}^t_{i,0}$, we have $\hat{\nu}^t_{i,0} = \hat{\nu}^{t+1}_{i,0} + \frac{1}{C_i}\,\phi^t_i(\hat{A}^t)\,\Delta^t_i \geq \nu^{t+1}_{i,0}$, which is the desired result. To see the claim that $\phi^t_i(\hat{A}^t)\,\Delta^t_i \geq 0$ for all $i \in \mathcal{N}$, assume on the contrary that there exists some $k \in \mathcal{N}$ such that $\phi^t_k(\hat{A}^t)\,\Delta^t_k < 0$. Let $\mathcal{N}^+ = \{i \in \mathcal{N} : \Delta^t_i \geq 0\}$ and $\mathcal{N}^- = \{i \in \mathcal{N} : \Delta^t_i < 0\}$. By our assumption, there exists some $k \in \mathcal{N}^-$ such that $\phi^t_k(\hat{A}^t) > 0$. Furthermore, by Assumption 2.1, $\phi^t_i(\hat{A}^t \cap \mathcal{N}^+) \geq \phi^t_i(\hat{A}^t)$ for all $i \in \hat{A}^t \cap \mathcal{N}^+$. By the same assumption, because $\hat{A}^t \in \mathcal{F}$, we have $\hat{A}^t \cap \mathcal{N}^+ \in \mathcal{F}$. So, we get

$$\begin{aligned}\sum_{i\in\mathcal{N}} \phi^t_i(\hat{A}^t)\,\Delta^t_i &= \sum_{i\in\mathcal{N}^+} \phi^t_i(\hat{A}^t)\,\Delta^t_i + \sum_{i\in\mathcal{N}^-} \phi^t_i(\hat{A}^t)\,\Delta^t_i < \sum_{i\in\mathcal{N}^+} \phi^t_i(\hat{A}^t)\,\Delta^t_i \\ &= \sum_{i\in\hat{A}^t\cap\mathcal{N}^+} \phi^t_i(\hat{A}^t)\,\Delta^t_i \leq \sum_{i\in\hat{A}^t\cap\mathcal{N}^+} \phi^t_i(\hat{A}^t\cap\mathcal{N}^+)\,\Delta^t_i = \sum_{i\in\mathcal{N}} \phi^t_i(\hat{A}^t\cap\mathcal{N}^+)\,\Delta^t_i,\end{aligned}$$

where the first inequality follows because there exists some $k \in \mathcal{N}^-$ such that $\phi^t_k(\hat{A}^t) > 0$, the second equality holds since $\phi^t_i(\hat{A}^t) = 0$ for all $i \notin \hat{A}^t$, and the second inequality uses the fact that $\phi^t_i(\hat{A}^t \cap \mathcal{N}^+) \geq \phi^t_i(\hat{A}^t)$ for all $i \in \hat{A}^t \cap \mathcal{N}^+$. Since $\hat{A}^t \cap \mathcal{N}^+ \in \mathcal{F}$, the chain of inequalities above contradicts the fact that $\hat{A}^t$ is an optimal solution to problem (5). ■

3.2 An Approximate Policy Using Marginal Values

We consider the greedy policy with respect to the value function approximations $\{\hat{J}^t : t \in \mathcal{T}\}$. If the system is in state $\boldsymbol{q}$ at time period t, then this policy offers the assortment $\hat{S}^t(\boldsymbol{q})$ given by

$$\begin{aligned}\hat{S}^t(\boldsymbol{q}) &= \arg\max_{S\in\mathcal{F}} \left\{ \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\geq 1\}} \phi^t_i(S) \Big[r^t_i + \pi^t_i - (1-\rho_{i,0})\, \mathbb{E}\Big\{ \hat{J}^{t+1}(\boldsymbol{X}(\boldsymbol{q})) - \hat{J}^{t+1}(\boldsymbol{X}(\boldsymbol{q}) - \boldsymbol{e}_{i,0} + \boldsymbol{e}_{i,1}) \Big\} \Big] \right\} \\ &= \arg\max_{S\in\mathcal{F}} \sum_{i=1}^{n} \mathbb{1}_{\{q_{i,0}\geq 1\}} \phi^t_i(S) \Big[r^t_i + \pi^t_i - (1-\rho_{i,0}) \left(\hat{\nu}^{t+1}_{i,0} - \hat{\nu}^{t+1}_{i,1}\right) \Big], \end{aligned} \tag{7}$$

where the second equality uses the definition of the value function approximations in (4). The next theorem is the main result of this section, giving a performance guarantee for this policy.

Theorem 3.2 (Performance of the Greedy Policy) *The total expected revenue of the greedy policy with respect to the value function approximations $\{\hat{J}^t : t \in \mathcal{T}\}$ is at least 50% of the optimal total expected revenue; that is, this policy is a half-approximation.*

The proof of this theorem makes use of the next lemma. Because we do not have any products in use at the beginning of the selling horizon, the initial state is $\sum_{i\in\mathcal{N}} C_i\, \boldsymbol{e}_{i,0}$. The next lemma relates the approximation $\hat{J}^1(\sum_{i\in\mathcal{N}} C_i\, \boldsymbol{e}_{i,0})$ to the optimal total expected revenue $J^1(\sum_{i\in\mathcal{N}} C_i\, \boldsymbol{e}_{i,0})$.

Lemma 3.3 (Expected Revenue Upper Bound) $J^1(\sum_{i\in\mathcal{N}} C_i\, \boldsymbol{e}_{i,0}) \le 2\, \hat{J}^1(\sum_{i\in\mathcal{N}} C_i\, \boldsymbol{e}_{i,0})$.

Proof: By Adelman (2007), we can obtain an upper bound on the optimal total expected revenue by using the objective value provided by any feasible solution to the linear program

$$
\begin{aligned}
\min \quad & \tilde{J}^1\left(\sum_{i\in\mathcal{N}} C_i \boldsymbol{e}_{i,0}\right) \\
\text{s.t.} \quad & \tilde{J}^t(\boldsymbol{q}) \;\ge\; \sum_{i\in\mathcal{N}} \pi_i^t \sum_{\ell=1}^{\infty} q_{i,\ell} \;+\; \mathbb{E}\left\{\tilde{J}^{t+1}(\boldsymbol{X}(\boldsymbol{q}))\right\} \\
& \qquad + \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\ge 1\}}\, \phi_i^t(S) \left[r_i^t + \pi_i^t - (1-\rho_{i,0})\, \mathbb{E}\left\{\tilde{J}^{t+1}\left(\boldsymbol{X}(\boldsymbol{q})\right) - \tilde{J}^{t+1}\left(\boldsymbol{X}(\boldsymbol{q}) - \boldsymbol{e}_{i,0} + \boldsymbol{e}_{i,1}\right)\right\}\right] \\
& \hspace{22em} \forall \boldsymbol{q} \in \mathcal{Q},\; S \in \mathcal{F},\; t \in \mathcal{T},
\end{aligned}
$$

where the decision variables are $\{\tilde{J}^t(\boldsymbol{q}) : \boldsymbol{q} \in \mathcal{Q},\; t \in \mathcal{T}\}$ and we follow the convention that $\tilde{J}^{T+1} = 0$. Define the constant $\hat{\beta}^t = \sum_{i\in\mathcal{N}} \hat{\nu}_{i,0}^t\, C_i$. We proceed to show that $\{\hat{\beta}^t + \hat{J}^t(\boldsymbol{q}) : \boldsymbol{q} \in \mathcal{Q},\; t \in \mathcal{T}\}$ with $\hat{J}^t(\boldsymbol{q})$ as in (4)-(6) is a feasible solution to the linear program above. (Without the constant $\hat{\beta}^t$, the solution $\{\hat{J}^t(\boldsymbol{q}) : \boldsymbol{q} \in \mathcal{Q},\; t \in \mathcal{T}\}$ is not necessarily feasible to the linear program.) As all units are available at the beginning of the selling horizon, a unit will never be in use for more than T time periods. Thus, we can assume that $\mathcal{Q}$ is a finite set, so the numbers of decision variables and constraints above are finite. By the definitions of $\hat{J}^{t+1}(\boldsymbol{q})$ in (4) and $\boldsymbol{X}(\boldsymbol{q})$ in (1), we get

$$
\begin{aligned}
\hat{\beta}^{t+1} + \mathbb{E}\left\{\hat{J}^{t+1}(\boldsymbol{X}(\boldsymbol{q}))\right\} &= \hat{\beta}^{t+1} + \sum_{i\in\mathcal{N}} \left\{ \hat{\nu}_{i,0}^{t+1}\left[q_{i,0} + \sum_{\ell=1}^{\infty} \rho_{i,\ell}\, q_{i,\ell}\right] + \sum_{\ell=1}^{\infty} \hat{\nu}_{i,\ell+1}^{t+1}\left[q_{i,\ell} - \rho_{i,\ell}\, q_{i,\ell}\right] \right\} \\
&= \hat{\beta}^{t+1} + \sum_{i\in\mathcal{N}} \left\{ q_{i,0}\, \hat{\nu}_{i,0}^{t+1} + \sum_{\ell=1}^{\infty} q_{i,\ell}\left[\rho_{i,\ell}\, \hat{\nu}_{i,0}^{t+1} + (1-\rho_{i,\ell})\, \hat{\nu}_{i,\ell+1}^{t+1}\right] \right\}.
\end{aligned}
$$

Similarly, $\mathbb{E}\left\{\hat{J}^{t+1}\left(\boldsymbol{X}(\boldsymbol{q})\right) - \hat{J}^{t+1}\left(\boldsymbol{X}(\boldsymbol{q}) - \boldsymbol{e}_{i,0} + \boldsymbol{e}_{i,1}\right)\right\} = \hat{\nu}_{i,0}^{t+1} - \hat{\nu}_{i,1}^{t+1}$. So, if we evaluate the right side of the constraint in the linear program above at $\{\hat{\beta}^t + \hat{J}^t(\boldsymbol{q}) : \boldsymbol{q} \in \mathcal{Q},\; t \in \mathcal{T}\}$, then we obtain

$$
\sum_{i\in\mathcal{N}} \pi_i^t \sum_{\ell=1}^{\infty} q_{i,\ell} \;+\; \hat{\beta}^{t+1} \;+\; \mathbb{E}\left\{\hat{J}^{t+1}(\boldsymbol{X}(\boldsymbol{q}))\right\}
$$

$$
\begin{aligned}
&+ \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\geq 1\}}\, \phi_i^t(S) \left[r_i^t + \pi_i^t - (1-\rho_{i,0})\, \mathbb{E}\left\{ \hat\beta^{t+1} + \hat J^{t+1}(\boldsymbol{X}(\boldsymbol{q})) - \hat\beta^{t+1} - \hat J^{t+1}(\boldsymbol{X}(\boldsymbol{q}) - \boldsymbol{e}_{i,0} + \boldsymbol{e}_{i,1}) \right\}\right] \\
=\;& \sum_{i\in\mathcal{N}} \pi_i^t \sum_{\ell=1}^\infty q_{i,\ell} + \hat\beta^{t+1} + \sum_{i\in\mathcal{N}} \left\{ q_{i,0}\, \hat\nu_{i,0}^{t+1} + \sum_{\ell=1}^\infty q_{i,\ell} \left[\rho_{i,\ell}\, \hat\nu_{i,0}^{t+1} + (1-\rho_{i,\ell})\, \hat\nu_{i,\ell+1}^{t+1} \right] \right\} \\
&\qquad + \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\geq 1\}}\, \phi_i^t(S) \left[r_i^t + \pi_i^t - (1-\rho_{i,0})\, (\hat\nu_{i,0}^{t+1} - \hat\nu_{i,1}^{t+1}) \right] \\
=\;& \sum_{i\in\mathcal{N}} \hat\nu_{i,0}^{t+1}\, C_i + \sum_{i\in\mathcal{N}} \left\{ q_{i,0}\, \hat\nu_{i,0}^{t+1} + \sum_{\ell=1}^\infty q_{i,\ell} \hat\nu_{i,\ell}^t \right\} + \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\geq 1\}}\, \phi_i^t(S) \left[r_i^t + \pi_i^t - (1-\rho_{i,0})\, (\hat\nu_{i,0}^{t+1} - \hat\nu_{i,1}^{t+1}) \right],
\end{aligned}
$$

where the second equality holds because we have $\hat\nu_{i,\ell}^t = \pi_i^t + \rho_{i,\ell}\, \hat\nu_{i,0}^{t+1} + (1-\rho_{i,\ell})\, \hat\nu_{i,\ell+1}^{t+1}$ by (6) and $\hat\beta^{t+1} = \sum_{i\in\mathcal{N}} \hat\nu_{i,0}^{t+1}\, C_i$. By a simple lemma, given as Lemma B.1 in Appendix B, if we let $\Delta_i^t = r_i^t + \pi_i^t - (1-\rho_{i,0})\,(\hat\nu_{i,0}^{t+1} - \hat\nu_{i,1}^{t+1})$, then $\sum_{i\in\mathcal{N}} \phi_i^t(\hat A^t)\, \Delta_i^t \geq \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\geq 1\}}\, \phi_i^t(S)\, \Delta_i^t$ for all $S\in\mathcal{F}$. Note that this inequality does not follow from the definition of $\hat A^t$ because although we have $\sum_{i\in\mathcal{N}} \phi_i^t(\hat A^t)\, \Delta_i^t \geq \sum_{i\in\mathcal{N}} \phi_i^t(S)\, \Delta_i^t$ by (5), we may have $\Delta_i^t < \mathbb{1}_{\{q_{i,0}\geq 1\}}\, \Delta_i^t$ when $\Delta_i^t < 0$. Thus, using the chain of equalities above, we upper bound the right side of the constraint in the linear program as

$$
\begin{aligned}
&\sum_{i\in\mathcal{N}} \hat\nu_{i,0}^{t+1}\, C_i + \sum_{i\in\mathcal{N}} \left\{ q_{i,0}\, \hat\nu_{i,0}^{t+1} + \sum_{\ell=1}^\infty q_{i,\ell}\, \hat\nu_{i,\ell}^t \right\} + \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\geq 1\}}\, \phi_i^t(S) \left[r_i^t + \pi_i^t - (1-\rho_{i,0})\, (\hat\nu_{i,0}^{t+1} - \hat\nu_{i,1}^{t+1}) \right] \\
&\quad\leq \sum_{i\in\mathcal{N}} \hat\nu_{i,0}^{t+1}\, C_i + \sum_{i\in\mathcal{N}} \sum_{\ell=0}^\infty q_{i,\ell}\, \hat\nu_{i,\ell}^t + \sum_{i\in\mathcal{N}} \phi_i^t(\hat A^t) \left[r_i^t + \pi_i^t - (1-\rho_{i,0})\, (\hat\nu_{i,0}^{t+1} - \hat\nu_{i,1}^{t+1}) \right] \\
&\quad= \sum_{i\in\mathcal{N}} \hat\nu_{i,0}^{t+1}\, C_i + \sum_{i\in\mathcal{N}} \sum_{\ell=0}^\infty q_{i,\ell}\, \hat\nu_{i,\ell}^t + \sum_{i\in\mathcal{N}} C_i\, (\hat\nu_{i,0}^t - \hat\nu_{i,0}^{t+1}) \\
&\quad= \sum_{i\in\mathcal{N}} \hat\nu_{i,0}^t\, C_i + \sum_{i\in\mathcal{N}} \sum_{\ell=0}^\infty q_{i,\ell}\, \hat\nu_{i,\ell}^t \quad=\quad \hat\beta^t + \hat J^t(\boldsymbol{q}),
\end{aligned}
$$

where the first inequality holds as $\hat\nu_{i,0}^t \geq \hat\nu_{i,0}^{t+1}$ by Lemma 3.1, the first equality follows from (6) and the last equality is by the definition of $\hat\beta^t$. By the chain of inequalities above, for any $\boldsymbol{q}\in\mathcal{Q}$, $S\in\mathcal{F}$ and $t\in\mathcal{T}$, if we evaluate the right side of the constraint at $\{\hat\beta^t + \hat J^t(\boldsymbol{q}) : \boldsymbol{q}\in\mathcal{Q},\ t\in\mathcal{T}\}$, then the right side of the constraint is upper bounded by $\hat\beta^t + \hat J^t(\boldsymbol{q})$. So, the solution $\{\hat\beta^t + \hat J^t(\boldsymbol{q}) : \boldsymbol{q}\in\mathcal{Q},\ t\in\mathcal{T}\}$ is feasible to the linear program, which implies that the objective value of the linear program evaluated at this solution is an upper bound on the optimal total expected revenue. The objective value of the linear program evaluated at the solution $\{\hat\beta^t + \hat J^t(\boldsymbol{q}) : \boldsymbol{q}\in\mathcal{Q},\ t\in\mathcal{T}\}$ is $\hat\beta^1 + \hat J^1(\sum_{i\in\mathcal{N}} C_i\, \boldsymbol{e}_{i,0}) = \hat\beta^1 + \sum_{i\in\mathcal{N}} \hat\nu_{i,0}^1\, C_i = 2\, \sum_{i\in\mathcal{N}} \hat\nu_{i,0}^1\, C_i = 2\, \hat J^1(\sum_{i\in\mathcal{N}} C_i\, \boldsymbol{e}_{i,0})$. Thus, $2\, \hat J^1(\sum_{i\in\mathcal{N}} C_i\, \boldsymbol{e}_{i,0})$ is an upper bound on the optimal total expected revenue. ■

The greedy policy with respect to the value function approximations $\{\hat J^t : t\in\mathcal{T}\}$ offers the assortment $\hat S^t(\boldsymbol{q})$ in (7) when the system is in state $\boldsymbol{q}$ at time period t. Let $U^t(\boldsymbol{q})$ denote the total expected revenue under this greedy policy over the time periods $t,\ldots,T$, given that we are in state $\boldsymbol{q}$ at time period t. We can compute $\{U^t : t\in\mathcal{T}\}$ by using the recursion

$$\begin{aligned}U^t(\boldsymbol{q}) &= \sum_{i\in\mathcal{N}}\pi_i^t\sum_{\ell=1}^{\infty}q_{i,\ell}\\ &+\sum_{i\in\mathcal{N}}\mathbb{1}_{\{q_{i,0}\,\geq\,1\}}\phi_i^t(\hat{S}^t(\boldsymbol{q}))\Big(r_i^t+\pi_i^t+\mathbb{E}\Big\{\mathsf{Z}(\rho_{i,0})\,U^{t+1}(\boldsymbol{X}(\boldsymbol{q}))+(1-\mathsf{Z}(\rho_{i,0}))\,U^{t+1}\left(\boldsymbol{X}(\boldsymbol{q})-\boldsymbol{e}_{i,0}+\boldsymbol{e}_{i,1}\right)\Big\}\Big)\\ &\qquad+\Big(1-\sum_{i\in\mathcal{N}}\mathbb{1}_{\{q_{i,0}\,\geq\,1\}}\phi_i^t(\hat{S}^t(\boldsymbol{q}))\Big)\,\mathbb{E}\Big\{U^{t+1}(\boldsymbol{X}(\boldsymbol{q}))\Big\},\end{aligned}$$

with the boundary condition that $U^{T+1}=0$. In the recursion above, we use the same line of reasoning that we used for the dynamic programming formulation in (2), but the decision is fixed as $\hat{S}^t(\boldsymbol{q})$. An observation that will shortly be useful is that U^{t+1} appears with a positive coefficient on the right side above. Therefore, if we replace U^{t+1} with a function H^{t+1} that satisfies $U^{t+1}(\boldsymbol{q})\geq H^{t+1}(\boldsymbol{q})$, then the right side of the expression above becomes smaller. By using the same sequence of manipulations that we used to obtain the dynamic program in (3), we can write the above recursion equivalently as

$$\begin{aligned}U^t(\boldsymbol{q}) \;&=\; \sum_{i\in\mathcal{N}}\pi_i^t\sum_{\ell=1}^{\infty}q_{i,\ell}\;+\;\mathbb{E}\Big\{U^{t+1}(\boldsymbol{X}(\boldsymbol{q}))\Big\}\\ &+\;\sum_{i\in\mathcal{N}}\mathbb{1}_{\{q_{i,0}\,\geq\,1\}}\,\phi_i^t(\hat{S}^t(\boldsymbol{q}))\left(r_i^t+\pi_i^t-(1-\rho_{i,0})\,\mathbb{E}\Big\{U^{t+1}\left(\boldsymbol{X}(\boldsymbol{q})\right)-U^{t+1}\left(\boldsymbol{X}(\boldsymbol{q})-\boldsymbol{e}_{i,0}+\boldsymbol{e}_{i,1}\right)\Big\}\right).\end{aligned}\tag{8}$$

The coefficients of U^{t+1} are not necessarily all positive on the right side above, but the last two recursions are equivalent. So, if we replace U^{t+1} on the right side above with a function H^{t+1} that satisfies $U^{t+1}(\boldsymbol{q})\geq H^{t+1}(\boldsymbol{q})$, then the right side of the expression above still gets smaller.

Here is the proof of Theorem 3.2.

Proof of Theorem 3.2: We will use induction over the time periods to show that $U^t(\boldsymbol{q})\geq\hat{J}^t(\boldsymbol{q})$ for all $\boldsymbol{q}\in\mathcal{Q}$ and $t\in\mathcal{T}$, where $\hat{J}^t(\boldsymbol{q})$ is as in (4). By definition, we have $\hat{\nu}_{i,\ell}^{T+1}=0$ for all $i\in\mathcal{N}$, $\ell=0,1,\ldots$, so that $\hat{J}^{T+1}=0$. Furthermore, we have $U^{T+1}=0$. Thus, the result holds at time period $T+1$. Assuming that $U^{t+1}(\boldsymbol{q})\geq\hat{J}^{t+1}(\boldsymbol{q})$ for all $\boldsymbol{q}\in\mathcal{Q}$, we proceed to show that $U^t(\boldsymbol{q})\geq\hat{J}^t(\boldsymbol{q})$ for all $\boldsymbol{q}\in\mathcal{Q}$. Using the same argument in the proof of Lemma 3.3, we have

$$\mathbb{E}\Big\{\hat{J}^{t+1}(\boldsymbol{X}(\boldsymbol{q}))\Big\}=\sum_{i\in\mathcal{N}}\left\{q_{i,0}\,\hat{\nu}_{i,0}^{t+1}+\sum_{\ell=1}^{\infty}q_{i,\ell}\left[\rho_{i,\ell}\,\hat{\nu}_{i,0}^{t+1}+(1-\rho_{i,\ell})\,\hat{\nu}_{i,\ell+1}^{t+1}\right]\right\}.$$

Similarly, we have $\mathbb{E}\Big\{\hat{J}^{t+1}(\boldsymbol{X}(\boldsymbol{q}))-\hat{J}^{t+1}\left(\boldsymbol{X}(\boldsymbol{q})-\boldsymbol{e}_{i,0}+\boldsymbol{e}_{i,1}\right)\Big\}=\hat{\nu}_{i,0}^{t+1}-\hat{\nu}_{i,1}^{t+1}$. Thus, by the inductive hypothesis and the recursion defining $U^t(\boldsymbol{q})$ in (8), we obtain

$$U^t(\boldsymbol{q}) \;\;\geq\;\; \sum_{i\in\mathcal{N}}\pi_i^t\sum_{\ell=1}^{\infty}q_{i,\ell}\;+\;\mathbb{E}\Big\{\hat{J}^{t+1}(\boldsymbol{X}(\boldsymbol{q}))\Big\}$$

$$
\begin{aligned}
& + \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\,\geq 1\}}\, \phi_i^t(\hat{S}^t(\boldsymbol{q}))\Big(r_i^t + \pi_i^t - (1-\rho_{i,0})\,\mathbb{E}\Big\{\hat{J}^{t+1}\left(\boldsymbol{X}(\boldsymbol{q})\right) - \hat{J}^{t+1}\left(\boldsymbol{X}(\boldsymbol{q}) - \boldsymbol{e}_{i,0} + \boldsymbol{e}_{i,1}\right)\Big\}\Big) \\
= \; & \sum_{i\in\mathcal{N}} \pi_i^t \sum_{\ell=1}^{\infty} q_{i,\ell} \; + \; \sum_{i\in\mathcal{N}} \left\{ q_{i,0}\,\hat{\nu}_{i,0}^{t+1} + \sum_{\ell=1}^{\infty} q_{i,\ell}\left[\rho_{i,\ell}\,\hat{\nu}_{i,0}^{t+1} + (1-\rho_{i,\ell})\,\hat{\nu}_{i,\ell+1}^{t+1}\right]\right\} \\
& + \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\,\geq 1\}}\, \phi_i^t(\hat{S}^t(\boldsymbol{q}))\left[r_i^t + \pi_i^t - (1-\rho_{i,0})(\hat{\nu}_{i,0}^{t+1} - \hat{\nu}_{i,1}^{t+1})\right] \\
= \; & \sum_{i\in\mathcal{N}} \left\{ q_{i,0}\,\hat{\nu}_{i,0}^{t+1} + \sum_{\ell=1}^{\infty} q_{i,\ell}\left[\pi_i^t + \rho_{i,\ell}\,\hat{\nu}_{i,0}^{t+1} + (1-\rho_{i,\ell})\,\hat{\nu}_{i,\ell+1}^{t+1}\right]\right\} \\
& + \max_{S\in\mathcal{F}} \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\,\geq 1\}}\, \phi_i^t(S)\left[r_i^t + \pi_i^t - (1-\rho_{i,0})\,(\hat{\nu}_{i,0}^{t+1} - \hat{\nu}_{i,1}^{t+1})\right],
\end{aligned}
$$

where the last equality follows from the fact that $\hat{S}^t(\boldsymbol{q})$ is, by (7), an optimal solution to the maximization problem on the right side above. Noting (6), for all $\ell \geq 1$, we have $\hat{\nu}_{i,\ell}^t = \pi_i^t + \rho_{i,\ell}\,\hat{\nu}_{i,0}^{t+1} + (1-\rho_{i,\ell})\,\hat{\nu}_{i,\ell+1}^{t+1}$. In this case, the expression on the right side of the chain of inequalities above can equivalently be written as

$$
\begin{aligned}
& \sum_{i\in\mathcal{N}} \left\{ q_{i,0}\,\hat{\nu}_{i,0}^{t+1} + \sum_{\ell=1}^{\infty} q_{i,\ell}\,\hat{\nu}_{i,\ell}^t \right\} \; + \; \max_{S\in\mathcal{F}} \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\,\geq 1\}}\, \phi_i^t(S)\left[r_i^t + \pi_i^t - (1-\rho_{i,0})\,(\hat{\nu}_{i,0}^{t+1} - \hat{\nu}_{i,1}^{t+1})\right] \\
& \qquad \geq \; \sum_{i\in\mathcal{N}} \left\{ q_{i,0}\,\hat{\nu}_{i,0}^{t+1} + \sum_{\ell=1}^{\infty} q_{i,\ell}\,\hat{\nu}_{i,\ell}^t \right\} \; + \; \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\,\geq 1\}}\, \phi_i^t(\hat{A}^t)\;\left[r_i^t + \pi_i^t - (1-\rho_{i,0})\,(\hat{\nu}_{i,0}^{t+1} - \hat{\nu}_{i,1}^{t+1})\right] \\
& \qquad \geq \; \sum_{i\in\mathcal{N}} \left\{ q_{i,0}\,\hat{\nu}_{i,0}^{t+1} + \sum_{\ell=1}^{\infty} q_{i,\ell}\,\hat{\nu}_{i,\ell}^t \right\} \; + \; \sum_{i\in\mathcal{N}} \frac{q_{i,0}}{C_i}\,\phi_i^t(\hat{A}^t)\left[r_i^t + \pi_i^t - (1-\rho_{i,0})\,(\hat{\nu}_{i,0}^{t+1} - \hat{\nu}_{i,1}^{t+1})\right] \\
& \qquad = \; \sum_{i\in\mathcal{N}} \left\{ q_{i,0}\,\hat{\nu}_{i,0}^{t+1} + \sum_{\ell=1}^{\infty} q_{i,\ell}\,\hat{\nu}_{i,\ell}^t \right\} \; + \; \sum_{i\in\mathcal{N}} q_{i,0}(\hat{\nu}_{i,0}^t - \hat{\nu}_{i,0}^{t+1}) \\
& \qquad = \; \sum_{i\in\mathcal{N}} \sum_{\ell=0}^{\infty} q_{i,\ell}\,\hat{\nu}_{i,\ell}^t \;\; = \;\; \hat{J}^t(\boldsymbol{q}).
\end{aligned}
$$

In the chain of inequalities above, to see that the second inequality holds, by the discussion in the proof of Lemma 3.1, we have $\phi_i^t(\hat{A}^t)\,[r_i^t + \pi_i^t - (1-\rho_{i,0})\left(\hat{\nu}_{i,0}^{t+1} - \hat{\nu}_{i,1}^{t+1}\right)] \geq 0$ for all $i\in\mathcal{N}$. Also, by the definition of $\mathcal{Q}$, we have $q_{i,0} \leq C_i$ for any $\boldsymbol{q}\in\mathcal{Q}$, which implies that $\mathbb{1}_{\{q_{i,0}\geq 1\}} \geq \frac{q_{i,0}}{C_i}$. The first equality follows from (6). The chain of inequalities above completes the induction argument so that we have $U^t(\boldsymbol{q}) \geq \hat{J}^t(\boldsymbol{q})$ for all $\boldsymbol{q}\in\mathcal{Q}$ and $t\in\mathcal{T}$. Because the initial state of the system is $\sum_{i\in\mathcal{N}} C_i\,\boldsymbol{e}_{i,0}$, the total expected revenue collected by the greedy policy is $U^1\left(\sum_{i\in\mathcal{N}} C_i\,\boldsymbol{e}_{i,0}\right)$. So, using the last inequality with $t=1$ and $\boldsymbol{q} = \sum_{i\in\mathcal{N}} C_i\,\boldsymbol{e}_{i,0}$, we get $U^1\left(\sum_{i\in\mathcal{N}} C_i\,\boldsymbol{e}_{i,0}\right) \geq \hat{J}^1\left(\sum_{i\in\mathcal{N}} C_i\,\boldsymbol{e}_{i,0}\right) \geq \frac{1}{2} J^1\left(\sum_{i\in\mathcal{N}} C_i\,\boldsymbol{e}_{i,0}\right)$, where the second inequality follows by Lemma 3.3. ∎

We note that simple myopic approaches that ignore the future customer arrivals can perform arbitrarily poorly, as we demonstrate in Appendix C. In contrast, by Theorem 3.2, the greedy policy with respect to the value function approximations $\{\hat{J}^t : t\in\mathcal{T}\}$ is guaranteed to obtain at least 50%

of the optimal total expected revenue. In our computational experiments, we demonstrate that the practical performance of this greedy policy can be substantially better than this theoretical performance guarantee. Despite having a performance guarantee, the greedy policy with respect to the value function approximations $\{\hat{J}^t : t \in \mathcal{T}\}$ has a somewhat undesirable feature. Consider two states $\boldsymbol{q} \in \mathcal{Q}$ and $\boldsymbol{q}' \in \mathcal{Q}$ such that $\{i \in \mathcal{N} : q_{i,0} \geq 1\} = \{i \in \mathcal{N} : q'_{i,0} \geq 1\}$. In other words, the set of products for which we have on-hand inventory is the same in the two states. In this case, by (7), we have $\hat{S}^t(\boldsymbol{q}) = \hat{S}^t(\boldsymbol{q}')$. Therefore, the decisions of the greedy policy depend on the set of products for which we have on-hand inventory, but not on the level of inventory for these products. The greedy policy does not differentiate between having too much or too little inventory of a product, as long as we have on-hand inventory for this product. In the next section, we develop a more sophisticated policy that explicitly takes the inventory levels into consideration, while still maintaining the performance guarantee of the greedy policy. Our computational experiments demonstrate that the latter policy can perform noticeably better than the greedy policy.

4. Improving the Policy Performance through Rollout

To develop a policy that explicitly takes the inventory levels of the products into consideration, we build on a static policy that offers a fixed assortment at each time period. With the assortment $\hat{A}^t$ defined in (5), the static policy always offers the assortment $\hat{A}^t$ at time period t. Using an analysis similar to the one for the greedy policy with respect to the linear value function approximations discussed in the previous section, we show that the static policy obtains at least 50% of the optimal total expected revenue. Furthermore, the value functions associated with the static policy are separable by the products. We perform rollout on the static policy to obtain a policy that takes the inventory levels of the products into consideration, while still maintaining the performance guarantee of the static policy. Exploiting the fact that the value functions associated with the static policy are separable by the products, we show that we can efficiently perform rollout on the static policy when the usage durations follow a negative binomial distribution or when the customers purchase the products outright without returning them at all.

4.1 Properties of the Static Policy

We consider a static policy that always offers the assortment $\hat{A}^t$ at time period t regardless of the product availabilities, where $\hat{A}^t$ is defined in (5). If a customer chooses a product that is not available, then she leaves the system. By the next lemma, the static policy obtains at least 50% of the optimal total expected revenue. The proof is similar to the analysis of the greedy policy with respect to the linear value function approximations. The details are in Appendix D.

Lemma 4.1 (Performance of the Static Policy) *The total expected revenue of the static policy that offers assortment $\hat{A}^t$ at time period t is at least 50% of the optimal total expected revenue.*

Let $V^t(\boldsymbol{q})$ denote the total expected revenue under the static policy over the time periods $t, \ldots, T$, given that we are in state $\boldsymbol{q}$ at time period t. Similar to the dynamic program in (3), we can compute $\{V^t : t \in \mathcal{T}\}$ by using the recursion

$$V^t(\boldsymbol{q}) = \sum_{i \in \mathcal{N}} \pi_i^t \sum_{\ell=1}^{\infty} q_{i,\ell} \; + \; \mathbb{E}\Big\{ V^{t+1}(\boldsymbol{X}(\boldsymbol{q})) \Big\} \\ + \sum_{i \in \mathcal{N}} \mathbb{1}_{\{q_{i,0} \geq 1\}} \, \phi_i^t(\hat{A}^t) \bigg(r_i^t + \pi_i^t - (1 - \rho_{i,0}) \mathbb{E}\Big\{ V^{t+1}(\boldsymbol{X}(\boldsymbol{q})) - V^{t+1}\left(\boldsymbol{X}(\boldsymbol{q}) - \boldsymbol{e}_{i,0} + \boldsymbol{e}_{i,1}\right) \Big\} \bigg),$$

with the boundary condition that $V^{T+1} = 0$. The following lemma shows that $V^t(\boldsymbol{q})$ decomposes by products. The proof is in Appendix E. To facilitate our exposition, let $\boldsymbol{e}_\ell$ be the standard unit vector with one in the ℓ-th coordinate. Let $\boldsymbol{q}_i = (q_{i,\ell} \; : \; \ell = 0, 1, \ldots)$ denote the numbers of units of product i that have been in use for different numbers of time periods. By (1), the state of the units of product i at the next time period depends on the state of the units of product i at the current time period, but not on other products. Thus, $X_{i,\ell}(\boldsymbol{q})$ is a function of $\boldsymbol{q}_i$ only, which implies that we can write $X_{i,\ell}(\boldsymbol{q})$ as $X_{i,\ell}(\boldsymbol{q}_i)$, so we can define the vector $\boldsymbol{X}_i(\boldsymbol{q}_i) = (X_{i,\ell}(\boldsymbol{q}_i) : \ell = 0, 1, \ldots)$.

Lemma 4.2 (Decomposability by Products) *For each $t \in \mathcal{T}$ and $\boldsymbol{q} \in \mathcal{Q}$, we have $V^t(\boldsymbol{q}) = \sum_{i \in \mathcal{N}} V_i^t(\boldsymbol{q}_i)$, where for each $i \in \mathcal{N}$, $\{V_i^t : t \in \mathcal{T}\}$ is computed by using the recursion*

$$V_i^t(\boldsymbol{q}_i) = \pi_i^t \sum_{\ell=1}^{\infty} q_{i,\ell} \; + \; \mathbb{E}\Big\{ V_i^{t+1}(\boldsymbol{X}_i(\boldsymbol{q}_i)) \Big\} \\ + \mathbb{1}_{\{q_{i,0} \geq 1\}} \, \phi_i^t(\hat{A}^t) \bigg(r_i^t + \pi_i^t - (1 - \rho_{i,0}) \, \mathbb{E}\Big\{ V_i^{t+1}(\boldsymbol{X}_i(\boldsymbol{q}_i)) - V_i^{t+1}\left(\boldsymbol{X}_i(\boldsymbol{q}_i) - \boldsymbol{e}_0 + \boldsymbol{e}_1\right) \Big\} \bigg), \quad (9)$$

with the boundary condition that $V_i^{T+1} = 0$.

4.2 Rollout Policy Based on the Static Policy

We perform rollout on the static policy to obtain a policy that takes the inventory levels of the products into consideration. To perform rollout on the static policy, given that we are in a particular state at the current time period, we choose the decision that maximizes the immediate expected revenue at the current time period plus the expected revenue from the static policy starting from the state at the next time period. We refer to the policy obtained by performing rollout on the static policy as the rollout policy. The rollout policy ultimately corresponds to using $V^t(\boldsymbol{q}) = \sum_{i \in \mathcal{N}} V_i^t(\boldsymbol{q}_i)$ as a separable nonlinear approximation to $J^t(\boldsymbol{q})$. Let $S^t_{\mathsf{rollout}}(\boldsymbol{q})$ be the assortment offered by the rollout policy given that we are in state $\boldsymbol{q}$ at time period t. As $V^{t+1}(\boldsymbol{q})$ is the total expected revenue obtained by the static policy starting in state $\boldsymbol{q}$ at time period $t+1$, $S^t_{\mathsf{rollout}}(\boldsymbol{q})$ is given by

$$
\begin{aligned}
&S^t_{\mathsf{rollout}}(\boldsymbol{q}) \\
&= \arg\max_{S\in\mathcal{F}} \Bigg\{ \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\geq 1\}}\, \phi^t_i(S) \Big(r^t_i + \pi^t_i + \mathbb{E}\Big\{ \mathsf{Z}(\rho_{i,0})\, V^{t+1}(\boldsymbol{X}(\boldsymbol{q})) + (1-\mathsf{Z}(\rho_{i,0}))\, V^{t+1}(\boldsymbol{X}(\boldsymbol{q}) - \boldsymbol{e}_{i,0} + \boldsymbol{e}_{i,1}) \Big\} \Big) \\
&\qquad\qquad + \Big(1 - \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\geq 1\}} \phi^t_i(S) \Big)\, \mathbb{E}\Big\{ V^{t+1}(\boldsymbol{X}(\boldsymbol{q})) \Big\} \Bigg\} \\
&= \arg\max_{S\in\mathcal{F}} \; \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\geq 1\}}\, \phi^t_i(S) \Big(r^t_i + \pi^t_i - (1-\rho_{i,0})\, \mathbb{E}\Big\{ V^{t+1}(\boldsymbol{X}(\boldsymbol{q})) - V^{t+1}\left(\boldsymbol{X}(\boldsymbol{q}) - \boldsymbol{e}_{i,0} + \boldsymbol{e}_{i,1}\right) \Big\} \Big) \\
&= \arg\max_{S\in\mathcal{F}} \; \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\geq 1\}} \phi^t_i(S) \Big(r^t_i + \pi^t_i - (1-\rho_{i,0})\, \mathbb{E}\Big\{ V^{t+1}_i(\boldsymbol{X}_i(\boldsymbol{q}_i)) - V^{t+1}_i\left(\boldsymbol{X}_i(\boldsymbol{q}_i) - \boldsymbol{e}_0 + \boldsymbol{e}_1\right) \Big\} \Big).
\end{aligned}
$$

In the first equality above, we follow the same argument that we used to construct the dynamic program in (2), in which we find an assortment that maximizes the immediate expected revenue and the expected value function at the next time period under the optimal policy, but above, we use the value function of the static policy at the next time period. We arrive at the second equality by the same reasoning that we used to obtain the dynamic program in (3) from the dynamic program in (2). The third equality follows from the fact that the value functions of the static policy decompose by the products, as shown in Lemma 4.2.

It is a well-known result that the policy obtained by performing rollout on a base policy always performs at least as well as the base policy itself; see Section 6.1.3 in Bertsekas and Tsitsiklis (1996). Therefore, the total expected revenue obtained by our rollout policy is at least as large as the total expected revenue obtained by the static policy. So, by Lemma 4.1, the rollout policy obtains at least 50% of the optimal total expected revenue as well. In many applications, a policy based on rollout tends to offer a dramatic improvement over the base policy. The key question is whether the rollout assortment $S^t_{\mathsf{rollout}}(\boldsymbol{q})$ can be computed efficiently. Lemma 4.2 shows that the value function of the static policy is separable by the products, indicating that computing the value functions of the static policy through the recursion in (9) is more manageable than computing the value functions of the optimal policy through the dynamic program in (3). As discussed earlier, without loss of generality, we can assume that the vector $\boldsymbol{q}_i = (q_{i,\ell} : \ell = 0, 1, \ldots)$ is finite-dimensional, because we start with no units in use so that we always have have $q_{i,\ell} = 0$ for all $\ell \geq T$. However, the state variable $\boldsymbol{q}_i = (q_{i,\ell} : \ell = 0, 1, \ldots)$ in the recursion in (9) is still a high-dimensional vector. In particular, the state space in this recursion is given by $\mathcal{Q}_i = \{(q_{i,\ell} \in \mathbb{Z}_+ : \ell = 0, 1, \ldots) \mid \sum_{\ell=0}^{\infty} q_{i,\ell} = C_i\}$ and computing the value function $V^t_i(\boldsymbol{q}_i)$ of the static policy for all $\boldsymbol{q}_i \in \mathcal{Q}_i$ is difficult.

In the remainder of this section, we consider two cases. First, if the usage duration follows a negative binomial distribution, then the value functions of the static policy can be computed efficiently. Second, if the customers purchase the products outright and never return them, then the value functions of the static policy can be computed efficiently as well. Once we compute the value functions $\{V^t : t \in \mathcal{T}\}$ of the static policy efficiently, we can solve the maximization problem above that defines $S^t_{\mathsf{rollout}}(\boldsymbol{q})$ to find the assortment offered by the rollout policy. Note that the maximization problem that we solve to obtain the assortment $S^t_{\mathsf{rollout}}(\boldsymbol{q})$ has the same structure as the maximization problem on the right side of the dynamic program in (3). Thus, once we compute the value functions $\{V^t : t \in \mathcal{T}\}$ of the static policy, as discussed at the end of Section 2, there are numerous choice models that render this maximization problem tractable. Lastly, we emphasize that even if we cannot compute the value functions $\{V^t : t \in \mathcal{T}\}$ of the static policy, we can use simulation to estimate the expected revenue of the static policy, which still allows performing rollout on the static policy. Section 6.1.3 in Bertsekas and Tsitsiklis (1996) discusses using simulation to perform rollout. Naturally, the computational requirements of performing rollout inflate when we use simulation to estimate the total expected revenue of the static policy. Next, we discuss how to perform rollout efficiently when the usage durations have a negative binomial distribution.

4.3 Negative Binomial Usage Duration

In this section, we assume that for each product $i \in \mathcal{N}$, the usage duration is given as $\mathsf{Duration}_i = 1 + \mathsf{NegBin}(s_i, \eta_i)$, where $\mathsf{NegBin}(s_i, \eta_i)$ denotes a negative binomial random variable with parameters $s_i \in \mathbb{Z}_{++}$ and $\eta_i \in [0,1]$ taking values over $\{0, 1, \ldots\}$. A negative binomial random variable with parameters (s_i, η_i) corresponds to the sum of s_i independent geometric random variables, each with parameter η_i. Thus, a negative binomial random variable with parameters $(1, \eta_i)$ is equivalent to a geometric random variable with parameter η_i. As s_i increases, the probability mass function of a negative binomial random variable with parameters (s_i, η_i) becomes more symmetric. Even with $s_i = 3$, the probability mass function is rather symmetric. Therefore, a negative binomial random variable is quite flexible for modeling usage durations.

Noting that a negative binomial random variable with parameters (s_i, η_i) corresponds to the sum of s_i geometric random variables, we provide the following interpretation for our use of a negative binomial random variable for modeling the usage durations. At each time period, a customer is satisfied with product i with probability η_i. As soon as a customer is dissatisfied with the product for s_i times, she returns the product, ending her rental duration. Naturally, we do not advocate this interpretation as a model of how a customer makes a decision for keeping the product, but this interpretation provides us with the vocabulary to explain our model more clearly, as follows. If the usage durations have negative binomial distributions, then our state variable does not need to keep track of the numbers of units of each product i that have been in use for a certain duration of time. It is enough to use a state variable that keeps track of the numbers of customers who are using each product i and have been dissatisfied for a certain number of times. In this case, we can efficiently compute the value functions of the static policy, as long as s_i is relatively small.

We discuss how we can compute the value functions of the static policy by using a recursion similar to the one in (9) when the usage durations are negative binomial random variables.

State and Transition Dynamics: To compute the value functions of the static policy through a recursion similar to the one used in (9), we define

$$w_{i,d} \;=\; \text{number of customers who are using product } i \text{ and have been dissatisfied for } d \text{ times.}$$

A customer using product i returns the product once she has been dissatisfied for s_i times, in which case, the product becomes available on-hand. Therefore, the s_i-dimensional vector $(w_{i,0},\ldots,w_{i,s_i-1})$ captures the state of the customers using product i. The on-hand inventory of product i is given by $C_i - \sum_{d=0}^{s_i-1} w_{i,d}$. Under negative binomial usage durations, we use $\boldsymbol{w}_i = (w_{i,d} : 0 \le d \le s_i - 1)$ to denote the state vector for product i at the beginning of a generic time period. With this state representation, if no purchase is made at the current time period, then the new random state $\boldsymbol{F}_i(\boldsymbol{w}_i) = (F_{i,d}(\boldsymbol{w}_i) \;:\; 0 \le d \le s_i - 1)$ at the next time period is given by

$$F_{i,d}(\boldsymbol{w}_i) = \begin{cases} \mathsf{Bin}(w_{i,0}, \eta_i) & \text{if} \quad d = 0, \\ \mathsf{Bin}(w_{i,d}, \eta_i) + (w_{i,d-1} - \mathsf{Bin}(w_{i,d-1}, \eta_i)) & \text{if} \quad d = 1, 2, \ldots, s_i - 1, \end{cases}$$

where we use the fact that for each d, the number of customers who continue to remain dissatisfied for d times at the next time period is equal to $\mathsf{Bin}(w_{i,d}, \eta_i)$, because each customer is satisfied with the product with probability η_i, independently of each other. Furthermore, $w_{i,d-1} - \mathsf{Bin}(w_{i,d-1}, \eta_i)$ captures the number of customers who were dissatisfied for $d-1$ times at the beginning of the current time period and they were dissatisfied one more time in the current time period; in that case, these customers are dissatisfied for a total of d times at the next time period. These customers add up to the number of customers dissatisfied d times at the next time period.

Dynamic Programming Formulation: With this state representation, we can compute the value functions of the static policy for each product i through the following recursion. We use $\boldsymbol{w}_i = (w_{i,0}, \ldots, w_{i,s_i-1})$ to capture the state of product i. Recall that the static policy offers the assortment $\hat{A}^t$ at each time period t. Given that the state of product i at time period t is $\boldsymbol{w}_i$, let $V_i^t(\boldsymbol{w}_i)$ be the total expected revenue from product i under the static policy over the time periods $t, \ldots, T$. Using the vectors $\boldsymbol{e}_0 = (1, 0, 0, \ldots, 0) \in \mathbb{R}^{s_i}$ and $\boldsymbol{e}_1 = (0, 1, 0, \ldots, 0) \in \mathbb{R}^{s_i}$, we can compute $\{V_i^t : t \in \mathcal{T}\}$ by using the recursion

$$V_i^t(\boldsymbol{w}_i) \;=\; \pi_i^t \sum_{d=0}^{s_i-1} w_{i,d} + \left(1 - \mathbb{1}_{\left\{\sum_{d=0}^{s_i-1} w_{i,d} \,<\, C_i\right\}} \phi_i^t(\hat{A}^t)\right) \mathbb{E}\left\{ V_i^{t+1}\left(\boldsymbol{F}_i(\boldsymbol{w}_i)\right) \right\}$$

$$
\begin{aligned}
&+\ \mathbb{1}_{\left\{\sum_{d=0}^{s_i-1} w_{i,d}\ <\ C_i\right\}} \phi_i^t(\hat{A}^t)\left(r_i^t+\pi_i^t+\eta_i\,\mathbb{E}\left\{V_i^{t+1}\left(\boldsymbol{F}_i(\boldsymbol{w}_i)+\boldsymbol{e}_0\right)\right\}+(1-\eta_i)\,\mathbb{E}\left\{V_i^{t+1}\left(\boldsymbol{F}_i(\boldsymbol{w}_i)+\boldsymbol{e}_1\right)\right\}\right)\\
=&\ \pi_i^t\sum_{d=0}^{s_i-1} w_{i,d}+\mathbb{E}\left\{V_i^{t+1}\left(\boldsymbol{F}_i(\boldsymbol{w}_i)\right)\right\}\\
&+\mathbb{1}_{\left\{\sum_{d=0}^{s_i-1} w_{i,d}\ <\ C_i\right\}} \phi_i^t(\hat{A}^t)\left(r_i^t+\pi_i^t-\eta_i\,\mathbb{E}\left\{V_i^{t+1}\left(\boldsymbol{F}_i(\boldsymbol{w}_i)\right)-V_i^{t+1}\left(\boldsymbol{F}_i(\boldsymbol{w}_i)+\boldsymbol{e}_0\right)\right\}\right.\\
&\qquad\qquad\left.-(1-\eta_i)\,\mathbb{E}\left\{V_i^{t+1}\left(\boldsymbol{F}_i(\boldsymbol{w}_i)\right)-V_i^{t+1}\left(\boldsymbol{F}_i(\boldsymbol{w}_i)+\boldsymbol{e}_1\right)\right\}\right),
\end{aligned}
\tag{10}
$$

with the boundary condition that $V_i^{T+1}=0$. In the first equality above, for a customer to rent a unit of product i, we need to have product i available on-hand and the customer needs to choose product i. The number of units of product i available on-hand is given by $C_i-\sum_{d=0}^{s_i-1} w_{i,d}$, so the expression $1-\mathbb{1}_{\left\{\sum_{d=0}^{s_i-1} w_{i,d}\ <\ C_i\right\}}\,\phi_i^t(\hat{A}^t)$ captures the probability that a customer does not rent product i when we offer the assortment $\hat{A}^t$. If $\sum_{d=0}^{s_i-1} w_{i,d}<C_i$, then we have product i available on-hand. If the customer chooses product i, then she rents this product. With probability η_i, the customer renting product i at the current time period is satisfied and she ends up being a customer with no dissatisfactions at the beginning of the next time period. With probability $1-\eta_i$, the customer renting product i at the current time period is dissatisfied and she becomes a customer who is dissatisfied for one time at the beginning of the next time period. The second equality follows by arranging the terms. If $s_i=1$, so that the usage durations for product i are geometric random variables, then the state variable $\boldsymbol{w}_i$ becomes the scalar $w_{i,0}$, in which case, the recursion above continues to hold as long as we set $\boldsymbol{e}_0=1$ and $\boldsymbol{e}_1=0$.

Discussion of the State Variable: We can reach the state variable $\boldsymbol{w}_i=(w_{i,d}:0\le d\le s_i-1)$ that we used in (10) by starting from the state variable $\boldsymbol{q}_i=(q_{i,\ell}:\ell=0,1,\ldots)$ that we used in (9). Recall that $q_{i,\ell}$ is the number of units of product i that have been in use for exactly ℓ time periods. Because the number of units of product i available on-hand is given by $q_{i,0}=C_i-\sum_{\ell=1}^{\infty} q_{i,\ell}$, we can use the state variable $(q_{i,\ell}:\ell=1,2,\ldots)$, instead of $(q_{i,\ell}:\ell=0,1,\ldots)$. Let $y_{i,d,\ell}$ be the number of customers who have been using product i for exactly ℓ time periods and have been dissatisfied for d times. By definition, we have $q_{i,\ell}=\sum_{d=0}^{s_i-1} y_{i,d,\ell}$. So, we can use the state variable $\boldsymbol{y}_i=(y_{i,d,\ell}:\ell=1,2,\ldots,\ 0\le d\le s_i-1)$ instead of $(q_{i,\ell}:\ell=1,2,\ldots)$, because given $(y_{i,d,\ell}:\ell=1,2,\ldots,\ 0\le d\le s_i-1)$, we can compute $(q_{i,\ell}:\ell=1,2,\ldots)$ as $q_{i,\ell}=\sum_{d=0}^{s_i-1} y_{i,d,\ell}$. Lastly, under negative binomial usage durations, from the perspective of immediate expected revenues and state transitions, if we know the number of times a customer has been dissatisfied, then we do not need to know how long she has been using the product. Thus, letting $w_{i,d}=\sum_{\ell=1}^{\infty} y_{i,d,\ell}$ be the number of customers who are using product i and have been dissatisfied d times, we can use $\boldsymbol{w}_i=(w_{i,d}:0\le d\le s_i-1)$ as the state variable, which is precisely the state variable in (10).

In the recursion in (10), the state variable is an s_i-dimensional vector $(w_{i,0},\ldots,w_{i,s_i-1})$ such that $\sum_{d=0}^{s_i-1} w_{i,d} \le C_i$, so the number of states is $O(C_i^{s_i})$. Thus, when s_i is relatively small, we can compute the value functions of the static policy efficiently. For example, in our computational experiments, using transaction data from the city of Seattle, we find that the negative binomial distribution provides a reasonably good model for the duration of time for which the drivers park their vehicles. In our experiments, the fitted value for the parameter s_i was 2.

4.4 Infinite Usage Duration

In this section, we focus on the case in which the usage duration is infinity. This case corresponds to the situation where the customers buy the products outright, never returning them. Infinite usage durations have a number of interesting applications. In the retail setting, customers make purchases among substitutable products, in which case, our model dynamically makes product assortment offerings to each individual customer as a function of the remaining product inventories (Topaloglu 2013, Golrezaei et al. 2014). Also, an important class of revenue management problems occurs on a flight network with parallel flights operating between the same origin-destination pair. In this setting, the customers make a purchase among multiple parallel flights on a particular departure date. Our model dynamically adjusts the assortment of flights offered to each individual customer as a function of the remaining flight capacities (Zhang and Cooper 2005, Liu and van Ryzin 2008, Dai et al. 2014). Under infinite usage durations, we proceed to discuss how we can compute the value functions of the static policy by using a recursion similar to the one in (9).

State and Transition Dynamics: Because the products are purchased outright, we assume that $\pi_i^t = 0$ for all $t \in \mathcal{T}$ so that there is no per-period rental fee. Since the products are not returned, we only need to keep track of the on-hand inventory of product i. We let $q_{i,0}$ be the number of units of product i on-hand and use $q_{i,0}$ as the state variable at the beginning of a time period. If the state of the system at the current time period is $q_{i,0}$ and a customer purchases product i, then the state of the system at the next time period is simply $q_{i,0}-1$.

Dynamic Programming Formulation: Given that we have $q_{i,0}$ units of product i on-hand, let $V_i^t(q_{i,0})$ be the total expected revenue from product i under the static policy over the time periods $t,\ldots,T$. We can compute $\{V_i^t : t \in \mathcal{T}\}$ by using the recursion

$$\begin{aligned} V_i^t(q_{i,0}) &= \left(1 - \mathbb{1}_{\{q_{i,0}\ge 1\}}\phi_i^t(\hat{A}^t)\right) V_i^{t+1}(q_{i,0}) \;+\; \mathbb{1}_{\{q_{i,0}\ge 1\}}\,\phi_i^t(\hat{A}^t)\left(r_i^t + V_i^{t+1}(q_{i,0}-1)\right) \\ &= V_i^{t+1}(q_{i,0}) \;+\; \mathbb{1}_{\{q_{i,0}\ge 1\}}\,\phi_i^t(\hat{A}^t)\left(r_i^t - \left\{V_i^{t+1}(q_{i,0}) - V_i^{t+1}(q_{i,0}-1)\right\}\right), \end{aligned} \tag{11}$$

with the boundary condition that $V_i^{T+1} = 0$. In the first equality above, if we have on-hand units of product i and a customer chooses product i, then we have one fewer on-hand unit at the next time period. The second equality follows by arranging the terms. Because the state variable $q_{i,0}$ in the recursion above is scalar, we can efficiently compute the value functions of the static policy under infinite usage durations. The recursion in (11) is similar to the one in revenue management problems with a single resource; see Section 2.6.2 in Talluri and van Ryzin (2005).

Thus, under both negative binomial and infinite usage durations, we can efficiently perform rollout on the static policy; in that case, we obtain a policy that takes the inventory levels of the products into consideration, while still obtaining at least 50% of the optimal total expected revenue. It turns out that we can further strengthen our performance guarantee under infinite usage durations. In particular, we let $C_{\min} = \min_{i \in \mathcal{N}} C_i$ to capture the smallest inventory of a product. Also, we let $R = \max_{i \in \mathcal{N}} \left\{ \frac{\max_{t \in \mathcal{T}} r_i^t}{\min_{t \in \mathcal{T}} r_i^t} \right\}$ to capture the largest relative deviation in the upfront fee for a product over the selling horizon. In Theorem F.2 in Appendix F, using a modified static policy based on a solution of a linear program, we can construct a tailored rollout policy that is guaranteed to obtain at least $\max \left\{ \frac{1}{2}, 1 - \frac{R}{2\sqrt[3]{C_{\min}}} \right\}$ fraction of the optimal total expected revenue. Therefore, the tailored variant of our rollout approach always provides at least a half-approximate performance guarantee, but it becomes near-optimal as the inventories of the products become large. This performance guarantee is not an asymptotic performance guarantee. It holds for any value of the product inventories and the number of time periods in the selling horizon. For example, if the smallest product inventory is 100 and the upfront fees are stationary so that $C_{\min} = 100$ and $R = 1$, then the tailored variant of our rollout policy is guaranteed to obtain at least 89% of the optimal total expected revenue, regardless of the other problem parameters. In addition, we consider a standard regime where the inventories of the products and the number of time periods in the selling horizon scale up linearly at the same rate κ (Gallego and van Ryzin 1994). In Theorem F.7 in Appendix F, we also show that the tailored variant of our rollout policy obtains at least $1 - \frac{B}{\sqrt{\kappa}}$ fraction of the optimal total expected revenue, where B is a constant that is independent of the scaling rate κ. Thus, the relative optimality gap of the tailored variant of the rollout policy is $O\left(1 - \frac{1}{\sqrt{\kappa}}\right)$ as the inventories of the products and the number of time periods scale up linearly at the same rate κ. These two performance guarantees do not generalize to arbitrary usage duration distributions.

5. Extensions

We give extensions to the case in which we have multiple customer types, we make pricing decisions instead of assortment offer decisions, and we can solve the assortment optimization problems only approximately. We show that our half-approximate performance guarantee continues to hold when we have multiple customer types and when we make pricing decisions. Furthermore, we show that if we can solve the assortment optimization problems approximately, then our performance guarantees hold with appropriate modifications to reflect the solution accuracy in the assortment problems. Some of these extensions are used in our computational experiments.

5.1 Heterogeneous Customer Types

We have m customer types indexed by $\mathcal{M} = \{1, 2, \ldots, m\}$. At time period $t \in \mathcal{T}$, a customer of type j arrives with probability $p^{t,j}$, where we have $\sum_{j \in \mathcal{M}} p^{t,j} = 1$, so that each time period has exactly one customer arrival. We observe the type of each arriving customer. Each customer type has its own choice model, reward structure, assortment constraints, and usage duration. If we offer the subset S of products, then a customer of type j arriving at time period t chooses product i with probability $\phi_i^{t,j}(S)$. Note that if we do not observe the type of each arriving customer, then we can continue using the model in Section 2, where the choice probability $\phi_i^t(S)$ is obtained by mixing the choice models corresponding to different customer types. If a customer of type j selects product i at time period t, then she pays a one-time upfront fee of $r_i^{t,j}$. Furthermore, if she rents this product during time period t, then she pays a per-period rental fee of $\pi_i^{t,j}$. The usage duration of product i by a customer of type j is given by the random variable $\mathsf{Duration}_i^j$. We let $\rho_{i,\ell}^j$ be the hazard rate of the usage duration of product i for a customer of type j, which is defined by $\rho_{i,\ell}^j = \Pr\{\mathsf{Duration}_i^j = \ell + 1 \mid \mathsf{Duration}_i^j > \ell\}$. Lastly, the assortments offered to customers of different types have different feasibility requirements. We use $\mathcal{F}^j$ to denote the set of feasible assortments that can be offered to customers of type j.

We can extend all of our results to the case with heterogeneous customer types. We will focus on the essentials in this section; the details are included in Appendix G. To capture the state of the system, because each customer type has its own reward structure and usage duration, we need to keep track of the number of units that are currently in use by each customer type. We use $q_{i,0}$ to denote the number of units of product i on-hand. For $\ell \geq 1$, we use $q_{i,\ell}^j$ to denote the number of units of product i that have been used for exactly ℓ time periods by a customer of type j. Therefore, we can describe the state of the system by using $\boldsymbol{q} = (q_{i,0}, q_{i,\ell}^j : i \in \mathcal{N},\ j \in \mathcal{M},\ \ell \geq 1)$. Using $\boldsymbol{q}$ as the state variable, we can give a dynamic programming formulation of the problem that resembles the one in (2). In this case, we use value function approximations of the form

$$\hat{J}^t(\boldsymbol{q}) = \sum_{i \in \mathcal{N}} \hat{\theta}_i^t \, q_{i,0} + \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{M}} \sum_{\ell=1}^{\infty} \hat{\nu}_{i,\ell}^{t,j} \, q_{i,\ell}^j,$$

where $\hat{\theta}_i^t$ captures the marginal value of a unit of product i on-hand at time period t and $\hat{\nu}_{i,\ell}^{t,j}$ captures the marginal value of a unit of product i that has been in use for ℓ periods by a customer of type j at time period t. We propose computing $\hat{\theta}_i^t$ and $\hat{\nu}_{i,\ell}^{t,j}$ recursively as follows.

- **Initialization:** Set $\hat{\theta}_i^{T+1} = 0$ and $\hat{\nu}_{i,\ell}^{T+1,j} = 0$ for all $i \in \mathcal{N},\ j \in \mathcal{M},\ \ell \geq 1$.
- **Recursion:** For $t = T, T-1, \ldots, 1$, we compute $\hat{\theta}_i^t$ and $\hat{\nu}_{i,\ell}^{t,j}$ by using $\{\hat{\theta}_i^{t+1} : i \in \mathcal{N}\}$ and $\{\hat{\nu}_{i,\ell}^{t+1,j} : i \in \mathcal{N},\ j \in \mathcal{M},\ \ell \geq 1\}$ as follows. For each $j \in \mathcal{M}$, let $\hat{A}^{t,j} \in \mathcal{F}^j$ be such that

$$\hat{A}^{t,j} \;=\; \arg\max_{S \in \mathcal{F}^j} \sum_{i \in \mathcal{N}} \phi_i^{t,j}(S) \Big[r_i^{t,j} + \pi_i^{t,j} - (1 - \rho_{i,0}^j) \left(\hat{\theta}_i^{t+1} - \hat{\nu}_{i,1}^{t+1,j} \right) \Big].$$

Once $\hat{A}^{t,j}$ is computed for all $j \in \mathcal{M}$, for each $i \in \mathcal{N}$ and $j \in \mathcal{M}$, let

$$\begin{aligned} \hat{\theta}_i^t \;&=\; \hat{\theta}_i^{t+1} + \frac{1}{C_i} \sum_{j \in \mathcal{M}} p^{t,j} \phi_i^{t,j}(\hat{A}^{t,j}) \Big[r_i^{t,j} + \pi_i^{t,j} - (1 - \rho_{i,0}^j) \left(\hat{\theta}_i^{t+1} - \hat{\nu}_{i,1}^{t+1,j} \right) \Big] \qquad (12) \\ \hat{\nu}_{i,\ell}^{t,j} \;&=\; \pi_i^{t,j} + \rho_{i,\ell}^j \, \hat{\theta}_i^{t+1} + (1 - \rho_{i,\ell}^j) \, \hat{\nu}_{i,\ell+1}^{t+1,j} \qquad \forall\, \ell = 1, 2, \ldots. \end{aligned}$$

The above discussion completes the specification of the approximate value function $\hat{J}^t$. The computation of the parameters $\{\hat{\theta}_i^t : i \in \mathcal{N},\ t \in \mathcal{T}\}$ and $\{\hat{\nu}_{i,\ell}^{t,j} : i \in \mathcal{N},\ j \in \mathcal{M},\ \ell \geq 1,\ t \in \mathcal{T}\}$ is similar to our approach in Section 3.1. Also, the intuition for the specification of the parameters above is similar to the one discussed in Section 3.1. Using an argument similar to the one in the previous two sections, we can show that the greedy policy with respect to the value function approximations $\{\hat{J}^t : t \in \mathcal{T}\}$ obtains at least 50% of the optimal total expected revenue. We can also perform rollout on a static policy to obtain a policy that takes the inventory levels of the products into consideration, while ensuring that we still obtain at least 50% of the optimal total expected revenue. We describe both of these results in Appendix G.

The use of heterogeneous customer types also allows us to model the case where the usage duration is revealed before offering an assortment. In our problem formulation, we observe the type of a customer before offering an assortment. Also, each customer type can have its own usage duration distribution. Thus, by associating different deterministic usage durations with different customer types, noting that we observe the type of a customer before offering an assortment, we can model the case where the usage duration is revealed before we offer an assortment.

5.2 Price Optimization with Discrete Prices

So far in the paper, we have assumed that the upfront and per-period rental fees for the products are fixed and we decide on the assortment of products to make available to the customers. It is not difficult to adopt our results to the case in which we decide the upfront and per-period rental fees for the products and the customers choose based on the prices we charge. In particular, we create multiple copies of each product i, where the different copies correspond to charging different prices for product i. We call each copy of a product a virtual product. Let $\mathcal{H}$ denote the set of possible copies of each product. We write $(i,h) \in \mathcal{N} \times \mathcal{H}$ to denote copy h of product i. Thus, the pairs $\{(i,h) : i \in \mathcal{N},\ h \in \mathcal{H}\}$ are the set of all virtual products that we can offer to the customers. Offering virtual product (i,h) means that we offer product i at the price level corresponding to copy h of this product. In this case, the question becomes that of choosing an assortment of virtual products to offer at each time period to maximize the total expected revenue. As we can offer a product at no more than one price level, among all virtual copies of a particular product, we can offer at most one virtual copy. Thus, the set of possible assortments of virtual products that we can offer at each time period is given by $\mathcal{F} = \{S \subseteq \mathcal{N} \times \mathcal{H} : |S \cap (\{i\} \times \mathcal{H})| \leq 1 \ \ \forall\, i \in \mathcal{N}\}$. Using $r^t_{i,h}$ to denote the upfront fee at time period t when we charge the price level corresponding to copy h for product i, and $\pi^t_{i,h}$ to denote the per-period fee at time period t when we charge the price level corresponding to copy h of product i, we can follow the same outline in the previous two sections to come up with a policy that obtains at least 50% of the optimal total expected revenue. The only difference is that we treat the virtual products $\mathcal{N} \times \mathcal{H}$ as the products.

5.3 Solving the Assortment Optimization Problem Approximately

The maximization problem in (5) is a combinatorial optimization problem. Under many choice models, we can solve this problem tractably, but it is not possible to solve this problem tractably under every choice model. In this section, we discuss how we can adapt our approach in principle to the case where we have a fully polynomial-time approximation scheme (FPTAS) for problem (5). For any $\epsilon > 0$, the FPTAS returns a $1/(1+\epsilon)$-approximate solution to problem (5), and the running time to do so is polynomial in n and $1/\epsilon$. It turns out that we can leverage the FPTAS to obtain a $1/(2(1+\epsilon))$-approximate policy, and the running time to obtain and execute the approximate policy is polynomial in n, $1/\epsilon$ and T. In particular, assume that we have an FPTAS such that for any $\epsilon > 0$, the FPTAS finds an assortment $\hat{A}^t$ satisfying

$$(1+\epsilon) \sum_{i \in \mathcal{N}} \phi^t_i(\hat{A}^t) \Big[r^t_i + \pi^t_i - (1-\rho_{i,0}) \left(\hat{\nu}^{t+1}_{i,0} - \hat{\nu}^{t+1}_{i,1} \right) \Big] \geq \max_{S \in \mathcal{F}} \sum_{i \in \mathcal{N}} \phi^t_i(S) \Big[r^t_i + \pi^t_i - (1-\rho_{i,0}) \left(\hat{\nu}^{t+1}_{i,0} - \hat{\nu}^{t+1}_{i,1} \right) \Big]$$

in running time that is polynomial in n and $1/\epsilon$. In the next theorem, we show how to leverage this FPTAS to find a $1/(2(1+\epsilon))$-approximate policy. The proof is in Appendix H.

Theorem 5.1 (Policies Through Approximate Solutions) *Assume that for any $\epsilon > 0$, we can find a $1/(1+\epsilon)$-approximate solution to problem (5) in running time that is polynomial in n*

and $1/\epsilon$. Then, we can construct value function approximations $\{\hat{J}^t : t \in \mathcal{T}\}$ such that the greedy policy with respect to these value function approximations is a $1/(2(1+\epsilon))$-approximate policy and the running time to obtain and execute the greedy policy is polynomial in n, $1/\epsilon$ and T.

A quick inspection of the proof of Theorem 5.1 shows that if the running time to obtain a $1/(1+\epsilon)$-approximate solution to problem (5) is $O(f(n, \frac{1}{\epsilon}))$ for some function f and the running time to compute the probabilities $\{\phi_i^t(S) : i \in \mathcal{N}\}$ for a fixed subset S and time period t is $O(g(n))$ for some function g, then the running time to obtain a $1/(2(1+\epsilon))$-approximate policy is $O(T \times f(n, \frac{4T}{\epsilon}) + T \times g(n) + T^2 n)$. Therefore, if we have an FPTAS for problem (5) so that $f(n, \frac{1}{\epsilon})$ is polynomial in $\frac{1}{\epsilon}$, then $f(n, \frac{4T}{\epsilon})$ is also a polynomial in n, $1/\epsilon$ and T, which corresponds to the case discussed in the theorem above. On the other hand, if we have only a polynomial-time approximation scheme for problem (5) so that $f(n, \frac{1}{\epsilon})$ is polynomial in n but exponential in $\frac{1}{\epsilon}$, then $f(n, \frac{4T}{\epsilon})$ is polynomial in n but exponential in $\frac{1}{\epsilon}$ and T.

6. Computational Experiments

We provide computational experiments to test the performance of our policies. In Section 6.1, we give an approach to obtain an upper bound on the optimal total expected revenue, which is useful for assessing the optimality gaps of our policies. In Sections 6.2 and 6.3, we give our computational results on retail assortment management and pricing parking spaces in the city of Seattle.

6.1 Upper Bound on the Optimal Total Expected Revenue

To compute an upper bound on the optimal total expected revenue, we formulate a linear program, in which the choices of the customers and the transition dynamics take on their expected values. We use the decision variables $(z^t(A) : A \in \mathcal{F},\ t \in \mathcal{T})$ and $(q_{i,\ell}^t : i \in \mathcal{N},\ \ell \geq 0,\ t \in \mathcal{T})$, where $z^t(A)$ is the frequency with which we offer assortment A at time period t and $q_{i,\ell}^t$ is the expected number of units of product i that have been in use for exactly ℓ time periods at time period t. To construct the constraints in our linear program, noting the dynamic programming formulation in (2), if the state of the system at the beginning of time period t is $\boldsymbol{q}^t = (q_{i,\ell}^t : i \in \mathcal{N},\ \ell \geq 0)$ and the customer arriving at this time period chooses product i, then the state of the system at the beginning of the next time period is given by the random variable $\mathsf{Z}(\rho_{i,0})\, \boldsymbol{X}(\boldsymbol{q}^t) + (1 - \mathsf{Z}(\rho_{i,0}))\, (\boldsymbol{X}(\boldsymbol{q}^t) - \boldsymbol{e}_{i,0} + \boldsymbol{e}_{i,1})$, where $\mathsf{Z}(\rho)$ is a Bernoulli random variable with parameter ρ. If the customer does not choose any of the products, then the state of the system is $\boldsymbol{X}(\boldsymbol{q}^t)$. Furthermore, if we offer the assortment A at time period t with frequency $z^t(A)$, then the probability that a customer chooses product i is $\sum_{A \in \mathcal{F}} \phi_i^t(A)\, z^t(A)$. In this case, if the state of the system at the beginning of time period t is $\boldsymbol{q}^t$ and we offer assortment A with frequency $z^t(A)$, then the expected state of the system at the beginning of the next time period is given by $\sum_{i\in\mathcal{N}}\left\{\sum_{A\in\mathcal{F}}\phi_i^t(A)\,z^t(A)\right\}\mathbb{E}\left\{\mathsf{Z}(\rho_{i,0})\,\boldsymbol{X}(\boldsymbol{q}^t)+(1-\mathsf{Z}(\rho_{i,0}))\,(\boldsymbol{X}(\boldsymbol{q}^t)-\boldsymbol{e}_{i,0}+\boldsymbol{e}_{i,1})\right\}$ + $\left\{1-\sum_{i\in\mathcal{N}}\left\{\sum_{A\in\mathcal{F}}\phi_i^t(A)\,z^t(A)\right\}\right\}\mathbb{E}\left\{\boldsymbol{X}(\boldsymbol{q}^t)\right\}$. Thus, using the fact that $\mathbb{E}\{\mathsf{Z}(\rho_{i,0})\}=\rho_{i,0}$, by arranging the terms, the expected state at the beginning of the next time period is given by $\mathbb{E}\left\{\boldsymbol{X}(\boldsymbol{q}^t)\right\}-\sum_{i\in\mathcal{N}}\left\{\sum_{A\in\mathcal{F}}\phi_i^t(A)\,z^t(A)\right\}\times(1-\rho_{i,0})\,(\boldsymbol{e}_{i,0}-\boldsymbol{e}_{i,1})$. By (1), $\mathbb{E}\{X_{i,0}(\boldsymbol{q}^t)\}=q_{i,0}^t+\sum_{s=1}^{\infty}\rho_{i,s}\,q_{i,s}^t$, $\mathbb{E}\{X_{i,1}(\boldsymbol{q}^t)\}=0$ and $\mathbb{E}\{X_{i,\ell}(\boldsymbol{q}^t)\}=q_{i,\ell-1}^t-\rho_{i,\ell-1}\,q_{i,\ell-1}^t$ for $\ell\geq 2$, which implies that the expected next state $\mathbb{E}\{\boldsymbol{X}(\boldsymbol{q}^t)\}$ in the last expression is linear in the decision variables $\boldsymbol{q}^t=(q_{i,\ell}^t: i\in\mathcal{N},\ \ell\geq 0)$. To obtain an upper bound on the optimal total expected revenue in our dynamic assortment problem, we use the linear program

$$
\begin{aligned}
\max\quad & \sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{N}}\sum_{A\in\mathcal{F}}(r_i^t+\pi_i^t)\,\phi_i^t(A)\,z^t(A)+\sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{N}}\pi_i^t\sum_{\ell=1}^{\infty}q_{i,\ell}^t \qquad (13)\\
\text{s.t.}\quad & \boldsymbol{q}^{t+1} \;=\; \mathbb{E}\Big\{\boldsymbol{X}(\boldsymbol{q}^t)\Big\}-\sum_{i\in\mathcal{N}}\left\{\sum_{A\in\mathcal{F}}\phi_i^t(A)\,z^t(A)\right\}(1-\rho_{i,0})\,(\boldsymbol{e}_{i,0}-\boldsymbol{e}_{i,1})\quad \forall\, t\in\mathcal{T}\setminus\{T\}\\
& \boldsymbol{q}^1=\sum_{i\in\mathcal{N}}C_i\,\boldsymbol{e}_{i,0}\\
& \sum_{A\in\mathcal{F}}z^t(A)=1\quad \forall\, t\in\mathcal{T}\\
& z^t(A)\geq 0\quad \forall A\in\mathcal{F},\ t\in\mathcal{T},\ q_{i,\ell}^t\geq 0\quad \forall\, i\in\mathcal{N},\ \ell\geq 0,\ t\in\mathcal{T}.
\end{aligned}
$$

From the discussion right before the above problem, the objective function and the constraints are linear in $(z^t(A): A\in\mathcal{F},\ t\in\mathcal{T})$ and $(q_{i,\ell}^t: i\in\mathcal{N},\ \ell\geq 0,\ t\in\mathcal{T})$. Therefore, the problem above is indeed a linear program. Since $\sum_{A\in\mathcal{F}}\phi_i^t(A)\,z^t(A)$ is the expected number of customers that choose product i at time period t, and $\sum_{\ell=1}^{\infty}q_{i,\ell}^t$ is the expected number of units of product i that are in use at time period t, the objective function computes the total expected revenue over the selling horizon. The first constraint keeps track of the expected numbers of products with different durations of use. The second constraint initializes the state of the system. The third constraint ensures that we offer an assortment at each time period, but this assortment can be empty. By the same argument in Section 2, because the products are all available on-hand at the beginning of the selling horizon, we have $q_{i,\ell}^t=0$ for all $\ell\geq T+1$ in a feasible solution to the linear program above. Thus, we do not need to define the decision variable $q_{i,\ell}^t$ for $\ell\geq T+1$, which indicates that the numbers of decision variables and constraints are finite. In the next proposition, we show that the optimal objective value of the linear program above is an upper bound on the optimal total expected revenue in our dynamic assortment problem. The proof follows from a standard argument in the revenue management literature. We defer the proof to Appendix I.

Proposition 6.1 *Letting Z^* be the optimal objective value of problem* (13)*, we have* $Z^* \geq J^1(\sum_{i \in \mathcal{N}} C_i \, \boldsymbol{e}_{i,0})$.

In problem (13), we have one decision variable $z^t(A)$ for each assortment $A \in \mathcal{F}$. Therefore, the number of decision variables increases exponentially with the number of products. Nevertheless, we can solve problem (13) by using column generation. In particular, noting that $\boldsymbol{q}^{t+1}$ in the first constraint in problem (13) corresponds to the vector $\boldsymbol{q}^{t+1} = (q^t_{i,\ell} : i \in \mathcal{N}, \ \ell \geq 0)$, we use $\boldsymbol{\alpha} = (\alpha^{t+1}_{i,\ell} : i \in \mathcal{N}, \ \ell \geq 0, \ t \in \mathcal{T} \setminus \{T\})$ to denote the dual variables associated with the first constraint. Similarly, noting that $\boldsymbol{q}^1$ in the second constraint in problem (13) corresponds to the vector $\boldsymbol{q}^1 = (q^1_{i,\ell} : i \in \mathcal{N}, \ \ell \geq 0)$, we use $\boldsymbol{\delta} = (\delta_{i,\ell} : i \in \mathcal{N}, \ \ell \geq 0)$ to denote the dual variables associated with the second constraint. Also, we use $\boldsymbol{\gamma} = (\gamma^t : t \in \mathcal{T})$ to denote the dual variables associated with the third constraint. In this case, the constraint associated with the decision variable $z^t(A)$ in the dual of problem (13) is $\sum_{i \in \mathcal{N}} \phi^t_i(A) \, (1 - \rho_{i,0}) \, (\alpha^{t+1}_{i,0} - \alpha^{t+1}_{i,1}) + \gamma^t \geq \sum_{i \in \mathcal{N}} \phi^t_i(A) \, (r^t_i + \pi^t_i)$. If we solve problem (13) with only a subset of the decision variables $(z^t(A) : A \in \mathcal{F}, \ t \in \mathcal{T})$ to obtain the dual solution $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\gamma}})$, then we can find which of the decision variables $(z^t(A) : A \in \mathcal{F}, \ t \in \mathcal{T})$ has the largest reduced cost by solving the problem

$$\max_{A \in \mathcal{F}} \left\{ \sum_{i \in \mathcal{N}} \phi^t_i(A) \, (r^t_i + \pi^t_i) - \sum_{i \in \mathcal{N}} \phi^t_i(A) \, (1 - \rho_{i,0}) \, (\hat{\alpha}^{t+1}_{i,0} - \hat{\alpha}^{t+1}_{i,1}) \right\}$$
$$= \max_{A \in \mathcal{F}} \sum_{i \in \mathcal{N}} \phi^t_i(A) \Big[r^t_i + \pi^t_i - (1 - \rho_{i,0}) \left(\hat{\alpha}^{t+1}_{i,0} - \hat{\alpha}^{t+1}_{i,1} \right) \Big]$$

for all $t \in \mathcal{T}$. In the problem above, we follow the convention that $\alpha^{T+1}_{i,0} = \alpha^{T+1}_{i,1} = 0$ for all $i \in \mathcal{N}$. The problem above is known as the column generation subproblem. The column generation subproblem above has the same structure as the maximization problem in (3). As discussed at the end of Section 2, this problem is tractable under a variety of choice models. Also, if the customers choose according to the multinomial logit model and there are no constraints on the assortments that we can offer, then we can build on the work of Gallego et al. (2015) to give an equivalent formulation for problem (13), whose numbers of decision variables and constraints increase linearly with the number of products. Therefore, we can directly solve the equivalent formulation without resorting to column generation. We discuss the equivalent formulation in Appendix J.

We formulate problem (13) under the assumption that there is a single customer type and we make assortment decisions. We can formulate analogues of problem (13) when we have multiple customer types and we make pricing decisions, which reflect the extensions we provided.

6.2 Dynamic Assortment Management

In our first set of computational experiments, the products are not reusable. We have access to a set of products with limited inventories. Customers arrive over time. Based on the remaining inventories of the products and the number of time periods left in the selling horizon, we offer an assortment to each arriving customer. The customer either purchases a product within the assortment or leaves without making a purchase. The purchased product is not returned, so the usage durations are infinite. Our goal is to find a policy to decide which assortment of products to offer to each customer so that we maximize the total expected revenue over the selling horizon.

Experimental Setup: In our test problems, we have six products indexed by $\mathcal{N} = \{1, \ldots, 6\}$ and six customer types indexed by $\mathcal{M} = \{1, \ldots, 6\}$. In Section 5.1, we discussed how to extend our model to the case with multiple customer types. Recalling that $\pi_i^{t,j}$ is the per-period rental fee that a customer of type j pays for product i at time period t, because the customers purchase the products outright, we set $\pi_i^{t,j} = 0$. The one-time upfront fee $r_i^{t,j}$ that a customer of type j pays for product i at time period t does not depend on the time period or the customer type. Thus, we use r_i to denote the upfront fee for product i.

To determine the upfront fees, we generate r_i from the uniform distribution over $[10, 25]$. After generating the upfront fees for all of the products, we reorder them so that $r_1 \geq r_2 \geq \ldots \geq r_n$. Thus, the first product has the largest upfront fee and the last product has the smallest upfront fee. The customers choose among the products according to the multinomial logit model. A customer of type j associates the preference weight v_i^j with product i and the preference weight v_0^j with the no-purchase option. If we offer the assortment S, then a customer of type j arriving at time period t chooses product $i \in S$ with probability $\phi_i^{t,j}(S) = v_i^j / (v_0^j + \sum_{\ell \in S} v_\ell^j)$. Note that the choice probabilities do not depend on the time period. To come up with the preference weights, we set the consideration set of customer type j as $\mathcal{C}_j = \{1, \ldots, j\}$. If $i \in \mathcal{C}_j$, then we generate v_i^j from the uniform distribution over $[0.9, 1.1]$, whereas if $i \notin \mathcal{C}_j$, then we set $v_i^j = 0$. Thus, a customer is interested in purchasing only the products in her consideration set. Among the products in her consideration set, she is somewhat indifferent. We calibrate the preference weight of the no-purchase option so that if we offer all products, then a customer leaves without a purchase with probability 0.1. Therefore, we calibrate v_0^j to satisfy $v_0^j / (v_0^j + \sum_{\ell \in \mathcal{C}^j} v_\ell^j) = 0.1$. Golrezaei et al. (2014) use a similar multinomial logit model with consideration sets in their computational experiments. Note that a customer of type n has the largest consideration set $\mathcal{C}_n = \{1, \ldots, n\}$, whereas a customer of type 1 has the smallest consideration set $\mathcal{C}_1 = \{1\}$. Therefore, customers of type n are the least choosy, whereas customers of type 1 are the most choosy.

In our test problems, the more choosy customers tend to arrive later in the selling horizon so that we need to carefully protect inventory for them. In particular, we choose equally-spaced time periods $\tau^n \le \tau^{n-1} \le \ldots \le \tau^1$ over the selling horizon. The probability $p^{t,j}$ that a customer of type j arrives at time period t is proportional to $e^{-\kappa\,|t-\tau^j|}$, where κ is a parameter that we vary. That is, we have $p^{t,j} = e^{-\kappa\,|t-\tau^j|} / \sum_{k\in\mathcal{M}} e^{-\kappa\,|t-\tau^k|}$. So, the arrival probability for a customer of type j peaks at around time period τ^j. Because $\tau^n \le \tau^{n-1} \le \ldots \le \tau^1$, as $\kappa \to \infty$, we obtain an arrival process where customers of type n arrive first, followed by customers of type $n-1$ and so on. As $\kappa \to 0$, we have $p^{t,j} \to 1/|\mathcal{M}|$, in which case, different customer types arrive with equal probability at each time period. Thus, we control the arrival order for the customer types through the parameter κ. The selling horizon has $T = 300$ time periods. The initial inventory of product i is $C_i = 30/\alpha$, where α is another parameter that we vary to control the inventory scarcity.

Varying the parameters (α, κ) over $\{0.7, 0.8, 0.9, 1.0\} \times \{0, 0.01, 0.03\}$, we obtain 12 test problems in our experimental setup.

Benchmarks: In our computational experiments, we compare the performance of the following seven benchmark strategies.

Greedy Policy (GR). In this benchmark, we use the greedy policy with respect to the linear value function approximations $\{\hat{J}^t : t \in \mathcal{T}\}$, as discussed in Section 3.

Rollout Policy (RO). This benchmark is the policy obtained by applying rollout on the static policy, as discussed in Section 4.

Bid-Prices (BP). We use the classical bid-price policy in this benchmark. We solve the linear program in (13) to estimate the value of a unit of inventory for each product, called its bid-price. We offer the revenue maximizing set of products at each time period, after adjusting the revenues from the products by their bid-prices; see Section 5.2 in Zhang and Adelman (2009).

Offer Sets (OS). We solve the linear program in (13) to obtain an optimal solution $(\hat{z}^t(A) : A \in \mathcal{F},\ t \in \mathcal{T})$ and $(\hat{q}^t_{i,\ell} : i \in \mathcal{N},\ \ell \ge 0,\ t \in \mathcal{T})$. Since $\sum_{A\in\mathcal{F}} \hat{z}^t(A) = 1$, letting N^t be the set of products with on-hand inventory at time period t, we sample an assortment S with respect to the probabilities $(\hat{z}^t(A) : A \in \mathcal{F})$ and offer the assortment $S \cap N^t$ at time period t. Offering the assortment $S \cap N^t$ ensures that we only offer products that are currently available.

Decomposition (DC). This benchmark is the classical dynamic programming decomposition method. The idea is to decompose the dynamic programming formulation of the problem by the products and to obtain value function approximations by solving a separate dynamic program for each product; see Section 6.2 in Liu and van Ryzin (2008). To our knowledge, this benchmark is one of the strongest heuristics in practice but it does not have a performance guarantee.

Myopic Policy (MY). We can construct myopic policies by ignoring the future customer arrivals altogether. In particular, if we are at time period t with $q^t_{i,0}$ units of product i on-hand, then the assortment that we offer to a customer of type j is given by an optimal solution to the problem $\max_{S\in\mathcal{F}} \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q^t_{i,0}\geq 1\}}\, \phi^{t,j}_i(S)\, r_i$. We use this benchmark to demonstrate the importance of considering the future customer arrivals when choosing an assortment to offer.

Inventory Balancing (IB). We implement the inventory balancing policy in Golrezaei et al. (2014). Letting $\Psi : [0,1] \to [0,1]$ be an increasing function with $\Psi(0) = 0$, if we are at time period t with $q^t_{i,0}$ units of product i on-hand, then the assortment that we offer to a customer of type j is given by an optimal solution to $\max_{S\in\mathcal{F}} \sum_{i\in\mathcal{N}} \Psi(q^t_{i,0}/C_i)\, \phi^{t,j}_i(S)\, r_i$. Following Golrezaei et al. (2014), we use $\Psi(x) = \frac{e}{e-1}\,(1-e^{-x})$. This policy has a half-approximation guarantee.

To further improve the performance of the benchmarks, we divide the selling horizon into three equal segments and recompute the policy parameters at the beginning of each segment. For GR, for example, if the remaining capacities of the products at the beginning of a segment are $(C'_i : i\in\mathcal{N})$ and the set of remaining time periods in the selling horizon is $\mathcal{T}' \subseteq \mathcal{T}$, then we apply the recursive computation at the beginning of Section 3.1 after replacing C_i with C'_i and $\mathcal{T}$ with $\mathcal{T}'$, which yields new value function approximations. We use the new value function approximations until we reach the next segment, at which point, we recompute the policy parameters. We use a similar approach to recompute the policy parameters for the other benchmarks, except for MY and IB. MY does not have any policy parameters to compute. The function Ψ in IB is fixed a priori.

Results: Table 1 shows our computational results. The first column in this table labels the test problems by using (α,κ), where α and κ are as discussed earlier in this section. The second column shows the upper bound on the optimal total expected revenue provided by the optimal objective value of problem (13). The third through ninth columns show the total expected revenues obtained by GR, RO, BP, OS, DC, MY and IB, which are estimated by simulating each benchmark over 1,000 sample paths. The remaining columns show the percent gaps between the total expected revenues obtained by RO and every other benchmark. The performance gaps except for those indicated with a star are statistically significant at the 95% level.

Our computational results indicate that RO performs quite well. By our use of separable and nonlinear value function approximations, RO noticeably improves the performance of GR, which uses linear value function approximations. Compared to BP and OS, which are based on the linear program in (13), RO provides average performance improvements of 4.7% and 1.5%, respectively. RO and DC are competitive, but to our knowledge, DC does not have a theoretical performance guarantee. Ignoring the future customer arrivals may result in inferior decisions, as indicated by

Params. (α, κ)	Upp. Bnd.	Total Expected Revenue GR	RO	BP	OS	DC	MY	IB	% Gain of RO over GR	BP	OS	DC	MY	IB
$(0.7, \quad 0)$	4,364	4,229	4,292	4,188	4,234	4,292	3,746	4,046	1.5	2.4	1.4	0.0*	12.7	5.7
$(0.7, 0.01)$	3,955	3,753	3,893	3,707	3,828	3,861	3,091	3,432	3.6	4.8	1.7	0.8	20.6	11.8
$(0.7, 0.03)$	3,925	3,726	3,854	3,572	3,783	3,858	2,691	3,168	3.3	7.3	1.8	-0.1*	30.2	17.8
$(0.8, \quad 0)$	4,026	3,924	3,959	3,865	3,907	3,951	3,510	3,763	0.9	2.4	1.3	0.2*	11.3	5.0
$(0.8, 0.01)$	3,937	3,790	3,873	3,668	3,820	3,882	3,051	3,454	2.1	5.3	1.4	-0.2	21.2	10.8
$(0.8, 0.03)$	3,934	3,801	3,882	3,526	3,810	3,901	3,151	3,318	2.1	9.2	1.9	-0.5	18.8	14.5
$(0.9, \quad 0)$	3,112	3,022	3,041	2,956	3,012	3,027	2,733	2,904	0.6	2.8	1.0	0.5	10.1	4.5
$(0.9, 0.01)$	2,809	2,718	2,773	2,645	2,732	2,779	2,325	2,487	2.0	4.6	1.5	-0.2	16.2	10.3
$(0.9, 0.03)$	3,084	2,977	3,040	2,914	2,994	3,073	2,511	2,603	2.1	4.1	1.5	-1.1	17.4	14.4
$(1.0, \quad 0)$	3,027	2,940	2,978	2,804	2,916	2,931	2,694	2,848	1.3	5.8	2.1	1.6	9.5	4.4
$(1.0, 0.01)$	2,483	2,415	2,467	2,389	2,425	2,463	2,070	2,212	2.1	3.2	1.7	0.2	16.1	10.3
$(1.0, 0.03)$	2,971	2,891	2,946	2,829	2,904	2,967	2,382	2,513	1.9	4.0	1.4	-0.7	19.1	14.7
Average									2.0	4.7	1.5	0.0	16.9	10.4

Table 1 Computational results for dynamic assortment management. Note GR = greedy policy, RO = rollout policy, BP = bid-prices, OS = offer sets, DC = decomposition, MY = myopic policy, and IB = inventory balancing.

the 16.9% average performance gap between RO and MY. The multiplicative revenue modifier $\Psi(q^t_{i,0}/C_i)$ used by IB does not depend on the future customer arrivals either. As a result, the average performance gap between RO and IB is 10.4%. When we compare RO with MY and IB, the performance gaps are particularly noticeable when κ is large; the more choosy customers in that case tend to arrive later, and we need to carefully protect inventory for these customers.

For GR, it takes 1.8 seconds on average to simulate its performance over one sample path. This computation time includes the time to recompute the policy parameters three times over the selling horizon and to solve problem (7) to find an assortment to offer at each time period. The same average computation time per sample path for RO is 4.1 seconds. The average computation times per sample path for BP, OS and DC are 147.3, 149.7 and 240.6 seconds, respectively. The average computation times per sample path for MY and IB are 0.8 and 0.7 seconds, respectively. Thus, beside its favorable revenues, RO has quite fast computation times. In Appendix K, we give the details of all computation times. Golrezaei et al. (2014) discuss possible variants of IB. We experimented with these variants but they did not provide qualitatively different results for our test problems. In Appendix L, we give our computational results on the variants of IB. In this section, the usage durations were infinite. In Appendix M, we test the performance of our policies under geometrically distributed usage durations with different means.

6.3 Street Parking Pricing in the City of Seattle

In our second set of computational experiments, we focus on the problem of dynamically pricing street parking spaces. We treat the parking spaces within close proximity to each other as one product. After having been used by a driver for a certain duration of time, a parking space can be used by another driver, so the parking spaces are reusable products. The dynamics of the problem are as follows. When a driver arrives into the system with an intention to park in a certain region, as a function of the remaining parking space inventory in the nearby regions, we decide on the prices to charge for the parking spaces in different regions. The driver is informed about the prices in real time, possibly through a smartphone application. The driver either parks at a particular parking space or decides to leave the system. If the driver parks, then the parking space generates revenue for a random usage duration. Our goal is to find a policy for deciding on the parking spaces to offer and their prices so that the total expected revenue is maximized.

Data: For brevity of discussion, we describe the essential elements of the data that we use, the approach that we use to augment the data for compliance with our modeling assumptions, and the methodology that we use to estimate the model parameters. We defer the details to Appendix N. We build on the data provided by the Open Data Program in the city of Seattle; see Seattle Open Data (2017). Seattle uses parking rates that are dependent on the location and the time of day. Through the Open Data Program, we have transaction data on the use of the street parking spaces during 20 weekdays of June 2017. Each transaction record shows a parking event, documenting the start time, duration, and location of the parking event, along with the rate paid. We focus on 40 blocks in the downtown area between the hours of 11AM and 4PM. We partition this area into 11 block clusters, each including approximately four blocks arranged in a two-by-two configuration. We refer to each two-by-two block cluster as a locale.

The street parking spaces in each locale correspond to a different product in our model. Thus, we have $n = 11$ products. To comply with our modeling assumptions, we augment the data from the Open Data Program as follows. We assume that each driver arrives into the system with the intention to park at a particular locale. The intended locale of a driver determines the type of the driver. In Section 5.1, we discussed the extension to multiple customer types. Because the intended locale of a driver determines her type, there are $m = 11$ customer types. In the data, we have access to the locale at which a driver actually parked, but we do not have access to the intended locale of a driver. For each driver, we randomly sample one of the five locales that are closest to the locale where she actually parked. We set the intended locale of the driver as this sampled locale. Once we augment the data in this way, each transaction record gives the start time, duration, intended locale, actual parked locale, and per-hour rate for each parking event. (The intended locale of a driver corresponds to the type of the customer, and the locale where a driver actually parks corresponds to the product the customer has chosen. If we had set the intended locale of a driver as the locale where she actually parked, then customers of a certain type would always be choosing the same product.) Because we augment the data from the Open Data Program, we caution the reader against comparing our results with the real operations in the city of Seattle.

The set of feasible locales we can offer to a driver are the five locales that are closest to her intended one. As a function of the remaining parking space inventories in these locales, we decide on the prices to charge for these locales. In Section 5.2, we discussed the extension of our model to the case in which we make pricing decisions. The driver either decides to park in one of these locales or leaves the system. If the driver parks, then we generate a certain revenue depending on the parking duration and the charged price. Although we have discussed the extensions of our model to multiple customer types and to pricing decisions separately, it is not difficult to combine these extensions and to come up with a variant of our model that makes pricing decisions under multiple customer types. It is also not difficult to extend the linear program in (13) to the case in which we make pricing decisions under multiple customer types.

Experimental Setup: As discussed in Section 5.2, when making pricing decisions, we create multiple copies of each product, whereby the different copies correspond to charging different prices for the product. As $\mathcal{N}$ corresponds to the set of possible parking locales, using $\mathcal{H}$ to denote the set of possible prices that we can charge for a parking space, offering product copy $(i,h) \in \mathcal{N} \times \mathcal{H}$ represents charging price level h for locale i. We use $\pi_{i,h}$ to denote the per-period fee when we charge the price level h for locale i. We assume that the choices of the drivers are governed by the multinomial logit model. So, if we offer the assortment $S \subseteq \mathcal{N} \times \mathcal{H}$ of locale and price combinations to a driver arriving at time period t with intended locale j, then she chooses to park in locale i with probability $\phi_i^{t,j}(S) = \frac{e^{\alpha^j + \beta\, \pi_{i,h}}}{1 + \sum_{(\ell,g) \in S} e^{\alpha^j + \beta\, \pi_{\ell,g}}}$ as long as $(i,h) \in S$. The parameter β, which captures the price sensitivity of the drivers, is assumed to be constant over all drivers.

Throughout the paper so far, we have assumed that there is one customer arrival at each time period. This assumption is not appropriate here because the arrival rate of the drivers vary during the day, but extending our model to the case in which there is at most one customer arrival at each time period is straightforward. We scale the time so that each time period in our model corresponds to a time interval of 30 seconds. A time interval of 30 seconds is short enough to ensure that there is at most one driver arrival in the region of our focus. We use $p^{t,j}$ to denote the probability that a driver with intended locale j arrives at time period t. We estimate the parameters β, $(\alpha^j : j \in \mathcal{M})$ and $(p^{t,j} : t \in \mathcal{T},\ j \in \mathcal{M})$ by using maximum likelihood.

We model the parking duration in locale i as $1 + \mathsf{NegBin}(s_i, \eta_i)$, where $\mathsf{NegBin}(s_i, \eta_i)$ is a negative binomial random variable with parameters $s_i \in \mathbb{Z}_{++}$ and $\eta_i \in [0,1]$. As discussed in Section 4.3, if s_i is small, then we can perform rollout on the static policy in a tractable fashion. For each locale i, a negative binomial distribution with the parameter $s_i = 2$ provides a sensible fit.

Ultimately, in our experimental setup, we vary the length of the selling horizon over two values, 11AM-2PM and 11AM-4PM. To obtain problems with different congestion levels, we scale the

Params. $(\mathcal{T}, \sigma, C)$	Upp. Bnd.	Total Expected Revenue				% Gain of RO over		
		RO	GR	OS	FP	GR	OS	FP
(11AM-2PM, 2.5, 79)	344	329	320	325	321	2.7	1.2	2.4
(11AM-2PM, 3.0, 79)	410	385	375	379	375	2.6	1.6	2.6
(11AM-2PM, 3.5, 79)	474	439	429	430	423	2.3	2.1	3.6
(11AM-2PM, 2.5, 55)	338	306	301	300	296	1.6	2.0	3.3
(11AM-2PM, 3.0, 55)	397	353	348	345	336	1.4	2.3	4.8
(11AM-2PM, 3.5, 55)	452	396	390	385	370	1.5	2.8	6.6
(11AM-4PM, 2.5, 79)	631	591	575	582	575	2.7	1.5	2.7
(11AM-4PM, 3.0, 79)	749	689	673	674	667	2.3	2.2	3.2
(11AM-4PM, 3.5, 79)	860	781	766	761	747	1.9	2.6	4.4
(11AM-4PM, 2.5, 55)	613	543	534	528	520	1.7	2.8	4.2
(11AM-4PM, 3.0, 55)	717	624	614	609	587	1.6	2.4	5.9
(11AM-4PM, 3.5, 55)	812	696	686	679	644	1.4	2.4	7.5
Average						2.0	2.1	4.3

Table 2 Computational results for street parking pricing in the city of Seattle.

arrival rates with three different factors, 2.5, 3.0 and 3.5. Also, we vary the number of parking spaces over two values, 55 and 79. This experimental setup yields 12 parameter combinations for our test problems. From the rates used by the city of Seattle, the possible rates that we can charge are within the menu of \$2, \$4 and \$6 per hour.

Benchmarks: We use the benchmarks greedy policy (GR), rollout policy (RO) and offer sets (OS), which are discussed in Section 6.2. We make the necessary modifications in these benchmarks to ensure that we can handle multiple customer types and we choose the prices of the offered products. The performances of bid-prices (BP) and myopic policy (MY) were not competitive. Decomposition (DC) and inventory balancing (IB) do not extend to reusable products. Thus, we drop these four benchmarks. We also add the following benchmark.

Fixed Price (FP). Here, we charge one fixed price for all locales at all time periods. We test the performance of the rates \$2, \$4 and \$6 per hour, which is the price menu used by the other benchmarks. We select the best constant price. This benchmark is not sophisticated but it serves as a baseline. In all test problems, the rate \$4 per hour provided the best performance.

Results: Table 2 shows our computational results. The first column in this table labels the test problems by using $(\mathcal{T}, \sigma, C)$, where $\mathcal{T} \in \{\text{11AM-2PM}, \text{11AM-4PM}\}$ is the selling horizon, $\sigma \in \{2.5, 3.0, 3.5\}$ is the multiplier for the arrival rates, and $C \in \{55, 79\}$ is the total number of parking spaces. The organization of the rest of the table closely mirrors that of Table 1. All of the performance gaps in Table 2 are statistically significant. To get a feel for the congestion in our test problems, noting that $\mathbb{E}\{\mathsf{Duration}_i\}$ is the expected parking duration in locale i, we can estimate the number of times that we can turn over a parking space in locale i as $T/\mathbb{E}\{\mathsf{Duration}_i\}$. The total expected demand for parking is $\sum_{t \in \mathcal{T}} \sum_{j \in \mathcal{M}} p^{t,j}$. Thus, with C_i parking spaces available in locale i, the ratio between the total expected demand and the total available capacity is

$\frac{\sum_{t\in\mathcal{T}}\sum_{j\in\mathcal{M}} p^{t,j}}{\sum_{i\in\mathcal{N}} C_i\, T/\mathbb{E}\{\mathsf{Duration}_i\}}$. For the test problems having the smallest demand and the largest capacity with $\sigma = 2.5$ and $C = 79$, this ratio is 0.72, whereas for the test problems having the largest demand and the smallest capacity with $\sigma = 3.5$ and $C = 55$, this ratio is 1.61.

Our results indicate that RO is consistently the strongest benchmark, providing average performance improvements of 2.0%, 2.1% and 4.3% over GR, OS and FP, respectively. Comparing RO with OS and FP, the performance gaps tend to be larger when σ is larger and C is smaller so that the system is more congested and the expected demand exceeds the available capacity by larger margins. For our test problems, depending on the length of the selling horizon, the time to compute the value functions $\{\hat{J}^t : t \in \mathcal{T}\}$ for GR ranges from 362 to 672 seconds. The time to compute the value functions $\{V_i^t : i \in \mathcal{N},\ t \in \mathcal{T}\}$ for RO ranges from 1,550 to 17,201 seconds. For OS, the time to solve the linear program in (13) ranges from 894 to 6,535 seconds. A few preliminary runs indicated that the performance did not noticeably improve for any of the benchmarks when we recomputed the policy parameters. Because the run times were relatively long, we did not recompute the policy parameters. Overall, the computation times for RO are significantly longer, but using nonlinear value function approximations, RO can provide significant revenue improvements.

7. Conclusions

We studied dynamic assortment problems with reusable products, and provided policies with half-approximate performance guarantees. A natural question that arises is what features of the rewards and transition dynamics make our half-approximation guarantees go through. In Appendix O, we give conditions on the rewards and transition dynamics that allow us to obtain our half-approximation guarantees. Our conditions on the transition dynamics, in particular, only require the expected transition dynamics to be linear in the state, along with a certain decoupling property between the effects of the actions and the current state on the transition dynamics. Considering future research, our rollout approach decomposes the problem by the products, which is reminiscent of dynamic programming decomposition techniques in revenue management. To our knowledge, existing decomposition techniques do not provide any performance guarantees. An exciting research area is to construct decomposition techniques with performance guarantees for other revenue management problems. We were able to give an improved performance guarantee for a tailored variant of our rollout policy under infinite usage durations. Our efforts to extend the tailored variant to arbitrary usage duration distributions were not yet fruitful. It would be interesting to give stronger performance guarantees for our rollout approach under arbitrary usage duration distributions.

Acknowledgements: We thank the area editor, Professor Yinyu Ye, the associate editor and three anonymous referees for their useful comments, which improved the paper in many ways. The first and third authors were supported in part by National Science Foundation grants CMMI-1825406 and CMMI-1824860, and the second author is supported by the NSF Graduate Research Fellowship. The work of the third author was partially funded by Schmidt Sciences.

References

Adelman, D. 2007. Dynamic bid prices in revenue management. *Operations Research* **55**(4) 647–661.

Aouad, A., R. Levi, D. Segev. 2015. Approximation algorithms for dynamic assortment optimization models. Tech. rep., Massachusetts Institute of Technology, Cambridge, MA.

Aouad, A., R. Levi, D. Segev. 2017. Greedy-like algorithms for dynamic assortment planning under multinomial logit preferences. Tech. rep., Massachusetts Institute of Technology, Cambridge, MA.

Bertsekas, D. P., J. N. Tsitsiklis. 1996. *Neuro-Dynamic Programming*. Athena Scientific, Belmont, MA.

Chen, Y., R. Levi, C. Shi. 2017. Revenue management of reusable resources with advanced reservations. *Production and Operations Management* **26**(5) 836–859.

Dai, J., W. Ding, A. J. Kleywegt, X. Wang, Y. Zhang. 2014. Choice based revenue management for parallel flights. Tech. rep., Georgia Tech, Atlanta, GA.

Davis, J. M., G. Gallego, H. Topaloglu. 2013. Assortment planning under the multinomial logit model with totally unimodular constraint structures. Tech. rep., Cornell University, Ithaca, NY.

Davis, J. M., G. Gallego, H. Topaloglu. 2014. Assortment optimization under variants of the nested logit model. *Operations Research* **62**(2) 250–273.

Desire, A., V. Goyal, J. Zhang. 2016. Near-optimal algorithms for capacity constrained assortment optimization. Tech. rep., Columbia University, New York, NY.

Feldman, J. B., H. Topaloglu. 2015. Technical note: Capacity constraints across nests in assortment optimization under the nested logit model. *Operations Research* **63**(4) 812–822.

Gallego, G. 1992. A minmax distribution free procedure for the (q, r) inventory model. *Operations Research Letters* **11**(1) 55–60.

Gallego, G., G. Iyengar, R. Phillips, A. Dubey. 2004. Managing flexible products on a network. CORC Technical Report TR-2004-01.

Gallego, G., A. Li, V. A. Truong, X. Wang. 2016. Online bipartite matching. Working Paper, Columbia University.

Gallego, G., R. Ratliff, S. Shebalov. 2015. A general attraction model and sales-based linear programming formulation for network revenue management under customer choice. *Operations Research* **63**(1) 212–232.

Gallego, G., H. Topaloglu. 2014. Constrained assortment optimization for the nested logit model. *Management Science* **60**(10) 2583–2601.

Gallego, G., G. van Ryzin. 1994. Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Management Science* **40**(8) 999–1020.

Golrezaei, N., H. Nazerzadeh, P. Rusmevichientong. 2014. Real-time optimization of personalized assortments. *Management Science* **60**(6) 1532–1551.

Goyal, V., R. Levi, D. Segev. 2016. Near-optimal algorithms for the assortment planning problem under dynamic substitution and stochastic demand. *Operations Research* **64**(1) 219–235.

Honhon, D., V. Gaur, S. Seshadri. 2010. Assortment planning and inventory decisions under stockout-based substitution. *Operations Research* **58**(5) 1364–1379.

Lei, Y., S. Jasin. 2016. Real-time dynamic pricing for revenue management with reusable resources and deterministic service time requirements. Tech. rep., University of Michigan, Ann Arbor, MI.

Levi, R., A. Radovanovic. 2010. Provably near-optimal LP-based policies for revenue management in systems with reusable resources. *Operations Research* **58**(2) 503–507.

Li, G., P. Rusmevichientong, H. Topaloglu. 2015. The *d*-level nested logit model: Assortment and price optimization problems. *Operations Research* **63**(2) 325–342.

Liu, Q., G. J. van Ryzin. 2008. On the choice-based linear programming model for network revenue management. *Manufacturing & Service Operations Management* **10**(2) 288–310.

Mahajan, S., G. van Ryzin. 2001. Stocking retail assortments under dynamic customer substitution. *Operations Research* **49**(3) 334–351.

Owen, Z., D. Simchi-Levi. 2017. Price and assortment optimization for reusable resources. Tech. rep., Massachusetts Institute of Technology, Cambridge, MA.

Ozer, O., R. Phillips, eds. 2012. *The Oxford Handbook of Pricing Management*. Oxford University Press, UK.

Rusmevichientong, P., D. Shmoys, C. Tong, H. Topaloglu. 2014. Assortment optimization under the multinomial logit model with random choice parameters. *Production and Operations Management* **23**(11) 2023–2039.

Seattle Open Data. 2017. Seattle parking transactions. URL `https://data.seattle.gov/Transportation/Seattle-Parking-Transactions/updn-y53g`.

Stein, C., V. A. Truong, X. Wang. 2016. Advance reservations with heterogeneous customers. Working Paper, Columbia University.

Talluri, K., G. J. van Ryzin. 2004. Revenue management under a general discrete choice model of consumer behavior. *Management Science* **50**(1) 15–33.

Talluri, K. T., G. J. van Ryzin. 2005. *The theory and practice of revenue management*. Kluwer Academic Publishers, Boston, MA.

Tong, C., H. Topaloglu. 2013. On the approximate linear programming approach for network revenue management problems. *INFORMS Journal on Computing* **26**(1) 121–134.

Topaloglu, H. 2013. Joint stocking and product offer decisions under the multinomial logit model. *Production and Operations Management* **22**(5) 1182–1199.

van Ryzin, G., S. Mahajan. 1999. On the relationship between inventory costs and variety benefits in retail assortments. *Management Science* **45**(11) 1496–1509.

Vossen, T. W. M., D. Zhang. 2015. Reductions of approximate linear programs for network revenue management. *Operations Research* **63**(6) 1352–1371.

Vulcano, G., G. J. van Ryzin, R. Ratliff. 2012. Estimating primary demand for substitutable products from sales transaction data. *Operations Research* **60**(2) 313–334.

Wang, X., V. A. Truong, D. Bank. 2016. Online advance admission scheduling for services, with customer preferences. Working Paper, Columbia University.

Zhang, D., D. Adelman. 2009. An approximate dynamic programming approach to network revenue management with customer choice. *Transportation Science* **43**(3) 381–394.

Zhang, D., W. L. Cooper. 2005. Revenue management for parallel flights with customer choice behavior. *Operations Research* **53**(3) 415–431.

Zhang, H., P. Rusmevichientong, H. Topaloglu. 2017. Assortment optimization under the paired combinatorial logit model. Tech. rep., Cornell University, Ithaca, NY.

This page is intentionally blank. Proper e-companion title page, with INFORMS branding and exact metadata of the main paper, will be produced by the INFORMS office when the issue is being assembled.

Online Appendix

Dynamic Assortment Optimization for Reusable Products with Random Usage Duration

Appendix A: Non-Separability of the Optimal Value Functions

Consider a problem instance with two products and one time period in the selling horizon, so that $\mathcal{N}=\{1,2\}$ and $\mathcal{T}=\{1\}$. The one-time upfront fees are $r_i^1=1$ for all $i\in\mathcal{N}$, whereas the per-period rental fees are $\pi_i^1=0$ for all $i\in\mathcal{N}$. The usage times for both products are deterministic and equal to one time period. If we offer both products, then the choice probabilities are $\phi_i^1(\{1,2\})=1/3$ for all $i\in\mathcal{N}$. If we offer only product 1, then the choice probability is $\phi_1^1(\{1\})=1/2$, whereas if we offer only product 2, then the choice probability is $\phi_2^1(\{2\})=1/2$. It is simple to check that these choice probabilities satisfy Assumption 2.1. It is feasible to offer all assortments, so the set of feasible assortments is $\mathcal{F}=\{\varnothing,\{1\},\{2\},\{1,2\}\}$. Since both products have the same upfront fee and rent, it is simple to check that the optimal policy offers all products for which we have on-hand inventory. Thus, using $J^1(a,b)$ to denote the optimal total expected revenue when we have a units of on-hand inventory for product 1 and b units of on-hand inventory for product 2, we have $J^1(0,0)=0$, $J^1(1,0)=\phi_1^1(\{1\})\times r_1^1=\frac{1}{2}$, $J^1(0,1)=\phi_2^1(\{2\})\times r_2^1=\frac{1}{2}$ and $J^1(1,1)=\phi_1^1(\{1,2\})\times r_1^1+\phi_2^1(\{1,2\})\times r_2^1=\frac{2}{3}$. Thus, we get $J^1(1,0)-J^1(0,0)=\frac{1}{2}\neq\frac{1}{6}=J^1(1,1)-J^1(0,1)$, indicating that the optimal value function is not separable by the products.

Appendix B: Expected Contribution of Ideal Assortment

For notational brevity, we let $\Delta_i^t=r_i^t+\pi_i^t-(1-\rho_{i,0})\left(\hat{\nu}_{i,0}^{t+1}-\hat{\nu}_{i,1}^{t+1}\right)$. In the next lemma, letting $\hat{A}^t$ be as in (5), we show that $\sum_{i\in\mathcal{N}}\phi_i^t(\hat{A}^t)\,\Delta_i^t\geq\sum_{i\in\mathcal{N}}\mathbb{1}_{\{q_{i,0}\geq 1\}}\,\phi_i^t(S)\,\Delta_i^t$ for all $S\in\mathcal{F}$.

Lemma B.1 *For all $S\in\mathcal{F}$, we have $\sum_{i\in\mathcal{N}}\phi_i^t(\hat{A}^t)\,\Delta_i^t\geq\sum_{i\in\mathcal{N}}\mathbb{1}_{\{q_{i,0}\geq 1\}}\,\phi_i^t(S)\,\Delta_i^t$.*

Proof: Note that we have $\mathbb{1}_{\{\Delta_i^t\geq 0\}}\,\Delta_i^t\geq\mathbb{1}_{\{\Delta_i^t\geq 0\}}\,\mathbb{1}_{\{q_{i,0}\geq 1\}}\,\Delta_i^t\geq\mathbb{1}_{\{q_{i,0}\geq 1\}}\,\Delta_i^t$, where the first inequality is by the fact that $\mathbb{1}_{\{\Delta_i^t\geq 0\}}\,\Delta_i^t\geq 0$ and the second inequality is by the fact that $\mathbb{1}_{\{\Delta_i^t\geq 0\}}\,\mathbb{1}_{\{q_{i,0}\geq 1\}}\,\Delta_i^t\geq 0$, but $\mathbb{1}_{\{q_{i,0}\geq 1\}}\,\Delta_i^t$ can be positive or negative. We will shortly establish the claim that $\sum_{i\in\mathcal{N}}\phi_i^t(\hat{A}^t)\,\Delta_i^t\geq\sum_{i\in\mathcal{N}}\phi_i^t(S)\,\mathbb{1}_{\{\Delta_i^t\geq 0\}}\,\Delta_i^t$ for all $S\in\mathcal{F}$. In this case, noting the chain of inequalities at the beginning of the proof, we obtain $\sum_{i\in\mathcal{N}}\phi_i^t(\hat{A}^t)\,\Delta_i^t\geq\sum_{i\in\mathcal{N}}\phi_i^t(S)\,\mathbb{1}_{\{\Delta_i^t\geq 0\}}\,\Delta_i^t\geq\sum_{i\in\mathcal{N}}\phi_i^t(S)\,\mathbb{1}_{\{\Delta_i^t\geq 0\}}\,\mathbb{1}_{\{q_{i,0}\geq 1\}}\,\Delta_i^t\geq\sum_{i\in\mathcal{N}}\phi_i^t(S)\,\mathbb{1}_{\{q_{i,0}\geq 1\}}\,\Delta_i^t$ for all $S\in\mathcal{F}$, which is the desired result. We proceed to establish the claim that $\sum_{i\in\mathcal{N}}\phi_i^t(\hat{A}^t)\,\Delta_i^t\geq\sum_{i\in\mathcal{N}}\phi_i^t(S)\,\mathbb{1}_{\{\Delta_i^t\geq 0\}}\,\Delta_i^t$ for all $S\in\mathcal{F}$. Assume on the contrary that there exists $\hat{S}\in\mathcal{F}$ such that $\sum_{i\in\mathcal{N}}\phi_i^t(\hat{A}^t)\,\Delta_i^t<$

$\sum_{i\in\mathcal{N}} \phi_i^t(\hat{S})\, \mathbb{1}_{\{\Delta_i^t \geq 0\}}\, \Delta_i^t$. We define the assortment S^* as $S^* = \{i \in \hat{S} : \Delta_i^t \geq 0\}$. Therefore, we have $S^* \subseteq \hat{S} \subseteq \mathcal{N}$, in which case, since $\hat{S} \in \mathcal{F}$, we have $S^* \in \mathcal{F}$ by Assumption 2.1. For all $i \in S^*$, we have $\Delta_i^t \geq 0$ by the definition of S^*, so $\sum_{i\in S^*} \phi_i^t(S^*)\, \Delta_i^t = \sum_{i\in S^*} \phi_i^t(S^*)\, \mathbb{1}_{\{\Delta_i^t \geq 0\}}\, \Delta_i^t$. Also, for all $i \in \hat{S} \setminus S^*$, we have $\phi_i^t(S^*) = 0$ and $\Delta_i^t < 0$, in which case, we have $\sum_{i\in\hat{S}\setminus S^*} \phi_i^t(S^*)\, \Delta_i^t = 0 = \sum_{i\in\hat{S}\setminus S^*} \phi_i^t(\hat{S})\, \mathbb{1}_{\{\Delta_i^t \geq 0\}}\, \Delta_i^t$. Lastly, for all $i \in \mathcal{N} \setminus \hat{S}$, we have $\phi_i^t(S^*) = 0 = \phi_i^t(\hat{S})$. Therefore, we have $\sum_{i\in\mathcal{N}\setminus\hat{S}} \phi_i^t(S^*)\, \Delta_i^t = 0 = \sum_{i\in\mathcal{N}\setminus\hat{S}} \phi_i^t(\hat{S})\, \mathbb{1}_{\{\Delta_i^t \geq 0\}}\, \Delta_i^t$. Noting Assumption 2.1, because $S^* \subseteq \hat{S}$, we have $\phi_i^t(S^*) \geq \phi_i^t(\hat{S})$ for all $i \in S^*$. In this case, we obtain

$$\begin{aligned}
\sum_{i\in\mathcal{N}} \phi_i^t(S^*)\, \Delta_i^t &= \sum_{i\in S^*} \phi_i^t(S^*)\, \Delta_i^t + \sum_{i\in\hat{S}\setminus S^*} \phi_i^t(S^*)\, \Delta_i^t + \sum_{i\in\mathcal{N}\setminus\hat{S}} \phi_i^t(S^*)\, \Delta_i^t \\
&= \sum_{i\in S^*} \phi_i^t(S^*)\, \mathbb{1}_{\{\Delta_i^t \geq 0\}}\, \Delta_i^t + \sum_{i\in\hat{S}\setminus S^*} \phi_i^t(\hat{S})\, \mathbb{1}_{\{\Delta_i^t \geq 0\}}\, \Delta_i^t + \sum_{i\in\mathcal{N}\setminus\hat{S}} \phi_i^t(\hat{S})\, \mathbb{1}_{\{\Delta_i^t \geq 0\}}\, \Delta_i^t \\
&\geq \sum_{i\in S^*} \phi_i^t(\hat{S})\, \mathbb{1}_{\{\Delta_i^t \geq 0\}}\, \Delta_i^t + \sum_{i\in\hat{S}\setminus S^*} \phi_i^t(\hat{S})\, \mathbb{1}_{\{\Delta_i^t \geq 0\}}\, \Delta_i^t + \sum_{i\in\mathcal{N}\setminus\hat{S}} \phi_i^t(\hat{S})\, \mathbb{1}_{\{\Delta_i^t \geq 0\}}\, \Delta_i^t \\
&= \sum_{i\in\mathcal{N}} \phi_i^t(\hat{S})\, \mathbb{1}_{\{\Delta_i^t \geq 0\}}\, \Delta_i^t.
\end{aligned}$$

Thus, noting the assumption $\sum_{i\in\mathcal{N}} \phi_i(\hat{S})\, \mathbb{1}_{\{\Delta_i^t \geq 0\}}\, \Delta_i^t > \sum_{i\in\mathcal{N}} \phi_i(\hat{A}^t)\, \Delta_i^t$, we get $\sum_{i\in\mathcal{N}} \phi_i(S^*)\, \Delta_i^t > \sum_{i\in\mathcal{N}} \phi_i(\hat{A}^t)\, \Delta_i^t$, which contradicts the fact that $\hat{A}^t$ is an optimal solution to problem (5). ■

Appendix C: Performance of Myopic Policies

Consider a problem instance with one product and two time periods in the selling horizon, so that $\mathcal{N} = \{1\}$ and $\mathcal{T} = \{1, 2\}$. The capacity of the product is $C_1 = 1$. For some $\epsilon > 0$, the one-time upfront fees for the product are $r_1^1 = \epsilon$ and $r_1^2 = 1$. The per-period rental fees are $\pi_1^t = 0$ for all $t \in \mathcal{T}$. The usage duration for the product is infinite, so once the product is rented, it is never returned. For all $t \in \mathcal{T}$, we have the choice probability $\phi_1^t(\{1\}) = \frac{1}{1+\epsilon}$. The set of feasible assortments is $\mathcal{F} = \{\varnothing, \{1\}\}$. Because the per-period rental fees are zero, at each time period t, the myopic policy chooses an assortment to offer by solving the problem $\max_{S\in\mathcal{F}} \sum_{i\in\mathcal{N}} \phi_i^t(S)\, r_i^t$. Thus, the myopic policy always offers the product, as long as there is on-hand inventory. Noting that we have only one unit of inventory for the product, the total expected revenue obtained by the myopic policy is

$$\frac{1}{1+\epsilon}\, r_1^1 + \left(1 - \frac{1}{1+\epsilon}\right) \frac{1}{1+\epsilon}\, r_1^2 = \frac{1}{1+\epsilon}\, \epsilon + \left(1 - \frac{1}{1+\epsilon}\right) \frac{1}{1+\epsilon} = \frac{\epsilon\,(2+\epsilon)}{(1+\epsilon)^2},$$

where we use the fact that obtaining any revenue at the second time period requires a customer not choosing the product at the first time period and choosing the product at the second time period. On the other hand, considering a policy that never offers the product at the first time period and always offers the product at the second time period, the expected revenue obtained by this policy is $\frac{1}{1+\epsilon}\, r_1^2 = \frac{1}{1+\epsilon}$. Thus, the ratio between the total expected revenues obtained by the optimal policy and the myopic policy is at least $\frac{1/(1+\epsilon)}{\epsilon\,(2+\epsilon)/(1+\epsilon)^2} = \frac{1+\epsilon}{\epsilon\,(2+\epsilon)}$. Since $\lim_{\epsilon\to 0} \frac{1+\epsilon}{\epsilon\,(2+\epsilon)} = \infty$, if we choose ϵ arbitrarily small, then the myopic policy performs arbitrarily poorly when compared with the optimal policy. In contrast, our greedy policy with respect to the linear value function approximations is always half-approximate. Thus, if ϵ is arbitrarily small, then the myopic policy also performs arbitrarily poorly when compared with our greedy policy with respect to the linear value function approximations.

Appendix D: Performance of the Static Policy

In this section, we give a proof of Lemma 4.1, which shows that the total expected revenue obtained by the static policy is at least 50% of the optimal total expected revenue.

Proof of Lemma 4.1: Under the static policy, we offer the assortment $\hat{A}^t$ at time period t regardless of the product availabilities, where $\hat{A}^t$ is given by an optimal solution to problem (5). If a customer chooses a product that does not have on-hand inventory, then the customer leaves without using the product. Let $V^t(\boldsymbol{q})$ denote the total expected revenue under this static policy over the time periods $t,\ldots,T$, given that we are in state $\boldsymbol{q}$ at time period t. Similar to the dynamic program in (3), we can compute $\{V^t : t \in \mathcal{T}\}$ by using the recursion

$$V^t(\boldsymbol{q}) = \sum_{i\in\mathcal{N}} \pi_i^t \sum_{\ell=1}^{\infty} q_{i,\ell} \;+\; \mathbb{E}\Big\{V^{t+1}(\boldsymbol{X}(\boldsymbol{q}))\Big\}$$
$$+ \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\,\geq\,1\}}\, \phi_i^t(\hat{A}^t)\bigg(r_i^t + \pi_i^t - (1-\rho_{i,0})\mathbb{E}\Big\{V^{t+1}(\boldsymbol{X}(\boldsymbol{q})) - V^{t+1}\left(\boldsymbol{X}(\boldsymbol{q}) - \boldsymbol{e}_{i,0} + \boldsymbol{e}_{i,1}\right)\Big\}\bigg),$$

with the boundary condition that $V^{T+1} = 0$. Consider the function $\hat{J}^t(\boldsymbol{q}) = \sum_{i\in\mathcal{N}} \sum_{\ell=0}^{\infty} q_{i,\ell}\, \hat{\nu}_{i,\ell}^t$, where the parameters $\{\hat{\nu}_{i,\ell}^t : i \in \mathcal{N},\ \ell \geq 0,\ t \in \mathcal{T}\}$ are obtained by using the recursion at the beginning of Section 3.1. We will use induction over the time periods to show that $V^t(\boldsymbol{q}) \geq \hat{J}^t(\boldsymbol{q})$ for all $\boldsymbol{q} \in \mathcal{Q}$ and $t \in \mathcal{T}$. Because $\hat{\nu}_{i,\ell}^{T+1} = 0$ for all $i \in \mathcal{N}$, $\ell = 0, 1, \ldots$ and $V^{T+1} = 0$, the result holds at time period $T+1$. Assuming that $V^{t+1}(\boldsymbol{q}) \geq \hat{J}^{t+1}(\boldsymbol{q})$ for all $\boldsymbol{q} \in \mathcal{Q}$, we will show that $V^t(\boldsymbol{q}) \geq \hat{J}^t(\boldsymbol{q})$ for all $\boldsymbol{q} \in \mathcal{Q}$. In the proof of Theorem 3.2, we show the equalities

$$\mathbb{E}\Big\{\hat{J}^{t+1}(\boldsymbol{X}(\boldsymbol{q}))\Big\} \;=\; \sum_{i\in\mathcal{N}} \bigg\{ q_{i,0}\, \hat{\nu}_{i,0}^{t+1} + \sum_{\ell=1}^{\infty} q_{i,\ell}\, \left[\rho_{i,\ell}\, \hat{\nu}_{i,0}^{t+1} + (1-\rho_{i,\ell})\, \hat{\nu}_{i,\ell+1}^{t+1}\right] \bigg\}.$$
$$\mathbb{E}\Big\{\hat{J}^{t+1}(\boldsymbol{X}(\boldsymbol{q})) - \hat{J}^{t+1}\left(\boldsymbol{X}(\boldsymbol{q}) - \boldsymbol{e}_{i,0} + \boldsymbol{e}_{i,1}\right)\Big\} \;=\; \hat{\nu}_{i,0}^{t+1} - \hat{\nu}_{i,1}^{t+1}\;.$$

Furthermore, by the claim that we establish in the proof of Lemma 3.1, recall that we have the inequality $\phi_i^t(\hat{A}^t)\,[r_i^t + \pi_i^t - (1-\rho_{i,0})\left(\hat{\nu}_{i,0}^{t+1} - \hat{\nu}_{i,1}^{t+1}\right)] \geq 0$ for all $i \in \mathcal{N}$. In this case, by the inductive hypothesis that $V^{t+1}(\boldsymbol{q}) \geq \hat{J}^{t+1}(\boldsymbol{q})$ for all $\boldsymbol{q} \in \mathcal{Q}$ and the above recursion defining $V^t(\boldsymbol{q})$, we obtain the chain of inequalities

$$\begin{aligned}
V^t(\boldsymbol{q}) \;\; & \geq \;\; \sum_{i\in\mathcal{N}} \pi_i^t \sum_{\ell=1}^{\infty} q_{i,\ell} \; + \; \mathbb{E}\Big\{ \hat{J}^{t+1}(\boldsymbol{X}(\boldsymbol{q})) \Big\} \\
& \quad + \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\,\geq\,1\}}\, \phi_i^t(\hat{A}^t) \Big(r_i^t + \pi_i^t - (1-\rho_{i,0})\, \mathbb{E}\Big\{ \hat{J}^{t+1}\left(\boldsymbol{X}(\boldsymbol{q})\right) - \hat{J}^{t+1}\left(\boldsymbol{X}(\boldsymbol{q}) - \boldsymbol{e}_{i,0} + \boldsymbol{e}_{i,1}\right) \Big\} \Big) \\
& = \;\; \sum_{i\in\mathcal{N}} \pi_i^t \sum_{\ell=1}^{\infty} q_{i,\ell} \; + \; \sum_{i\in\mathcal{N}} \left\{ q_{i,0}\, \hat{\nu}_{i,0}^{t+1} + \sum_{\ell=1}^{\infty} q_{i,\ell} \left[\rho_{i,\ell}\, \hat{\nu}_{i,0}^{t+1} + (1-\rho_{i,\ell})\, \hat{\nu}_{i,\ell+1}^{t+1} \right] \right\} \\
& \qquad\quad + \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\,\geq\,1\}}\, \phi_i^t(\hat{A}^t) \left[r_i^t + \pi_i^t - (1-\rho_{i,0})(\hat{\nu}_{i,0}^{t+1} - \hat{\nu}_{i,1}^{t+1}) \right] \\
& \geq \;\; \sum_{i\in\mathcal{N}} \pi_i^t \sum_{\ell=1}^{\infty} q_{i,\ell} \; + \; \sum_{i\in\mathcal{N}} \left\{ q_{i,0}\, \hat{\nu}_{i,0}^{t+1} + \sum_{\ell=1}^{\infty} q_{i,\ell} \left[\rho_{i,\ell}\, \hat{\nu}_{i,0}^{t+1} + (1-\rho_{i,\ell})\, \hat{\nu}_{i,\ell+1}^{t+1} \right] \right\} \\
& \qquad\quad + \sum_{i\in\mathcal{N}} \frac{q_{i,0}}{C_i}\, \phi_i^t(\hat{A}^t) \left[r_i^t + \pi_i^t - (1-\rho_{i,0})(\hat{\nu}_{i,0}^{t+1} - \hat{\nu}_{i,1}^{t+1}) \right],
\end{aligned}$$

where the last inequality uses the fact that $\phi_i^t(\hat{A}^t)\left[r_i^t + \pi_i^t - (1-\rho_{i,0})(\hat{\nu}_{i,0}^{t+1} - \hat{\nu}_{i,1}^{t+1}) \right] \geq 0$ and $\mathbb{1}_{\{q_{i,0}\,\geq\,1\}} \geq q_{i,0}/C_i$ for all $\boldsymbol{q} \in \mathcal{Q}$. By the definition of $\hat{\nu}_{i,0}^t$ in (6), we have $\hat{\nu}_{i,0}^t - \hat{\nu}_{i,0}^{t+1} = \frac{1}{C_i}\,\phi_i^t(\hat{A}^t) \times [r_i^t + \pi_i^t - (1-\rho_{i,0})(\hat{\nu}_{i,0}^{t+1} - \hat{\nu}_{i,1}^{t+1})]$. In this case, the expression on the right side of the chain of inequalities above can equivalently be written as

$$\begin{aligned}
& \sum_{i\in\mathcal{N}} \pi_i^t \sum_{\ell=1}^{\infty} q_{i,\ell} \; + \; \sum_{i\in\mathcal{N}} \left\{ q_{i,0}\, \hat{\nu}_{i,0}^{t+1} + \sum_{\ell=1}^{\infty} q_{i,\ell} \left[\rho_{i,\ell}\, \hat{\nu}_{i,0}^{t+1} + (1-\rho_{i,\ell})\, \hat{\nu}_{i,\ell+1}^{t+1} \right] \right\} + \sum_{i\in\mathcal{N}} q_{i,0}\,(\hat{\nu}_{i,0}^t - \hat{\nu}_{i,0}^{t+1}) \\
& \quad = \; \sum_{i\in\mathcal{N}} \left\{ q_{i,0}\, \hat{\nu}_{i,0}^{t+1} + \sum_{\ell=1}^{\infty} q_{i,\ell} \left[\pi_i^t + \rho_{i,\ell}\, \hat{\nu}_{i,0}^{t+1} + (1-\rho_{i,\ell})\, \hat{\nu}_{i,\ell+1}^{t+1} \right] \right\} + \sum_{i\in\mathcal{N}} q_{i,0}\,(\hat{\nu}_{i,0}^t - \hat{\nu}_{i,0}^{t+1}) \\
& \quad = \; \sum_{i\in\mathcal{N}} \left\{ q_{i,0}\, \hat{\nu}_{i,0}^{t+1} + \sum_{\ell=1}^{\infty} q_{i,\ell}\, \hat{\nu}_{i,\ell}^t \right\} + \sum_{i\in\mathcal{N}} q_{i,0}\,(\hat{\nu}_{i,0}^t - \hat{\nu}_{i,0}^{t+1}) \;\; = \;\; \sum_{i\in\mathcal{N}} \sum_{\ell=0}^{\infty} \hat{\nu}_{i,\ell}^t\, q_{i,\ell} \;\; = \;\; \hat{J}^t(\boldsymbol{q}),
\end{aligned}$$

where the second equality uses the fact that $\hat{\nu}_{i,\ell}^t = \pi_i^t \; + \; \rho_{i,\ell}\, \hat{\nu}_{i,0}^{t+1} + (1-\rho_{i,\ell})\, \hat{\nu}_{i,\ell+1}^{t+1}$ by (6). The two chains of equalities and inequalities above complete our induction argument, so that $V^t(\boldsymbol{q}) \geq \hat{J}^t(\boldsymbol{q})$ for all $\boldsymbol{q} \in \mathcal{Q}$ and $t \in \mathcal{T}$. By Lemma 3.3, we also have $J^1(\sum_{i\in\mathcal{N}} C_i\, \boldsymbol{e}_{i,0}) \leq 2\hat{J}^1(\sum_{i\in\mathcal{N}} C_i\, \boldsymbol{e}_{i,0})$, where $J^1(\sum_{i\in\mathcal{N}} C_i\, \boldsymbol{e}_{i,0})$ is the optimal total expected revenue. Thus, we obtain

$$V^1\left(\sum_{i\in\mathcal{N}} C_i\, \boldsymbol{e}_{i,0} \right) \; \geq \; \hat{J}^1\left(\sum_{i\in\mathcal{N}} C_i\, \boldsymbol{e}_{i,0} \right) \; \geq \; \frac{1}{2} J^1\left(\sum_{i\in\mathcal{N}} C_i\, \boldsymbol{e}_{i,0} \right).$$

■

Appendix E: Decomposability of the Value Functions of the Static Policy

In this section, we give a proof of Lemma 4.2, which shows that the value functions of the static policy are separable by products.

Proof of Lemma 4.2: We will prove the result by using induction over the time periods. The result holds at time period $T+1$ because $V^{T+1}=0=\sum_{i\in\mathcal{N}} V_i^{T+1}$ by definition. Assuming that the result holds at time period $t+1$, we proceed to show that the result holds at time period t. Letting $\boldsymbol{e}_\ell$ be the standard unit vector with one in the ℓ-th coordinate, by the inductive hypothesis, we have $\mathbb{E}\Big\{V^{t+1}(\boldsymbol{X}(\boldsymbol{q}))-V^{t+1}\left(\boldsymbol{X}(\boldsymbol{q})-\boldsymbol{e}_{i,0}+\boldsymbol{e}_{i,1}\right)\Big\}=\mathbb{E}\Big\{V_i^{t+1}(\boldsymbol{X}_i(\boldsymbol{q}_i))-V_i^{t+1}\left(\boldsymbol{X}_i(\boldsymbol{q}_i)-\boldsymbol{e}_0+\boldsymbol{e}_1\right)\Big\}$. In this case, by the recursion that we use to compute $\{V^t : t\in\mathcal{T}\}$ in Section 4.1, we obtain

$$
\begin{aligned}
V^t(\boldsymbol{q}) = &\sum_{i\in\mathcal{N}}\pi_i^t\sum_{\ell=1}^{\infty}q_{i,\ell} \;+\; \sum_{i\in\mathcal{N}}\mathbb{E}\Big\{V_i^{t+1}(\boldsymbol{X}_i(\boldsymbol{q}_i))\Big\}\\
&+\sum_{i\in\mathcal{N}}\mathbb{1}_{\{q_{i,0}\geq 1\}}\,\phi_i^t(\hat{A}^t)\bigg(r_i^t+\pi_i^t-(1-\rho_{i,0})\,\mathbb{E}\Big\{V_i^{t+1}(\boldsymbol{X}_i(\boldsymbol{q}_i))-V_i^{t+1}\left(\boldsymbol{X}_i(\boldsymbol{q}_i)-\boldsymbol{e}_0+\boldsymbol{e}_1\right)\Big\}\bigg).
\end{aligned}
$$

By (9), the expression on the right side above is equal to $\sum_{i\in\mathcal{N}}V_i^t(\boldsymbol{q}_i)$. Therefore, the result holds at time period t as well. ■

Appendix F: A Tailored Rollout Policy under Infinite Usage Durations

Under infinite usage durations, we give a tailored variant of our rollout policy that is guaranteed to obtain at least $\max\Big\{\frac{1}{2},1-\frac{R}{2\sqrt[3]{C_{\min}}}\Big\}$ fraction of the optimal total expected revenue, where $C_{\min}=\min_{i\in\mathcal{N}}C_i$ and $R=\max_{i\in\mathcal{N}}\Big\{\frac{\max_{t\in\mathcal{T}}r_i^t}{\min_{t\in\mathcal{T}}r_i^t}\Big\}$. The tailored variant of our rollout policy is based on performing rollout on a static policy that is obtained from a linear programming approximation. Since the usage durations are infinite, the products are purchased outright. Thus, we set $\pi_i^t=0$ for all $i\in\mathcal{N}$ and $t\in\mathcal{T}$ in this section, so there is no per-period rental fee.

F.1 Problem Formulation and Upper Bound on the Optimal Total Expected Revenue

We give a dynamic programming formulation for our dynamic assortment problem under infinite usage durations. Following this dynamic program, we formulate a linear programming approximation that we use to obtain an upper bound on the optimal total expected revenue. The linear programming approximation will be useful to derive a static policy. Our notation closely follows the one in Section 2. Indeed, as the case with infinite usage durations is a special case of the model in Section 2, we do not need to define any additional notation. We use $q_{i,0}$ to denote the remaining units of product i on-hand at the beginning of a generic time period. Since the products are not returned and there is no per-period rental fee, we do not need to keep track of the products that are currently in use. In this case, we can use the vector $\boldsymbol{q}=(q_{i,0}:i\in\mathcal{N})$ to capture the state of the remaining on-hand product inventories. The state space is $\mathcal{Q}=\{\boldsymbol{q}:q_{i,0}\in\{0,1,\ldots,C_i\}\;\forall\,i\in\mathcal{N}\}$. Let $J^t(\boldsymbol{q})$ denote the maximum total expected revenue over the time periods $t,\ldots,T$, given that the system is in state $\boldsymbol{q}$ at the beginning of time period t. Thus, using $\boldsymbol{e}_i\in\mathbb{R}_+^n$ to denote the standard unit vector with a one in the i-th coordinate, we can compute the optimal value functions $\{J^t : t \in \mathcal{T}\}$ by solving the dynamic program

$$\begin{aligned} J^t(\boldsymbol{q}) &= \max_{S \in \mathcal{F}} \left\{ \sum_{i \in \mathcal{N}} \mathbb{1}_{\{q_{i,0} \geq 1\}} \, \phi_i^t(S) \left(r_i^t + J^{t+1}(\boldsymbol{q} - \boldsymbol{e}_i) \right) + \left(1 - \sum_{i \in \mathcal{N}} \mathbb{1}_{\{q_{i,0} \geq 1\}} \, \phi_i^t(S) \right) J^{t+1}(\boldsymbol{q}) \right\} \\ &= \max_{S \in \mathcal{F}} \left\{ \sum_{i \in \mathcal{N}} \mathbb{1}_{\{q_{i,0} \geq 1\}} \, \phi_i^t(S) \left(r_i^t - \left\{ J^{t+1}(\boldsymbol{q}) - J^{t+1}(\boldsymbol{q} - \boldsymbol{e}_i) \right\} \right) \right\} + J^{t+1}(\boldsymbol{q}), \qquad (14) \end{aligned}$$

with the boundary condition that $J^{T+1} = 0$. In the dynamic program above, if we offer the assortment S, then a customer arriving at time period t chooses product i with probability $\phi_i^t(S)$. If there is on-hand inventory for product i, then we make a revenue of r_i^t and the state of the system at the next time period is $\boldsymbol{q} - \boldsymbol{e}_i$.

Since all units are available on-hand at the beginning of the selling horizon, the initial state is $\sum_{i \in \mathcal{N}} C_i \, \boldsymbol{e}_i$, so the optimal total expected revenue is given by $J^1(\sum_{i \in \mathcal{N}} C_i \, \boldsymbol{e}_i)$. Defining the decision variable $z^t(A)$ as the frequency with which we offer assortment A at time period t, we consider the linear programming approximation

$$\begin{aligned} \max \quad & \sum_{t \in \mathcal{T}} \sum_{i \in \mathcal{N}} \sum_{A \in \mathcal{F}} r_i^t \, \phi_i^t(A) \, z^t(A) \qquad (15) \\ \text{s.t.} \quad & \sum_{t \in \mathcal{T}} \sum_{A \in \mathcal{F}} \phi_i^t(A) \, z^t(A) \leq C_i \qquad \forall \, i \in \mathcal{N} \\ & \sum_{A \in \mathcal{F}} z^t(A) = 1 \qquad \forall \, t \in \mathcal{T} \\ & z^t(A) \geq 0 \qquad \forall \, A \in \mathcal{F}, \; t \in \mathcal{T}. \end{aligned}$$

Problem (15) is a linear programming approximation to our dynamic assortment problem that is formulated under the assumption that the choices of the customers take on their expected values. The objective function in problem (15) computes the total expected revenue over the selling horizon. Noting that the products are not returned, the first constraint ensures that the total expected number of units sold for product i does not exceed the initial inventory. The second constraint ensures that we offer an assortment at each time period. It is well-known that the optimal objective value of problem (15) is an upper bound on the optimal total expected revenue; see Proposition 1 in **?**. We will not only use the fact that the optimal objective value of problem (15) is an upper bound on the optimal total expected revenue, but also build on the solution to problem (15) to construct a static policy, as discussed next.

F.2 Static Policy

We construct a static policy by building on an optimal solution to problem (15). We show that we can efficiently compute the value functions of this static policy. Ultimately, the tailored variant of our rollout policy corresponds to the greedy policy with respect to the value functions of the static policy. Let $(\hat{z}^t(A) : A \in \mathcal{F},\ t \in \mathcal{T})$ be an optimal solution to problem (15). In our static policy, we offer assortment A at time period t with probability $\hat{z}^t(A)$. The assortment offer decisions at different time periods are independent of each other. If we offer some assortment A at time period t and the customer arriving at this time period chooses a product for which we do not have any on-hand units, then the customer leaves. Also, if we offer some assortment A at time period t and the customer arriving at this time period chooses a product for which we have on-hand units, then we can still reject the customer and leave her empty handed. The fact that we can reject a customer is not a concern because we only use the value functions of the static policy to come up with a rollout policy that actually never rejects a customer. Let $V^t(\boldsymbol{q})$ denote the total expected revenue under the static policy over the time periods $t,\ldots,T$, given that the system is in state $\boldsymbol{q}$ at the beginning of time period t. We can compute $\{V^t : t \in \mathcal{T}\}$ by using the recursion

$$\begin{aligned} V^t(\boldsymbol{q}) \;&=\; \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\ge 1\}} \Bigg(\sum_{A\in\mathcal{F}} \phi_i^t(A)\, \hat{z}^t(A) \Bigg) \max\Big\{ r_i^t + V^{t+1}(\boldsymbol{q}-\boldsymbol{e}_i),\, V^{t+1}(\boldsymbol{q}) \Big\} \\ &\qquad + \Bigg(1 - \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\ge 1\}} \Bigg(\sum_{A\in\mathcal{F}} \phi_i^t(A)\, \hat{z}^t(A) \Bigg) \Bigg) V^{t+1}(\boldsymbol{q}) \\ &=\; \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\ge 1\}} \Bigg(\sum_{A\in\mathcal{F}} \phi_i^t(A)\, \hat{z}^t(A) \Bigg) \Big[r_i^t - \Big\{ V^{t+1}(\boldsymbol{q}) - V^{t+1}(\boldsymbol{q}-\boldsymbol{e}_i) \Big\} \Big]^+ + V^{t+1}(\boldsymbol{q}), \end{aligned} \tag{16}$$

with the boundary condition that $V^{T+1} = 0$. In the recursion above, a customer chooses product i at time period t with probability $\sum_{A\in\mathcal{F}} \phi_i^t(A)\, \hat{z}^t(A)$. If we have on-hand units for product i, then we decide whether to serve or reject the customer. If we serve the customer, then we make a revenue of r_i^t and the state of the products at the next time period is $\boldsymbol{q} - \boldsymbol{e}_i$. In the next lemma, we show that $V^t(\boldsymbol{q})$ decomposes by products.

Lemma F.1 *For each $t \in \mathcal{T}$ and $\boldsymbol{q} \in \mathcal{Q}$, we have $V^t(\boldsymbol{q}) = \sum_{i\in\mathcal{N}} V_i^t(q_{i,0})$, where for each $i \in \mathcal{N}$, $\{V_i^t : t \in \mathcal{T}\}$ is computed by using the recursion*

$$\begin{aligned} V_i^t(q_{i,0}) \;&=\; \mathbb{1}_{\{q_{i,0}\ge 1\}} \Bigg(\sum_{A\in\mathcal{F}} \phi_i^t(A)\, \hat{z}^t(A) \Bigg) \max\Big\{ r_i^t + V_i^{t+1}(q_{i,0}-1),\; V_i^{t+1}(q_{i,0}) \Big\} \\ &\qquad + \Bigg(1 - \mathbb{1}_{\{q_{i,0}\ge 1\}} \Bigg(\sum_{A\in\mathcal{F}} \phi_i^t(A)\, \hat{z}^t(A) \Bigg) \Bigg) V_i^{t+1}(q_{i,0}) \\ &= \mathbb{1}_{\{q_{i,0}\ge 1\}} \Bigg(\sum_{A\in\mathcal{F}} \phi_i^t(A)\, \hat{z}^t(A) \Bigg) \Big[r_i^t - \Big\{ V_i^{t+1}(q_{i,0}) - V_i^{t+1}(q_{i,0}-1) \Big\} \Big]^+ \;+\; V_i^{t+1}(q_{i,0}), \end{aligned} \tag{17}$$

with the boundary condition that $V_i^{T+1} = 0$.

The proof of the lemma above is similar to that of Lemma 4.2 and it is omitted. Note that the state variable in the recursion in (17) is a scalar. Therefore, we can compute $\{V_i^t : t \in \mathcal{T}\}$ efficiently. In this case, by Lemma F.1, we can compute $\{V^t : t \in \mathcal{T}\}$ efficiently as well.

F.3 Rollout Policy and Performance Guarantee

The tailored variant of our rollout policy corresponds to the greedy policy with respect to the value functions $\{V^t : t \in \mathcal{T}\}$. In other words, the tailored variant of our rollout policy corresponds to using $V^t(\boldsymbol{q}) = \sum_{i\in\mathcal{N}} V_i^t(\boldsymbol{q}_{i,0})$ as a separable approximation to the optimal value function $J^t(\boldsymbol{q})$. Let $S^t_{\mathsf{rollout}}(\boldsymbol{q})$ be the assortment offered by the tailored variant of our rollout policy given that we are in state $\boldsymbol{q}$ at time period t. Replacing the optimal value function $J^{t+1}(\boldsymbol{q})$ on the right side of (14) with $V^{t+1}(\boldsymbol{q})$, $S^t_{\mathsf{rollout}}(\boldsymbol{q})$ is given by

$$\begin{aligned}
S^t_{\mathsf{rollout}}(\boldsymbol{q}) &= \arg\max_{S\in\mathcal{F}} \left\{ \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\geq 1\}}\, \phi_i^t(S) \Big(r_i^t + V^{t+1}(\boldsymbol{q}-\boldsymbol{e}_i)\Big) + \Big(1 - \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\geq 1\}}\, \phi_i^t(S)\Big) V^{t+1}(\boldsymbol{q}) \right\} \\
&= \arg\max_{S\in\mathcal{F}} \left\{ \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\geq 1\}}\, \phi_i^t(S) \Big(r_i^t - \Big\{ V^{t+1}(\boldsymbol{q}) - V^{t+1}(\boldsymbol{q}-\boldsymbol{e}_i)\Big\}\Big)\right\} \\
&= \arg\max_{S\in\mathcal{F}} \left\{ \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\geq 1\}}\, \phi_i^t(S) \Big(r_i^t - \Big\{ V_i^{t+1}(q_{i,0}) - V_i^{t+1}(q_{i,0}-1)\Big\}\Big)\right\}, \qquad (18)
\end{aligned}$$

where the second equality follows from Lemma F.1. If the state of the system at time period t is $\boldsymbol{q}$, then the tailored variant of our rollout policy offers the assortment $S^t_{\mathsf{rollout}}(\boldsymbol{q})$. Note that if a customer chooses a product in the assortment $S^t_{\mathsf{rollout}}(\boldsymbol{q})$, then the tailored variant does not reject the customer. Let $U^t(\boldsymbol{q})$ be the total expected revenue obtained by the tailored variant of our rollout policy over the time periods $t,\ldots,T$, given that we are in state $\boldsymbol{q}$ at time period t. We can compute $\{U^t : t\in\mathcal{T}\}$ by solving the recursion

$$\begin{aligned}
U^t(\boldsymbol{q}) &= \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\geq 1\}}\, \phi_i^t(S^t_{\mathsf{rollout}}(\boldsymbol{q})) \Big(r_i^t + U^{t+1}(\boldsymbol{q}-\boldsymbol{e}_i)\Big) + \Big(1 - \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\geq 1\}}\, \phi_i^t(S^t_{\mathsf{rollout}}(\boldsymbol{q}))\Big) U^{t+1}(\boldsymbol{q}) \\
&= \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\geq 1\}}\, \phi_i^t(S^t_{\mathsf{rollout}}(\boldsymbol{q})) \Big(r_i^t - \Big\{ U^{t+1}(\boldsymbol{q}) - U^{t+1}(\boldsymbol{q}-\boldsymbol{e}_i)\Big\}\Big) + U^{t+1}(\boldsymbol{q}), \qquad (19)
\end{aligned}$$

with the boundary condition that $U^{T+1} = 0$. In the recursion above, if a customer chooses a product for which we have on-hand inventory, then we must serve this customer. Therefore, the tailored variant of our rollout policy never rejects a customer.

Since the initial state is $\sum_{i\in\mathcal{N}} C_i\, \boldsymbol{e}_i$, the total expected revenue obtained by the tailored variant of our rollout policy is $U^1(\sum_{i\in\mathcal{N}} C_i\, \boldsymbol{e}_i)$. The next theorem is our main result, which gives a performance guarantee for the tailored variant.

Theorem F.2 *Letting* $C_{\min} = \min_{i\in\mathcal{N}} C_i$ *and* $R = \max_{i\in\mathcal{N}} \left\{ \frac{\max_{t\in\mathcal{T}} r_i^t}{\min_{t\in\mathcal{T}} r_i^t} \right\}$, *the total expected revenue of our rollout policy tailored to infinite usage durations is at least* $\max\left\{\frac{1}{2}, 1 - \frac{R}{2\sqrt[3]{C_{\min}}}\right\}$ *fraction of the optimal total expected revenue.*

The proof of the theorem above uses a sequence of lemmas. We devote the next section to the proof of this theorem.

F.4 Proof of Theorem F.2

Recalling that $\{V^t : t \in \mathcal{T}\}$ are the value functions of the static policy computed through (16), the next lemma compares the total expected revenue of the static policy with the optimal objective value of problem (15). Throughout this section, we use $(\hat{z}^t(A) : A \in \mathcal{F},\ t \in \mathcal{T})$ to denote an optimal solution to problem (15) with the corresponding optimal objective value Z^*.

Lemma F.3 *We have $V^1(\sum_{i\in\mathcal{N}} C_i\, \boldsymbol{e}_i) \geq Z^*/2$.*

Proof: First, we give a lower bound on $V^t(\boldsymbol{q})$. Consider the scalars $\{v_i^t : t \in \mathcal{T}\}$ that are computed by using the recursion

$$v_i^t = \frac{1}{C_i}\left(\sum_{A\in\mathcal{F}} \phi_i^t(A)\, \hat{z}^t(A)\right)[r_i^t - v_i^{t+1}]^+ + v_i^{t+1}, \tag{20}$$

with the boundary condition that $v_i^{T+1} = 0$. We claim that $V_i^t(q_{i,0}) \geq v_i^t\, q_{i,0}$. We will prove the claim by using induction over the time periods. The result holds at time period $T+1$ because $V_i^{T+1} = 0 = v_i^{T+1}$. Assuming that the result holds at time period $t+1$, we proceed to show that the result holds at time period t as well. We have

$$\begin{aligned}
V_i^t(q_{i,0}) &= \mathbb{1}_{\{q_{i,0}\geq 1\}}\left(\sum_{A\in\mathcal{F}} \phi_i^t(A)\, \hat{z}^t(A)\right)\left[r_i^t - \left\{V_i^{t+1}(q_{i,0}) - V_i^{t+1}(q_{i,0}-1)\right\}\right]^+ + V_i^{t+1}(q_{i,0}) \\
&\geq \mathbb{1}_{\{q_{i,0}\geq 1\}}\left(\sum_{A\in\mathcal{F}} \phi_i^t(A)\, \hat{z}^t(A)\right)[r_i^t - v_i^{t+1}]^+ + v_i^{t+1}\, q_{i,0} \\
&\geq \frac{q_{i,0}}{C_i}\left(\sum_{A\in\mathcal{F}} \phi_i^t(A)\, \hat{z}^t(A)\right)[r_i^t - v_i^{t+1}]^+ + v_i^{t+1}\, q_{i,0} \\
&= v_i^t\, q_{i,0}.
\end{aligned}$$

Since $\sum_{A\in\mathcal{F}} \hat{z}^t(A) = 1$ by the second constraint in problem (15), we have $\sum_{A\in\mathcal{F}} \phi_i^t(A)\, \hat{z}^t(A) \leq 1$, in which case, the first inequality above follows from the inductive hypothesis. The second inequality holds because $\mathbb{1}_{\{q_{i,0}\geq 1\}} \geq q_{i,0}/C_i$. The last equality is by (20). The chain of inequalities above establishes the claim. So, we have $V^t(\boldsymbol{q}) = \sum_{i\in\mathcal{N}} V_i^t(q_{i,0}) \geq \sum_{i\in\mathcal{N}} v_i^t\, q_{i,0}$.

Second, we give a lower bound on $\sum_{i\in\mathcal{N}} v_i^1\, C_i$. By (20), we have $v_i^t\, C_i = \sum_{A\in\mathcal{F}} \phi_i^t(A)\, \hat{z}^t(A) \times [r_i^t - v_i^{t+1}]^+ + v_i^{t+1}\, C_i$. Adding this equality over all $t \in \mathcal{T}$, since $v_i^{T+1} = 0$, we get

$$\begin{aligned}
v_i^1\, C_i &= \sum_{t\in\mathcal{T}}\sum_{A\in\mathcal{F}} \phi_i^t(A)\, \hat{z}^t(A)\, [r_i^t - v_i^{t+1}]^+ \\
&\geq \sum_{t\in\mathcal{T}}\sum_{A\in\mathcal{F}} \phi_i^t(A)\, \hat{z}^t(A)\, [r_i^t - v_i^{t+1}] \;\geq\; \sum_{t\in\mathcal{T}}\sum_{A\in\mathcal{F}} \phi_i^t(A)\, \hat{z}^t(A)\, r_i^t - \sum_{t\in\mathcal{T}}\sum_{A\in\mathcal{F}} \phi_i^t(A)\, \hat{z}^t(A)\, v_i^1.
\end{aligned}$$

Noting that $[a]^+ \geq 0$ for all $a \in \mathbb{R}$, the second inequality above holds because we have $v_i^1 \geq v_i^2 \geq \ldots \geq v_i^{T+1}$ by (20). Adding the chain of inequalities above over all $i \in \mathcal{N}$, it follows that

$\sum_{i\in\mathcal{N}} v_i^1 C_i \geq \sum_{t\in\mathcal{T}} \sum_{i\in\mathcal{N}} \sum_{A\in\mathcal{F}} r_i^t \phi_i^t(A) \hat{z}^t(A) - \sum_{i\in\mathcal{N}} v_i^1 \sum_{t\in\mathcal{T}} \sum_{A\in\mathcal{F}} \phi_i^t(A) \hat{z}^t(A)$. By the second constraint in problem (15), we have $\sum_{t\in\mathcal{T}} \sum_{A\in\mathcal{F}} \phi_i^t(A) \hat{z}^t(A) \leq C_i$. In this case, the last inequality implies that $\sum_{i\in\mathcal{N}} v_i^1 C_i \geq \sum_{t\in\mathcal{T}} \sum_{i\in\mathcal{N}} \sum_{A\in\mathcal{F}} r_i^t \phi_i^t(A) \hat{z}^t(A) - \sum_{i\in\mathcal{N}} v_i^1 C_i$. Therefore, noting that the optimal objective value of problem (15) is $Z^* = \sum_{t\in\mathcal{T}} \sum_{i\in\mathcal{N}} \sum_{A\in\mathcal{F}} r_i^t \phi_i^t(A) \hat{z}^t(A)$, it follows that $2 \sum_{i\in\mathcal{N}} v_i^1 C_i \geq Z^*$, showing that $Z^*/2$ is a lower bound on $\sum_{i\in\mathcal{N}} v_i^1 C_i$. By the first part of the proof, we have $V^1(\sum_{i\in\mathcal{N}} C_i \boldsymbol{e}_i) = \sum_{i\in\mathcal{N}} V_i^1(C_i) \geq \sum_{i\in\mathcal{N}} v_i^1 C_i$, whereas by the second part of the proof, we have $\sum_{i\in\mathcal{N}} v_i^1 C_i \geq Z^*/2$. So, the desired result follows. ∎

Under the static policy, we offer assortment A at time period t with probability $\hat{z}^t(A)$. On the other hand, if we offer assortment A at time period t, then a customer chooses product i with probability $\phi_i^t(A)$. Therefore, under the static policy, a customer chooses product i with probability $\sum_{A\in\mathcal{F}} \phi_i^t(A) \hat{z}^t(A)$. For notational brevity, we let

$$\hat{\alpha}_i^t = \sum_{A\in\mathcal{F}} \phi_i^t(A) \hat{z}^t(A),$$

in which case, using Y_i^t to denote a Bernoulli random variable with parameter $\hat{\alpha}_i^t$, the random variable Y_i^t corresponds to the demand for product i at time period t under the static policy. Also, we define $W_i^t = \sum_{s=t}^T Y_i^s$ with $W_i^{T+1} = 0$, which corresponds to the total demand for product i under the static policy over the time periods $t, \ldots, T$.

Next, we establish two preliminary bounds. In the next lemma, we give a lower bound on $\{V_i^t : t\in\mathcal{T}\}$ by using the random variables $\{W_i^t : t\in\mathcal{T}\}$.

Lemma F.4 *Letting $R_{i,\max} = \max_{t\in\mathcal{T}} r_i^t$, for each $q_{i,0} = 0, 1, \ldots, C_i$ and $t\in\mathcal{T}$, we have*

$$V_i^t(q_{i,0}) \geq \sum_{s=t}^T r_i^s \hat{\alpha}_i^s - R_{i,\max} \mathbb{E}\{[W_i^t - q_{i,0}]^+\}. \tag{21}$$

Proof: Because $\{Y_i^t : t\in\mathcal{T}\}$ are independent of each other, Y_i^t and W_i^{t+1} are also independent of each other. Noting that Y_i^t is a Bernoulli random variable with parameter $\hat{\alpha}_i^t$, we get

$$\mathbb{E}\{[W_i^t - q_{i,0}]^+\} = \mathbb{E}\{[W_i^{t+1} + Y_i^t - q_{i,0}]^+\} = \hat{\alpha}_i^t \mathbb{E}\{[W_i^{t+1} + 1 - q_{i,0}]^+\} + (1 - \hat{\alpha}_i^t) \mathbb{E}\{[W_i^{t+1} - q_{i,0}]^+\}.$$

We will use induction over the time periods to show that the inequality in (21) holds for all $t\in\mathcal{T}$. The inequality holds at time period $T+1$ because $V_i^{T+1} = 0$ and the right side of the inequality in (21) with $t = T+1$ is also zero. Assuming that the inequality in (21) holds at time period $t+1$, we proceed to show that this inequality holds at time period t as well. First, consider the case $q_{i,0} \geq 1$.

Because $q_{i,0} \geq 1$, we have $\mathbb{1}_{\{q_{i,0} \geq 1\}} = 1$. Furthermore, noting the definition of $\hat{\alpha}_i^t$, we can write the recursion in (17) as

$$\begin{aligned}
V_i^t(q_{i,0}) &= \hat{\alpha}_i^t \left[r_i^t - \left\{ V_i^{t+1}(q_{i,0}) - V_i^{t+1}(q_{i,0}-1) \right\} \right]^+ + V_i^{t+1}(q_{i,0}) \\
&\geq \hat{\alpha}_i^t \left(r_i^t - \left\{ V_i^{t+1}(q_{i,0}) - V_i^{t+1}(q_{i,0}-1) \right\} \right) + V_i^{t+1}(q_{i,0}) \\
&\geq \hat{\alpha}_i^t \left(r_i^t + R_{i,\max} \left\{ \mathbb{E}\{[W_i^{t+1} - q_{i,0}]^+\} - \mathbb{E}\{[W_i^{t+1} + 1 - q_{i,0}]^+\} \right\} \right) \\
&\qquad\qquad + \sum_{s=t+1}^{T} r_i^s \, \hat{\alpha}_i^s - R_{i,\max} \, \mathbb{E}\{[W_i^{t+1} - q_{i,0}]^+\} \\
&= \sum_{s=t}^{T} r_i^s \, \hat{\alpha}_i^s - R_{i,\max} \, \hat{\alpha}_i^t \, \mathbb{E}\{[W_i^{t+1} + 1 - q_{i,0}]^+\} - R_{i,\max} \, (1 - \hat{\alpha}_i^t) \, \mathbb{E}\{[W_i^{t+1} - q_{i,0}]^+\} \\
&= \sum_{s=t}^{T} r_i^s \, \hat{\alpha}_i^s - R_{i,\max} \, \mathbb{E}\{[W_i^t - q_{i,0}]^+\},
\end{aligned}$$

where the second inequality is by the inductive hypothesis. Thus, (21) holds at time period t as well. Second, consider the case $q_{i,0} = 0$, so $\mathbb{1}_{\{q_{i,0} \geq 1\}} = 0$. Therefore, by (17), we get

$$\begin{aligned}
V_i^t(q_{i,0}) &= V_i^{t+1}(q_{i,0}) \geq \sum_{s=t+1}^{T} r_i^s \, \hat{\alpha}_i^s - R_{i,\max} \, \mathbb{E}\{W_i^{t+1}\} \\
&\geq \sum_{s=t}^{T} r_i^s \, \hat{\alpha}_i^s - R_{i,\max} \, \mathbb{E}\{Y_i^t + W_i^{t+1}\} = \sum_{s=t}^{T} r_i^s \, \hat{\alpha}_i^s - R_{i,\max} \, \mathbb{E}\{[W_i^t - q_{i,0}]^+\},
\end{aligned}$$

where the first inequality is by the inductive hypothesis and the second inequality uses the fact that $\mathbb{E}\{Y_i^t\} = \hat{\alpha}_i^t$ and $R_{i,\max} \geq r_i^t$. Thus, (21) holds at time period t as well. ■

By the lemma above, we have $V_i^1(C_i) \geq \sum_{t \in \mathcal{T}} r_i^t \, \hat{\alpha}_i^t - R_{i,\max} \, \mathbb{E}\{[W_i^1 - C_i]^+\}$. Next, we upper bound the expectation on the right side of this inequality.

Lemma F.5 *We have* $\mathbb{E}\{[W_i^1 - C_i]^+\} \leq \frac{1}{2\sqrt[3]{C_i}} \sum_{t \in \mathcal{T}} \hat{\alpha}_i^t$.

Proof: We have $\mathbb{E}\{Y_i^t\} = \hat{\alpha}_i^t$ and $Var(Y_i^t) = \hat{\alpha}_i^t \, (1 - \hat{\alpha}_i^t) \leq \hat{\alpha}_i^t$. Since $W_i^1 = \sum_{t \in \mathcal{T}} Y_i^t$ and $\{Y_i^t : t \in \mathcal{T}\}$ are independent of each other, we have $\mathbb{E}\{W_i^1\} = \sum_{t \in \mathcal{T}} \hat{\alpha}_i^t$ and $Var(W_i^1) \leq \sum_{t \in \mathcal{T}} \hat{\alpha}_i^t$. First, consider the case $\sum_{t \in \mathcal{T}} \hat{\alpha}_i^t \leq (C_i)^{2/3}$. It is simple to check that $[x - a]^+ \leq \frac{1}{4a} x^2$ for all $x, a \in \mathbb{R}_+$, which implies that $\mathbb{E}\{[W_i^1 - C_i]^+\} \leq \frac{1}{4C_i} \mathbb{E}\{(W_i^1)^2\} = \frac{1}{4C_i} Var(W_i^1) + \frac{1}{4C_i} (E\{W_i^1\})^2$. Thus, we get

$$\mathbb{E}\{[W_i^1 - C_i]^+\} \leq \frac{1}{4C_i} \sum_{t \in \mathcal{T}} \hat{\alpha}_i^t + \frac{1}{4C_i} \left(\sum_{t \in \mathcal{T}} \hat{\alpha}_i^t \right)^2 \leq \frac{1}{4C_i} \sum_{t \in \mathcal{T}} \hat{\alpha}_i^t + \frac{1}{4\sqrt[3]{C_i}} \sum_{t \in \mathcal{T}} \hat{\alpha}_i^t \leq \frac{1}{2\sqrt[3]{C_i}} \sum_{t \in \mathcal{T}} \hat{\alpha}_i^t,$$

where the second inequality uses the fact that $\sum_{t \in \mathcal{T}} \hat{\alpha}_i^t \leq (C_i)^{2/3}$ and the third inequality holds because $C_i \geq \sqrt[3]{C_i}$. Second, consider the case $\sum_{t \in \mathcal{T}} \hat{\alpha}_i^t \geq (C_i)^{2/3}$. If Z is a random variable with

$\mathbb{E}\{Z\} \le a$, then $\mathbb{E}\{[Z-a]^+\} \le \frac{1}{2}\sqrt{Var(Z)}$; see **?**. By the first constraint in problem (15), we have $\mathbb{E}\{W_i^1\} = \sum_{t\in\mathcal{T}} \hat{\alpha}_i^t = \sum_{t\in\mathcal{T}} \sum_{A\in\mathcal{F}} \phi_i^t(A)\,\hat{z}^t(A) \le C_i$. Thus, we get

$$\mathbb{E}\{[W_i^1 - C_i]^+\} \;\le\; \frac{1}{2}\sqrt{Var(W_i^1)} \;\le\; \frac{1}{2}\sqrt{\sum_{t\in\mathcal{T}} \hat{\alpha}_i^t} \;\le\; \frac{1}{2\sqrt[3]{C_{\min}}} \sum_{t\in\mathcal{T}} \hat{\alpha}_i^t,$$

where the second inequality uses the fact that $Var(W_i^1) \le \sum_{t\in\mathcal{T}} \hat{\alpha}_i^t$ and the third inequality uses the fact that $\sum_{t\in\mathcal{T}} \hat{\alpha}_i^t \ge (C_i)^{2/3}$. ■

In the next lemma, we compare the total expected revenue of the static policy with the optimal objective value of problem (15). Lemma F.3 will ultimately allow us to show that the tailored variant of our rollout policy always obtains at least half of the optimal total expected revenue, whereas the next lemma will ultimately allow us to show that the tailored variant of our rollout policy always obtains at least $1 - \frac{R}{2\sqrt[3]{C_{\min}}}$ fraction of the optimal total expected revenue.

Lemma F.6 *We have* $V^1(\sum_{i\in\mathcal{N}} C_i\, \boldsymbol{e}_i) \ge \left(1 - \frac{R}{2\sqrt[3]{C_{\min}}}\right) Z^*$.

Proof: By Lemmas F.4 and F.5, we have $V_i^1(C_i) \ge \sum_{t\in\mathcal{T}} r_i^t\, \hat{\alpha}_i^t - R_{i,\max}\, \frac{1}{2\sqrt[3]{C_i}} \sum_{t\in\mathcal{T}} \hat{\alpha}_i^t$. For notational brevity, we let $R_{i,\min} = \min_{t\in\mathcal{T}} r_i^t$. Adding the last inequality over all $i\in\mathcal{N}$, we obtain

$$\begin{aligned}
V^1\left(\sum_{i\in\mathcal{N}} C_i\, \boldsymbol{e}_i\right) \;&=\; \sum_{i\in\mathcal{N}} V_i^1(C_i) \;\ge\; \sum_{i\in\mathcal{N}}\sum_{t\in\mathcal{T}} r_i^t\, \hat{\alpha}_i^t - \sum_{i\in\mathcal{N}} R_{i,\max}\, \frac{1}{2\sqrt[3]{C_i}} \sum_{t\in\mathcal{T}} \hat{\alpha}_i^t \\
&\ge\; \sum_{i\in\mathcal{N}}\sum_{t\in\mathcal{T}} r_i^t\, \hat{\alpha}_i^t - \sum_{i\in\mathcal{N}} \frac{R_{i,\max}}{R_{i,\min}}\, \frac{1}{2\sqrt[3]{C_i}} \sum_{t\in\mathcal{T}} r_i^t\, \hat{\alpha}_i^t \\
&\ge \sum_{i\in\mathcal{N}}\sum_{t\in\mathcal{T}} r_i^t\, \hat{\alpha}_i^t - \frac{R}{2\sqrt[3]{C_{\min}}} \sum_{i\in\mathcal{N}}\sum_{t\in\mathcal{T}} r_i^t\, \hat{\alpha}_i^t \\
&=\; \left(1 - \frac{R}{2\sqrt[3]{C_{\min}}}\right) \sum_{i\in\mathcal{N}}\sum_{t\in\mathcal{T}}\sum_{A\in\mathcal{F}} r_i^t\, \phi_i^t(A)\, \hat{z}^t(A) \;=\; \left(1 - \frac{R}{2\sqrt[3]{C_{\min}}}\right) Z^*,
\end{aligned}$$

where the second inequality uses the fact that $r_i^t / R_{i,\min} \ge 1$ and the third inequality holds because $R = \max_{i\in\mathcal{N}} \left\{ \frac{\max_{t\in\mathcal{T}} r_i^t}{\min_{t\in\mathcal{T}} r_i^t} \right\} = \max_{i\in\mathcal{N}} \frac{R_{i,\max}}{R_{i,\min}}$ and $C_{\min} = \min_{i\in\mathcal{N}} C_i$. ■

Here is the proof of Theorem F.2.

Proof of Theorem F.2: Shortly, we will prove the claim that $U^t(\boldsymbol{q}) \ge V^t(\boldsymbol{q})$ for all $q \in \mathcal{Q}$ and $t \in \mathcal{T}$, where $\{U^t : t\in\mathcal{T}\}$ is computed through the recursion in (19), so that $U^t(\boldsymbol{q})$ is the total expected revenue of the tailored variant of our rollout policy over the time periods $t,\ldots,T$ given that we are in state $\boldsymbol{q}$ at time period t. In this case, we get $U^1(\sum_{i\in\mathcal{N}} C_i\, \boldsymbol{e}_i) \ge V^1(\sum_{i\in\mathcal{N}} C_i\, \boldsymbol{e}_i)$. Furthermore, by Lemmas F.3 and F.6, we have $V^1(\sum_{i\in\mathcal{N}} C_i\, \boldsymbol{e}_i) \ge \max\left\{\frac{1}{2}, 1 - \frac{R}{2\sqrt[3]{C_{\min}}}\right\} Z^*$. Therefore, we get $U^1(\sum_{i\in\mathcal{N}} C_i\, \boldsymbol{e}_i) \ge \max\left\{\frac{1}{2}, 1 - \frac{R}{2\sqrt[3]{C_{\min}}}\right\} Z^*$, in which case, the desired result follows by the fact that $U^1(\sum_{i\in\mathcal{N}} C_i\, \boldsymbol{e}_i)$ corresponds to the total expected revenue of the tailored variant of our rollout policy and the optimal objective value of problem (15) is an upper bound on the optimal total expected revenue. In the rest of the proof, we use induction over the time periods to prove the claim that we have at the beginning of this paragraph. Because $U^{T+1}=0=V^{T+1}$, the claim holds at time period $T+1$. Assuming that the claim holds at time period $t+1$, we proceed to show that the claim holds at time period t as well. Noting (19), we obtain

$$\begin{aligned}
U^t(\boldsymbol{q}) &= \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\ge 1\}}\, \phi_i^t(S^t_{\mathsf{rollout}}(\boldsymbol{q}))\left(r_i^t-\left\{U^{t+1}(\boldsymbol{q})-U^{t+1}(\boldsymbol{q}-\boldsymbol{e}_i)\right\}\right) + U^{t+1}(\boldsymbol{q})\\
&\ge \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\ge 1\}}\, \phi_i^t(S^t_{\mathsf{rollout}}(\boldsymbol{q}))\left(r_i^t-\left\{V^{t+1}(\boldsymbol{q})-V^{t+1}(\boldsymbol{q}-\boldsymbol{e}_i)\right\}\right) + V^{t+1}(\boldsymbol{q})\\
&= \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\ge 1\}}\, \phi_i^t(S^t_{\mathsf{rollout}}(\boldsymbol{q}))\left(r_i^t-\left\{V_i^{t+1}(q_{i,0})-V_i^{t+1}(q_{i,0}-1)\right\}\right) + V^{t+1}(\boldsymbol{q})\\
&= \max_{S\in\mathcal{F}}\left\{\sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\ge 1\}}\, \phi_i^t(S)\left(r_i^t-\left\{V_i^{t+1}(q_{i,0})-V_i^{t+1}(q_{i,0}-1)\right\}\right)\right\} + V^{t+1}(\boldsymbol{q})\\
&= \max_{S\in\mathcal{F}}\left\{\sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\ge 1\}}\, \phi_i^t(S)\left[r_i^t-\left\{V_i^{t+1}(q_{i,0})-V_i^{t+1}(q_{i,0}-1)\right\}\right]^+\right\} + V^{t+1}(\boldsymbol{q})\\
&\ge \sum_{A\in\mathcal{F}} \hat{z}^t(A)\left\{\sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\ge 1\}}\, \phi_i^t(A)\left[r_i^t-\left\{V_i^{t+1}(q_{i,0})-V_i^{t+1}(q_{i,0}-1)\right\}\right]^+\right\} + V^{t+1}(\boldsymbol{q})\\
&= \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\ge 1\}}\left(\sum_{A\in\mathcal{F}}\phi_i^t(A)\,\hat{z}^t(A)\right)\left[r_i^t-\left\{V^{t+1}(\boldsymbol{q})-V^{t+1}(\boldsymbol{q}-\boldsymbol{e}_i)\right\}\right]^+\Bigg\} + V^{t+1}(\boldsymbol{q}) = V^t(\boldsymbol{q}).
\end{aligned}$$

In the chain of inequalities above, the first inequality uses the inductive hypothesis. The second equality holds because we have $V^t(\boldsymbol{q})=\sum_{i\in\mathcal{N}} V_i^t(q_{i,0})$ by Lemma F.1. The third equality uses the fact that $S^t_{\mathsf{rollout}}(\boldsymbol{q})$ is, by definition, an optimal solution to the maximization problem on the right side of (18). By using an argument that is similar to the one in the proof of Lemma B.1, we can show that $\max_{S\in\mathcal{F}}\sum_{i\in\mathcal{N}}\phi_i^t(S)\,\Delta_i^t=\max_{S\in\mathcal{F}}\sum_{i\in\mathcal{N}}\phi_i^t(S)\,[\Delta_i^t]^+$ for any $\{\Delta_i^t : i\in\mathcal{N}\}$. In this case, the fourth equality follows by identifying Δ_i^t with $\mathbb{1}_{\{q_{i,0}\ge 1\}}\left[r_i^t-\left\{V_i^{t+1}(q_{i,0})-V_i^{t+1}(q_{i,0}-1)\right\}\right]$. Noting that we have $\sum_{A\in\mathcal{F}}\hat{z}^t(A)=1$ by the second constraint in problem (15), the last inequality holds since the optimal objective value of a maximization problem is at least as large as any convex combination of the objective values of the problem evaluated at all feasible solutions. The fifth equality follows by, once again, noting that $V^t(\boldsymbol{q})=\sum_{i\in\mathcal{N}}V_i^t(q_{i,0})$ and arranging the terms. The last equality is by the dynamic program in (16). The chain of inequalities above shows that the claim holds at time period t as well, completing the induction argument. ■

F.5 Asymptotic Scaling Regime

By Theorem F.2, as the initial inventories of the products get large, the tailored variant of our rollout policy becomes near-optimal. Note that this result holds irrespective of the other parameters of the problem. In this section, we consider a standard regime where the inventories of the products and the number of time periods in the selling horizon scale up linearly at the same rate κ (**?**). We show that the tailored variant of our rollout policy obtains at least $1-\frac{B}{\sqrt{\kappa}}$ fraction of the optimal total expected revenue, where B is a constant that is independent of the scaling rate κ. In particular, consider a sequence of instances of our dynamic assortment problem $\{\mathcal{P}^\kappa : \kappa \in \mathbb{Z}_{++}\}$ indexed by $\kappa \in \mathbb{Z}_+$. Instance $\mathcal{P}^1$ corresponds to the instance that we consider throughout the paper. In instance $\mathcal{P}^\kappa$, there are κT time periods in the selling horizon indexed by $\mathcal{T}^\kappa = \{1,\ldots,\kappa T\}$. The initial inventory of product i is κC_i. Using $\lceil\cdot\rceil$ to denote the round-up function, the one-time upfront fee for product i at time period t is $r_i^{t,\kappa} = r_i^{\lceil t/\kappa\rceil}$. If we offer the assortment S at time period t, then an arriving customer chooses product i with probability $\phi_i^{t,\kappa}(S) = \phi_i^{\lceil t/\kappa\rceil}(S)$. Therefore, for any $t=1,\ldots,T$, the one-period fees and the choice probabilities in instance $\mathcal{P}^1$ at time period t are the same as the one-period fees and the choice probabilities in instance $\mathcal{P}^\kappa$ at time periods $\{\kappa(t-1)+1,\ldots,\kappa t\}$. Intuitively speaking, time period t in instance $\mathcal{P}^1$ is repeated κ times in instance $\mathcal{P}^\kappa$. The next theorem is the main result of this section and gives a performance guarantee for the tailored rollout policy for instance $\mathcal{P}^\kappa$.

Theorem F.7 *For instance $\mathcal{P}^\kappa$, the total expected revenue of our rollout policy tailored to infinite usage durations is at least $1-\frac{B}{\sqrt{\kappa}}$ fraction of the optimal total expected revenue, where B is a constant that is independent of κ.*

Proof: The proof follows from an outline similar to the one in the previous section. We let $\hat{Z}^\kappa$ be the optimal objective value of problem (15) for instance $\mathcal{P}^\kappa$. Noting that time period t in instance $\mathcal{P}^1$ is repeated κ times in instance $\mathcal{P}^\kappa$, it is simple to check that $Z^\kappa = \kappa Z^1$. Using $(\hat{z}^{t,\kappa}(A) : A \in \mathcal{F},\ t \in \mathcal{T}^\kappa)$ to denote an optimal solution to problem (15) for instance $\mathcal{P}^\kappa$, for notational brevity, we let $\hat{\alpha}_i^{t,k} = \sum_{A\in\mathcal{F}} \phi_i^{t,\kappa}(A)\,\hat{z}^{t,\kappa}(A)$. In this case, if we let $Y_i^{t,\kappa}$ be a Bernoulli random variable with parameter $\hat{\alpha}_i^{t,\kappa}$, $Y_i^{t,\kappa}$ captures the demand for product i at time period t under the static policy for instance $\mathcal{P}^\kappa$. We use $\{V^{t,\kappa} : t \in \mathcal{T}^\kappa\}$ to denote the value functions of the static policy for instance $\mathcal{P}^\kappa$, which are computed through (16). By Lemma F.1, $V^{1,\kappa}(\sum_{i\in\mathcal{N}} \kappa C_i\, \boldsymbol{e}_i) = \sum_{i\in\mathcal{N}} V_i^{1,\kappa}(\kappa C_i)$, where $\{V_i^t : t \in \mathcal{T}\}$ are computed through (17) for instance $\mathcal{P}^\kappa$. Note that $R_{i,\max} = \frac{\max_{t\in\mathcal{T}} r_i^t}{\min_{t\in\mathcal{T}} r_i^t} = \frac{\max_{t\in\mathcal{T}^\kappa} r_i^{t,\kappa}}{\min_{t\in\mathcal{T}^\kappa} r_i^{t,\kappa}}$. In this case, letting $W_i^{1,\kappa} = \sum_{t\in\mathcal{T}^\kappa} Y_i^{t,\kappa}$ and using Lemma F.4 for instance $\mathcal{P}^\kappa$, we have

$$V_i^{1,\kappa}(\kappa C_i) \geq \sum_{t\in\mathcal{T}^\kappa} r_i^{t,\kappa}\, \hat{\alpha}_i^{t,\kappa} - R_{i,\max}\, \mathbb{E}\{[W_i^{1,\kappa} - \kappa C_i]^+\}. \tag{22}$$

Since the variance of a Bernoulli random variable is at most $1/4$, we have $Var(W_i^{1,\kappa}) = \sum_{t\in\mathcal{T}^\kappa} Var(Y_i^{t,\kappa}) \leq \frac{1}{4}\kappa T$. As in the proof of Lemma F.5, for a random variable Z with $\mathbb{E}\{Z\} \leq a$, we have $\mathbb{E}\{[Z-a]^+\} \leq \frac{1}{2}\sqrt{Var(Z)}$. By the first constraint in problem (15) for instance $\mathcal{P}^\kappa$, we have $\mathbb{E}\{W_i^{1,\kappa}\} = \sum_{t\in\mathcal{T}^\kappa} \mathbb{E}\{Y_i^{t,\kappa}\} = \sum_{t\in\mathcal{T}^\kappa} \hat{\alpha}_i^{t,\kappa} = \sum_{t\in\mathcal{T}^\kappa} \sum_{A\in\mathcal{F}} \phi_i^{t,\kappa}\, \hat{z}^{t,\kappa}(A) \le \kappa\, C_i$, in which case, we obtain $\mathbb{E}\{[W_i^{1,\kappa} - \kappa\, C_i]^+\} \le \frac{1}{2}\sqrt{Var(W_i^{1,\kappa})} \le \frac{\sqrt{\kappa T}}{4}$. So, by (22), we get

$$\begin{aligned} V^{1,\kappa}\left(\sum_{i\in\mathcal{N}} \kappa\, C_i\, \boldsymbol{e}_i\right) \;&=\; \sum_{i\in\mathcal{N}} V_i^{1,\kappa}(\kappa\, C_i) \;\ge\; \sum_{i\in\mathcal{N}}\sum_{t\in\mathcal{T}^\kappa} r_i^{t,\kappa}\, \hat{\alpha}_i^{t,\kappa} - \sum_{i\in\mathcal{N}} R_{i,\max}\, \mathbb{E}\{[W_i^{1,\kappa} - \kappa\, C_i]^+\} \\ &\ge\; \sum_{i\in\mathcal{N}}\sum_{t\in\mathcal{T}^\kappa} r_i^{t,\kappa}\, \hat{\alpha}_i^{t,\kappa} - \frac{\sqrt{\kappa T}}{4} \sum_{i\in\mathcal{N}} R_{i,\max} \\ &=\; \sum_{i\in\mathcal{N}}\sum_{t\in\mathcal{T}^k}\sum_{A\in\mathcal{F}} r_i^{t,\kappa}\, \phi_i^{t,\kappa}(A)\, \hat{z}^{t,\kappa}(A) - \frac{\sqrt{\kappa T}}{4} \sum_{i\in\mathcal{N}} R_{i,\max} \;=\; \left(1 - \frac{\sqrt{\kappa T}}{4\,\kappa\, \hat{Z}^1} \sum_{i\in\mathcal{N}} R_{i,\max}\right) \hat{Z}^\kappa, \end{aligned}$$

where the last equality uses the fact that the optimal objective value of problem (15) for instance $\mathcal{P}^\kappa$ is given by $\sum_{i\in\mathcal{N}}\sum_{t\in\mathcal{T}^k}\sum_{A\in\mathcal{F}} r_i^{t,\kappa}\, \phi_i^{t,\kappa}(A)\, \hat{z}^{t,\kappa}(A) = \hat{Z}^\kappa = \kappa\, \hat{Z}^1$. Thus, letting $B = \frac{\sqrt{T}}{4\,\hat{Z}^1} \sum_{i\in\mathcal{N}} R_{i,\max}$, we get $V^{1,\kappa}(\sum_{i\in\mathcal{N}} \kappa\, C_i\, \boldsymbol{e}_i) \ge \left(1 - \frac{B}{\sqrt{\kappa}}\right) \hat{Z}^\kappa$. In this case, the rest of the proof follows by using precisely the same argument in the proof of Theorem F.2. ■

Appendix G: Heterogeneous Customer Types

In this section, we discuss the extension of our approach to the case where there are multiple customer types. In Section 5.1, we already discuss the notation that we use under heterogeneous customer types. We do not repeat the discussion of the notation here. We proceed to give a dynamic programming formulation under heterogeneous customer types. We use the vector $\boldsymbol{q} = ((q_{i,0}, q_{i,\ell}^j) : i \in \mathcal{N},\ j \in \mathcal{M},\ \ell \ge 1)$ as the state variable, where $q_{i,0}$ is the number of units of product i available on-hand and $q_{i,\ell}^j$ is the number of units of product i that have been used for exactly ℓ time periods by a customer of type j. In this case, the state space is given by $\mathcal{Q} = \{\big((q_{i,0} \in \mathbb{Z}_+, q_{i,\ell}^j \in \mathbb{Z}_+) \,:\, i \in \mathcal{N},\ j \in \mathcal{M},\ \ell \ge 1\big) \,:\, q_{i,0} + \sum_{j\in\mathcal{M}} \sum_{\ell=1}^\infty q_{i,\ell}^j = C_i\ \forall\, i \in \mathcal{N}\}$. We capture the decision at each time period by $(S^1, \ldots, S^m)$, where $S^j \subseteq \mathcal{N}$ is the assortment that we offer to customers of type j. The set of feasible assortments that we can offer to customers of type j is given by $\mathcal{F}^j$. As in Assumption 2.1, we assume that if $A \in \mathcal{F}^j$, then $S \in \mathcal{F}^j$ for all $S \subseteq A$. Similarly, the choice model $\{\phi_i^{t,j}(S) : i \in \mathcal{N},\ S \subseteq \mathcal{N}\}$ that drives the choices of customers of type j at time period t satisfies $\phi_i^{t,j}(S \cup \{k\}) \le \phi_i^{t,j}(S)$ for all $S \subseteq \mathcal{N}$, $k \in \mathcal{N}$ and $i \in S$. Given that the state is $\boldsymbol{q} \in \mathcal{Q}$ at the current time period, if there is no purchase, then the state at the next time period is given by the random vector $\boldsymbol{X}(\boldsymbol{q}) = \big(X_{i,0}(\boldsymbol{q}), X_{i,\ell}^j(\boldsymbol{q}) : i \in \mathcal{N},\ j \in \mathcal{M},\ \ell \ge 1\big)$, where we have

$$\begin{aligned} X_{i,0}(\boldsymbol{q}) \;&=\; q_{i,0} + \sum_{j\in\mathcal{M}} \sum_{s=1}^\infty \mathsf{Bin}\left(q_{i,s}^j, \rho_{i,s}^j\right), \\ X_{i,\ell}^j(\boldsymbol{q}) \;&=\; \begin{cases} 0 & \text{if } \ell = 1, \\ q_{i,\ell-1}^j - \mathsf{Bin}(q_{i,\ell-1}^j, \rho_{i,\ell-1}^j) & \text{if } \ell \ge 2. \end{cases} \end{aligned}$$

The transition dynamics above are similar to the one in (1). The only difference is that we need to keep track of the types of the customers using the units. Let $J^t(\boldsymbol{q})$ denote the maximum total expected revenue over the time periods $t,\dots,T$, given that the system is in state $\boldsymbol{q}$ at time period t. We can compute $\{J^t : t\in\mathcal{T}\}$ by solving the dynamic program

$$J^t(\boldsymbol{q}) = \sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{M}} \pi_i^{t,j}\sum_{\ell=1}^{\infty} q_{i,\ell}^j \;+\; \max_{(S^1,\dots,S^m)\in\mathcal{F}^1\times\dots\times\mathcal{F}^m}\Bigg\{\sum_{j\in\mathcal{M}} p^{t,j}\sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\geq 1\}}\,\phi_i^{t,j}(S^j)\times$$
$$\Big(r_i^{t,j}+\pi_i^{t,j}+\mathbb{E}\Big\{\mathsf{Z}(\rho_{i,0}^j)\,J^{t+1}(\boldsymbol{X}(\boldsymbol{q}))+(1-\mathsf{Z}(\rho_{i,0}^j))\,J^{t+1}\big(\boldsymbol{X}(\boldsymbol{q})-\boldsymbol{e}_{i,0}+\boldsymbol{e}_{i,1}^j\big)\Big\}\Big)$$
$$+\Big(1-\sum_{j\in\mathcal{M}} p^{t,j}\sum_{i\in\mathcal{N}}\mathbb{1}_{\{q_{i,0}\geq 1\}}\phi_i^{t,j}(S^j)\Big)\mathbb{E}\Big\{J^{t+1}(\boldsymbol{X}(\boldsymbol{q}))\Big\}\Bigg\},$$

with the boundary condition that $J^{T+1}=0$. Here, $\boldsymbol{e}_{i,1}^j$ is the unit vector with a one in the $(i,1)$-th coordinate associated with a customer of type j and zero everywhere else. Note that the dynamic program above is very similar to the dynamic program in (2). As in Section 2, we can write the dynamic program above equivalently as

$$J^t(\boldsymbol{q}) = \sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{M}} \pi_i^{t,j}\sum_{\ell=1}^{\infty} q_{i,\ell}^j \;+\; \mathbb{E}\Big\{J^{t+1}(\boldsymbol{X}(\boldsymbol{q}))\Big\}$$
$$+\sum_{j\in\mathcal{M}} p^{t,j}\max_{S^j\in\mathcal{F}^j}\Bigg\{\sum_{i\in\mathcal{N}}\mathbb{1}_{\{q_{i,0}\geq 1\}}\phi_i^{t,j}(S^j)\Big(r_i^{t,j}+\pi_i^{t,j}-(1-\rho_{i,0}^j)\mathbb{E}\Big\{J^{t+1}(\boldsymbol{X}(\boldsymbol{q}))-J^{t+1}\big(\boldsymbol{X}(\boldsymbol{q})-\boldsymbol{e}_{i,0}+\boldsymbol{e}_{i,1}^j\big)\Big\}\Big)\Bigg\}. \tag{23}$$

Once we observe that the maximization problem in our initial dynamic programming formulation decomposes by the customer types, the way we obtain the dynamic program above from the initial one is similar to the way we obtain the dynamic program in (3) from (2). Next, we construct an approximation to the optimal value function and bound the optimal total expected revenue.

Assuming that we have no products in use at the beginning of the selling horizon, the initial state of the system is given by $\boldsymbol{q}^1 = ((q_{i,0}^1, q_{i,\ell}^{j,1}) : i\in\mathcal{N},\ j\in\mathcal{M},\ \ell\geq 1) = ((C_i,0) : i\in\mathcal{N},\ j\in\mathcal{M},\ \ell\geq 1)$, so that the optimal total expected revenue is $J^1(\boldsymbol{q}^1)$. We use a value function approximation of the form $\hat{J}^t(\boldsymbol{q}) = \sum_{i\in\mathcal{N}}\hat{\theta}_i^t\,q_{i,0} + \sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{M}}\sum_{\ell=1}^{\infty}\hat{\nu}_{i,\ell}^{t,j}\,q_{i,\ell}^j$, where we compute $\hat{\theta}_i^t$ and $\hat{\nu}_{i,\ell}^{t,j}$ as discussed in Section 5.1. We consider the greedy policy with respect to the value function approximations $\{\hat{J}^t : t\in\mathcal{T}\}$. If the system is in state $\boldsymbol{q}$ at time period t, then this policy offers the assortment $\hat{S}^{t,j}(\boldsymbol{q})$ to a customer of type j, which is given by

$$\hat{S}^{t,j}(\boldsymbol{q}) = \arg\max_{S\in\mathcal{F}^j}\Bigg\{\sum_{i\in\mathcal{N}}\mathbb{1}_{\{q_{i,0}\geq 1\}}\phi_i^{t,j}(S)\Big(r_i^{t,j}+\pi_i^{t,j}-(1-\rho_{i,0}^j)\,\mathbb{E}\Big\{\hat{J}^{t+1}(\boldsymbol{X}(\boldsymbol{q}))-\hat{J}^{t+1}\big(\boldsymbol{X}(\boldsymbol{q})-\boldsymbol{e}_{i,0}+\boldsymbol{e}_{i,1}^j\big)\Big\}\Big)\Bigg\}$$
$$= \arg\max_{S\in\mathcal{F}^j}\Bigg\{\sum_{i=1}^{n}\mathbb{1}_{\{q_{i,0}\geq 1\}}\phi_i^{t,j}(S)\Big(r_i^{t,j}+\pi_i^{t,j}-(1-\rho_{i,0}^j)\Big[\hat{\theta}_i^{t+1}-\hat{\nu}_{i,1}^{t+1,j}\Big]\Big)\Bigg\}, \tag{24}$$

where the second equality follows from the definition of $\hat{J}^t$. Our main result under heterogeneous customer types is stated in the following theorem.

Theorem G.1 *Under heterogeneous customer types, the total expected revenue of the greedy policy with respect to the value function approximations $\{\hat{J}^t : t \in \mathcal{T}\}$ is at least 50% of the optimal total expected revenue.*

To show Theorem G.1, we use the next lemma. Note that $\boldsymbol{q}^1 = ((q^1_{i,0}, q^{j,1}_{i,\ell}) : i \in \mathcal{N},\ j \in \mathcal{M},\ \ell \geq 1) = ((C_i, 0) : i \in \mathcal{N},\ j \in \mathcal{M},\ \ell \geq 1)$ is the initial state of the system.

Lemma G.2 $J^1(\boldsymbol{q}^1) \leq 2 \sum_{i \in \mathcal{N}} \hat{\theta}^1_i\, C_i$.

Proof: We can obtain an upper bound on the optimal total expected revenue by using the objective value provided by any feasible solution to the linear program

$$
\begin{aligned}
\min \quad & \tilde{J}^1(\boldsymbol{q}^1) \\
\text{s.t.} \quad & \tilde{J}^t(\boldsymbol{q}) \;\geq\; \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{M}} \pi^{t,j}_i \sum_{\ell=1}^{\infty} q^j_{i,\ell} \;+\; \mathbb{E}\Big\{ \tilde{J}^{t+1}(\boldsymbol{X}(\boldsymbol{q})) \Big\} \\
& + \sum_{j \in \mathcal{M}} p^{t,j} \sum_{i \in \mathcal{N}} \mathbb{1}_{\{q_{i,0} \geq 1\}}\, \phi^{t,j}_i(S^j) \bigg(r^{t,j}_i + \pi^{t,j}_i - (1 - \rho^j_{i,0})\, \mathbb{E}\Big\{ \tilde{J}^{t+1}(\boldsymbol{X}(\boldsymbol{q})) - \tilde{J}^{t+1}\big(\boldsymbol{X}(\boldsymbol{q}) - \boldsymbol{e}_{i,0} + \boldsymbol{e}^j_{i,1}\big) \Big\} \bigg) \\
& \qquad \forall\, \boldsymbol{q} \in \mathcal{Q},\ (S^1, \ldots, S^m) \in \mathcal{F}^1 \times \cdots \times \mathcal{F}^m,\ t \in \mathcal{T},
\end{aligned}
$$

where the decision variables in the linear program above are $\{\tilde{J}^t(\boldsymbol{q}) : \boldsymbol{q} \in \mathcal{Q},\ t \in \mathcal{T}\}$ and we follow the convention that $\tilde{J}^{T+1} = 0$. Define the constant $\hat{\beta}^t = \sum_{i \in \mathcal{N}} \hat{\theta}^t_i\, C_i$, where $\hat{\theta}^t_i$ is computed as in (12). Letting the linear value function approximations $\{\hat{J}^t : t \in \mathcal{T}\}$ be defined through the algorithm in Section 5.1, we claim that $\{\hat{\beta}^t + \hat{J}^t(\boldsymbol{q}) : \boldsymbol{q} \in \mathcal{Q},\ t \in \mathcal{T}\}$ is a feasible solution to the linear program above. (The solution $\{\hat{J}^t(\boldsymbol{q}) : \boldsymbol{q} \in \mathcal{Q},\ t \in \mathcal{T}\}$ without the constant $\hat{\beta}^t$ is not necessarily feasible to the linear program.) To establish the claim, because $\hat{J}^{t+1}(\boldsymbol{q})$ is a linear function of $\boldsymbol{q}$ of the form $\hat{J}^{t+1}(\boldsymbol{q}) = \sum_{i \in \mathcal{N}} \hat{\theta}^{t+1}_i\, q_{i,0} + \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{M}} \sum_{\ell=1}^{\infty} \hat{\nu}^{t+1,j}_{i,\ell}\, q^j_{i,\ell}$, by the definition of $\boldsymbol{X}(\boldsymbol{q})$, we get

$$
\begin{aligned}
\hat{\beta}^{t+1} + \mathbb{E}\Big\{ \hat{J}^{t+1}(\boldsymbol{X}(\boldsymbol{q})) \Big\} &= \hat{\beta}^{t+1} + \sum_{i \in \mathcal{N}} \bigg\{ \hat{\theta}^{t+1}_i \Big[q_{i,0} + \sum_{j \in \mathcal{M}} \sum_{\ell=1}^{\infty} \rho^j_{i,\ell}\, q^j_{i,\ell} \Big] + \sum_{j \in \mathcal{M}} \sum_{\ell=1}^{\infty} \hat{\nu}^{t+1,j}_{i,\ell+1} \, [q^j_{i,\ell} - \rho^j_{i,\ell}\, q^j_{i,\ell}] \bigg\} \\
&= \hat{\beta}^{t+1} + \sum_{i \in \mathcal{N}} \bigg\{ q_{i,0}\, \hat{\theta}^{t+1}_i + \sum_{j \in \mathcal{M}} \sum_{\ell=1}^{\infty} q^j_{i,\ell} \, [\rho^j_{i,\ell}\, \hat{\theta}^{t+1}_i + (1 - \rho^j_{i,\ell})\, \hat{\nu}^{t+1,j}_{i,\ell+1}] \bigg\} \\
&= \hat{\beta}^{t+1} + \sum_{i \in \mathcal{N}} \bigg\{ q_{i,0}\, \hat{\theta}^{t+1}_i + \sum_{j \in \mathcal{M}} \sum_{\ell=1}^{\infty} q^j_{i,\ell} \, [\hat{\nu}^{t,j}_{i,\ell} - \pi^{t,j}_i] \bigg\}, \qquad (25)
\end{aligned}
$$

where the last equality follows from the way we compute $\hat{\nu}^{t,j}_{i,\ell}$ in (12). Similarly, using the fact that $\hat{J}^{t+1}(\boldsymbol{q})$ is linear in $\boldsymbol{q}$, we also have $\mathbb{E}\Big\{ \hat{J}^{t+1}(\boldsymbol{X}(\boldsymbol{q})) - \hat{J}^{t+1}(\boldsymbol{X}(\boldsymbol{q}) - \boldsymbol{e}_{i,0} + \boldsymbol{e}^j_{i,1}) \Big\} = \hat{\theta}^{t+1}_i - \hat{\nu}^{t+1,j}_{i,1}$ by the definitions of $\hat{J}^{t+1}$ and $X(\boldsymbol{q})$. Thus, if we evaluate the right side of the constraint in the linear program above at $\{\hat{\beta}^t + \hat{J}^t(\boldsymbol{q}) : \boldsymbol{q} \in \mathcal{Q},\ t \in \mathcal{T}\}$, then we get

$$\begin{aligned}
&\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{M}} \pi_i^{t,j} \sum_{\ell=1}^{\infty} q_{i,\ell}^j \;+\; \hat{\beta}^{t+1} \;+\; \mathbb{E}\Big\{\tilde{J}^{t+1}(\boldsymbol{X}(\boldsymbol{q}))\Big\} \\
&\quad + \sum_{j\in\mathcal{M}} p^{t,j} \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\geq 1\}}\, \phi_i^{t,j}(S^j)\bigg(r_i^{t,j} + \pi_i^{t,j} - (1-\rho_{i,0}^j)\,\mathbb{E}\Big\{\tilde{J}^{t+1}(\boldsymbol{X}(\boldsymbol{q})) - \tilde{J}^{t+1}\big(\boldsymbol{X}(\boldsymbol{q}) - \boldsymbol{e}_{i,0} + \boldsymbol{e}_{i,1}^j\big)\Big\}\bigg) \\
&= \sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{M}} \pi_i^{t,j} \sum_{\ell=1}^{\infty} q_{i,\ell}^j \;+\; \hat{\beta}^{t+1} + \sum_{i\in\mathcal{N}}\bigg\{q_{i,0}\,\hat{\theta}_i^{t+1} + \sum_{j\in\mathcal{M}}\sum_{\ell=1}^{\infty} q_{i,\ell}^j\,[\hat{\nu}_{i,\ell}^{t,j} - \pi_i^{t,j}]\bigg\} \\
&\quad + \sum_{j\in\mathcal{M}} p^{t,j} \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\geq 1\}}\, \phi_i^{t,j}(S^j)\bigg(r_i^{t,j} + \pi_i^{t,j} - (1-\rho_{i,0}^j)\,(\hat{\theta}_i^{t+1} - \hat{\nu}_{i,1}^{t+1,j})\bigg) \\
&= \sum_{i\in\mathcal{N}} \hat{\theta}_i^{t+1}\,C_i + \sum_{i\in\mathcal{N}}\bigg\{q_{i,0}\,\hat{\theta}_i^{t+1} + \sum_{j\in\mathcal{M}}\sum_{\ell=1}^{\infty} q_{i,\ell}^j\,\hat{\nu}_{i,\ell}^{t,j}\bigg\} \\
&\quad + \sum_{j\in\mathcal{M}} p^{t,j} \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\geq 1\}}\, \phi_i^{t,j}(S^j)\bigg(r_i^{t,j} + \pi_i^{t,j} - (1-\rho_{i,0}^j)\,(\hat{\theta}_i^{t+1} - \hat{\nu}_{i,1}^{t+1,j})\bigg),
\end{aligned}$$

where the second equality uses the definition of $\hat{\beta}^{t+1}$ at the beginning of the proof. Using the same argument in the proof of Lemma 3.1, we can show that $\hat{\theta}_i^t \geq \hat{\theta}_i^{t+1}$ under heterogeneous customer types. Furthermore, using the same argument in the proof of Lemma B.1, we can show that $\sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\geq 1\}}\, \phi_i^{t,j}(S^j)\Big[r_i^{t,j} + \pi_i^{t,j} - (1-\rho_{i,0}^j)\,(\hat{\theta}_i^{t+1} - \hat{\nu}_{i,1}^{t+1,j})\Big] \leq \sum_{i\in\mathcal{N}} \phi_i^{t,j}(\hat{A}^{t,j}) \times \Big[r_i^{t,j} + \pi_i^{t,j} - (1-\rho_{i,0}^j)\,(\hat{\theta}_i^{t+1} - \hat{\nu}_{i,1}^{t+1,j})\Big]$ for all $S^j \in \mathcal{F}^j$, where the assortment $\hat{A}^{t,j}$ is as given in the algorithm in Section 5.1. In this case, by the chain of equalities above, we can bound the right side of the constraint in the linear program as

$$\begin{aligned}
&\sum_{i\in\mathcal{N}} \hat{\theta}_i^{t+1}\,C_i + \sum_{i\in\mathcal{N}}\bigg\{q_{i,0}\,\hat{\theta}_i^{t+1} + \sum_{j\in\mathcal{M}}\sum_{\ell=1}^{\infty} q_{i,\ell}^j\,\hat{\nu}_{i,\ell}^{t,j}\bigg\} \\
&\qquad\qquad + \sum_{j\in\mathcal{M}} p^{t,j} \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\geq 1\}}\, \phi_i^{t,j}(S^j)\bigg(r_i^{t,j} + \pi_i^{t,j} - (1-\rho_{i,0}^j)\,(\hat{\theta}_i^{t+1} - \hat{\nu}_{i,1}^{t+1,j})\bigg) \\
&\quad\leq \sum_{i\in\mathcal{N}} \hat{\theta}_i^{t+1}\,C_i + \sum_{i\in\mathcal{N}}\bigg\{q_{i,0}\,\hat{\theta}_i^{t} + \sum_{j\in\mathcal{M}}\sum_{\ell=1}^{\infty} q_{i,\ell}^j\,\hat{\nu}_{i,\ell}^{t,j}\bigg\} \\
&\qquad\qquad + \sum_{j\in\mathcal{M}} p^{t,j} \sum_{i\in\mathcal{N}} \phi_i^{t,j}(\hat{A}^{t,j})\bigg(r_i^{t,j} + \pi_i^{t,j} - (1-\rho_{i,0}^j)\,(\hat{\theta}_i^{t+1} - \hat{\nu}_{i,1}^{t+1,j})\bigg) \\
&\quad= \sum_{i\in\mathcal{N}} \hat{\theta}_i^{t+1}\,C_i + \sum_{i\in\mathcal{N}}\bigg\{q_{i,0}\,\hat{\theta}_i^{t} + \sum_{j\in\mathcal{M}}\sum_{\ell=1}^{\infty} q_{i,\ell}^j\,\hat{\nu}_{i,\ell}^{t,j}\bigg\} + \sum_{i\in\mathcal{N}} C_i\,(\hat{\theta}_i^t - \hat{\theta}_i^{t+1}) \\
&\quad= \sum_{i\in\mathcal{N}} \hat{\theta}_i^{t}\,C_i + \sum_{i\in\mathcal{N}}\bigg\{q_{i,0}\,\hat{\theta}_i^{t} + \sum_{j\in\mathcal{M}}\sum_{\ell=1}^{\infty} q_{i,\ell}^j\,\hat{\nu}_{i,\ell}^{t,j}\bigg\} = \hat{\beta}^t + \sum_{i\in\mathcal{N}}\bigg\{q_{i,0}\,\hat{\theta}_i^{t} + \sum_{j\in\mathcal{M}}\sum_{\ell=1}^{\infty} q_{i,\ell}^j\,\hat{\nu}_{i,\ell}^{t,j}\bigg\} = \hat{\beta}^t + \hat{J}^t(\boldsymbol{q}),
\end{aligned}$$

where the first equality follows from the way we compute $\hat{\theta}_i^t$ in (12). By the chain of inequalities above, for any $\boldsymbol{q} \in \mathcal{Q}$, $(S^1,\ldots,S^m) \in \mathcal{F}^1 \times \ldots \times \mathcal{F}^m$, $t \in \mathcal{T}$, if we evaluate the right side of the constraint in the linear program at $\{\hat{\beta}^t + \hat{J}^t(\boldsymbol{q}) : \boldsymbol{q} \in \mathcal{Q},\ t \in \mathcal{T}\}$, then the right side of the constraint is upper bounded by $\hat{\beta}^t + \hat{J}^t(\boldsymbol{q})$. So, the solution $\{\hat{\beta}^t + \hat{J}^t(\boldsymbol{q}) : \boldsymbol{q} \in \mathcal{Q},\ t \in \mathcal{T}\}$ is feasible to the linear program, in which case, the objective value provided by this solution is an upper bound on the optimal total expected revenue. Noting the definition of $\boldsymbol{q}^1$, the objective value provided by the solution $\{\hat{\beta}^t + \hat{J}^t(\boldsymbol{q}) : \boldsymbol{q} \in \mathcal{Q},\ t \in \mathcal{T}\}$ is $\hat{\beta}^1 + \hat{J}^1(\boldsymbol{q}^1) = \hat{\beta}^1 + \sum_{i \in \mathcal{N}} C_i\, \hat{\theta}_i^1 = 2 \sum_{i \in \mathcal{N}} C_i\, \hat{\theta}_i^1$. Thus, $2 \sum_{i \in \mathcal{N}} C_i\, \hat{\theta}_i^1$ is an upper bound on the optimal total expected revenue. ■

If we are in state $\boldsymbol{q}$ at time period t, then the greedy policy offers the assortment $\hat{S}^{t,j}(\boldsymbol{q})$ in (24) to a customer of type j. Let $U^t(\boldsymbol{q})$ be the total expected revenue obtained by the greedy policy over the time periods $t, \ldots, T$, given that we are in state $\boldsymbol{q}$ at time period t. Using an argument similar to the one right before the proof of Theorem 3.2 and noting the dynamic program under heterogeneous customer types in (23), we can compute $\{U^t : t \in \mathcal{T}\}$ by using the recursion

$$U^t(\boldsymbol{q}) = \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{M}} \pi_i^{t,j} \sum_{\ell=1}^{\infty} q_{i,\ell}^j \ +\ \mathbb{E}\Big\{U^{t+1}(\boldsymbol{X}(\boldsymbol{q}))\Big\}$$
$$+ \sum_{j \in \mathcal{M}} p^{t,j} \sum_{i \in \mathcal{N}} \mathbb{1}_{\{q_{i,0} \geq 1\}}\, \phi_i^{t,j}(\hat{S}^{t,j}(\boldsymbol{q})) \Bigg(r_i^{t,j} + \pi_i^{t,j} - (1-\rho_{i,0}^j)\, \mathbb{E}\Big\{ U^{t+1}(\boldsymbol{X}(\boldsymbol{q})) - U^{t+1}\left(\boldsymbol{X}(\boldsymbol{q}) - \boldsymbol{e}_{i,0} + \boldsymbol{e}_{i,1}^j\right) \Big\} \Bigg),$$

with the boundary condition that $U^{T+1} = 0$. The coefficient of $\mathbb{E}\{U^{t+1}(\boldsymbol{X}(\boldsymbol{q}))\}$ above is $1 - \sum_{j \in \mathcal{M}} p^{t,j} \sum_{i \in \mathcal{N}} \mathbb{1}_{\{q_{i,0} \geq 1\}}\, \phi_i^{t,j}(\hat{S}^{t,j}(\boldsymbol{q}))(1-\rho_{i,0}^j)$. Since $\sum_{j \in \mathcal{M}} p^{t,j} = 1$ and $\sum_{i \in \mathcal{N}} \phi_i^{t,j}(\hat{S}^{t,j}(\boldsymbol{q})) \leq 1$, this coefficient is positive. The coefficient of $\mathbb{E}\left\{U^{t+1}\left(\boldsymbol{X}(\boldsymbol{q}) - \boldsymbol{e}_{i,0} + \boldsymbol{e}_{i,1}^j\right)\right\}$ is positive as well. Thus, if we replace the U^{t+1} on the right side above with a function H^{t+1} that satisfies $U^{t+1}(\boldsymbol{q}) \geq H^{t+1}(\boldsymbol{q})$, then the right side of the expression above gets smaller.

Here is the proof of Theorem G.1.

Proof of Theorem G.1: We will use induction over the time periods to show that $U^t(\boldsymbol{q}) \geq \hat{J}^t(\boldsymbol{q})$ for all $\boldsymbol{q} \in \mathcal{Q}$ and $t \in \mathcal{T}$, where $\hat{J}^t(\boldsymbol{q}) = \sum_{i \in \mathcal{N}} \hat{\theta}_i^t\, q_{i,0} + \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{M}} \sum_{\ell=1}^{\infty} \hat{\nu}_{i,\ell}^{t,j}\, q_{i,\ell}^j$ is the linear value function approximation computed through the algorithm in Section 5.1. In this case, noting that the initial state is given by $\boldsymbol{q}^1 = ((q_{i,0}^1, q_{i,\ell}^{j,1}) : i \in \mathcal{N},\ j \in \mathcal{M},\ \ell \geq 1) = ((C_i, 0) : i \in \mathcal{N},\ j \in \mathcal{M},\ \ell \geq 1)$, we obtain $U^1(\boldsymbol{q}^1) \geq \hat{J}^1(\boldsymbol{q}^1) = \sum_{i \in \mathcal{N}} \hat{\theta}_i^1\, C_i \geq \frac{1}{2} J^1(\boldsymbol{q}^1)$, where the second inequality follows from Lemma G.2. Therefore, the desired result follows. In the rest of the proof, we use induction over the time periods to show that $U^t(\boldsymbol{q}) \geq \hat{J}^t(\boldsymbol{q})$. Since $\hat{\theta}_i^{T+1} = 0$ and $\hat{\nu}_{i,\ell}^{T+1,j} = 0$ for all $i \in \mathcal{N}$, $j \in \mathcal{M}$, $\ell = 1, 2, \ldots$, we have $\hat{J}^{T+1} = 0$. We have $U^{T+1} = 0$ as well. Thus, the result holds at time period $T+1$. Assuming that $U^{t+1}(\boldsymbol{q}) \geq \hat{J}^{t+1}(\boldsymbol{q})$ for all $\boldsymbol{q} \in \mathcal{Q}$, we proceed to show that $U^t(\boldsymbol{q}) \geq \hat{J}^t(\boldsymbol{q})$ for all $\boldsymbol{q} \in \mathcal{Q}$ as well. By (25), we have $\mathbb{E}\{\hat{J}^{t+1}(\boldsymbol{X}(\boldsymbol{q}))\} = \sum_{i \in \mathcal{N}} \Big\{ q_{i,0}\, \hat{\theta}_i^{t+1} + \sum_{j \in \mathcal{M}} \sum_{\ell=1}^{\infty} q_{i,\ell}^j\, [\hat{\nu}_{i,\ell}^{t,j} - \pi_i^{t,j}] \Big\}$. Also, since $\hat{J}^t(\boldsymbol{q}) = \sum_{i \in \mathcal{N}} \hat{\theta}_i^t\, q_{i,0} + \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{M}} \sum_{\ell=1}^{\infty} \hat{\nu}_{i,\ell}^{t,j}\, q_{i,\ell}^j$, we have $\hat{J}^{t+1}(\boldsymbol{X}(\boldsymbol{q})) - \hat{J}^{t+1}(\boldsymbol{X}(\boldsymbol{q}) - \boldsymbol{e}_{i,0} + \boldsymbol{e}_{i,1}^j) =$

$\hat\theta_i^{t+1} - \hat\nu_{i,1}^{t+1,j}$, so that $\mathbb{E}\{\hat J^{t+1}(\boldsymbol{X}(\boldsymbol{q})) - \hat J^{t+1}(\boldsymbol{X}(\boldsymbol{q}) - \boldsymbol{e}_{i,0} + \boldsymbol{e}_{i,1}^j)\} = \hat\theta_i^{t+1} - \hat\nu_{i,1}^{t+1,j}$. In this case, by the inductive hypothesis and the recursion that defines $U^t(\boldsymbol{q})$, we get

$$
\begin{aligned}
U^t(\boldsymbol{q}) \;\geq\; & \sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{M}} \pi_i^{t,j} \sum_{\ell=1}^{\infty} q_{i,\ell}^j \;+\; \mathbb{E}\Big\{\hat J^{t+1}(\boldsymbol{X}(\boldsymbol{q}))\Big\} \\
& + \sum_{j\in\mathcal{M}} p^{t,j} \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\geq 1\}}\, \phi_i^{t,j}(\hat S^{t,j}(\boldsymbol{q}))\Big(r_i^{t,j} + \pi_i^{t,j} - (1-\rho_{i,0}^j)\,\mathbb{E}\Big\{\hat J^{t+1}(\boldsymbol{X}(\boldsymbol{q})) - \hat J^{t+1}\left(\boldsymbol{X}(\boldsymbol{q}) - \boldsymbol{e}_{i,0} + \boldsymbol{e}_{i,1}^j\right)\Big\}\Big) \\
= \;& \sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{M}} \pi_i^{t,j} \sum_{\ell=1}^{\infty} q_{i,\ell}^j \;+\; \sum_{i\in\mathcal{N}} \Big\{ q_{i,0}\,\hat\theta_i^{t+1} + \sum_{j\in\mathcal{M}}\sum_{\ell=1}^{\infty} q_{i,\ell}^j\,[\hat\nu_{i,\ell}^{t,j} - \pi_i^{t,j}] \Big\} \\
& \qquad + \sum_{j\in\mathcal{M}} p^{t,j} \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\geq 1\}}\, \phi_i^{t,j}(\hat S^{t,j}(\boldsymbol{q}))\Big(r_i^{t,j} + \pi_i^{t,j} - (1-\rho_{i,0}^j)\,(\hat\theta_i^{t+1} - \hat\nu_{i,1}^{t+1,j})\Big) \\
= \;& \sum_{i\in\mathcal{N}} \Big\{ q_{i,0}\,\hat\theta_i^{t+1} + \sum_{j\in\mathcal{M}}\sum_{\ell=1}^{\infty} q_{i,\ell}^j\,\hat\nu_{i,\ell}^{t,j} \Big\} \\
& \qquad + \sum_{j\in\mathcal{M}} p^{t,j} \max_{S^j\in\mathcal{F}^j} \Big\{ \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\geq 1\}}\, \phi_i^{t,j}(S^j)\Big(r_i^{t,j} + \pi_i^{t,j} - (1-\rho_{i,0}^j)\,(\hat\theta_i^{t+1} - \hat\nu_{i,1}^{t+1,j})\Big)\Big\},
\end{aligned}
$$

where the last equality is by the definition of $\hat S^{t,j}(\boldsymbol{q})$ in (24). Noting that $\hat A^{t,j}$ is a feasible but not necessarily an optimal solution to the last maximization problem, we have

$$
\begin{aligned}
& \sum_{i\in\mathcal{N}} \Big\{ q_{i,0}\,\hat\theta_i^{t+1} + \sum_{j\in\mathcal{M}}\sum_{\ell=1}^{\infty} q_{i,\ell}^j\,\hat\nu_{i,\ell}^{t,j} \Big\} \\
& \qquad\qquad + \sum_{j\in\mathcal{M}} p^{t,j} \max_{S^j\in\mathcal{F}^j} \Big\{ \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\geq 1\}}\, \phi_i^{t,j}(S^j)\Big(r_i^{t,j} + \pi_i^{t,j} - (1-\rho_{i,0}^j)\,(\hat\theta_i^{t+1} - \hat\nu_{i,1}^{t+1,j})\Big)\Big\} \\
\geq & \sum_{i\in\mathcal{N}} \Big\{ q_{i,0}\,\hat\theta_i^{t+1} + \sum_{j\in\mathcal{M}}\sum_{\ell=1}^{\infty} q_{i,\ell}^j\,\hat\nu_{i,\ell}^{t,j} \Big\} + \sum_{j\in\mathcal{M}} p^{t,j} \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\geq 1\}}\, \phi_i^{t,j}(\hat A^{t,j})\Big(r_i^{t,j} + \pi_i^{t,j} - (1-\rho_{i,0}^j)\,(\hat\theta_i^{t+1} - \hat\nu_{i,1}^{t+1,j})\Big) \\
\geq & \sum_{i\in\mathcal{N}} \Big\{ q_{i,0}\,\hat\theta_i^{t+1} + \sum_{j\in\mathcal{M}}\sum_{\ell=1}^{\infty} q_{i,\ell}^j\,\hat\nu_{i,\ell}^{t,j} \Big\} + \sum_{j\in\mathcal{M}} p^{t,j} \sum_{i\in\mathcal{N}} \frac{q_{i,0}}{C_i}\, \phi_i^{t,j}(\hat A^{t,j})\Big(r_i^{t,j} + \pi_i^{t,j} - (1-\rho_{i,0}^j)\,(\hat\theta_i^{t+1} - \hat\nu_{i,1}^{t+1,j})\Big) \\
= & \sum_{i\in\mathcal{N}} \Big\{ q_{i,0}\,\hat\theta_i^{t+1} + \sum_{j\in\mathcal{M}}\sum_{\ell=1}^{\infty} q_{i,\ell}^j\,\hat\nu_{i,\ell}^{t,j} \Big\} + \sum_{i\in\mathcal{N}} q_{i,0}\,(\hat\theta_i^t - \hat\theta_i^{t+1}) \\
= & \sum_{i\in\mathcal{N}} \Big\{ \hat\theta_i^t\, q_{i,0} + \sum_{j\in\mathcal{M}}\sum_{\ell=1}^{\infty} \hat\nu_{i,\ell}^{t,j}\, q_{i,\ell}^j \Big\} \;=\; \hat J^t(\boldsymbol{q}).
\end{aligned}
$$

In the second inequality above, we can use an argument similar to the one in the proof of Lemma 3.1 to show that $\phi_i^{t,j}(\hat A^{t,j})(r_i^{t,j} + \pi_i^{t,j} - (1-\rho_{i,0}^j)\,(\hat\theta_i^{t+1} - \hat\nu_{i,1}^{t+1,j})) \geq 0$ for all $i\in\mathcal{N}$, in which case, the second inequality follows by the fact that $\mathbb{1}_{\{q_{i,0}\geq 1\}} \geq q_{i,0}/C_i$ for any $q_{i,0}\leq C_i$. The first equality follows from the way we compute $\hat\theta_i^t$ in (12). The two chains of inequalities above show that $U^t(\boldsymbol{q}) \geq \hat J^t(\boldsymbol{q})$, establishing the desired claim. ■

Appendix H: Solving the Assortment Problem Approximately

We consider the case where we can solve problem (5) only approximately. Assume that we have an FPTAS such that for any $\epsilon > 0$, the FPTAS finds an assortment $\hat{A}^t$ that satisfies

$$(1+\epsilon)\sum_{i\in\mathcal{N}} \phi_i^t(\hat{A}^t)\Big[r_i^t + \pi_i^t - (1-\rho_{i,0})\left(\hat{\nu}_{i,0}^{t+1} - \hat{\nu}_{i,1}^{t+1}\right)\Big] \ge \max_{S\in\mathcal{F}} \sum_{i\in\mathcal{N}} \phi_i^t(S)\Big[r_i^t + \pi_i^t - (1-\rho_{i,0})\left(\hat{\nu}_{i,0}^{t+1} - \hat{\nu}_{i,1}^{t+1}\right)\Big]$$

in running time that is polynomial in n and $1/\epsilon$. We compute $\hat{\nu}_{i,\ell}^t$ for all $i\in\mathcal{N}$, $\ell\ge 0$ and $t\in\mathcal{T}$ as in (6), but the assortment $\hat{A}^t$ satisfies the inequality above, rather than being an optimal solution to problem (5). In other words, the assortment $\hat{A}^t$ is a $1/(1+\epsilon)$-approximate solution to problem (5). In the next lemma, we generalize Lemma 3.3.

Lemma H.1 *Assume that $\hat{A}^t$ is a $1/(1+\epsilon)$-approximate solution to problem (5) for all $t\in\mathcal{T}$. Then, we have $J^1\left(\sum_{i\in\mathcal{N}} C_i\, \boldsymbol{e}_{i,0}\right) \le 2\,(1+\epsilon)^T \sum_{i\in\mathcal{N}} C_i\, \hat{\nu}_{i,0}^1$.*

Proof: Assume that $\{\hat{\nu}_{i,\ell}^t : i\in\mathcal{N},\ \ell\ge 0,\ t\in\mathcal{T}\}$ are computed as in (6), but $\hat{A}^t$ is a $1/(1+\epsilon)$-approximate solution to problem (5). Define the constant $\hat{\beta}^t = \sum_{i\in\mathcal{N}} \hat{\nu}_{i,0}^t\, C_i$. For the value function approximation $\hat{J}^t(\boldsymbol{q}) = \sum_{i\in\mathcal{N}}\sum_{\ell=0}^\infty \hat{\nu}_{i,\ell}^t\, q_{i,\ell}$, we claim that $\{(1+\epsilon)^{T-t+1}(\hat{\beta}^t + \hat{J}^t(\boldsymbol{q})) : \boldsymbol{q}\in\mathcal{Q},\ t\in\mathcal{T}\}$ is a feasible solution to the linear program in the proof of Lemma 3.3. To show the claim, from the discussion in the proof of Lemma 3.3, recall that $\mathbb{E}\{\hat{J}^{t+1}(\boldsymbol{X}(\boldsymbol{q})) - \hat{J}^{t+1}(\boldsymbol{X}(\boldsymbol{q}) - \boldsymbol{e}_{i,0} + \boldsymbol{e}_{i,1})\} = \hat{\nu}_{i,0}^{t+1} - \hat{\nu}_{i,1}^{t+1}$ and $\hat{\beta}^{t+1} + \mathbb{E}\{\hat{J}^{t+1}(\boldsymbol{X}(\boldsymbol{q}))\} = \hat{\beta}^{t+1} + \sum_{i\in\mathcal{N}}\{q_{i,0}\,\hat{\nu}_{i,0}^{t+1} + \sum_{\ell=1}^\infty q_{i,\ell}\,[\rho_{i,\ell}\,\hat{\nu}_{i,0}^{t+1} + (1-\rho_{i,\ell})\,\hat{\nu}_{i,\ell+1}^{t+1}]\}$. In this case, if we evaluate the right side of the constraint in the linear program in the proof of Lemma 3.3 at the solution $\{(1+\epsilon)^{T-t+1}(\hat{\beta}^t + \hat{J}^t(\boldsymbol{q})) : \boldsymbol{q}\in\mathcal{Q},\ t\in\mathcal{T}\}$, then we obtain

$$\begin{aligned}
&\sum_{i\in\mathcal{N}} \pi_i^t \sum_{\ell=1}^\infty q_{i,\ell} \;+\; (1+\epsilon)^{T-t}\,\mathbb{E}\Big\{\hat{\beta}^{t+1} + \hat{J}^{t+1}(\boldsymbol{X}(\boldsymbol{q}))\Big\} \\
&\qquad + \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\ge 1\}}\,\phi_i^t(S)\,\Big[r_i^t + \pi_i^t - (1-\rho_{i,0})\,(1+\epsilon)^{T-t}\,\mathbb{E}\Big\{\hat{J}^{t+1}(\boldsymbol{X}(\boldsymbol{q})) - \hat{J}^{t+1}(\boldsymbol{X}(\boldsymbol{q}) - \boldsymbol{e}_{i,0} + \boldsymbol{e}_{i,1})\Big\}\Big] \\
&= \sum_{i\in\mathcal{N}} \pi_i^t \sum_{\ell=1}^\infty q_{i,\ell} + (1+\epsilon)^{T-t}\,\hat{\beta}^{t+1} + (1+\epsilon)^{T-t}\sum_{i\in\mathcal{N}}\left\{q_{i,0}\,\hat{\nu}_{i,0}^{t+1} + \sum_{\ell=1}^\infty q_{i,\ell}\left[\rho_{i,\ell}\,\hat{\nu}_{i,0}^{t+1} + (1-\rho_{i,\ell})\,\hat{\nu}_{i,\ell+1}^{t+1}\right]\right\} \\
&\qquad + \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\ge 1\}}\,\phi_i^t(S)\,\left[r_i^t + \pi_i^t - (1+\epsilon)^{T-t}\,(1-\rho_{i,0})\,(\hat{\nu}_{i,0}^{t+1} - \hat{\nu}_{i,1}^{t+1})\right] \\
&\le (1+\epsilon)^{T-t}\,\hat{\beta}^{t+1} + (1+\epsilon)^{T-t}\sum_{i\in\mathcal{N}}\left\{q_{i,0}\,\hat{\nu}_{i,0}^{t+1} + \sum_{\ell=1}^\infty q_{i,\ell}\,\hat{\nu}_{i,\ell}^t\right\} \\
&\qquad + \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\ge 1\}}\,\phi_i^t(S)\,\left[r_i^t + \pi_i^t - (1-\rho_{i,0})\,(1+\epsilon)^{T-t}\,(\hat{\nu}_{i,0}^{t+1} - \hat{\nu}_{i,1}^{t+1})\right],
\end{aligned}$$

where the inequality holds because we have $\hat{\nu}_{i,\ell}^t = \pi_i^t + \rho_{i,\ell}\,\hat{\nu}_{i,0}^{t+1} + (1-\rho_{i,\ell})\,\hat{\nu}_{i,\ell+1}^{t+1}$ by (6). By Lemma B.1, we have $\sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\ge 1\}}\,\phi_i^t(S)\,\left[r_i^t + \pi_i^t - (1-\rho_{i,0})\,(\hat{\nu}_{i,0}^{t+1} - \hat{\nu}_{i,1}^{t+1})\right] \le \max_{A\in\mathcal{F}} \sum_{i\in\mathcal{N}} \phi_i^t(A)\ \times$

$\left[r_i^t + \pi_i^t - (1-\rho_{i,0})\,(\hat\nu_{i,0}^{t+1} - \hat\nu_{i,1}^{t+1})\right]$ for all $S \in \mathcal{F}$, since this lemma assumes that $\hat A^t$ is the optimal solution to the last maximization problem. So, we continue the chain of inequalities above as

$$
\begin{aligned}
&(1+\epsilon)^{T-t}\,\hat\beta^{t+1} + (1+\epsilon)^{T-t}\sum_{i\in\mathcal{N}}\left\{q_{i,0}\,\hat\nu_{i,0}^{t+1} + \sum_{\ell=1}^{\infty} q_{i,\ell}\hat\nu_{i,\ell}^t\right\} \\
&\qquad\qquad + \sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\,\geq\,1\}}\,\phi_i^t(S)\left[r_i^t+\pi_i^t-(1-\rho_{i,0})\,(1+\epsilon)^{T-t}\,(\hat\nu_{i,0}^{t+1}-\hat\nu_{i,1}^{t+1})\right] \\
&\leq\; (1+\epsilon)^{T-t}\,\hat\beta^{t+1} + (1+\epsilon)^{T-t}\sum_{i\in\mathcal{N}}\left\{q_{i,0}\,\hat\nu_{i,0}^{t+1} + \sum_{\ell=1}^{\infty} q_{i,\ell}\hat\nu_{i,\ell}^t\right\} \\
&\qquad\qquad + (1+\epsilon)^{T-t}\sum_{i\in\mathcal{N}} \mathbb{1}_{\{q_{i,0}\,\geq\,1\}}\,\phi_i^t(S)\left[r_i^t+\pi_i^t-(1-\rho_{i,0})\,(\hat\nu_{i,0}^{t+1}-\hat\nu_{i,1}^{t+1})\right] \\
&\leq\; (1+\epsilon)^{T-t}\,\hat\beta^{t+1} + (1+\epsilon)^{T-t}\sum_{i\in\mathcal{N}}\left\{q_{i,0}\,\hat\nu_{i,0}^{t+1} + \sum_{\ell=1}^{\infty} q_{i,\ell}\hat\nu_{i,\ell}^t\right\} \\
&\qquad\qquad + (1+\epsilon)^{T-t+1}\frac{1}{1+\epsilon}\max_{A\in\mathcal{F}}\left\{\sum_{i\in\mathcal{N}}\phi_i^t(A)\left[r_i^t+\pi_i^t-(1-\rho_{i,0})\,(\hat\nu_{i,0}^{t+1}-\hat\nu_{i,1}^{t+1})\right]\right\} \\
&\leq\; (1+\epsilon)^{T-t}\,\hat\beta^{t+1} + (1+\epsilon)^{T-t}\sum_{i\in\mathcal{N}}\left\{q_{i,0}\,\hat\nu_{i,0}^{t+1} + \sum_{\ell=1}^{\infty} q_{i,\ell}\hat\nu_{i,\ell}^t\right\} \\
&\qquad\qquad + (1+\epsilon)^{T-t+1}\sum_{i\in\mathcal{N}}\phi_i^t(\hat A^t)\left[r_i^t+\pi_i^t-(1-\rho_{i,0})\,(\hat\nu_{i,0}^{t+1}-\hat\nu_{i,1}^{t+1})\right], \\
&=\; (1+\epsilon)^{T-t}\sum_{i\in\mathcal{N}}\hat\nu_{i,0}^{t+1}\,C_i + (1+\epsilon)^{T-t}\sum_{i\in\mathcal{N}}\left\{q_{i,0}\,\hat\nu_{i,0}^{t+1} + \sum_{\ell=1}^{\infty} q_{i,\ell}\hat\nu_{i,\ell}^t\right\} \\
&\qquad\qquad + (1+\epsilon)^{T-t+1}\sum_{i\in\mathcal{N}} C_i\,(\hat\nu_{i,0}^t - \hat\nu_{i,0}^{t+1}),
\end{aligned}
$$

where the last inequality is by the fact that $\hat A^t$ is $1/(1+\epsilon)$-approximate solution to problem (5) and the last equality holds since $\hat\nu_{i,0}^t = \hat\nu_{i,0}^{t+1} + \frac{1}{C_i}\phi_i^t(\hat A^t)[r_i^t+\pi_i^t-(1-\rho_{i,0})\left(\hat\nu_{i,0}^{t+1}-\hat\nu_{i,1}^{t+1}\right)]$ by (6) and $\hat\beta^{t+1} = \sum_{i\in\mathcal{N}}\hat\nu_{i,0}^{t+1}\,C_i$. Even if we choose $\hat A^t$ as an approximate solution to problem (5), we can follow precisely the same reasoning in the proof of Lemma 3.1 to show that we can drop each product i with $\phi_i^t(\hat A^t)\,\left(r_i^t+\pi_i^t-(1-\rho_{i,0})\left(\hat\nu_{i,0}^{t+1}-\hat\nu_{i,1}^{t+1}\right)\right)\leq 0$ from $\hat A^t$ without deteriorating the objective value of problem (5) provided by the solution $\hat A^t$. Thus, by the reasoning in the proof of Lemma 3.1, we can assume that $\hat\nu_{i,0}^t \geq \hat\nu_{i,0}^{t+1}$. So, we continue the last chain of inequalities as

$$
\begin{aligned}
&(1+\epsilon)^{T-t}\sum_{i\in\mathcal{N}}\hat\nu_{i,0}^{t+1}\,C_i + (1+\epsilon)^{T-t}\sum_{i\in\mathcal{N}}\left\{q_{i,0}\,\hat\nu_{i,0}^{t+1} + \sum_{\ell=1}^{\infty} q_{i,\ell}\hat\nu_{i,\ell}^t\right\} + (1+\epsilon)^{T-t+1}\sum_{i\in\mathcal{N}} C_i\,(\hat\nu_{i,0}^t - \hat\nu_{i,0}^{t+1}) \\
&\quad\leq\; (1+\epsilon)^{T-t+1}\sum_{i\in\mathcal{N}}\hat\nu_{i,0}^{t+1}\,C_i + (1+\epsilon)^{T-t+1}\sum_{i\in\mathcal{N}}\left\{q_{i,0}\,\hat\nu_{i,0}^{t} + \sum_{\ell=1}^{\infty} q_{i,\ell}\hat\nu_{i,\ell}^t\right\} + (1+\epsilon)^{T-t+1}\sum_{i\in\mathcal{N}} C_i\,(\hat\nu_{i,0}^t - \hat\nu_{i,0}^{t+1}) \\
&\quad=\; (1+\epsilon)^{T-t+1}\sum_{i\in\mathcal{N}}\left\{\hat\nu_{i,0}^t\,C_i + \sum_{\ell=0}^{\infty} q_{i,\ell}\hat\nu_{i,\ell}^t\right\} \;=\; (1+\epsilon)^{T-t+1}\,(\hat\beta^t + \hat J^t(\boldsymbol{q})).
\end{aligned}
$$

By the discussion so far, for any $\boldsymbol{q}\in\mathcal{Q}$, $S\in\mathcal{F}$ and $t\in\mathcal{T}$, if we evaluate the right side of the constraint at $\{(1+\epsilon)^{T-t+1}\,(\hat\beta^t + \hat J^t(\boldsymbol{q})) : \boldsymbol{q}\in\mathcal{Q},\ t\in\mathcal{T}\}$, then the right side of the constraint is upper bounded by $(1+\epsilon)^{T-t+1}(\hat{\beta}^t + \hat{J}^t(\boldsymbol{q}))$. Thus, the solution $\{(1+\epsilon)^{T-t+1}(\hat{\beta}^t + \hat{J}^t(\boldsymbol{q})) : \boldsymbol{q} \in \mathcal{Q},\ t \in \mathcal{T}\}$ is feasible to the linear program, so the objective value of the linear program at this solution is an upper bound on the optimal total expected revenue. The desired result follows by noting that the objective value of the linear program at the solution $\{(1+\epsilon)^{T-t+1}(\hat{\beta}^t + \hat{J}^t(\boldsymbol{q})) : \boldsymbol{q} \in \mathcal{Q},\ t \in \mathcal{T}\}$ is $(1+\epsilon)^T\,(\hat{\beta}^1 + \hat{J}^1(\sum_{i\in\mathcal{N}} C_i\, \boldsymbol{e}_{i,0})) = (1+\epsilon)^T\,(\sum_{i\in\mathcal{N}} \hat{\nu}^1_{i,0}\, C_i + \sum_{i\in\mathcal{N}} \hat{\nu}^1_{i,0}\, C_i)$. ∎

Consider a greedy policy with respect to the value function approximations $\{\hat{J}^t : t \in \mathcal{T}\}$. To compute the decision of this policy, we need to solve the combinatorial optimization problem in (7), which has the same structure as the one in (5). Therefore, we assume that we can obtain only an approximate solution to problem (7). In particular, if the state of the system at time period t is $\boldsymbol{q}$, then the greedy policy offers the assortment $\hat{S}^t(\boldsymbol{q})$ such that

$$\begin{aligned}(1+\epsilon)\sum_{i=1}^n \mathbb{1}_{\{q_{i,0}\ge 1\}}\phi^t_i(\hat{S}^t(\boldsymbol{q}))\Big[r^t_i + \pi^t_i - (1-\rho_{i,0})(\hat{\nu}^{t+1}_{i,0} - \hat{\nu}^{t+1}_{i,1})\Big] \\ \ge \max_{S\in\mathcal{F}} \sum_{i=1}^n \mathbb{1}_{\{q_{i,0}\ge 1\}}\phi^t_i(S)\Big[r^t_i + \pi^t_i - (1-\rho_{i,0})(\hat{\nu}^{t+1}_{i,0} - \hat{\nu}^{t+1}_{i,1})\Big].\end{aligned}$$

We can compute the total expected revenue obtained by this greedy policy through the recursion in (8). The only difference is that the assortment $\hat{S}^t(\boldsymbol{q})$ is a $1/(1+\epsilon)$-approximate solution to problem (7), rather than the optimal solution. We let $U^t(\boldsymbol{q})$ be the total expected revenue obtained by the greedy policy over the time periods $t,\dots,T$, given that the system is in state $\boldsymbol{q}$ at time period t. We have the following lemma for the total expected revenue of the greedy policy.

Lemma H.2 *For all $\boldsymbol{q} \in \mathcal{Q}$ and $t \in \mathcal{T}$, we have $(1+\epsilon)^{T-t+1}\, U^t(\boldsymbol{q}) \ge \sum_{i\in\mathcal{N}} \sum_{\ell=0}^{\infty} \hat{\nu}^t_{i,\ell}\, q_{i,\ell}$.*

The proof of this lemma is omitted and it follows from induction over the time periods by using the ideas in the proofs of Theorem 3.2 and Lemma H.1. Here is the proof of Theorem 5.1.

Proof of Theorem 5.1: For any $\delta > 0$, we will show that we can obtain a $1/(2\,(1+\delta))$-approximate policy and the running time to obtain and execute the approximate policy is polynomial in n, $1/\delta$ and T. Assume for the moment that $\delta \le 1$. Given such δ, set $\epsilon = \delta/(4T)$, choose $\hat{A}^t$ as a $1/(1+\epsilon)$-approximate solution to problem (5) and choose $\hat{S}^t(\boldsymbol{q})$ as a $1/(1+\epsilon)$-approximate solution to problem (7). Since we can obtain these approximate solutions in running times polynomial in n and $1/\epsilon$, the running times involved are polynomial in n and T/δ, which are, in turn, polynomial in n, $1/\delta$ and T, establishing the desired running time. By Lemmas H.1 and H.2, we have

$$2\,(1+\epsilon)^{2T}\, U^1\big(\textstyle\sum_{i\in\mathcal{N}} C_i\, \boldsymbol{e}_{i,0}\big) \ge 2\,(1+\epsilon)^T \displaystyle\sum_{i\in\mathcal{N}} \hat{\nu}^1_{i,0}\, C_i \ge J^1\big(\textstyle\sum_{i\in\mathcal{N}} C_i\, \boldsymbol{e}_{i,0}\big)\,.$$

Letting Opt be the optimal total expected revenue and Gre be the total expected revenue from the greedy policy, noting that $\epsilon = \delta/(4T)$, the chain of inequalities above yields Opt $\le$

$2\,(1+\frac{\delta}{4T})^{2T}\,\textsc{Gre} \leq 2\,\exp(\delta/2)\,\textsc{Gre} \leq 2\,(1+\delta)\,\textsc{Gre}$, where the last inequality follows from the fact that $\exp(\delta/2) \leq 1+\delta$ for all $\delta \in [0,1]$. The last chain of inequalities shows that the greedy policy is a $1/(2\,(1+\delta))$-approximate policy. Lastly, if $\delta > 1$, then we can simply choose $\epsilon = 1/(4T)$. ■

Appendix I: Upper Bound on the Optimal Total Expected Revenue

In this section, we give a proof for Proposition 6.1.

Proof of Proposition 6.1: Under the optimal policy, we use $Q^t_{i,\ell}$ to denote the number of units of product i that have been in use for ℓ time periods at time period t. Also, under the optimal policy, we let $Z^t(A) = 1$ if we offer assortment A at time period t; otherwise, we have $Z^t(A) = 0$. Lastly, under the optimal policy, we let $\Phi^t_i = 1$ if the customer arriving at time period t chooses product i; otherwise, we have $\Phi^t_i = 0$. Note that $Q^t_{i,\ell}$, $Z^t(A)$ and Φ^t_i are random variables. Furthermore, $\Pr\{\Phi^t_i = 1 \,|\, Z^t(A) = 1\} = \phi^t_i(A)$. Using the vector $\boldsymbol{Q}^t = (Q^t_{i,\ell} : i \in \mathcal{N},\ \ell \geq 0)$, by the transition dynamics of our dynamic assortment problem, we have

$$\begin{aligned}
\boldsymbol{Q}^{t+1} &= \sum_{i\in\mathcal{N}} \Phi^t_i \Big[\mathsf{Z}(\rho_{i,0})\, \boldsymbol{X}(\boldsymbol{Q}^t) + (1-\mathsf{Z}(\rho_{i,0}))\,(\boldsymbol{X}(\boldsymbol{Q}^t) - \boldsymbol{e}_{i,0} + \boldsymbol{e}_{i,1}) \Big] + \Big\{ 1 - \sum_{i\in\mathcal{N}} \Phi^t_i \Big\} \boldsymbol{X}(\boldsymbol{Q}^t) \\
&= \boldsymbol{X}(\boldsymbol{Q}^t) - \sum_{i\in\mathcal{N}} \Phi^t_i\,(1-\mathsf{Z}(\rho_{i,0}))\,(\boldsymbol{e}_{i,0} - \boldsymbol{e}_{i,1}),
\end{aligned}$$

where the first equality uses an argument similar to the one that we use to justify the first constraint in problem (13). Since $\Pr\{\Phi^t_i = 1 \,|\, Z^t(A) = 1\} = \phi^t_i(A)$, we have $\mathbb{E}\{\Phi^t_i\} = \sum_{A\in\mathcal{F}} \Pr\{Z^t(A) = 1\}\, \Pr\{\Phi^t_i = 1 \,|\, Z^t(A) = 1\} = \sum_{A\in\mathcal{F}} \phi^t_i(A)\, \Pr\{Z^t(A) = 1\}$. In this case, letting $\bar{q}^t_{i,\ell} = \mathbb{E}\{Q^t_{i,\ell}\}$ and $\bar{z}^t(A) = \mathbb{E}\{Z^t(A)\}$, taking expectations in the chain of equalities above and noting that $\mathbb{E}\{\boldsymbol{X}(\boldsymbol{q})\}$ is linear in $\boldsymbol{q}$, it follows that the solution $(\bar{z}^t(A) : A \in \mathcal{F},\ t \in \mathcal{T})$ and $(\bar{q}^t_{i,\ell} : i \in \mathcal{N},\ \ell \geq 0,\ t \in \mathcal{T})$ satisfies the first constraint in problem (13). Under the optimal policy, we start with the initial state $\sum_{i\in\mathcal{N}} C_i\, \boldsymbol{e}_{i,0}$ and offer one assortment at each time period, so $\boldsymbol{Q}^1 = \sum_{i\in\mathcal{N}} C_i\, \boldsymbol{e}_{i,0}$ and $\sum_{A\in\mathcal{F}} Z^t(A) = 1$. Taking expectations in the last two equalities indicates that the solution $(\bar{z}^t(A) : A \in \mathcal{F},\ t \in \mathcal{T})$ and $(\bar{q}^t_{i,\ell} : i \in \mathcal{N},\ \ell \geq 0,\ t \in \mathcal{T})$ satisfies the second and third constraints in problem (13) as well. Therefore, this solution is feasible to problem (13). Furthermore, noting that $\mathbb{E}\{\Phi^t_i\} = \sum_{A\in\mathcal{F}} \phi^t_i(A)\, \Pr\{Z^t(A) = 1\} = \sum_{A\in\mathcal{F}} \phi^t_i(A)\, \bar{z}^t(A)$, the total expected revenue under the optimal policy is

$$\begin{aligned}
J^1\left(\sum_{i\in\mathcal{N}} C_i\, \boldsymbol{e}_{i,0} \right) &= \mathbb{E}\left\{ \sum_{t\in\mathcal{T}} \sum_{i\in\mathcal{N}} (r^t_i + \pi^t_i)\, \Phi^t_i + \sum_{t\in\mathcal{T}} \sum_{i\in\mathcal{N}} \pi^t_i \sum_{\ell=1}^{\infty} Q^t_{i,\ell} \right\} \\
&= \sum_{t\in\mathcal{T}} \sum_{i\in\mathcal{N}} (r^t_i + \pi^t_i) \sum_{A\in\mathcal{F}} \phi^t_i(A)\, \bar{z}^t(A) + \sum_{t\in\mathcal{T}} \sum_{i\in\mathcal{N}} \pi^t_i \sum_{\ell=1}^{\infty} \bar{q}^t_{i,\ell},
\end{aligned}$$

which is the objective value that the solution $(\bar{z}^t(A) : A \in \mathcal{F},\ t \in \mathcal{T})$ and $(\bar{q}^t_{i,\ell} : i \in \mathcal{N},\ \ell \geq 0,\ t \in \mathcal{T})$ provides for problem (13). So, there exists a feasible solution to problem (13) that provides an objective value of $J^1\left(\sum_{i\in\mathcal{N}} C_i\,\boldsymbol{e}_{i,0}\right)$. Therefore, the optimal objective value of problem (13) must be at least $J^1\left(\sum_{i\in\mathcal{N}} C_i\,\boldsymbol{e}_{i,0}\right)$. ■

Appendix J: Linear Programming Approximation under the Multinomial Logit Model

We consider the case where the customers choose according to the multinomial logit model. Under the multinomial logit model, a customer arriving at time period t associates the preference weight v_i^t with product i and the preference weight v_0^t with the no-purchase option. If we offer the assortment S, then a customer arriving at time period t chooses product i with probability $\phi_i^t(S)=v_i^t/(v_0^t+\sum_{j\in S} v_j^t)$. In this section, we show that if the customer arriving at each time period chooses according to the multinomial logit model, then we can give an equivalent formulation for problem (13), whose numbers of decision variables and constraints increase only linearly with the number of products. In the equivalent formulation, we use the decision variables $(y_i^t : i\in\mathcal{N}\cup\{0\},\ t\in\mathcal{T})$, where y_i^t captures the probability that a customer arriving at time period t chooses product i and y_0^t captures the probability that a customer arriving at time period t leaves without choosing any of the products. In this case, we show that problem (13) has the same optimal objective value as the linear program

$$
\begin{aligned}
\max \quad & \sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{N}}(r_i^t+\pi_i^t)\,y_i^t+\sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{N}}\pi_i^t\sum_{\ell=1}^{\infty} q_{i,\ell}^t \qquad\qquad (26)\\
\text{s.t.} \quad & \boldsymbol{q}^{t+1} \;=\; \mathbb{E}\Big\{\boldsymbol{X}(\boldsymbol{q}^t)\Big\}-\sum_{i\in\mathcal{N}} y_i^t\,(1-\rho_{i,0})\,(\boldsymbol{e}_{i,0}-\boldsymbol{e}_{i,1}) \quad \forall\, t\in\mathcal{T}\setminus\{T\}\\
& \boldsymbol{q}^1=\sum_{i\in\mathcal{N}} C_i\,\boldsymbol{e}_{i,0}\\
& \sum_{i\in\mathcal{N}} y_i^t+y_0^t \;=\; 1 \quad \forall\, t\in\mathcal{T}\\
& \frac{y_i^t}{v_i^t}\le\frac{y_0^t}{v_0^t} \quad \forall\, i\in\mathcal{N},\ t\in\mathcal{T}\\
& y_i^t\ge 0 \quad \forall\, i\in\mathcal{N}\cup\{0\},\ t\in\mathcal{T},\ q_{i,\ell}^t\ge 0 \quad \forall\, i\in\mathcal{N},\ \ell\ge 0,\ t\in\mathcal{T}.
\end{aligned}
$$

Noting that y_i^t captures the probability that a customer arriving at time period t chooses product i, the objective function and the first constraint above have the same interpretation as the objective function and the first constraint in problem (13) after replacing $\sum_{A\in\mathcal{F}}\phi_i^t(A)\,z^t(A)$ with y_i^t. The second constraint above initializes the state of the system. The third constraint ensures that the customer arriving at each time period either chooses a product or leaves without choosing any of the products. The fourth constraint ensures that the choice probabilities of the customer arriving at each time period are consistent with the multinomial logit model.

To show that problems (13) and (26) have the same optimal objective values, noting that $\boldsymbol{q}^{t+1}$ in the first constraint in problem (13) corresponds to the vector $\boldsymbol{q}^{t+1}=(q_{i,\ell}^t : i\in\mathcal{N},\ \ell\ge 0)$, we use $\boldsymbol{\alpha} = (\alpha_{i,\ell}^{t+1} : i \in \mathcal{N},\ \ell \geq 0,\ t \in \mathcal{T} \setminus \{T\})$ to denote the dual variables associated with the first constraint. Defining the vector $\boldsymbol{\alpha}^{t+1} = (\alpha_{i,\ell}^{t+1} : i \in \mathcal{N},\ \ell \geq 0)$ and letting $\boldsymbol{x} \cdot \boldsymbol{y}$ to denote the scalar product of the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, if we dualize the first constraint in problem (13) by associating the dual variables $\boldsymbol{\alpha}$, then we can write the objective function of problem (13) as

$$
\begin{aligned}
&\sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{N}}\sum_{A\in\mathcal{F}} (r_i^t + \pi_i^t)\, \phi_i^t(A)\, z^t(A) + \sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{N}} \pi_i^t \sum_{\ell=1}^{\infty} q_{i,\ell}^t \\
&\qquad + \sum_{t\in\mathcal{T}\setminus\{T\}} \boldsymbol{\alpha}^{t+1} \cdot \Big(\mathbb{E}\Big\{ \boldsymbol{X}(\boldsymbol{q}^t) \Big\} - \sum_{i\in\mathcal{N}} \Big\{ \sum_{A\in\mathcal{F}} \phi_i^t(A)\, z^t(A) \Big\} (1-\rho_{i,0})\, (\boldsymbol{e}_{i,0} - \boldsymbol{e}_{i,1}) - \boldsymbol{q}^{t+1} \Big) \\
&= \sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{N}}\sum_{A\in\mathcal{F}} \Big[r_i^t + \pi_i^t - (1-\rho_{i,0})\, \big(\alpha_{i,0}^{t+1} - \alpha_{i,1}^{t+1}\big) \Big]\, \phi_i^t(A)\, z^t(A) + \sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{N}} \pi_i^t \sum_{\ell=1}^{\infty} q_{i,\ell}^t \\
&\qquad + \sum_{t\in\mathcal{T}\setminus\{T\}} \boldsymbol{\alpha}^{t+1} \cdot \Big(\mathbb{E}\Big\{ \boldsymbol{X}(\boldsymbol{q}^t) \Big\} - \boldsymbol{q}^{t+1} \Big),
\end{aligned}
$$

where we follow the convention that $\alpha_{i,\ell}^{T+1} = 0$ and use the fact that $\boldsymbol{\alpha}^{t+1} \cdot \boldsymbol{e}_{i,0} = \alpha_{i,0}^{t+1}$ and $\boldsymbol{\alpha}^{t+1} \cdot \boldsymbol{e}_{i,1} = \alpha_{i,1}^{t+1}$. Therefore, by linear programming duality, if we define $F(\boldsymbol{\alpha})$ as

$$
\begin{aligned}
F(\boldsymbol{\alpha}) = \max\ & \sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{N}}\sum_{A\in\mathcal{F}} \Big[r_i^t + \pi_i^t - (1-\rho_{i,0})\, \big(\alpha_{i,0}^{t+1} - \alpha_{i,1}^{t+1}\big) \Big]\, \phi_i^t(A)\, z^t(A) + \sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{N}} \pi_i^t \sum_{\ell=1}^{\infty} q_{i,\ell}^t \\
&\qquad + \sum_{t\in\mathcal{T}\setminus\{T\}} \boldsymbol{\alpha}^{t+1} \cdot \Big(\mathbb{E}\Big\{ \boldsymbol{X}(\boldsymbol{q}^t) \Big\} - \boldsymbol{q}^{t+1} \Big) \qquad (27) \\
\text{s.t. }\ & \boldsymbol{q}^1 = \sum_{i\in\mathcal{N}} C_i\, \boldsymbol{e}_{i,0} \\
& \sum_{A\in\mathcal{F}} z^t(A) = 1 \quad \forall\, t \in \mathcal{T} \\
& z^t(A) \geq 0 \quad \forall\, A \in \mathcal{F},\ t \in \mathcal{T},\ q_{i,\ell}^t \geq 0 \quad \forall\, i \in \mathcal{N},\ \ell \geq 0,\ t \in \mathcal{T},
\end{aligned}
$$

then the optimal objective value of problem (13) is given by $\min_{\boldsymbol{\alpha}} F(\boldsymbol{\alpha})$. Also, using precisely the same sequence of steps but by working with problem (26), if we define $G(\boldsymbol{\alpha})$ as

$$
\begin{aligned}
G(\boldsymbol{\alpha}) = \max\ & \sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{N}} \Big[r_i^t + \pi_i^t - (1-\rho_{i,0})\, \big(\alpha_{i,0}^{t+1} - \alpha_{i,1}^{t+1}\big) \Big]\, y_i^t + \sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{N}} \pi_i^t \sum_{\ell=1}^{\infty} q_{i,\ell}^t \\
&\qquad + \sum_{t\in\mathcal{T}\setminus\{T\}} \boldsymbol{\alpha}^{t+1} \cdot \Big(\mathbb{E}\Big\{ \boldsymbol{X}(\boldsymbol{q}^t) \Big\} - \boldsymbol{q}^{t+1} \Big) \qquad (28) \\
\text{s.t. }\ & \boldsymbol{q}^1 = \sum_{i\in\mathcal{N}} C_i\, \boldsymbol{e}_{i,0} \\
& \sum_{i\in\mathcal{N}} y_i^t + y_0^t \ = \ 1 \quad \forall\, t \in \mathcal{T} \\
& y_i^t / v_i^t \leq y_0^t / v_0^t \quad \forall\, i \in \mathcal{N},\ t \in \mathcal{T} \\
& y_i^t \geq 0 \quad \forall\, i \in \mathcal{N} \cup \{0\},\ t \in \mathcal{T},\ q_{i,\ell}^t \geq 0 \quad \forall\, i \in \mathcal{N},\ \ell \geq 0,\ t \in \mathcal{T},
\end{aligned}
$$

then the optimal objective value of problem (26) is given by $\min_{\boldsymbol{\alpha}} G(\boldsymbol{\alpha})$. Therefore, if we can show that $F(\boldsymbol{\alpha}) = G(\boldsymbol{\alpha})$ for any $\boldsymbol{\alpha}$, then it follows that the optimal objective values of problem (13) and

(26) are equal to each other, as desired. In the next lemma, we build on the work of **?** to show that we indeed have $F(\boldsymbol{\alpha}) = G(\boldsymbol{\alpha})$.

Lemma J.1 *If $\mathcal{F} = \{S : S \subseteq \mathcal{N}\}$, then we have $F(\boldsymbol{\alpha}) = G(\boldsymbol{\alpha})$.*

Proof: Given index sets $\mathcal{L}$ and $\{\mathcal{N}^\ell : \ell \in \mathcal{L}\}$, along with the constants $\{p_i^\ell : i \in \mathcal{N}_\ell,\ \ell \in \mathcal{L}\}$ and $\{\phi_i^\ell(S) : i \in \mathcal{N}^\ell,\ S \subseteq \mathcal{N}^\ell,\ \ell \in \mathcal{L}\}$, **?** consider the linear program

$$\begin{aligned} \max \quad & \sum_{\ell \in \mathcal{L}} \sum_{i \in \mathcal{N}^\ell} \sum_{S \subseteq \mathcal{N}^\ell} p_i^\ell \, \phi_i^\ell(S) \, z^\ell(S) \\ \text{s.t.} \quad & \sum_{S \subseteq \mathcal{N}^\ell} z^\ell(S) = 1 \qquad \forall \, \ell \in \mathcal{L} \\ & z^\ell(S) \geq 0 \qquad \forall \, S \subseteq N^\ell,\ \ell \in \mathcal{L}. \end{aligned} \tag{29}$$

By Theorem 3 in **?**, if $\phi_i^\ell(S)$ is of the form $\phi_i^\ell(S) = v_i^\ell / (v_0^\ell + \sum_{j \in S} v_j^\ell)$, then the problem above has the same optimal objective value as the linear program

$$\begin{aligned} \max \quad & \sum_{\ell \in \mathcal{L}} \sum_{i \in \mathcal{N}^\ell} p_i^\ell \, y_i^\ell \\ \text{s.t.} \quad & \sum_{i \in \mathcal{N}^\ell} y_i^\ell + y_0^\ell = 1 \qquad \forall \, \ell \in \mathcal{L} \\ & y_i^\ell / v_i^\ell \leq y_0^\ell / v_0^\ell \qquad \forall \, i \in \mathcal{N}^\ell,\ \ell \in \mathcal{L} \\ & y_i^\ell \geq 0 \qquad \forall \, i \in \mathcal{N}^\ell \cup \{0\},\ \ell \in \mathcal{L}. \end{aligned} \tag{30}$$

In problem (27), the decision variables $(z^t(A) : A \in \mathcal{F},\ t \in \mathcal{T})$ and $(q_{i,\ell}^t : i \in \mathcal{N},\ \ell \geq 0,\ t \in \mathcal{T})$ do not interact in the objective function or the constraints. Therefore, problem (27) decomposes into two subproblems, one subproblem involving the decision variables $(z^t(A) : A \in \mathcal{F},\ t \in \mathcal{T})$ and the other subproblem involving the decision variables $(q_{i,\ell}^t : i \in \mathcal{N},\ \ell \geq 0,\ t \in \mathcal{T})$. Similarly, problem (28) decomposes into two subproblems as well, one subproblem involving the decision variables $(y_i^t : i \in \mathcal{N} \cup \{0\},\ t \in \mathcal{T})$ and the other subproblem involving the decision variables $(q_{i,\ell}^t : i \in \mathcal{N},\ \ell \geq 0,\ t \in \mathcal{T})$. For problems (27) and (28), we observe that the subproblems that involve the decision variables $(q_{i,\ell}^t : i \in \mathcal{N},\ \ell \geq 0,\ t \in \mathcal{T})$ are identical to each other. Therefore, these subproblems have the same optimal objective value. Also, as $\mathcal{F} = \{S : S \subseteq \mathcal{N}\}$, letting $\mathcal{L} = \mathcal{T}$, $\mathcal{N}^\ell = \mathcal{N}$ and $p_i^\ell = r_i^t + \pi_i^t - (1 - \rho_{i,0})\,(\alpha_{i,0}^{t+1} - \alpha_{i,1}^{t+1})$, the subproblem for problem (27) that involves the decision variables $(z^t(A) : A \in \mathcal{F},\ t \in \mathcal{T})$ has precisely the same form as problem (29), whereas the subproblem for problem (28) that involves the decision variables $(y_i^t : i \in \mathcal{N} \cup \{0\},\ t \in \mathcal{T})$ has precisely the same form as problem (30). So, by Theorem 3 in **?**, the subproblem for problem (27) that involves the decision variables $(z^t(A) : A \in \mathcal{F},\ t \in \mathcal{T})$ has the same optimal objective value as

Params.	Computation Time per Sample Path						
(α,κ)	GR	RO	BP	OS	DC	MY	IB
$(0.7,\quad 0)$	1.69	5.05	155.27	150.28	264.11	0.84	0.83
$(0.7,0.01)$	2.69	4.30	164.68	172.77	286.14	1.89	1.15
$(0.7,0.03)$	1.62	4.44	140.92	168.38	271.44	0.72	0.66
$(0.8,\quad 0)$	1.57	4.20	152.69	146.11	241.85	0.68	0.77
$(0.8,0.01)$	1.61	4.05	141.74	149.04	253.53	0.60	0.63
$(0.8,0.03)$	1.78	4.07	175.45	202.73	281.71	0.97	0.58
$(0.9,\quad 0)$	2.01	4.01	124.29	120.28	197.96	0.68	0.62
$(0.9,0.01)$	1.54	3.76	133.97	131.64	220.03	0.67	0.62
$(0.9,0.03)$	1.79	3.92	178.06	156.26	243.24	0.73	0.81
$(1.0,\quad 0)$	1.51	3.50	118.90	101.63	176.94	0.56	0.53
$(1.0,0.01)$	1.96	3.80	134.95	132.23	207.62	0.52	0.62
$(1.0,0.03)$	1.78	3.48	146.91	165.31	243.25	1.16	1.02
Average	1.79	4.05	147.32	149.72	240.65	0.84	0.74

Table EC.1 Detailed computation times.

the subproblem for problem (28) that involves the decision variables $(y_i^t : i \in \mathcal{N} \cup \{0\},\ t \in \mathcal{T})$. Thus, problems (27) and (28) have the same optimal objective value. ■

Lastly, we note that **?** give an approach to obtain an optimal solution to problem (13) by using an optimal solution to problem (26).

Appendix K: Details of Computation Times

We give the details of the computation times for the benchmarks that we use for the test problems in Table 1. The computation times that we report correspond to the average time to simulate the performance of each benchmark over one sample path. Recall that we divide the selling horizon into three equal segments and recompute the policy parameters for each benchmark at the beginning of each segment. Therefore, the computation times that we report include the time to recompute the policy parameters three times and making the assortment offer decisions for all of the customers that arrive over the selling horizon. We give our results in Table EC.1. The first column in this table labels the test problems by using (α,κ). The remaining seven columns show the computation time in seconds to simulate the performance of each benchmark over a sample path.

Appendix L: Experimentation with Inventory Balancing

In our implementation of IB in Section 6.2, we use a revenue modifier of the form $\Psi(x) = \frac{e}{e-1}(1-e^{-x})$. Following **?**, we also work with a revenue modifier of the form $\Psi(x) = x$. In addition, **?** propose a hybrid policy that judiciously picks between the assortment proposed by IB and the assortment proposed by some other secondary policy at each time period. We implemented the hybrid policy, where we use OS discussed in Section 6.2 as the secondary policy and set the revenue modifier in IB as $\Psi(x) = \frac{e}{e-1}(1-e^{-x})$. We also tried using BP discussed in Section 6.2 as the

Params. (α,κ)	Total Expected Revenue RO	IB-Exp	IB-Lin	HY	% Gain of RO over IB-Exp	IB-Lin	HY
$(0.7,\quad 0)$	4,292	4,046	4,086	4,176	5.7	4.8	2.7
$(0.7, 0.01)$	3,893	3,432	3,476	3,563	11.8	10.7	8.5
$(0.7, 0.03)$	3,854	3,168	3,204	3,274	17.8	16.9	15.0
$(0.8,\quad 0)$	3,959	3,763	3,795	3,868	5.0	4.1	2.3
$(0.8, 0.01)$	3,873	3,454	3,490	3,603	10.8	9.9	7.0
$(0.8, 0.03)$	3,882	3,318	3,320	3,369	14.5	14.5	13.2
$(0.9,\quad 0)$	3,041	2,904	2,932	2,984	4.5	3.6	1.9
$(0.9, 0.01)$	2,773	2,487	2,513	2,589	10.3	9.4	6.6
$(0.9, 0.03)$	3,040	2,603	2,637	2,670	14.4	13.3	12.2
$(1.0,\quad 0)$	2,978	2,848	2,869	2,932	4.4	3.7	1.5
$(1.0, 0.01)$	2,467	2,212	2,235	2,310	10.3	9.4	6.4
$(1.0, 0.03)$	2,946	2,513	2,552	2,598	14.7	13.4	11.8
Average					10.4	9.5	7.4

Table EC.2 Performance of inventory balancing.

secondary policy or setting the revenue modifier in IB as $\Psi(x)=x$, but we did not obtain better results. In the description of the hybrid policy in **?**, there is a parameter γ that determines the tendency of the hybrid policy to deviate from the decisions of IB and switch to the secondary policy. After some experimentation, we set $\gamma=1.5$, which is consistent with the choice of γ in **?**. We use IB-Exp to refer to IB with the revenue modifier $\Psi(x)=\frac{e}{e-1}(1-e^{-x})$, IB-Lin to refer to IB with the revenue modifier $\Psi(x)=x$ and HY to refer to the hybrid policy. We give our results in Table EC.2. The layout of this table is similar to that of Table 1. The first column labels the test problems by using (α,κ). The second through fifth columns show the total expected revenues obtained by RO, IB-Exp, IB-Lin and HY. Recall that RO refers to our rollout policy. The remaining columns show the percent gaps between the total expected revenues obtained by RO and every other benchmark. Our results indicate that IB-Lin improves upon IB-Exp by about 1% on average. The performance of HY is noticeably better than that of IB-Exp and IB-Lin, but HY still lags behind RO significantly. All of the performance gaps in Table EC.2 are statistically significant at the 95% level.

Appendix M: Computational Experiments under Geometrically Distributed Usage Durations

We give computational experiments on test problems with reusable products. In our test problems, we have six products and six customer types. The choices of the customers are governed by the multinomial logit model. We use precisely the same approach discussed in Section 6.2 to generate the one-time upfront fees $\{r_i^{t,j} : i \in \mathcal{N},\ j \in \mathcal{M},\ t \in \mathcal{T}\}$, the parameters of the multinomial logit model $\{v_i^j : i \in \mathcal{N} \cup \{0\},\ j \in \mathcal{M}\}$ and the customer arrival probabilities $\{p^{t,j} : j \in \mathcal{M},\ t \in \mathcal{T}\}$. We set the per-period rental fee to zero. The number of time periods in the selling horizon is $T=1,200$. The usage time of a product is geometrically distributed with mean γ, where γ is a parameter that we vary. The initial inventory of a product i is $C_i=\gamma/(2\,|\mathcal{N}|)$. With this choice of initial inventories, the total product capacity we have available for use is $\sum_{i \in \mathcal{N}} C_i = \gamma/2$ per time period. By the discussion in Section 6.2, we generate the parameters of the multinomial logit model so that if we offer all products, then a customer arriving at a time period leaves without choosing any of the products with probability 0.1. Since each customer uses a product for γ time periods on average, if we offer all products at all time periods, then the expected demand for the products is $0.9\,\gamma$ per time period. Thus, even if we offer all products at all time periods, the expected demand exceeds the expected capacity by a factor of $\frac{0.9\,\gamma}{\gamma/2} = 1.8$. We experimented with different values for the ratio between the expected demand and the capacity and our results qualitatively remained unchanged. In our approach for generating the customer arrival probabilities $\{p^{t,j} : j \in \mathcal{M},\ t \in \mathcal{T}\}$ in Section 6.2, recall that we use a parameter κ that controls the degree to which the customers of different types arrive over non-overlapping time intervals. Varying the parameters (κ, γ) over $\{0, 0.01, 0.03\} \times \{200, 400, 600, 1{,}200\}$, we obtain 12 test problems. We test the performance of GR, RO and OS, where GR is the greedy policy with respect to our linear value function approximations, RO is our rollout approach performed on the static policy, and OS is the randomized offer set policy that is based on the linear programming approximation. In Table 1, apart from DC, OS is the strongest benchmark and DC does not apply in the presence of reusable products. Therefore, we provide comparisons against OS.

We give our computational results in Table EC.3. The layout of this table is similar to that of Table 1. The first column labels the test problems by using (κ, γ). The second column shows the upper bound on the optimal total expected revenue provided by the optimal objective value of problem (13). The third through fifth columns show the total expected revenues obtained by GR, RO and OS. The last two columns show the percent gaps between the total expected revenues obtained by RO and the other two benchmarks. Our results indicate that both GR and RO provide noticeable improvements over OS, especially when the average usage duration is relatively small so that we can use a product multiple times over the selling horizon. The performance of GR is generally competitive to that of RO. In Table EC.3, the performance gaps except for those indicated with a star are statistically significant at the 95% level.

Appendix N: Data and Experimental Setup for Street Parking Pricing in the City of Seattle

As discussed in Section 6.3, we augmented the data provided by the Open Data Program in Seattle to ensure that we have an intended parking locale for each driver. In this case, each transaction record gives the start time, duration, intended locale, actual parked local, and per-hour rate for each parking event. Note that the intended and the actual parked locales may be the same. The parking duration in the data reflects the duration of time for which each driver made a payment,

Params. (κ, γ)	Upp. Bnd.	Total Expected Revenue GR	RO	OS	% Gain of RO over GR	OS
$(0, 200)$	8,787	8,248	8,404	7,608	1.85	9.46
$(0, 400)$	11,604	11,141	11,283	10,641	1.25	5.69
$(0, 600)$	12,654	12,251	12,354	11,918	0.83	3.53
$(0, 1200)$	17,012	16,708	16,647	16,412	-0.37	1.41
$(0.01, 200)$	9,414	9,049	9,011	8,345	-0.42	7.40
$(0.01, 400)$	10,444	10,128	10,120	9,684	-0.07*	4.31
$(0.01, 600)$	11,942	11,677	11,626	11,284	-0.44	2.94
$(0.01, 1200)$	16,731	16,353	16,326	16,183	-0.17*	0.87
$(0.03, 200)$	6,856	6,692	6,632	6,039	-0.90	8.95
$(0.03, 400)$	8,422	8,265	8,234	7,817	-0.38*	5.07
$(0.03, 600)$	10,252	10,061	10,067	9,668	0.06*	3.97
$(0.03, 1200)$	12,470	12,244	12,290	12,069	0.37*	1.80
Average					0.14	4.62

Table EC.3 Computational results under geometrically distributed usage durations.

but the driver may not occupy the parking space for this whole duration. Nevertheless, we assume that a driver indeed occupies the parking space for the whole duration of time for which she made a payment. Payments can be made by using a smart phone application that allows extending a parking session remotely, so we can treat the usage duration of a parking space as a random quantity not known to the system operator at the time a driver parks.

The city of Seattle imposes parking time limits that prevent drivers from creating transactions with a duration greater than the maximum time limit. Such time limits result in an abnormally large fraction of transactions with durations that are exactly equal to the time limit, which created difficulties when estimating the parking duration distributions. Thus, we eliminated the transaction records whose durations are exactly equal to the time limit. After eliminating these transactions, a negative binomial distribution with parameters (s_i, η_i) with $s_i = 2$ gave reasonable fits. After eliminating the transactions, the load in the system was small enough that taking the future driver arrivals into consideration did not make an impact and simple policies performed remarkably well. To alleviate this problem, we artificially multiplied the arrival rates estimated from the data by a constant factor and decreased the number of parking spaces by another constant factor to obtain a reasonably large load. The multipliers that we use are given in Section 6.3.

We assume that the drivers with intended locale of j arrive into the system according to a Poisson process with the arrival rate function $\{\Lambda^{\tau,j} : \tau \geq 0\}$, where the time τ is measured in seconds. Recall that each time period in our model corresponds to a time interval of 30 seconds. In this case, a driver with intended locale j arrives at time period t approximately with probability $p^{t,j} = 30 \times \Lambda^{f(t),j}$, where $f(t)$ is the time in the day corresponding to time period t in the selling horizon of our model. For estimation purposes, we assume that the arrival rate function $\{\Lambda^{\tau,j} : \tau \geq 0\}$ is constant over each 15 minute time interval. When a driver with intended locale j arrives into the system, we offer a price menu for the five locales that are closest to her intended locale. The driver makes a choice among these locales or decides to leave the system without parking. The latter decision may correspond to using a parking space that is not street parking. In our choice model, if we offer the assortment S at time period t, then a driver whose intended locale is j chooses to park in locale i with probability

$$\phi_i^{t,j}(S) = \frac{e^{\alpha^j + \beta\, \pi_{i,h}}}{1 + \sum_{(\ell,g)\in S} e^{\alpha^j + \beta\, \pi_{\ell,g}}}$$

as long as $(i,h) \in S$. The parameter β is the price sensitivity of the drivers and it is constant across all drivers. This assumption helps us keep the number of parameters that we need to estimate manageable. We estimate the parameters β, $\{\alpha^j : j \in \mathcal{M}\}$ and $\{p^{t,j} : j \in \mathcal{M},\ t \in \mathcal{T}\}$ through maximum likelihood. The likelihood function that we use for this purpose closely mirrors the one used by **?**.

As discussed in Section 3.3 of **?**, when estimating the parameters of the choice model and the arrival rates, there is a continuum of choices for the parameters that yield the same value for the likelihood function. Therefore, we fix the no-purchase probability of each driver. In particular, we focus on the time period 11AM to 4PM in our numerical study. The per-hour parking rate for each locale in the data is fixed during this time period, but each locale has a different rate. Fixing the no-purchase probability at 0.1, the no-purchase probability for a driver with intended locale j needs to satisfy $1/(1+\sum_{(\ell,g)\in S^j} e^{\alpha^j+\beta\,\pi_{\ell,g}}) = 0.1$, where S^j is the set of locale and rate combinations offered to a driver with intended locale j. If we fix the parameter β, then the value of the parameter α^j is fixed by the last equality. Therefore, we estimate β and $\{p^{t,j} : j \in \mathcal{M},\ t \in \mathcal{T}\}$ through maximum likelihood and determine the values of the parameters $\{\alpha^j : j \in \mathcal{M}\}$ by the last equality. We estimated the parameters of the choice model and the arrival rates by using the data from 15 weekdays of June 2017. Using the data from the remaining five days of June 2017, we checked the percent deviation in the expected number of parkings according to our demand model and the number of parkings in the data over each hourly interval in each locale. The average absolute percent deviation was 27.05%.

When we estimated the parameters of the choice model through the data, the price sensitivity parameter estimate came out to be $\beta = -0.191$ with a standard error of 0.008. Following the magnitudes of the fares in the data, we allow the price of a parking space to take values \$2, \$4 or \$6 per hour. With these settings, the price sensitivity parameter turned out to be so small that changing the price of a parking space did not make a discernible difference in the choices of the drivers. Therefore, we bumped the price sensitivity parameter to $\beta = -0.5$ in all of our computational experiments.

Appendix O: Conditions on the Rewards and Transition Dynamics

In this section, we give conditions on the rewards and transition dynamics that allow us to extend our half-approximation guarantees to somewhat more general settings. We have n products indexed by $\mathcal{N}=\{1,\dots,n\}$. For each product i, let $C_i\in\mathbb{Z}_+$ denote its initial inventory level. There are T time periods in the selling horizon indexed by $\mathcal{T}=\{1,\dots,T\}$. Each unit of product i can be in one of the $k+1$ states indexed by $\mathcal{K}=\{0,1,\dots,k\}$. We refer to state 0 as the base state. To capture the state of the system at the beginning of a generic time period, we use $q_{i,\ell}$ to denote the number of units of product i in state ℓ. Therefore, we can describe the state of the system by using the vector $\boldsymbol{q}=(q_{i,\ell}:i\in\mathcal{N},\ \ell\in\mathcal{K})$. At each time period t, we choose an action in the set $\mathcal{F}$. If we choose the action $A\in\mathcal{F}$ at time period t and we have at least one unit of product i in the base state, then we generate an immediate expected revenue of $R_i^t(A)$. Also, for each unit of product i that is in state ℓ other than the base state, we generate an immediate expected revenue of $\Pi^t_{i,\ell}$. Note that the latter expected revenue does not depend on the action we choose. Thus, if the system is in state $\boldsymbol{q}$ and we take the action A, then we obtain an immediate expected revenue of

$$\sum_{i\in\mathcal{N}}\mathbb{1}_{\{q_{i,0}\geq 1\}}\,R_i^t(A)+\sum_{i\in\mathcal{N}}\sum_{\ell=1}^{k}\Pi^t_{i,\ell}\,q_{i,\ell}.$$

Next, we consider the transition dynamics. If the state of the units of product i at the current time period is $\boldsymbol{q}_i=(q_{i,\ell}:\ell\in\mathcal{K})$ and we take action A at this time period, then the number of units of product i in state ℓ at the next time period is given by the random variable

$$Q_{i,\ell}(\boldsymbol{q}_i,A)=Y_{i,\ell}(\boldsymbol{q}_i)+\mathbb{1}_{\{q_{i,0}\geq 1\}}\,\Delta_{i,\ell}(A),$$

where $Y_{i,\ell}$ and $\Delta_{i,\ell}(A)$ are random variables. By the transition dynamics above, the number of units $Q_{i,\ell}(\boldsymbol{q}_i,A)$ of product i in state ℓ at the next time period can be decomposed into two terms. The first term $Y_{i,\ell}(\boldsymbol{q}_i)$ is independent of the action that we take. The second term $\mathbb{1}_{\{q_{i,0}\geq 1\}}\,\Delta_{i,\ell}(A)$ depends on the action that we take, but this term takes effect only if we have at least one unit of product i in the base state. Observe that the number of units $Q_{i,\ell}(\boldsymbol{q}_i,A)$ of product i in state ℓ at the next time period can depend on the full state $\boldsymbol{q}_i$ of the units of product i at the current time period. We assume that $\sum_{\ell\in\mathcal{K}}(Y_{i,\ell}(\boldsymbol{q}_i)+\mathbb{1}_{\{q_{i,0}\geq 1\}}\,\Delta_{i,\ell}(A))\leq C_i$ with probability one, so that the number of units of product i does not increase over time. Also, we naturally assume that $Q_{i,\ell}(\boldsymbol{q}_i,A)\geq 0$, so that the number of units does not turn negative.

We argue that the problem primitives described so far generalize those that we use in our problem formulation in Section 2. In Section 2, the set of states $\mathcal{K}$ corresponds to the set of possible number of time periods that a unit of a product can be in use. A unit of product i in the base state corresponds to a unit on-hand. The action corresponds to the assortment that we offer at a time period. In Section 2, if we offer the assortment A at time period t and the customer chooses product i, then we generate a one-time upfront fee of r_i^t, as long as we have a unit of product i available on-hand. Furthermore, we generate a per-period rental fee of π_i^t from this product at the time period that it was just rented. Therefore, we have $R_i^t(A) = \phi_i^t(A)\,(r_i^t + \pi_i^t)$. For each unit of product i that is in use at time period t, we generate a per-period rental fee of π_i^t. So, we have $\Pi_{i,\ell}^t = \pi_i^t$. Letting $Y_{i,\ell}(\boldsymbol{q}_i) = X_{i,\ell}(\boldsymbol{q}_i)$, where $X_{i,\ell}(\boldsymbol{q}_i)$ is as in (1), along with

$$(\Delta_{i,0}(A), \Delta_{i,1}(A)) = \begin{cases} (-1,1) & \text{with probability } \phi_i^t(A)\,(1-\rho_{i,0}), \\ (0,0) & \text{with probability } 1-\phi_i^t(A)\,(1-\rho_{i,0}), \end{cases}$$

and $\Delta_{i,\ell}(A) = 0$ for all $\ell \in \mathcal{K} \setminus \{0,1\}$, in the model in Section 2, the state of the product i at the next time period can be captured by $Q_{i,\ell}(\boldsymbol{q}_i, A) = Y_{i,\ell}(\boldsymbol{q}_i) + \mathbb{1}_{\{q_{i,0} \geq 1\}}\,\Delta_{i,\ell}(A)$. Therefore, the problem primitives in Section 2 can be captured by using the notation defined so far in this section. We make three assumptions to be able to obtain a half-approximation guarantee.

- The expectation $\mathbb{E}\{Y_{i,\ell}(\boldsymbol{q}_i)\}$ is linear in $\boldsymbol{q}_i$. In particular, there exist constants $\{\theta_{i,\ell,s} : s \in \mathcal{K}\}$ such that

$$\mathbb{E}\{Y_{i,\ell}(\boldsymbol{q}_i)\} = \sum_{s \in \mathcal{K}} \theta_{i,s,\ell}\, q_{i,s}.$$

  Intuitively, the parameter $\theta_{i,s,\ell}$ captures the "tendency" of a unit of product i in state s to end up in state ℓ at the next time period, in expectation.
- In expectation, a unit of product i in the base state has a tendency to stay in the base state. In particular, we have

$$\theta_{i,0,0} = 1 \qquad \text{and} \qquad \theta_{i,0,\ell} = 0 \quad \forall\, \ell \in \mathcal{K} \setminus \{0\}.$$

- We can round the transition-adjusted revenue from each product i up to zero without loss of generality. In particular, for any choice of constants $\{u_\ell : \ell \in \mathcal{K}\}$, we have

$$\max_{A \in \mathcal{F}} \left\{ \sum_{i \in \mathcal{N}} \left(R_i^t(A) + \sum_{\ell \in \mathcal{K}} u_\ell\, \mathbb{E}\{\Delta_{i,\ell}(A)\} \right) \right\} = \max_{A \in \mathcal{F}} \left\{ \sum_{i \in \mathcal{N}} \left[R_i^t(A) + \sum_{\ell \in \mathcal{K}} u_\ell\, \mathbb{E}\{\Delta_{i,\ell}(A)\} \right]^+ \right\}. \quad (31)$$

All of these three assumptions are satisfied by the problem primitives in Section 2.

This problem setup generalizes the one in Section 2 to some extent, especially in terms of the transition dynamics. In particular, the random variable $Y_{i,\ell}(\boldsymbol{q}_i)$ can have any distribution as long as it satisfies the first two assumptions above. Furthermore, the random variable $\Delta_{i,\ell}(A)$ can take values other than $\{-1, 0, +1\}$ as long as it satisfies (31). It may be difficult to ensure that (31) is satisfied in general, but in Section 2, we give at least one specific case for $R_i^t(A)$ and $\Delta_{i,\ell}(A)$ that ensures that (31) is indeed satisfied.

Our goal is to find a policy to choose an action at each time period so that we maximize the total expected revenue over the selling horizon. We can find the optimal policy by formulating a dynamic program that uses $\boldsymbol{q} = (q_{i,\ell} : i \in \mathcal{N},\ k \in \mathcal{K})$ as the state variable. To obtain an approximate policy, we can use linear value function approximations of the form $\hat{J}^t(\boldsymbol{q}) = \sum_{i \in \mathcal{N}} \sum_{\ell \in \mathcal{K}} \hat{\nu}_{i,\ell}^t \, q_{i,\ell}$. In this case, we can follow the approach in Section 3 with minor modifications to choose the parameters $\{\hat{\nu}_{i,\ell}^t : i \in \mathcal{N},\ \ell \in \mathcal{K},\ t \in \mathcal{T}\}$ in the value function approximations so that the greedy policy with respect to the linear value function approximations is half-approximate. Furthermore, we can follow the approach in Section 4 with minor modifications to perform rollout on a static policy, yielding a policy that is also half-approximate.

Online Appendix References