Pricing Problems under the Markov Chain Choice Model

James Dong
School of Operations Research and Information Engineering,
Cornell University, Ithaca, New York 14853, USA
jd748@cornell.edu

A. Serdar Simsek
Naveen Jindal School of Management,
The University of Texas at Dallas, Richardson, TX 75080, USA
serdar.simsek@utdallas.edu

Huseyin Topaloglu
School of Operations Research and Information Engineering,
Cornell Tech, New York, New York 10011, USA
topaloglu@orie.cornell.edu

April 23, 2018

Abstract

We consider pricing problems when customers choose under the Markov chain choice model. In this choice model, a customer arriving into the system is interested in a certain product with a certain probability. Depending on the price charged for this product, the customer decides whether to purchase the product. If the customer purchases the product, then she leaves the system. Otherwise, the customer transitions to another product or to the no purchase option with certain transition probabilities. In this way, the customer transitions between the products until she purchases a product or reaches the no purchase option. We study three fundamental pricing problems under this choice model. First, for the monopolistic pricing problem, we show how to compute the optimal prices efficiently. Second, for the competitive pricing problem, we show that a Nash equilibrium exists, prove that Nash equilibrium prices are no larger than the prices computed by a central planner controlling all prices and characterize a Nash equilibrium that Pareto dominates all other Nash equilibria. Third, for the dynamic pricing problem with a single resource, we show that the optimal prices decrease as we have more resource capacity or as we get closer to the end of the selling horizon. We also consider a deterministic approximation formulated under the assumption that the demand for each product takes on its expected value. Although the objective function and constraints in this approximation do not have explicit expressions, we develop an equivalent reformulation with explicit expressions for the objective function and constraints.

Discrete choice models have been gaining attention in the revenue management literature as they capture the demand for a particular product as a joint function of the features of all products that are made available to the customers. By using discrete choice models, we can capture the fact that increasing the price for a certain product may not only decrease the demand for this product, but may also increase the demand for other products, since the customers may substitute for the more expensive product by using less expensive alternatives. In this case, the demand for a certain product depends not only on its price, but also on the prices of all other products. Although discrete choice models allow us to construct rich models of the customer demand, if we model the customer demand by using a discrete choice model, then solving the corresponding optimization problems to find the optimal prices to charge can be challenging.

In this paper, we consider a Markov chain choice model to describe how the customers choose among the products as a function of the prices of all of the available products and we solve pricing problems under this choice model. In our Markov chain choice model, a customer arriving into the system is interested in a certain product with a certain probability. Depending on the price charged for this product, the customer decides whether to purchase the product. If the customer purchases the product, then she leaves the system. If the customer does not purchase the product, then she transitions to another product or to the no purchase option with certain transition probabilities. If the customer transitions to another product, then she decides whether to purchase the other product depending on the price of this product. In this way, the customer transitions between the products until she purchases a product or she reaches the no purchase option. We study three fundamental multi-product pricing problems when customers choose according to the Markov chain choice model. First, we study monopolistic pricing problems, where the prices for the products are controlled by a single firm. The goal is to set the prices for the products to maximize the expected profit from each customer. Second, we study competitive pricing problems with multiple firms, where each firm controls the prices for a different subset of the products. The customers choose among all products offered by all firms. The goal of each firm is to set the prices for its own products to maximize the expected profit from each customer. Third, we study dynamic pricing problems with a single resource, where the sale of each product consumes a unit of the resource. The goal is to find a policy to dynamically set the prices for the products to maximize the total expected profit over a finite selling horizon. We proceed to explaining our main findings.

Main Results and Contributions. First, we study monopolistic pricing problems, where there is a single firm that controls the prices for all of the products. Customers choose among the products according to the Markov chain choice model. Following the literature on pricing problems with multiple products, there is also a unit cost incurred when a sale of a product occurs. The goal is to set the prices for the products to maximize the expected profit obtained from each customer, where the profit is given by the difference between the revenue and the cost associated with the sold product. A standard formulation of the problem presents critical difficulties. The objective function of this formulation turns out to be nonconcave. Thus, it is not clear how to obtain a global maximizer of the expected profit function. Furthermore, if we charge certain prices for the products, then we need to solve a system of equations to compute the probability that a customer chooses each one of the products. Thus, simply computing the objective value of the formulation at particular prices requires solving a system of equations. We develop an approach to find the prices for the products that maximize the expected profit obtained from each customer. Although the objective function of the standard formulation is not concave, we show that our approach finds a global maximizer of the objective function (Theorem 3). Surprisingly, our approach requires solving a sequence of single dimensional optimization problems, which can be done efficiently. Also, we give comparative statistics for the optimal prices in the monopolistic pricing problem. In particular, we show that if the unit cost associated with a certain product increases, then the optimal price of this product increases, whereas the optimal prices of all other products decrease (Lemma 4). To interpret these results, note that if we increase the unit cost of a product, then the optimal price of this product increases to make up for the increase in the unit cost, but this increase in the optimal price results in a decrease in the expected number of customers making a purchase. To make up for the decrease in the expected number of customers making a purchase, the optimal prices of all other products decrease to generate sales.

Second, we study competitive pricing problems with multiple firms. Each firm owns a certain subset of the products and controls the prices for the products that it owns. Customers choose among all products owned by all firms. The goal of each firm is to set the prices for the products that it owns to maximize the expected profit it obtains from each customer. We show that a Nash equilibrium exists (Theorem 6). Our existence proof uses first principles and also allows us to derive structural properties. In particular, we show that the prices in any Nash equilibrium are no larger than those charge by a central planner, who maximizes the expected profit obtained from each customer (Theorem 7). Thus, competition between the firms tends to lower the prices. Also, we show that there exists a Pareto dominant equilibrium, where the expected profit of each firm is at least as large as its expected profit in any other Nash equilibria (Theorem 8). Thus, the Pareto dominant equilibrium is simultaneously preferred by all firms. We show that the prices at the Pareto dominant equilibrium decrease as the control of the products are split among a larger number of firms and the intensity of competition increases (Theorem 9). Lastly, we show that if each firm owns a single product, then all prices in the Pareto dominant equilibrium increase when the unit cost of any product increases (Lemma 10).

Third, we study dynamic pricing problems with a single resource. Customers arrive randomly over time and choose among the products according to the Markov chain choice model. There is limited inventory of the resource. The sale of a product consumes a unit of the resource. The goal is to find a policy to dynamically decide what prices to charge for the products to maximize the total expected profit over a finite selling horizon. We show that if we have more units of the resource at a particular time period, then the optimal prices for the products decrease. Also, if we get closer to the end of the selling horizon with a certain inventory of the resource, then the optimal prices for the products decrease as well (Lemma 12). Thus, if we have more units of the resource or we get closer to the end of the selling horizon, then the pressure to liquidate the resource inventory takes precedence and we charge lower prices. Furthermore, we consider a deterministic approximation formulated under the assumption that the demands for the products take on their expected values. As we discuss in our literature review, there is work on constructing approximate policies from such a deterministic approximation. Under the Markov chain choice model, the deterministic approximation has a nonconcave objective function and a nonconvex feasible set. Also, the objective function and the constraints do not have closed form expressions. We give an equivalent reformulation for the deterministic approximation with closed form expressions for the objective function and the constraints (Theorem 13). The feasible set for the equivalent reformulation is a polytope. We characterize when the objective function is concave.

Beside the three classes of pricing problems above, our formulation of the Markov chain choice model makes useful contributions. The Markov chain choice model was proposed by Blanchet et al. (2016) under the assumption that the prices for the products are fixed. It is not a priori clear how to use this choice model when the choice process of the customers reacts to the prices. One can make the transition probabilities a function of the prices, but this approach makes the corresponding pricing problems intractable. In our approach, the transition probabilities are independent of the prices, but when a customer visits a certain product, she decides whether to purchase this product based on the price of the product. Although the transition probabilities are independent of the prices, the ultimate purchase probability of a product depends jointly on all prices. We show that our extension of the Markov chain choice model is compatible with the random utility maximization principle, where each customer associates random utilities with all alternatives, choosing the alternative with the largest utility (Theorem 1). We show that we can calibrate the Markov chain choice model so that the purchase probabilities under this choice model become identical to those under generalized attraction models, where the purchase probability of a product can be written as the ratio of the attraction of the product to the total attraction of all alternatives and the attraction of the no purchase option can increase as we charge larger prices (Lemma 2). Thus, although the derivation of the Markov chain choice model is different from that of the generalized attraction model, their choice probabilities can be made compatible. The multinomial logit model is a subclass of generalized attraction models. We give a numerical study to demonstrate that the flexibility provided by the Markov chain choice model can be beneficial and this choice model can do a better job of predicting the customer purchases when compared with the multinomial logit model.

Literature Review. The Markov chain choice model is proposed in Blanchet et al. (2016). The authors show how to solve assortment optimization problems under this choice model. In the assortment optimization setting, the prices for the products are fixed and the goal is to decide which assortment of products to offer to customers to maximize the expected profit from each customer. Feldman and Topaloglu (2017) consider various assortment optimization problems under the Markov chain choice model and they characterize the structure of the optimal assortments. Desir et al. (2015) solve assortment optimization problems under the Markov chain choice model when there is a constraint that limits the capacity consumption of the offered products. All of the work that has been done so far under the Markov chain choice model is under the assumption that the prices for the products are fixed and there is a single firm that chooses the assortment of products to offer to the customers.

There is work on estimating the parameters of the Markov chain choice model by using maximum likelihood estimation. Feldman and Topaloglu (2017) give an efficient approach to compute the gradient of the likelihood function with respect to the parameters of the Markov chain choice model, allowing the use of a gradient ascent algorithm to find a local maximizer of the likelihood function. Their computational experiments demonstrate that the Markov chain choice model can provide noticeable benefits in predicting the purchase behavior of the customers, when compared with the multinomial logit model. Simsek and Topaloglu (2017) give an expectation-maximization algorithm to find a stationary point of the likelihood function. Their algorithm requires only solving systems of linear equations. Their computational experiments are partly based on real data and they also demonstrate the potential benefits from using the Markov chain choice model. Wang and Yuan (2017) compare the performance of the expectation-maximization and gradient ascent algorithms to estimate the parameters of the Markov chain choice model. Thus, there is recent work indicating that it is possible to calibrate the Markov chain choice model in a computationally tractable fashion so that we can predict the customer purchase behavior more accurately when compared with simpler models, such as the multinomial logit model. All of this work is under the assumption that the prices are fixed, but our numerical study indicates that we can calibrate the Markov chain choice model to obtain similar benefits when the prices are adjustable.

Markov chains are also used to describe the choice process in other settings. Jeuland (1979) gives a model to capture the brand loyalty behavior. In each purchase, the customer either stays with the brand in her last purchase or chooses a brand according to a fixed probability distribution. Givon (1984) uses a similar model to capture the brand variety seeking behavior. In each purchase, the customer either switches to a new brand uniformly over all brands that are not in her last purchase or chooses a brand according to a fixed probability distribution. Bawa (1990) postulates that the utility of a customer from the brand that is in her last purchase is a quadratic function of the number of successive times she purchased this brand. The utilities of the brands that are not in her last purchase are fixed. In her next purchase, the customer chooses according to a multinomial logit model among all brands. In the model used by Gilboa and Pazgal (1995), the customer keeps preference rankings of all brands. In each purchase, the customer purchases the most preferred brand and updates only the preference ranking of the purchased brand. Craswell et al. (2008) propose the cascade model to capture the choice process within search engine results. In the cascade model, a user scans the results starting from the top one. With a certain probability, she either clicks this result or moves on to the next one. Guo et al. (2009) extend the cascade model to allow multiple clicks during a search session. The papers discussed in this paragraph focus on parameter estimation, but not on optimizing product prices or rankings

Considering pricing problems under choice models other than the Markov chain choice model, Hanson and Martin (1996) work with the multinomial logit model and observe that the expected profit is not a concave function of the prices. Gallego et al. (2006) study competitive pricing problems under the multinomial logit model. They show that there exists a unique and stable equilibrium. Song and Xue (2007) consider pricing problems under the multinomial logit model and show that the expected profit is concave in the market shares of the products. They solve the pricing problem by using the market shares of the products as decision variables. Chen and Hausman (2000) and Wang (2012) study joint assortment and pricing problems under the multinomial logit model, where the set of offered products, as well as their corresponding prices are decision variables. Keller et al. (2014) consider pricing problems when there are constraints on the expected sales of a product. Gallego et al. (2015) propose generalized attraction models, which can be viewed as a generalization of the multinomial logit model, where the attraction of the no purchase option increases as we offer a more restricted subset of products.

Li and Huh (2011) study pricing problems under the nested logit model when the products in the same nest have the same price sensitivity. The authors show that the pricing problem can be formulated as a convex program. They consider the competitive pricing problem, as well as the monopolistic one. Gallego and Wang (2014) relax the assumption that the products in the same nest have the same price sensitivity and show that the pricing problem can be solved as a single dimensional optimization problem. They make extensions to the case where there are multiple firms, each controlling the prices for the products in a different nest. Gallego and Topaloglu (2014) study pricing problems under the nested logit model when the price for a product is chosen within a finite set of possible prices and formulate the problem as a linear program. Li et al. (2015) study pricing problems under the nested logit model with multiple levels of nests.

Our study of dynamic pricing problems with a single resource is motivated by Maglaras and Meissner (2006), where the authors draw parallels between control mechanisms that are based on adjusting the prices of the products or the set of available products. Gallego and van Ryzin (1994) use a deterministic approximation to develop approximate policies for dynamic pricing problems with a single resource. Gallego and van Ryzin (1997) extend this work to dynamic pricing problems over a network of resources, where the sale of a product consumes a combination of resources. Since dynamic programming formulations of capacity control problems over a network of resources involve high dimensional state variables, it is common to formulate deterministic approximations by assuming that the demands for the products take on their expected values. Beside Gallego and van Ryzin (1997), such approximations appear in Talluri and van Ryzin (1998), Gallego et al. (2004), Liu and van Ryzin (2008), Vossen and Zhang (2015) and Zhang and Lu (2013).

Organization. In Section 1, we describe the Markov chain choice model. We show that this choice model is compatible with the random utility maximization principle and we can calibrate its parameters to ensure that the choice probabilities under the Markov chain choice model are identical to those under the generalized attraction model. In Section 2, we focus on monopolistic pricing. In Section 3, we focus on competitive pricing. In Section 4, we focus on dynamic pricing with a single resource. In Section 5, we give our numerical study. In Section 6, we conclude.

1 Markov Chain Choice Model

There are n products indexed by $N = \{1, \dots, n\}$. We use p_i to denote the price charged for product i. The set of feasible prices for product i is $\mathcal{P}_i = [L_i, U_i]$. With probability λ_i, a customer arriving into the system visits product i. A customer visiting product i purchases product i with probability $\theta_i(p_i)$, where the function $\theta_i(\cdot) : \mathcal{P}_i \to [0, 1]$ maps the price of product i to the probability that a customer visiting product i purchases this product. With probability $1 - \theta_i(p_i)$, a customer visiting product i does not purchase product i, in which case, she transitions from product i to product j with probability ρ_{ij} and visits product j. If a customer visiting product i does not purchase this product, then she transitions to the no purchase option and leaves the system without making a purchase with probability $1 - \sum_{j \in N} \rho_{ij}$. In this way, the customer transitions between the different products until she purchases one of the products or decides to leave the system without a purchase. We proceed to computing the probability that a customer purchases a certain product as a function of the prices $\boldsymbol{p} = \{p_i : i \in N\}$ charged for the products. We use $v_i(\boldsymbol{p})$ to denote the expected number of times that a customer visits product i when the prices charged for the products are given by $\boldsymbol{p}$. We can compute $\{v_i(\boldsymbol{p}) : i \in N\}$ by solving the system of equations

$$v_i(\boldsymbol{p}) = \lambda_i + \sum_{j \in N} \rho_{ji} \, (1 - \theta_j(p_j)) \, v_j(\boldsymbol{p}) \quad \forall \, i \in N. \tag{1}$$

We interpret the system of equations in (1) as follows. By definition, the term $v_i(\boldsymbol{p}_i)$ on the left side corresponds to the expected number of times that a customer visits product i. Each customer arriving into the system visits product i with probability λ_i. Thus, the expected number of times that a customer visits product i on arrival is λ_i, yielding the term λ_i on the right side. The expected number of times that a customer visits some product j is $v_j(\boldsymbol{p})$, but each time the customer visits some product j, she does not purchase product j with probability $1 - \theta_j(p_j)$ and transitions from product j to product i with probability ρ_{ji}. In this case, she ends up visiting product i, yielding the term $\rho_{ji} \, (1 - \theta_j(p_j)) \, v_j(\boldsymbol{p})$ on the right side. We can solve the n linear equations in (1) for the n unknowns in $\{v_i(\boldsymbol{p}) : i \in N\}$ to compute the expected number of times a customer visits each product. Each time a customer visits product i, she purchases this product with probability $\theta_i(p_i)$. Therefore, if the prices charged for the products are given by $\boldsymbol{p}$, then a customer purchases product i with probability $\theta_i(p_i) \, v_i(\boldsymbol{p})$. Note that the parameters of the Markov chain choice model are $\{\lambda_i : i \in N\}$, $\{\theta_i(\cdot) : i \in N\}$ and $\{\rho_{ij} : i, j \in N\}$. Next, we describe the assumptions that we make regarding the parameters of the Markov chain choice model.

Regarding $\{\lambda_i : i \in N\}$, we assume that $\lambda_i > 0$ for all $i \in N$, in which case, by (1), we get $v_i(\boldsymbol{p}) > 0$ for all $i \in N$. Our results in the paper hold with minor modifications when $\lambda_i = 0$ for some $i \in N$, but assuming that $\lambda_i > 0$ for all $i \in N$ allows us to avoid degenerate cases. We allow having $\sum_{i \in N} \lambda_i < 1$, in which case, we have no customer arrival with probability $1 - \sum_{i \in N} \lambda_i$. Regarding $\{\theta_i(\cdot) : i \in N\}$, we assume that $\theta_i(\cdot)$ is differentiable and strictly decreasing, so that the probability that a customer visiting product i purchases this product is strictly decreasing in the price of product i. Also, we assume that $\theta_i(p_i) \, (p_i - x)$ is strictly quasiconcave in $p_i \in \mathcal{P}_i$ for any $x \in \Re$, which ensures that the optimization problems that we solve have unique optimal solutions. Lastly, we assume that $\lim_{p_i \to U_i} \theta_i(p_i) = 0$ and $\lim_{p_i \to U_i} p_i \, \theta_i(p_i) = 0$ for all $i \in N$. Thus, there exists a large enough price that yields a purchase probability of zero for each product and we cannot obtain arbitrarily large expected profit with arbitrarily large prices. There are several choices of $\theta_i(\cdot)$ and $\mathcal{P}_i$ that satisfy this assumption, including $\theta_i(p_i) = \mathrm{e}^{-\beta_i \, p_i}$ with $\mathcal{P}_i = [0, \infty)$ and $\theta_i(p_i) = 1 - \beta_i \, p_i$ with $\mathcal{P}_i = [0, 1/\beta_i]$, where we have $\beta_i > 0$. Regarding $\{\rho_{ij} : i, j \in N\}$, we assume that $\sum_{j \in N} \rho_{ij} < 1$ for all $i \in N$, in which case, if a customer visiting product i decides not to purchase this product, then she transitions to the no purchase option with the strictly positive probability $1 - \sum_{j \in N} \rho_{ij}$. Therefore, the choice process is always guaranteed to terminate.

Since we use the system of equations in (1) to compute the choice probabilities, a natural question is whether this system of equations always has a unique solution with nonnegative entries. Using the matrix $\mathbf{Q}(\boldsymbol{p}) = \{\rho_{ij} \, (1 - \theta_i(p_i)) : i, j \in N\}$ and the vectors $\boldsymbol{v}(\boldsymbol{p}) = \{v_i(\boldsymbol{p}) : i \in N\}$ and $\boldsymbol{\lambda} = \{\lambda_i : i \in N\}$, we write (1) equivalently as $\boldsymbol{v}(\boldsymbol{p}) = \boldsymbol{\lambda} + \mathbf{Q}(\boldsymbol{p})^\top \, \boldsymbol{v}(\boldsymbol{p})$. Since $\sum_{j \in N} \rho_{ij} < 1$, the sum of the entries in each row of $\mathbf{Q}(\boldsymbol{p})$ is strictly less than one, in which case, using $\mathbf{I} \in \Re^{n \times n}$ to denote the identity matrix, by Theorem 3.2.c in Puterman (1994), $\mathbf{I} - \mathbf{Q}(\boldsymbol{p})$ is invertible and its inverse has nonnegative entries. Thus, $(\mathbf{I} - \mathbf{Q}(\boldsymbol{p}))^\top = \mathbf{I} - \mathbf{Q}(\boldsymbol{p})^\top$ is invertible and its inverse has nonnegative entries as well, which implies that the system of equations in (1) always has a unique solution given by $\boldsymbol{v}(\boldsymbol{p}) = (\mathbf{I} - \mathbf{Q}^\top(\boldsymbol{p}))^{-1} \, \boldsymbol{\lambda}$ and this solution has nonnegative entries.

Random Utility Maximization and Relationship to Other Choice Models. A standard way to construct choice models is based on the random utility maximization principle. Under this principle, a customer associates random utilities with all products and the no purchase option. The distribution of the utility for each product depends on its price. The customer chooses the alternative that provides the largest utility. We use the random variable $U_i(p_i)$ to denote the utility of product i given that we charge the price p_i for this product. For notational uniformity, we use the random variable $U_0(p_0)$ to denote the utility of the no purchase option, but the no purchase option does not have a price under our control. Under the random utility maximization principle, if we charge the prices $\boldsymbol{p}$ for the products, then a customer chooses product i with probability $\mathbb{P}\{U_i(p_i) = \max_{j \in N \cup \{0\}} U_j(p_j)\}$, where we assume that there is always a unique maximum element of the set $\{U_j(p_j) : j \in N \cup \{0\}\}$. In the next theorem, we show that the choice probabilities under the Markov chain choice model can be captured by using appropriately defined utility random variables $\{U_i(\cdot) : i \in N \cup \{0\}\}$. The proofs of all results in the paper are in Appendix A.

Theorem 1 *For any Markov chain choice model, there exist utility random variables $\{U_i(\cdot) : i \in N \cup \{0\}\}$ such that if we charge the prices $\boldsymbol{p}$, then the purchase probability of product i under the Markov chain choice model can be written as $\mathbb{P}\{U_i(p_i) = \max_{j \in N \cup \{0\}} U_j(p_j)\}$.*

The proof of Theorem 1 is constructive, where we explicitly construct the utility random variables by using the first visit times in a Markov chain with the initial distribution $\{\lambda_i : i \in N\}$ and the transition probabilities $\{\rho_{ij} : i, j \in N\}$. The utility $U_i(p_i)$ of product i in our construction depends only on the price of product i, but not on the prices of the other products. It is not a priori clear that the Markov chain choice model can be represented by using one utility random variable for each alternative, whose distribution depends on the price of only that alternative. An inspection of the proof of Theorem 1 also indicates that the utility $U_i(p_i)$ of product i satisfies $U_i(p_i^0) \geq U_i(p_i^+)$ with probability one for any $p_i^0 \leq p_i^+$. Therefore, the utility random variable for each product satisfies the intuitive property that it increases as the price for the product decreases. Lastly, viewing $\mathbb{E}\{U_i(p_i)\}$ as the deterministic component and $U_i(p_i) - \mathbb{E}\{U_i(p_i)\}$ as the random shock, we can view the utility of product i as the sum of a deterministic component and a random shock. Here, the distribution of the random shock depends on the price as well.

Under certain choices of the parameters $\{\lambda_i : i \in N\}$, $\{\theta_i(\cdot) : i \in N\}$ and $\{\rho_{ij} : i, j \in N\}$, we can relate the Markov chain choice model to other choice models. Using the vector $\boldsymbol{\lambda} = \{\lambda_i : i \in N\}$ and the matrix $\mathbf{R} = \{\rho_{ij} : i, j \in N\}$, we consider the case where $\mathbf{R}$ is a rank one matrix of the form $\mathbf{R} = \boldsymbol{\beta}\, \boldsymbol{\lambda}^\top$ for some vector $\boldsymbol{\beta} = \{\beta_i : i \in N\}$. So, we have $\rho_{ij} = \beta_i\, \lambda_j$. Since we need to have $\sum_{j \in N} \rho_{ij} < 1$, we assume that $\beta_i < 1$ for all $i \in N$, in which case, we get $\sum_{j \in N} \rho_{ij} = \beta_i \sum_{j \in N} \lambda_j < 1$, where the inequality uses the fact that $\sum_{i \in N} \lambda_i \leq 1$. In this case, the parameters of the Markov chain choice model are $\{\lambda_i : i \in N\}$, $\{\theta_i(\cdot) : i \in N\}$ and $\{\beta_i : i \in N\}$. We refer to this Markov chain choice model as the rank one Markov chain choice model. In the next lemma, we give a closed form expression for the purchase probabilities under the rank one Markov chain choice model.

Lemma 2 *If we charge the prices $\boldsymbol{p}$, then the purchase probability of product i under the rank one Markov chain choice model is given by*

$$\frac{\lambda_i\, \theta_i(p_i)}{1 - \sum_{j \in N} \lambda_j + \sum_{j \in N} \lambda_j\, (1 - \beta_j)\, (1 - \theta_j(p_j)) + \sum_{j \in N} \lambda_j\, \theta_j(p_j)}.$$

The choice probability in the lemma above has an intuitive interpretation. We view $\lambda_i\, \theta_i(p_i)$ as the attractiveness of product i. Since $\theta_i(\cdot)$ is decreasing, increasing the price of product i decreases its attractiveness. We view $1 - \sum_{j \in N} \lambda_j + \sum_{j \in N} \lambda_j\, (1 - \beta_j)\, (1 - \theta_j(p_i))$ as the attractiveness of the no purchase option. Noting that $1 - \theta_j(\cdot)$ is increasing, increasing the price of a product increases the attractiveness of the no purchase option. The parameter $1 - \beta_j$ captures how much increasing the price of product j increases the attractiveness of the no purchase option. If $\beta_j = 1$, then the attractiveness of the no purchase option does not increase with an increase in the price of product j. So, the purchase probability of product i in the lemma is given by the ratio of the attractiveness of product i to the total attractiveness of all alternatives, including the no purchase option.

Gallego et al. (2015) develop the generalized attraction model when the prices of the products are fixed. In their generalized attraction model, the attractiveness of the no purchase option increases as some products are not offered to the customers. Lemma 2, in essence, gives a generalized attraction model when the prices of the products are adjustable. For fixed parameters $\{\mu_i : i \in N\}$ and $\{\alpha_i : i \in N\}$ with $\alpha_i > 0$, if we set $\lambda_i = \mathrm{e}^{\mu_i} / (1 + \sum_{j \in N} \mathrm{e}^{\mu_j})$, $\theta_i(p_i) = \mathrm{e}^{-\alpha_i\, p_i}$ and $\beta_i = 1$, then the purchase probability in Lemma 2 becomes $e^{\mu_i - \alpha_i p_i}/(1 + \sum_{j \in N} e^{\mu_j - \alpha_j p_j})$, which is the purchase probability under the multinomial logit model. Therefore, we can set the parameters of the Markov chain choice model so that the purchase probabilities under the Markov chain choice model become identical to those under the multinomial logit model. Blanchet et al. (2016) show a similar result, but they consider the case where the prices of the products are fixed.

We should not view the Markov chain choice model as a faithful model of the mental thought process of the customers when they purchase a product. In other words, by using the Markov chain choice model, we do not insist that the customers transition from one product to another in their minds until they make a choice. In the same vein, the cascade model captures the click process of a user by using transitions from a higher ranked document to a lower ranked one, but the users do not necessarily click on the search engine results by going through such a thought process. Also, we did not construct the Markov chain choice model by using the random utility maximization principle, but after the construction, we can establish its compatibility with the random utility maximization principle. This principle ensures that if an individual prefers option 1 to option 2 and option 2 to option 3, then this individual prefers option 1 to option 3. So, individuals are consistent. Since the utilities are random, different individuals may have different preference orders.

2 Monopolistic Pricing

We consider the monopolistic pricing problem, where the prices of all products are controlled by a single firm and the goal is to set the prices for the products to maximize the expected profit from each customer. Similar to our notation in the previous section, we index the products by $N = \{1, \ldots, n\}$. We use $p_i \in \mathcal{P}_i$ to denote the price charged for product i and c_i to denote the unit cost incurred for the sale of product i. Therefore, if we sell one unit of product i, then we obtain a profit of $p_i - c_i$. It is standard to include unit costs in multi-product pricing problems; see Gallego and Wang (2014). Customers choose among the products according to the Markov chain choice model. In other words, if the prices charged for the products are given by $\boldsymbol{p} = \{p_i : i \in N\}$, then a customer purchases product i with probability $\theta_i(p_i)\, v_i(\boldsymbol{p})$, where $\{v_i(\boldsymbol{p}) : i \in N\}$ is the solution to the system of equations in (1). The goal is to set the prices for the products to maximize the expected profit obtained from each customer, yielding the optimization problem

$$\max_{\boldsymbol{p} \in \times_{i \in N} \mathcal{P}_i} \left\{ \sum_{i \in N} \theta_i(p_i)\, v_i(\boldsymbol{p})\, (p_i - c_i) \right\}. \tag{2}$$

We can come up with counterexamples to show that the objective function of problem (2) is not necessarily concave in the prices. Also, we do not have an explicit expression for $v_i(\boldsymbol{p})$. Thus, maximizing the objective function of problem (2) directly can be difficult.

To obtain an optimal solution to problem (2), we follow an alternative approach based on dynamic programming. We use $\hat{r}_i$ to denote the optimal expected profit obtained from a customer currently visiting product i. If we charge the price p_i for product i, then a customer visiting product i purchases this product with probability $\theta_i(p_i)$, in which case, we obtain an optimal expected profit of simply $p_i - c_i$. On the other hand, if we charge the price p_i for product i, then a customer visiting product i does not purchase this product with probability $1 - \theta_i(p_i)$ and she transitions from product i to product j with probability ρ_{ij}, in which case, we obtain an optimal expected profit of $\hat{r}_j$. Therefore, if we charge the price p_i for product i, then we obtain an optimal expected profit of $\theta_i(p_i)\,(p_i - c_i) + (1 - \theta_i(p_i)) \sum_{j \in N} \rho_{ij}\, \hat{r}_j$ from a customer currently visiting product i. The preceding discussion intuitively suggests that if we use $\hat{r}_i$ to denote the optimal expected profit obtained from a customer visiting product i, then $\hat{\boldsymbol{r}} = \{\hat{r}_i : i \in N\}$ should satisfy $\hat{r}_i = \max_{p_i \in \mathcal{P}_i} \{\theta_i(p_i)\,(p_i - c_i) + (1 - \theta_i(p_i)) \sum_{j \in N} \rho_{ij}\, \hat{r}_j\}$ for all $i \in N$. To write the last equality succinctly, for all $i \in N$, we define the operator $f_i(\cdot) : \Re^n \to \Re$ as

$$f_i(\boldsymbol{r}) = \max_{p_i \in \mathcal{P}_i} \left\{ \theta_i(p_i)\,(p_i - c_i) + (1 - \theta_i(p_i)) \sum_{j \in N} \rho_{ij}\, r_j \right\}, \tag{3}$$

in which case, $\hat{\boldsymbol{r}}$ should satisfy $\hat{r}_i = f_i(\hat{\boldsymbol{r}})$ for all $i \in N$. For any two vectors $\boldsymbol{r} = \{r_i : i \in N\}$ and $\boldsymbol{q} = \{q_i : i \in N\}$, in Appendix B, we establish that $|f_i(\boldsymbol{r}) - f_i(\boldsymbol{q})| \leq \max_{i \in N}\{\sum_{j \in N} \rho_{ij}\} \|\boldsymbol{r} - \boldsymbol{q}\|$ for all $i \in N$, where we define the norm $\| \cdot \|$ as $\|\boldsymbol{r}\| = \max_{i \in N}\{|r_i|\}$. Since $\sum_{j \in N} \rho_{ij} < 1$ for all $i \in N$, the last inequality implies that the operator $\{f_i(\cdot) : i \in N\} : \Re^n \to \Re^n$ is a contraction with respect to the norm $\| \cdot \|$, in which case, by Theorem 6.2.3.a in Puterman (1994), there exists $\hat{\boldsymbol{r}} = \{\hat{r}_i : i \in N\}$ that satisfies $\hat{r}_i = f_i(\hat{\boldsymbol{r}})$ for all $i \in N$. On the surface, there is no immediate relationship between problem (2) and the operator $f_i(\cdot)$ in (3). For example, the choice probability $\theta_i(p_i)\, v_i(\boldsymbol{p})$ under the prices $\boldsymbol{p}$ explicitly appears in problem (2) but not in the definition of the operator $f_i(\cdot)$ in (3). In the next theorem, we formally show that we can obtain an optimal solution to problem (2) by using $\hat{\boldsymbol{r}} = \{\hat{r}_i : i \in N\}$ that satisfies $\hat{r}_i = f_i(\hat{\boldsymbol{r}})$ for all $i \in N$.

Theorem 3 *The prices $\hat{\boldsymbol{p}} = \{\hat{p}_i : i \in N\}$ are an optimal solution to problem (2) if and only if the price $\hat{p}_i$ is an optimal solution to the problem*

$$\max_{p_i \in \mathcal{P}_i} \left\{ \theta_i(p_i)\,(p_i - c_i) + (1 - \theta_i(p_i)) \sum_{j \in N} \rho_{ij}\, \hat{r}_j \right\} \tag{4}$$

for all $i \in N$, where $\hat{\boldsymbol{r}} = \{\hat{r}_i : i \in N\}$ satisfies $\hat{r}_i = f_i(\hat{\boldsymbol{r}})$ for all $i \in N$.

Assuming that $\hat{\boldsymbol{r}}$ satisfies $\hat{r}_i = f_i(\hat{\boldsymbol{r}})$ for all $i \in N$, if we let $\hat{p}_i$ be an optimal solution to problem (4) for all $i \in N$, then by Theorem 3, $\hat{\boldsymbol{p}} = \{\hat{p}_i : i \in N\}$ is an optimal solution to problem (2). Since we assume that $\theta_i(p_i)\,(p_i - x)$ is quasiconcave in p_i for any $x \in \Re$, the objective function of problem (4) is quasiconcave, so that we can obtain an optimal solution to problem (4) by checking its first order condition. Therefore, if we can find $\hat{\boldsymbol{r}}$ satisfying $\hat{r}_i = f_i(\hat{\boldsymbol{r}})$ for all $i \in N$, then we can efficiently obtain an optimal solution to problem (2) by using problem (4). To find $\hat{\boldsymbol{r}}$ satisfying $\hat{r}_i = f_i(\hat{\boldsymbol{r}})$ for all $i \in N$, we generate the sequence $\{\boldsymbol{r}(t) : t \in \mathbb{N}\}$ by initializing $\boldsymbol{r}(0) \in \Re^n$ arbitrarily and using the relationship $r_i(t+1) = f_i(\boldsymbol{r}(t))$ for all $i \in N$. Since the operator $\{f_i(\cdot) : i \in N\} : \Re^n \to \Re^n$

is a contraction, Theorem 6.2.3.b in Puterman (1994) implies that the sequence $\{\boldsymbol{r}(t) : t \in \mathbb{N}\}$ converges to $\hat{\boldsymbol{r}}$ that satisfies $\hat{r}_i = f_i(\hat{\boldsymbol{r}})$ for all $i \in N$.

Using problem (4), we can also characterize how the optimal solution to problem (2) changes as a function of the unit costs. As a function of the unit costs $\boldsymbol{c} = \{c_i : i \in N\}$, we use $\hat{\boldsymbol{p}}(\boldsymbol{c})$ to denote an optimal solution to problem (2). We let $\boldsymbol{e} \in \Re^n$ be the vector of all ones and $\boldsymbol{e}_i \in \Re^n$ be the unit vector with a one for product i. In the next lemma, we show that if the unit cost of a product increases, then its optimal price increases, but the optimal prices of the other products decrease. If all unit costs increase by the same amount, then all optimal prices increase.

Lemma 4 *For any $\epsilon > 0$, we have $\hat{p}_i(\boldsymbol{c} + \epsilon\, \boldsymbol{e}_i) \geq \hat{p}_i(\boldsymbol{c})$, $\hat{p}_j(\boldsymbol{c} + \epsilon\, \boldsymbol{e}_i) \leq \hat{p}_j(\boldsymbol{e})$ for all $j \in N \setminus \{i\}$ and $\hat{p}_j(\boldsymbol{c} + \epsilon\, \boldsymbol{e}) \geq \hat{p}_j(\boldsymbol{c})$ for all $j \in N$.*

To provide some intuition for the result in Lemma 4, if we increase the unit cost of product i, then we increase the optimal price for product i to make up for the increase in the unit cost, which results in a decrease in the expected number of customers making a purchase. To make up for this decrease, we expect a decrease in the optimal prices of the other products. The lemma above may be of independent interest, but we also use this lemma to develop structural properties for the optimal policy in the dynamic pricing problem with a single resource.

3 Competitive Pricing

In this section, we consider the competitive pricing problem, where there are multiple firms and each firm sets the prices of its products to maximize its own expected profit.

3.1 Best Response

There are multiple firms. Different firms own different partitions of the products. The price of a product is controlled by the firm that owns the product. Customers choose among all of the products according to the Markov chain choice model. If a customer purchases a product, then the firm that owns the product obtains the profit. The goal of each firm is to set the prices for the products that it owns to maximize the expected profit that it obtains from each customer. We pursue the following outline. In this section, we show that we can use a dynamic programming idea to find the best response of each firm to the others. In Section 3.2, we show that there is a Nash equilibrium for competitive pricing. In Section 3.3, we give structural properties for the Nash equilibrium, where we compare the prices in a Nash equilibrium under different levels of competition in a sense that we make precise. In Section 3.4, we give a computational method to check the uniqueness of the Nash equilibrium for a particular problem instance.

Our notation is similar to the one used in the previous two sections, but we introduce some new notation to capture the products owned by each firm. We index the products by $N = \{1, \ldots, n\}$

and the firms by $M = \{1, \ldots, m\}$. The set of products owned by firm k is N^k, where we have $\cup_{k \in M} N^k = N$. Since different firms own different partitions of the products, we have $N^k \cap N^l = \varnothing$ for $k \neq l$. Letting $p_i \in \mathcal{P}_i$ be the price charged for product i, the prices charged for the products owned by firm k are given by $\boldsymbol{p}^k = \{p_i : i \in N^k\}$, whereas the prices charged for the products owned by firms other than firm k are given by $\boldsymbol{p}^{-k} = \{p_i : i \notin N^k\}$. Since the customers choose among all of the products according to the Markov chain choice model, if firm k charges the prices $\boldsymbol{p}^k = \{p_i : i \in N^k\}$ for the products that it owns and the other firms charge the prices $\hat{\boldsymbol{p}}^{-k} = \{\hat{p}_i : i \notin N^k\}$ for the products that they own, then a customer purchases product $i \in N^k$ with probability $\theta_i(p_i)\, v_i(\boldsymbol{p}^k, \hat{\boldsymbol{p}}^{-k})$, where $v_i(\boldsymbol{p}^k, \hat{\boldsymbol{p}}^{-k})$ is the expected number of times a customer visits product i when the prices charged for all of the products are given by $(\boldsymbol{p}^k, \hat{\boldsymbol{p}}^{-k})$. In particular, $\{v_i(\boldsymbol{p}^k, \hat{\boldsymbol{p}}^{-k}) : i \in N^k,\ k \in M\}$ satisfies a slightly modified version of the system of equations in (1) given by $v_i(\boldsymbol{p}^k, \hat{\boldsymbol{p}}^{-k}) = \lambda_i + \sum_{j \in N^k} \rho_{ji}\, (1 - \theta_j(p_j))\, v_j(\boldsymbol{p}^k, \hat{\boldsymbol{p}}^{-k}) + \sum_{j \notin N^k} \rho_{ji}\, (1 - \theta_j(\hat{p}_j))\, v_j(\boldsymbol{p}^k, \hat{\boldsymbol{p}}^{-k})$ for all $i \in N$, where the prices charged by firm k are fixed at $\boldsymbol{p}^k$ and the prices charged by the other firms are fixed at $\hat{\boldsymbol{p}}^{-k}$. In this case, if the firms other than firm k charge the prices $\hat{\boldsymbol{p}}^{-k}$ for their products, then firm k can find the prices to charge for its products to maximize the expected profit that it obtains from each customer by solving the problem

$$\max_{\boldsymbol{p}^k \in \times_{i \in N^k} \mathcal{P}_i} \left\{ \sum_{i \in N^k} \theta_i(p_i)\, v_i(\boldsymbol{p}^k, \hat{\boldsymbol{p}}^{-k})\, (p_i - c_i) \right\}. \tag{5}$$

To solve problem (5), we follow an approach based on dynamic programming. We use $\hat{r}_i^k$ to denote the optimal expected profit that firm k obtains from a customer visiting product i, given that the other firms charge the prices $\hat{\boldsymbol{p}}^{-k}$ for their products. First, we consider the case $i \in N^k$ so that firm k owns product i. If firm k charges the price p_i for product i, then a customer visiting product i purchases this product with probability $\theta_i(p_i)$, in which case, firm k obtains an optimal expected profit of simply $p_i - c_i$. Also, a customer visiting product i does not purchase this product with probability $1 - \theta_i(p_i)$ and transitions from product i to product j with probability ρ_{ij}, in which case, firm k obtains an optimal expected profit of $\hat{r}_j^k$. Thus, if firm k charges the price p_i for product i, then it obtains an optimal expected profit of $\theta_i(p_i)\, (p_i - c_i) + (1 - \theta_i(p_i)) \sum_{j \in N} \rho_{ij}\, \hat{r}_j^k$ from a customer visiting product $i \in N^k$. Second, we consider the case $i \notin N^k$ so that firm k does not own product i. Since firm k does not own product i, if a customer visiting product i purchases it, then firm k does not obtain a profit. If the firms other than firm k charge the prices $\hat{\boldsymbol{p}}^{-k}$, then a customer visiting product i does not purchase this product with probability $1 - \theta_i(\hat{p}_i)$ and transitions from product i to product j with probability ρ_{ij}, in which case, firm k obtains an optimal expected profit of $\hat{r}_j^k$. Thus, if the firms other than firm k charge the prices $\hat{\boldsymbol{p}}^{-k}$, then firm k obtains an optimal expected profit of $(1 - \theta_i(\hat{p}_i)) \sum_{j \in N} \rho_{ij}\, \hat{r}_j^k$ from a customer visiting product $i \notin N^k$.

The discussion in the previous paragraph intuitively suggests that if we use $\hat{r}_i^k$ to denote the optimal expected profit that firm k obtains from a customer visiting product i given that the other firms charge the prices $\hat{\boldsymbol{p}}^{-k}$ for their products, then $\hat{\boldsymbol{r}}^k = \{\hat{r}_i^k : i \in N\}$ should satisfy $\hat{r}_i^k = \max_{p_i \in \mathcal{P}_i} \{\theta_i(p_i)\, (p_i - c_i) + (1 - \theta_i(p_i)) \sum_{j \in N} \rho_{ij}\, \hat{r}_j^k\}$ for all $i \in N^k$ and $\hat{r}_i^k = (1 - \theta_i(\hat{p}_i)) \sum_{j \in N} \rho_{ij}\, \hat{r}_j^k$ for all $i \notin N^k$. Shortly, we make this intuition formal. In the last equality, we observe that since $i \notin N^k$, the price of product i is controlled by a firm other than firm k and its price is fixed at $\hat{p}_i$. To write the last two equalities succinctly, we define the operator $g_i^k(\cdot \,|\, \hat{\boldsymbol{p}}^{-k}) : \Re^n \to \Re$ as

$$g_i^k(\boldsymbol{r}^k \,|\, \hat{\boldsymbol{p}}^{-k}) = \begin{cases} \max\limits_{p_i \in \mathcal{P}_i} \left\{ \theta_i(p_i)\,(p_i - c_i) + (1 - \theta_i(p_i)) \sum\limits_{j \in N} \rho_{ij}\, r_j^k \right\} & \text{if } i \in N^k \\ (1 - \theta_i(\hat{p}_i)) \sum\limits_{j \in N} \rho_{ij}\, r_j^k & \text{if } i \notin N^k, \end{cases} \tag{6}$$

in which case, $\hat{\boldsymbol{r}}^k$ should satisfy $\hat{r}_i^k = g_i^k(\hat{\boldsymbol{r}}^k \,|\, \hat{\boldsymbol{p}}^{-k})$ for all $i \in N$. For any two vectors $\boldsymbol{r}^k = \{r_i^k : i \in N\}$ and $\boldsymbol{q}^k = \{q_i^k : i \in N\}$, we can follow the same approach in Appendix B to show that $|g_i^k(\boldsymbol{r}^k \,|\, \hat{\boldsymbol{p}}^{-k}) - g_i^k(\boldsymbol{q}^k \,|\, \hat{\boldsymbol{p}}^{-k})| \le \max_{i \in N}\{\sum_{j \in N} \rho_{ij}\}\, \|\boldsymbol{r}^k - \boldsymbol{q}^k\|$ for all $i \in N$, which implies that the operator $\{g_i^k(\cdot \,|\, \hat{\boldsymbol{p}}^{-k}) : i \in N\} : \Re^n \to \Re^n$ is a contraction with respect to the norm $\|\cdot\|$. In this case, there exists $\hat{\boldsymbol{r}}^k = \{\hat{r}_i^k : i \in N\}$ that satisfies $\hat{r}_i^k = g_i^k(\hat{\boldsymbol{r}}^k \,|\, \hat{\boldsymbol{p}}^{-k})$ for all $i \in N^k$. In the next lemma, we formally show that we can indeed obtain an optimal solution to problem (5) by using $\hat{\boldsymbol{r}}^k = \{\hat{r}_i^k : i \in N\}$ that satisfies $\hat{r}_i^k = g_i^k(\hat{\boldsymbol{r}}^k \,|\, \hat{\boldsymbol{p}}^{-k})$ for all $i \in N$. The proof of this lemma follows the same approach in the proof of Theorem 3 and we do not give a proof.

Lemma 5 *The prices $\hat{\boldsymbol{p}}^k = \{\hat{p}_i : i \in N^k\}$ are an optimal solution to problem* (5) *if and only if the price $\hat{p}_i$ is an optimal solution to the problem*

$$\max_{p_i \in \mathcal{P}_i} \left\{ \theta_i(p_i)\,(p_i - c_i) + (1 - \theta_i(p_i)) \sum_{j \in N} \rho_{ij}\, \hat{r}_j^k \right\}$$

for all $i \in N^k$, where $\hat{\boldsymbol{r}}^k = \{\hat{r}_i^k : i \in N\}$ satisfies $\hat{r}_i^k = g_i^k(\hat{\boldsymbol{r}}^k \,|\, \hat{\boldsymbol{p}}^{-k})$ for all $i \in N$.

Problem (5) computes the best response of firm k to the prices $\hat{\boldsymbol{p}}^{-k}$ charged by the other firms. Lemma 5 uses $\hat{\boldsymbol{r}}^k$ satisfying $\hat{r}_i^k = g_i^k(\hat{\boldsymbol{r}} \,|\, \hat{\boldsymbol{p}}^{-k})$ for all $i \in N$ to find the best response. This lemma ultimately becomes useful to establish the existence of a Nash equilibrium.

We comment on the assumption that different firms own different partitions of the products. In our model, two different products i and j owned by two different firms could correspond to the same product sold by two different firms. For example, two products i and j could correspond to the same soap sold by two different supermarket chains. A customer purchasing product i would correspond to a customer purchasing the soap from the first supermarket chain, whereas a customer purchasing product j would correspond to a customer purchasing the soap from the second supermarket chain. The approach that we use in our model is consistent with the one in Gallego et al. (2006), Li and Huh (2011) and Gallego and Wang (2014). In Appendix C, we discuss an alternative model. In this model, if the same soap is sold by two supermarket chains, then there is only one product corresponding to this soap in the Markov chain choice model. If a customer visits this product, then she decides which supermarket chain to purchase from, based on the prices charged by the two supermarket chains. Depending on the parameters of the model, we demonstrate that the same soap may or may not be offered by the two supermarket chains simultaneously.

3.2 Existence of Equilibrium

In this section, we show that there exists a Nash equilibrium for competitive pricing. We use $\boldsymbol{p} = \{p_i : i \in N^k,\ k \in M\}$ to denote the prices charged for the products, where $\boldsymbol{p}^k = \{p_i : i \in N^k\}$ captures the prices that firm k charges for its products and $\boldsymbol{p}^{-k} = \{p_i : i \notin N^k\}$ captures the prices that the other firms charge for their products. The prices $\hat{\boldsymbol{p}} = \{\hat{p}_i : i \in N^k,\ k \in M\}$ are a Nash equilibrium if and only if, for all $k \in M$, the prices $\hat{\boldsymbol{p}}^k = \{\hat{p}_i : i \in N^k\}$ are an optimal solution to problem (5) for firm k when the other firms charge the prices $\hat{\boldsymbol{p}}^{-k} = \{\hat{p}_i : i \notin N^k\}$ for their products. In other words, the prices $\hat{\boldsymbol{p}} = \{\hat{p}_i : i \in N^k,\ k \in M\}$ are a Nash equilibrium if and only if, for all $k \in M$, the prices $\hat{\boldsymbol{p}}^k = \{\hat{p}_i : i \in N^k\}$ are a best response of firm k to the prices $\hat{\boldsymbol{p}}^{-k} = \{\hat{p}_i : i \notin N^k\}$ that the other firms charge for their products. One approach for establishing the existence of a Nash equilibrium is based on supermodular games; see Topkis (1979), Milgrom and Roberts (1990) and Vives (1990). However, we can generate counterexamples to show that our competitive pricing setting does not yield a supermodular game. In particular, consider a Markov chain choice model with two products. There are two firms, each of which owning one of the products. The unit costs c_1 and c_2 are both zero. We use $\pi_1(p_1, p_2)$ to denote the expected profit of the first firm when the firms charge the prices (p_1, p_2), so that $\pi_1(p_1, p_2) = \theta_1(p_1)\, v_1(p_1, p_2)\, p_1$. The parameters of the Markov chain choice model are

$$\lambda_1 = 0.1, \quad \lambda_2 = 0.9, \quad \theta_1(p_1) = \mathrm{e}^{-0.1\, p_1}, \quad \theta_2(p_2) = \mathrm{e}^{-0.4\, p_2},$$
$$\rho_{12} = 0.2, \quad \rho_{21} = 0.8, \quad \rho_{11} = \rho_{22} = 0.$$

For $p_1^+ = 15$, $p_1^0 = 8$, $p_2^+ = 4$ and $p_2^0 = 2$, we have $\pi_1(p_1^+, p_2^+) - \pi_1(p_1^0, p_2^+) < \pi_1(p_1^+, p_2^0) - \pi_1(p_1^0, p_2^0)$, indicating that the expected profit of the first firm $\pi_1(p_1, p_2)$ does not have increasing differences in (p_1, p_2). Thus, our competitive pricing setting does not yield a supermodular game.

Although the competitive pricing setting does not yield a supermodular game, it is possible to use first principles to characterize a Nash equilibrium and to show that a Nash equilibrium exists. We follow a dynamic programming approach that is similar to the one in the previous section to give a characterization of a Nash equilibrium. We use $\hat{r}_i^k$ to denote the expected profit that firm k obtains from a customer visiting product i given that each firm charges the prices that are its best response to the others. First, we consider the case where $i \in N^k$ so that firm k owns product i. Similar to the discussion in the previous section, if firm k charges the price p_i for product i and the other firms charge the prices that are their best responses, then firm k obtains an expected profit of $\theta_i(p_i)\,(p_i - c_i) + (1 - \theta_i(p_i)) \sum_{j \in N} \rho_{ij}\, \hat{r}_j^k$ from a customer visiting product $i \in N^k$. Thus, firm k can find its best response to the other firms by solving the problem $\max_{p_i \in \mathcal{P}_i} \{\theta_i(p_i)\,(p_i - c_i) + (1 - \theta_i(p_i)) \sum_{j \in N} \rho_{ij}\, \hat{r}_j^k\}$. We use $\hat{p}_i$ to denote an optimal solution to this problem. Second, we consider the case where $i \notin N^k$ so that firm k does not own product i. In this case, if the firms other than firm k charge the prices that are their best responses, then firm k obtains an expected profit of $(1 - \theta_i(\hat{p}_i)) \sum_{j \in N} \rho_{ij}\, \hat{r}_j^k$ from a customer visiting product $i \notin N^k$. The preceding discussion intuitively suggests that if we use $\hat{r}_i^k$ to denote the expected profit that firm k obtains from a customer visiting product i given that each firm charges the prices that are its best response to the others, then $\hat{\boldsymbol{r}}^k = \{\hat{r}_i^k : i \in N^k\}$ for all $k \in M$ should satisfy $\hat{r}_i^k = \max_{p_i \in \mathcal{P}_i}\{\theta_i(p_i)\,(p_i - c_i) + (1-\theta_i(p_i)) \sum_{j\in N} \rho_{ij}\, \hat{r}_j^k\}$ for all $i \in N^k$, $k \in M$ and $\hat{r}_i^k = (1-\theta_i(\hat{p}_i)) \sum_{j \in N} \rho_{ij}\, \hat{r}_j^k$ for all $i \notin N^k$, $k \in M$, where, for all $i \in N^k$, $k \in M$, $\hat{p}_i$ is an optimal solution to problem $\max_{p_i\in\mathcal{P}_i}\{\theta_i(p_i)\,(p_i - c_i) + (1-\theta_i(p_i)) \sum_{j\in N} \rho_{ij}\, \hat{r}_j^k\}$. (Since $\cup_{k\in M} N^k = N$, the prices $\{\hat{p}_i : i \in N^k,\ k \in M\}$ provide a price $\hat{p}_i$ for each product $i \in N$.) Shortly, we formally show that we can use these equalities to characterize a Nash equilibrium. To write these equalities succinctly, for all $i \in N$, $k \in M$, we define the operator $h_i^k(\cdot,\ldots,\cdot) : \Re^{n\times m} \to \Re_+$ as

$$h_i^k(\boldsymbol{r}^1,\ldots,\boldsymbol{r}^m) = \begin{cases} \max\limits_{p_i\in\mathcal{P}_i}\left\{\theta_i(p_i)\,(p_i-c_i) + (1-\theta_i(p_i)) \sum\limits_{j\in N}\rho_{ij}\, r_j^k\right\} & \text{if } i \in N^k \\ (1-\theta_i(\hat{p}_i)) \sum\limits_{j\in N} \rho_{ij}\, r_j^k & \text{if } i \notin N^k, \end{cases} \tag{7}$$

where, for all $i \in N^k$, $k \in M$, the price $\hat{p}_i$ used in the second case above is given by an optimal solution to the problem $\max_{p_i\in\mathcal{P}_i}\{\theta_i(p_i)\,(p_i - c_i) + (1-\theta_i(p_i)) \sum_{j\in N} \rho_{ij}\, r_j^k\}$. Thus, the preceding discussion implies that $(\hat{\boldsymbol{r}}^1,\ldots,\hat{\boldsymbol{r}}^m)$ should satisfy $\hat{r}_i^k = h_i^k(\hat{\boldsymbol{r}}^1,\ldots,\hat{\boldsymbol{r}}^m)$ for all $i \in N$, $k \in M$. To compute $h_i^k(\boldsymbol{r}^1,\ldots,\boldsymbol{r}^m)$, we begin by solving the maximization problem in the first case in (7) for all $i \in N^k$, $k \in M$ to obtain $\{\hat{p}_i : i \in N\}$. Once we have $\{\hat{p}_i : i \in N\}$, we can compute $(1-\theta_i(\hat{p}_i)) \sum_{j\in N}\rho_{ij}\, r_j^k$ as in the second case in (7). We observe that the operator $h_i^k(\cdot,\ldots,\cdot)$ is similar to the operator $g_i^k(\cdot \,|\, \hat{\boldsymbol{p}}^{-k})$ defined in the previous section, but the price $\hat{p}_i$ used in the second case in the definition of $h_i^k(\cdot,\ldots,\cdot)$ is given by an optimal solution to the problem $\max_{p_i\in\mathcal{P}_i}\{\theta_i(p_i)\,(p_i - c_i) + (1-\theta_i(p_i)) \sum_{j\in N} \rho_{ij}\, r_j^k\}$, whereas, the price $\hat{p}_i$ used in the second case in the definition of $g_i^k(\cdot \,|\, \hat{\boldsymbol{p}}^{-k})$ is fixed by the prices $\hat{\boldsymbol{p}}^{-k}$. Due to this difference, it is possible to generate counterexamples to show that the operator $\{h_i^k(\cdot,\ldots,\cdot) : i \in N,\ k \in M\} : \Re^{n\times m} \to \Re^{n\times m}$ is not a contraction. However, as we demonstrate shortly, we can start from first principles to show that there exists $(\hat{\boldsymbol{r}}^1,\ldots,\hat{\boldsymbol{r}}^m)$ that satisfies $\hat{r}_i^k = h_i^k(\hat{\boldsymbol{r}}^1,\ldots,\hat{\boldsymbol{r}}^m)$ for all $i \in N$, $k \in M$.

In the next theorem, we formally show that we can use $(\hat{\boldsymbol{r}}^1,\ldots,\hat{\boldsymbol{r}}^m)$ that satisfies $\hat{r}_i^k = h_i^k(\hat{\boldsymbol{r}}^1,\ldots,\hat{\boldsymbol{r}}^m)$ for all $i \in N$, $k \in M$ to obtain a Nash equilibrium. After this theorem, we argue that there indeed exists $(\hat{\boldsymbol{r}}^1,\ldots,\hat{\boldsymbol{r}}^m)$ that satisfies $\hat{r}_i^k = h_i^k(\hat{\boldsymbol{r}}^1,\ldots,\hat{\boldsymbol{r}}^m)$ for all $i \in N$, $k \in M$.

Theorem 6 *The prices $\hat{\boldsymbol{p}} = \{\hat{p}_i : i \in N^k,\ k \in M\}$ are a Nash equilibrium for competitive pricing if and only if the price $\hat{p}_i$ is an optimal solution to the problem*

$$\max_{p_i\in\mathcal{P}_i}\left\{\theta_i(p_i)\,(p_i-c_i) + (1-\theta_i(p_i)) \sum_{j\in N}\rho_{ij}\, \hat{r}_j^k\right\} \tag{8}$$

for all $i \in N^k$, $k \in M$, where $(\hat{\boldsymbol{r}}^1,\ldots,\hat{\boldsymbol{r}}^m)$ satisfies $\hat{r}_i^k = h_i^k(\hat{\boldsymbol{r}}^1,\ldots,\hat{\boldsymbol{r}}^m)$ for all $i \in N$, $k \in M$.

In the proof of Theorem 6, we use the characterization of the best response that we give in Lemma 5. Assuming $(\hat{\boldsymbol{r}}^1,\ldots,\hat{\boldsymbol{r}}^m)$ satisfies $\hat{r}_i^k = h_i^k(\hat{\boldsymbol{r}}^1,\ldots,\hat{\boldsymbol{r}}^m)$ for all $i \in N$, $k \in M$, if we use $\hat{p}_i$ to denote an optimal solution to problem (8) for all $i \in N^k$, $k \in M$, then Theorem 6 implies that $\hat{\boldsymbol{p}} = \{\hat{p}_i : i \in N^k,\ k \in M\}$ is a Nash equilibrium. Therefore, if we can show that there exists $(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ that satisfies $\hat{r}_i^k = h_i^k(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ for all $i \in N$, $k \in M$, then it follows that there exists a Nash equilibrium for competitive pricing. Next, we proceed to arguing that there indeed exists $(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ that satisfies $\hat{r}_i^k = h_i^k(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ for all $i \in N$, $k \in M$. In Appendix D, we show that the operator $h_i^k(\cdot, \ldots, \cdot)$ is monotone. In other words, if $r_i^k \leq q_i^k$ for all $i \in N$, $k \in M$, then we have $h_i^k(\boldsymbol{r}^1, \ldots, \boldsymbol{r}^m) \leq h_i^k(\boldsymbol{q}^1, \ldots, \boldsymbol{q}^m)$ for all $i \in N$, $k \in M$. We generate the sequence $\{(\bar{\boldsymbol{r}}^1(t), \ldots, \bar{\boldsymbol{r}}^m(t)) : t \in \mathbb{N}\}$ by using the relationship $\bar{r}_i^k(t+1) = h_i^k(\bar{\boldsymbol{r}}^1(t), \ldots, \bar{\boldsymbol{r}}^m(t))$ for all $i \in N$, $k \in M$ and starting with the initial condition that $\bar{r}_i^k(0) = \bar{u}$ for all $i \in N$, $k \in M$ for some $\bar{u} \in \Re_+$. In this case, in Appendix E, we use the monotonicity of the operator $h_i^k(\cdot, \ldots, \cdot)$ to show that if we choose $\bar{u}$ large enough, then the sequence $\{(\bar{\boldsymbol{r}}^1(t), \ldots, \bar{\boldsymbol{r}}^m(t)) : t \in \mathbb{N}\}$ is decreasing and bounded from below, satisfying $\bar{r}_i^k(t) \geq \bar{r}_i^k(t+1) \geq 0$ for all $i \in N$, $k \in M$, $t \in \mathbb{N}$. In particular, letting $\rho_{\max} = \max_{i \in N}\{\sum_{j \in N} \rho_{ij}\}$ and $\Delta = \max_{i \in N}\{\max_{p_i \in \mathcal{P}_i}\{\theta_i(p_i)\,(p_i - c_i)\}\}$, it suffices to fix $\bar{u}$ as $\bar{u} = \Delta/(1 - \rho_{\max})$ to ensure that the sequence $\{(\bar{\boldsymbol{r}}^1(t), \ldots, \bar{\boldsymbol{r}}^m(t)) : t \in \mathbb{N}\}$ is decreasing and bounded from below. Since the sequence is decreasing and bounded from below, it has a limit. Using $(\bar{\boldsymbol{r}}^1, \ldots, \bar{\boldsymbol{r}}^m)$ to denote the limit, noting that the sequence $\{(\bar{\boldsymbol{r}}^1(t), \ldots, \bar{\boldsymbol{r}}^m(t)) : t \in \mathbb{N}\}$ is generated by using the relationship $\bar{r}_i^k(t+1) = h_i^k(\bar{\boldsymbol{r}}^1(t), \ldots, \bar{\boldsymbol{r}}^m(t))$ for all $i \in N$, $k \in M$, its limit $(\bar{\boldsymbol{r}}^1, \ldots, \bar{\boldsymbol{r}}^m)$ must satisfy $\bar{r}_i^k = h_i^k(\bar{\boldsymbol{r}}^1, \ldots, \bar{\boldsymbol{r}}^m)$ for all $i \in N$, $k \in M$, which establishes that there exists $(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ that satisfies $\hat{r}_i^k = h_i^k(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ for all $i \in N$, $k \in M$. Therefore, by the discussion right after Theorem 6, there exists a Nash equilibrium.

3.3 Properties of Equilibrium

We show four structural properties of the Nash equilibrium. First, we show that the price for each product in any Nash equilibrium is no larger than its price when a central planner computes the prices to maximize the expected profit obtained from each customer. Second, we show that there exists a Nash equilibrium that Pareto dominates any other Nash equilibria. Third, we show that if the competition gets more intense in a sense that we make precise, then all prices in the Pareto dominant equilibrium get smaller. Fourth, we show that if each firm owns one product, then all prices in the Pareto dominant equilibrium increase when the unit cost of any product increases. In the next theorem, we show that the price for each product in any Nash equilibrium is no larger than its price in an optimal solution to problem (2). Note that problem (2) corresponds to the case where a central planner computes all prices without any competition.

Theorem 7 *If the prices $\hat{\boldsymbol{p}} = \{\hat{p}_i : i \in N^k,\ k \in M\}$ are a Nash equilibrium for competitive pricing and the prices $\tilde{\boldsymbol{p}} = \{\tilde{p}_i : i \in N\}$ are an optimal solution to problem* (2)*, then we have $\hat{p}_i \leq \tilde{p}_i$ for all $i \in N^k$, $k \in M$.*

In the proof of Theorem 7, we show that if $(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ satisfies $\hat{r}_i^k = h_i^k(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ for all $i \in N$, $k \in M$ and $\tilde{\boldsymbol{r}} = \{\tilde{r}_i : i \in N\}$ satisfies $\tilde{r}_i = f_i(\tilde{\boldsymbol{r}})$ for all $i \in N$, then $\hat{r}_i^k \leq \tilde{r}_i$ for all $i \in N$, $k \in M$. In this case, the result follows by arguing that the optimal solutions to problems (4) and (8) are respectively increasing in $\hat{r}_j$ and $\hat{r}_j^k$. Theorem 7 shows that the prices charged for the products in any Nash equilibrium do not exceed their corresponding prices computed by a central planner. Therefore, competition has the effect of lowering the prices of the products.

Another property of the Nash equilibrium for competitive pricing is that there exists a Nash equilibrium that Pareto dominates any other Nash equilibria. In other words, there exists a Nash equilibrium where the expected profit obtained by each firm is at least as large as its corresponding expected profit in any other Nash equilibria. Thus, the Pareto dominant equilibrium is simultaneously preferred by all firms. To show this result, we consider the sequence $\{(\bar{\boldsymbol{r}}^1(t), \ldots, \bar{\boldsymbol{r}}^m(t)) : t \in \mathbb{N}\}$ used in the previous section. Letting $\bar{u} = \Delta/(1-\rho_{\max})$ be as defined right after Theorem 6, this sequence is generated by using the relationship $\bar{r}_i^k(t+1) = h_i^k(\bar{\boldsymbol{r}}^1(t), \ldots, \bar{\boldsymbol{r}}^m(t))$ for all $i \in N$, $k \in M$ with the initial condition that $\bar{r}_i^k(0) = \bar{u}$ for all $i \in N$, $k \in M$. In the previous section, we argue that the sequence $\{(\bar{\boldsymbol{r}}^1(t), \ldots, \bar{\boldsymbol{r}}^m(t)) : t \in \mathbb{N}\}$ has a limit $(\bar{\boldsymbol{r}}^1, \ldots, \bar{\boldsymbol{r}}^m)$ satisfying $\bar{r}_i^k = h_i^k(\bar{\boldsymbol{r}}^1, \ldots, \bar{\boldsymbol{r}}^m)$ for all $i \in N$, $k \in M$. In this case, if we let $\bar{p}_i$ be an optimal solution to the problem $\max_{p_i \in \mathcal{P}_i} \{\theta_i(p_i)\,(p_i - c_i) + (1 - \theta_i(p_i)) \sum_{j \in N} \rho_{ij}\, \bar{r}_j^k\}$, then Theorem 6 implies that $\bar{\boldsymbol{p}} = \{\bar{p}_i : i \in N^k,\ k \in M\}$ is a Nash equilibrium. We refer to the Nash equilibrium $\bar{\boldsymbol{p}} = \{\bar{p}_i : i \in N^k,\ k \in M\}$ as the Nash equilibrium induced by the sequence $\{(\bar{\boldsymbol{r}}^1(t), \ldots, \bar{\boldsymbol{r}}^m(t)) : t \in \mathbb{N}\}$. Noting (5), the expected profit that firm k obtains from each customer in this Nash equilibrium is $\sum_{i \in N^k} \theta_i(p_i)\, v_i(\bar{\boldsymbol{p}}^k, \bar{\boldsymbol{p}}^{-k})\,(\bar{p}_i - c_i)$, where $\bar{\boldsymbol{p}}^k = \{\bar{p}_i : i \in N^k\}$ captures the prices that firm k charges and $\bar{\boldsymbol{p}}^{-k} = \{\bar{p}_i : i \notin N^k\}$ captures the prices that the other firms charge. In the next theorem, we show that the expected profit that each firm obtains in this Nash equilibrium is at least as large as the one that it obtains in any other Nash equilibria.

Theorem 8 *If the prices $\bar{\boldsymbol{p}} = \{\bar{p}_i : i \in N^k,\ k \in M\}$ are the Nash equilibrium induced by the sequence $\{(\bar{\boldsymbol{r}}^1(t), \ldots, \bar{\boldsymbol{r}}^m(t)) : t \in \mathbb{N}\}$ and the prices $\hat{\boldsymbol{p}} = \{\hat{p}_i : i \in N^k, k \in M\}$ are any Nash equilibrium, then $\sum_{i \in N^k} \theta_i(\bar{p}_i)\, v_i(\bar{\boldsymbol{p}}^k, \bar{\boldsymbol{p}}^{-k})\,(\bar{p}_i - c_i) \geq \sum_{i \in N^k} \theta_i(\hat{p}_i)\, v_i(\hat{\boldsymbol{p}}^k, \hat{\boldsymbol{p}}^{-k})\,(\hat{p}_i - c_i)$ for all $k \in M$.*

By Theorem 6, if $\hat{\boldsymbol{p}}$ is some Nash equilibrium, then $\hat{p}_i^k$ is an optimal solution to problem (8), where $(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ satisfies $\hat{r}_i^k = h_i^k(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ for all $i \in N$, $k \in M$. Letting $(\bar{\boldsymbol{r}}^1, \ldots, \bar{\boldsymbol{r}}^m)$ be the limit of the sequence $\{(\bar{\boldsymbol{r}}^1(t), \ldots, \bar{\boldsymbol{r}}^m(t)) : t \in \mathbb{N}\}$, the proof of Theorem 8 is based on showing that $\bar{r}_i^k \geq \hat{r}_i^k$ for all $i \in N$, $k \in M$ and relating the expected profit of a firm in the Nash equilibrium $\hat{\boldsymbol{p}}$ to $(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$. By Theorem 8, the Nash equilibrium induced by the sequence $\{(\bar{\boldsymbol{r}}^1(t), \ldots, \bar{\boldsymbol{r}}^m(t)) : t \in \mathbb{N}\}$ is Pareto dominant. Next, we study the prices in this Pareto dominant equilibrium as the competition gets more intense. We consider two systems. In both systems, the set of products is N. In the first system, the set of firms is $M = \{1, \ldots, m\}$. The sets of products owned by the firms are $N^1, \ldots, N^{m-1}, N^m$. In the second system, the set of firms is $\tilde{M} = \{1, \ldots, m+1\}$. The sets of products owned by the firms are $N^1, \ldots, N^{m-1}, \tilde{N}^m, \tilde{N}^{m+1}$, where $\tilde{N}^m \cup \tilde{N}^{m+1} = N^m$. So, there is an additional firm in the second system. The firms $1, \ldots, m-1$ own the same sets of products in the two systems. The set of products owned by firm m in the first system is split between firms m and $m+1$ in the second system. Thus, the second system is more competitive than the first one in some sense. In both systems, the customers choose according to Markov chain choice model with the same parameters. We say that second system is marginally more competitive than the first one. In the next theorem, we show that the prices in the Pareto dominant equilibrium in the second system are no larger than those in the first system.

Theorem 9 *Consider two systems, where the second system is marginally more competitive than the first system. In this case, letting the prices* $\hat{\boldsymbol{p}} = \{\hat{p}_i : i \in N^k,\ k \in M\}$ *and* $\tilde{\boldsymbol{p}} = \{\tilde{p}_i : i \in N^k,\ k \in \tilde{M}\}$ *respectively be the Pareto dominant equilibria in the first and second systems, we have* $\hat{p}_i \geq \tilde{p}_i$ *for all* $i \in N$.

The proof of the theorem above follows from an argument similar to the one that we use in the proof of Theorem 7. We can successively apply the result in Theorem 9. In particular, we can start with a system with only one firm owning all products, corresponding to a system controlled by a central planner. We can split the set of products owned by this firm to obtain a second system marginally more competitive than the first one. We can keep on splitting the products owned by the firms to obtain any competitive setting with any desired division of the products among the firms. By Theorem 9, the optimal prices in the first system, which is controlled by a central planner, are at least as large as the prices in the Pareto dominant equilibrium in the competitive pricing setting. This observation resembles the result in Theorem 7, but neither of these results is more general than the other. In particular, Theorem 7 compares the prices charged by a central planner with those in any Nash equilibrium, which may or may not be the Pareto dominant equilibrium. In contrast, Theorem 9 compares the prices in the Pareto dominant equilibria in two competitive pricing settings, one of which may or may not involve a central planner.

In Figure 1, we consider an example with eight products and m firms. Each firm owns $8/m$ products. The products owned by firm k are $\{(k-1)\,8/m + 1, \ldots, k\,8/m\}$. We vary the number of firms over $m \in \{1, 2, 4, 8\}$. In Appendix F, we give all parameters of the Markov chain choice model governing the choices of the customers. For each value of m, we compute the Pareto dominant equilibrium. On the left side of Figure 1, we show prices in the equilibrium when there are different numbers of competing firms. The horizontal axis shows the products. Different data series plot prices in the equilibrium when there are different numbers of competing firms. Competition can have a significant effect on the prices. The price for product 1 ranges between 8.43 to 3.63 depending on the number of firms. On the right side of Figure 1, we show the sum of the expected profits obtained by all firms from a customer when there are different numbers of competing firms. The horizontal axis shows the number of firms. Different bars show the total expected profit obtained by all firms when there are different numbers of competing firms. The expected profit obtained from a customer ranges between 5.21 to 3.25 depending on the number of firms.

Lastly, we consider how the unit costs affect the prices in a Nash equilibrium. In the next lemma, we focus on the case where each firm owns a single product and show that if the unit cost of a product increases, then the prices charged by all firms for all products in the Pareto dominant

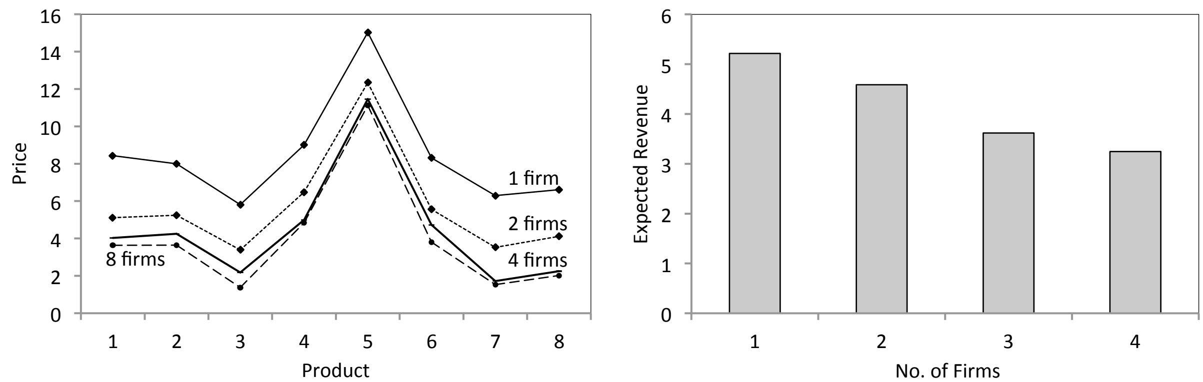

Figure 1: Prices and total expected profit from a customer in the Pareto dominant equilibrium.

equilibrium increase. In this lemma, we use $\hat{\boldsymbol{p}}(\boldsymbol{c}) = \{\hat{p}_i(\boldsymbol{c}) : i \in N^k,\ k \in M\}$ to denote the prices in the Pareto dominant equilibrium as a function of the unit costs $\boldsymbol{c} = \{c_i : i \in N\}$.

Lemma 10 *Consider a system where each firm owns one product. For any* $\epsilon > 0$*, we have* $\hat{p}_i(\boldsymbol{c} + \epsilon\, \boldsymbol{e}_j) \geq \hat{p}_i(\boldsymbol{c})$ *for all* $i \in N^k$*,* $k \in M$*,* $j \in N$*.*

Intuitively speaking, if the unit cost of a product increases, then the firm owning this product increases the price for this product to make up for the increase in the unit cost. This increase in the price creates an opportunity for all other firms to increase their prices as well. Thus, if the unit cost of a product increases, then the prices charged by all firms for all of the products in the Pareto dominant equilibrium increases. In Appendix G, we give a counterexample to show that if each firm owns an arbitrary number of products, then an increase in the unit cost of a product owned by a firm may result in an increase or a decrease in the prices charged by its competitors in the Pareto dominant equilibrium. Therefore, it is not possible to extend the result in the lemma above to the case where each firm owns an arbitrary number of products.

3.4 Checking Uniqueness of Equilibrium

There exists a Nash equilibrium in the competitive pricing setting, but we leave the question of whether the equilibrium is, in general, unique open. In particular, a counterexample with multiple Nash equilibria or a proof of uniqueness for the Nash equilibrium have both been elusive to us. Nevertheless, we can give an efficient procedure that can be used to definitively check whether a particular problem instance has a unique Nash equilibrium. This procedure eliminates the need for an exhaustive search to check whether a particular problem instance has multiple Nash equilibria. In our procedure, we consider the sequence $\{(\bar{\boldsymbol{r}}^1(t), \ldots, \bar{\boldsymbol{r}}^m(t)) : t \in \mathbb{N}\}$ defined in Section 3.2. Letting $\bar{u} = \Delta/(1 - \rho_{\max})$ be as defined right after Theorem 6, this sequence is generated by initializing $(\bar{\boldsymbol{r}}^1(0), \ldots, \bar{\boldsymbol{r}}^m(0)) \in \Re^{n \times m}$ as $\bar{r}_i^k(0) = \bar{u}$ for all $i \in N$, $k \in M$ and using the relationship

$\bar{r}_i^k(t+1) = h_i^k(\bar{\boldsymbol{r}}^1(t), \ldots, \bar{\boldsymbol{r}}^m(t))$ for all $i \in N$, $k \in M$. In Section 3.2, we argue that the sequence $\{(\bar{\boldsymbol{r}}^1(t), \ldots, \bar{\boldsymbol{r}}^m(t)) : t \in \mathbb{N}\}$ has a limit $(\bar{\boldsymbol{r}}^1, \ldots, \bar{\boldsymbol{r}}^m)$ satisfying $\bar{r}_i^k = h_i^k(\bar{\boldsymbol{r}}^1, \ldots, \bar{\boldsymbol{r}}^m)$ for all $i \in N$, $k \in M$. Here, we also consider an analogue of this sequence generated by using a different initial condition. In particular, we consider the sequence $\{(\underline{\boldsymbol{r}}^1(t), \ldots, \underline{\boldsymbol{r}}^m(t)) : t \in \mathbb{N}\}$ that is generated by initializing $(\underline{\boldsymbol{r}}^1(0), \ldots, \underline{\boldsymbol{r}}^m(0)) \in \Re^{n \times m}$ as $\underline{r}_i^k(0) = 0$ for all $i \in N$, $k \in M$ and using the relationship $\underline{r}_i^k(t+1) = h_i^k(\underline{\boldsymbol{r}}^1(t), \ldots, \underline{\boldsymbol{r}}^m(t))$ for all $i \in N$, $k \in M$. In Appendix H, we show that the sequence $\{(\underline{\boldsymbol{r}}^1(t), \ldots, \underline{\boldsymbol{r}}^m(t)) : t \in \mathbb{N}\}$ has a limit, which we denote by $(\underline{\boldsymbol{r}}^1, \ldots, \underline{\boldsymbol{r}}^m)$. In the next lemma, we show that if the limits of the sequences $\{(\bar{\boldsymbol{r}}^1(t), \ldots, \bar{\boldsymbol{r}}^m(t)) : t \in \mathbb{N}\}$ and $\{(\underline{\boldsymbol{r}}^1(t), \ldots, \underline{\boldsymbol{r}}^m(t)) : t \in \mathbb{N}\}$ are the same, then there is a unique Nash equilibrium.

Theorem 11 *Letting $(\bar{\boldsymbol{r}}^1, \ldots, \bar{\boldsymbol{r}}^m)$ and $(\underline{\boldsymbol{r}}^1, \ldots, \underline{\boldsymbol{r}}^m)$ respectively be the limits of the sequences $\{(\bar{\boldsymbol{r}}^1(t), \ldots, \bar{\boldsymbol{r}}^m(t)) : t \in \mathbb{N}\}$ and $\{(\underline{\boldsymbol{r}}^1(t), \ldots, \underline{\boldsymbol{r}}^m(t)) : t \in \mathbb{N}\}$, if $\bar{r}_i^k = \underline{r}_i^k$ for all $i \in N$, $k \in M$, then there is a unique Nash equilibrium for competitive pricing.*

We can use Theorem 11 to give a procedure to check whether the Nash equilibrium for a particular problem instance is unique. We can generate the sequence $\{(\bar{\boldsymbol{r}}^1(t), \ldots, \bar{\boldsymbol{r}}^m(t)) : t \in \mathbb{N}\}$ starting with $\bar{r}_i^k(0) = \bar{u}$ for all $i \in N$, $k \in M$ and using the relationship $\bar{r}_i^k(t+1) = h_i^k(\bar{\boldsymbol{r}}^1(t), \ldots, \bar{\boldsymbol{r}}^m(t))$ for all $i \in N$, $k \in M$. In this way, we can compute the limit $(\bar{\boldsymbol{r}}^1, \ldots, \bar{\boldsymbol{r}}^m)$ of this sequence. We can also compute the limit $(\underline{\boldsymbol{r}}^1, \ldots, \underline{\boldsymbol{r}}^m)$ of the sequence $\{(\underline{\boldsymbol{r}}^1(t), \ldots, \underline{\boldsymbol{r}}^m(t)) : t \in \mathbb{N}\}$ similarly. If we have $\bar{r}_i^k = \underline{r}_i^k$ for all $i \in N$, $k \in M$, then there is a unique Nash equilibrium. Naturally, we cannot compute the limits $(\bar{\boldsymbol{r}}^1, \ldots, \bar{\boldsymbol{r}}^m)$ and $(\underline{\boldsymbol{r}}^1, \ldots, \underline{\boldsymbol{r}}^m)$ exactly, but we can estimate them quite accurately. If the estimates are close, then the proof of Theorem 11 indicates that there do not exist Nash equilibria that significantly differ from each other.

4 Dynamic Pricing with a Single Resource

In this section, we study dynamic pricing problems with a single resource. We give structural properties of the optimal policy. We show that a deterministic approximation often used to develop heuristic policies has a tractable formulation. We index the products by $N = \{1, \ldots, n\}$. The set of time periods in the selling horizon is $T = \{1, \ldots, \tau\}$. At the beginning of the selling horizon, we have q units of resource available, which is exogenously given. The sale of each product consumes one unit of the resource. A time period in the selling horizon corresponds to a small enough interval of time that there is at most one customer arrival at each time period. The customer arriving at time period t chooses among the products according to the Markov chain choice model with parameters $\{\lambda_{it} : i \in N\}$, $\{\theta_{it}(\cdot) : i \in N\}$ and $\{\rho_{ijt} : i, j \in N\}$. Since we allow having $\sum_{i \in N} \lambda_{it} < 1$, with probability $1 - \sum_{i \in N} \lambda_{it}$, there is no customer arrival at time period t. We use $\mathcal{P}_{it} = [L_{it}, U_{it}]$ to denote the set of feasible prices for product i at time period t. The goal is to find a policy to decide what prices to charge at each time period to maximize the total expected revenue. Letting $\boldsymbol{p}_t = \{p_{it} : i \in N\}$ be the prices that we charge for the products at time period t, we use $\{v_{it}(\boldsymbol{p}_t) : i \in N\}$ to denote the expected number of times a customer visits product i during the course of her choice process, given that the customer chooses according to the Markov chain choice model with parameters $\{\lambda_{it} : i \in N\}$, $\{\theta_{it}(\cdot) : i \in N\}$ and $\{\rho_{ijt} : i, j \in N\}$. We can obtain $\{v_{it}(\boldsymbol{p}_t) : i \in N\}$ by solving the system of equations in (1). In this case, if we charge the prices $\boldsymbol{p}_t$ for the products at time period t, then a customer purchases product i with probability $\theta_{it}(p_{it})\, v_{it}(\boldsymbol{p}_t)$. Using x_t to denote the number of units of remaining resource at the beginning of time period t, we can compute the value functions $\{V_t(\cdot) : t \in T\}$ by solving the dynamic program

$$\begin{aligned} V_t(x_t) &= \max_{\boldsymbol{p}_t \in \times_{i \in N} \mathcal{P}_{it}} \left\{ \sum_{i \in N} \theta_{it}(p_{it})\, v_{it}(\boldsymbol{p}_t) \Big\{ p_{it} + V_{t+1}(x_t - 1) \Big\} + \Big\{ 1 - \sum_{i \in N} \theta_{it}(p_{it})\, v_{it}(\boldsymbol{p}_t) \Big\} V_{t+1}(x_t) \right\} \\ &= \max_{\boldsymbol{p}_t \in \times_{i \in N} \mathcal{P}_{it}} \left\{ \sum_{i \in N} \theta_{it}(p_{it})\, v_{it}(\boldsymbol{p}_t) \Big\{ p_{it} - \Delta V_{t+1}(x_t) \Big\} \right\} + V_{t+1}(x_t), \end{aligned} \tag{9}$$

where we use $\Delta V_{t+1}(x_t) = V_{t+1}(x_t) - V_{t+1}(x_t - 1)$. In the dynamic program above, the boundary conditions are $V_t(0) = 0$ for all $t \in T$ and $V_{\tau+1}(\cdot) = 0$.

In the next lemma, we show that the optimal policy obtained through the dynamic program above satisfies certain intuitive structural properties. In particular, we show that the optimal prices to charge for the products at a certain time period decreases as we have more units of the resource. Furthermore, if the Markov chain choice models that govern the choice behavior of the customers at different time periods have the same parameters, then the optimal prices to charge decrease as we get closer to the end of the selling horizon.

Lemma 12 *Letting $\hat{\boldsymbol{p}}_t(x) = \{\hat{p}_{it}(x) : i \in N\}$ be the optimal prices to charge at time period t when we have x units of remaining resource, we have $\hat{p}_{it}(x+1) \le \hat{p}_{it}(x)$ for all $i \in N$. Furthermore, if the Markov chain choice models at the different time periods have the same parameters, then we have $\hat{p}_{i,t+1}(x) \le \hat{p}_{it}(x)$ for all $i \in N$.*

By a standard result in the revenue management literature, the marginal value $\Delta V_t(x)$ of a unit of resource is decreasing in x and t; see Proposition 2-2.A.4 in Talluri and van Ryzin (2005). In the proof of Lemma 12, we use this fact and Lemma 4. By the lemma above, as we have more units of the resource at a certain time period, we have more incentive to sell products, motivating us to charge lower prices to sell products. Also, as we get closer to the end of the selling horizon, we run out of opportunities to sell products, also motivating us to charge lower prices to sell products. Although the structural properties in Lemma 12 are intuitive, they do not hold under arbitrary choice models. In Appendix I, we give a counterexample where the prices may increase as we have more units of the resource or as we get closer to the end of the selling horizon. In this counterexample, the same choice model governs the customer choices at different time periods and the choice model is compatible with the random utility maximization principle.

Gallego and van Ryzin (1994) use a deterministic approximation to come up with an a priori fixed price trajectory to charge over the selling horizon. In this deterministic approximation, we use the decision variables $\boldsymbol{p} = \{p_{it} : i \in N,\ t \in T\}$, where p_{it} is the price charged for product i at time period t. In this case, we consider the deterministic approximation

$$\max_{\boldsymbol{p} \in \Re^{n \times \tau}} \left\{ \sum_{t \in T} \sum_{i \in N} \theta_{it}(p_{it})\, v_{it}(\boldsymbol{p}_t)\, p_{it} \ : \ \sum_{t \in T} \sum_{i \in N} \theta_{it}(p_{it})\, v_{it}(\boldsymbol{p}_t) \leq q,\ p_{it} \in \mathcal{P}_{it} \ \ \forall i \in N,\ t \in T \right\}. \quad (10)$$

In problem (10), $\sum_{i \in N} \theta_{it}(p_{it})\, v_{it}(\boldsymbol{p})\, p_{it}$ is the expected revenue at time period t. In the constraint, $\sum_{i \in N} \theta_{it}(p_{it})\, v_{it}(\boldsymbol{p}_t)$ is the expected resource consumption at time period t. Thus, we enforce the capacity constraint only in the expected sense. Note that $\{v_{it}(\boldsymbol{p}_t) : i \in N\}$ is given by the solution to the system of equations in (1), so there are no closed form expressions for the objective function and the constraint above. We give an equivalent formulation for problem (10) with closed form expressions for the objective function and the constraint. Also, under certain assumptions on $\{\theta_{it}(\cdot) : i \in N,\ t \in T\}$, the equivalent formulation is a convex program.

In the equivalent formulation, we use the decision variables $\boldsymbol{x} = \{x_{it} : i \in N,\ t \in T\}$ and $\boldsymbol{y} = \{y_{it} : i \in N,\ t \in T\}$, where x_{it} is the number of times that a customer arriving at time period t visits product i during the course of her choice process and y_{it} is the probability that a customer arriving at time period t purchases product i. Therefore, the decision variables $(\boldsymbol{x}, \boldsymbol{y})$ relate to the prices $\boldsymbol{p}$ in problem (10) as $x_{it} = v_{it}(\boldsymbol{p}_t)$ and $y_{it} = \theta_{it}(p_{it})\, v_{it}(\boldsymbol{p}_t)$, so that $y_{it}/x_{it} = \theta_{it}(p_{it})$. Since $\theta_{it}(\cdot)$ is strictly decreasing, its inverse exists and we get $p_{it} = \theta_{it}^{-1}(y_{it}/x_{it})$. Substituting $p_{it} = \theta_{it}^{-1}(y_{it}/x_{it})$ and $\theta_{it}(p_{it})\, v_{it}(\boldsymbol{p}_t) = y_{it}$ in problem (10), we obtain the formulation

$$\max_{(\boldsymbol{x}, \boldsymbol{y}) \in \Re_+^{2 \times n \times \tau}} \left\{ \sum_{t \in T} \sum_{i \in N} y_{it}\, \theta_{it}^{-1}(y_{it}/x_{it}) \ : \ \sum_{t \in T} \sum_{i \in N} y_{it} \leq q, \ \ \theta_{it}^{-1}(y_{it}/x_{it}) \in \mathcal{P}_{it} \ \ \forall\, i \in N,\ t \in T, \right.$$
$$\left. x_{it} = \lambda_{it} + \sum_{j \in N} \rho_{jit}\, (x_{jt} - y_{jt}) \ \ \forall\, i \in N,\ t \in T \right\}. \quad (11)$$

In the problem above, the objective function and the first and second constraints follow by setting $p_{it} = \theta_{it}^{-1}(y_{it}/x_{it})$ and $\theta_{it}(p_{it})\, v_{it}(\boldsymbol{p}_t) = y_{it}$ in (10). By (1), since $v_{it}(\boldsymbol{p}_t)$ satisfies $v_{it}(\boldsymbol{p}_t) = \lambda_{it} + \sum_{j \in N} \rho_{jit}\, (1 - \theta_{jt}(p_{jt}))\, v_{jt}(\boldsymbol{p}_t)$, writing the last expression as $\lambda_{it} + \sum_{j \in N} \rho_{jit}\, (v_{jt}(\boldsymbol{p}_t) - \theta_{jt}(p_{jt})\, v_{jt}(\boldsymbol{p}_t))$ and substituting $x_{it} = v_{it}(\boldsymbol{p}_t)$ and $y_{it} = \theta_{it}(p_{it})\, v_{it}(\boldsymbol{p}_t)$, we get the third constraint above. The first and third constraints are linear in $(\boldsymbol{x}, \boldsymbol{y})$. Since $\mathcal{P}_{it} = [L_{it}, U_{it}]$ and $\theta_{it}(\cdot)$ is decreasing, we write the second constraint as $y_{it}/x_{it} \in [\theta_{it}(U_{it}), \theta_{it}(L_{it})]$ or $\theta_{it}(U_{it})\, x_{it} \leq y_{it} \leq \theta_{it}(L_{it})\, x_{it}$, which is also linear in $(\boldsymbol{x}, \boldsymbol{y})$. In the next theorem, we show that problems (10) and (11) are equivalent.

Theorem 13 *If $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ is an optimal solution to problem* (11)*, then letting $\hat{p}_{it} = \theta_{it}^{-1}(y_{it}/x_{it})$ for all $i \in N$, $t \in T$, $\hat{\mathbf{p}} = \{\hat{p}_i : i \in N,\ t \in T\}$ is an optimal solution to problem* (10)*. Conversely, if $\hat{\mathbf{p}}$ is an optimal solution to problem* (10)*, then letting $\hat{x}_{it} = v_{it}(\hat{\mathbf{p}}_t)$ and $\hat{y}_{it} = \theta_{it}(\hat{p}_{it})\, v_{it}(\hat{\mathbf{p}}_t)$ for all $i \in N$, $t \in T$, $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) = \{(\hat{x}_{it}, \hat{y}_{it}) : i \in N,\ t \in T\}$ is an optimal solution to problem* (11).

In the proof of Theorem 13, we explicitly construct a feasible solution to one of problems (10) and (11) by using an optimal solution to the other one such that the objective values provided by the two solutions for their respective problems are equal to each other. The objective function and the constraints in problem (11) have closed form expressions. Also, the constraints are linear in $(\boldsymbol{x}, \boldsymbol{y})$. The objective function of problem (11) is not always concave in $(\boldsymbol{x}, \boldsymbol{y})$, but we can come up with conditions to ensure concavity. In particular, in Appendix J, we show that if $1/\theta_{it}(p_{it})$ is convex in p_{it}, then $y_{it}\,\theta_{it}^{-1}(y_{it}/x_{it})$ is concave in (x_{it}, y_{it}). So, if $\theta_{it}(p_{it}) = \mathrm{e}^{-\beta_{it}\,p_{it}}$ with $\beta_{it} > 0$, then $1/\theta_{it}(p_{it}) = \mathrm{e}^{\beta_{it}\,p_{it}}$, which is convex in p_{it}, in which case, by the discussion in Appendix J, $y_{it}\,\theta_{it}^{-1}(y_{it}/x_{it})$ is concave in (x_{it}, y_{it}). Similarly, if $\theta_{it}(p_{it}) = 1-\beta_{it}\,p_{it}$ with $\beta_{it} > 0$, then $1/\theta_{it}(p_{it}) = 1/(1-\beta_{it}\,p_{it})$, which is also convex in p_{it}, in which case, $y_{it}\,\theta_{it}^{-1}(y_{it}/x_{it})$ is concave in (x_{it}, y_{it}) as well. So, there are some choices of $\theta_{it}(\cdot)$ that render the objective function of problem (11) concave in $(\boldsymbol{x}, \boldsymbol{y})$. Since the constraints of this problem are linear in $(\boldsymbol{x}, \boldsymbol{y})$, problem (11) becomes a convex program. Instead of trying to solve problem (10) directly, solving the equivalent formulation in (11) can provide significant benefits in the quality of the solutions that we obtain. In Appendix K, we provide a numerical study to demonstrate such benefits.

Gallego and van Ryzin (1994) use a problem similar to the one in (10) to come up with a heuristic policy and they show a performance guarantee for their heuristic policy. Gallego and van Ryzin (1997) study a similar heuristic policy in the network revenue management setting, where there are a set of resources with finite inventories and the sale of a product consumes a combination of resources. In that setting, we index the resources by $M = \{1, \ldots, m\}$. We have q_ℓ units of resource ℓ at the beginning of the selling horizon. A sale of product i consumes $a_{\ell i}$ units of resource ℓ. In this case, Gallego and van Ryzin (1997) consider a deterministic approximation similar to the one in (10), but the only difference is that the first constraint is replaced by $\sum_{t\in T}\sum_{i\in N} a_{\ell i}\,\theta_{it}(p_{it})\,v_{it}(\boldsymbol{p}_t) \le q_\ell$ for all $\ell \in M$, ensuring that the total expected capacity consumption of each resource does not exceed its availability. In this case, we can also give an equivalent formulation similar to the one in (11). All we need to do is to modify the first constraint in (11) as $\sum_{t\in T}\sum_{i\in N} a_{\ell i}\,y_{it} \le q_\ell$ for all $\ell \in M$. We can use the same proof technique in the proof of Theorem 13 to show that the two formulations are equivalent to each other, even when we have multiple resources.

5 Numerical Experiments

We give numerical experiments to demonstrate the potential benefits from using the Markov chain choice model to capture the customer choice process, instead of simpler models, such as the multinomial logit model. As discussed in the introduction section, there is work by Feldman and Topaloglu (2017), Simsek and Topaloglu (2017) and Wang and Yuan (2017) to demonstrate that it is possible to calibrate the Markov chain choice model in a computationally tractable fashion so that we can predict the customer purchase behavior more accurately when compared with simpler models, such as the multinomial logit model. The work by these authors focuses on the assortment optimization setting, where the prices of the products are fixed and the sets of products that are offered to the customers are decision variables. Our goal is demonstrate that such numerical benefits from using the Markov chain choice model can carry from the assortment optimization to the pricing setting. The focus of our paper is not on estimating the parameters of the Markov chain choice model, so a full set of computational experiments on estimating the parameters of the Markov chain choice model is naturally beyond our scope.

In our numerical setup, we assume that the customer choices are governed by a ground choice model that is not related to the multinomial logit or the Markov chain choice model. In particular, we assume that the customer choices are governed by the exponomial choice model. In this choice model, the utility of product i is $U_i(p_i) = \mu_i - \gamma_i\, p_i - Z_i$, where μ_i and γ_i are fixed parameters and Z_i is an exponential random variable. For notational uniformity, we let $U_0(p_0) = \mu_0 - Z_0$ be the utility of the no purchase option, but there is, naturally, no price for the no purchase option. A customer chooses product i with probability $\mathbb{P}\{U_i(p_i) = \max_{j \in N \cup \{0\}} U_j(p_j)\}$. Alptekinoglu and Semple (2016) propose the exponomial choice model. They show that if the exponential random variables $\{Z_i : i \in N \cup \{0\}\}$ are independent and have the same mean, then there is a closed form expression for the purchase probabilities of the products. In our numerical setup, there are five products. To come up with $\{\mu_i : i \in N\}$, we sample μ_i from the uniform distribution over $[\mu_L, \mu_U]$, where μ_L and μ_U are parameters that we vary. We fix $\mu_0 = \mu_L$. To come up with $\{\gamma_i : i \in N\}$, we set $\gamma_i = \Gamma$ for all $i \in N$, where Γ is another parameter that we vary. The exponential random variables $\{Z_i : i \in N \cup \{0\}\}$ have mean κ, where we also vary the parameter κ. By using the approach in this paragraph, we generate the parameters of the exponomial model that governs the choices of the customers and fix this choice model as the ground choice model.

Once we fix the ground choice model, we generate the purchase history for τ customers. We denote this purchase history by $\{(\hat{\boldsymbol{p}}^\ell, i^\ell) : \ell = 1, \ldots, \tau\}$, where $\hat{\boldsymbol{p}}^\ell = \{\hat{p}_i^\ell : i \in N\}$ are the prices offered to customer ℓ and i^ℓ is the product purchased by customer ℓ. If customer ℓ leaves without a purchase, then we have $i^\ell = 0$. To come up with the prices $\{\hat{\boldsymbol{p}}^\ell : \ell = 1, \ldots, \tau\}$ in the purchase history, we solve the problem $\max_{p_i \in \Re_+} \mathbb{P}\{U_i(p_i) > U_0(p_0)\}\, p_i$, which finds the price that maximizes the expected profit under the exponomial choice model when the customers choose only between product i and the no purchase option. Letting p_i^* be the optimal solution to this problem, we sample $\hat{p}_i^\ell$ from the uniform distribution over $[\frac{1}{2} p_i^*, \frac{3}{2} p_i^*]$. After generating the prices, we generate the choice i^ℓ of customer ℓ in the purchase history according to the ground choice model, given that the customer ℓ is offered the prices $\hat{\boldsymbol{p}}^\ell$. In other words, i^ℓ takes value i with probability $\mathbb{P}\{U_i(\hat{p}_i^\ell) = \max_{j \in N \cup \{0\}} U_j(\hat{p}_j^\ell)\}$. By following the approach described so far in this paragraph, we generate the purchase history for τ customers and use this purchase history as the training data. We vary τ over {1,000, 2,500, 5,000} to capture different levels of data availability. Using the same approach, we also generate the purchase history for another 10,000 customers and use this purchase history as the testing data. We fit our choice models to the training data and test the performance of the fitted choice models on the testing data.

We fit a Markov chain choice model to the training data by using maximum likelihood estimation. To fit a Markov chain choice model, we use the function $\theta_i(p_i) = \mathrm{e}^{-\gamma_i\, p_i}$ in our Markov chain choice model. Letting $\mathbf{1}(\cdot)$ be the indicator function, we estimate the parameters

$\boldsymbol{\lambda} = \{\lambda_i : i \in N\}$, $\boldsymbol{\gamma} = \{\gamma_i : i \in N\}$ and $\boldsymbol{\rho} = \{\rho_{ij} : i, j \in N\}$ of the Markov chain choice model by maximizing the log-likelihood of the training data given by

$$\sum_{\ell=1}^{\tau} \sum_{i \in N} \mathbf{1}(i^\ell = i) \, \log(\theta_i(\hat{p}_i^\ell \,|\, \boldsymbol{\gamma}) \, v_i(\hat{\boldsymbol{p}}^\ell \,|\, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\rho})) + \sum_{\ell=1}^{\tau} \mathbf{1}(i^\ell = 0) \log \Big(1 - \sum_{i \in N} \theta_i(\hat{p}_i^\ell \,|\, \boldsymbol{\gamma}) \, v_i(\hat{\boldsymbol{p}}^\ell \,|\, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\rho})\Big)$$

over $(\boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\rho}) \in \Re_+^{2n+n^2}$ subject to the constraints that $\sum_{i \in N} \lambda_i \leq 1$ and $\sum_{j \in N} \rho_{ij} \leq 1$ for all $i \in N$. In the log-likelihood function above, we make the dependence of $v_i(\boldsymbol{p})$ satisfying (1) on $(\boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\rho})$ and the dependence of $\theta_i(p_i)$ on $\boldsymbol{\gamma}$ explicit. The expression $\theta_i(\hat{p}_i^\ell \,|\, \boldsymbol{\gamma}) \, v_i(\hat{\boldsymbol{p}}^\ell \,|\, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\rho})$ is the purchase probability of product i by customer ℓ. We use the `fmincon` routine in `Matlab` to maximize the log-likelihood function above. We estimate the parameters of the multinomial logit model by maximizing a similar log-likelihood function, but by using the purchase probabilities under the multinomial logit model. In the multinomial logit model, we assume that the mean utility of product i is given by $\sigma_i - \alpha_i \, p_i$ for fixed parameters σ_i and α_i. Once we fit a Markov chain choice model and a multinomial logit model to the training data, we use the testing data to check the ability of these fitted choice models to predict the purchase behavior of the customers that are not in the training data. In particular, for $\mathsf{CM} \in \{\mathsf{MC}, \mathsf{ML}\}$, we use $\mathsf{Choice}_i^{\mathsf{CM}}(\boldsymbol{p})$ to denote the purchase probability of product i under the fitted choice model CM when we charge the prices $\boldsymbol{p}$. Having $\mathsf{CM} = \mathsf{MC}$ corresponds to the Markov chain choice model, whereas having $\mathsf{CM} = \mathsf{ML}$ corresponds to the multinomial logit model. In this case, we compute the out of sample log-likelihood for the fitted choice models on the testing data, which is given by $\sum_{\ell=1}^{10{,}000} \sum_{i \in N} \mathbf{1}(i^\ell = i) \, \log \mathsf{Choice}_i^{\mathsf{CM}}(\tilde{\boldsymbol{p}}^\ell) + \sum_{\ell=1}^{10{,}000} \mathbf{1}(i^\ell = 0) \log(1 - \sum_{i \in N} \log \mathsf{Choice}_i^{\mathsf{CM}}(\tilde{\boldsymbol{p}}^\ell))$ for $\mathsf{CM} \in \{\mathsf{MC}, \mathsf{ML}\}$, where $\{(\tilde{\boldsymbol{p}}^\ell, i^\ell) : \ell = 1, \ldots, 10{,}000\}$ are the prices charged to and the choices of the customers in the testing data. The out of sample log-likelihood is of the same form as the log-likelihood of the training data given above. A larger value for the out of sample log-likelihood indicates that the choice model under consideration more accurately predicts the purchase behavior of the customers that are not in the training data.

We vary (μ_L, μ_U) over $\{(100, 150), (150, 200)\}$, Γ over $\{1, 2\}$ and κ over $\{10, 15\}$ to obtain eight configurations for the ground choice model. For each configuration, we generate the ground choice model as an exponomial choice model as discussed above and fix it. Once we fix the ground choice model, we generate the training and testing data sets from the ground choice model. We use the training data to fit a Markov chain choice model and a multinomial logit model. We check the out of sample log-likelihoods for the fitted choice models on the testing data. In Table 1, we show our results. In this table, the first column shows the configuration for the ground choice model. In the rest of the table, there are three blocks of three columns. Each block corresponds to a different number of customers in the training data, capturing different levels of data availability for estimation. In each block, the first column shows the out of sample log-likelihoods of the fitted Markov chain choice model, the second column shows the out of sample log-likelihoods of the fitted multinomial logit model and the third column shows the percent gap between the two log-likelihoods. In 22 out of 24 cases, the out of sample log-likelihoods of the fitted Markov chain

Ground. Ch. Config. $(\mu_L, \mu_U, \Gamma, \kappa)$	$\tau = 1,000$ Likeli. MC	Likeli. ML	Perc. Gap	$\tau = 2,500$ Likeli. MC	Likeli. ML	Perc. Gap	$\tau = 5,000$ Likeli. MC	Likeli. ML	Perc. Gap
$(100, 150, 1, 10)$	-7,865	-7,837	-0.35	-7,734	-7,767	0.42	-7,723	-7,757	0.45
$(100, 150, 1, 15)$	-11,677	-11,716	0.34	-11,611	-11,687	0.66	-11,571	-11,664	0.81
$(100, 150, 2, 10)$	-9,802	-9,830	0.29	-9,726	-9,818	0.94	-9,703	-9,811	1.12
$(100, 150, 2, 15)$	-9,994	-10,005	0.12	-9,941	-9,971	0.30	-9,926	-9,959	0.33
$(150, 200, 1, 10)$	-9,459	-9,541	0.86	-9,384	-9,480	1.02	-9,374	-9,474	1.06
$(150, 200, 1, 15)$	-11,935	-11,918	-0.15	-11,851	-11,896	0.38	-11,811	-11,881	0.59
$(150, 200, 2, 10)$	-11,596	-11,617	0.18	-11,514	-11,584	0.61	-11,507	-11,586	0.68
$(150, 200, 2, 15)$	-11,208	-11,229	0.18	-11,101	-11,184	0.75	-11,090	-11,182	0.83

Table 1: Out of sample log-likelihoods for the fitted Markov chain choice model and the fitted multinomial logit model on the testing data.

choice model are larger than those of the fitted multinomial logit model. Small differences in log-likelihoods can still correspond to significant differences in the purchase probability predictions, since the logarithmic scale has the effect of shrinking significant differences in the purchase probability predictions. The two cases where the log-likelihoods of the fitted multinomial logit model are larger correspond to the cases with the smallest amount of training data. The multinomial logit model has $O(n)$ parameters given by $\{(\sigma_i, \alpha_i) : i \in N\}$, whereas the Markov chain choice model has $O(n^2)$ parameters given by $\{(\lambda_i, \gamma_i) : i \in N\}$ and $\{\rho_{ij} : i, j \in N\}$. When we have smaller amount of training data, it can be difficult to estimate a larger number of parameters due to overfitting. Our numerical experiments indicate that the Markov chain choice model can provide noticeably better predictions of the purchase behavior of the customers when compared with the multinomial logit model, but we, of course, cannot claim that the Markov chain choice model is universally better than the multinomial logit model. In this section, we compare the fitted choice models using out of sample log-likelihoods. In Appendix L, we compare the fitted choice models using two additional performance measures, which are the errors in the purchase probability and expected revenue predictions of the fitted choice models. We observe similar results.

6 Conclusions

In this paper, we gave a Markov chain choice model to deal with the case where the prices for the products are adjustable and the choice process for the customers reacts to the prices charged for the products. We focused on three fundamental pricing problems under this choice model, which are monopolistic pricing, competitive pricing and dynamic pricing with a single resource. We can give slight generalizations of our results. In particular, we can allow the transition probability $\rho_{ij}(p_i)$ to depend on the price of product i, but not on the price of product j. In this case, our approach in Section 2 to compute the optimal prices and our approach in Section 3.1 to compute the best response still work, but we lose the structural properties of the optimal prices and the prices at an equilibrium. However, even without allowing the transition probabilities to depend on the prices, the Markov chain choice model can provide useful modeling flexibility in capturing the customer purchase behavior. More sophisticated versions of the Markov chain choice model is certainly one avenue for future work. Another avenue for future work is to consider the case where there are constraints on the prices. In quality consistent pricing problems, for example, the prices for the products of higher quality must be larger than the prices for the products of lower quality. When we have constraints that link the prices charged for the different products, our approach for solving pricing problems do not work and extensions in this direction appear to be nontrivial. Also, there has been recent work on estimating the parameters of the Markov chain choice model. Another useful research direction is to use the Markov chain choice model in different application areas to check its benefits in predicting the customer purchase behavior. On the theory side, one can develop computationally efficient methods to estimate the parameters of the Markov chain choice model. On the practice side, one can use the Markov chain choice model to predict the purchase behavior of customers in real data. Both activities will likely help improve the practical appeal of the Markov chain choice model. Lastly, a useful feature of the multinomial logit model is that we can parameterize the mean utilities and the price sensitivities of the products by using a small number of features. In this case, the number of parameters in the multinomial logit model grows linearly with the number of features, instead of the number of products. Also, we do not have to refit the model when a new product is introduced, as long as we know the features of the new product. It would be useful to investigate whether we can characterize the parameters of the Markov chain choice model by using a small number of features.

Acknowledgments. The authors thank the department editor, senior editor and two anonymous referees whose comments improved the contents and the exposition of the paper.

References

Alptekinoglu, A. and Semple, J. H. (2016), 'The exponomial choice model: A new alternative for assortment and price optimization', *Operations Research* **64**(1), 79–93.

Bawa, K. (1990), 'Modeling inertia and variety seeking tendencies in brand choice behavior', *Marketing Science* **9**(3), 263–278.

Blanchet, J., Gallego, G. and Goyal, V. (2016), 'A Markov chain approximation to choice modeling', *Operations Research* **64**(4), 886–905.

Chen, K. D. and Hausman, W. H. (2000), 'Technical note: Mathematical properties of the optimal product line selection problem using choice-based conjoint analysis', *Management Science* **46**(2), 327–332.

Craswell, N., Zoeter, O., Taylor, M. and Ramsey, B. (2008), An experimental comparison of click position-bias models, *in* '*Proceedings of the 2008 International Conference of Web Search and Data Mining*', Palo Alto, CA, pp. 87–94.

Desir, A., Goyal, V., Segev, D. and Ye, C. (2015), Constrained assortment optimization under the Markov chain based choice model, Technical report, Columbia University, New York, NY.

Feldman, J. B. and Topaloglu, H. (2017), 'Revenue management under the Markov chain choice model', *Operations Research* **65**(5), 1322–1342.

Gallego, G., Huh, T., Kang, W. and Phillips, R. (2006), 'Price competition with the attraction demand model: Existence of unique equilibrium and its stability', *Manufacturing & Service Operations Management* **8**(4), 359–375.

Gallego, G., Iyengar, G., Phillips, R. and Dubey, A. (2004), Managing flexible products on a network, Computational Optimization Research Center Technical Report TR-2004-01, Columbia University.

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

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

Gallego, G. and van Ryzin, G. (1997), 'A multiproduct dynamic pricing problem and its applications to yield management', *Operations Research* **45**(1), 24–41.

Gallego, G. and Wang, R. (2014), 'Multi-product price optimization and competition under the nested attraction model', *Operations Research* **62**(2), 450–461.

Gilboa, I. and Pazgal, A. (1995), History dependent brand switching: Theory and evidence, Technical report, Northwestern University, Kellogg School of Management, Evanston, IL.

Givon, M. (1984), 'Variety seeking through brand switching', *Marketing Science* **3**(1), 1–22.

Guo, F., Liu, C. and Wang, Y. M. (2009), Efficient multiple-click models in web search, *in* '*Proceedings of the Second ACM International Conference on Web Search and Data Mining*', Barcelona, Spain, pp. 124–131.

Hanson, W. and Martin, K. (1996), 'Optimizing multinomial logit profit functions', *Management Science* **42**(7), 992–1003.

Jeuland, A. P. (1979), 'Brand choice inertia as one aspect of the notion of brand loyalty', *Management Science* **25**(7), 671–682.

Keller, P. W., Levi, R. and Perakis, G. (2014), 'Efficient formulations for pricing under attraction demand models', *Math. Programming A.* **145**, 223–261.

Li, G., Rusmevichientong, P. and Topaloglu, H. (2015), 'The d-level nested logit model: Assortment and price optimization problems', *Operations Research* **63**(2), 325–352.

Li, H. and Huh, W. T. (2011), 'Pricing multiple products with the multinomial logit and nested models: Concavity and implications', *Manufacturing & Service Operations Management* **13**(4), 549–563.

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

Maglaras, C. and Meissner, J. (2006), 'Dynamic pricing strategies for multiproduct revenue management problems', *Manufacturing & Service Operations Management* **8**(2), 136–148.

Milgrom, P. and Roberts, J. (1990), 'Rationalizability, learning, and equilibrium in games with strategic complementarities', *Econometrica* **58**(6), 1255–1277.

Puterman, M. L. (1994), *Markov Decision Processes*, John Wiley and Sons, Inc., New York.

Simsek, A. S. and Topaloglu, H. (2017), 'Technical note: An expectation-maximization algorithm to estimate the parameters of the Markov chain choice model', *Operations Research* (to appear).

Song, J.-S. and Xue, Z. (2007), Demand management and inventory control for substitutable products, Technical report, Duke University, Durham, NC.

Talluri, K. T. and van Ryzin, G. J. (2005), *The Theory and Practice of Revenue Management*, Springer, New York, NY.

Talluri, K. and van Ryzin, G. (1998), 'An analysis of bid-price controls for network revenue management', *Management Science* **44**(11), 1577–1593.

Topkis, D. M. (1979), 'Equilibrium points in nonzero-sum n-person submodular games', *SIAM Journal on Control and Optimization* **17**(6), 773–787.

Vives, X. (1990), 'Nash equilibrium with strategic complementarities', *Journal of Mathematical Economics* **19**(3), 305–321.

Vossen, T. W. M. and Zhang, D. (2015), 'Reductions of approximate linear programs for network revenue management', *Operations Research* **63**(6), 1352–1371.

Wang, R. (2012), 'Capacitated assortment and price optimization under the multinomial logit model', *Operations Research Letters* **40**, 492–497.

Wang, Z. and Yuan, H. (2017), Technical note: A gradient descent method for estimating the Markov chain choice model, Technical report, University of Minnesota, Minneapolis, MN.

Zhang, D. and Lu, Z. (2013), 'Assessing the value of dynamic pricing in network revenue management', *INFORMS Journal on Computing* **25**(1), 102–115.

A Appendix: Omitted Proofs

In this section, we provide the proofs that are omitted in the paper.

A.1 Proof of Theorem 1

Consider the Markov chain $\{X_t : t \in \mathbb{N}\}$ over the state space $N \cup \{0\}$. Visiting state $i \in N$ corresponds to visiting product i, whereas visiting state 0 corresponds to visiting the no purchase option. The probability law for the Markov chain is $\mathbb{P}\{X_0 = i\} = \lambda_i$ and $\mathbb{P}\{X_{t+1} = j \mid X_t = i\} = \rho_{ij}$, where, for notational brevity, we let $\rho_{00} = 1$ and $\rho_{i0} = 1 - \sum_{j \in N} \rho_{ij}$ for all $i \in N$. Note that state 0 is absorbing. For all $i \in N \cup \{0\}$, we consider the sequence of independent random variables $\{Y_{it}(p_i) : t \in T\}$, each taking value zero or one, where we have $\mathbb{P}\{Y_{it}(p_i) = 1\} = \theta_i(p_i)$ and $\mathbb{P}\{Y_{0t}(p_0) = 1\} = 1$. Having $Y_{it}(p_i) = 1$ corresponds to a customer purchasing product i whenever she visits product i at transition t. Note that the Markov chain $\{X_t : t \in \mathbb{N}\}$ corresponds to the states visited by a customer making a choice under the Markov chain choice model and the sequence of random variables $\{Y_{it}(p_i) : t \in \mathbb{N}\}$ captures the purchase decisions of a customer under the Markov chain choice model whenever she visits product i. Under the Markov chain choice model, for a customer to purchase product i, this product should be the first product the customer visits for which she makes a purchase decision. In particular, we use the random variable $\tau_i(p_i)$ to denote the first visit time for product i when the customer makes a purchase decision for this product. That is, we have $\tau_i(p_i) = \min\{t \in \mathbb{N} : X_t = i \text{ and } Y_{it}(p_i) = 1\}$. If the last set is empty, then we set $\tau_i(p_i) = \infty$. Similarly, we define $\tau_0(p_0) = \min\{t \in \mathbb{N} : X_t = 0 \text{ and } Y_{0t}(p_0) = 1\} = \min\{t \in \mathbb{N} : X_t = 0\}$, since we have $Y_{0t}(p_0) = 1$ with probability one. For a customer to purchase product i, product i needs to be the first product that the customer visits for which she makes a purchase decision. So, the purchase probability of product i is $\mathbb{P}\{\tau_i(p_i) = \arg\min_{j \in N \cup \{0\}} \tau_j(p_j)\}$. If we define the utility of product i as $U_i(p_i) = -\tau_i(p_i)$, then the last expression implies that the purchase probability of product i is $\mathbb{P}\{-\tau_i(p_i) = \arg\max_{j \in N \cup \{0\}} -\tau_j(p_j)\} = \mathbb{P}\{U_i(p_i) = \arg\max_{j \in N \cup \{0\}} U_j(p_j)\}$. Thus, there exist utility random variables $\{U_i(\cdot) : i \in N\}$ such that if we charge the prices $\boldsymbol{p}$, then purchase probability of product i under the Markov chain choice model can be written as $\mathbb{P}\{U_i(p_i) = \max_{j \in N \cup \{0\}} U_j(p_j)\}$, which is the desired result.

A.2 Proof of Lemma 2

For the rank one Markov chain choice model, the system of equations in (1) takes the form $v_i(\boldsymbol{p}) = \lambda_i + \lambda_i \sum_{j \in N} \beta_j \, (1 - \theta_j(p_i)) \, v_j(\boldsymbol{p})$ for all $i \in N$. It is simple to check that the solution to this system of equations is given by $v_i(\boldsymbol{p}) = \lambda_i / (1 - \sum_{j \in N} \lambda_j \, \beta_j \, (1 - \theta_j(p_j))$ for all $i \in N$. In this case, the choice probability of product i is given by $\theta_i(p_i) \, v_i(\boldsymbol{p}) = \lambda_i \, \theta_i(p_i) / (1 - \sum_{j \in N} \lambda_j \, \beta_j \, (1 - \theta_j(p_j))$. The desired result holds since we can use simple algebraic manipulations to verify that the expression in the denominator of the last fraction is equivalently given by $1 - \sum_{j \in N} \lambda_j \, \beta_j \, (1 - \theta_j(p_j)) = 1 - \sum_{j \in N} \lambda_j + \sum_{j \in N} \lambda_j \, (1 - \beta_j) \, (1 - \theta_j(p_j)) + \sum_{j \in N} \lambda_j \, \theta_j(p_j)$.

A.3 Proof of Theorem 3

Assume that $\hat{\boldsymbol{r}} = \{\hat{r}_i : i \in N\}$ satisfies $\hat{r}_i = f_i(\hat{\boldsymbol{r}})$ for all $i \in N$. First, we show that if $\hat{p}_i$ is an optimal solution to problem (4) for all $i \in N$, then $\hat{\boldsymbol{p}} = \{\hat{p}_i : i \in N\}$ is an optimal solution to problem (2). Since $\hat{p}_i$ is an optimal solution to problem (4) for all $i \in N$ and $\hat{r}_i = f_i(\hat{\boldsymbol{r}})$, we have $\hat{r}_i = f_i(\hat{\boldsymbol{r}}) = \theta_i(\hat{p}_i)\,(\hat{p}_i - c_i) + (1 - \theta_i(\hat{p}_i)) \sum_{j \in N} \rho_{ij}\, \hat{r}_j$ for all $i \in N$. Multiplying this equality with $v_i(\hat{\boldsymbol{p}})$ and adding over all $i \in N$, we get $\sum_{i \in N} v_i(\hat{\boldsymbol{p}})\, \hat{r}_i = \sum_{i \in N} \theta_i(\hat{p}_i)\, v_i(\hat{\boldsymbol{p}})\,(\hat{p}_i - c_i) + \sum_{j \in N} \sum_{i \in N} \rho_{ij}\,(1 - \theta_i(\hat{p}_i))\, v_i(\hat{\boldsymbol{p}})\, \hat{r}_j$. Noting that $\sum_{i \in N} \rho_{ij}\,(1 - \theta_i(\hat{p}_i))\, v_i(\hat{\boldsymbol{p}}) = v_j(\hat{\boldsymbol{p}}) - \lambda_j$ by (1), the last equality is equivalent to $\sum_{i \in N} v_i(\hat{\boldsymbol{p}})\, \hat{r}_i = \sum_{i \in N} \theta_i(\hat{p}_i)\, v_i(\hat{\boldsymbol{p}})\,(\hat{p}_i - c_i) + \sum_{j \in N}(v_j(\hat{\boldsymbol{p}}) - \lambda_j)\, \hat{r}_j$, in which case, canceling the terms on both sides yields $\sum_{i \in N} \theta_i(\hat{p}_i)\, v_i(\hat{\boldsymbol{p}})\,(\hat{p}_i - c_i) = \sum_{i \in N} \lambda_i\, \hat{r}_i$. For any $\tilde{\boldsymbol{p}} = \{\tilde{p}_i : i \in N\} \in \times_{i \in N} \mathcal{P}_i$, since $\tilde{p}_i$ is a feasible, but not necessarily an optimal solution to problem (4), we have $\hat{r}_i = f_i(\hat{\boldsymbol{r}}) \geq \theta_i(\tilde{p}_i)\,(\tilde{p}_i - c_i) + (1 - \theta_i(\tilde{p}_i)) \sum_{j \in N} \rho_{ij}\, \hat{r}_j$ for all $i \in N$. Multiplying this inequality with $v_i(\tilde{\boldsymbol{p}})$ and adding over all $i \in N$, we get $\sum_{i \in N} v_i(\tilde{\boldsymbol{p}})\, \hat{r}_i \geq \sum_{i \in N} \theta_i(\tilde{p}_i)\, v_i(\tilde{\boldsymbol{p}})\,(\tilde{p}_i - c_i) + \sum_{j \in N} \sum_{i \in N} \rho_{ij}\,(1 - \theta_i(\tilde{p}_i))\, v_i(\tilde{\boldsymbol{p}})\, \hat{r}_j$. Since $\sum_{i \in N} \rho_{ij}\,(1 - \theta_i(\tilde{p}_i))\, v_i(\tilde{\boldsymbol{p}}) = v_j(\tilde{\boldsymbol{p}}) - \lambda_j$ by (1), the last inequality is equivalent to $\sum_{i \in N} v_i(\tilde{\boldsymbol{p}})\, \hat{r}_i \geq \sum_{i \in N} \theta_i(\tilde{p}_i)\, v_i(\tilde{\boldsymbol{p}})\,(\tilde{p}_i - c_i) + \sum_{j \in N}(v_j(\tilde{\boldsymbol{p}}) - \lambda_j)\, \hat{r}_j$ and we can cancel the terms on both sides to obtain $\sum_{i \in N} \theta_i(\tilde{p}_i)\, v_i(\tilde{\boldsymbol{p}})\,(\tilde{p}_i - c_i) \leq \sum_{i \in N} \lambda_i\, \hat{r}_i$. Thus, we have $\sum_{i \in N} \theta_i(\hat{p}_i)\, v_i(\hat{\boldsymbol{p}})\,(\hat{p}_i - c_i) = \sum_{i \in N} \lambda_i\, \hat{r}_i \geq \sum_{i \in N} \theta_i(\tilde{p}_i)\, v_i(\tilde{\boldsymbol{p}})\,(\tilde{p}_i - c_i)$ for any $\tilde{\boldsymbol{p}} \in \times_{i \in N} \mathcal{P}_i$, which shows that $\hat{\boldsymbol{p}}$ is an optimal solution to problem (2). We observe that the last chain of inequalities also implies that the optimal objective value of problem (2) is $\sum_{i \in N} \lambda_i\, \hat{r}_i$.

Second, we show that if $\hat{\boldsymbol{p}} = \{\hat{p}_i : i \in N\}$ is an optimal solution to problem (2), then $\hat{p}_i$ is an optimal solution to problem (4) for all $i \in N$. Since $\hat{p}_i$ is a feasible, but not necessarily an optimal solution to problem (4), we have $\hat{r}_i = f_i(\hat{\boldsymbol{r}}) \geq \theta_i(\hat{p}_i)\,(\hat{p}_i - c_i) + (1 - \theta_i(\hat{p}_i)) \sum_{j \in N} \rho_{ij}\, \hat{r}_j$ for all $i \in N$. To get a contradiction, assume that $\hat{p}_i$ is not an optimal solution to problem (4) for some $i \in N$ so that the last inequality is strict for some $i \in N$. Since $\lambda_i > 0$, by (1), we have $v_i(\hat{\boldsymbol{p}}) > 0$ for all $i \in N$. If we multiply the last inequality by $v_i(\hat{\boldsymbol{p}})$ and add over all $i \in N$, then noting that $v_i(\hat{\boldsymbol{p}}) > 0$ for all $i \in N$ and the last inequality is strict for some $i \in N$, we obtain $\sum_{i \in N} v_i(\hat{\boldsymbol{p}})\, \hat{r}_i > \sum_{i \in N} \theta_i(\hat{p}_i)\, v_i(\hat{\boldsymbol{p}})\,(\hat{p}_i - c_i) + \sum_{j \in N} \sum_{i \in N} \rho_{ij}\,(1 - \theta_i(\hat{p}_i))\, v_i(\hat{\boldsymbol{p}})\, \hat{r}_j$. As before, since $\sum_{i \in N} \rho_{ij}\,(1 - \theta_i(\hat{p}_i))\, v_i(\hat{\boldsymbol{p}}) = v_j(\hat{\boldsymbol{p}}) - \lambda_j$ by (1), the last inequality is equivalent to $\sum_{i \in N} v_i(\hat{\boldsymbol{p}})\, \hat{r}_i > \sum_{i \in N} \theta_i(\hat{p}_i)\, v_i(\hat{\boldsymbol{p}})\,(\hat{p}_i - c_i) + \sum_{j \in N}(v_j(\hat{\boldsymbol{p}}) - \lambda_j)\, \hat{r}_j$. Canceling the terms on both sides yields $\sum_{i \in N} \theta_i(\hat{p}_i)\, v_i(\hat{\boldsymbol{p}})\,(\hat{p}_i - c_i) < \sum_{i \in N} \lambda_i\, \hat{r}_i$. The desired result follows by noting that the last inequality contradicts the fact that $\hat{\boldsymbol{p}}$ is an optimal solution to problem (2) and the optimal objective value of this problem is given by $\sum_{i \in N} \lambda_i\, \hat{r}_i$.

A.4 Proof of Lemma 4

We use the notation $f_i(\cdot \,|\, \boldsymbol{c})$ to make the dependence of the operator $f_i(\cdot)$ in (3) on the unit costs explicit. Let $\hat{\boldsymbol{r}}(\boldsymbol{c}) = \{\hat{r}_i(\boldsymbol{c}) : i \in N\}$ be such that $\hat{r}_i(\boldsymbol{c}) = f_i(\hat{\boldsymbol{r}}(\boldsymbol{c}) \,|\, \boldsymbol{c})$ for all $i \in N$. We claim that the chain of inequalities $\hat{r}_j(\boldsymbol{c}) - \epsilon \leq \hat{r}_j(\boldsymbol{c} + \epsilon\, \boldsymbol{e}) \leq r_j(c + \epsilon\, \boldsymbol{e}_i) \leq \hat{r}_j(\boldsymbol{c})$ holds for all $j \in N$. For economy of space, we only show that the first inequality in this chain of inequalities holds. The other two inequalities follow from a similar argument. We consider the sequence $\{\tilde{\boldsymbol{r}}(t) : t \in \mathbb{N}\}$

that is generated by using the relationship $\tilde{r}_i(t+1) = f_i(\tilde{\boldsymbol{r}}(t)\,|\,\boldsymbol{c})$ and $\tilde{r}_i(0) = 0$ for all $i \in N$. By the discussion right after (3), the sequence $\{\tilde{\boldsymbol{r}}(t) : t \in \mathbb{N}\}$ converges to $\hat{\boldsymbol{r}}(\boldsymbol{c}) = \{\hat{r}_i(\boldsymbol{c}) : i \in N\}$ that satisfies $\hat{r}_i(\boldsymbol{c}) = f_i(\hat{\boldsymbol{r}}(\boldsymbol{c})\,|\,\boldsymbol{c})$ for all $i \in N$. Similarly, we consider the sequence $\{\bar{\boldsymbol{r}}(t) : t \in \mathbb{N}\}$ that is generated by using the relationship $\bar{r}_i(t+1) = f_i(\bar{\boldsymbol{r}}(t)\,|\,\boldsymbol{c} + \epsilon\,\boldsymbol{e})$ and $\bar{r}_i(0) = 0$ for all $i \in N$. The sequence $\{\bar{\boldsymbol{r}}(t) : t \in \mathbb{N}\}$ converges to $\hat{\boldsymbol{r}}(\boldsymbol{c} + \epsilon\,\boldsymbol{e}) = \{\hat{r}_i(\boldsymbol{c} + \epsilon\,\boldsymbol{e}) : i \in N\}$ that satisfies $\hat{r}_i(\boldsymbol{c}+\epsilon\,\boldsymbol{e}) = f_i(\hat{\boldsymbol{r}}(\boldsymbol{c}+\epsilon\,\boldsymbol{e})\,|\,\boldsymbol{c}+\epsilon\,\boldsymbol{e})$ for all $i \in N$. We use induction to show that $\tilde{r}_i(t) - \epsilon \leq \bar{r}_i(t)$ for all $i \in N$, $t \in \mathbb{N}$. In this case, taking limits on both sides, we obtain $\hat{r}_i(\boldsymbol{c}) - \epsilon \leq \hat{r}_i(c + \epsilon\,\boldsymbol{e})$ for all $i \in N$, as desired. Since $\tilde{r}_i(0) = 0 = \bar{r}_i(0)$, we have $\tilde{r}_i(0) - \epsilon \leq \bar{r}_i(0)$. Assuming that $\tilde{r}_i(t) - \epsilon \leq \bar{r}_i(t)$ for all $i \in N$, we show that $\tilde{r}_i(t+1) - \epsilon \leq \bar{r}_i(t+1)$ for all $i \in N$. We use $\tilde{p}_i$ to denote an optimal solution to problem (3) when we solve this problem after replacing $\boldsymbol{r}$ with $\tilde{\boldsymbol{r}}(t)$, so that the optimal objective value of this problem gives $f_i(\tilde{\boldsymbol{r}}(t)\,|\,\boldsymbol{c})$. Since $\tilde{r}_i(t+1) = f_i(\tilde{\boldsymbol{r}}(t)\,|\,\boldsymbol{c})$, we get

$$\begin{aligned}
\tilde{r}_i(t+1) = f_i(\tilde{\boldsymbol{r}}(t)\,|\,\boldsymbol{c}) &= \theta_i(\tilde{p}_i)\,(\tilde{p}_i - c_i) + (1 - \theta_i(\tilde{p}_i)) \sum_{j \in N} \rho_{ij}\,\tilde{r}_j(t) \\
&\leq \theta_i(\tilde{p}_i)\,(\tilde{p}_i - c_i) + (1 - \theta_i(\tilde{p}_i)) \sum_{j \in N} \rho_{ij}\,(\bar{r}_j(t) + \epsilon) \\
&\leq \theta_i(\tilde{p}_i)\,(\tilde{p}_i - c_i - \epsilon) + (1 - \theta_i(\tilde{p}_i)) \sum_{j \in N} \rho_{ij}\,\bar{r}_j(t) + \epsilon \leq f_i(\bar{\boldsymbol{r}}(t)\,|\,\boldsymbol{c} + \epsilon\,\boldsymbol{e}) + \epsilon = \bar{r}_i(t+1) + \epsilon.
\end{aligned}$$

Here, the first inequality is by the induction assumption. The second inequality is by the fact that $(1 - \theta_i(\tilde{p}_i)) \sum_{j \in N} \rho_{ij} \leq (1 - \theta_i(\tilde{p}_i))$. The third inequality is by the fact that $\tilde{p}_i$ is a feasible, but not necessarily an optimal, solution to problem (3) when we solve this problem with the unit costs $c + \epsilon\,\boldsymbol{e}$. Thus, we have $\tilde{r}_i(t+1) - \epsilon \leq \bar{r}_i(t+1)$, completing the induction argument.

Dropping the constant terms, an optimal solution to problem (3) can be obtained by solving the problem $\max_{p_i \in \mathcal{P}_i} \{\theta_i(p_i)\,(p_i - c_i - \sum_{j \in N} \rho_{ij}\,r_j)\}$. By Theorem 3, $\hat{p}_i(\boldsymbol{c} + \epsilon\,\boldsymbol{e}_i)$ is an optimal solution to problem (3) after replacing c_i with $c_i + \epsilon$ and r_j with $\hat{r}_j(\boldsymbol{c} + \epsilon\,\boldsymbol{e}_i)$. Thus, $\hat{p}_i(\boldsymbol{c} + \epsilon\,\boldsymbol{e}_i)$ is an optimal solution to the problem $\max_{p_i \in \mathcal{P}_i} \{\theta_i(p_i)\,(p_i - c_i - \epsilon - \sum_{j \in N} \rho_{ij}\,\hat{r}_j(\boldsymbol{c} + \epsilon\,\boldsymbol{e}_i))\}$. By a similar argument, $\hat{p}_i(\boldsymbol{c})$ is an optimal solution to the problem $\max_{p_i \in \mathcal{P}_i} \{\theta_i(p_i)\,(p_i - c_i - \sum_{j \in N} \rho_{ij}\,\hat{r}_j(\boldsymbol{c}))\}$. Note that $\epsilon + \sum_{j \in N} \rho_{ij}\,\hat{r}_j(\boldsymbol{c} + \epsilon\,\boldsymbol{e}_i) \geq \epsilon + \sum_{j \in N} \rho_{ij}\,(\hat{r}_j(\boldsymbol{c}) - \epsilon) \geq \sum_{i \in N} \rho_{ij}\,\hat{r}_j(\boldsymbol{c})$, where the first inequality follows from the chain of inequalities established in the previous paragraph. A simple lemma, given as Lemma 19 in Appendix M, shows that the unique optimal solution to the problem $\max_{p_i \in \mathcal{P}_i} \{\theta_i(p_i)(p_i - c_i - x)\}$ is increasing in x. So, since $\epsilon + \sum_{j \in N} \rho_{ij}\,\hat{r}_j(\boldsymbol{c} + \epsilon\,\boldsymbol{e}_i) \geq \sum_{j \in N} \rho_{ij}\,\hat{r}_j(\boldsymbol{c})$, it follows that $\hat{p}_i(\boldsymbol{c} + \epsilon\,\boldsymbol{e}_i) \geq \hat{p}_i(\boldsymbol{c})$, establishing the first inequality in the lemma. The inequalities $\hat{p}_j(\boldsymbol{c} + \epsilon\,\boldsymbol{e}_i) \leq \hat{p}_j(\boldsymbol{c})$ and $\hat{p}_j(\boldsymbol{c} + \epsilon\,\boldsymbol{e}) \geq \hat{p}_j(\boldsymbol{c})$ follow from a similar reasoning.

A.5 Proof of Theorem 6

Fix some $(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m) \in \Re^{n \times m}$ arbitrarily and let $\hat{p}_i$ be an optimal solution to the problem $\max_{p_i \in \mathcal{P}_i} \{\theta_i(p_i)\,(p_i - c_i) + (1 - \theta_i(p_i)) \sum_{j \in N} \rho_{ij}\,\hat{r}_j^k\}$ for all $i \in N^k$, $k \in M$. We claim that having $\hat{r}_i^k = g_i^k(\hat{\boldsymbol{r}}^k\,|\,\hat{\boldsymbol{p}}^{-k})$ for all $i \in N$, $k \in M$ is identical to having $\hat{r}_i^k = h_i^k(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ for all $i \in N$, $k \in M$. To see the claim, if $\hat{r}_i^k = g_i^k(\hat{\boldsymbol{r}}^k\,|\,\hat{\boldsymbol{p}}^{-k})$ for all $i \in N$, $k \in M$, then by (6), it follows that $\hat{r}_i^k = \max_{p_i \in \mathcal{P}_i} \{\theta_i(p_i)\,(p_i - c_i) + (1-\theta_i(p_i)) \sum_{j \in N} \rho_{ij}\, \hat{r}_j^k\}$ for all $i \in N^k$, $k \in M$ and $\hat{r}_i^k = (1-\theta_i(\hat{p}_i)) \sum_{j \in N} \rho_{ij}\, \hat{r}_j^k$ for all $i \notin N^k$, $k \in M$. Noting that $\hat{p}_i$ is defined as an optimal solution to the problem $\max_{p_i \in \mathcal{P}_i} \{\theta_i(p_i)\,(p_i - c_i) + (1-\theta_i(p_i)) \sum_{j \in N} \rho_{ij}\, \hat{r}_j^k\}$ for all $i \in N^k$, $k \in M$, by (7), the last two equalities are precisely the conditions to ensure that $\hat{r}_i^k = h_i^k(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ for all $i \in N$, $k \in M$, which establishes the claim.

We show the desired result in two parts. First, we assume that $\hat{\boldsymbol{p}} = \{\hat{p}_i : i \in N^k,\ k \in M\}$ is a Nash equilibrium. We show that $\hat{p}_i$ is an optimal solution to problem (8) for all $i \in N^k$, $k \in M$, where $(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ satisfies $\hat{r}_i^k = h_i^k(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ for all $i \in N$, $k \in M$. If $\hat{\boldsymbol{p}}$ is a Nash equilibrium, then $\hat{\boldsymbol{p}}^k$ is a best response of firm k to the prices $\hat{\boldsymbol{p}}^{-k}$ for all $k \in M$. In this case, by Lemma 5, $\hat{p}_i$ is an optimal solution to the problem $\max_{p_i \in \mathcal{P}_i} \{\theta_i(p_i)\,(p_i - c_i) + (1-\theta_i(p_i)) \sum_{j \in N} \rho_{ij}\, \hat{r}_j^k\}$ for all $i \in N^k$, $k \in M$, where $\hat{\boldsymbol{r}}^k$ satisfies $\hat{r}_i^k = g_i^k(\hat{\boldsymbol{r}} \,|\, \hat{\boldsymbol{p}}^{-k})$ for all $i \in N$. By the claim in the previous paragraph, having $\hat{p}_i$ be an optimal solution to the problem $\max_{p_i \in \mathcal{P}_i} \{\theta_i(p_i)\,(p_i - c_i) + (1-\theta_i(p_i)) \sum_{j \in N} \rho_{ij}\, \hat{r}_j^k\}$ for all $i \in N^k$, $k \in M$ and $\hat{r}_i^k = g_i^k(\hat{\boldsymbol{r}} \,|\, \hat{\boldsymbol{p}}^{-k})$ for all $i \in N$, $k \in M$ is identical to having $\hat{r}_i^k = h_i^k(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ for all $i \in N$, $k \in M$. Thus, if $\hat{\boldsymbol{p}}$ is a Nash equilibrium, then $\hat{p}_i$ is an optimal solution to the problem $\max_{p_i \in \mathcal{P}_i} \{\theta_i(p_i)\,(p_i - c_i) + (1-\theta_i(p_i)) \sum_{j \in N} \rho_{ij}\, \hat{r}_j^k\}$ for all $i \in N^k$, $k \in M$, where $(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ satisfies $\hat{r}_i^k = h_i^k(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ for all $i \in N$, $k \in M$. Second, we assume that $(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ satisfies $\hat{r}_i^k = h_i^k(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ for all $i \in N$, $k \in M$ and $\hat{p}_i$ is an optimal solution to problem (8) for all $i \in N^k$, $k \in M$. We show that $\hat{\boldsymbol{p}}$ is a Nash equilibrium. By the claim in the previous paragraph, if $\hat{p}_i$ is an optimal solution to the problem $\max_{p_i \in \mathcal{P}_i} \{\theta_i(p_i)\,(p_i - c_i) + (1-\theta_i(p_i)) \sum_{j \in N} \rho_{ij}\, \hat{r}_j^k\}$ for all $i \in N^k$, $k \in M$, then having $\hat{r}_i^k = h_i^k(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ for all $i \in N$, $k \in M$ is identical to having $\hat{r}_i^k = g_i^k(\hat{\boldsymbol{r}}^k \,|\, \hat{\boldsymbol{p}}^{-k})$ for all $i \in N$, $k \in M$. In this case, it follows that $\hat{p}_i$ is an optimal solution to the problem $\max_{p_i \in \mathcal{P}_i} \{\theta_i(p_i)\,(p_i - c_i) + (1-\theta_i(p_i)) \sum_{j \in N} \rho_{ij}\, \hat{r}_j^k\}$ for all $i \in N^k$, $k \in M$, where $\hat{\boldsymbol{r}}^k$ satisfies $\hat{r}_i^k = g_i^k(\hat{\boldsymbol{r}}^k \,|\, \hat{\boldsymbol{p}}^{-k})$ for all $i \in N$, $k \in M$. By Lemma 5, we observe that the last statement is equivalent to the prices $\hat{\boldsymbol{p}}^k$ being the best response to the prices $\hat{\boldsymbol{p}}^{-k}$ for all $k \in M$, which is, in turn, equivalent to the prices $\hat{\boldsymbol{p}}$ being a Nash equilibrium.

A.6 Proof of Theorem 7

Since the prices $\hat{\boldsymbol{p}} = \{\hat{p}_i : i \in N^k, k \in M\}$ are a Nash equilibrium, by Theorem 6, for all $i \in N^k$, $k \in M$, $\hat{p}_i$ is an optimal solution to the problem $\max_{p_i \in \mathcal{P}_i} \{\theta_i(p_i)\,(p_i - c_i) + (1-\theta_i(p_i)) \sum_{j \in N} \rho_{ij}\, \hat{r}_j^k\}$, where $(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ satisfies $\hat{r}_i^k = h_i^k(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ for all $i \in N$, $k \in M$. Dropping the constant terms, $\hat{p}_i$ is also given by an optimal solution to $\max_{p_i \in \mathcal{P}_i} \{\theta_i(p_i)\,(p_i - c_i - \sum_{j \in N} \rho_{ij}\, \hat{r}_j^k)\}$. Similarly, since the prices $\tilde{\boldsymbol{p}} = \{\tilde{p}_i : i \in N\}$ are an optimal solution to problem (2), by Theorem 3, $\tilde{p}_i$ is an optimal solution to the problem $\max_{p_i \in \mathcal{P}_i} \{\theta_i(p_i)\,(p_i - c_i) + (1-\theta_i(p_i)) \sum_{j \in N} \rho_{ij}\, \tilde{r}_j\}$ for all $i \in N$, where $\tilde{\boldsymbol{r}} = \{\tilde{r}_i : i \in N\}$ satisfies $\tilde{r}_i = f_i(\tilde{\boldsymbol{r}})$ for all $i \in N$. Dropping the constant terms, $\tilde{p}_i$ is also given by an optimal solution to $\max_{p_i \in \mathcal{P}_i} \{\theta_i(p_i)\,(p_i - c_i - \sum_{j \in N} \rho_{ij}\, \tilde{r}_j)\}$. By Lemma 19 in Appendix M, the unique optimal solution to the problem $\max_{p_i \in \mathcal{P}_i} \{\theta_i(p_i)\,(p_i - c_i - x)\}$ is increasing in x. So, if we can show that $\hat{r}_i^k \leq \tilde{r}_i$ for all $i \in N$, $k \in M$, then we obtain $\hat{p}_i \leq \tilde{p}_i$, which is the desired result. To show that $\hat{r}_i^k \leq \tilde{r}_i$ for all $i \in N$, $k \in M$, we observe that since $\hat{r}_i^k = h_i^k(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ for all $i \in N$, $k \in M$

and $\hat{p}_i$ is an optimal solution to the problem $\max_{p_i \in \mathcal{P}_i}\{\theta_i(p_i)\,(p_i - c_i) + (1-\theta_i(p_i))\sum_{j\in N}\rho_{ij}\,\hat{r}^k_j\}$ for all $i \in N^k$, $k \in M$, by (7), we have $\hat{r}^k_i = \max_{p_i \in \mathcal{P}_i}\{\theta_i(p_i)\,(p_i - c_i) + (1-\theta_i(p_i))\sum_{j\in N}\rho_{ij}\,\hat{r}^k_j\} = \theta_i(\hat{p}_i)\,(\hat{p}_i - c_i) + (1-\theta_i(\hat{p}_i))\sum_{j\in N}\rho_{ij}\,\hat{r}^k_j$ for all $i \in N^k$, $k \in M$ and $\hat{r}^k_i = (1-\theta_i(\hat{p}_i))\sum_{j\in N}\rho_{ij}\,\hat{r}^k_j$ for all $i \notin N^k$, $k \in M$. Using $\mathbf{1}(\cdot)$ to denote the indicator function, for fixed $i \in N$, if we add the last two equalities over all $k \in M$, then we get

$$
\begin{aligned}
&\sum_{k\in M}\hat{r}^k_i = \sum_{k\in M}\Big\{\mathbf{1}(i\in N^k)\,\hat{r}^k_i + \mathbf{1}(i\notin N^k)\,\hat{r}^k_i\Big\}\\
&= \sum_{k\in M}\left\{\mathbf{1}(i\in N^k)\left\{\theta_i(\hat{p}_i)\,(\hat{p}_i-c_i) + (1-\theta_i(\hat{p}_i))\sum_{j\in N}\rho_{ij}\,\hat{r}^k_j\right\} + \mathbf{1}(i\notin N^k)\left\{(1-\theta_i(\hat{p}_i))\sum_{j\in N}\rho_{ij}\,\hat{r}^k_j\right\}\right\}\\
&= \sum_{k\in M}\Big\{\mathbf{1}(i\in N^k)\,\theta_i(\hat{p}_i)\,(\hat{p}_i-c_i)\Big\} + (1-\theta_i(\hat{p}_i))\sum_{j\in N}\rho_{ij}\left\{\sum_{k\in M}\Big\{\mathbf{1}(i\in N^k)\,\hat{r}^k_j + \mathbf{1}(i\notin N^k)\,\hat{r}^k_j\Big\}\right\}\\
&= \theta_i(\hat{p}_i)\,(\hat{p}_i-c_i) + (1-\theta_i(\hat{p}_i))\sum_{j\in N}\rho_{ij}\left\{\sum_{k\in M}\hat{r}^k_j\right\},
\end{aligned}
$$

where the last equality uses the fact that there is only one firm that owns each product i so that $\sum_{k\in M}\mathbf{1}(i\in N^k)\,\theta_i(\hat{p}_i)\,(\hat{p}_i-c_i) = \theta_i(\hat{p}_i)\,(\hat{p}_i-c_i)$. Letting $\hat{q}_i = \sum_{k\in M}\hat{r}^k_i$ for notational brevity, the chain of equalities above implies that $\hat{q}_i = \theta_i(\hat{p}_i)\,(\hat{p}_i-c_i) + (1-\theta_i(\hat{p}_i))\sum_{j\in N}\rho_{ij}\,\hat{q}_j$ for all $i\in N$. Also, since $\tilde{r}_i = f_i(\tilde{\boldsymbol{r}})$, by (3), we have $\tilde{r}_i = \max_{p_i\in\mathcal{P}_i}\{\theta_i(p_i)\,(p_i-c_i) + (1-\theta_i(p_i))\sum_{j\in N}\rho_{ij}\,\tilde{r}_j\} \geq \theta_i(\hat{p}_i)\,(\hat{p}_i-c_i) + (1-\theta_i(\hat{p}_i))\sum_{j\in N}\rho_{ij}\,\tilde{r}_j$. Subtracting the last equality from the last inequality, we obtain $\tilde{r}_i - \hat{q}_i \geq (1-\theta_i(\hat{p}_i))\sum_{j\in N}\rho_{ij}\,(\tilde{r}_j - \hat{q}_j)$ for all $i\in N$.

We claim that $\tilde{r}_i - \hat{q}_i \geq 0$ for all $i \in N$. To get a contradiction, we let $\Delta = \min_{i\in N}\{\tilde{r}_i - \hat{q}_i\}$ and assume that $\Delta < 0$. Since $\tilde{r}_i - \hat{q}_i \geq \Delta$ for all $i\in N$, we have $\sum_{j\in N}\rho_{ij}\,(\tilde{r}_j - \hat{q}_j) \geq \sum_{j\in N}\rho_{ij}\,\Delta > \Delta$, where the last inequality uses the fact that $\sum_{j\in N}\rho_{ij} < 1$ and $\Delta < 0$. Thus, we have $\sum_{j\in N}\rho_{ij}\,(\tilde{r}_j - \hat{q}_j) > \Delta$ for all $i\in N$. If $\sum_{j\in N}\rho_{ij}\,(\tilde{r}_j - \hat{q}_j) < 0$, then having $\sum_{j\in N}\rho_{ij}\,(\tilde{r}_j - \hat{q}_j) > \Delta$ implies that $(1-\theta_i(\hat{p}_i))\sum_{j\in N}\rho_{ij}\,(\tilde{r}_j - \hat{q}_j) > \Delta$. If, on the other hand, $\sum_{j\in N}\rho_{ij}\,(\tilde{r}_j - \hat{q}_j) \geq 0$, then since $\Delta < 0$, it immediately follows that $(1-\theta_i(\hat{p}_i))\sum_{j\in N}\rho_{ij}\,(\tilde{r}_j - \hat{q}_j) \geq 0 > \Delta$. Therefore, we obtain $(1-\theta_i(\hat{p}_i))\sum_{j\in N}\rho_{ij}\,(\tilde{r}_j - \hat{q}_j) > \Delta$ for all $i\in N$, irrespective of whether $\sum_{j\in N}\rho_{ij}\,(\tilde{r}_j - \hat{q}_j) < 0$ or $\sum_{j\in N}\rho_{ij}\,(\tilde{r}_j - \hat{q}_j) \geq 0$. In this case, noting the inequality at the end of the previous paragraph, we obtain $\tilde{r}_i - \hat{q}_i \geq (1-\theta_i(\hat{p}_i))\sum_{j\in N}\rho_{ij}\,(\tilde{r}_j - \hat{q}_j) > \Delta$ for all $i \in N$, which contradicts the fact that $\Delta = \min_{i\in N}\{\tilde{r}_i - \hat{q}_i\}$ and establishes the claim that $\tilde{r}_i - \hat{q}_i \geq 0$ for all $i\in N$. Therefore, we have $\tilde{r}_i - \hat{q}_i = \tilde{r}_i - \sum_{k\in M}\hat{r}^k_i \geq 0$ for all $i \in N$. By a simple lemma, given as Lemma 20 in Appendix M, if $(\hat{\boldsymbol{r}}^1,\ldots,\hat{\boldsymbol{r}}^m)$ satisfies $\hat{r}^k_i = h^k_i(\hat{\boldsymbol{r}}^1,\ldots,\hat{\boldsymbol{r}}^m)$ for all $i\in N$, $k\in M$, then we must have $\hat{r}^k_i \geq 0$ for all $i\in N$, $k\in M$. Therefore, since $\hat{r}^k_i \geq 0$ for all $i \in N$, $k\in M$, we obtain $\tilde{r}_i - \hat{r}^k_i \geq \tilde{r}_i - \sum_{l\in M}\hat{r}^l_i \geq 0$ for all $i\in N$, $k \in M$, as desired.

A.7 Proof of Theorem 8

Since $\hat{\boldsymbol{p}} = \{\hat{p}_i : i\in N^k,\, k\in M\}$ is a Nash equilibrium, by Theorem 6, $\hat{p}_i$ is an optimal solution to the problem $\max_{p_i\in\mathcal{P}_i}\{\theta_i(p_i)\,(p_i-c_i) + (1-\theta_i(p_i))\sum_{j\in N}\rho_{ij}\,\hat{r}^k_j\}$, where $(\hat{\boldsymbol{r}}^1,\ldots,\hat{\boldsymbol{r}}^m)$ satisfies

$\hat{r}_i^k = h_i^k(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ for all $i \in N$, $k \in M$. In this case, by the definition of $h_i^k(\cdot, \ldots, \cdot)$ in (7), we have $\hat{r}_i^k = \max_{p_i \in \mathcal{P}_i}\{\theta_i(p_i)\,(p_i - c_i) + (1-\theta_i(p_i))\sum_{j\in N}\rho_{ij}\,\hat{r}_j^k\} = \theta_i(\hat{p}_i)\,(\hat{p}_i - c_i) + (1-\theta_i(\hat{p}_i))\sum_{j\in N}\rho_{ij}\,\hat{r}_j^k$ for all $i \in N^k$ and $\hat{r}_i^k = (1-\theta_i(\hat{p}_i))\sum_{j\in N}\rho_{ij}\,\hat{r}_j^k$ for all $i \notin N^k$. Multiplying the last two equalities by $v_i(\hat{\boldsymbol{p}}^k, \hat{\boldsymbol{p}}^{-k})$, adding the first equality over all $i \in N^k$ and adding the second equality over all $i \notin N^k$, we obtain $\sum_{i\in N} v_i(\hat{\boldsymbol{p}}^k, \hat{\boldsymbol{p}}^{-k})\,\hat{r}_i^k = \sum_{i\in N^k} v_i(\hat{\boldsymbol{p}}^k, \hat{\boldsymbol{p}}^{-k})\,\theta_i(\hat{p}_i)\,(\hat{p}_i - c_i) + \sum_{j\in N}\sum_{i\in N}\rho_{ij}\,(1-\theta_i(\hat{p}_i))\,v_i(\hat{\boldsymbol{p}}^k, \hat{\boldsymbol{p}}^{-k})\,\hat{r}_j^k$. By the discussion right before problem (5), we observe that $\{v_i(\hat{\boldsymbol{p}}^k, \hat{\boldsymbol{p}}^{-k}) : i \in N\}$ satisfies $v_i(\hat{\boldsymbol{p}}^k, \hat{\boldsymbol{p}}^{-k}) = \lambda_i + \sum_{j\in N}\rho_{ji}\,(1-\theta_j(\hat{p}_j))\,v_j(\hat{\boldsymbol{p}}^k, \hat{\boldsymbol{p}}^{-k})$ for all $i \in N$, which implies that $\sum_{i\in N}\rho_{ij}\,(1-\theta_i(\hat{p}_i))\,v_i(\hat{\boldsymbol{p}}^k, \hat{\boldsymbol{p}}^{-k}) = v_j(\hat{\boldsymbol{p}}^k, \hat{\boldsymbol{p}}^{-k}) - \lambda_j$. Thus, the last equality is equivalent to $\sum_{i\in N} v_i(\hat{\boldsymbol{p}}^k, \hat{\boldsymbol{p}}^{-k})\,\hat{r}_i^k = \sum_{i\in N^k} v_i(\hat{\boldsymbol{p}}^k, \hat{\boldsymbol{p}}^{-k})\,\theta_i(\hat{p}_i)\,(\hat{p}_i - c_i) + \sum_{j\in N}(v_j(\hat{\boldsymbol{p}}^k, \hat{\boldsymbol{p}}^{-k}) - \lambda_j)\,\hat{r}_j^k$. Canceling the terms on both sides yields

$$\sum_{i\in N^k} v_i(\hat{\boldsymbol{p}}^k, \hat{\boldsymbol{p}}^{-k})\,\theta_i(\hat{p}_i)\,(\hat{p}_i - c_i) = \sum_{i\in N}\lambda_i\,\hat{r}_i^k. \tag{12}$$

Similarly, by the discussion right before the theorem, $\bar{\boldsymbol{p}}_i$ is an optimal solution to the problem $\max_{p_i\in\mathcal{P}_i}\{\theta_i(p_i)\,(p_i - c_i) + (1-\theta_i(p_i))\sum_{j\in N}\rho_{ij}\,\bar{r}_j^k\}$, where $(\bar{\boldsymbol{r}}^1, \ldots, \bar{\boldsymbol{r}}^m)$ is the limit of the sequence $\{(\bar{\boldsymbol{r}}^1(t), \ldots, \bar{\boldsymbol{r}}^m(t)) : t \in \mathbb{N}\}$ and this limit satisfies $\bar{r}_i^k = h_i^k(\bar{\boldsymbol{r}}^1, \ldots, \bar{\boldsymbol{r}}^m)$ for all $i \in N$, $k \in M$. In this case, we can follow the same argument in the previous paragraph to show that $\sum_{i\in N^k} v_i(\bar{\boldsymbol{p}}^k, \bar{\boldsymbol{p}}^{-k})\,\theta_i(\bar{p}_i)\,(\bar{p}_i - c_i) = \sum_{i\in N}\lambda_i\,\bar{r}_i^k$. Therefore, if we can show that $\hat{r}_i^k \le \bar{r}_i^k$ for all $i \in N$, then using (12), we get $\sum_{i\in N^k} v_i(\hat{\boldsymbol{p}}^k, \hat{\boldsymbol{p}}^{-k})\,\theta_i(\hat{p}_i)\,(\hat{p}_i - c_i) = \sum_{i\in N}\lambda_i\,\hat{r}_i^k \le \sum_{i\in N}\lambda_i\,\bar{r}_i^k = \sum_{i\in N^k} v_i(\bar{\boldsymbol{p}}^k, \bar{\boldsymbol{p}}^{-k})\,\theta_i(\bar{p}_i)\,(\bar{p}_i - c_i)$, which is the desired result.

To conclude the proof, we show that $\hat{r}_i^k \le \bar{r}_i^k$ for all $i \in N$, $k \in M$. In Lemma 20 in Appendix M, we show that if $(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ satisfies $\hat{r}_i^k = h_i^k(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ for all $i \in N$, $k \in M$, then we must have $\hat{r}_i^k \le \bar{u}$ for all $i \in N$, $k \in M$, where $\bar{u}$ is as defined right after Theorem 6. So, we have $\hat{r}_i^k \le \bar{u} = \bar{r}_i^k(0)$, where the equality holds since we initialize the sequence $\{(\bar{\boldsymbol{r}}^1(t), \ldots, \bar{\boldsymbol{r}}^m(t)) : t \in \mathbb{N}\}$ as $\bar{r}_i^k(0) = \bar{u}$. Also, if we assume that $\hat{r}_i^k \le \bar{r}_i^k(t)$ for all $i \in N$, $k \in M$, then we obtain $\hat{r}_i^k = h_i^k(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m) \le h_i^k(\bar{\boldsymbol{r}}^1(t), \ldots, \bar{\boldsymbol{r}}^m(t)) = \bar{r}_i^k(t+1)$ for all $i \in N$, $k \in M$, where the first equality uses the fact that $(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ satisfies $\hat{r}_i^k = h_i^k(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ for all $i \in N$, $k \in M$ and the inequality is by the monotonicity of $h_i^k(\cdot, \ldots, \cdot)$ in Lemma 15 in Appendix D and the assumption that $\hat{r}_i^k \le \bar{r}_i^k(t)$ for all $i \in N$, $k \in M$. So, by induction, we get $\hat{r}_i^k \le \bar{r}_i^k(t)$ for all $i \in N$, $k \in M$, $t \in \mathbb{N}$. So, the limit $(\bar{\boldsymbol{r}}_1, \ldots, \bar{\boldsymbol{r}}^m)$ of the sequence $\{(\bar{\boldsymbol{r}}^1(t), \ldots, \bar{\boldsymbol{r}}^m(t)) : t \in \mathbb{N}\}$ also satisfies $\hat{r}_i^k \le \bar{r}_i^k$ for all $i \in N$, $k \in M$.

A.8 Proof of Theorem 9

Consider two systems. In both systems, the set of products is N. In the first system, the set of firms is $M = \{1, \ldots, m\}$. The sets of products owned by these firms are $N^1, \ldots, N^m$. In the second system, the set of firms is $\tilde{M} = \{1, \ldots, m+1\}$. The sets of products owned by these firms are $\tilde{N}^1, \ldots, \tilde{N}^{m+1}$, where we have $\tilde{N}^1 = N^1, \ldots, \tilde{N}^{m-1} = N^{m-1}$ and $\tilde{N}^m \cup \tilde{N}^{m+1} = N^m$. We observe that we have $N^1 \cup \ldots \cup N^m = N = \tilde{N}^1 \cup \ldots \cup \tilde{N}^{m+1}$. For the first system, we consider the sequence $\{(\hat{\boldsymbol{r}}^1(t), \ldots, \hat{\boldsymbol{r}}^m(t)) : t \in \mathbb{N}\}$, where we initialize $(\hat{\boldsymbol{r}}^1(0), \ldots, \hat{\boldsymbol{r}}^m(0)) \in \Re^{n\times m}$ as $\hat{r}_i^k(0) = \bar{u}$ and use the relationship $\hat{r}_i^k(t+1) = h_i^k(\hat{\boldsymbol{r}}^1(t), \ldots, \hat{\boldsymbol{r}}^m(t))$ for all $i \in N$, $k \in M$. Letting $(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ be the limit of the sequence $\{(\hat{\boldsymbol{r}}^1(t), \ldots, \hat{\boldsymbol{r}}^m(t)) : t \in \mathbb{N}\}$, by the discussion right before Theorem 8, for $i \in N^k$, $k \in M$, the price $\hat{p}_i$ in the Pareto dominant equilibrium $\hat{\boldsymbol{p}}$ is given by an optimal solution to the problem $\max_{p_i \in \mathcal{P}_i}\{\theta_i(p_i)\,(p_i - c_i) + (1 - \theta_i(p_i)) \sum_{j \in N} \rho_{ij}\, \hat{r}_j^k\}$. Dropping the constant terms, $\hat{p}_i$ is also given by an optimal solution to $\max_{p_i \in \mathcal{P}_i}\{\theta_i(p_i)\,(p_i - c_i - \sum_{j \in N} \hat{r}_j^k)\}$. Note that since $\lim_{p_i \to U_i} \theta_i(p_i) = 0$ and $\lim_{p_i \to U_i} \theta_i(p_i)\, p_i = 0$, if $r_i^k \geq 0$ for all $i \in N$, $k \in M$, then by (7), $h_i^k(\boldsymbol{r}^1, \ldots, \boldsymbol{r}^m) \geq 0$. Therefore, $\hat{r}_i^k(t) \geq 0$ for all $i \in N$, $k \in M$, $t \in \mathbb{N}$.

For the second system, we consider the sequence $\{(\tilde{\boldsymbol{r}}^1(t), \ldots, \tilde{\boldsymbol{r}}^{m+1}(t)) : t \in \mathbb{N}\}$, where we initialize $(\tilde{\boldsymbol{r}}^1(0), \ldots, \tilde{\boldsymbol{r}}^{m+1}(0)) \in \Re^{n \times (m+1)}$ as $\tilde{r}_i^k(0) = \bar{u}$ and use the relationship $\tilde{r}_i^k(t+1) = h_i^k(\tilde{\boldsymbol{r}}^1(t), \ldots, \tilde{\boldsymbol{r}}^{m+1}(t))$ for all $i \in N$, $k \in \tilde{M}$. Letting $(\tilde{\boldsymbol{r}}^1, \ldots, \tilde{\boldsymbol{r}}^{m+1})$ be the limit of the sequence $\{(\tilde{\boldsymbol{r}}^1(t), \ldots, \tilde{\boldsymbol{r}}^{m+1}(t)) : t \in \mathbb{N}\}$, for $i \in \tilde{N}^k$, $k \in \tilde{M}$, the price $\tilde{p}_i$ in the Pareto dominant equilibrium $\tilde{\boldsymbol{p}}$ is given by an optimal solution to $\max_{p_i \in \mathcal{P}_i}\{\theta_i(p_i)\,(p_i - c_i - \sum_{i \in N} \rho_{ij}\, \tilde{r}_j^k)\}$. As above, we also have $\tilde{r}_i^k(t) \geq 0$ for all $i \in N$, $k \in \tilde{M}$. By Lemma 19 in Appendix M, the optimal solution to the problem $\max_{p_i \in \mathcal{P}_i}\{\theta_i(p_i)\,(p_i - c_i - x)\}$ is increasing in x. Therefore, if we can show that $\hat{r}_i^k \geq \tilde{r}_i^k$ for all $i \in N$, $k \in \{1, \ldots, m-1\}$ and $\hat{r}_i^m \geq \max\{\tilde{r}_i^m, \tilde{r}_i^{m+1}\}$ for all $i \in N$, then comparing the problems that we solve to obtain $\hat{p}_i$ and $\tilde{p}_i$, it follows that $\hat{p}_i \geq \tilde{p}_i$ for all $i \in N$, which is the desired result. In the rest of the proof, we use induction to show that $\hat{r}_i^k(t) \geq \tilde{r}_i^k(t)$ for all $i \in N$, $k \in \{1, \ldots, m-1\}$, $t \in \mathbb{N}$ and $\hat{r}_i^m(t) \geq \max\{\tilde{r}_i^m(t), \tilde{r}_i^{m+1}(t)\}$ for all $i \in N$, $t \in \mathbb{N}$, in which case, the limits of the sequences $\{(\hat{\boldsymbol{r}}^1(t), \ldots, \hat{\boldsymbol{r}}^m(t)) : t \in \mathbb{N}\}$ and $\{(\tilde{\boldsymbol{r}}^1(t), \ldots, \tilde{\boldsymbol{r}}^{m+1}(t)) : t \in \mathbb{N}\}$ satisfy the inequalities $\hat{r}_i^k \geq \tilde{r}_i^k$ for all $i \in N$, $k \in \{1, \ldots, m-1\}$ and $\hat{r}_i^m \geq \max\{\tilde{r}_i^m, \tilde{r}_i^{m+1}\}$ for all $i \in N$ as well.

We have $\theta_i(p_i)\,(p_i - c_i - x) \geq 0$ in the optimal solution to $\max_{p_i \in \mathcal{P}_i}\{\theta_i(p_i)\,(p_i - c_i - x)\}$, since noting that $\lim_{p_i \to U_i} \theta_i(p_i) = 0$ and $\lim_{p_i \to U_i} \theta_i(p_i)\, p_i = 0$, we can always make the objective value arbitrarily close to zero. Moving on to the induction argument, since $\hat{r}_i^k(0) = \bar{u}$ and $\tilde{r}_i^k(0) = \bar{u}$, we have $\hat{r}_i^k(0) \geq \tilde{r}_i^k(0)$ for all $i \in N$, $k \in \{1, \ldots, m-1\}$ and $\hat{r}_i^m(0) \geq \max\{\tilde{r}_i^m(0), \tilde{r}_i^{m+1}(0)\}$ for all $i \in N$. Assuming that $\hat{r}_i^k(t) \geq \tilde{r}_i^k(t)$ for all $i \in N$, $k \in \{1, \ldots, m-1\}$ and $\hat{r}_i^m(t) \geq \max\{\tilde{r}_i^m(t), \tilde{r}_i^{m+1}(t)\}$ for all $i \in N$, we proceed to showing that $\hat{r}_i^k(t+1) \geq \tilde{r}_i^k(t+1)$ for all $i \in N$, $k \in \{1, \ldots, m-1\}$ and $\hat{r}_i^m(t+1) \geq \max\{\tilde{r}_i^m(t+1), \tilde{r}_i^{m+1}(t)\}$ for all $i \in N$. For all $i \in N^k$, $k \in M$, we let $\hat{p}_i(t)$ be an optimal solution to the problem $\max_{p_i \in \mathcal{P}_i}\{\theta_i(p_i)\,(p_i - c_i) + (1 - \theta_i(p_i)) \sum_{j \in N} \rho_{ij}\, \hat{r}_j^k(t)\}$. For all $i \in \tilde{N}^k$, $k \in \tilde{M}$, we use $\tilde{p}_i(t)$ be an optimal solution to the problem $\max_{p_i \in \mathcal{P}_i}\{\theta_i(p_i)\,(p_i - c_i) + (1 - \theta_i(p_i)) \sum_{j \in N} \rho_{ij}\, \tilde{r}_j^k(t)\}$. Since $\hat{r}_i^k(t) \geq \tilde{r}_i^k(t)$ for all $i \in N$, $k \in \{1, \ldots, m-1\}$ and $\hat{r}_i^m(t) \geq \max\{\tilde{r}_i^m(t), \tilde{r}_i^{m+1}(t)\}$ for all $i \in N$, by the same reasoning at the end of the previous paragraph, it holds that $\hat{p}_i(t) \geq \tilde{p}_i(t)$ for all $i \in N^k$, $k \in M$. We consider the four cases given below.

First, consider the case $i \notin N^k$, $k \in \{1, \ldots, m-1\}$. By (7), we have the chain of inequalities $\hat{r}_i^k(t+1) = h_i^k(\hat{\boldsymbol{r}}^1(t), \ldots, \hat{\boldsymbol{r}}^m(t)) = (1 - \theta_i(\hat{p}_i(t)) \sum_{j \in N} \rho_{ij}\, \hat{r}_j^k(t) \geq (1 - \theta_i(\tilde{p}_i(t)) \sum_{j \in N} \rho_{ij}\, \tilde{r}_j^k(t) = h_i^k(\tilde{\boldsymbol{r}}^1(t), \ldots, \tilde{\boldsymbol{r}}^m(t)) = \tilde{r}_i^k(t+1)$, where the inequality follows from the fact that $\hat{p}_i(t) \geq \tilde{p}_i(t)$ and $\theta_i(\cdot)$ is decreasing and noting the induction assumption. Second, consider the case $i \in N^k$, $k \in \{1, \ldots, m-1\}$. By (7), $\hat{r}_i^k(t+1) = \max_{p_i \in \mathcal{P}_i}\{\theta_i(p_i)\,(p_i - c_i) + (1 - \theta_i(p_i)) \sum_{j \in N} \rho_{ij}\, \hat{r}_j^k(t)\} \geq \max_{p_i \in \mathcal{P}_i}\{\theta_i(p_i)\,(p_i - c_i) + (1 - \theta_i(p_i)) \sum_{j \in N} \rho_{ij}\, \tilde{r}_j^k(t)\} = \tilde{r}_i^k(t+1)$, where the inequality is by the induction assumption. Third, consider the case $i \notin N^m$. Since $N^m = \tilde{N}^m \cup \tilde{N}^{m+1}$, having $i \notin N^m$ implies that $i \notin \tilde{N}^m$ and $i \notin \tilde{N}^{m+1}$. So, by (7), we obtain

$$\hat{r}_i^m(t+1) = (1 - \theta_i(\hat{p}_i(t)) \sum_{j \in N} \rho_{ij}\, \hat{r}_j^m(t) \geq (1 - \theta_i(\tilde{p}_i(t)) \sum_{j \in N} \rho_{ij} \max\{\tilde{r}_j^m(t), \tilde{r}_j^{m+1}(t))\}$$

$$\geq \max\left\{(1 - \theta_i(\tilde{p}_i(t)) \sum_{j \in N} \rho_{ij}\, \tilde{r}_j^m(t), (1 - \theta_i(\tilde{p}_i(t)) \sum_{j \in N} \rho_{ij}\, \tilde{r}_j^{m+1}(t)\right\} = \max\{\tilde{r}_i^m(t+1), \tilde{r}_i^{m+1}(t+1)\},$$

where the first inequality is by the induction assumption and the fact that $\hat{p}_i(t) \geq \tilde{p}_i(t)$. Fourth, consider the case $i \in N^m$. For the moment, assume that $i \in \tilde{N}^m$ and $i \notin \tilde{N}^{m+1}$. So, by (7), since $i \in \tilde{N}^m$, we have $\tilde{r}_i^m(t+1) = \max_{p_i \in \mathcal{P}_i}\{\theta_i(p_i)\,(p_i - c_i) + (1 - \theta_i(p_i)) \sum_{j \in N} \rho_{ij}\, \tilde{r}_j^m(t)\}$ and $\tilde{p}_i(t)$ is an optimal solution to the last problem. By the discussion at the beginning of the previous paragraph, we have $\theta_i(\tilde{p}_i(t))\,(\tilde{p}_i(t) - c_i) \geq \theta_i(\tilde{p}_i(t))\,(\tilde{p}_i(t) - c_i - \sum_{j \in N} \rho_{ij}\, \tilde{r}_j^m(t)) \geq 0$ Also, by (7), since $i \notin \tilde{N}^{m+1}$, we have $\tilde{r}_i^{m+1}(t+1) = (1 - \theta_i(\tilde{p}_i(t))) \sum_{j \in N} \rho_{ij}\, \tilde{r}_j^{m+1}(t)$. Thus, we get

$$\begin{aligned}
&\max\{\tilde{r}_i^m(t+1), \tilde{r}_i^{m+1}(t+1)\} \\
&\quad = \max\left\{\theta_i(\tilde{p}_i(t))\,(\tilde{p}_i(t) - c_i) + (1 - \theta_i(\tilde{p}_i(t))) \sum_{j \in N} \rho_{ij}\, \tilde{r}_j^m(t),\ (1 - \theta_i(\tilde{p}_i(t))) \sum_{j \in N} \rho_{ij}\, \tilde{r}_j^{m+1}(t)\right\} \\
&\quad \leq \theta_i(\tilde{p}_i(t))\,(\tilde{p}_i(t) - c_i) + (1 - \theta_i(\tilde{p}_i(t))) \sum_{j \in N} \rho_{ij}\, \max\{\tilde{r}_j^m(t), \tilde{r}_j^{m+1}(t)\} \\
&\quad \leq \max_{p_i \in \mathcal{P}_i}\left\{\theta_i(p_i)\,(p_i - c_i) + (1 - \theta_i(p_i)) \sum_{j \in N} \rho_{ij}\, \hat{r}_j^m(t)\right\} = \hat{r}_i^m(t+1),
\end{aligned}$$

where the first inequality is by the fact that $\theta_i(\tilde{p}_i(t))\,(\tilde{p}_i(t) - c_i) \geq 0$, the second inequality is by the induction assumption and the last equality is by the fact that $i \in N^m$ so that we can compute $\hat{r}_i^m(t+1)$ by using the first case in (7). If $i \notin \tilde{N}^m$ and $i \in \tilde{N}^{m+1}$, we can repeat the same argument to show that $\max\{\tilde{r}_i^m(t+1), \tilde{r}_i^{m+1}(t+1)\} \leq \hat{r}_i^m(t+1)$. The four cases above collectively show that $\hat{r}_i^k(t+1) \geq \tilde{r}_i^k(t+1)$ for all $i \in N$, $k \in \{1, \ldots, m-1\}$ and $\hat{r}_i^m(t+1) \geq \max\{\tilde{r}_i^m(t+1), \tilde{r}_i^{m+1}(t+1)\}$ for all $i \in N$, which completes the induction argument.

A.9 Proof of Lemma 10

Fixing some firm $\ell \in M$, we let product $q \in N^\ell$ be the only product that firm ℓ owns. Throughout the proof, we use the notation $h_i^k(\cdot, \ldots, \cdot \,|\, \boldsymbol{c})$ to make the dependence of the operator $h_i^k(\cdot, \ldots, \cdot)$ in (7) on unit costs explicit. We generate the sequence $\{(\hat{\boldsymbol{r}}^1(t), \ldots, \hat{\boldsymbol{r}}^m(t)) : t \in \mathbb{N}\}$ by using the relationship $\hat{r}_i^k(t+1) = h_i^k(\hat{\boldsymbol{r}}^1(t), \ldots, \hat{\boldsymbol{r}}^m(t) \,|\, \boldsymbol{c})$ for all $i \in N$, $k \in M$ and starting with the initial condition that $\hat{r}_i^k(0) = \bar{u}$ for all $i \in N$, $k \in M$. By the discussion right before Theorem 8, the sequence $\{(\hat{\boldsymbol{r}}^1(t), \ldots, \hat{\boldsymbol{r}}^m(t)) : t \in \mathbb{N}\}$ has a limit. Furthermore, using $(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ to denote this limit, given that the unit costs are $\boldsymbol{c}$, for all $i \in N^k$, $k \in M$, the price $\hat{p}_i$ in the Pareto dominant equilibrium $\hat{\boldsymbol{p}}$ is given by an optimal solution to the problem $\max_{p_i \in \mathcal{P}_i}\{\theta_i(p_i)\,(p_i - c_i) + (1 - \theta_i(p_i)) \sum_{j \in N} \rho_{ij}\, \hat{r}_j^k\}$. Thus, dropping the constant terms,

$\hat{p}_i$ is also given by an optimal solution to the problem $\max_{p_i \in \mathcal{P}_i}\{\theta_i(p_i)\,(p_i - c_i - \sum_{j\in N}\rho_{ij}\,\hat{r}_j^k)\}$. Similarly, we generate the sequence $\{(\tilde{\boldsymbol{r}}^1(t),\ldots,\tilde{\boldsymbol{r}}^m(t)) : t \in \mathbb{N}\}$ by using the relationship $\tilde{r}_i^k(t+1) = h_i^k(\tilde{\boldsymbol{r}}^1(t),\ldots,\tilde{\boldsymbol{r}}^m(t)\,|\,\boldsymbol{c} + \epsilon\,\boldsymbol{e}_q)$ for all $i \in N$, $k \in M$ and starting with the initial condition that $\tilde{r}_i^k(0) = \bar{u}$ for all $i \in N$, $k \in M$. The sequence $\{(\tilde{\boldsymbol{r}}^1(t),\ldots,\tilde{\boldsymbol{r}}^m(t)) : t \in \mathbb{N}\}$ has a limit. Using $(\tilde{\boldsymbol{r}}^1,\ldots,\tilde{\boldsymbol{r}}^m)$ to denote this limit, given that the unit costs are $\boldsymbol{c} + \epsilon\,\boldsymbol{e}_q$, for all $i \in N^k$, $k \in M$, the price $\tilde{p}_i$ in the Pareto dominant equilibrium $\tilde{\boldsymbol{p}}$ is given by an optimal solution to the problem $\max_{p_i\in\mathcal{P}_i}\{\theta_i(p_i)\,(p_i - c_i - \mathbf{1}(i=q)\,\epsilon) + (1-\theta_i(p_i))\sum_{j\in N}\rho_{ij}\,\tilde{r}_j^k\}$. Since each firm owns one product, considering the fixed firm ℓ and the product q that this firm owns, we have $i = q$ in the last problem if and only if $k = \ell$, in which case, we can replace the expression $\mathbf{1}(i=q)$ with $\mathbf{1}(k=\ell)$. Thus, dropping the constant terms in the last problem, $\tilde{p}_i$ is also given by an optimal solution to the problem $\max_{p_i\in\mathcal{P}_i}\{\theta_i(p_i)\,(p_i - c_i - \mathbf{1}(k=\ell)\,\epsilon - \sum_{j\in N}\rho_{ij}\,\tilde{r}_j^k)\}$. By Lemma 19 in Appendix M, the unique optimal solution to the problem $\max_{p_i\in\mathcal{P}_i}\{\theta_i(p_i)\,(p_i - c_i - x)\}$ is increasing in x. Therefore, if we can show that $\hat{r}_i^k - \mathbf{1}(k=\ell)\,\epsilon \le \tilde{r}_i^k$ for all $i \in N$, $k \in M$, then we obtain

$$\sum_{j\in N}\rho_{ij}\,\hat{r}_j^k \le \sum_{j\in N}\rho_{ij}\,(\tilde{r}_j^k + \mathbf{1}(k=\ell)\,\epsilon) \le \sum_{j\in N}\rho_{ij}\,\tilde{r}_j^k + \mathbf{1}(k=\ell)\,\epsilon.$$

In this case, comparing the problems that we solve to obtain $\hat{p}_i$ and $\tilde{p}_i$, it follows that $\tilde{p}_i \ge \hat{p}_i$ for all $i \in N$, which is the desired result. In the rest of the proof, we use induction to show that the sequences $\{(\hat{\boldsymbol{r}}^1(t),\ldots,\hat{\boldsymbol{r}}^m(t)) : t\in\mathbb{N}\}$ and $\{(\tilde{\boldsymbol{r}}^1(t),\ldots,\tilde{\boldsymbol{r}}^m(t)) : t\in\mathbb{N}\}$ satisfy $\hat{r}_i^k(t) - \mathbf{1}(k=\ell)\,\epsilon \le \tilde{r}_i^k(t)$ for all $i\in N$, $k \in M$, $t \in \mathbb{N}$, in which case, the limits $(\hat{\boldsymbol{r}}^1,\ldots,\hat{\boldsymbol{r}}^m)$ and $(\tilde{\boldsymbol{r}}^1,\ldots,\tilde{\boldsymbol{r}}^m)$ of these sequences satisfy $\hat{r}_i^k - \mathbf{1}(k=\ell)\,\epsilon \le \tilde{r}_i^k$ for all $i \in N$, $k\in M$ as well.

Since $\hat{r}_i^k(0) = \bar{u} = \tilde{r}_i^k(0)$ for all $i\in N$, $k\in M$, we get $\hat{r}_i^k(0) - \mathbf{1}(k=\ell)\,\epsilon \le \tilde{r}_i^k(0)$. Assuming that $\hat{r}_i^k(t) - \mathbf{1}(k=\ell)\,\epsilon \le \tilde{r}_i^k(t)$ for all $i\in N$, $k\in M$, we proceed to showing that $\hat{r}_i^k(t+1) - \mathbf{1}(k=\ell)\,\epsilon \le \tilde{r}_i^k(t+1)$ for all $i \in N$, $k \in M$. We consider two cases. First, consider the case $i \in N^k$, $k \in M$. Since $\hat{r}_i^k(t+1) = h_i^k(\hat{\boldsymbol{r}}^1(t),\ldots,\hat{\boldsymbol{r}}^m(t)\,|\,\boldsymbol{c})$ and $\tilde{r}_i^k(t+1) = h_i^k(\tilde{\boldsymbol{r}}^1(t),\ldots,\tilde{\boldsymbol{r}}^m(t)\,|\,\boldsymbol{c}+\epsilon\,\boldsymbol{e}_q)$, noting the first case in (7), for all $i\in N^k$, $k \in M$, we obtain

$$\begin{aligned}
\hat{r}_i^k(t+1) &= \max_{p_i\in\mathcal{P}_i}\left\{\theta_i(p_i)\,(p_i - c_i) + (1-\theta_i(p_i))\sum_{j\in N}\rho_{ij}\,\hat{r}_j^k(t)\right\}\\
&\le \max_{p_i\in\mathcal{P}_i}\left\{\theta_i(p_i)\,(p_i - c_i) + (1-\theta_i(p_i))\sum_{j\in N}\rho_{ij}\,(\tilde{r}_j^k(t) + \mathbf{1}(k=\ell)\,\epsilon)\right\}\\
&\le \max_{p_i\in\mathcal{P}_i}\left\{\theta_i(p_i)\,(p_i - c_i - \mathbf{1}(k=\ell)\,\epsilon) + (1-\theta_i(p_i))\sum_{j\in N}\rho_{ij}\,\tilde{r}_j^k(t)\right\} + \mathbf{1}(k=\ell)\,\epsilon\\
&= \tilde{r}_i^k(t+1) + \mathbf{1}(k=\ell)\,\epsilon,
\end{aligned}$$

where the first inequality is by the induction assumption, the second inequality is by the fact that $\sum_{j\in N}\rho_{ij} < 1$ so that $(1-\theta_i(p_i))\sum_{j\in N}\rho_{ij}\,\mathbf{1}(k=\ell)\,\epsilon \le (1-\theta_i(p_i))\,\mathbf{1}(k=\ell)\,\epsilon$ and the second equality is by the fact that $\tilde{r}_i^k(t+1) = h_i^k(\tilde{\boldsymbol{r}}^1(t),\ldots,\tilde{\boldsymbol{r}}^m(t)\,|\,\boldsymbol{c}+\epsilon\,\boldsymbol{e}_q)$. Therefore, we have $\hat{r}_i^k(t+1) - \mathbf{1}(k=\ell)\,\epsilon \le \tilde{r}_i^k(t+1)$ for all $i\in N^k$, $k\in M$. Second, consider the case $i \notin N^k$, $k\in M$. In the chain of inequalities above, we let $\hat{p}_i(t)$ be an optimal solution to the first maximization problem and $\tilde{p}_i(t)$ be an optimal solution to the third maximization problem. Using precisely the same reasoning at the end of the previous paragraph, but replacing $\hat{r}_i^k$ and $\tilde{r}_i^k$ with $\hat{r}_i^k(t)$ and $\tilde{r}_i^k(t)$ and noting the induction assumption that $\hat{r}_i^k(t) - \mathbf{1}(k=\ell) \leq \tilde{r}_i^k(t)$, we obtain $\tilde{p}_i(t) \geq \hat{p}_i(t)$. Therefore, since $\hat{r}_i^k(t+1) = h_i^k(\hat{\boldsymbol{r}}^1(t), \ldots, \hat{\boldsymbol{r}}^m(t) \,|\, \boldsymbol{c})$ and $\tilde{r}_i^k(t+1) = h_i^k(\tilde{\boldsymbol{r}}^1(t), \ldots, \tilde{\boldsymbol{r}}^m(t) \,|\, \boldsymbol{c} + \epsilon\, \boldsymbol{e}_q)$, noting the second case in (7), for all $i \notin N^k$, $k \in M$, we obtain the chain of inequalities

$$\begin{aligned}\hat{r}_i^k(t+1) = (1-\theta_i(\hat{p}_i(t))) \sum_{j\in N} \rho_{ij}\, \hat{r}_j^k(t) &\leq (1-\theta_i(\hat{p}_i(t))) \sum_{j\in N} \rho_{ij}\, (\tilde{r}_j^k(t) + \mathbf{1}(k=\ell)\,\epsilon) \\ &\leq (1-\theta_i(\tilde{p}_i(t))) \sum_{j\in N} \rho_{ij}\, \tilde{r}_j^k(t) + \mathbf{1}(k=\ell)\,\epsilon = \tilde{r}_i^k(t+1) + \mathbf{1}(k=\ell)\,\epsilon,\end{aligned}$$

where the first inequality is by the induction assumption and the second inequality is by the fact that $\tilde{p}_i(t) \geq \hat{p}_i(t)$, in which case, we have $1-\theta_i(\tilde{p}_i(t)) \geq 1-\theta_i(\hat{p}_i(t))$, along with the fact that $(1-\theta_i(\hat{p}_i(t))) \sum_{j\in N} \rho_{ij} \leq 1$. Therefore, we have $\hat{r}_i^k(t+1) - \mathbf{1}(k=\ell)\,\epsilon \leq \tilde{r}_i^k(t+1)$ for all $i \notin N^k$, $k \in M$ as well, competing the induction argument. Putting the two cases together, it follows that $\hat{r}_i^k(t) - \mathbf{1}(k=\ell)\,\epsilon \leq \tilde{r}_i^k(t)$ for all $i \in N$, $k \in M$, $t \in \mathbb{N}$.

A.10 Proof of Theorem 11

Consider any $(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m) \in \Re^{n\times m}$ satisfying $\hat{r}_i^k = h_i^k(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ for all $i \in N$, $k \in M$. We claim that if $\bar{r}_i^k = \underline{r}_i^k$ for all $i \in N$, $k \in M$, then we have $\bar{r}_i^k = \hat{r}_i^k = \underline{r}_i^k$. In particular, Lemma 20 in Appendix M shows that if $(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ satisfies $\hat{r}_i^k = h_i^k(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ for all $i \in N$, $k \in M$, then we must have $\bar{u} \geq \hat{r}_i^k \geq 0$ for all $i \in N$, $k \in M$. Thus, since $\bar{r}_i^k(0) = \bar{u}$ and $\underline{r}_i^k(0) = 0$, we have $\bar{r}_i^k(0) \geq \hat{r}_i^k \geq \underline{r}_i^k(0)$ for all $i \in N$, $k \in M$. In Lemma 15 in Appendix D, we show that the operator $h_i^k(\cdot, \ldots, \cdot)$ is monotone. Thus, applying this operator on all sides of the last inequality, we get $\bar{r}_i^k(1) = h_i^k(\bar{\boldsymbol{r}}^1(0), \ldots, \bar{\boldsymbol{r}}^m(0)) \geq h_i^k(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m) \geq h_i^k(\underline{\boldsymbol{r}}^1(0), \ldots, \underline{\boldsymbol{r}}^m(0)) = \underline{r}_i^k(1)$. Noting that $\hat{r}_i^k = h_i^k(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$, the last inequality yields $\bar{r}_i^k(1) \geq \hat{r}_i^k \geq \underline{r}_i^k(1)$ for all $i \in N$, $k \in M$. Repeating the same argument, we get $\bar{r}_i^k(t) \geq \hat{r}_i^k \geq \underline{r}_i^k(t)$ for all $i \in N$, $k \in M$, $t \in \mathbb{N}$. Since the limits of the sequences $\{(\bar{\boldsymbol{r}}^1(t), \ldots, \bar{\boldsymbol{r}}^m(t)) : t \in \mathbb{N}\}$ and $\{(\underline{\boldsymbol{r}}^1(t), \ldots, \underline{\boldsymbol{r}}^m(t)) : t \in \mathbb{N}\}$ are respectively $(\bar{\boldsymbol{r}}^1, \ldots, \bar{\boldsymbol{r}}^m)$ and $(\underline{\boldsymbol{r}}^1, \ldots, \underline{\boldsymbol{r}}^m)$, we get $\bar{r}_i^k \geq \hat{r}_i^k \geq \underline{r}_i^k$ for all $i \in N$, $k \in M$, establishing the claim. By the claim that we just established, if $\bar{r}_i^k = \underline{r}_i^k$ for all $i \in N$, $k \in M$, then there exists a unique $(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ that satisfies $\hat{r}_i^k = h_i^k(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ for all $i \in N$, $k \in M$ and this unique $(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ is given by the common value of $(\bar{\boldsymbol{r}}^1, \ldots, \bar{\boldsymbol{r}}^m)$ and $(\underline{\boldsymbol{r}}^1, \ldots, \underline{\boldsymbol{r}}^m)$.

By Theorem 6, if the prices $\hat{\boldsymbol{p}} = \{\hat{p}_i : i \in N^k,\ k \in M\}$ are a Nash equilibrium, then $\hat{p}_i$ must be an optimal solution to the problem $\max_{p_i \in \mathcal{P}_i} \{\theta_i(p_i)\,(p_i - c_i) + (1-\theta_i(p_i)) \sum_{j\in N} \rho_{ij}\, \hat{r}_j^k\}$ for all $i \in N^k$, $k \in M$, where $(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ satisfies $\hat{r}_i^k = h_i^k(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ for all $i \in N$, $k \in M$. Dropping the constant terms, an optimal solution to the last optimization problem is also given by an optimal solution to the problem $\max_{p_i \in \mathcal{P}_i} \{\theta_i(p_i)\,(p_i - c_i - \sum_{j\in N} \rho_{ij}\, \hat{r}_j^k)\}$. Since $\theta_i(p_i)\,(p_i - x)$ is strictly quasiconcave in p_i for any $x \in \Re$, there exists a unique solution to this problem. In this case, the desired result follows from the fact that if $\bar{r}_i^k = \underline{r}_i^k$ for all $i \in N$, $k \in M$, then there exists a unique $(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ that satisfies $\hat{r}_i^k = h_i^k(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ for all $i \in N$, $k \in M$.

A.11 Proof of Lemma 12

We can follow a standard argument in the revenue management literature to show that if the value functions $\{V_t(\cdot) : t \in T\}$ are computed through the dynamic program in (9), then we have $\Delta V_t(x) \geq \Delta V_t(x+1)$ for all $x = 1, \ldots, q-1$, $t \in T$; see the proof of Proposition 2-2.A.4 in Talluri and van Ryzin (2005) for the proof technique. Noting (9), the prices $\hat{\boldsymbol{p}}_t(x+1)$ are an optimal solution to the problem $\max_{\boldsymbol{p} \in \times_{i \in N} \mathcal{P}_{it}} \{\sum_{i \in N} \theta_{it}(p_i)\, v_{it}(\boldsymbol{p})\, (p_i - \Delta V_{t+1}(x+1))\}$. Defining the cost vector $\boldsymbol{c} = \Delta V_{t+1}(x+1)\, \boldsymbol{e}$, using the notation in Lemma 4, the optimal solution to this problem is denoted by $\hat{\boldsymbol{p}}(\boldsymbol{c})$. Also, the prices $\hat{\boldsymbol{p}}_t(x)$ are an optimal solution to the problem

$$\max_{\boldsymbol{p} \in \times_{i \in N} \mathcal{P}_{it}} \left\{ \sum_{i \in N} \theta_{it}(p_i)\, v_{it}(\boldsymbol{p})\, (p_i - \Delta V_{t+1}(x)) \right\}$$
$$= \max_{\boldsymbol{p} \in \times_{i \in N} \mathcal{P}_{it}} \left\{ \sum_{i \in N} \theta_{it}(p_i)\, v_{it}(\boldsymbol{p})\, (p_i - (\Delta V_{t+1}(x+1) + \Delta V_{t+1}(x) - \Delta V_{t+1}(x+1)) \right\}.$$

Letting $\epsilon = \Delta V_{t+1}(x) - \Delta V_{t+1}(x+1)$, we have $\epsilon \geq 0$ since $\Delta V_{t+1}(x) \geq \Delta V_{t+1}(x+1)$. Letting the unit cost vector $\boldsymbol{c} = \Delta V_{t+1}(x+1)\, \boldsymbol{e}$ as above and using the notation in Lemma 4, the optimal solution to the problem above is denoted by $\hat{\boldsymbol{p}}(\boldsymbol{c} + \epsilon\, \boldsymbol{e})$. Thus, by Lemma 4, we have $\hat{p}_i(\boldsymbol{c} + \epsilon\, \boldsymbol{e}) \geq \hat{p}_i(\boldsymbol{c})$, which implies that $\hat{p}_{it}(x) \geq \hat{p}_{it}(x+1)$. On the other hand, by using the approach in the proof of Proposition 2-2.A.4 in Talluri and van Ryzin (2005), we can also show that $\Delta V_t(x) \geq \Delta V_{t+1}(x)$ for all $x = 1, \ldots, q$, $t \in T$, in which case, we can follow a reasoning similar to the one that we used earlier in this proof to show that $\hat{p}_{it}(x) \geq \hat{p}_{i,t+1}(x)$.

A.12 Proof of Theorem 13

First, given an optimal solution to problem (11), we use the transformation in the theorem to construct a feasible solution to problem (10) such that the objective values provided by the two solutions match. Assume that $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ is an optimal solution to problem (11). We define the solution $\hat{\boldsymbol{p}}$ for problem (10) as $\hat{p}_{it} = \theta_{it}^{-1}(\hat{y}_{it}/\hat{x}_{it})$ for all $i \in N$, $t \in T$. Since $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ is a feasible solution to problem (11), we have $\hat{x}_{it} = \lambda_{it} + \sum_{j \in N} \rho_{jit}\, (1 - \hat{y}_{jt}/\hat{x}_{jt})\, \hat{x}_{jt}$ for all $i \in N$. Noting that $\hat{p}_{it} = \theta_{it}^{-1}(\hat{y}_{it}/\hat{x}_{it})$, we have $\hat{y}_{it}/\hat{x}_{it} = \theta_{it}(\hat{p}_{it})$, in which case, we write the last equality as $\hat{x}_{it} = \lambda_{it} + \sum_{j \in N} \rho_{jit}\, (1 - \theta_{it}(\hat{p}_{it}))\, \hat{x}_{jt}$ for all $i \in N$. Thus, $\{\hat{x}_{it} : i \in N\}$ solves the system of equations in (1) under the prices $\hat{\boldsymbol{p}}_t = \{\hat{p}_{it} : i \in N\}$. As discussed in Section 1, this system of equations has a unique solution. Thus, it must be the case that $\hat{x}_{it} = v_{it}(\hat{\boldsymbol{p}}_t)$ for all $i \in N$, where $v_{it}(\hat{\boldsymbol{p}}_t)$ is the expected number of times a customer visits product i under the Markov chain choice model with parameters $\{\lambda_{it} : i \in N\}$, $\{\theta_{it}(\cdot) : i \in N\}$ and $\{\rho_{ijt} : i, j \in N\}$ when we charge the prices $\hat{\boldsymbol{p}}_t$. Also, since $\hat{y}_{it}/\hat{x}_{it} = \theta_{it}(\hat{p}_{it})$, we have $\hat{y}_{it} = \theta_{it}(\hat{p}_{it})\, v_{it}(\hat{\boldsymbol{p}}_t)$. In this case, the optimal objective value of problem (11) is given by $\sum_{t \in T} \sum_{i \in N} \hat{y}_{it}\, \theta_{it}^{-1}(\hat{y}_{it}/\hat{x}_{it}) = \sum_{t \in T} \sum_{i \in N} \theta_{it}(\hat{p}_{it})\, v_{it}(\hat{\boldsymbol{p}}_t)\, \hat{p}_{it}$. Therefore, the solution $\hat{\boldsymbol{p}}$ provides the same objective value for problem (10) as does the solution $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ for problem (11). To see that the solution $\hat{\boldsymbol{p}}$ is feasible for problem (10), note that $\sum_{t \in T} \sum_{i \in N} \theta_{it}(\hat{p}_{it})\, v_{it}(\boldsymbol{p}_t) = \sum_{t \in T} \sum_{i \in N} \hat{y}_{it} \leq q$, where the inequality follows from the fact that

$(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ is a feasible solution to problem (11). Since $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ is a feasible solution to problem (11), we have $\theta_{it}^{-1}(\hat{y}_{it}/\hat{x}_{it}) \in \mathcal{P}_{it}$, which implies that $\hat{p}_{it} \in \mathcal{P}_{it}$. Thus, $\hat{\boldsymbol{p}}$ is feasible to problem (10).

Second, given an optimal solution to problem (10), we use the transformation in the theorem to construct a feasible solution to problem (11) such that the objective values provided by the two solutions match. Assume that $\hat{\boldsymbol{p}}$ is an optimal solution to problem (10). We define the solution $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ to problem (11) as $\hat{x}_{it} = v_{it}(\hat{\boldsymbol{p}}_t)$ and $\hat{y}_{it} = \theta_{it}(\hat{p}_{it})\, v_{it}(\hat{\boldsymbol{p}}_t)$, where $v_{it}(\hat{\boldsymbol{p}}_t)$ is as defined in the previous paragraph. In this case, we have $\theta_{it}(\hat{p}_{it}) = \hat{y}_{it}/\hat{x}_{it}$, which implies that $\theta_{it}^{-1}(\hat{y}_{it}/\hat{x}_{it}) = \hat{p}_{it}$. Thus, the optimal objective value of problem (10) is given by $\sum_{t\in T}\sum_{i\in N}\theta_{it}(\hat{p}_{it})\, v_{it}(\hat{\boldsymbol{p}}_t)\,\hat{p}_{it} = \sum_{t\in T}\sum_{i\in N}\hat{y}_{it}\,\theta_{it}^{-1}(\hat{y}_{it}/\hat{x}_{it})$. Therefore, the solution $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ provides the same objective value for problem (11) as does the solution $\hat{\boldsymbol{p}}$ for problem (10). To see that the solution $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ is feasible to problem (11), note that $\sum_{t\in T}\sum_{i\in N}\hat{y}_{it} = \sum_{t\in T}\sum_{i\in N}\theta_{it}(\hat{p}_{it})\, v_{it}(\hat{\boldsymbol{p}}_t) \le q$, where the inequality follows from the fact that $\hat{\boldsymbol{p}}$ is a feasible solution to problem (10). Similarly, since $\hat{\boldsymbol{p}}$ is a feasible solution to problem (10), we have $\hat{p}_{it} \in \mathcal{P}_{it}$, which implies that $\theta_{it}^{-1}(\hat{y}_{it}/\hat{x}_{it}) \in \mathcal{P}_{it}$. Lastly, the expected number of visits to product i under the prices $\hat{\boldsymbol{p}}_t$ satisfies the system of equations in (1), so that $v_{it}(\hat{\boldsymbol{p}}_t) = \lambda_{it} + \sum_{j\in N}\rho_{jit}\,(1-\theta_{jt}(\hat{p}_{jt}))\, v_{jt}(\hat{\boldsymbol{p}}_t)$. By the definitions of $\hat{x}_{it}$ and $\hat{y}_{it}$, this equality is equivalent to $\hat{x}_{it} = \lambda_{it} + \sum_{j\in N}\rho_{jit}\,(\hat{x}_{jt} - \hat{y}_{jt})$. Thus, $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ is feasible to problem (11), in which case, the desired result follows.

B Appendix: Contraction Property

In the next lemma, we show that the operator $\{f_i(\cdot) : i \in N\} : \Re^n \to \Re^n$ is a contraction. This result is used in Section 2.

Lemma 14 *For any $\boldsymbol{r} = \{r_i : i \in N\} \in \Re^n$ and $\boldsymbol{q} = \{q_i : i \in N\} \in \Re^n$, we have $|f_i(\boldsymbol{r}) - f_i(\boldsymbol{q})| \le \max_{i\in N}\{\sum_{j\in N}\rho_{ij}\}\|\boldsymbol{r}-\boldsymbol{q}\|$ for all $i \in N$.*

Proof. By (3), we have $f_i(\boldsymbol{r}) = \max_{p_i\in\mathcal{P}_i}\{\theta_i(p_i)\,(p_i - c_i) + (1-\theta_i(p_i))\,\sum_{j\in N}\rho_{ij}\, r_j\}$. Letting $\hat{p}_i$ be an optimal solution to this problem, we obtain $f_i(\boldsymbol{r}) = \theta_i(\hat{p}_i)\,(\hat{p}_i - c_i) + (1-\theta_i(\hat{p}_i))\,\sum_{j\in N}\rho_{ij}\, r_j$. Also, we have $f_i(\boldsymbol{q}) = \max_{p_i\in\mathcal{P}_i}\{\theta_i(p_i)\,(p_i - c_i) + (1-\theta_i(p_i))\,\sum_{j\in N}\rho_{ij}\, q_j\}$. As $\hat{p}_i$ is a feasible, but not necessarily an optimal, solution to this problem, we get $f_i(\boldsymbol{q}) \ge \theta_i(\hat{p}_i)\,(\hat{p}_i - c_i) + (1-\theta_i(\hat{p}_i))\,\sum_{j\in N}\rho_{ij}\, q_j$. In this case, we obtain $f_i(\boldsymbol{r}) - f_i(\boldsymbol{q}) \le (1-\theta_i(\hat{p}_i))\,\sum_{j\in N}\rho_{ij}\,(r_j - q_j) \le (1-\theta_i(\hat{p}_i))\,\sum_{j\in N}\rho_{ij}\,\|\boldsymbol{r}-\boldsymbol{q}\| \le \max_{i\in N}\{\sum_{j\in N}\rho_{ij}\}\,\|\boldsymbol{r}-\boldsymbol{q}\|$. Repeating the same argument after interchanging the role of $\boldsymbol{r}$ and $\boldsymbol{q}$, we obtain $f_i(\boldsymbol{q}) - f_i(\boldsymbol{r}) \le \max_{i\in N}\{\sum_{j\in N}\rho_{ij}\}\,\|\boldsymbol{r}-\boldsymbol{q}\|$ as well. Thus, we have $|f_i(\boldsymbol{r}) - f_i(\boldsymbol{q})| \le \max_{i\in N}\{\sum_{j\in N}\rho_{ij}\}\|\boldsymbol{r}-\boldsymbol{q}\|$. □

C Appendix: An Alternative Model with Common Products

At the end of Section 3.1, we discuss the assumption that different firms own different partitions of the products. Although different firms own different partitions of the products, two products i

and j that are owned by two different firms may still, for example, correspond to the same soap sold by two different supermarket chains. In our model, if the two supermarket chains charge different prices for this soap, then both supermarket chains can garner demand for it. Naturally, if one supermarket chain decreases the price that it charges for this soap, then the probability that a customer purchases the soap from the other supermarket chain decreases. This approach is consistent with the one in Gallego et al. (2006), Li and Huh (2011) and Gallego and Wang (2014). In this section, we discuss an alternative model where if the same soap is sold by two supermarket chains, then there is only one product corresponding to this soap in the Markov chain choice model. If a customer visits this product and she decides to purchase, then she decides which supermarket chain to purchase from, if any, based on the prices charged by the two supermarket chains.

We consider the case with two firms, but we can extend the model to the case with more than two firms. We index the firms by $M = \{1, -1\}$ and the products by $N = \{1, \dots, n\}$. For notational uniformity, we follow the convention that all products are offered by both firms, but if a customer visits a certain product during the course of her choice process, then the probability that the customer decides to purchase this product from one of the firms may be zero irrespective of the prices. In this case, the product in question is effectively offered only by the other firm. Let $\boldsymbol{p} = \{p_i^k : i \in N,\ k \in M\}$ be the prices charged for the products, where p_i^k is the price charged for product i by firm k. Since all products are offered by both firms, we have a price for each product and for each firm. In contrast, we have a single price for each product in Section 3.1.

The set of feasible prices that firm k can charge for product i is $\mathcal{P}_i^k = [L_i^k, U_i^k]$. With probability λ_i, a customer arriving into the system visits product i. A customer visiting product i decides to purchase this product from firm k with probability $\theta_i^k(p_i^k, p_i^{-k})$. If product i is a product that is not offered by firm k, then it must be the case that $\theta_i^k(p_i^k, p_i^{-k}) = 0$ and $\theta_i^{-k}(p_i^{-k}, p_i^k)$ is independent of p_i^k. With probability $1 - \theta_i^k(p_i^k, p_i^{-k}) - \theta_i^{-k}(p_i^{-k}, p_i^k)$, a customer visiting product i decides not to purchase this product, in which case, she transitions to product j with probability ρ_{ij}. If a customer visiting product i decides not to purchase this product, then she transitions to the no purchase option with probability $1 - \sum_{j \in N} \rho_{ij}$. Therefore, the parameters of the Markov chain choice model are $\{\lambda_i : i \in N\}$, $\{\theta_i^k(\cdot, \cdot) : i \in N,\ k \in M\}$ and $\{\rho_{ij} : i, j \in N\}$.

We use $\boldsymbol{p}^k = \{p_i^k : i \in N\}$ to denote the prices charged by firm k. Given that firm k charges the prices $\boldsymbol{p}^k$ and the other firm charges the prices $\hat{\boldsymbol{p}}^{-k}$, we use $v_i(\boldsymbol{p}^k, \hat{\boldsymbol{p}}^{-k})$ to denote the expected number of times that a customer visits product i during the course of her choice process. In this case, $\{v_i(\boldsymbol{p}^k, \hat{\boldsymbol{p}}^{-k}) : i \in N\}$ satisfies a slightly modified version of the system of equations in (1), where we have $v_i(\boldsymbol{p}^k, \hat{\boldsymbol{p}}^{-k}) = \lambda_i + \sum_{j \in N} \rho_{ji}\,(1 - \theta_j^k(p_j^k, \hat{p}_j^{-k}) - \theta_j^{-k}(\hat{p}_j^{-k}, p_j^k))\, v_j(\boldsymbol{p}^k, \hat{\boldsymbol{p}}^{-k})$ for all $i \in N$. Thus, if the prices charged by both firms are given by $(\boldsymbol{p}^k, \hat{\boldsymbol{p}}^{-k})$, then a customer purchases product i from firm k with probability $\theta_i^k(p_i^k, \hat{p}_i^{-k})\, v_i(\boldsymbol{p}^k, \hat{\boldsymbol{p}}^{-k})$. Thus, firm k computes its best response to the prices $\hat{\boldsymbol{p}}^{-k}$ charged by firm $-k$ by maximizing $\sum_{i \in N} \theta_i^k(p_i^k, \hat{p}_i^{-k})\, v_i(\boldsymbol{p}^k, \hat{\boldsymbol{p}}^{-k})\,(p_i^k - c_i)$ over $\boldsymbol{p}^k \in \times_{i \in N} \mathcal{P}_i^k$. In this expression, as before, c_i is the unit cost of product i. We use $\hat{r}_i^k$ to denote the optimal expected profit of firm k from a customer visiting product i, given that the other firm charges the prices $\hat{\boldsymbol{p}}^{-k}$. By using a dynamic programming argument similar to the one in Section 3.1, if firm k charges the price p_i^k for product i and the other firm charges the prices $\hat{\boldsymbol{p}}^{-k}$ for the products, then the optimal expected profit that firm k obtains from a customer visiting product i is given by $\theta_i^k(p_i^k, \hat{p}_i^{-k})\,(p_i^k - c_i) + (1 - \theta_i^k(p_i^k, \hat{p}_i^{-k}) - \theta_i^{-k}(\hat{p}_i^{-k}, p_i^k)) \sum_{j \in N} \rho_{ij}\, \hat{r}_j^k$. Therefore, firm k can find the price of product i in the best response to the prices $\hat{\boldsymbol{p}}^{-k}$ charged by the other firm by maximizing the last expression over $p_i^k \in \mathcal{P}_i^k$. To write this problem succinctly, we define the operator $g_i^k(\cdot \,|\, \hat{\boldsymbol{p}}^{-k}) : \Re^n \to \Re$ as

$$g_i^k(\boldsymbol{r}^k \,|\, \hat{\boldsymbol{p}}^{-k}) \;=\; \max_{p_i^k \in \mathcal{P}_i^k} \left\{ \theta_i^k(p_i^k, \hat{p}_i^{-k})\,(p_i^k - c_i) + (1 - \theta_i^k(p_i^k, \hat{p}_i^{-k}) - \theta_i^{-k}(\hat{p}_i^{-k}, p_i^k)) \sum_{j \in N} \rho_{ij}\, r_j^k \right\}. \qquad (13)$$

In this case, $\hat{\boldsymbol{r}}^k = \{\hat{r}_i^k : i \in N\}$ needs to satisfy $\hat{r}_i^k = g_i^k(\hat{\boldsymbol{r}}^k \,|\, \hat{\boldsymbol{p}}^{-k})$ for all $i \in N$. We can show that the operator $\{g_i^k(\cdot \,|\, \hat{\boldsymbol{p}}^{-k}) : i \in N\} : \Re^n \to \Re^n$ is a contraction. Therefore, we can use precisely the same approach right before Lemma 5 to compute the best response of firm k to the prices $\hat{\boldsymbol{p}}^{-k}$ charged by firm $-k$. All we need to do is to find the value of $\hat{\boldsymbol{r}}^k = \{\hat{r}_i^k : i \in N\}$ that satisfies $\hat{r}_i^k = g_i^k(\hat{\boldsymbol{r}}^k \,|\, \hat{\boldsymbol{p}}^{-k})$ for all $i \in N$ and solve the problem on the right side above after replacing r_j^k with $\hat{r}_j^k$. The approach discussed in this paragraph provides a tractable way to compute the best response of firm k to the prices $\hat{\boldsymbol{p}}^{-k}$ charged by firm $-k$.

Although it is tractable to compute the best response of one firm to the prices charged by the other, characterizing the structural properties of the Nash equilibrium under the model that we develop in this section is difficult. The difficulty is that if there are products that are offered by both firms, then it is not clear how to define an operator similar to the operator $h_i^k(\cdot, \ldots, \cdot)$ in (7) to characterize a Nash equilibrium. In the operator in (7), different firms own different partitions of the products. In this case, to compute $h_i^k(\boldsymbol{r}^1, \ldots, \boldsymbol{r}^m)$ for all $i \in N$, $k \in M$, we can solve the maximization problem in the first case in (7) to compute $\hat{p}_i$ for all $i \in N^k$, $k \in M$. In this way, we obtain a price for each product. Furthermore, the optimal objective value of the maximization problem in the first case yields $h_i^k(\boldsymbol{r}^1, \ldots, \boldsymbol{r}^m)$ for all $i \in N^k$, $k \in M$ as well. Once we have the price $\hat{p}_i$ for all $i \in N^k$, $k \in M$, we can use these prices in the second case in (7) to compute $h_i^k(\boldsymbol{r}^1, \ldots, \boldsymbol{r}^m)$ for all $i \notin N^k$, $k \in M$. In other words, when applying the operator in (7), since different firms own different partitions of the products, we can compute the prices charged by each firm for the products that it offers. In the model that we develop in this section, however, since the probability that a customer visiting product i purchases this product depends on the prices charged by both firms, it is not clear how to define an analogue of the operator $h_i^k(\cdot, \ldots, \cdot)$. One can show the existence of a Nash equilibrium by appealing to general results that exploit the continuity of the payoff function for each firm and the compactness of the action space, but it is difficult to come up with structural properties for the Nash equilibrium similar to those in Section 3.3, which compare the prices in a Nash equilibrium as the market gets more competitive and give comparative statistics for the prices as a function of the unit costs.

Nevertheless, we can still use the operator in (13) to numerically try to reach a Nash equilibrium by successively computing the best response of each firm to the other. One interesting observation is that depending on the problem parameters, the two firms may or may not be simultaneously in the market for the common products in a Nash equilibrium. In other words, using $\hat{\boldsymbol{p}} = \{\hat{p}_i^k : i \in N,\ k \in M\}$ to denote the prices charged by the two firms in a Nash equilibrium, in some problem instances, we may have some product i such that $\theta_i^k(\hat{p}_i^k, \hat{p}_i^{-k}) > 0$ and $\theta_i^{-k}(\hat{p}_i^{-k}, \hat{p}_i^k) > 0$, indicating that both firms garner demand for product i in a Nash equilibrium. In some other problem instances, we may have some product i such that $\theta_i^k(\hat{p}_i^k, \hat{p}_i^{-k}) > 0$ and $\theta_i^{-k}(\hat{p}_i^{-k}, \hat{p}_i^k) = 0$, indicating that firm k garners demand for product i in a Nash equilibrium, but not firm $-k$. To give a specific example, consider a problem instance where we have two firms and three products. The unit costs are zero. The parameters of the underlying Markov chain choice model are

$$\lambda_1 = \lambda_2 = \lambda_3 = \frac{1}{3},$$
$$\theta_1^1(p_1^1, p_1^{-1}) = 1 - p_1^1, \quad \theta_1^{-1}(p_1^{-1}, p_1^1) = 0,$$
$$\theta_2^1(p_2^1, p_2^{-1}) = \frac{1}{2} - \frac{1}{4}p_2^1, \quad \theta_2^{-1}(p_2^{-1}, p_2^1) = \frac{1}{2} - \frac{1}{4}p_2^{-1},$$
$$\theta_3^1(p_3^1, p_3^{-1}) = 0, \quad \theta_3^{-1}(p_3^{-1}, p_3^1) = 1 - \frac{1}{10}p_3^{-1},$$
$$\rho_{11} = 0.1, \quad \rho_{12} = 0, \quad \rho_{13} = 0.8,$$
$$\rho_{21} = 0, \quad \rho_{22} = 0.4, \quad \rho_{23} = 0.4,$$
$$\rho_{31} = 0.1, \quad \rho_{32} = 0, \quad \rho_{33} = 0.8.$$

The set of feasible prices for each product i and for each firm k are those that ensure that the function $\theta_i^k(\cdot, \cdot)$ takes nonnegative values. Note that firm -1 cannot garner demand for product 1, whereas firm 1 cannot garner demand for product 3. Thus, these products are, in essence, offered by one firm. In contrast, both firms can garner demand for product 2. For this problem instance, considering the prices $\hat{\boldsymbol{p}}^1 = (0.533, 1.086, 0)$ charged by firm 1 and the prices $\hat{\boldsymbol{p}}^{-1} = (0, 2, 7.176)$ charged by firm -1, the prices $\hat{\boldsymbol{p}}^k$ are a best response to the prices $\hat{\boldsymbol{p}}^{-k}$ for all $k \in \{1, -1\}$. Therefore, these prices form a Nash equilibrium. In this Nash equilibrium, we have $\theta_2^1(\hat{p}_2^1, \hat{p}_2^{-1}) = 0.229$, but $\theta_2^{-1}(\hat{p}_2^{-1}, \hat{p}_2^1) = 0$, indicating that firm 1 garners demand for product 2 in the Nash equilibrium, whereas firm -1 shuts off the demand for product 2 by charging a high price.

To see the intuitive reason for firm -1 to shut off the demand for product 2, since we have $\theta_3^{-1}(p_3^{-1}, p_3^1) = 1 - \frac{1}{10}p_3^{-1}$, note that firm -1 can increase the price of product 3 up to 10. In other words, firm -1 has access to a product whose demand is somewhat price insensitive. By shutting off the demand for product 2, firm -1 tries to minimize the probability that a customer visiting product 2 purchases this product. In this case, if the customer visiting product 2 happens to transition from product 2 to product 3, then firm -1 has a chance to obtain a relatively large profit from this customer. Therefore, firm -1 does not have to rely on product 2 to maximize its expected profit. In contrast, firm 1 does not have access to a product whose demand is as price insensitive and it relies on product 2 to maximize its expected profit.

In Figure 2, we numerically verify the Nash equilibrium. Given that firm $-k$ charges the prices $\hat{\boldsymbol{p}}^{-k}$, we use $\pi^k(\boldsymbol{p}^k \,|\, \hat{\boldsymbol{p}}^{-k})$ to denote the expected profit of firm k as a function of the prices $\boldsymbol{p}^k$ that

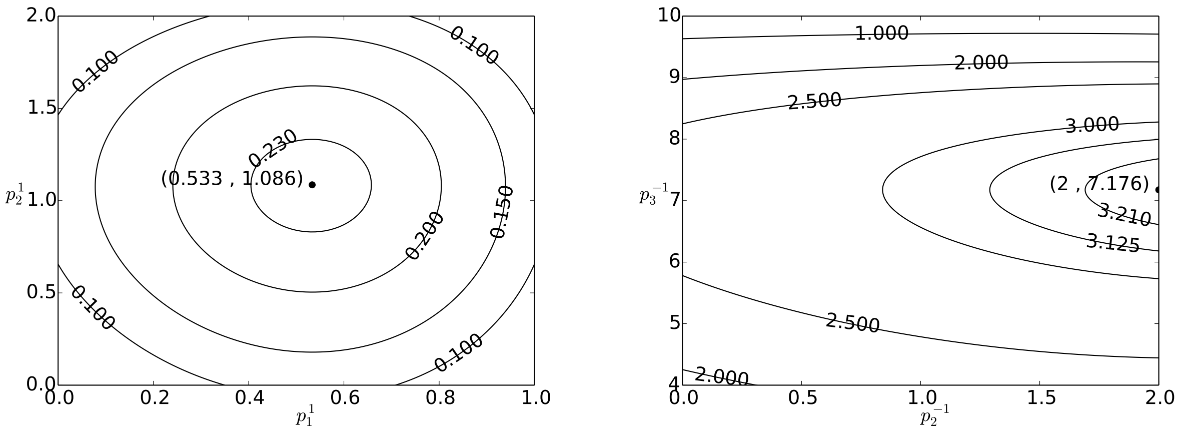

Figure 2: Expected profits of firms 1 and -1 as a function of their prices.

it charges. On the left side of Figure 2, we give a contour plot for $\pi^1((p_1^1, p_2^1, 0) \,|\, (0, 2, 7.176))$ as a function of (p_1^1, p_2^1). Since $\theta_3^1(p_3^1, p_3^{-1}) = 0$, firm 1 cannot garner demand from product 3, so we can fix the price $p_3^1 = 0$. In the figure, if firm -1 charges its equilibrium prices, then the expected profit of firm 1 is maximized at $p_1^1 = 0.533$ and $p_2^1 = 1.086$, which are the equilibrium prices of firm 1. On the right side of Figure 2, we give a contour plot for $\pi^{-1}((0, p_2^{-1}, p_3^{-1}) \,|\, (0.533, 1.086, 0))$ as a function of (p_2^{-1}, p_3^{-1}). Since $\theta_1^{-1}(p_1^{-1}, p_1^1) = 0$, we can fix the price $p_1^{-1} = 0$. In the figure, if firm 1 charges its equilibrium prices, then the expected profit of firm -1 is maximized at $p_2^{-1} = 2$ and $p_3^{-1} = 7.176$, which are the equilibrium prices of firm -1.

We can also give a specific example where the two firms are simultaneously in the market for the common products in a Nash equilibrium. We consider a problem instance with two firms and three products. All unit costs are zero. The parameters of the Markov chain choice model are symmetric for the two firms and they are given by

$$\lambda_1 = \lambda_2 = \lambda_3 = \frac{1}{3},$$
$$\theta_1^1(p_1^1, p_1^{-1}) = 1 - p_1^1, \quad \theta_1^{-1}(p_1^{-1}, p_1^1) = 0,$$
$$\theta_2^1(p_2^1, p_2^{-1}) = \frac{1}{2} - \frac{1}{2}\,p_2^1, \quad \theta_2^{-1}(p_2^{-1}, p_2^1) = \frac{1}{2} - \frac{1}{2}\,p_2^{-1},$$
$$\theta_3^1(p_3^1, p_3^{-1}) = 0, \quad \theta_3^{-1}(p_3^{-1}, p_3^1) = 1 - p_3^{-1},$$
$$\rho_{11} = 0.1, \quad \rho_{12} = 0.4, \quad \rho_{13} = 0,$$
$$\rho_{21} = 0, \quad \rho_{22} = 0.4, \quad \rho_{23} = 0,$$
$$\rho_{31} = 0, \quad \rho_{32} = 0.4, \quad \rho_{33} = 0.1.$$

For this problem instance, the prices $\hat{\boldsymbol{p}}^1 = (0.547, 0.532, 0)$ and $\hat{\boldsymbol{p}}^{-1} = (0, 0.532, 0.547)$ are a Nash equilibrium. In this Nash equilibrium, we have $\theta_2^1(\hat{p}_2^1, \hat{p}_2^{-1}) = \theta_2^{-1}(\hat{p}_2^{-1}, \hat{p}_2^1) = 0.234$, indicating that both firms garner demand from product 2 in the Nash equilibrium. Indeed, firm 1 garners demand from products 1 and 2, whereas firm -1 garners demand from products 2 and 3.

D Appendix: Monotonicity of the Equilibrium Operator

In the next lemma, we show that the operator $h_i^k(\cdot,\ldots,\cdot)$ in (7) is monotone. We use this result at the end of Section 3.2 when showing the existence of a Nash equilibrium.

Lemma 15 *If $(\boldsymbol{r}^1,\ldots,\boldsymbol{r}^m) \in \Re_+^{n\times m}$ and $(\boldsymbol{q}^1,\ldots,\boldsymbol{q}^m) \in \Re_+^{n\times m}$ satisfy $r_i^k \le q_i^k$ for all $i \in N$, $k \in M$, then we have $h_i^k(\boldsymbol{r}^1,\ldots,\boldsymbol{r}^m) \le h_i^k(\boldsymbol{q}^1,\ldots,\boldsymbol{q}^m)$ for all $i \in N$, $k \in M$.*

Proof. First, consider the case $i \in N^k$. We have $\max_{p_i\in\mathcal{P}_i}\{\theta_i(p_i)\,(p_i-c_i)+(1-\theta_i(p_i))\,\sum_{j\in N}\rho_{ij}\,r_j^k\} \le \max_{p_i\in\mathcal{P}_i}\{\theta_i(p_i)\,(p_i-c_i)+(1-\theta_i(p_i))\,\sum_{j\in N}\rho_{ij}\,q_j^k\}$, since $r_i^k \le q_i^k$ for all $i\in N$, $k\in M$. In this case, by (7), we get $h_i^k(\boldsymbol{r}^1,\ldots,\boldsymbol{r}^m) \le h_i^k(\boldsymbol{q}^1,\ldots,\boldsymbol{q}^m)$. Second, consider the case $i \notin N^k$. We use $\hat{p}_i$ to denote an optimal solution to the problem $\max_{p_i\in\mathcal{P}_i}\{\theta_i(p_i)\,(p_i-c_i)+(1-\theta_i(p_i))\,\sum_{j\in N}\rho_{ij}\,r_j^k\}$. If we drop the constant terms, then an optimal solution to the last problem is also given by an optimal solution to $\max_{p_i\in\mathcal{P}_i}\{\theta_i(p_i)\,(p_i-c_i-\sum_{j\in N}\rho_{ij}\,r_j^k)\}$. Similarly, we use $\hat{s}_i$ to denote an optimal solution to the problem $\max_{p_i\in\mathcal{P}_i}\{\theta_i(p_i)\,(p_i-c_i)+(1-\theta_i(p_i))\,\sum_{j\in N}\rho_{ij}\,q_j^k\}$, which is also given by an optimal solution to $\max_{p_i\in\mathcal{P}_i}\{\theta_i(p_i)\,(p_i-c_i-\sum_{j\in N}\rho_{ij}\,q_j^k)\}$. By Lemma 19 in Appendix M, the unique optimal solution to the problem $\max_{p_i\in\mathcal{P}_i}\{\theta_i(p_i)\,(p_i-c_i-x)\}$ is increasing in x. In this case, since $r_i^k \le q_i^k$ for all $i\in N$, $k\in M$, noting the maximization problems that we solve to compute $\hat{p}_i$ and $\hat{s}_i$, it follows that $\hat{p}_i \le \hat{s}_i$ for all $i \in N$. Since we consider the case $i \notin N^k$, by (7), we get $h_i^k(\boldsymbol{r}^1,\ldots,\boldsymbol{r}^m) = (1-\theta_i(\hat{p}_i))\sum_{j\in N}\rho_{ij}\,r_j^k \le (1-\theta_i(\hat{s}_i))\sum_{j\in N}\rho_{ij}\,q_j^k = h_i^k(\boldsymbol{q}^1,\ldots,\boldsymbol{q}^m)$, where the inequality is by the fact that $\theta_i(\cdot)$ is decreasing, $\hat{p}_i \le \hat{s}_i$ and $r_j^k \le q_j^k$ for all $j \in N$. □

E Appendix: Monotonicity of the Expected Profit Sequence

Letting $\rho_{\max} = \max_{i\in N}\{\sum_{j\in N}\rho_{ij}\}$ and $\Delta = \max_{i\in N}\{\max_{p_i\in\mathcal{P}_i}\{\theta_i(p_i)\,(p_i-c_i)\}\}$, we define $\bar{u}$ as $\bar{u} = \Delta/(1-\rho_{\max})$. We generate the sequence $\{(\bar{\boldsymbol{r}}^1(t),\ldots,\bar{\boldsymbol{r}}^m(t)) : t\in\mathbb{N}\}$ by using the recursion $\bar{r}_i^k(t+1) = h_i^k(\bar{\boldsymbol{r}}^1(t),\ldots,\bar{\boldsymbol{r}}^m(t))$ for all $i\in N$, $k\in M$ with $\bar{r}_i^k(0)=\bar{u}$. In the next lemma, we show that the sequence $\{(\bar{\boldsymbol{r}}^1(t),\ldots,\bar{\boldsymbol{r}}^m(t)) : t\in\mathbb{N}\}$ is decreasing and bounded from below.

Lemma 16 *If the sequence $\{(\bar{\boldsymbol{r}}^1(t),\ldots,\bar{\boldsymbol{r}}^m(t)) : t\in\mathbb{N}\}$ is generated by $\bar{r}_i^k(0)=\bar{u}$ and $\bar{r}_i^k(t+1) = h_i^k(\bar{\boldsymbol{r}}^1(t),\ldots,\bar{\boldsymbol{r}}^m(t))$ for all $i\in N$, $k\in M$, then $\bar{r}_i^k(t) \ge \bar{r}_i^k(t+1) \ge 0$ for all $i\in N$, $k\in M$, $t\in\mathbb{N}$.*

Proof. First, we show that $\bar{r}_i^k(t) \ge 0$ for all $i\in N$, $k\in M$, $t\in\mathbb{N}$. Since $\lim_{p_i\to U_i}\theta_i(p_i)=0$ and $\lim_{p_i\to U_i}\theta_i(p_i)\,p_i = 0$, if $r_i^k \ge 0$ for all $i\in N$, $k\in M$, then the optimal objective value of the maximization problem in the first case in (7) is nonnegative. If $r_i^k\ge 0$ for all $i\in N$, $k\in M$, then the quantity in the second case in (7) is nonnegative as well. Therefore, if $(\boldsymbol{r}^1,\ldots,\boldsymbol{r}^m)$ satisfies $r_i^k \ge 0$ for all $i\in N$, $k\in M$, then it follows that $h_i^k(\boldsymbol{r}^1,\ldots,\boldsymbol{r}^m)\ge 0$ for all $i\in N$, $k\in M$. Similarly, since $\lim_{p_i\to U_i}\theta_i(p_i)=0$, we have $\max_{p_i\in\mathcal{P}_i}\{\theta_i(p_i)\,(p_i-c_i)\}\ge 0$, which implies that $\Delta\ge 0$. Noting that $\sum_{j\in N}\rho_{ij} < 1$ for all $i\in N$, we get $\bar{u}\ge 0$. Since $\bar{r}_i^k(0)=\bar{u}\ge 0$ for all $i\in N$, $k\in M$, by the discussion at the beginning of the proof, we have $\bar{r}_i^k(1) = h_i^k(\bar{\boldsymbol{r}}^1(0), \ldots, \bar{\boldsymbol{r}}^m(0)) \geq 0$. Repeating the argument recursively, we obtain $\bar{r}_i^k(t) \geq 0$ for all $i \in N$, $k \in M$, $t \in \mathbb{N}$. Thus, the sequence $\{(\bar{\boldsymbol{r}}^1(t), \ldots, \bar{\boldsymbol{r}}^m(t)) : t \in \mathbb{N}\}$ is bounded from below by zero. Second, we use induction to show that $\bar{r}_i^k(t) \geq \bar{r}_i^k(t+1)$ for all $i \in N$, $k \in M$, $t \in \mathbb{N}$. Considering the case $i \notin N^k$, by (7), we have $\bar{r}_i^k(1) = h_i^k(\bar{\boldsymbol{r}}^1(0), \ldots, \bar{\boldsymbol{r}}^m(0)) = (1 - \theta_i(\hat{p}_i)) \sum_{j \in N} \rho_{ij}\, \bar{r}_j^k(0) = (1 - \theta_i(\hat{p}_i)) \sum_{j \in N} \rho_{ij}\, \bar{u} \leq \bar{u} = \bar{r}_i^k(0)$. On the other hand, considering the case $i \in N^k$, we have

$$\begin{aligned}
\bar{r}_i^k(1) = h_i^k(\bar{\boldsymbol{r}}^1(0), \ldots, \bar{\boldsymbol{r}}^m(0)) &= \max_{p_i \in \mathcal{P}_i} \left\{ \theta_i(p_i)\,(p_i - c_i) + (1 - \theta_i(p_i)) \sum_{j \in N} \rho_{ij}\, \bar{r}_j^k(0) \right\} \\
&\leq \max_{p_i \in \mathcal{P}_i} \left\{ \theta_i(p_i)\,(p_i - c_i) \right\} + \sum_{j \in N} \rho_{ij}\, \bar{u} \leq \Delta + \rho_{\max}\, \bar{u} = \bar{u} = \bar{r}_i^k(0),
\end{aligned}$$

where the second inequality uses the definition of Δ and $\rho_{\max}$, whereas the second equality uses the fact that $\bar{u} = \Delta/(1 - \rho_{\max})$. Therefore, we obtain $\bar{r}_i^k(1) \leq \bar{r}_i^k(0)$ in both cases. Next, assuming that $\bar{r}_i^k(t+1) \leq \bar{r}_i^k(t)$ for all $i \in N$, $k \in M$, we proceed to showing that $\bar{r}_i^k(t+2) \leq \bar{r}_i^k(t+1)$ for all $i \in N$, $k \in M$. In particular, since $\bar{r}_i^k(t+1) \leq \bar{r}_i^k(t)$ for all $i \in N$, $k \in M$, by Lemma 15 in Appendix D, we immediately get $\bar{r}_i^k(t+2) = h_i^k(\bar{\boldsymbol{r}}^1(t+1), \ldots, \bar{\boldsymbol{r}}^m(t+1)) \leq h_i^k(\bar{\boldsymbol{r}}^1(t), \ldots, \bar{\boldsymbol{r}}^m(t)) = \bar{r}_i^k(t+1)$, which completes the induction argument. $\square$

F Appendix: Parameters of the Markov Chain Choice Model

We give the parameters of the Markov chain choice model in the numerical example at the end of Section 3.3. For each product i, the function $\theta_i(\cdot)$ is of the form $\theta_i(p_i) = \mathrm{e}^{-\alpha_i\, p_i}$. So, the parameters of the Markov chain choice model are $\{\lambda_i : i \in N\}$, $\{\alpha_i : i \in N\}$ and $\{\rho_{ij} : i, j \in N\}$. We give the values for these parameters in Table 2. We generate the parameters in the table randomly by using the following approach. To come up with $\{\lambda_i : i \in N\}$, we sample β_i uniformly over the interval $[0.1, 1]$ and set $\lambda_i = \beta_i / \sum_{j \in N} \beta_j$. To come up with $\{\alpha_i : i \in N\}$, we sample α_i uniformly over the interval $[0.1, 1]$. To come up with $\{\rho_{ij} : i, j \in N\}$, we sample γ_{ij} from the uniform distribution over $[0, 1]$ and ζ_i from the uniform distribution over $[0.85, 1]$, in which case, we set $\rho_{ij} = \gamma_{ij}\, \zeta_i / \sum_{k \in N} \gamma_{ik}$. Since $\sum_{j \in N} \rho_{ij} = \zeta_i$, if a customer visiting product i decides not to purchase this product, then she transitions to the no purchase option with probability $1 - \zeta_i$. The set of feasible prices for each product i is $\mathcal{P}_i = [0, \infty)$. The unit costs $\{c_i : i \in N\}$ are zero.

G Appendix: Effect of the Unit Costs on the Equilibrium Prices

We give a counterexample to demonstrate that if each firm owns an arbitrary number of products, then an increase in the unit cost of a product owned by a firm may result in an increase or a decrease in the prices charged by its competitors in the Pareto dominant equilibrium. This counterexample indicates that we cannot extend the result in Lemma 10 to the case where each firm owns an arbitrary number of products. We consider a problem instance with two firms and four products. Firm 1 owns products 1 and 2, whereas firm 2 owns products 3 and 4. The unit

$\{\lambda_i : i \in N\}$							
0.12	0.15	0.12	0.09	0.09	0.13	0.11	0.19

$\{\alpha_i : i \in N\}$							
0.32	0.30	0.82	0.24	0.10	0.28	0.73	0.55

$\{\rho_{ij} : i, j \in N\}$							
0.15	0.08	0.04	0.07	0.21	0.00	0.21	0.20
0.21	0.09	0.07	0.06	0.11	0.02	0.17	0.14
0.10	0.03	0.19	0.22	0.00	0.14	0.20	0.03
0.20	0.22	0.07	0.23	0.07	0.00	0.06	0.06
0.27	0.06	0.11	0.06	0.12	0.06	0.24	0.01
0.16	0.02	0.16	0.24	0.05	0.09	0.11	0.08
0.10	0.20	0.12	0.18	0.10	0.09	0.08	0.06
0.19	0.04	0.03	0.11	0.11	0.11	0.16	0.14

Table 2: Parameters of the Markov chain choice model that we use in the numerical example in Section 3.3.

costs of all products are zero. The parameters of the Markov chain choice model that governs the customer choice process are given by

$$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \frac{1}{4}$$

$$\theta_1(p_1) = \mathrm{e}^{-0.1\,p_1}, \quad \theta_2(p_2) = \mathrm{e}^{-0.2\,p_2}, \quad \theta_3(p_3) = \mathrm{e}^{-0.2\,p_3}, \quad \theta_4(p_4) = \mathrm{e}^{-0.3\,p_4},$$

$$\rho_{11} = 0.3, \quad \rho_{12} = 0.3, \quad \rho_{13} = 0, \quad \rho_{14} = 0.1,$$

$$\rho_{21} = 0.3, \quad \rho_{22} = 0.1, \quad \rho_{23} = 0.4, \quad \rho_{24} = 0,$$

$$\rho_{31} = 0, \quad \rho_{32} = 0.5, \quad \rho_{33} = 0.4, \quad \rho_{34} = 0,$$

$$\rho_{41} = 0.4, \quad \rho_{42} = 0.3, \quad \rho_{43} = 0.1, \quad \rho_{44} = 0.1.$$

For this problem instance, when the unit costs of all products are zero, the price of products 3 and 4 in the Pareto dominant equilibrium are respectively 6.869 and 4.428. When the unit cost of product 1 is 36 and the unit costs of all other products are still zero, the prices of products 3 and 4 in the Pareto dominant equilibrium are respectively 6.794 and 4.485. Thus, if the unit cost of product 1 increases, then the price of product 3 decreases, but the price of product 4 increases.

H Appendix: Limit of the Expected Profit Sequence

We generate the sequence $\{(\underline{\boldsymbol{r}}^1(t), \ldots, \underline{\boldsymbol{r}}^m(t)) : t \in \mathbb{N}\}$ by using the recursion $\underline{r}_i^k(t+1) = h_i^k(\underline{\boldsymbol{r}}^1(t), \ldots, \underline{\boldsymbol{r}}^m(t))$ for all $i \in N$, $k \in M$ with $\underline{r}_i^k(0) = 0$. In the next lemma, we show that the sequence $\{(\underline{\boldsymbol{r}}^1(t), \ldots, \underline{\boldsymbol{r}}^m(t)) : t \in \mathbb{N}\}$ is increasing and bounded from above, which immediately implies that this sequence has a limit. We use this result in the discussion at the beginning of Section 3.4. In this lemma, $\bar{u} = \Delta/(1 - \rho_{\max})$ is as defined right after Theorem 6.

Lemma 17 *If the sequence $\{(\underline{\boldsymbol{r}}^1(t), \ldots, \underline{\boldsymbol{r}}^m(t)) : t \in \mathbb{N}\}$ is generated by $\underline{r}_i^k(0) = 0$ and $\underline{r}_i^k(t+1) = h_i^k(\underline{\boldsymbol{r}}^1(t), \ldots, \underline{\boldsymbol{r}}^m(t))$ for all $i \in N$, $k \in M$, then $\underline{r}_i^k(t) \le \underline{r}_i^k(t+1) \le \bar{u}$ for all $i \in N$, $k \in M$, $t \in \mathbb{N}$.*

Proof. First, we use induction to show that $\underline{r}_i^k(t) \le \bar{u}$ for all $i \in N$, $k \in N$, $t \in \mathbb{N}$. We have $\underline{r}_i^k(0) = 0 \le \bar{u}$ for all $i \in N$, $k \in M$. Next, assuming that $\underline{r}_i^k(t) \le \bar{u}$ for all $i \in N$, $k \in M$, we show that $\underline{r}_i^k(t+1) \le \bar{u}$. Considering the case $i \notin N^k$, by the second case in (7), we get $\underline{r}_i^k(t+1) = h_i^k(\underline{\boldsymbol{r}}^m(t), \ldots, \underline{\boldsymbol{r}}^m(t)) = (1-\theta_i(\hat{p}_i)) \sum_{j \in N} \rho_{ij}\, \underline{r}_j^k(t) \le (1-\theta_i(\hat{p}_i)) \sum_{j \in N} \rho_{ij}\, \bar{u} \le \bar{u}$, where the first inequality is by the induction assumption. Considering the case $i \in N^k$, by (7), we have $\underline{r}_i^k(t+1) = h_i^k(\underline{\boldsymbol{r}}^1(t), \ldots, \underline{\boldsymbol{r}}^m(t)) = \max_{p_i \in \mathcal{P}_i}\{\theta_i(p_i)\,(p_i - c_i) + (1-\theta_i(p_i)) \sum_{j \in N} \rho_{ij}\, \underline{r}_j^k(t)\} \le \max_{p_i \in \mathcal{P}_i}\{\theta_i(p_i)\,(p_i - c_i)\} + \sum_{j \in N} \rho_{ij}\, \bar{u} \le \Delta + \rho_{\max}\, \bar{u} = \bar{u}$, where the first inequality follows from the induction assumption and the last equality is by the fact that $\bar{u} = \Delta/(1-\rho_{\max})$. We get $\underline{r}_i^k(t+1) \le \bar{u}$ in both cases, which completes the induction argument. Therefore, it follows that $\underline{r}_i^k(t) \le \bar{u}$ for all $i \in N$, $k \in M$, $t \in \mathbb{N}$.

Second, we show that $\underline{r}_i^k(t) \le \underline{r}_i^k(t+1)$ for all $i \in N$, $k \in M$, $t \in \mathbb{N}$. Since $\lim_{p_i \to U_i} \theta_i(p_i) = 0$ and $\lim_{p_i \to U_i} \theta_i(p_i)\, p_i = 0$, if $r_i^k \ge 0$ for all $i \in N$, $k \in M$, then the optimal objective value of the maximization problem in the first case in (7) is nonnegative. In this case, noting that $\underline{r}_i^k(0) = 0$ for all $i \in N$, $k \in M$, we obtain $\underline{r}_i^k(1) = h_i^k(\underline{\boldsymbol{r}}^1(0), \ldots, \underline{\boldsymbol{r}}^m(0)) \ge 0 = \underline{r}_i^k(0)$ for all $i \in N$, $k \in M$. Next, if we assume that $\underline{r}_i^k(t+1) \ge \underline{r}_i^k(t)$ for all $i \in N$, $k \in M$, then applying the operator $h_i^k(\cdot, \ldots, \cdot)$ on both sides of the last inequality, by Lemma 15 in Appendix D, we get $\underline{r}_i^k(t+2) = h_i^k(\underline{\boldsymbol{r}}^1(t+1), \ldots, \underline{\boldsymbol{r}}^m(t+1)) \ge h_i^k(\underline{\boldsymbol{r}}^1(t), \ldots, \underline{\boldsymbol{r}}^m(t)) = \underline{r}_i^k(t+1)$. Therefore, by induction, we can conclude that $\underline{r}_i^k(t+1) \ge \underline{r}_i^k(t)$ for all $i \in N$, $k \in M$, $t \in \mathbb{N}$. □

I Appendix: Structural Properties of the Optimal Policy

We give a counterexample to show that Lemma 12 may not hold under arbitrary choice models, even if these choice models are compatible with the random utility maximization principle.

Choice Model. We consider a choice model where each customer associates random willingness to pay amounts with the products. The surplus of a product is the difference between the willingness to pay amount that the customer associates with the product and the price of the product. The customer chooses the product with the largest nonnegative surplus. If there are no products with nonnegative surplus, then the customer leaves without a purchase. We index the products by $N = \{1, \ldots, n\}$. Let the random variable W_i be the willingness to pay amount that a customer associates with product i. Thus, if we charge the prices $\boldsymbol{p} = \{p_i \in N\}$ for the products, then a customer purchases product i with probability $\Theta_i(\boldsymbol{p}) = \mathbb{P}\{W_i - p_i = \max_{j \in N}\{W_j - p_j\} \text{ and } W_i - p_i \ge 0\}$, where we assume that there is a unique maximum element of the set $\{W_j - p_j \,:\, j \in N\}$ with probability one. If $\{W_i : i \in N\}$ have a continuous distribution, then this assumption is satisfied. In our discussion, we focus on the case with two products. The joint density function for the willingness to pay amounts (W_1, W_2) is given by

$$f(w_1, w_2) = \begin{cases} \frac{42}{11} & \text{if } w_1 \in [0, \frac{1}{2}] \text{ and } w_2 \in [\frac{1}{2}, 1] \\ \frac{2}{11} & \text{if } w_1 \in [\frac{1}{2}, 1] \text{ and } w_2 \in [0, \frac{1}{2}] \\ 0 & \text{otherwise.} \end{cases} \tag{14}$$

We proceed to deriving a formula for the probability $\Theta_i(p_1, p_2)$ that a customer purchases product i as a function of the prices (p_1, p_2) that we charge for the products. Note that the willingness to pay amount for each product is between zero and one. Therefore, there is no reason to charge a price that is outside the interval from zero to one, so we derive the purchase probabilities for the products when the prices are between zero and one. In Figure 3, we show seven regions A, B, C, D, E, F and G that partition the unit square $[0,1] \times [0,1]$. Focusing on each region separately, given that the prices (p_1, p_2) that we charge for the products are in each one of the regions, we derive the probabilities that a customer purchases products 1 and 2. We begin by considering the case where the prices (p_1, p_2) for the two products are in region A.

Consider the prices (p_1, p_2) that are in region A. On the left side of Figure 4, the horizontal and vertical axes respectively correspond to the willingness to pay amounts for products 1 and 2. The black dot shows a possible value for the prices (p_1, p_2) in region A. The thick dotted line is a 45 degree line starting at the point (p_1, p_2). Note that if the willingness to pay amounts (w_1, w_2) for the two products fall into region A_0, then we have $w_1 - p_1 < 0$ and $w_2 - p_2 < 0$, in which case, the customer leaves without a purchase. If the willingness to pay amounts (w_1, w_2) for the two products fall into region A_1, then we have $w_1 - p_1 > w_2 - p_2$ and $w_1 - p_1 > 0$, in which case, the customer purchases product 1. Lastly, if the willingness to pay amounts (w_1, w_2) for the two products fall into region A_2, then we have $w_2 - p_2 > w_1 - p_1$ and $w_2 - p_2 > 0$, in which case, the customer purchases product 2. Therefore, if we charge the prices (p_1, p_2) in region A, then the probabilities that a customer purchases products 1 and 2 are given by $\Theta_1(p_1, p_2) = \mathbb{P}\{(W_1, W_2) \in A_1\}$ and $\Theta_2(p_1, p_2) = \mathbb{P}\{(W_1, W_2) \in A_2\}$. Next, we explicitly compute these probabilities.

On the right side of Figure 4, we partition region A_1 into subregions A_{11}, A_{12}, A_{13} and A_{14}. Noting the form of the joint density function in (14), this density function takes the value $42/11$ over subregion A_{11}, takes the value $2/11$ over subregion A_{14} and takes the value zero over subregions A_{12} and A_{13}. Therefore, using $\mathsf{Area}(A_{11})$ and $\mathsf{Area}(A_{14})$ to respectively denote the areas of subregions A_{11} and A_{14}, the probability that the willingness to pay amounts (W_1, W_2) take a value in region A_1 is $\frac{42}{11}\,\mathsf{Area}(A_{11}) + \frac{2}{11}\,\mathsf{Area}(A_{14})$. By a simple geometric reasoning, we have $\mathsf{Area}(A_{11}) = \frac{1}{2}\,(p_2 - 2\,p_1)\,(1 - p_2) + (p_2 - \frac{1}{2})\,(\frac{1}{2} - p_1)$ and $\mathsf{Area}(A_{14}) = \frac{1}{4}$. Therefore, we obtain $\Theta_1(p_1, p_2) = \mathbb{P}\{(W_1, W_2) \in A_1\} = \frac{42}{11} \times \frac{1}{2}\,(p_2 - 2\,p_1)\,(1 - p_2) + \frac{42}{11} \times (p_2 - \frac{1}{2})\,(\frac{1}{2} - p_1) + \frac{2}{11} \times \frac{1}{4}$. Similarly, the density function in (14) takes the value $42/11$ over region A_2. Using $\mathsf{Area}(A_2)$ to denote the area of region A_2, the probability that the willingness to pay amounts (W_1, W_2) take a value in region A_2 is $\frac{42}{11}\,\mathsf{Area}(A_2)$. It is simple to check that $\mathsf{Area}(A_2) = \frac{1}{2}\,(1 - p_2)\,(1 + 2\,p_1 - p_2)$. Thus, we obtain $\Theta_2(p_1, p_2) = \mathbb{P}\{(W_1, W_2) \in A_2\} = \frac{42}{11} \times \frac{1}{2}\,(1 - p_2)\,(1 + 2\,p_1 - p_2)$.

By the discussion in the previous two paragraphs, we have the purchase probability of each product as a function of the prices that we charge, when the prices of the products are in region A. Using the same argument, when the prices of the products are in regions B, C, D, E, F and G in Figure 3, we can derive the purchase probability of each product as a function of the prices that we charge. In Table 3, we show the purchase probability of each product, given that the prices of the products are in a certain region. For example, if the prices (p_1, p_2) that we charge are in region E, then the purchase probability of product 2 is $\frac{2}{11} \times \frac{1}{2}\,(p_1 - p_2)^2 + \frac{42}{11} \times \frac{1}{4}$. Thus,

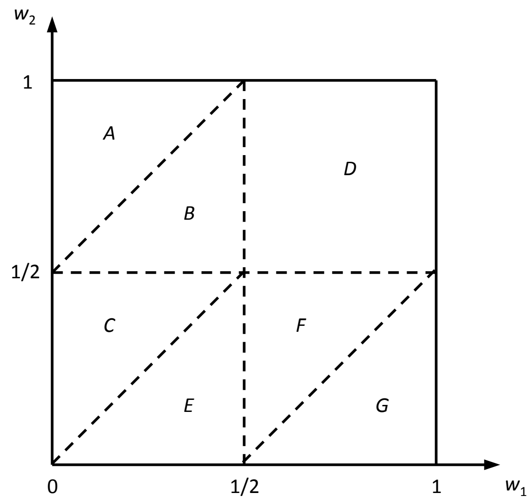

Figure 3: Seven regions partitioning the unit square $[0, 1] \times [0, 1]$.

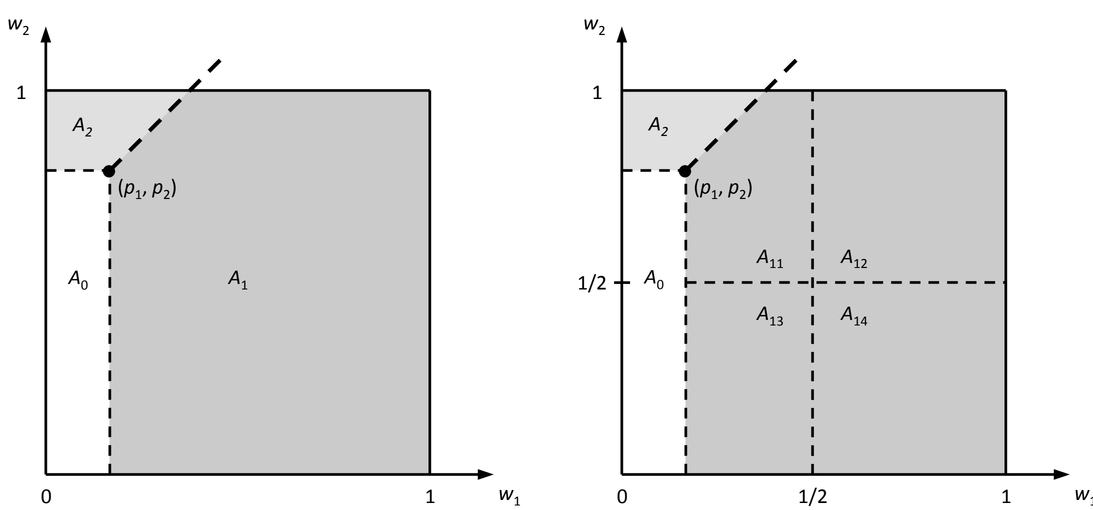

Figure 4: Computation of the purchase probabilities of the products given that the prices (p_1, p_2) are in region A.

Region for prices	$\Theta_1(p_1,p_2)$	$\Theta_2(p_1,p_2)$
A	$\frac{42}{11}\times\frac{1}{2}\,(p_2-2\,p_1)\,(1-p_2)$ $+\frac{42}{11}\times(p_2-\frac{1}{2})\,(\frac{1}{2}-p_1)+\frac{2}{11}\times\frac{1}{4}$	$\frac{42}{11}\times\frac{1}{2}\,(1-p_2)\,(1-p_2+2\,p_1)$
B	$\frac{42}{11}\times\frac{1}{2}\,(2\,p_2-p_1-\frac{1}{2})\,(\frac{1}{2}-p_1)+\frac{2}{11}\times\frac{1}{4}$	$\frac{42}{11}\times p_1\,(1-p_2)$ $+\frac{42}{11}\times\frac{1}{2}\,(\frac{1}{2}-p_1)\,(\frac{3}{2}+p_1-2\,p_2)$
C	$\frac{42}{11}\times\frac{1}{2}\,(p_2-p_1)^2+\frac{2}{11}\times\frac{1}{4}$	$\frac{42}{11}\times\frac{1}{2}\,(1-p_2+p_1)\,(p_2-p_1)$ $+\frac{42}{11}\times\frac{1}{2}\,(\frac{1}{2}-p_2+p_1)$
D	$\frac{2}{11}\times\frac{1}{2}\,(1-p_1)$	$\frac{42}{11}\times\frac{1}{2}\,(1-p_2)$
E	$\frac{2}{11}\times\frac{1}{2}\,(1-p_1+p_2)\,(p_1-p_2)$ $+\frac{2}{11}\times\frac{1}{2}\,(\frac{1}{2}-p_1+p_2)$	$\frac{2}{11}\times\frac{1}{2}\,(p_1-p_2)^2+\frac{42}{11}\times\frac{1}{4}$
F	$\frac{2}{11}\times p_2\,(1-p_1)$ $+\frac{2}{11}\times\frac{1}{2}\,(\frac{1}{2}-p_2)\,(\frac{3}{2}+p_2-2\,p_1)$	$\frac{2}{11}\times\frac{1}{2}\,(2\,p_1-p_2-\frac{1}{2})\,(\frac{1}{2}-p_2)+\frac{42}{11}\times\frac{1}{4}$
G	$\frac{2}{11}\times\frac{1}{2}\,(1-p_1)\,(1-p_1+2\,p_2)$	$\frac{2}{11}\times\frac{1}{2}\,(p_1-2\,p_2)\,(1-p_1)$ $+\frac{2}{11}\times(p_1-\frac{1}{2})\,(\frac{1}{2}-p_2)+\frac{42}{11}\times\frac{1}{4}$

Table 3: Purchase probabilities of the products when the prices (p_1,p_2) are in each one of the seven regions.

Table 3 gives a complete characterization of the purchase probabilities $\Theta_1(p_1,p_2)$ and $\Theta_2(p_1,p_2)$ of the two products as a function of the prices (p_1,p_2).

Under the choice model described so far, we let $\pi(p_1,p_2\,|\,\bar{c})$ be the expected profit from a customer as a function of the prices (p_1,p_2) of the two products, given that both products have the same unit cost $\bar{c}$. So, we have $\pi(p_1,p_2\,|\,\bar{c})=(p_1-\bar{c})\,\Theta_1(p_1,p_2)+(p_2-\bar{c})\,\Theta_2(p_1,p_2)$. Using the expressions for the choice probabilities in Table 3, we can check that if the unit costs of the products are zero, then the optimal solution to the problem $\max_{(p_1,p_2)\in[0,1]^2}\pi(p_1,p_2\,|\,0)$ is $(\hat{p}^1,\hat{p}^2)=(0.5,0.5)$ yielding the optimal objective value of $\pi(\hat{p}_1,\hat{p}_2\,|\,0)=0.5$. If the unit costs of the products are 0.1, then the optimal solution to the problem $\max_{(p_1,p_2)\in[0,1]^2}\pi(p_1,p_2\,|\,0.1)$ is $(\tilde{p}^1,\tilde{p}^2)=(0.447,0.567)$ yielding the optimal objective value of $\pi(\tilde{p}_1,\tilde{p}_2\,|\,0)=0.406$. Thus, under this choice model, if the unit costs of all products increase by the same amount, then the optimal price of a product may increase or decrease. Under the Markov chain choice model, by Lemma 4, if the unit costs of all products increase by the same amount, then the optimal prices of all products increase.

Dynamic Pricing with a Single Resource. Using the observations in the previous paragraph, we construct a counterexample to show that the structural properties of the optimal policy in Lemma 12 do not necessarily hold under arbitrary choice models. Our notation closely follows the one in Section 4. We consider a problem instance where there are $\tau = 2$ time periods in the selling horizon and $n = 2$ products. At the beginning of the selling horizon, we have $q = 2$ units of resource available. At each time period, a customer arrives into the system with probability $\Lambda = 0.2$. As a function of the prices of the products, an arriving customer chooses among the products according to the choice model described earlier in this section. If we charge the prices (p_{1t}, p_{2t}) for the products at time period t, then a customer purchases product i with probability $\Theta_i(p_{1t}, p_{2t})$. The sale of a product consumes one unit of the resource. Note that the choice model governing the choice process of the customers is stationary. Our goal is to find a policy that maximizes the total expected revenue over the selling horizon. Using x_t to denote the number of units of remaining resource at the beginning of time period t, we can find the optimal policy by computing the value functions $\{V_t(\cdot) : t \in T\}$ through the dynamic program

$$\begin{aligned} V_t(x_t) &= \max_{(p_{1t},p_{2t})\in[0,1]^2} \Bigg\{ \sum_{i\in\{1,2\}} \Lambda\, \Theta_i(p_{1t}, p_{2t}) \Big\{ p_{it} + V_{t+1}(x_t - 1) \Big\} \\ &\qquad\qquad + \Big\{ 1 - \Lambda + \Lambda \Big(1 - \sum_{i\in\{1,2\}} \Theta_i(p_{1t}, p_{2t}) \Big) \Big\} V_{t+1}(x_t) \Bigg\} \\ &= \max_{(p_{1t},p_{2t})\in[0,1]^2} \Bigg\{ \sum_{i\in\{1,2\}} \Lambda\, \Theta_i(p_{1t}, p_{2t}) \Big\{ p_{it} - \Delta V_{t+1}(x_t) \Big\} \Bigg\} + V_{t+1}(x_t), \end{aligned} \tag{15}$$

with the boundary conditions that $V_t(0) = 0$ and for all $t \in T$ and $V_{\tau+1}(\cdot) = 0$. In the first equality, we use the fact that if there is no customer arrival at a time period or a customer arrives and purchases nothing, then there is no change in the resource inventory.

Let $\hat{p}_{it}(x)$ be the optimal price to charge for product i at time period t when we have x units of remaining resource. We show that if the customers choose under the choice model described earlier in this section, then $\hat{p}_{11}(1) = 0.447 \leq 0.5 = \hat{p}_{11}(2)$. Therefore, the optimal price to charge for product 1 at time period 1 increases as we have more units of the resource. Under the Markov chain choice model, by Lemma 12, the optimal prices to charge at a certain time period decrease as we have more units of the resource. Also, we show that if the customers choose under the choice model described earlier in this section, then $\hat{p}_{11}(1) = 0.447 \leq 0.5 = \hat{p}_{12}(1)$. Therefore, the optimal price to charge for product 1 increases as we get closer to the end of the selling horizon with one unit of resource. Under the Markov chain choice model, by Lemma 12, if the choice models governing the choice behavior of the customers at different time periods are the same, then the optimal prices to charge decrease as we get closer to the end of the selling horizon. Therefore, our counterexample shows that Lemma 12 does not necessarily hold under arbitrary choice models.

We compute the optimal prices at time period 2 when there is one unit of resource. By the boundary conditions of the dynamic program in (15), we have $V_3(\cdot) = 0$, in which case, if there is one unit of resource at time period 2, then we can compute the optimal prices by solving the problem $\max_{(p_{12},p_{22})\in[0,1]^2} \sum_{i\in\{1,2\}} \Lambda\, \Theta_i(p_{12}, p_{22})\, p_{i2}$. Ignoring the constant Λ, the last maximization problem is the same as the profit maximization problem that we solved earlier in this section with the unit costs of the products being zero. We know that the optimal solution to this problem is obtained by setting $p_{12} = 0.5$ and $p_{22} = 0.5$. Thus, we have $\hat{p}_{12}(1) = 0.5$ and $\hat{p}_{22}(1) = 0.5$. Ignoring the constant Λ, we know that the optimal objective value of the last maximization problem is 0.5. Since $\Lambda = 0.2$, the optimal objective value of the last maximization problem is $0.2 \times 0.5 = 0.1$. This calculation also shows that the value function at time period 2 satisfies $V_2(1) = 0.1$. Using precisely the same computation, we have $V_2(2) = 0.1$. Lastly, by the boundary conditions of the dynamic program in (15), we have $V_2(0) = 0$. To summarize our computations for time period 2, we have $\hat{p}_{12}(1) = 0.5$, $V_2(2) = V_2(1) = 0.1$ and $V_2(0) = 0$.

Next, we compute the optimal prices at time period 1 when there is one unit of resource. By (15), if we have one unit of remaining resource at time period 1, then we can compute the optimal prices by solving the problem $\max_{(p_{11},p_{21})\in[0,1]^2} \sum_{i\in\{1,2\}} \Lambda\, \Theta_i(p_{11}, p_{21})\, (p_{i1} - \Delta V_2(1))$. Ignoring the constant Λ, since $\Delta V_2(1) = V_2(1) - V_2(0) = 0.1$, the last maximization problem is the same as the profit maximization problem that we solved earlier in this section with the unit costs of the products being 0.1. We know that the optimal solution to this problem is obtained by setting $p_{11} = 0.447$ and $p_{21} = 0.567$. Thus, we have $\hat{p}_{11}(1) = 0.447$ and $\hat{p}_{21}(1) = 0.567$. Similarly, if we have two units of remaining resource at time period 1, then we can compute the optimal prices by solving the problem $\max_{(p_{11},p_{21})\in[0,1]^2} \sum_{i\in\{1,2\}} \Lambda\, \Theta_i(p_{11}, p_{21})\, (p_{i1} - \Delta V_2(2))$. Since $\Delta V_2(2) = V_2(2) - V_2(1) = 0$, the last maximization problem is the same as the profit maximization problem that we solved earlier in this section with the unit costs of the products being zero, so $\hat{p}_{11}(2) = 0.5$ and $\hat{p}_{21}(2) = 0.5$. Thus, we have $\hat{p}_{11}(1) = 0.447 \le 0.5 = \hat{p}_{11}(2)$ and $\hat{p}_{11}(1) = 0.447 \le 0.5 = \hat{p}_{12}(1)$.

J Appendix: Concavity of the Objective Function

Consider a differentiable and strictly decreasing function $\theta(\cdot)$ such that $1/\theta(x)$ is convex in x. In the next lemma, we show that $y\, \theta^{-1}(y/x)$ is concave in (x, y).

Lemma 18 *For an interval $\mathcal{P} \subset \Re$, consider a differentiable and strictly decreasing function $\theta(\cdot) : \mathcal{P} \to [0, 1]$ such that $1/\theta(\alpha)$ is convex in α for $\alpha \in \mathcal{P}$. Then, defining $f(\cdot) : [0, 1] \to \mathcal{P}$ as $f(x) = \theta^{-1}(x)$, the function $F(x, y) = y\, f(y/x)$ is concave in (x, y) for $y \in \Re_+$ and $y/x \in [0, 1]$.*

Proof. We use $\partial^2_{XX} F(x, y)$ to denote the second derivative of $F(\cdot, \cdot)$ with respect to the first argument evaluated at (x, y). We use $\partial^2_{XY} F(x, y)$ and $\partial^2_{YY} F(x, y)$ with similar interpretations. By direct differentiation, it is simple to verify that

$$\partial^2_{XX} F(x,y) = \frac{y^2}{x^3}\Big\{2\, f'(y/x) + \frac{y}{x}\, f''(y/x)\Big\}, \qquad \partial^2_{XY} F(x,y) = -\frac{y}{x^2}\Big\{2\, f'(y/x) + \frac{y}{x}\, f''(y/x)\Big\},$$
$$\partial^2_{YY} F(x,y) = \frac{1}{x}\Big\{2\, f'(y/x) + \frac{y}{x}\, f''(y/x)\Big\},$$

which implies that $\partial^2_{XX} F(x, y)\, \partial^2_{YY} F(x, y) - (\partial^2_{XY} F(x, y))^2 = 0$. Therefore, to show that $F(x, y) = y\, f(y/x)$ is concave in (x, y), it is enough to show that $\partial^2_{XX} F(x, y) \le 0$ and $\partial^2_{YY} F(x, y) \le 0$. To show the last two inequalities, by the expressions for the partial derivatives above, it is enough to show that $2\, f'(y/x) + \frac{y}{x} f''(y/x) \leq 0$ for all $y/x \in [0,1]$. Letting $z = y/x$, we show that $2\, f'(z) + z\, f''(z) \leq 0$ for all $z \in [0,1]$. Since $f(z) = \theta^{-1}(z)$, we have $\theta(f(z)) = z$. Differentiating, we get $\theta'(f(z))\, f'(z) = 1$ so that $f'(z) = 1/\theta'(f(z))$. Differentiating one more time, we get $f''(z) = -\theta''(f(z))\, f'(z)/\theta'(f(z))^2 = -\theta''(f(z))/\theta'(f(z))^3$. Letting $\Theta(\alpha) = 1/\theta(\alpha)$, by the assumption in the lemma, $\Theta(\cdot)$ is convex. Differentiating, we get $\Theta'(\alpha) = -\theta'(\alpha)/\theta(\alpha)^2$. Differentiating one more time, we have $\Theta''(\alpha) = 2\, \frac{\theta'(\alpha)^2}{\theta(\alpha)^3} - \frac{\theta''(\alpha)}{\theta(\alpha)^2} \geq 0$, where the inequality is by the fact that $\Theta(\cdot)$ is convex. Using the last inequality with $\alpha = f(z)$ and noting that $\theta(f(z)) = z$, we get

$$\Theta''(f(z)) = 2\, \frac{\theta'(f(z))^2}{\theta(f(z))^3} - \frac{\theta''(f(z))}{\theta(f(z))^2} = \left(\frac{\theta'(f(z))}{z} \right)^3 \left\{ \frac{2}{\theta'(f(z))} - z\, \frac{\theta''(f(z))}{\theta'(f(z))^3} \right\}$$
$$= \left(\frac{\theta'(f(z))}{z} \right)^3 \left\{ 2\, f'(z) + z\, f''(z) \right\} \geq 0,$$

where the third equality uses the fact that $f'(z) = 1/\theta'(f(z))$ and $f''(z) = -\theta''(f(z))/\theta'(f(z))^3$ and the inequality follows from the fact that $\Theta(\cdot)$ is convex. Since $\theta(\cdot)$ is strictly decreasing, the chain of inequalities above implies that $2\, f'(z) + z\, f''(z) \leq 0$. □

K Appendix: Numerical Study for the Equivalent Formulation

We give a numerical study to check the benefit in the quality of the solutions that we obtain by solving the equivalent formulation in (11), instead of trying to solve problem (10) directly. In our numerical study, we randomly generate a large number of instances of the dynamic pricing problem with a single resource. For each problem instance, we solve problem (10) directly by using a nonconvex optimization software with 10 different initial solutions. Since problem (10) is not a convex program, the solutions that we obtain may not be global optima. Also, we solve the equivalent formulation in (11) for each problem instance. In all of our test problems, we have $\theta_{it}(p_{it}) = \mathrm{e}^{-\alpha_{it}\, p_{it}}$, in which case, by the discussion right after Theorem 13, problem (11) is a convex program. Our goal is to demonstrate that the nonconvexity of problem (10) can be a concern and we can get stuck at inferior local optima when we try to solve problem (10) directly.

We use the following approach to generate our problem instances. The Markov chain choice models that govern the choices of the customers at different time periods are stationary. Therefore, we use $\{\lambda_i : i \in N\}$, $\{\theta_i(\cdot) : i \in N\}$ and $\{\rho_{ij} : i, j \in N\}$ to denote the parameters of the Markov chain choice model. For each product i, the function $\theta_i(\cdot)$ is of the form $\theta_i(p_i) = \mathrm{e}^{-\alpha_i\, p_i}$ with $\mathcal{P}_i = [0, \infty)$. To come up with the parameters of the Markov chain choice model, we sample β_i from the uniform distribution over $[0,1]$ and set $\lambda_i = \beta_i / \sum_{j \in N} \beta_j$. Also, we sample α_i from the uniform distribution over $[0.1, 1]$. Lastly, we sample γ_{ij} from the uniform distribution over $[0,1]$ and ζ_i from the uniform distribution over $[0, 0.001]$, in which case, we set $\rho_{ij} = \gamma_{ij}\,(1 - \zeta_i)/\sum_{k \in N} \gamma_{ik}$. Under this setup, note that $\sum_{j \in N} \rho_{ij} = 1 - \zeta_i$, where the largest value of ζ_i is 0.001. In all of our test problems, there are $\tau = 100$ time periods in the selling horizon. To come up with the initial capacity of the resource q, we compute the prices that maximize the expected revenue when there is no capacity constraint on the resource. To compute such prices, we solve the problem $\max_{\boldsymbol{p} \in \times_{i \in N} \mathcal{P}_i} \sum_{i \in N} \theta_i(p_i)\, \boldsymbol{v}_i(\boldsymbol{p})\, p_i$ by using the approach in Section 2. We use $\hat{\boldsymbol{p}}$ to denote an optimal solution to the last problem. If there is no capacity constraint on the resource, then it is optimal to charge the prices $\hat{\boldsymbol{p}}$ at each time period. If we charge the prices $\hat{\boldsymbol{p}}$ at each time period, then the total expected consumption of the resource over the whole selling horizon is $\tau \sum_{i \in N} \theta_i(\hat{p}_i)\, \boldsymbol{v}_i(\hat{\boldsymbol{p}})$, so we set the initial capacity of the resource as $q = \Delta\, \tau \sum_{i \in N} \theta_i(\hat{p}_i)\, \boldsymbol{v}_i(\hat{\boldsymbol{p}})$, where Δ is a parameter that we vary. The parameter Δ captures the tightness of the resource capacity. Recalling that we use n to denote the number of products, we vary $(n, \Delta) \in \{5, 10, 20\} \times \{0.4, 0.6, 0.8, 1.0\}$, which provides 12 parameter combinations in our numerical setup.

In each parameter combination, we randomly generate 100 individual problem instances by using the approach described in the previous paragraph. For each problem instance, we obtain locally optimal solutions directly by solving problem (10) by using the `fmincon` routine in `Matlab` with 10 different initial solutions. In the initial solutions, the price of each product is sampled from the uniform distribution over $[0, 100]$. For problem instance k, we let Local_ℓ^k be the objective value in (10) at the local optimum obtained by starting from the ℓ-th initial solution. For each problem instance, we also solve the equivalent formulation in (11). Since problem (11) is a convex program, we can obtain its globally optimal solution, which, in turn, yields the optimal objective value for problem (10). We solve problem (11) by using the `fmincon` routine in `Matlab` as well. For problem instance k, we use Global^k to denote the objective value of problem (10) at the global optimum.

We show our results in Table 4. In this table, the first column shows the parameter configuration by using the pair (n, Δ). The second column shows the average percent optimality gap of the locally optimal solutions, where the average is computed over all 100 problem instances in a parameter configuration and all locally optimal solutions that we obtain by starting from 10 different initial solutions. In particular, recall that we generate 100 problem instances in each parameter configuration. For problem instance k in a parameter configuration, the percent optimality gap of the locally optimal solution that we obtain by starting from the ℓ-th initial solution is $\mathsf{Gap}_\ell^k = 100 \times \frac{\mathsf{Global}^k - \mathsf{Local}_\ell^k}{\mathsf{Global}^k}$. Therefore, the second column shows $\frac{1}{1000} \sum_{k=1}^{100} \sum_{\ell=1}^{10} \mathsf{Gap}_\ell^k$. The third column shows the frequency of locally optimal solutions with optimality gaps exceeding 1%. In other words, the third column shows $\frac{1}{1000} \sum_{k=1}^{100} \sum_{\ell=1}^{10} \mathbf{1}(\mathsf{Gap}_\ell^k \geq 1)$.

Over all of our test problems, the average optimality gap of the locally optimal solutions is about 10.56%. In about 0.404 fraction of the locally optimal solutions, the optimality gaps exceed 1%. As the number of products increases and the number of decision variables in problem (10) gets larger, the optimality gaps tend to increase slightly. In particular, considering the problem instances with 5, 10 and 25 products separately, the average percent optimality gaps of the locally optimal solutions are respectively 10.05%, 10.39% and 11.23%. Similarly, considering the problem instances with 5, 10 and 25 products separately, the fractions of locally optimal solutions with more than 1% optimality gaps are respectively 0.283, 0.346 and 0.585. Considering the CPU seconds to solve

Param. Conf. (n, Δ)	Avg. Opt. Gap	Freq. Gap $\geq 1\%$
$(5, 0.4)$	10.43%	0.319
$(5, 0.6)$	9.38%	0.301
$(5, 0.8)$	9.56%	0.292
$(5, 1.0)$	10.83%	0.219
Average	10.05%	0.283

Param. Conf. (n, Δ)	Avg. Opt. Gap	Freq. Gap $\geq 1\%$
$(10, 0.4)$	10.91%	0.303
$(10, 0.6)$	11.14%	0.419
$(10, 0.8)$	10.12%	0.292
$(10, 1.0)$	9.41%	0.369
Average	10.39%	0.346

Param. Conf. (n, Δ)	Avg. Opt. Gap	Freq. Gap $\geq 1\%$
$(25, 0.4)$	9.81%	0.565
$(25, 0.6)$	10.06%	0.544
$(25, 0.8)$	12.38%	0.596
$(25, 1.0)$	12.68%	0.634
Average	11.23%	0.585

Table 4: Performance of the locally optimal solutions that we obtain by solving problem (10) directly.

problems (10) and (11), for the problem instances with 25 products, which are the largest ones in our numerical study, solving problem (10) directly takes about 1.36 second on average, whereas solving problem (11) takes about about 0.58 seconds on average. Overall, our results indicate that we can get stuck at local optima with noticeably inferior solution quality when we try to solve problem (10) directly, instead of using the equivalent formulation in (11).

L Appendix: Comparison of the Fitted Choice Models

In Section 5, we compare the fitted choice models from the perspective of out of sample log-likelihoods. In this section, we compare the fitted choice models from the perspective of two additional performance measures, which are the errors in the purchase probability and expected revenue predictions of the fitted choice models. Recall that we have eight different configurations of the ground choice model, which we indicate by using the tuple $(\mu_L, \mu_U, \Gamma, \kappa)$ with $(\mu_L, \mu_U) \in \{(100, 150), (150, 200)\}$, $\Gamma \in \{1, 2\}$ and $\kappa \in \{10, 15\}$. For each configuration of the ground choice model, we generate the purchase history of τ customers and use this purchase history as the training data. We have $\tau \in \{1{,}000,\ 2{,}500,\ 5{,}000\}$, corresponding to different levels of data availability. We also generate the past purchase history for another 10,000 customers to use as the testing data. We fit a Markov chain choice model and a multinomial logit model to the training data and test the performance of the fitted choice models on the testing data.

We begin by considering the error in the purchase probability predictions of the fitted choice models. For $\mathsf{CM} \in \{\mathsf{GR}, \mathsf{MC}, \mathsf{ML}\}$, we use $\mathsf{Choice}_i^{\mathsf{CM}}(\boldsymbol{p})$ to denote the purchase probability of product i under the choice model CM when we charge the prices $\boldsymbol{p}$. Having $\mathsf{CM} = \mathsf{GR}$ corresponds to the ground choice model that actually governs the choices of the customers in the training and testing data, having $\mathsf{CM} = \mathsf{MC}$ corresponds to the fitted Markov chain choice model and having $\mathsf{CM} = \mathsf{ML}$ corresponds to the fitted multinomial logit model. Note that we have access to the ground choice model. We capture the testing data by $\{(\tilde{\boldsymbol{p}}^\ell, i^\ell) : \ell = 1, \ldots, 10{,}000\}$, where the prices offered to customer ℓ are $\tilde{\boldsymbol{p}}^\ell$ and the product purchased by customer ℓ is i^ℓ. To check the error in the purchase probability predictions of the fitted Markov chain choice model, we compute $\frac{1}{10{,}000 \times n} \sum_{\ell=1}^{10{,}000} \sum_{i \in N} |\mathsf{Choice}_i^{\mathsf{GR}}(\tilde{\boldsymbol{p}}^\ell) - \mathsf{Choice}_i^{\mathsf{MC}}(\tilde{\boldsymbol{p}}^\ell)|$, which is the average absolute error in the purchase probabilities of the products under the fitted Markov chain choice model, given that we

Ground. Ch. Config. $(\mu_L, \mu_U, \Gamma, \kappa)$	$\tau = 1,000$ Err. MC	Err. ML	Perc. Gap	$\tau = 2,500$ Err. MC	Err. ML	Perc. Gap	$\tau = 5,000$ Err. MC	Err. ML	Perc. Gap
$(100, 150, 1, 10)$	0.0205	0.0196	-4.96	0.0164	0.0169	3.01	0.0157	0.0161	2.26
$(100, 150, 1, 15)$	0.0225	0.0245	8.03	0.0198	0.0235	15.86	0.0164	0.0218	25.01
$(100, 150, 2, 10)$	0.0249	0.0228	-8.96	0.0202	0.0220	8.51	0.0183	0.0218	15.99
$(100, 150, 2, 15)$	0.0215	0.0207	-3.82	0.0182	0.0191	4.24	0.0171	0.0183	6.69
$(150, 200, 1, 10)$	0.0205	0.0235	13.05	0.0173	0.0197	12.51	0.0171	0.0196	12.59
$(150, 200, 1, 15)$	0.0251	0.0226	-11.29	0.0174	0.0203	14.17	0.0145	0.0193	24.86
$(150, 200, 2, 10)$	0.0224	0.0223	-0.28	0.0175	0.0214	18.03	0.0170	0.0213	20.03
$(150, 200, 2, 15)$	0.0234	0.0233	-0.43	0.0175	0.0216	18.96	0.0164	0.0210	21.96

Table 5: Average absolute error in the purchase probabilities under the fitted Markov chain choice model and the fitted multinomial logit model on the testing data.

charge the prices in the testing data. We check the error in the purchase probability predictions of the fitted multinomial logit model in a similar fashion. In Table 5, we show the average absolute error in the purchase probabilities of the fitted choice models. The layout of this table is identical to that of Table 1. After the first column, there are three blocks of three columns. In each of the three blocks, we focus on the Markov chain choice model and the multinomial logit model fitted to the training data by using three different levels of data availability. In each block, the first column shows the average absolute error in the purchase probabilities of the fitted Markov chain choice model, the second column shows the average absolute error in the purchase probabilities of the fitted multinomial logit model and the third column shows the percent gap between the two average absolute errors. In 18 out of 24 cases, the average absolute error of the fitted Markov chain choice model is smaller than that of the fitted multinomial logit model. The remaining six cases where the average absolute error of the fitted multinomial logit model is smaller correspond to the cases with the smallest amount of training data with $\tau = 1{,}000$.

Next, we consider the error in the expected revenue predictions of the fitted choice models. For $\mathsf{CM} \in \{\mathsf{GR}, \mathsf{MC}, \mathsf{ML}\}$, we use $\mathsf{Rev}^{\mathsf{CM}}(\boldsymbol{p})$ to denote the expected revenue under the choice model CM when we charge the prices $\boldsymbol{p}$. In other words, we have $\mathsf{Rev}^{\mathsf{CM}}(\boldsymbol{p}) = \sum_{i \in N} \mathsf{Choice}_i^{\mathsf{CM}}(\boldsymbol{p})\, p_i$. To check the error in the expected revenue predictions of the fitted Markov chain choice model, we compute $\frac{1}{10{,}000} \sum_{\ell=1}^{10{,}000} |\mathsf{Rev}^{\mathsf{GR}}(\tilde{\boldsymbol{p}}^\ell) - \mathsf{Rev}^{\mathsf{MC}}(\tilde{\boldsymbol{p}}^\ell)|$, which is the average absolute error in the expected revenues under the fitted Markov chain choice model, given that we charge the prices in the testing data. We check the error in the expected revenue predictions of the fitted multinomial logit model in a similar fashion. In Table 6, we show the average absolute error in the expected revenues of the fitted choice models. The layout of this table is identical to that of Table 5. The only difference is that we focus on the average absolute error in the expected revenues, rather than the average absolute error in the purchase probabilities. Our results indicate that the fitted Markov chain choice model compares quite favorably with the fitted multinomial logit model from the perspective of the errors in the expected revenue predictions as well.

Ground. Ch. Config. $(\mu_L, \mu_U, \Gamma, \kappa)$	$\tau = 1,000$ Err. MC	Err. ML	Perc. Gap	$\tau = 2,500$ Err. MC	Err. ML	Perc. Gap	$\tau = 5,000$ Err. MC	Err. ML	Perc. Gap
$(100, 150, 1, 10)$	0.7795	0.7852	0.73	0.6075	0.7663	20.72	0.6411	0.8105	20.89
$(100, 150, 1, 15)$	0.5944	0.6950	14.47	0.5916	0.8002	26.07	0.4964	0.7912	37.26
$(100, 150, 2, 10)$	0.3885	0.4109	5.44	0.3008	0.4060	25.90	0.2897	0.4032	28.15
$(100, 150, 2, 15)$	0.3122	0.4035	22.62	0.2387	0.3204	25.48	0.2260	0.3025	25.30
$(150, 200, 1, 10)$	0.5488	0.7892	30.45	0.5685	0.6576	13.55	0.5744	0.6854	16.21
$(150, 200, 1, 15)$	0.6384	1.0087	36.71	0.4793	0.8287	42.17	0.4904	0.7945	38.27
$(150, 200, 2, 10)$	0.1785	0.2469	27.71	0.1313	0.2083	36.94	0.1389	0.2156	35.59
$(150, 200, 2, 15)$	0.3614	0.5552	34.91	0.3435	0.4983	31.06	0.3038	0.4751	36.07

Table 6: Average absolute error in the expected revenues under the fitted Markov chain choice model and the fitted multinomial logit model on the testing data.

M Appendix: Omitted Results

In this section, we give two auxiliary results used in our proofs. In the next lemma, we show that the unique optimal solution to the problem $\max_{p_i \in \mathcal{P}_i} \{\theta_i(p_i)\,(p_i - c_i - x)\}$ is increasing in x.

Lemma 19 *There exists a unique optimal solution to the problem* $\max_{p_i \in \mathcal{P}_i} \{\theta_i(p_i)\,(p_i - c_i - x)\}$ *and this unique optimal solution is increasing in* x*.*

Proof. By the assumption that $\theta_i(p_i)\,(p_i - c_i - x)$ is strictly quasiconcave in p_i for any $x \in \Re$, there exists a unique optimal solution to the problem in the lemma. Fix x^+ and x^0 with $x^+ > x^0$. Let $\hat{p}_i^+$ be the optimal solution to the problem in the lemma with $x = x^+$ and $\hat{p}_i^0$ be the optimal solution to the problem in the lemma with $x = x^0$. To get a contradiction, assume that $\hat{p}_i^+ < \hat{p}_i^0$. Since $\hat{p}_i^+ < \hat{p}_i^0$ and both $\hat{p}_i^+$ and $\hat{p}_i^0$ are in the interval $\mathcal{P}_i$, $\hat{p}_i^+$ is not the upper bound of the interval. Thus, since $\hat{p}_i^+$ is the optimal solution with $x = x^+$, the first order condition implies that $\theta_i'(\hat{p}_i^+)\,(\hat{p}_i^+ - c_i - x^+) + \theta_i(\hat{p}_i^+) \leq 0$. Furthermore, since $\theta_i(p_i)\,(p_i - c_i - x)$ is strictly quasiconcave in x, its derivative with respect to p_i changes sign from positive to negative only once as p_i increases. Once the derivative changes sign from positive to negative, the sign of the derivative stays negative. Thus, since $\theta_i'(\hat{p}_i^+)\,(\hat{p}_i^+ - c_i - x^+) + \theta_i(\hat{p}_i^+) \leq 0$ and $\hat{p}_i^+ < \hat{p}_i^0$,

$$0 > \theta_i'(\hat{p}_i^0)\,(\hat{p}_i^0 - c_i - x^+) + \theta_i(\hat{p}_i^0) = \theta_i'(\hat{p}_i^0)\,(\hat{p}_i^0 - c_i - x^0) + \theta_i(\hat{p}_i^0) + \theta_i'(\hat{p}_i^0)\,(x^0 - x^+).$$

Noting that $\theta_i(\cdot)$ is decreasing and $x^+ > x^0$, we have $\theta_i'(\hat{p}_i^0)\,(x^0 - x^+) \geq 0$, in which case, the chain of inequalities above yields $\theta_i'(\hat{p}_i^0)\,(\hat{p}_i^0 - c_i - x^0) + \theta_i(\hat{p}_i^0) < 0$. So, the derivative of $\theta_i(p_i)\,(p_i - c_i - x^0)$ with respect to p_i at $\hat{p}_i^0$ is strictly negative. Since $\hat{p}_i^+ < \hat{p}_i^0$, $\hat{p}_i^0$ is not the lower bound of the interval $\mathcal{P}_i$. Thus, we can decrease $\hat{p}_i^0$ by an infinitesimal amount to increase the value of the function $\theta_i(p_i)\,(p_i - c_i - x^0)$, contradicting the fact that $\hat{p}_i^0$ maximizes this function over the interval $\mathcal{P}_i$. $\Box$

In the next lemma, we show that if $\hat{r}_i^k = h_i^k(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ for all $i \in N$, $k \in M$, then we have $0 \leq \hat{r}_i^k \leq \bar{u}$ for all $i \in N$, $k \in M$, where $\bar{u} = \Delta/(1 - \rho_{\max})$ is as defined right after Theorem 6.

Lemma 20 *If $(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ satisfies $\hat{r}_i^k = h_i^k(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ for all $i \in N$, $k \in M$, then we must have $0 \leq \hat{r}_i^k \leq \bar{u}$ for all $i \in N$, $k \in M$.*

Proof. Assume that $\hat{r}_i^k = h_i^k(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m)$ for all $i \in N$, $k \in M$. Letting $\mu = \min_{i \in N,\ k \in M}\{\hat{r}_i^k\}$, we show that $\mu \geq 0$. To get a contradiction, assume that $\mu < 0$. First, consider some $i \in N^k$. By (7), we have $\hat{r}_i^k = h_i^k(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m) = \max_{p_i \in \mathcal{P}_i} \{\theta_i(p_i)\,(p_i - c_i) + (1 - \theta_i(p_i)) \sum_{j \in N} \rho_{ij}\, \hat{r}_j^k\} \geq \sum_{j \in N} \rho_{ij}\, \hat{r}_j^k \geq \sum_{j \in N} \rho_{ij}\, \mu > \mu$, where the first inequality is by the fact $\lim_{p_i \to U_i} \theta_i(p_i) = 0$ and $\lim_{p_i \to U_i} \theta_i(p_i)\, p_i = 0$ and the third inequality is by the fact that $\sum_{j \in N} \rho_{ij} < 1$ and $\mu < 0$. Second, consider some $i \notin N^k$. By (7), we have $\hat{r}_i^k = h_i^k(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m) = (1 - \theta_i(\hat{p}_i)) \sum_{j \in N} \rho_{ij}\, \hat{r}_j^k \geq (1 - \theta_i(\hat{p}_i)) \sum_{j \in N} \rho_{ij}\, \mu > \mu$, where we, again, use the fact that $\sum_{j \in N} \rho_{ij} < 1$ and $\mu < 0$. Therefore, we have $\hat{r}_i^k > \mu$ for all $i \in N$, $k \in M$, which contradicts the fact that $\mu = \min_{i \in N,\ k \in M}\{\hat{r}_i^k\}$. Thus, we must have $\mu \geq 0$, which implies that $\hat{r}_i^k \geq 0$ for all $i \in N$, $k \in M$.

Next, letting $\zeta = \max_{i \in N,\ k \in M}\{\hat{r}_i^k\}$, we show that $\zeta \leq \bar{u}$. To get a contradiction, assume that $\zeta > \bar{u}$. First, consider some $i \in N^k$. Recall that we define $\bar{u}$ as $\bar{u} = \Delta/(1 - \rho_{\max})$, where $\Delta = \max_{i \in N}\{\max_{p_i \in \mathcal{P}_i}\{\theta_i(p_i)\,(p_i - c_i)\}\}$ and $\rho_{\max} = \max_{i \in N}\{\sum_{j \in N} \rho_{ij}\}$. In this case, by (7), we get $\hat{r}_i^k = h_i^k(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m) = \max_{p_i \in \mathcal{P}_i}\{\theta_i(p_i)\,(p_i - c_i) + (1 - \theta_i(p_i)) \sum_{j \in N} \rho_{ij}\, \hat{r}_j^k\} \leq \Delta + \rho_{\max}\, \zeta = (1 - \rho_{\max})\, \bar{u} + \rho_{\max}\, \zeta < \zeta$, where the first inequality uses the fact that $\zeta = \max_{i \in N,\ k \in M}\{\hat{r}_i^k\}$ and the last inequality follows from the fact that $\rho_{\max} \in [0, 1)$ and $\zeta > \bar{u}$. Second, consider some $i \notin N^k$. By (7), we have $\hat{r}_i^k = h_i^k(\hat{\boldsymbol{r}}^1, \ldots, \hat{\boldsymbol{r}}^m) = (1 - \theta_i(\hat{p}_i)) \sum_{j \in N} \rho_{ij}\, \hat{r}_j^k \leq \rho_{\max}\, \zeta < \zeta$, where the second inequality uses the fact that $\rho_{\max} \in [0, 1)$. Therefore, we have $\hat{r}_i^k < \zeta$ for all $i \in N$, $k \in M$, which contradicts the fact that $\zeta = \max_{i \in N,\ k \in M}\{\hat{r}_i^k\}$. Thus, we must have $\zeta \leq \bar{u}$, which implies that $\hat{r}_i^k \leq \bar{u}$ for all $i \in N$, $k \in M$. $\square$