10-2014

Stability and Endogenous Formation of Inventory Transshipment Networks

Xin FANG
Singapore Management University, XFANG@smu.edu.sg

Soo-Haeng CHO
Carnegie Mellon University

Published in Operations Research, November 2014, Volume 62, Issue 6, Pages 1316-1334.
http://doi.org/10.1287/opre.2014.1324

Stability and Endogenous Formation of Inventory Transshipment Networks

Xin Fang · Soo-Haeng Cho
Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA 15213
xfang@andrew.cmu.edu · soohaeng@andrew.cmu.edu

Abstract: This paper studies a cooperative game of inventory transshipment among multiple firms. In this game, firms first make their inventory decisions independently, and then decide collectively how to transship excess inventories to satisfy unmet demands. In modeling transshipment, we use networks of firms as the primitive, which offer a richer representation of relationships among firms by taking the coalitions used in all previous studies as special cases. For any given cooperative network, we construct a dual price allocation under which the network is stable for any residual demands and supplies in the sense that no firms find it more profitable to form subnetworks. Under the allocation based on the marginal contribution of each firm to its network (called the MJW value), we show that various network structures such as complete, hub-spoke, and chain networks are stable only under certain conditions on residual amounts. Moreover, these conditions differ across network structures, implying that a network structure plays an important role in establishing the stability of a decentralized transshipment system. While the previous coalition-based approach examines only the grand coalition (i.e., the complete network), we find the complete network tends to be less stable than incomplete networks under the MJW value. Finally, we consider the case when firms establish networks endogenously, and show that pairwise Nash stable networks underperform the corresponding networks in centralized systems.
Subject classifications: Games/group decisions: Cooperative. Networks. Inventory.
Area of review: Manufacturing, Service, and Supply Chain Operations.

1 Introduction

Networks are often used to represent relationships among multiple firms. In this paper, we investigate networks of firms that share inventory through transshipment. When firms are connected in these networks, they can transship excess inventories (also called "residual supplies") between each other to satisfy unmet demands (also called "residual demands"). The benefit of inventory transshipment is straightforward. By joining a network of transshipment, firms can generate additional revenues by utilizing their otherwise unused inventories and demands. According to Narus and Anderson (1996), transshipment can reduce inventory cost by 15% to 20% and the amount of lost sales by as much as 75%. Due to these benefits of transshipment, many manufacturers promote inventory transshipment in their networks of independent retailers (Shao et al. 2011).

The importance of transshipment has been recognized well in the operations management (OM) literature. In particular, our paper is related to the stream of research that studies transshipment of inventories in *decentralized* systems. In these systems, multiple independent firms cooperate in order to maximize their own profits. Thus a proper mechanism needs to be developed to ensure that firms have incentives to participate in transshipment.[1] The extant literature analyzes the incentives of independent firms using the cooperative game theory based on the concept of coalitions. A *coalition* is a set of firms, within which firms can share their excess inventories and unmet demands, and thus generate additional profits from transshipment. When all firms belong to one coalition, such a coalition is called the *grand coalition*. The profits from transshipment are allocated among firms within a coalition according to a predetermined allocation rule. A central question of this research stream is whether a certain allocation is in the core. When a core allocation is used, no subset of firms has incentives to form subcoalitions by seceding from the grand coalition.

Our paper is distinguished from the previous research stream by analyzing *transshipment networks in decentralized systems*. As compared with a coalition-based cooperative game studied in the literature, we consider a network-based cooperative game. Our network-based approach is fundamentally different from the coalition-based approach in the following two important aspects. First, in a coalition-based game, any two firms within a coalition can share their residual demands or supplies *directly* with each other, meaning that transshipment between two firms can occur without the cooperation of other firms. In contrast, in a network-based game, a firm can share its residual demands or supplies *directly* with the other firms to which the firm has links, and *indirectly* with the other firms to which the firm has no direct links but is connected via other intermediate firms. In the latter case, transshipment occurs only when all intermediate firms cooperate as well. To illustrate this difference, consider a market with three firms. Figure 1(a) shows the grand coalition in this market. In a coalition-based game, all three firms can share their residuals with one another only in the grand coalition. On the other hand, Figure 1(b)-(e) show that all three firms can share their residuals with one another in four different types of networks: the *complete network* in which all three firms are linked to each other (which is essentially the same as the grand coalition), and three *incomplete networks* in which there is one link missing with respect to the complete network. For example, in the incomplete network shown in Figure 1(c), firm 1 and firm 3 are connected to each other through firm 2, and they can share residuals only when firm 2 cooperates. Second, within a coalition, firms are distinguished by the amount of their residual demands or supplies. In a network, firms are distinguished by their positions within the network as well as their residual amounts. For example, in the network shown in Figure 1(c), firm 2 is positioned better than firms 1 and 3 because firms 1 and 3 can share their residuals only through firm 2, while firm 2 can share

[1] A similar incentive problem can occur even among different branches within the same firm. For example, in our industry project with an energy distribution company, we observed that inventory transshipment occurred between two branches that are located in different regions. In this company, each individual manager who operates a local branch is evaluated based primarily on the performance of his/her own branch. Thus, in order for transshipment to occur, two managers have to agree upon how the additional profit generated from transshipment is allocated.

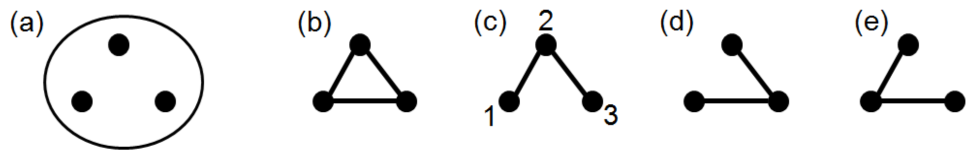

Figure 1: Coalition vs. network: (a) grand coalition, (b) complete network, and (c)-(e) incomplete networks in which all three firms are connected.

its residuals directly with firm 1 or 3.

As illustrated in Figure 1, networks represent richer relationships among firms than coalitions by taking coalitions as special cases. For this reason, there have been advances in the theory of a network-based cooperative game, starting with Myerson (1977) who notes that "there are many intermediate possibilities between universal cooperation and no cooperation." Such partial cooperation structures can be captured by networks, but not by coalitions. For example, in the network shown in Figure 1(c), firm 1 and firm 3 have partnerships with firm 2, but not between each other. A comprehensive review for the development of the theory and applications of a network-based cooperative game can be found in Jackson (2006). Our paper is among the first papers in the OM literature that apply this theory to firms' operational decisions.

The analysis of various networks is also important from a practical viewpoint. The complete network is often too expensive to implement in practice due to the cost of establishing connections (Jackson 2006). Several papers reviewed by Paterson et al. (2011) have analyzed the efficiency of incomplete networks in a *centralized* system where a central coordinator optimally determines transshipment based on those networks given exogenously. In our paper, we do not assume the existence of such a central coordinator; instead, we examine firms' individual incentives to join transshipment networks in a *decentralized* setting, and also analyze a situation in which firms build their networks *endogenously.* The following examples further motivate our research:

– In a typical regional blood management system (e.g., Prastacos 1984, Fontaine et al. 2009), multiple independent hospital blood banks set their own inventory levels for fresh units of whole blood or red cells. When one blood bank experiences a shortage and has an urgent need for blood, it needs to have it transshipped from another blood bank who has it in inventory. Such transshipment occurs through a regional blood center (usually a large hospital) who re-distributes the unused units of blood among smaller blood banks. Although a detailed arrangement of transshipment is fairly complex, this system resembles a hub-spoke network depicted in Figure 1(c), in which a regional blood center works as the hub and smaller blood banks work as spokes.

– Many states in the United States require auto manufacturers to sell their cars through independent dealers (Ramsey and Bauerlein 2013). The availability of various models and colors are different across dealers. If a dealer does not have a stock of the specific model desired by a consumer, a common practice is to have it transshipped from another dealer who has it in inventory (Zhao and Atkins 2009). To initiate such transshipment, a salesperson in the dealer shop first needs to find out which dealers have a desired model in inventory. This can be done either through shared web-based inquiry tools or through phone calls by the salesperson. Once the salesperson has found the model available in others' inventories, he needs to contact those other dealers and agrees upon various terms and conditions for its transshipment. This process involves costs for labor and administrative arrangements (Lien et al. 2011).

Similarly, decentralized transshipment networks are also used in industries such as machine tools and repair parts (Narus and Anderson 1996) and trucking industry (Zhao and Atkins 2009). Duvall (2000) reports that companies in a variety of sectors explore combining their inventories in either physical or virtual warehouses.

To analyze a decentralized transshipment system, we develop a two-stage model that is similar to those of Anupindi et al. (2001) and Granot and Sošić (2003), but we consider the formation of *networks* instead of *coalitions* in the prior work. In the first stage, firms make their inventory decisions independently under uncertain demands. After the realization of the demands, some firms may have leftover inventories, while others may have unsatisfied demands. In the second stage, the firms cooperate by transshipping residual supplies to satisfy residual demands. Then the additional profits generated from transshipment are allocated among participants according to a predetermined rule. We analyze this model backwards with emphasis on the second-stage network analysis. Our second-stage analysis contains two parts: (1) examining the stability of an existing network, and (2) predicting networks to be established by firms when there is no existing network. Specifically, in the first part, we extend the concept of the core defined in a coalition-based cooperative game of transshipment into a network-based cooperative game. In a coalition-based cooperative game, to determine whether an allocation is in the core, one needs to show that no subset of firms has an incentive to form subcoalitions by seceding from the grand coalition. On the other hand, in a network-based cooperative game, we need to examine the incentives of firms to form *subnetworks* by seceding from a given network. In the second part, we derive equilibrium network structures when firms form networks *endogenously* under the allocation rule based on firms' marginal contributions to their networks. In this case, each firm simultaneously announces a set of firms to which it wants to set up a link, and a link is established when two firms have announced each other. For our first-stage analysis, we investigate the inventory decisions of the firms, and determine the conditions under which the Nash equilibrium inventory levels coincide with the inventory levels in a centralized system (i.e., the first-best inventory levels).

Our main contributions are three-fold. First, we present a novel model of inventory transshipment networks, and construct a dual price allocation that is in the core for any network structures and for any residual amounts. Second, we examine the allocation based on the marginal contribution of each firm to its network that is proposed first by Myerson (1977) and is refined later by Jackson and Wolinsky (1996) (which we call the "MJW value"). Under this allocation, various network structures such as complete, hub-spoke, and chain networks are stable only under certain conditions on residual amounts. Moreover, these conditions differ across network structures, implying that a network structure plays an important role in establishing the stability of a decentralized distribution system. While the previous coalition-based approach examines only the grand coalition (i.e., complete network), we find that the complete network tends to be less stable than incomplete networks under the MJW value. Finally, while the previous OM literature on supply chain networks (as reviewed by Netessine 2009) commonly assumes that a network of firms is given exogenously, we analyze the case when firms establish networks endogenously, and show that pairwise Nash stable networks underperform the corresponding networks in centralized systems.

2 Related Literature

Our paper is related to the stream of research that studies transshipment of inventories in *decentralized* systems. For a comprehensive survey of the literature on inventory transshipment, the reader is referred to Paterson et al. (2011).

Two pioneering papers in this stream are Anupindi et al. (2001) and Granot and Sošić (2003). Similar to our model (except that we take a network approach), Anupindi et al. (2001) analyze a two-stage game among multiple firms that sell a common product. They show that a dual price allocation is always in the core in the second stage of transshipment. In addition, for the first-stage decisions, they establish conditions under which the Nash equilibrium inventory levels coincide with the inventory levels in a centralized system. While Anupindi et al. (2001) assume that firms share all residual demands and supplies, Granot and Sošić (2003) consider a model in which each firm decides the amount of its residuals it wants to share with other firms. They show that the dual price allocation, although it is in the core, fails to induce the firms to share all of their residuals. Alternatively, the allocation based on the Shapley value (Shapley 1953), which allocates profits according to firms' marginal contributions to their coalitions, induces the firms to share all of their residuals, but it is not always in the core.

Several researchers have studied transshipment in vertically decentralized supply chains. Dong and Rudi (2004) and Zhang (2005) consider transshipment in a two-tier supply chain, where a supplier sells to a single downstream firm with several locations. Slikker et al. (2005) consider a newsvendor game in a supply chain with a single supplier and multiple retailers who coordinate their orders to the supplier. They show that the game has a non-empty core. Özen et al. (2008) consider a supply chain in which multiple firms coordinate their orders supplied from multiple warehouses to increase their joint profits. Kemahlıoğlu-Ziya and Bartholdi (2011) study the use of the Shapley value when a supplier and multiple retailers whose orders are filled from the common pool can form an inventory-pooling coalition. Shao et al. (2011) analyze the case in which a supplier sells to multiple independent firms, and compare their results with those in the case where the firms are under joint ownership.

There are some papers including Zhao et al. (2005), Rong et al. (2010), and Huang and Sošić (2010a) that analyze transshipment games in multiple periods. In particular, Huang and Sošić (2010a) extend Granot and Sošić (2003) into a repeated game, and show that it is a subgame-perfect Nash equilibrium for firms to share all of their residuals under the dual price allocation when a discount factor is sufficiently high. Sošić (2006) extends Granot and Sošić (2003) by proving that, although the Shapley value is not always in the core, the grand coalition is farsighted stable in the sense that no firms will secede from the grand coalition if they consider how others would react to their actions.

Another stream of research studies the role of transshipment prices for distribution of residual profits. Using non-cooperative game theory, Rudi et al. (2001) and Hu et al. (2007) analyze a model in which the residual profit from transshipment is allocated between two firms using predetermined prices. Rudi et al. (2001) develop the expressions for transshipment prices that achieve a first-best outcome. Hu et al. (2007) extend Rudi et al. (2001) by considering the case when there are finite and uncertain capacities. When multiple firms cooperate in transshipment, Huang and Sošić (2010b) compare the performance of the two methods: the use of predetermined prices and the dual price allocation. They show that neither allocation method dominates the other.

To our knowledge, Hezarkhani and Kubiak (2010) is the only work that analyzes the decentralized transshipment problem using the concept of pairwise stability. They study a matching problem in a two-sided market in which a link is established only between a firm with residual supplies and a firm with residual demands. The core allocation they consider is from the coalition-based cooperative game, which does not take into account any network structure. In contrast, we allow for general network structures under which any two firms can build links, and both direct transshipment between two firms and indirect transshipment through intermediate firms are possible. Thus, in our model, firms' positions in a network play a critical role in establishing stability results.

While our paper focuses on the inventory transshipment problem, cooperative game theory has been applied to several other problems in the OM literature. These include bargaining in supply chains (e.g., Reyniers and Tapiero 1995), alliance formation among competing firms (e.g., Nagarajan and Sošić 2007), decentralized assembly systems (e.g., Granot and Yin 2008, Nagarajan and Sošić. 2009, Yin 2010), group buying (e.g., Chen and Yin 2010, Nagarajan et al. 2010), and capacity allocation and scheduling (e.g., Hall and Liu 2010). Readers are referred to Nagarajan and Sošić (2008) for the review of early work. Note that all these papers analyze coalition-based cooperative games.

Our paper contributes to this literature by applying the theory of social and economic networks to the decentralized inventory transshipment problem. Our network-based approach provides a richer form of representing relationships among firms than the previous coalition-based approach, and enables us to establish the stability of partial cooperation structures based on networks.

3 Coalition-Based Cooperative Transshipment Games

In this section, we first describe the simplified variant of the model in Anupindi et al. (2001), which excludes the possibility of storing stocks at commonly shared warehouses. This variant has also been used by Granot and Sošić (2003) in comparing their work with Anupindi et al. (2001). We then introduce some concepts of coalition-based cooperative game theory, and summarize the main results of those two papers. This will help understand how the previous coalition-based approach differs from our network-based approach that will be presented in subsequent sections.

Consider a set $N = \{1, 2, ..., n\}$ of firms who sell a common product. The firms make decisions in two stages as follows. In the first stage, each firm $i \in N$ determines its order quantity X_i under uncertain demand D_i by taking into consideration other firms' inventory decisions as well as transshipment in the following stage. Define inventory profile $X = \{X_1, ..., X_n\}$ and demand profile $D = \{D_1, ..., D_n\}$.[2] Let c_i denote a unit ordering cost. After the firms receive what they have ordered, demands are realized. Firm i generates revenue r_i for each unit of the demand satisfied, and any excess inventory can be salvaged at u_i $(< r_i)$ per unit. The model assumes that each firm satisfies its own demand first using its local inventory, and then ship any excess inventories to satisfy unmet demands of other firms. Let $H_i = \max\{X_i - D_i, 0\}$ and $E_i = \max\{D_i - X_i, 0\}$ denote the residual supply and demand of firm i, respectively. In the second stage, the firms transship the residual supplies to satisfy the residual demands. Throughout the paper, we limit our interest to the case where at least one firm has residual demand and at least one firm has residual supply because otherwise no transshipment will occur. Let Y_{ij} denote the number of units shipped from firm i

[2] This model allows the demand of some firm i, D_i, to be zero. Such firms can be viewed as independent warehouses that are built only for transshipment.

to firm j, and let t_{ij} denote the unit transportation cost associated with this shipment. Further, define transshipment pattern $Y = \{Y_{ij} : i, j = 1, ..., n\}$. Since both firms with residual supplies and firms with residual demands contribute to the additional profits generated from transshipment, the profits are allocated among these firms according to a predetermined allocation rule.

To characterize the transshipment in the second stage of the game, Anupindi et al. (2001) and Granot and Sošić (2003) use the concepts from coalition-based cooperative game theory. A pair (N, w) is called a *cooperative game*, in which $w : 2^N \rightarrow R$ is a *characteristic function* of the game. A subset S of N is called a *coalition* and N itself is the *grand coalition.* The characteristic function $w(S)$ captures the value generated by a coalition S. An *allocation rule* in a cooperative game is a function $\phi : W(N) \rightarrow R^n$, where $W(N)$ is a set of all such games on the set N. An allocation rule is efficient if $\sum_{i \in N} \phi_i(w) = w(N)$. An efficient allocation rule specifies how much of the value generated by the grand coalition is attributed to each firm. We consider only efficient allocation rules in this paper. When an allocation rule applies to a particular game (N, w), we call it an *allocation*, denoted as $\varphi = \{\varphi_1, ..., \varphi_n\}$. The *core* defines a set of allocations with a stability property. We say that an efficient allocation φ is a member of the core of (N, w) if $\sum_{i \in S} \varphi_i \geq w(S)$ for all $S \subseteq N$; i.e., for any subset of firms, the sum of allocations they receive in the grand coalition is at least as large as the value that they can generate by forming a subcoalition. A core allocation leads to a stable outcome of the cooperative game in the sense that no subset of firms has an incentive to secede from the grand coalition.

The allocation mainly considered by Anupindi et al. (2001) is a dual price allocation. To find dual prices to be used in this allocation, they first formulate the centralized transshipment problem in which a single decision-maker optimizes a transshipment pattern of all firms. Specifically, given inventory profile X and demand profile D, the central decision-maker solves the following program:

$$CTP(X, D) = \max_{Y} \sum_{i,j \in N} (r_j - u_i - t_{ij}) Y_{ij} \qquad (1)$$

$$s.t. \sum_{j \in N} Y_{ji} \leq E_i \quad \forall i \in N \qquad (2)$$

$$\sum_{j \in N} Y_{ij} \leq H_i \quad \forall i \in N \qquad (3)$$

$$Y_{ij} \geq 0 \quad \forall i, j \in N. \qquad (4)$$

In (1), $(r_j - u_i - t_{ij})$ represents the profit of transshipping one unit of inventory from i to j. The constraints in (2) ensure that the number of units transshipped to firm i does not exceed its residual demand E_i. The constraints in (3) ensure that the number of units transshipped from firm i does not exceed its residual supply H_i. Let γ_i and δ_i be the dual prices associated with the constraints in (2) and (3), respectively. According to the dual price allocation, each firm $i \in N$ is allocated

$\gamma_i E_i + \delta_i H_i$. The main result of Anupindi et al. (2001) is that the dual allocation is in the core of the cooperative transshipment game for any residual supplies and demands.

Instead of the dual price allocation, Granot and Sošić (2003) consider the Shapley value (Shapley 1953). The Shapley value has many desirable properties such as symmetry, efficiency, and additivity (see details in Shapley (1953) or Granot and Sošić (2003)), and hence it is commonly used in the literature. The allocation rule based on the Shapley value is defined as follows: for firm $i \in N$,

$$\phi_i^{SV}(w) = \sum_{S \subseteq N \setminus \{i\}} \{w(S \cup \{i\}) - w(S)\} \left\{ \frac{\#S!(n - \#S - 1)!}{n!} \right\}, \tag{5}$$

where $\#S$ is the number of firms in S. The standard interpretation of the Shapley value is as follows (e.g., in Jackson (2006)). Consider all possible orderings of firms. For each ordering, consider building a society by adding one firm at a time into that order. A firm obtains the marginal contribution that she/he makes to the society when added to the coalition formed by the firms before her/him in the order. So, a firm i whose place in the order follows a coalition S receives value $w(S \cup \{i\}) - w(S)$. Since there are $\#S!(n - \#S - 1)!$ such orderings, averaging over all possible orderings $n!$ leads to the Shapley value. Granot and Sošić (2003) have shown that the Shapley value is not always in the core of the cooperative transshipment game although it induces the firms to share all their residuals.

4 Network-Based Cooperative Transshipment Games

This section is organized as follows. In §4.1, we present our network model of the decentralized transshipment problem, and describe the concepts of network-based cooperative game theory that we use to analyze our model. In §4.2, we examine whether the pair of a given network and a dual price allocation is always in the core of the transshipment game. Finally, in §4.3, we examine the conditions under which the MJW value is in the core for various network structures. Throughout this section, we highlight the difference between our network-based approach and the coalition-based approach reviewed in §3.

4.1 Model of Transshipment Networks

Our model has the same sequence of the decisions as in the previous coalition-based model described in §3. The main difference is that transshipment in the second stage of our model occurs based on a given network g. A node in g represents a firm in $N = \{1, 2, ..., n\}$. We call a firm with residual supply a *supply node*, and call a firm with residual demand a *demand node*. A bidirectional link ij between nodes i and j represents a partnership between firm i and firm j, so that two firms can transship inventory directly between each other. We write $ij \in g$ to indicate that nodes i and j are linked in the network g, and describe a network with a set of links. For example, for

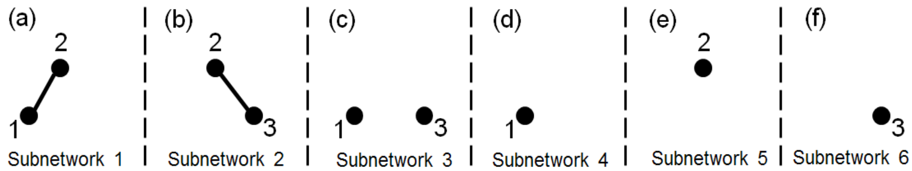

Figure 2: Subnetworks of the network $g' = \{12, 23\}$: (a) $g'|_{\{1,2\}}$, (b) $g'|_{\{2,3\}}$, (c) $g'|_{\{1,3\}}$, (d) $g'|_{\{1\}}$, (e) $g'|_{\{2\}}$, and (f) $g'|_{\{3\}}$.

$N = \{1, 2, 3\}$, $g = \{12, 23\}$ represents a network with two links: one link between nodes 1 and 2 and the other link between nodes 2 and 3. In the case of a regional blood management system, we may represent a regional blood center as node 2, and two blood banks as nodes 1 and 3 that share their residual supplies and demands through the regional blood center. We denote the network obtained by adding link ij to g by $g + ij$, and denote the network obtained by deleting link ij from g by $g - ij$. Further, we use g^N to denote the *complete network on* N in which a link exists between any two nodes $i, j \in N$. A *path* between firms i and j in g is a sequence of firms $i_1, i_2, ..., i_\kappa$ such that $i_k i_{k+1} \in g$ for each $k \in \{1, 2, ..., \kappa - 1\}$ with $i_1 = i$ and $i_\kappa = j$. Transshipment can occur between two firms *directly* through a link between them or *indirectly* through a path between them. We use Y_{ij} to denote the number of units shipped *directly* from firm i to firm j. If H_i units of inventories are shipped *indirectly* from firm i to firm j through a path $i_1, i_2, ..., i_\kappa$ with $i_1 = i$ and $i_\kappa = j$, then we have $Y_{i_k i_{k+1}} = H_i$ for $k = \{1, 2, .., \kappa - 1\}$.

We use the solution concepts in a network-based cooperative game proposed by Jackson and Wolinsky (1996). In this game, the value of a coalition not only depends on the amount of residuals of its members, but also on the structure of the network formed by its members. Similar to the characteristic function in the coalition-based approach, the value function in the network setting is defined as $v : G(N) \to R$, where $G(N)$ denotes a set of all possible networks over N. An allocation rule in a cooperative game can be extended naturally into the network setting as a function $\phi : G(N) \times V(N) \to R^n$, in which $V(N)$ represents the set of all possible value functions for a society N. Since allocations depend on the structure of a network, so does the specification of the core. We say that a pair of a network and an efficient allocation, (g, φ), is a member of the core of (N, v) if $\sum_{i \in S} \varphi_i(g) \geq v(g|_S)$ for all $S \subseteq N$, where $g|_S$ means a subnetwork of g restricted to the firms in S. For example, Figure 2 shows all subnetworks of the network $\{12, 23\}$ (denote as g') shown earlier in Figure 1(c). The requirement of the core is that no subnetworks can deviate and generate a higher value than what they are being allocated in the initial network g. So, the core allocation is stable to deviations from the network g.

Given inventory profile X and demand profile D, the value function v in our model specifies the maximum profit that any subnetwork $g|_S$ can generate through transshipment. Specifically, we define v as follows:

$$v(g|_S) = \max_Y \sum_{i \in S} \left[a_i \left(\sum_{j \in B_i(g|_S)} Y_{ji} - \sum_{j \in B_i(g|_S)} Y_{ij} \right) - \sum_{j \in B_i(g|_S)} Y_{ij}(t_{ij} + u_i - u_j) \right] \tag{6}$$

$$s.t. \quad \sum_{j \in B_i(g|_S)} Y_{ji} - \sum_{j \in B_i(g|_S)} Y_{ij} \leq E_i \quad \forall i \in S \tag{7}$$

$$\sum_{j \in B_i(g|_S)} Y_{ij} - \sum_{j \in B_i(g|_S)} Y_{ji} \leq H_i \quad \forall i \in S \tag{8}$$

$$Y_{ij} \geq 0 \quad \forall i, j \in S, \tag{9}$$

where $B_i(g)$ denotes the set of nodes that have a link to node i in g, and a_i is defined as $r_i - u_i$ for $i \in S$ with $E_i > 0$, and 0 for $i \in S$ with $E_i = 0$. In (6), $a_i(\sum_{j \in B_i(g|_S)} Y_{ji} - \sum_{j \in B_i(g|_S)} Y_{ij})$ represents the net revenue from transshipment, and $\sum_{j \in B_i(g|_S)} Y_{ij}(t_{ij} + u_i - u_j)$ represents the cost of transshipment which includes transportation costs and differences in salvage values. The constraints given in (7) (resp., in (8)) ensure that the difference between the number of units shipped to firm i and that from firm i does not exceed the residual demand (resp., supply) of firm i. Note that $v(g|_S)$ depends on inventory profile X and demand profile D, but we suppress (X, D) for notational convenience. In the rest of this paper, we refer to the program (6)-(9) as *the transshipment problem within* S, and denote its optimal transshipment pattern by $Y^{g|_S}$.

4.2 Dual Price Allocation in Transshipment Networks

This subsection examines whether or not a pair of a given network g and the dual price allocation is in the core of our network-based transshipment game. Anupindi et al. (2001) have shown that the dual price allocation is always in the core when transshipment occurs among members of a coalition. Because the grand coalition is essentially the same as the complete network, a pair of the *complete* network and the dual price allocation is also in the core of our network-based game.

To examine whether or not a pair of an *incomplete* network and the dual allocation is in the core, we refine the dual price allocation of Anupindi et al. (2001) such that it depends on the structure of a given network g. Similar to $CTP(X, D)$ given in (1)-(4), we define the centralized transshipment problem $CTP(g, X, D)$ for a network g as the program (6)-(9) with $S = N$. Then the dual price allocation $\varphi^{DP}(g)$ allocates $\gamma_i^g E_i + \delta_i^g H_i$ to firm $i \in N$, where γ_i^g and δ_i^g are the dual prices associated with the constraints (7) and (8) with $S = N$ in the network g, respectively. To illustrate the dual price allocation $\varphi^{DP}(g)$ in the network-based transshipment game, consider the following example from Anupindi et al. (2001):

Example 1: Suppose that $r_i = 10$ (\$/unit), $u_i = 5$ (\$/unit), $t_{ij} = 1$ (\$/unit) $\forall i, j \in N =$

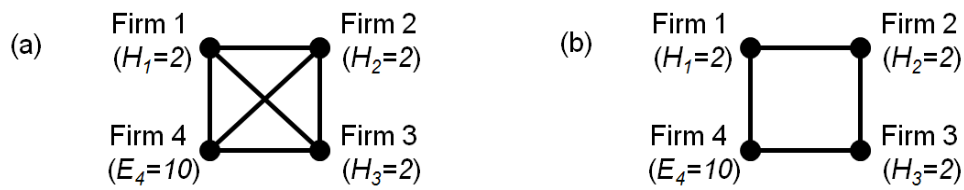

Figure 3: Examples of networks when four firms exist in the market: (a) the complete network $g^{\{1,2,3,4\}} = \{12, 13, 14, 23, 24, 34\}$, and (b) chain $\{12, 23, 34, 14\}$.

$\{1, 2, 3, 4\}$, and that $H_1 = H_2 = H_3 = 2$ (units) and $E_4 = 10$ (units).

When the grand coalition or the complete network is formed as shown in Figure 3(a), the total profit from transshipment is: $\min\{H_1+H_2+H_3, E_4\}\times(r_i-u_i-t_{ij}) = 6$ (units) $\times$ 4 (\$/unit) = \$24. Anupindi et al. (2001) have shown that the dual price allocation based on the program given in (1)-(4) is $\{\$8, \$8, \$8, 0\}$. This allocation assigns all profits from transshipment to firms 1, 2, and 3 with residual supplies, and none to firm 4 with residual demands because the total residual demands of 10 units are much larger than the total residual supplies of 6 units. Next, consider a "chain" network of $\{12, 23, 34, 14\}$ as shown in Figure 3(b). In this network, the maximum profit from transshipment is \$22 because firm 2 has to ship its residual supplies of 2 units to firm 4 via firm 1 or firm 3, incurring an additional transportation cost of \$2. Now suppose that firms 1, 2, and 3 still receive the same allocation, so that each of these firms receives \$22/3. Then this network is no longer stable because a subnetwork of firms 1, 3, and 4 can generate a total profit of 4 (units) $\times$ 4 (\$/unit) = \$16, which is larger than the sum of allocations, \$44/3, they receive under the initial network. To prevent firms 1, 3 and 4 from forming a subnetwork, our dual price allocation based on $CTP(g, X, D)$ decreases the allocation of firm 2 by \$2, so that $\varphi^{DP}(\{12, 23, 34, 14\}) = \{\$8, \$6, \$8, 0\}$. It is important to observe that firm 2 has a worse position in the network than firms 1 and 3 because firm 2 is not linked directly to firm 4 which is the only demand node in this example.

The above example highlights that the dual price allocation $\varphi^{DP}(g)$ takes into account the position of a firm in the network g as well as the amount of its residual supply or demand. More generally, we can establish the following result about the stability of a network under the dual price allocation. All proofs are provided in Appendix.

Proposition 1 *For any given network g, the pair $(g, \varphi^{DP}(g))$ is in the core of the network-based cooperative game (N, v) for any residual supplies and demands.*

Proposition 1 demonstrates that the dual price allocation is an ideal allocation for the stability of incomplete networks as well as the complete network. This result may not be surprising if one could note that the dual price allocation based on $CTP(X, D)$ under the coalition-based approach could yield the same allocation as φ^{DP} based on $CTP(g, X, D)$ under our network-based approach. To understand this, note the following difference between $CTP(X, D)$ and $CTP(g, X, D)$. Our network-based approach explicitly models the actual path $i_1, i_2, ..., i_\kappa$ with $i_1 = i$ and $i_\kappa = j$ over which inventories are transshipped between any two firms i and j in a network. As a result, the solution of $CTP(g, X, D)$ provides us with the *optimal* path for transshipment between any two firms i and j among a number of potential paths between those two firms. In contrast, the coalition-based approach specifies *one* transportation cost t_{ij} of transshipment between any two firms i and j; this means that a decision-maker needs to choose one path a priori among a number of potential paths between firms i and j. If we redefine transportation costs under the coalition-based approach such that they reflect actual transshipment paths (i.e., set t_{ij} equal to a minimum transportation cost among all paths between every pair of nodes i and j in a network), then this approach yields the same allocation as φ^{DP} under our network-based approach.[3] Therefore, the dual price allocation is in the core of our cooperative game for any network g.

The implication of Proposition 1 is as follows: If the goal of an allocation mechanism is to maximize the cooperation of independent firms, thereby maximizing the total value generated from transshipment, then the dual price allocation can achieve this goal. For example, in a blood management system, the transshipment of unused units of blood can potentially save the lives of people who urgently need those units. For a non-profit association of local blood banks, the dual price allocation can incentivize independent blood banks to participate in this cooperative network.

Despite the core property of the dual price allocation, the dual price allocation may not be intuitively appealing. For instance, in Example 1, firm 4 is the only demand node in the network (without which no subnetwork can generate any profit from transshipment), but receives zero allocation. In fact, by the definition of dual prices, if the constraints (7) and (8) associated with a firm are not binding, then the firm receives zero under the dual price allocation. Thus, as shown by Granot and Sošić (2003), if firms can choose how much residual demand or supply to share, then they may not share all of their residuals under the dual price allocation; on the other hand, the allocation based on the Shapley value induces firms to share all of their residuals because firms can receive a larger allocation by doing so when the allocation is determined according to firms' marginal contributions. (We can easily show that this result continues to hold in our network setting.) For this reason, the allocation based on firms' marginal contributions is more commonly

[3] For instance, if we set $t_{13} = t_{24} = t_{31} = t_{42} = 2$ (\$/unit) in Example 1, the coalition-based approach yields the same dual price allocation as the network-based approach. However, the computation of minimum transportation costs may become complicated as the number of firms and the complexity of a network structure increase because one needs to solve an optimization problem for every pair of nodes in the network.

used in the literature (see §2). We next examine the stability of a network under the allocation based on the marginal contribution of each firm to its network, i.e., the MJW value.

4.3 Allocation Based on Marginal Contributions in Transshipment Networks

In this subsection, we examine whether or not the pair of a given network g and the MJW value is in the core of our network-based transshipment game. In the rest of this paper, we say that a network g is *MJW-stable* if the pair of g and the MJW value is in the core of the transshipment game. The allocation rule based on the MJW value for a given network g is defined as follows:

$$\phi_i^{MJW}(v,g) = \sum_{S \subseteq N \setminus \{i\}} \{v(g|_{S\cup\{i\}}) - v(g|_S)\} \left\{ \frac{\#S!(n - \#S - 1)!}{n!} \right\} \text{ for } i \in N. \tag{10}$$

Similar to the Shapley value, the MJW value captures firms' marginal contributions to the network g, and preserves many desirable properties of the Shapley value (Jackson and Wolinsky 1996). The MJW value can be computed using the procedure similar to the Shapley value calculation, but it is based on how the value changes as firm i is added to the network comprising firms in S; i.e., $v(g|_{S\cup\{i\}}) - v(g|_S)$. To illustrate how the MJW value differs from the Shapley value, we consider the following example of three firms from Sošić (2006):

Example 2: Suppose that $r_i = 1$ (\$/unit), $u_i = 0$ (\$/unit), and $t_{ij} = 0$ (\$/unit) for $i, j \in N = \{1, 2, 3\}$, and that firm 1 and firm 2 have residual demands E_1 and E_2, respectively, with $E_1 \leq E_2$, while firm 3 has the residual supply H_3.

In this example, there are four possible scenarios: (1) $H_3 \geq E_1 + E_2$, (2) $E_2 \leq H_3 < E_1 + E_2$, (3) $E_1 \leq H_3 < E_2$, and (4) $H_3 < E_1$. For these scenarios, Table 1 presents the allocations based on the Shapley value as well as the allocations based on the MJW value with the network $g' = \{12, 23\}$ shown earlier in Figure 1(c). (See Appendix for the procedure of computing the MJW value.)

Table 1: Allocations based on the Shapley value and the MJW value

	Shapley value (= MJW value in $g^{\{1,2,3\}}$)			MJW value in $g' = \{12, 23\}$		
Scenario	Firm 1	Firm 2	Firm 3	Firm 1	Firm 2	Firm 3
(1) $H_3 \geq E_1 + E_2$	$\frac{E_1}{2}$	$\frac{E_2}{2}$	$\frac{E_1+E_2}{2}$	$\frac{E_1}{3}$	$\frac{2E_1+3E_2}{6}$	$\frac{2E_1+3E_2}{6}$
(2) $E_2 \leq H_3 < E_1 + E_2$	$\frac{E_1+2H_3-2E_2}{6}$	$\frac{E_2+2H_3-2E_1}{6}$	$\frac{E_1+E_2+2H_3}{6}$	$\frac{H_3-E_2}{3}$	$\frac{2H_3+E_2}{6}$	$\frac{2H_3+E_2}{6}$
(3) $E_1 \leq H_3 < E_2$	$\frac{E_1}{6}$	$\frac{3H_3-2E_1}{6}$	$\frac{E_1+3H_3}{6}$	0	$\frac{H_3}{2}$	$\frac{H_3}{2}$
(4) $H_3 < E_1$	$\frac{H_3}{6}$	$\frac{H_3}{6}$	$\frac{4H_3}{6}$	0	$\frac{H_3}{2}$	$\frac{H_3}{2}$

Observe from Table 1 that the allocations based on the MJW value in g' are different from those based on the Shapley value under all four scenarios. With the Shapley value, the allocation to firm 3 is strictly larger than the allocation to firm 1 or firm 2 in all four scenarios. This is because there exists only one supply node (firm 3) while there are two demand nodes (firm 1 and firm 2), and thus the marginal contribution of the supply node is more significant. However, under the MJW value, firms 2 and 3 in the network g' receive the same allocations in all four scenarios. The intuition is that both firms play equally important roles in the transshipment network in the following sense: without firm 2, no transshipment can occur because firm 1 and firm 3 are not connected to each other, and without firm 3, there is no residual supply to transship. As shown in this example, the MJW value takes into account the structure of a network as well as the residual amounts of firms.

In the following, we examine whether or not the pair $(g, \varphi^{MJW}(g))$ is in the core: (1) when g is the complete network in §4.3.1, and (2) when g is an incomplete network in §4.3.2.

4.3.1 Complete Network

In the complete network, any two firms can transship inventory directly between each other. Thus one can expect that our network-based approach yields the same result as the coalition-based approach. For Example 2, Sošić (2006) has shown that the Shapley value is in the core in Scenario 1, whereas it is not in the core in Scenarios 2, 3, and 4. Thus the Shapley value is not always in the core. Because the allocations under the MJW value in the complete network are the same as those under the Shapley value, in general, the pair $(g^N, \varphi^{MJW}(g^N))$ is *not always* in the core of our network-based transshipment game.

However, we can show that the pair $(g^N, \varphi^{MJW}(g^N))$ is in the core *under a certain condition* on the residual supplies and demands. To establish this result, we use the notion of *convexity* in cooperative game theory. Let the cooperative game (N, v) be convex if $v(g|_{S''\cup\{i\}}) - v(g|_{S''}) \leq v(g|_{S'\cup\{i\}}) - v(g|_{S'})$ for any subsets S' and S'' of N that satisfy $S'' \subset S'$ and any firm $i \notin S'$. Convexity refers to the property that, as the number of firms in the subnetwork the firm joins increases, the marginal contribution of a firm (hence its allocation under the MJW value) weakly increases. Thus, if (N, v) is convex, then no subnetworks have incentives to secede from the grand coalition, and therefore the pair $(g^N, \varphi^{MJW}(g^N))$ is in the core of (N, v) (Jackson 2006). In our transshipment game, suppose there exist only one demand node and multiple supply nodes. The marginal contribution of the demand node is non-decreasing with the number of firms in any subnetwork because there are more residual supplies within the subnetwork having more firms. For the marginal contribution of a supply node to be non-decreasing with the number of firms in a subnetwork, the subnetwork should have sufficient residual demands to generate additional profits from the supply of the node. The next proposition bears this intuition.

Proposition 2 *The pair $(g^N, \varphi^{MJW}(g^N))$ is in the core if there is one supply (resp., demand) node i and its residual supply H_i (resp., demand E_i) is greater than or equal to the sum of all* *residual demands (resp., supplies), i.e.* $H_i \geq \sum_{j \in N\setminus\{i\}} E_j$ *(resp.,* $E_i \geq \sum_{j \in N\setminus\{i\}} H_j$*).*

In Example 2, Scenario 1 satisfies the condition given in Proposition 2, whereas Scenarios 2, 3, and 4 do not satisfy this condition because there is only one supply node in the network but its residual supply (H_3) is not sufficient to satisfy all residual demands ($E_1 + E_2$). In those scenarios, a subset of firms will form a subnetwork to earn more profit than the allocations they would receive in the initial network $g^{\{1,2,3\}}$ under the MJW value. For example, in Scenario 3, firms 2 and 3 will form a subnetwork to earn the total profit of H_3 which is greater than the sum of their allocations under the MJW value in g^N, $H_3 - E_1/6$. For the case where multiple demand nodes and multiple supply nodes exist in g^N, we show in the proof that the game is no longer convex.

By noting that the condition given in Proposition 2 is very restrictive, we can conclude that the complete network is not MJW-stable in most scenarios. The primary reason for such instability is that any two firms can transship inventory directly between each other in the complete network, and firms often find it more profitable to secede from the complete network. In a regional blood management system, for example, there is often a designated blood center through which other blood banks transship their unused units of blood. One might temp to think that a complete network, which allows direct transshipment between blood banks without going through a regional blood center, may facilitate transshipment further because such direct shipment can potentially save the costs of transportation and coordination. However, our result shows that such a complete network may not be MJW-stable because a subset of blood centers may form their own subnetwork.

Because direct transshipment between every pair of firms is not possible in an *incomplete* network, we may expect that incomplete networks could be more stable for transshipment among independent firms. We examine this issue formally next.

4.3.2 Incomplete Networks

When a network g is incomplete, the contribution of each firm to transshipment depends on the links associated with the firm as well as the amount of its residual supply or demand. As the allocations based on the MJW value vary across different network structures, so do the stability results. We first consider examine the MJW-stability of various incomplete network structures that appear in the literature: (1) a hub-spoke network, (2) a line network, and (3) a chain. Then we demonstrate in (4) how a MJW-stable network with a general network structure can be derived from another (known) MJW-stable network.

(1) Hub-spoke network To illustrate this network, let us first consider Example 2 in the incomplete network $g'' = \{13, 23\}$ shown in Figure 4(a) in which firm 3 can be viewed as the hub. Observe that even if the link 12 existed in this network, it would not change the profits from

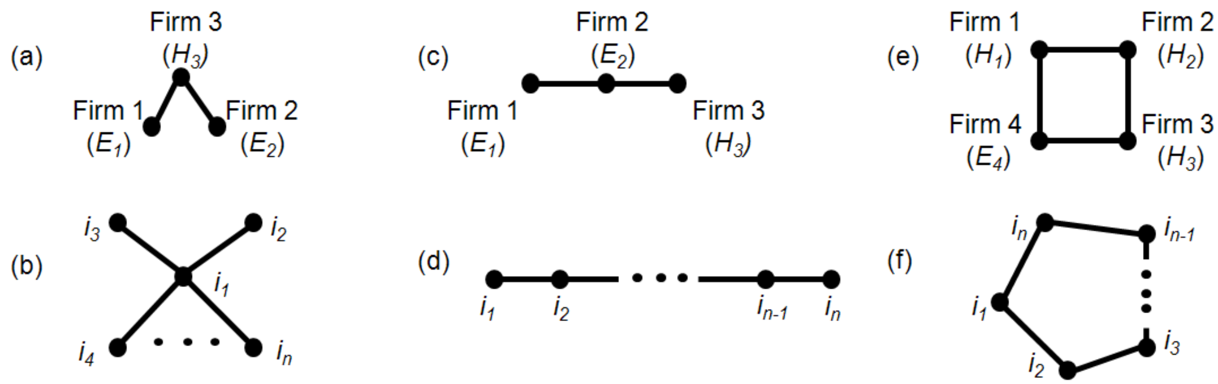

Figure 4: Examples of incomplete networks: (a)-(b) hub-spoke networks, (c)-(d) line networks, and (e)-(f) chain networks.

transshipment among any subset of firms. This is because transportation costs are zero in this example and both firms 1 and 2 are demand nodes. This suggests that the allocations under g'' are the same as those under the complete network. Therefore, from the result of the complete network in §4.3.1, the pair $(g'', \varphi^{MJW}(g''))$ is in the core in Scenario 1, but not in Scenarios 2, 3, and 4.

For a general hub-spoke network g^H with n firms as shown in Figure 4(b), the following proposition presents a sufficient condition under which $(g^H, \varphi^{MJW}(g^H))$ is in the core.

Proposition 3 *Consider a hub-spoke network* $g^H = \{i_1 i_2, i_1 i_3, ..., i_1 i_n\}$ *where* i_1 *is the hub. The pair* $(g^H, \varphi^{MJW}(g^H))$ *is in the core if* $H_{i_1} \geq \sum_{j \in N \setminus \{i_1\}} E_j$ *or* $E_{i_1} \geq \sum_{j \in N \setminus \{i_1\}} H_j$.

Different from Proposition 2 for the complete network, Proposition 3 shows that the number of supply nodes or demand nodes in a hub-spoke network need not be one in order for $(g^H, \varphi^{MJW}(g^H))$ to be in the core. This difference arises due to the special role of the hub in this network; i.e., any subnetwork that does not include the hub generates no value. Recall from §1 that a regional blood management system with one large blood center and multiple smaller blood banks forms a hub-spoke network. Proposition 3 suggests that the regional blood center as the hub needs to have sufficient residual supplies to satisfy all residual demands of blood banks in the spokes in order for this network to be MJW-stable. This is because a subnetwork having the hub should receive at least as large allocation in its initial network as the profit it can generate alone.

(2) Line network To illustrate this network, let us first consider Example 2 in $g' = \{12, 23\}$ shown in Figure 4(c). We refer to such a network as g' in which all demand nodes are placed on one side and all supply nodes are placed on the other side as a "line." Table 1 (earlier) shows the allocations based on the MJW value for g'. Apparently, subnetworks $g'|_{\{1,3\}}$ and $g'|_{\{1,2\}}$ have no incentives to secede from g' because they cannot generate any profit from transshipment. Thus, to determine whether the pair $(g', \varphi^{MJW}(g'))$ is in the core under each scenario, we only need to examine whether a subnetwork $g'|_{\{2,3\}}$ has an incentive to secede from g'. If the subnetwork $g'|_{\{2,3\}}$ is formed, then it can generate a profit of E_2 in Scenarios 1 and 2, and a profit of H_3 in Scenarios 3 and 4. On the other hand, when firms 2 and 3 stay in g', Table 1 shows that the sum of allocations to both firms is $\frac{2E_1+3E_2}{3}$, $\frac{2H_3+E_2}{3}$, H_3, or H_3 in Scenario 1, 2, 3, or 4, respectively. By comparing these profits under each scenario, we find that firms 2 and 3 do not earn more by forming the subnetwork $g'|_{\{2,3\}}$. Therefore, the pair $(g', \varphi^{MJW}(g'))$ is in the core under all four scenarios.

By comparing the stability result of g' with those of the complete network g^N and the hub-spoke network g'', we can generate the following insight. In g^N or g'', we have shown that some demand node is excluded from transshipment (e.g., demand node 2 in Scenario 4) when the residual supply is not sufficient to satisfy total residual demands in Scenarios 2, 3, and 4. In contrast, in g', such a demand node is valuable to the network by providing a path without which indirect transshipment between the other nodes is not possible. For example, demand node 2 in g' provides a path for transshipment between demand node 1 and supply node 3. In this case, we say that the node provides "useful links" to the network.

Building on the intuition from three-firm networks above, we can establish a necessary and sufficient condition for the pair $(g, \varphi^{MJW}(g))$ to be *in the core for any residual amounts.* As shown in Proposition 4, this condition requires that every node in the network g provides useful links to the network, hence contributing to the connectivity of the entire network. Furthermore, the only network structure that satisfies this condition is the line, denoted by g^L, as shown in Figure 4(d).

Proposition 4 *The only network structure, under which any two firms are connected and the pair $(g, \varphi^{MJW}(g))$ is in the core for any residual supply H_i, residual demand E_i, and (r_i, t_{ij}, u_i) $(i, j \in N)$, is the line $g^L = \{i_1i_2, i_2i_3, ..., i_ki_{k+1}, ..., i_{n-1}i_n\}$ where $E_{i_j} = 0$ for $j \in \{1, ..., k\}$ and $H_{i_j} = 0$ for $j \in \{k+1, ..., n\}$ with $1 \leq k \leq n-1$.*

(3) Chain For a network g having a structure other than the line, Proposition 4 suggests that there always exist some amounts of residual supplies and demands with which some subnetwork has an incentive to secede from the initial network g under the MJW value. Let us illustrate this result using chain networks that were studied previously under a centralized system (e.g., Lien et al. 2011). For example, in the chain network $g''' \equiv \{12, 23, 34, 14\}$ shown in Figure 4(e), transshipment between firm 2 and firm 4 can occur either through firm 1 or firm 3. Thus, there are situations in which firm 1 (resp., firm 3) does not provide useful links to the network and a subnetwork $g'''|_{\{2,3,4\}}$ (resp., $g'''|_{\{1,2,4\}}$) deviates from its initial network. As a result, additional conditions on the residual supplies and demands are required for this chain to be MJW-stable. Specifically, the following corollary shows such conditions when link $i_n i_1$ is added to the line g^L described in Proposition 4 so that the network becomes the chain g^C; see Figure 4(f).

Corollary 1 *Consider the chain network* $g^C = \{i_1 i_2, i_2 i_3, ..., i_k i_{k+1}, ..., i_{n-1} i_n, i_n i_1\}$ *where* $E_{i_j} = 0$ *for* $j \in \{1, ..., k\}$ *and* $H_{i_j} = 0$ *for* $j \in \{k+1, ..., n\}$ *with* $1 \leq k \leq n-1$. *The pair* $(g^C, \varphi^{MJW}(g^C))$ *is in the core if* $\sum_{i \in N} H_i = \sum_{i \in N} E_i$ *and* $\varphi_i^{MJW}(g^C) \leq v(g^C) - v(g^C|_{N \setminus \{i\}})$ $\forall i \in N$.

Corollary 1 shows that two conditions are required for the pair $(g^C, \varphi^{MJW}(g^C))$ to be in the core. The first condition requires that the total residual supplies equal the total residual demands. Suppose the total residual supplies are greater than the total residual demands. Unlike the line, there are two paths between any two firms in this chain. Thus, some supply nodes are excluded from transshipment because neither their residual supplies nor their links are valuable to the network. The second condition requires the allocation to each firm is lower than the firm's marginal contribution when the firm joins the subnetwork of $(n-1)$ firms. This condition ensures that firms have no incentives to cut one path for transshipment by excluding one firm.

(4) Other networks Although the line network presented in Proposition 4 guarantees the MJW value to be in the core for any residual amounts, it is not very likely to observe such a network in practice – especially when a large number of firms exist. Then what other connected networks could be stable under the MJW value? In the following proposition, we demonstrate how a MJW-stable network with an arbitrary network structure can be derived from another MJW-stable network (e.g., a line network).

Proposition 5 *Suppose the pair* $(g, \varphi^{MJW}(g))$ *is in the core. Then the pair* $(g+ij, \varphi^{MJW}(g+ij))$ *is in the core if* $v(g+ij|_S) - v(g|_S) \leq \sum_{k \in S} \{\varphi_k^{MJW}(g+ij) - \varphi_k^{MJW}(g)\}$ *for all* $S \subseteq N$.

We illustrate the result stated in Proposition 5 using an example shown in Figure 5. There are four firms in the market: firms 1, 2 and 3 have residual demands $E_1 = E_2 = E_3 = 1$ (unit), while firm 4 has residual supply $H_4 = 3$ (units). Suppose that $r_i = 1$ (\$/unit) and $u_i = 0$ (\$/unit) for $i \in \{1, 2, 3, 4\}$, and $t_{ij} = 0$ (\$/unit) for $i, j \in \{1, 2, 3, 4\}$ except that $t_{34} = t_{43} = 0.9$ (\$/unit). The line network $g = \{12, 23, 34\}$ in Figure 5(a) is MJW-stable according to Proposition 4. By adding link 24 to this network g, we construct another network $g + 24 = \{12, 23, 34, 24\}$ shown in Figure 5(b). In the following, we show that the initial network g and link 24 satisfy the condition given in Proposition 5, so that the new network $g + 24$ is MJW-stable as well. To begin, consider $S = \{1, 2, 4\}$. This subset contains both firms 2 and 4 that are connected through the new link 24. The initial network $g|_S$ generates no profits from transshipment because firms 1 and 2 are

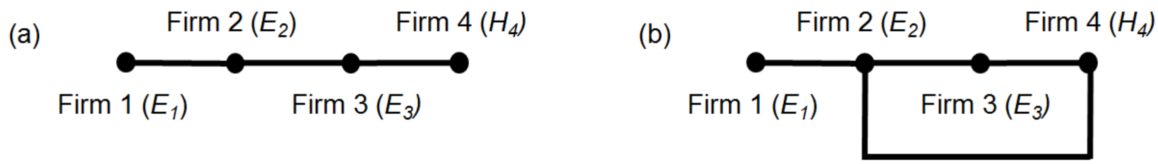

Figure 5: Examples of networks when four firms exist in the market: (a) line $\{12, 23, 34\}$, (b) network $\{12, 23, 34, 24\}$ obtained by adding link 24 to line $\{12, 23, 34\}$.

not connected to firm 4 without firm 3, whereas the new network $g + 24|_S$ generates a profit of 2 (units) $\times$ 1 (\$/unit) $=$ \$2; hence, $v(g + 24|_S) - v(g|_S) =$ \$2. On the other hand, we can show that the total allocation to S under $g|_S$ is \$23/120, while that under $g + 24|_S$ is \$151/60; hence, $\sum_{k\in\{1,2,4\}} \{\varphi_k^{MJW}(g + 24) - \varphi_k^{MJW}(g)\} =$ \$279/120 $(> v(g + 24|_S) - v(g|_S))$. This means that link 24 increases the total allocation to S more than the additional profit it generates for S. This happens because link 24 benefits firms 1, 2, and 4 in subsets other than S, and the total allocation to S reflects the overall impact of link 24 across all subsets of $\{1, 2, 3, 4\}$. Similarly, we can show that the condition given in Proposition 5 is satisfied for $S = \{2, 3, 4\}$ or $\{2, 4\}$. For any subset S that does not contain firm 2, firm 4 or both, link 24 does not generate any additional profit from transshipment (i.e., $v(g + 24|_S) - v(g|_S) = 0$), nor does it decrease the total allocation to S (i.e., $\sum_{k\in S}\{\varphi_k^{MJW}(g + 24) - \varphi_k^{MJW}(g)\} \geq 0$); hence, satisfying the condition given in Proposition 5 as well. Therefore, by Proposition 5, we can conclude that the network $g + 24$ shown in Figure 5(b) is MJW-stable.

Let us summarize our main findings in this subsection. Under the MJW value, the line network is the only network structure that is MJW-stable for any residual amounts. For other networks, MJW-stability requires certain conditions on residual amounts. Interestingly, the complete network requires more stringent conditions than some incomplete networks. Therefore, the complete network is not only more expensive in establishing all the links among firms, but can also be less stable in a decentralized transshipment system. This result bears important implication because the previous coalition-based approach examines only the complete network. Finally, note that this section has examined the stability of an existing network g, assuming that the cost of building a link is a sunk cost. In the next section, we examine the endogenous formation of networks, in which the cost of building a link counts.

5 Endogenous Formation of Transshipment Networks

Suppose that a network g is formed endogenously through a partnership announcement subgame. This subgame happens between the first stage of inventory decisions and the second stage of transshipment in the model described in §4.1. After observing the realized demand, each firm simultaneously announces a set of firms to which it wants to set up a link. In making this decision, a firm takes into account subsequent transshipment based on its network. When both firm i and firm j have announced each other, a link ij is established in the network g, incurring a link cost l_{ij} to firm i and l_{ji} to firm j. In the example of car dealers discussed in §1, dealers with residual demand or supply contact other dealers, and transshipment can occur between two dealers who are mutually interested in sharing their residuals. This process usually involves costs for labor and administrative arrangements (i.e., link costs).

Our objective is to find equilibrium network structures in this subgame. As for the equilibrium concept for this subgame, we use *pairwise Nash stability* that refines Nash equilibrium (Jackson and Wolinsky 1996). Given an allocation φ and link costs, a transshipment network g is *pairwise Nash stable* if the following two conditions are satisfied:

$$\varphi_i(g) \geq \varphi_i(g - ij) + l_{ij} \text{ and } \varphi_j(g) \geq \varphi_j(g - ij) + l_{ji} \text{ for all } ij \in g; \tag{11}$$

$$\text{if } \varphi_i(g + ij) > \varphi_i(g) + l_{ij}, \text{ then } \varphi_j(g + ij) < \varphi_j(g) + l_{ji} \text{ for all } ij \notin g. \tag{12}$$

A network is pairwise Nash stable if no firm wants to sever a link (ensured by (11)) and no two firms want to add a link between them (ensured by (12)). For ease of exposition, we first consider a special case of zero link costs in §5.1, and then analyze a general case of arbitrary link costs in §5.2. In both cases, we assume that the profit generated from transshipment is allocated according to the MJW value, $\varphi^{MJW}(g)$.[4]

5.1 Zero Link Costs

In order to find a pairwise Nash stable network, we choose the complete network as an initial network (which is created when every firm announces all the other firms), and examine if some firms have incentives to deviate. Alternatively, one can start from any network other than the complete network, and analyze whether any firm has an incentive to sever a link or any two firms have an incentive to establish a link between them.

To begin, we use Example 2 to illustrate some properties of a pairwise Nash stable network under the MJW value. We compare the allocations to firms 1 and 3 in the complete network $g^{\{1,2,3\}}$ (shown in the left column of Table 1) with those in the incomplete network $g' = \{12, 23\}$ without link 13 (shown in the right column of Table 1). Table 1 reveals that the allocations to both firms 1 and 3 in $g^{\{1,2,3\}}$ with link 13 are greater than or equal to those in g' without link 13

[4] The MJW value has also been used in establishing pairwise Nash stability of networks in various other applications because it not only has several desirable properties mentioned earlier but also guarantees the existence of a pairwise Nash stable network (Jackson 2006).

in all four scenarios. This is because link 13 increases the marginal contributions of firms 1 and 3 to the network. Similarly, we can show that the allocations to firms 2 and 3 with link 23 are greater than or equal to those without link 23. When both links 13 and 23 exist as in $g^{\{1,2,3\}}$ and $g'' = \{13, 23\}$, whether or not link 12 exists has no impact on the allocations because transshipment between two demand nodes 1 and 2 does not add value to the network (when transportation cost is zero). Therefore, both complete network $g^{\{1,2,3\}}$ and incomplete network g'' are pairwise Nash stable under the MJW value.

We can generalize the above procedure developed for a three-firm network ($n = 3$) to a network with any size ($n \geq 3$). Define *the networks derived from the complete network* as follows: A network g is derived from the complete network g^N if there exists a sequence of networks $g_1, ..., g_\kappa$ with $g_1 = g^N$ and $g_\kappa = g$ such that $g_{k+1} = g_k - ij$ for each $k \in \{1, ..., \kappa - 1\}$, where link ij satisfies $\varphi_i^{MJW}(g_{k+1}) \geq \varphi_i^{MJW}(g_k) - l_{ij}$ or $\varphi_j^{MJW}(g_{k+1}) \geq \varphi_j^{MJW}(g_k) - l_{ji}$. The existence of this sequence of networks implies that, starting from the complete network, there exists a sequence of firms that find it weakly profitable to sever their links. As illustrated above, multiple pairwise Nash stable networks may exist (e.g., $g^{\{1,2,3\}}$ and g'' in Example 2). Using this procedure, we can obtain the following proposition which describes a set of pairwise Nash stable networks for $N = \{1, 2, ..., n\}$.

Proposition 6 *Suppose that a transshipment network is formed by n firms endogenously with no link costs. Then the complete network g^N is always pairwise Nash stable under the allocation based on the MJW value. Furthermore, any pairwise Nash stable network g derived from the complete network g^N contains any link ij with $Y_{ij}^{g^N|_S} > 0$ for some $S \subseteq N$, where $Y^{g^N|_S}$ is the unique optimal transshipment pattern of the transshipment problem within S given in (6)-(9).*

Proposition 6 states that firms have no incentives to sever any link ij which is useful for transshipment within some $S \subseteq N$ (i.e., $Y_{ij}^{g^N|_S} > 0$). For example, in our discussion of Example 2 above, link 13 is useful for transshipment within $S = \{1, 3\}$, and link 23 is useful within $S = \{2, 3\}$; consequently, $g'' = \{13, 23\}$ is pairwise Nash stable under the MJW value. Furthermore, Proposition 6 implies that there is *over-connection* in the pairwise Nash stable networks under the MJW value relative to an optimal transshipment pattern of the centralized transshipment problem. Firms have no incentives to sever any link ij which is useful for *some* S (i.e., $Y_{ij}^{g^N|_S} > 0$) even if that link is not utilized in the optimal transshipment pattern of the centralized transshipment problem for N (i.e., $Y_{ij}^{g^N} = 0$ and $Y_{ji}^{g^N} = 0$). For example, one optimal transshipment pattern of the centralized transshipment problem in Example 2 under Scenario 1 is that $Y_{32} = E_1 + E_2$ and $Y_{21} = E_1$; i.e., firm 3 transships its residual supply of $E_1 + E_2$ ($\leq H_3$) to firm 2, and then firm 2 transships the supply of E_1 to firm 1 after satisfying its residual demand E_2. In this transshipment pattern, link

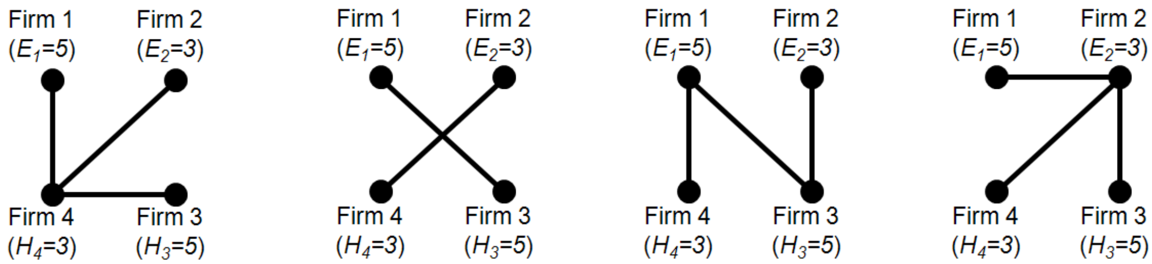

Figure 6: Pairwise Nash stable networks under the MJW value in Example 3.

13 is not utilized. However, the complete network $g^{\{1,2,3\}}$, which is pairwise Nash stable under the MJW value, includes link 13. From a bargaining perspective, firms 1 and 3 have stronger bargaining powers with link 13 when negotiating how to allocate profits from transshipment with firm 2. This is because firms 1 and 3 can still transship E_1 without firm 2; in other words, they have a better outside option with link 13. In the example of car dealers above, link costs may be negligible when a web-based inventory system as well as an industry protocol for transshipment is established well. In this case, a dealer may attempt to build connections with many other dealers so as to increase his bargaining power against others even though he may need only few partners for transshipment eventually. Because there is no cost to improve one's bargaining power in this case, the complete network is always pairwise Nash stable under the MJW value.

5.2 Positive Link Costs

Suppose that link cost l_{ij} or l_{ji} is positive for at least one pair of firms i and j. Following the procedure developed in §5.1, we can identify a set of pairwise Nash stable networks. For example, Figure 6 shows all pairwise Nash stable networks in the following example under the MJW value:

Example 3: Suppose that $r_i = 1$ (\$/unit), $u_i = 0$ (\$/unit), $t_{ij} = 0$ (\$/unit) $\forall i, j \in N$, and that $E_1 = 5$ (units), $E_2 = 3$ (units), $H_3 = 5$ (units), $H_4 = 3$ (units) and $l_{ij} = l_{ji} = 1$ (\$/link) $\forall ij \in g$.

As shown in Figure 6, unlike the case with zero link costs, the complete network is not always pairwise Nash stable under the MJW value because firms face trade-offs between better network positions and additional link costs.

For the case with zero link costs, Proposition 6 in §5.1 has shown that firms tend to be over-connected in pairwise Nash stable networks under the MJW value. Since a positive link cost discourages firms from establishing unnecessary links, one may wonder if firms are still over-connected in this case. To examine this question, we introduce the notion of *efficiency.* We say that a network g is *efficient* if it maximizes the profit from transshipment less the total link cost, $v(g) - \sum_{ij \in g} (l_{ij} + l_{ji})$. Proposition 7 below shows that pairwise Nash stable networks are not, in general, efficient in transshipment under the MJW value. This demonstrates the tension between the stability and efficiency of a network.[5]

Proposition 7 *Suppose that there exist some $S \subset N$ and a link ij with $l_{ij} > 0$ or $l_{ji} > 0$ such that: (i) $v(g^N|_S)$ has a unique optimal transshipment pattern $Y^{g^N|_S}$ with $Y_{ij}^{g^N|_S} > 0$, and (ii) $v(g^N)$ has a unique optimal transshipment pattern Y^{g^N} with $Y_{ij}^{g^N} = 0$ and $Y_{ji}^{g^N} = 0$. Then there exists $\bar{l} > 0$ such that, if $l_{ij} < \bar{l}$ and $l_{ji} < \bar{l}$ $\forall ij \in g$, every pairwise Nash stable network g derived from the complete network g^N is not efficient under the allocation based on the MJW value because there is a subnetwork g^* of g which satisfies $v(g) - \sum_{ij \in g}(l_{ij} + l_{ji}) < v(g^*) - \sum_{ij \in g^*}(l_{ij} + l_{ji})$.*

Proposition 7 suggests that a pairwise Nash stable transshipment network under the MJW value is still over-connected with positive link costs. The inefficiency of a pairwise Nash stable network arises because individual firms try to maximize their own allocations instead of maximizing the aggregate profit of the entire network. The condition under which a stable network is inefficient requires the existence of a link with a positive link cost, which is utilized for transshipment within at least one subset $S \subset N$ (i.e., $Y_{ij}^{g^N|_S} > 0$), but is not utilized in the optimal transshipment pattern of the centralized problem among all firms (i.e., $Y_{ij}^{g^N} = 0$ and $Y_{ji}^{g^N} = 0$). In the example of a car dealer network, this result means that, even if the cost of establishing an connection is not negligible, dealers still build connections with a larger number of other dealers than necessary so as to improve their positions. Such excess connections hurt the efficiency of a decentralized transshipment system. To improve the efficiency, a third-party organization such as a local automobile dealers association or an auto manufacturer may intervene a decentralized dealer network to coordinate the incentives of independent dealers.

6 Inventory Decisions

In this section, we analyze firms' inventory decisions in the first stage of the game. Following Anupindi et al. (2001) and Granot and Sošić (2003), we focus our analysis on whether a certain allocation rule used in the second transshipment stage would lead to inventory decisions in the first stage which are optimal for the centralized system (i.e., achieve the first-best). We will first discuss the case when a network g is given as in §4, and then discuss the case when a network g is formed endogenously as in §5.

[5] This tension is first discovered by Jackson (2006) in the setting where any link in the network generates a positive value; i.e., $v(\{ij\}) > 0$ for any $i, j \in N$. In that case, the complete network is the only pairwise Nash stable network with a small link cost. Our model is different in that $v(\{ij\}) = 0$ is possible when $E_i = E_j = 0$ or $H_i = H_j = 0$ for any $i, j \in N$, so that incomplete networks can also be pairwise Nash stable.

Given a network g, each firm determines its inventory level independently by considering uncertain demands and ex-post transshipment of its residual demand or supply with other firms in the network g. The ex-post total profit of firm i can be expressed as follows: given the network g, demand profile D, its own inventory level X_i, and a vector of other firms' inventory levels X_{-i},

$$\pi_i(g, X_i, X_{-i}, D) = [r_i \min\{X_i, D_i\} + u_i H_i - c_i X_i] + \varphi_i(g, X_i, X_{-i}, D),$$

where the first term in the bracket represents the profit from satisfying its local demand D_i, and the second term φ_i represents the allocation it receives from subsequent transshipment. A Nash equilibrium inventory profile X^{NE} satisfies: $X_i^{NE}(g) = \arg\max_{X_i} E_D[\pi_i(g, X_i, X_{-i}^{NE}, D)]$ for all i. In the coalition-based transshipment game, Anupindi et al. (2001) and Granot and Sošić (2003) have shown that the dual price allocation and the Shapley value do not always lead to the first-best inventory levels, respectively. Because coalitions are special cases of networks, the same results apply to our network-based transshipment game, so that our dual price allocation $\varphi^{DP}(g)$ and the MJW value $\varphi^{MJW}(g)$ do not lead to the first-best inventory levels.

Following the lead of Anupindi et al. (2001) and Granot and Sošić (2003), however, we can construct a new allocation $\varphi^{FB}(g)$ that leads to the first-best inventory levels. While $\varphi^{DP}(g)$ and $\varphi^{MJW}(g)$ determine allocations of firms based on the ex-post profit generated from the second transshipment stage, $\varphi^{FB}(g)$ uses the ex-post profit generated from *both* stages. Let $\Pi(g, X, D)$ denote the sum of ex-post profits of all firms from both stages. We can express Π as follows:

$$\Pi(g, X, D) = \sum_{i \in N} [r_i \min\{X_i, D_i\} + u_i H_i - c_i X_i] + CTP(g, X, D). \tag{13}$$

Corollary 2 *Let $\lambda_i \in (0, 1)$ such that $\sum_{i \in N} \lambda_i = 1$. When a network g is given, the allocation defined by $\varphi_i^{FB}(g) = \lambda_i \Pi(g, X, D) - [r_i \min\{X_i, D_i\} + u_i H_i - c_i X_i]$ for $i \in N$ induces the inventory levels in a first-best solution to be a Nash equilibrium profile X^{NE}.*

The intuition from Corollary 2 is straightforward as follows. The allocation $\varphi^{FB}(g)$ gives each firm i a fixed fraction λ_i of the total profit $\Pi(g, X, D)$ of all firms in both stages less the firm's own profit from the first stage of the game. Then the total profit of a firm from both stages is simply a fixed fraction λ_i of the total profit $\Pi(g, X, D)$ generated by all firms in the network g. Since λ_i is independent of a firm's inventory decision, the firm chooses the inventory level to maximize the total profit $\Pi(g, X, D)$ under this allocation.

When a network g is formed endogenously as in §5, a firm needs to consider the strategic formation of transshipment networks in determining its optimal inventory level, in addition to uncertain demands and ex-post transshipment based on the network g. The total profit $\Pi(g, X, D)$

in this case can be written as

$$\Pi(g,X,D)=\sum_{i\in N}[r_i\min\{X_i,D_i\}+u_iH_i-c_iX_i-\sum_{j\in B_i(g)}l_{ij}]+CTP(g,X,D). \tag{14}$$

Compared with (13), $\Pi(g,X,D)$ in (14) contains the total link cost $\sum_{i\in N}\sum_{j\in B_i(g)}l_{ij}$. Similar to the allocation presented in Corollary 2, we can construct an allocation $\varphi^{FB}(g)$ that induces firms to choose their first-best inventory levels. Furthermore, under $\varphi^{FB}(g)$, there always exists a pairwise Nash stable network that is also efficient. Corollary 3 summarizes the results.

Corollary 3 *Let $\lambda_i\in(0,1)$ such that $\sum_{i\in N}\lambda_i=1$. When a network g is formed endogenously, there exists an efficient network g^* that is pairwise Nash stable under the allocation defined by $\varphi_i^{FB}(g)=\lambda_i\Pi(g,X,D)+\sum_{j\in B_i(g)}l_{ij}-[r_i\min\{X_i,D_i\}+u_iH_i-c_iX_i]$ for $i\in N$. Furthermore, $\varphi_i^{FB}(g^*)$ induces the inventory levels in a first-best solution to be a Nash equilibrium profile X^{NE}.*

Corollary 3 shows that, unlike the MJW value, there is no tension between stability and efficiency under $\varphi_i^{FB}(g)$. However, $\varphi_i^{FB}(g)$ may violate some desired properties of the MJW value such as symmetry when λ_i is not chosen properly.

7 Conclusion

This paper studies a cooperative game of inventory transshipment among multiple firms. As firms try to maximize their own profits, the value generated through transshipment needs to be allocated properly to coordinate the incentives of firms to participate in transshipment. To analyze this problem, the extant literature uses the concept of coalitions in cooperative game theory, while we use networks of firms as the primitive. Our network-based approach explicitly models the actual paths over which inventories are transshipped, and provides a richer form of representing relationships among firms than the previous coalition-based approach. This enables us to analyze partial cooperation structures based on networks in which partnership may exist between some but not among all firms. Our results provide the following managerial insights.

First, if the primary objective of a decentralized transshipment network is to make all firms participate in transshipment, then the dual price allocation can achieve this goal by providing firms with proper incentives. This might be the case for a non-profit association of local blood banks, since the transshipment of unused units of blood can potentially save the lives of people who urgently need those units. The dual price allocation we construct in the network setting takes into account the positions of firms in their networks as well as the amounts of their residual supplies or demands. By doing so, this allocation prevents a subset of firms from forming their subnetwork, and thus maximizes the value of pooling all residuals.

Second, compared to the dual price allocation, the MJW value provides a more intuitive way of allocating profits from transshipment to individual firms by considering their marginal contributions to a network. However, one should be cautious in implementing this allocation rule because it may induce some firms to form their own subnetwork, hurting the total value that can be generated from transshipment of all residuals. For example, in a regional blood management system, there is often a designated blood center through which other blood banks transship their unused units of blood. If the MJW value is implemented in this hub-spoke network, then the regional blood center as the hub needs to have sufficient residual supplies to satisfy all residual demands of blood banks in the spokes in order for this transshipment system to be stable.

Third, when the profits from transshipment are allocated according to the MJW value, firms' incentives to participate in transshipment depend crucially on how they are connected in a network. For example, one might temp to think that the complete network, which allows direct transshipment between blood banks without going through a regional blood center, may facilitate transshipment further because direct shipments can potentially save the costs of transportation and coordination. However, our result shows that such a transshipment system can be less stable than a hub-spoke network having a regional blood center as the hub. Thus, firms should consider building an incomplete network for their transshipment network because it is not only cheaper to institute than the complete network, but can also lead to a more stable cooperative transshipment system.

Fourth, an efficient network in a centralized system may not be stable in a decentralized system of independent firms. For example, a chain of retail stores or warehouses is known to be efficient in a centralized system by allowing all members to share their residual supplies and demands with a small number of links (e.g., Gerchak and Kalikhman 2011, Lien et al. 2011). However, our result shows that a chain of independent firms can incentivize firms to participate in transshipment only under some restrictive conditions. Thus, managers should be careful in applying the insights obtained for a centralized system to a decentralized system.

Finally, when firms are able to establish connections with each other, they tend to build connections with a large number of other firms even though they may need only few partners for actual transshipments. Such over-connection in a network exists because firms try to increase their bargaining powers against other firms, leading to an inefficient transshipment system. For example, when car dealers search for their partners to share their residual demand or supply, they may attempt to negotiate with many other dealers so as to increase their bargaining positions. This may happen even when the cost of search and negotiation is not negligible. To improve the efficiency of a transshipment network, a third-party organization such as a local automobile dealers association or an auto manufacturer may intervene a decentralized transshipment system to coordinate the incentives of independent firms.

There are several future research avenues. In our model of endogenous network formation, firms announce their partners simultaneously. One may consider sequential announcements. Modeling repeated interactions and negotiations between firms is an interesting dimension to investigate. In addition to the concept of pairwise Nash stability we used in this paper, there also exist other stability concepts such as stochastic or farsighted stability. Such concepts can be applied to analyze the inventory transshipment problem in a richer setting. Besides the inventory transshipment networks studied in this paper, one can potentially apply the theory of economic and social networks to analyze the stability and formation of networks in various other operational problems. For example, in a supply chain with disruption risks, over-connection in a network might be beneficial for firms in terms of risk-hedging. This paper may serve as the first step to many such analyses.

References

Anupindi, R., Y. Bassok, E. Zemel. 2001. A general framework for the study of decentralized distribution systems. *Manufacturing and Service Operations Management* **3**(4) 349-368.

Chen, R., S. Yin. 2010. The equivalence of uniform and Shapley-valued based allocations in a specific game. Forthcoming at *Operations Research Letters*.

Dong, L., N. Rudi. 2004. Who benefits from transshipment? exogenous vs. endogenous wholesale prices. *Management Science* **50**(5) 645–657.

Duvall, M. 2000. Cost-saving idea: Virtual warehouses. *Inter@ctive Week*, **7**(42) 18-18.

Fontaine, M. J., Y.T. Chung, W.M. Rogers, H.D. Sussmann, P. Quach, S.A. Galel, L.T. Goodnough, F. Erhun. 2009. Improving platelet supply chains through collaborations between blood centers and transfusion services. *Transfusion* **49**(10) 2040–2047.

Gerchak, Y., D. Kalikhman. 2011. Inventory sharing: chaining vs. complete pooling. Working paper, Tel-Aviv University.

Granot, D., G. Sošić. 2003. A three-stage model for a decentralized distribution system of retailers. *Operations Research* **51**(5) 771-784.

Granot, D., S. Yin. 2008. Competition and cooperation in decentralized push and pull assembly systems. *Management Science* **54**(4) 733-747.

Hall, N., Z. Liu. 2010. Capacity allocation and scheduling in supply chains. *Operations Research* **58**(6) 1711-1725.

Hezarkhani, B., W. Kubiak. 2010. Transshipment prices and pair-wise stability in coordinating the decentralized transshipment problem. Proceedings of Behavioral and Quantitative Game Theory Conference on Future Directions.

Hu, X., I. Duenyas, R. Kapuscinski. 2007. Existence of coordinating transshipment prices in a two-location inventory model. *Management Science* **53**(8) 1289–1302.

Huang, X., G. Sošić. 2010a. Repeated newsvendor game with transshipments under dual allocations. *European Journal of Operational Research* **204** 274-284.

Huang, X., G. Sošić. 2010b. Transshipment of inventories: Dual allocations vs. transshipment prices. *Manufacturing and Service Operations Management* **12**(2) 299–318.

Jackson, M.O. 2006. *Social and economic networks.* Princeton University Press, Princeton, NJ.

Jackson, M.O., J. Wolinsky. 1996. A strategic model of social and economic networks. *Journal of Economic Theory* **71**(1) 44-74.

Kemahlıoğlu-Ziya, E., J.J. Bartholdi. 2011. Centralizing inventory in supply chains by using Shapley value to allocate the profits. *Manufacturing and Service Operations Management* **13**(2) 146-162.

Lien, R.W., S.M.R. Iravani, K. Smilowitz, M. Tzur. 2011. An efficient and robust design for transshipment networks. *Production and Operations Management* **20**(5) 699–713.

Myerson, R.B. 1977. Graphs and cooperation in games. *Mathematics of Operations Research* **2**(3) 225-229.

Nagarajan, M., G. Sošić. 2007. Stable farsighted coalitions in competitive markets. *Management Science* **53**(1) 29-45.

Nagarajan, M., G. Sošić. 2008. Game-theoretic analysis of cooperation among supply chain agents: Review and extensions. *European Journal of Operational Research* **187**(3) 719-745.

Nagarajan, M., G. Sošić. 2009. Coalition stability in assembly models. *Operations Research* **57**(1) 131-145.

Nagarajan, M., G. Sošić, H. Zhang. 2010. Stable group purchasing organization. Working Paper, University of British Columbia.

Narus, J. A., J. C. Anderson. 1996. Rethinking distribution: adaptive channels. *Harvard Business Review* 112-120.

Netessine, S. 2009. Supply webs: managing, organizing, and capitalizing on global networks of suppliers. Chapter 13 in P. Kleindorfer, Y. J. Wind, eds. *The Network Challenge: Strategy, Profit, and Risk in an Interlinked World*, Wharton School Publishing, PA.

Owen, G. 1975. On the core of linear production games. *Math. Programming* **9** 358–370.

Özen, U., J. Fransoo, H. Norde, M. Slikker. 2008. Cooperation between multiple newsvendors with warehouses. *Manufacturing and Service Operations Management* **10**(2) 311-324.

Paterson, C., G. Kiesmuler, R. Teunter, K. Glazebrook. 2011. Inventory models with lateral transshipments: A review. *European Journal of Operational Research* **210** 125-136.

Prastacos, G.P. 1984. Blood inventory management: An overview of theory and practice. *Management Science* **30**(7) 777-800.

Ramsey, M., V. Bauerlein. 2013. Tesla clashes with car dealers. *The Wall Street Journal.* June 18, 2013.

Reyniers, D.J., C.S. Tapiero. 1995. The delivery and control of quality in supplier-producer contracts. *Management Science* **41**(10) 1581-1589.

Rong, Y., L.V. Snyder, Y. Sun. 2010. Inventory sharing under decentralized preventive transshipments. Working Paper, Lehigh University.

Rudi, N., S. Kapur, D.F. Pyke. 2001. A two-location inventory model with transshipment and local decision making. *Management Science* **47**(12) 1668–1680.

Shapley, L., M. Shubik. 1975. Competitive outcomes in the cores of market games. The Rand Corporation, Santa Monica, CA.

Shapley, L. 1953. A value for n-person games. *Annals of Mathematical Studies* **28** 307–317.

Shao, J., H. Krishnan, S.T. McCormick. 2011. Incentives for transshipment in a supply chain with decentralized retailers. *Manufacturing and Service Operations Management* **13**(3) 361-372.

Slikker, M., J. Fransoo, M. Wouters. 2005. Cooperation between multiple news-vendors with transshipments. *European Journal of Operational Research* **167**(2) 370–380.

Sošić, G. 2006. Transshipment of inventories among retailers: Myopic vs. farsighted stability. *Management Science* **52**(10) 1493-1508.

Yin, S. 2010. Alliance formation among perfectly complementary suppliers in a price-sensitive assembly system. *Manufacturing and Service Operations Management* **12**(3) 527-544.

Zhang, J. 2005. Transshipment and its impact on supply chain members' performance. *Management Science* **51**(10) 1534–1539.

Zhao, H., V. Deshpande, J.K. Ryan. 2005. Inventory sharing and rationing in decentralized dealer networks. *Management Science* **51**(4) 531–547.

Zhao, X., D. Atkins. 2009. Transshipment between competing retailers. *IIE Transactions* **41**(8) 665-676.

Appendix

Proof of Proposition 1: For a given network g, the value function v given in (6)-(9) is obtained by solving a linear program. So, the cooperative transshipment game given a network g can be viewed as the linear production game considered by Owen (1975). Moreover, this game can be modeled as the market game of Shapley and Shubik (1975) in which players trade products with each other. Then the result follows from Shapley and Shubik (1975) who have shown that the dual price allocation is in the core of a market game. Q.E.D.

Remark: Essentially, this proof is the same as the proof of Theorem 4.1 in Anupindi et al. (2001) because the linearity in (6)-(9) is preserved in our network setting.

Proof of Proposition 2: Consider the case when only one supply node exists in the network. The proof for the case when only one demand node exists follows the same procedure. To prove that the pair $(g^N, \varphi^{MJW}(g^N))$ is in the core of (N, v), it suffices to show that (N, v) is convex (Jackson 2006). We prove that (N, v) is convex by showing that $v(g|_{S''\cup\{i\}}) - v(g|_{S''}) \leq v(g|_{S'\cup\{i\}}) - v(g|_{S'})$ for any subsets S' and S'' of N that satisfy $S'' \subset S'$ and any node $i \notin S'$ in each of the following two cases: (Case 1) node i is the supply node, and (Case 2) node i is not the supply node.

(Case 1): If i is the supply node, then neither S' nor S'' contains any supply nodes, so $v(g^N|_{S''}) = v(g^N|_{S'}) = 0$. Thus, if $v(g^N|_{S'\cup\{i\}}) \geq v(g^N|_{S''\cup\{i\}})$, then $v(g^N|_{S'\cup\{i\}}) - v(g^N|_{S'}) \geq v(g^N|_{S''\cup\{i\}}) - v(g^N|_{S''})$. To prove that $v(g^N|_{S'\cup\{i\}}) \geq v(g^N|_{S''\cup\{i\}})$, we define $f(g|_S, Y) \equiv \sum\limits_{i\in S}\{a_i(\sum\limits_{j\in B_i(g|_S)} Y_{ji} - \sum\limits_{j\in B_i(g|_S)} Y_{ij}) - \sum\limits_{j\in B_i(g|_S)} Y_{ij}(t_{ij} + u_i - u_j)\}$, which is the objective function of the program given in (6). Then, $v(g^N|_{S'\cup\{i\}}) = f(g^N|_{S'\cup\{i\}}, Y^{g^N|_{S'\cup\{i\}}}) \geq f(g^N|_{S'\cup\{i\}}, Y^{g^N|_{S''\cup\{i\}}}) = f(g^N|_{S''\cup\{i\}}, Y^{g^N|_{S''\cup\{i\}}}) = v(g^N|_{S''\cup\{i\}})$ because:

- By the definition of optimal transshipment patterns, $v(g^N|_{S'\cup\{i\}}) = f(g^N|_{S'\cup\{i\}}, Y^{g^N|_{S'\cup\{i\}}})$ and $v(g^N|_{S''\cup\{i\}}) = f(g^N|_{S''\cup\{i\}}, Y^{g^N|_{S''\cup\{i\}}})$;
- Since $S'' \subset S'$, the transshipment pattern $Y^{g^N|_{S''\cup\{i\}}}$ is feasible to the program (6)-(9) with $g = g^N$ and $S = S' \cup \{i\}$, so that $f(g^N|_{S'\cup\{i\}}, Y^{g^N|_{S''\cup\{i\}}}) = f(g^N|_{S''\cup\{i\}}, Y^{g^N|_{S''\cup\{i\}}})$;
- Because $Y^{g^N|_{S'\cup\{i\}}}$ is the optimal transshipment pattern among all the feasible transshipment patterns given the network $g^N|_{S'\cup\{i\}}$, $f(g^N|_{S'\cup\{i\}}, Y^{g^N|_{S'\cup\{i\}}}) \geq f(g^N|_{S'\cup\{i\}}, Y^{g^N|_{S''\cup\{i\}}})$.

(Case 2): Suppose j $(\neq i)$ is the supply node. There are three possibilities:

(a) If $j \notin S'$ (hence $j \notin S''$), then $v(g^N|_{S''\cup\{i\}}) = v(g^N|_{S''}) = v(g^N|_{S'\cup\{i\}}) = v(g^N|_{S'}) = 0$ because $S' \cup \{i\}$, S', $S'' \cup \{i\}$, and S'' have no supply nodes.

(b) If $j \in S'$ and $j \notin S''$, then $v(g^N|_{S'\cup\{i\}}) - v(g^N|_{S'}) \geq 0 = v(g^N|_{S''\cup\{i\}}) - v(g^N|_{S''})$, where the inequality is due to the fact that $S' \cup \{i\}$ and S' include the supply node j and $\sum\limits_{k\in S'\cup\{i\}} E_k \geq \sum\limits_{k\in S'} E_k$, and the equality is because $S'' \cup \{i\}$ and S'' have no supply nodes.

(c) If $j \in S''$ (hence $j \in S'$), $v(g^N|_{S' \cup \{i\}}) - v(g^N|_{S'}) = v(g^N|_{S'' \cup \{i\}}) - v(g^N|_{S''}) = (r_i - u_j - t_{ji}^{\min})E_i$, where $t_{ji}^{\min}$ is the minimum of the transportation costs from j to i among all possible paths. The equations hold because $S' \cup \{i\}$, S', $S'' \cup \{i\}$, and S'' include the supply node j with $H_j \geq \sum_{k \in N \setminus \{j\}} E_k$.
Q.E.D.

Remark: Suppose direct shipment incurs a minimum transportation cost among all paths between every pair of two firms and transshipment is always profitable (i.e., $r_\alpha - u_\beta - t_{\beta\alpha} > 0$ for all $\alpha, \beta \in N$). Then, the necessary and sufficient condition for the convexity to hold is that there is only one supply (resp., demand) node j with $H_j \geq \sum_{k \in N \setminus \{j\}} E_k$ (resp., $E_j \geq \sum_{k \in N \setminus \{j\}} H_k$). Sufficiency is shown in the proof of Proposition 2. In the following, we show the necessity in two steps.

First, we show that the convexity does not hold when multiple demand nodes and multiple supply nodes exist in the complete network. We prove this by contradiction. Consider the complete network g^N with multiple demand nodes and multiple supply nodes. Without loss of generality, suppose E_1, E_2, H_3, and H_4 are all positive. If the convexity holds, then $v(g^N|_{\{1,2,3\} \cup \{4\}}) - v(g^N|_{\{1,2,3\}}) \geq v(g^N|_{\{1,2\} \cup \{4\}}) - v(g^N|_{\{1,2\}})$. This indicates that $E_1 + E_2 \geq H_3 + H_4$ because otherwise there are more residual demands available within $\{1, 2\}$ than within $\{1, 2, 3\}$ (i.e., $E_1 + E_2 > E_1 + E_2 - H_3$) and $v(g^N|_{\{1,2,3\} \cup \{4\}}) - v(g^N|_{\{1,2,3\}}) < v(g^N|_{\{1,2\} \cup \{4\}}) - v(g^N|_{\{1,2\}})$. Similarly, since $v(g^N|_{\{2,3,4\} \cup \{1\}}) - v(g^N|_{\{2,3,4\}}) \geq v(g^N|_{\{3,4\} \cup \{1\}}) - v(g^N|_{\{3,4\}})$, $E_1 + E_2 \leq H_3 + H_4$. To satisfy these two inequalities simultaneously, $E_1 + E_2 = H_3 + H_4$. Further, since $v(g^N|_{\{2,3\} \cup \{4\}}) - v(g^N|_{\{2,3\}}) \geq v(g^N|_{\{2\} \cup \{4\}}) - v(g^N|_{\{2\}})$, $E_2 \geq H_3 + H_4$. Because $E_2 = H_3 + H_4 - E_1 \geq H_3 + H_4$ and $E_1 \geq 0$, $E_1 = 0$, which contradicts our premise that $E_1 > 0$.

Second, we show that, given there exists only one supply node j, the convexity does not hold when $H_j < \sum_{k \in N \setminus \{j\}} E_k$. (The case when only one demand node exists follows the same procedure, and hence omitted.) We prove this by finding S' and S'' ($\subset S'$) that do not satisfy $v(g|_{S'' \cup \{i\}}) - v(g|_{S''}) \leq v(g|_{S' \cup \{i\}}) - v(g|_{S'})$. Since $H_j < \sum_{k \in N \setminus \{j\}} E_k$, there exist S' and S'' such that $H_j - \sum_{k \in S'} E_k < E_i$ and $H_j - \sum_{k \in S''} E_k > 0$ (S'' can be the empty set) for $i \notin S'$. Then, $v(g^N|_{S' \cup \{i\}}) - v(g^N|_{S'}) = (r_i - u_j - t_{ji}) \max(H_j - \sum_{k \in S'} E_k, 0)$ and $v(g^N|_{S'' \cup \{i\}}) - v(g^N|_{S''}) = (r_i - u_j - t_{ji}) \min(H_j - \sum_{k \in S''} E_k, E_i)$ because direct shipment incurs a minimum transportation cost and transshipment is always profitable. Therefore, $v(g^N|_{S' \cup \{i\}}) - v(g^N|_{S'}) < v(g^N|_{S'' \cup \{i\}}) - v(g^N|_{S''})$ because $H_j - \sum_{k \in S'} E_k < H_j - \sum_{k \in S''} E_k$, $H_j - \sum_{k \in S'} E_k < E_i$, $H_j - \sum_{k \in S''} E_k > 0$ and $E_i > 0$.

Proof of Proposition 3: Consider the case when the hub is a supply node. (The proof for the case when the hub is a demand node follows the same procedure.) Similar to the proof of Proposition 2, we prove that the pair $(g^H, \varphi^{MJW}(g^H))$ is in the core by showing that $v(g^H|_{S'' \cup \{i\}}) - v(g^H|_{S''}) \leq v(g^H|_{S' \cup \{i\}}) - v(g^H|_{S'})$ for any subsets S' and S'' of N that satisfy $S'' \subset S'$, and for any node $i \notin S'$.

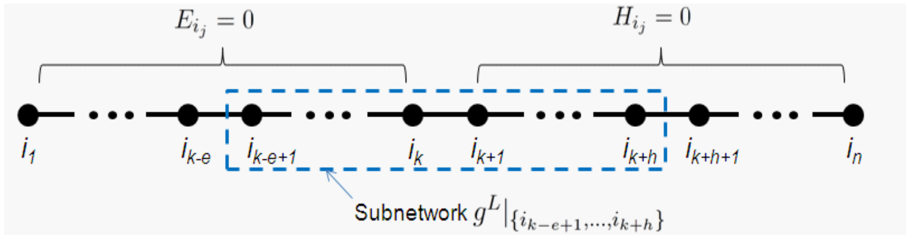

Figure 7: Line g^L and its subnetwork $g^L|_{\{i_{k-e+1},\ldots,i_{k+h}\}}$.

We consider the following three cases: (Case 1) node i is the hub, (Case 2) node i is a supply node but not the hub, and (Case 3) node i is a demand node.

(Case 1): The proof is similar to that in (Case 1) of Proposition 2 because, given $i \notin S'$, we have $B_k(g^H|_{S'}) = \emptyset \ \forall k \in S'$ and $B_k(g^H|_{S''}) = \emptyset \ \forall k \in S''$ so that $v(g^N|_{S''}) = v(g^N|_{S'}) = 0$.

(Case 2): Suppose $j \ (\neq i)$ is the hub. There are three possibilities:

(a) If $j \notin S'$ (hence $j \notin S''$), then $v(g^H|_{S'' \cup \{i\}}) = v(g^H|_{S''}) = v(g^H|_{S' \cup \{i\}}) = v(g^H|_{S'}) = 0$ because $B_k(g^H|_{S' \cup \{i\}}) = \emptyset \ \forall k \in S' \cup \{i\}$ and $B_k(g^H|_{S'' \cup \{i\}}) = \emptyset \ \forall k \in S'' \cup \{i\}$ without the hub.

(b) If $j \in S'$ and $j \notin S''$, then $v(g^H|_{S' \cup \{i\}}) - v(g^H|_{S'}) \geq 0 = v(g^H|_{S'' \cup \{i\}}) = v(g^H|_{S''})$ where the inequality is due to the fact that $S' \cup \{i\}$ and S' include the hub, and the equality is because $S'' \cup \{i\}$ and S'' do not include the hub.

(c) If $j \in S''$ (hence $j \in S'$), $v(g^N|_{S' \cup \{i\}}) - v(g^N|_{S'}) = v(g^N|_{S'' \cup \{i\}}) - v(g^N|_{S''}) = 0$ because H_i generates zero value given $H_j \geq \sum_{k \in N \setminus \{j\}} E_k$.

(Case 3): The proof is similar to the proof of (Case 2) of Proposition 2, in which the hub acts as the supply node j which ships inventory to node i. Q.E.D.

Proof of Proposition 4: For sufficiency, we show that the pair $(g^L, \varphi^{MJW}(g^L))$ is always in the core, so any subnetwork of g^L has no incentive to secede from g^L under the MJW value. Clearly, any subnetwork with no supply nodes or no demand nodes has no incentive to secede because they generate no profit. Thus, in the rest of the proof, we consider subnetworks that contain at least one supply node and one demand node. Without loss of generality, consider the subnetwork $g^L|_{\{i_{k-e+1},\ldots,i_{k+h}\}}$ of the line g^L shown in Figure 7. Note that g^L contains k firms with $E_{i_j} = 0$ for $j = 1, 2, \ldots, k$, and $(n-k)$ firms with $H_{i_j} = 0$ for $j = k+1, k+2, \ldots, n$, while $g^L|_{\{i_{k-e+1},\ldots,i_{k+h}\}}$ contains $e \ (\leq k)$ firms with $E_{i_j} = 0$ for $j = k-e+1, \ldots, k$, and $h \ (\leq n-k)$ firms with $H_{i_j} = 0$ for $j = k+1, \ldots, k+h$, where $e = 1, 2, \ldots, k$, and $h = 1, 2, \ldots, n-k$.

To prove that $g^L|_{\{i_{k-e+1},\ldots,i_{k+h}\}}$ has no incentive to secede from g^L, we show that the allocations to the firms in $g^L|_{\{i_{k-e+1},\ldots,i_{k+h}\}}$ are non-decreasing when an outside firm is added to this subnetwork. Consider adding firm i_{k-e} with no residual demand to this subnetwork. (The proof for adding a firm with no residual supply is similar, and hence omitted.) If $H_{i_{k-e}} = 0$, the allocations are unchanged, so we focus on the case when $H_{i_{k-e}} > 0$. First, consider the allocations to firm i_j for $j = k - e + 1, ..., k$. We shall show that, when i_{k-e} is added, the marginal contributions of firm i_j are non-decreasing in all the orderings of the firms in $g^L|_{\{i_{k-e}, i_{k-e+1}, ..., i_{k+h}\}}$. Without loss of generality, consider two orderings before and after firm i_{k-e} is added: (1) ordering $\{i_1^{'}, ..., i_{e+h}^{'}\}$, and (2) ordering $\{i_1^{''}, ..., i_\alpha^{''}, ..., i_{e+h+1}^{''}\}$ with $i_\alpha^{''} = i_{k-e}$, $i_j^{''} = i_j^{'}$ for $j < \alpha$, and $i_j^{''} = i_{j-1}^{'}$ for $j > \alpha$; i.e., α ($\leq e + h + 1$) is the position of firm i_{k-e} in the second ordering and the two orderings have the same sequence of firms except i_{k-e}. Suppose the positions of firm i_j, whose allocations are under consideration, in the two orderings are β_1 and β_2, respectively. We can define the sets of firms before firm i_j in the two orderings, respectively, $S^{'} = \{i_1^{'}, ..., i_{\beta_1 - 1}^{'}\}$ and $S^{''} = \{i_1^{''}, ..., i_{\beta_2 - 1}^{''}\}$. If $i_{k-e} \notin S^{''}$, then we have $S^{'} = S^{''}$, so $v(g^L|_{S' \cup \{i_j\}}) - v(g^L|_{S'}) = v(g^L|_{S'' \cup \{i_j\}}) - v(g^L|_{S''})$. If $i_{k-e} \in S^{''}$, it can be shown that $v(g^L|_{S'}) = v(g^L|_{S''})$ because firm i_{k-e} is not connected to any demand node without i_j. Furthermore, we have $v(g^L|_{S'' \cup \{i_j\}}) \geq v(g^L|_{S' \cup \{i_j\}})$ because more residual supplies are available within the subnetwork after firm i_{k-e} is added. Therefore, we obtain $v(g^L|_{S'' \cup \{i_j\}}) - v(g^L|_{S''}) \geq v(g^L|_{S' \cup \{i_j\}}) - v(g^L|_{S'})$. Second, it can be shown in a similar way that, when i_{k-e} is added, the allocations to i_j for $j = k + 1, ..., k + h$ are non-decreasing as well. As a result, no subnetwork in g^L has an incentive to secede.

For necessity, we show that if the pair $(g, \varphi^{MJW}(g))$ is in the core for any residual amounts (and thus for any numbers of supply and demand nodes), network g must be the line g^L. We conduct the proof by showing the following two properties must hold for the network g such that the pair $(g, \varphi^{MJW}(g))$ is in the core for any residual amounts: (Property 1) every demand (resp., supply) node has at most one link to supply (resp., demand) nodes, and (Property 2) every firm has at most two links. Then, the only network structure that satisfies these two properties for any residual amounts and any numbers of supply and demand nodes is the line.

First, we show that if g does not satisfy Property 1, then $(g, \varphi^{MJW}(g))$ is not always in the core for any residual amounts. For this proof, it suffices to find a network g having a demand node with more than one link to supply nodes, and show that a subnetwork of g has an incentive to secede from g for certain quantities of residual supplies and demands. Consider g that contains firm 1 with $E_1 = 1$ and $B_1(g) > 1$, and $(n-1)$ firms with residual supplies. Let firm 2 denote a firm with the minimum value of $u_i + t_{i1}$ among all $i \in B_1(g)$, and assume $H_2 = 1$. The profit generated by transshipment is $v(g) = r_1 - u_2 - t_{21}$. If there exists any other firm $i \in B_1(g)$ such that $H_i > 0$ and $r_1 - u_i - t_{i1} > 0$, it receives a strictly positive allocation based on the MJW value according to (10) because firm i makes a positive marginal contribution to g. Then the sum of allocations to firms 1 and 2 is strictly less than $r_1 - u_2 - t_{21}$ because the presence of another supply node i reduces the

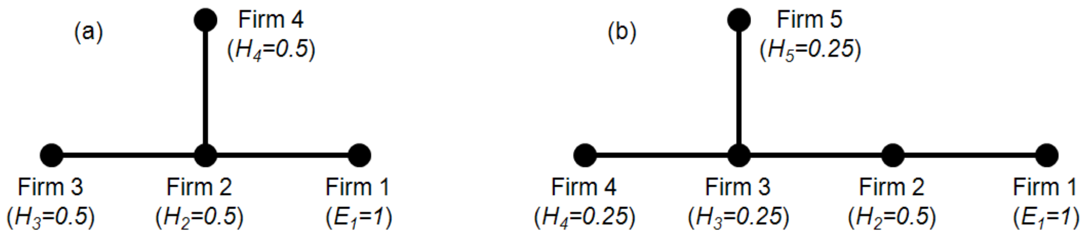

Figure 8: Subnetworks that have incentives to secede from the network g.

marginal contribution of firm 2. On the other hand, if firms 1 and 2 secede, they can generate a profit of $r_1 - u_2 - t_{21}$ which is larger than the allocations they receive in the initial network g.

Second, similarly to Property 1, we show that if g does not satisfy Property 2, then $(g, \varphi^{MJW}(g))$ is not always in the core for any residual amounts. As in the proof for Property 1, consider firm 1 with $E_1 = 1$ and $(n-1)$ firms with residual supplies. Given firm 1 is only linked to firm 2 to prevent any subnetwork from seceding from g, we show that there exists at most one firm i with $H_i > 0$ linked to firm 2; otherwise, there exists a subnetwork which has an incentive to secede. As shown in Figure 8(a), denote the firm $i \in B_2(g)$ in the network g with the minimum value of $u_i + t_{i2}$ as firm 3. Set $H_2 = 0.5$ and $H_3 = 0.5$. Then the profit generated by transshipment is $v(g) = r_1 - 0.5u_2 - t_{21} - 0.5u_3 - 0.5t_{32}$. If there exists any other firm $i \in B_2(g)$ such that $H_i > 0$ and $r_1 - u_i - t_{i2} - t_{21} > 0$, it receives strictly positive allocation based on the MJW value, e.g. firm 4. On the other hand, if firms 1, 2 and 3 secede, they can still generate a profit of $r_1 - 0.5u_2 - t_{21} - 0.5u_3 - 0.5t_{32}$ which is greater than the allocations they receive in the initial network g. Next, we consider the subnetwork in Figure 8(b) to show that, other than firm 2, there exists at most one firm i with $H_i > 0$ which is linked to firm 3. We denote firm 4 as the firm with the minimum value of $u_i + t_{i3}$ for $i \in B_3(g)$ in the network g. Set $H_2 = 0.5$, $H_3 = 0.25$ and $H_4 = 0.25$. By the same argument as above, we know firms 1, 2, 3 and 4 have an incentive to secede if firm 5 is also linked to firm 3. Repeating the same procedure for firms $4, ..., n$, we can show that network g must be a line if the pair $(g, \varphi^{MJW}(g))$ is always in the core. Q.E.D.

Remark: The chain does not satisfy Property 1 while it satisfies Property 2, because, in the case with only one demand (resp., supply) node, the demand (resp., supply) node is linked to two supply (resp., demand) nodes in the chain.

Proof of Corollary 1: First, since $\varphi_i^{MJW}(g^C) \leq v(g^C) - v(g^C|_{N\setminus\{i\}})$ $\forall i \in N$, we know that any subnetworks with $n-1$ firms have no incentive to secede from g^C. Next, similar to the proof of the sufficiency of Proposition 4, we consider a subnetwork of g^C with less than $n-1$ firms and show that the allocations to the firms in this subnetwork are non-decreasing when an outside firm is added.

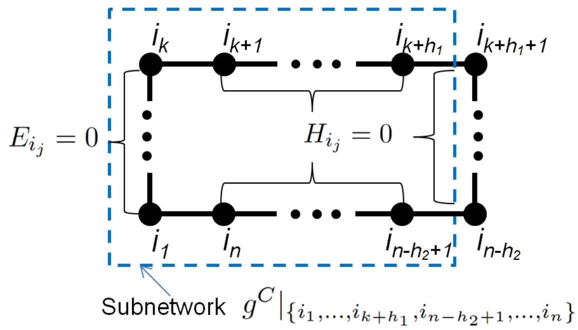

Figure 9: Chain g^C and its subnetwork $g^C|_{\{i_1,\dots,i_{k+h_1},i_{n-h_2+1},\dots,i_n\}}$.

For the subnetworks in which all the demand nodes are on one side and all the supply nodes are on the other side, the proof is the same as that for the line. Without loss of generality, we consider the subnetwork $g^C|_{\{i_1,\dots,i_{k+h_1},i_{n-h_2+1},\dots,i_n\}}$ of the chain g^C shown in Figure 9. Note that g^C contains k firms with $E_{i_j} = 0$ for $j = 1, 2, \dots, k$, and $(n-k)$ firms with $H_{i_j} = 0$ for $j = k+1, k+2, \dots, n$, while $g^C|_{\{i_1,\dots,i_{k+h_1},i_{n-h_2+1},\dots,i_n\}}$ contains k firms with $E_{i_j} = 0$ for $j = 1, 2, \dots, k$, and $h_1 + h_2$ $(\leq n-k-2)$ firms with $H_{i_j} = 0$ for $j = k+1, \dots, k+h_1$ and $j = n-h_2+1, \dots, n$ where $h_1 + h_2 = 2, 3, \dots, n-k-2$.

Consider adding firm i_{n-h_2} to this subnetwork. If $E_{i_{n-h_2}} = 0$, the allocations are unchanged, so we focus on the case when $E_{i_{n-h_2}} > 0$. Adding residual demands does not reduce the marginal values of residual supplies and i_{n-h_2} is not connected to any supply node without i_j for $j = n-h_2+1, \dots, n$. Thus, it can be shown in the same way as the proof of Proposition 4 that the allocations to firm i_j for $j = 1, \dots, k$ and $j = n-h_2+1, \dots, n$ are non-decreasing when i_{n-h_2} is added. In the rest of this proof, we show that the marginal contributions of firm i_j for $j = k+1, \dots, k+h_1$ are also non-decreasing in all the orderings of the firms in $g^C|_{\{i_1,\dots,i_{k+h_1},i_{n-h_2+1},\dots,i_n\}}$. Without loss of generality, consider two orderings before and after firm i_{n-h_2} is added with the sequence of other firms being the same: (1) ordering $\{i'_1, \dots, i'_{k+h_1+h_2}\}$, and (2) ordering $\{i''_1, \dots, i''_\alpha, \dots, i''_{k+h_1+h_2+1}\}$ with $i''_\alpha = i_{n-h_2}$, $i''_j = i'_j$ for $j < \alpha$, and $i''_j = i'_{j-1}$ for $j > \alpha$. Suppose the positions of firm i_j, whose allocations are under consideration, in the two orderings are β_1 and β_2, respectively. We can define the sets of firms before firm i_j in the two orderings, respectively, $S' = \{i'_1, \dots, i'_{\beta_1-1}\}$ and $S'' = \{i''_1, \dots, i''_{\beta_2-1}\}$. If $i_{n-h_2} \notin S''$, then we have $S' = S''$, so $v(g^C|_{S'\cup\{i_j\}}) - v(g^C|_{S'}) = v(g^C|_{S''\cup\{i_j\}}) - v(g^C|_{S''})$. If $i_{n-h_2} \in S''$ and there exists $i_m \notin S''$ for $m = 1, \dots, k$, then $v(g^C|_{S'\cup\{i_j\}}) - v(g^C|_{S'}) = v(g^C|_{A_{i_j}\cup\{i_j\}}) - v(g^C|_{A_{i_j}}) = v(g^C|_{S''\cup\{i_j\}}) - v(g^C|_{S''})$ where A_{i_j} is the set of firms that are directly or indirectly connected to i_j in $g^C|_{S'\cup\{i_j\}}$. The first equality holds because the marginal value of i_j depends only on the firms it is connected to. The second equality holds because i_j and i_{n-h_2} are not connected in $g^C|_{S''\cup\{i_j\}}$. Finally, if $i_{n-h_2} \in S''$ and $i_m \in S''$ $\forall$ $m = 1, \dots, k$, then $v(g^C|_{S''\cup\{i_j\}}) - v(g^C|_{S''}) =$

$v(g^C|_{S'\cup\{i_j\}}) - v(g^C|_{S'})$ because $\sum_{i\in N} H_i = \sum_{i\in N} E_i$. Therefore, no subnetwork of g^C has an incentive to secede from g^C. Q.E.D.

Proof of Proposition 5: Similar to the proof of Proposition 2, we can show that $v(g+ij|_S) - v(g|_S) \geq 0$ for all $S \subseteq N$. When $v(g+ij|_S)-v(g|_S) = 0$ for all $S \subseteq N$, $\varphi_k^{MJW}(g+ij)-\varphi_k^{MJW}(g) = 0$ for all $k \in N$ due to (10). Therefore, $\sum_{k\in S}\varphi_k^{MJW}(g+ij) = \sum_{k\in S}\varphi_k^{MJW}(g) \geq v(g|_S) = v(g+ij|_S)$, where the inequality holds because the pair $(g, \varphi^{MJW}(g))$ is in the core. When $v(g+ij|_S) - v(g|_S) > 0$, $\sum_{k\in S}\varphi_k^{MJW}(g+ij) - v(g+ij|_S) \geq \sum_{k\in S}\varphi_k^{MJW}(g) - v(g|_S) \geq 0$ because $v(g+ij|_S) - v(g|_S) \leq \sum_{k\in S}\{\varphi_k^{MJW}(g+ij) - \varphi_k^{MJW}(g)\}$. Therefore, the pair $(g+ij, \varphi^{MJW}(g+ij))$ is in the core. Q.E.D.

Proof of Proposition 6: First, we prove that the complete network is pairwise Nash stable under the MJW value by showing that any firms i and j have no incentives to sever link ij in the complete network. Consider ordering $\{i_1, i_2, ..., i_\alpha, ..., i_n\}$ used to calculate the MJW value with $i_\alpha = i$ and firm j can be anywhere in this ordering. Define set $S' = \{i_1, i_2, ..., i_{\alpha-1}\}$, i.e. the set of firms before firm i in this ordering. Then we have $v(g^N|_{S'}) - v(g^N - ij|_{S'}) = 0$ because S' does not contain firm i. Also, $v(g^N|_{S'\cup\{i\}}) - v(g^N - ij|_{S'\cup\{i\}}) \geq 0$ because $g^N|_{S'\cup\{i\}}$ has one more link than $g^N - ij|_{S'\cup\{i\}}$. With different orderings, S' can be any subset of $N\backslash\{i\}$. Therefore, from the definition of the MJW value given in (10), we obtain $\varphi_i^{MJW}(g^N) - \varphi_i^{MJW}(g^N - ij) = \sum_{S'\subseteq N\backslash\{i\}} \left\{v(g^N|_{S'\cup\{i\}}) - v(g^N - ij|_{S'\cup\{i\}})\right\} \left\{\frac{\#S'!(n-\#S'-1)!}{n!}\right\} \geq 0$. The same argument holds for firm j, so the complete network is pairwise Nash stable under the MJW value.

Second, we show that no firm would find it weakly profitable to sever link ij in the complete network if there exists $S \subseteq N$ such that the transshipment problem within S given in (6)-(9) has a unique optimal transshipment pattern $Y^{g^N|_S}$ with $Y_{ij}^{g^N|_S} > 0$. Consider the ordering $\{i_1, i_2, ..., i_\alpha, ..., i_n\}$ with $i_\alpha = i$ and the set $S' = \{i_1, ..., i_{\alpha-1}\}$. When $S' = S$, we have $v(g^N|_{S'\cup\{i\}}) - v(g^N - ij|_{S'\cup\{i\}}) > 0$ because $Y_{ij}^{g^N|_S} > 0$ implies that link ij is used for transshipment in g^N but cannot be used in $g^N - ij$. Thus, we obtain the following inequalities $\varphi_i^{MJW}(g^N) - \varphi_i^{MJW}(g^N - ij) \geq \left\{v(g^N|_{S\cup\{i\}}) - v(g^N - ij|_{S\cup\{i\}})\right\} \left\{\frac{\#S!(n-\#S-1)!}{n!}\right\} > 0$, in which the first inequality follows from the first part of the proof that shows $v(g^N|_{S'\cup\{i\}}) - v(g^N - ij|_{S'\cup\{i\}}) \geq 0$ for any $S' \subseteq N\backslash\{i\}$. The same argument holds for firm j.

Finally, we show that no firm would find it weakly profitable to sever such link ij in any network derived from the complete network. Consider the network $g^N - i'j'$ derived from the complete network g^N by removing link $i'j'$ such that $\varphi_{i'}^{MJW}(g^N) \leq \varphi_{i'}^{MJW}(g^N - i'j')$ and $\varphi_{j'}^{MJW}(g^N) \leq \varphi_{j'}^{MJW}(g^N - i'j')$ according to the definition of the networks derived from the complete network. From the second part of the proof, we know $Y^{g^N - i'j'|_S} = Y^{g^N|_S}$ for any $S \subseteq N$ because otherwise $\varphi_{i'}^{MJW}(g^N) > \varphi_{i'}^{MJW}(g^N - i'j')$ and $\varphi_{j'}^{MJW}(g^N) > \varphi_{j'}^{MJW}(g^N - i'j')$. Using the same argument as in the second part of the proof, we can prove that no firm would find it weakly profitable to sever link ij in the network $g^N - i'j'$ if the transshipment problem within S given in (6)-(9) has a unique optimal transshipment pattern $Y^{g^N|_S}$ with $Y_{ij}^{g^N|_S} > 0$ for some $S \subseteq N$. We can repeat this procedure (for example, consider the network derived from the complete network $g^N - i'j' - i''j''$ with $\varphi_{i''}^{MJW}(g^N - i'j') \leq \varphi_{i''}^{MJW}(g^N - i'j' - i''j'')$ and $\varphi_{j''}^{MJW}(g^N - i'j') \leq \varphi_{j''}^{MJW}(g^N - i'j' - i''j'')$), and show that any pairwise Nash stable network g derived from the complete network contains such ij. Q.E.D.

Proof of Proposition 7: First, we prove that, with small link cost, any pairwise Nash stable network derived from the complete network contains link ij such that the transshipment problem within S given in (6)-(9) has a unique optimal transshipment pattern $Y^{g^N|_S}$ with $Y_{ij}^{g^N|_S} > 0$. For such ij, we know from the proof of Proposition 6 that $\varphi_i^{MJW}(g^N) - \varphi_i^{MJW}(g^N - ij) > 0$. Let $\bar{l}$ (> 0) denote the minimum of $\varphi_i^{MJW}(g^N) - \varphi_i^{MJW}(g^N - ij) > 0$ among all such ij. Then, for any ij that satisfies the above condition and $l_{ij} < \bar{l}$, we have $\varphi_i^{MJW}(g^N) - \varphi_i^{MJW}(g^N - ij) - l_{ij} > 0$. Next, we show that no firm would find it weakly profitable to sever such link ij in any network derived from the complete network. Consider the network $g^N - i'j'$ derived from the complete network g^N by removing link $i'j'$ such that $\varphi_{i'}^{MJW}(g^N) - l_{i'j'} \leq \varphi_{i'}^{MJW}(g^N - i'j')$ and $\varphi_{j'}^{MJW}(g^N) - l_{j'i'} \leq \varphi_{j'}^{MJW}(g^N - i'j')$. By the same argument as in the proof of Proposition 6: $Y^{g^N - i'j'|_S} = Y^{g^N|_S}$ for any $S \subseteq N$, and we can obtain $\varphi_i^{MJW}(g^N - i'j') - \varphi_i^{MJW}(g^N - i'j' - ij) - l_{ij} > 0$ and $\varphi_j^{MJW}(g^N - i'j') - \varphi_j^{MJW}(g^N - i'j' - ij) - l_{ji} > 0$. We can repeat this procedure (for example, consider the network derived from the complete network $g^N - i'j' - i''j''$ with $\varphi_{i''}^{MJW}(g^N - i'j') - l_{i''j''} \leq \varphi_{i''}^{MJW}(g^N - i'j' - i''j'')$ and $\varphi_{j''}^{MJW}(g^N - i'j') - l_{j''i''} \leq \varphi_{j''}^{MJW}(g^N - i'j' - i''j'')$), and show that $\varphi_i^{MJW}(g) - \varphi_i^{MJW}(g - ij) - l_{ij} > 0$ for any network g derived from the complete network.

Second, we show the existence of the subnetwork g^* stated in the proposition. From the premise of the proposition, the centralized transshipment problem has a unique optimal transshipment pattern Y^{g^N} with $Y_{ij}^{g^N} = 0$ and $Y_{ji}^{g^N} = 0$. Suppose that there is no link cost. Since link ij is not used in this problem, there exists an efficient network g^* such that $ij \notin g^*$ and $v(g^*) = v(g^N)$. This network g^* is a subnetwork of the pairwise Nash stable networks derived from the complete network because any pairwise Nash stable network derived from the complete network contains the links that are used in the centralized transshipment problem. Because $l_{ij} > 0$ or $l_{ji} > 0$, any pairwise Nash stable networks under the MJW value that contain ij are dominated by g^*. Q.E.D.

Proof of Corollary 2: Under $\varphi^{FB}(g)$, the expected profit of firm i from both stages is $\lambda_i \cdot E_D[\Pi(g, X, D)]$. Since λ_i is a constant, the best-response functions of all the firms coincide with the first-order conditions in the centralized system. Q.E.D.

Proof of Corollary 3: Under $\varphi^{FB}(g)$, the ex-post profit of firm i generated from *both* stages is $\lambda_i \cdot \Pi(g, X, D) = \lambda_i\{\sum_{i \in N}(r_i \min\{X_i, D_i\} + u_i H_i - c_i X_i - \sum_{j \in B_i(g)} l_{ij}) + CTP(g, X, D)\}$. Since λ_i is a constant, for given inventory profile X and demand profile D, the objective of firm i is to maximize $CTP(g, X, D) + \sum_{i \in N}(-\sum_{j \in B_i(g)} l_{ij}) = v(g) - \sum_{ij \in g}(l_{ij} + l_{ji})$ in the partnership announcement subgame. Therefore, the efficient network g^*, which maximizes $v(g) - \sum_{ij \in g}(l_{ij} + l_{ji})$, is pairwise Nash stable because there is no profitable deviation for any firm. The rest of the proof is similar to the proof of Corollary 2. Q.E.D.

Computing the MJW value in Table 1: In the following, we show how we have obtained the allocation to firm 1 in Table 1 based on the MJW value in $g' = \{12, 23\}$ under Scenario 1. The rest of the allocations presented in Table 1 can be calculated by following the same procedure. Consider the following 6 orderings of the three firms: (1) $\{1, 2, 3\}$, (2) $\{1, 3, 2\}$, (3) $\{2, 1, 3\}$, (4) $\{3, 1, 2\}$, (5) $\{2, 3, 1\}$, and (6) $\{3, 2, 1\}$. The marginal values of firm 1 are all zero in the orderings (1)-(3), i.e., $v(g'|_{S \cup \{1\}}) - v(g'|_S) = 0$, where S is the set of firms before firm 1 in the orderings. This is because there exists no firm with residual supplies in S, so E_1 generates no value. In the ordering (4), the marginal value of firm 1 is also zero because firms 1 and 3 are not connected with each other without firm 2. In the ordering (5), $v(g'|_{\{2,3\} \cup \{1\}}) = E_1 + E_2$ and $v(g'|_{\{2,3\}}) = E_2$ because $H_3 \geq E_1 + E_2$. So the marginal value of firm 1 is E_1. Similarly, in the ordering (6), the marginal value of firm 1 is E_1. Therefore, the allocation to firm 1 under the MJW value is the average marginal value, $(E_1 + E_1)/6 = E_1/3$.

Second Paper Below

Combating Strategic Counterfeiters in Licit and Illicit Supply Chains

Soo-Haeng Cho · Xin Fang · Sridhar Tayur

Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA 15213
soohaeng@andrew.cmu.edu · xfang@andrew.cmu.edu · stayur@andrew.cmu.edu

Abstract: Counterfeit goods are becoming more sophisticated from shoes to infant milk powder and aircraft parts, creating problems for consumers, firms, and governments. By comparing two types of counterfeiters - deceptive, so infiltrating a licit (but complicit) distributor, or non-deceptive in an illicit channel, we provide insights into the impact of anti-counterfeiting strategies on a brand-name company, a counterfeiter, and consumers. Our analysis highlights that the effectiveness of these strategies depends critically on whether a brand-name company faces a non-deceptive or deceptive counterfeiter. For example, by improving quality, the brand-name company can improve her expected profit against a non-deceptive counterfeiter when the counterfeiter steals an insignificant amount of brand value. However, the same strategy does not work against the deceptive counterfeiter who can get a free ride on the improved quality. Reducing price works well in combating the non-deceptive counterfeiter, but it could be ineffective against the deceptive counterfeiter. Moreover, the strategies that improve the profit of the brand-name company may benefit the counterfeiter inadvertently and even hurt consumer welfare. Therefore, firms and governments should carefully consider a trade-off among different objectives in implementing an anti-counterfeiting strategy.

Key words: intellectual property, illegal operations, supply chain management

1 Introduction

Trademarks, also called brands, represent the most valuable assets of many firms, requiring significant investment in research and development as well as years of efforts in maintaining high product quality and careful brand management. Famous global brands such as GE, Nike and Nestlé are popular because they offer a guarantee of quality, which is vital to consumers when they make purchasing decisions. For those goods for which the mere display of a particular brand confers prestige on their owners, such as luxury watches and fashion apparel, many consumers purchase branded goods to demonstrate that they are consumers of the particular good. These intrinsic values of trademarks create incentives for counterfeiting.

Nowadays counterfeits have developed into a substantial threat to many industries. The OECD estimates that international trade in counterfeit could amount to up to \$250 billion or 1.95% of world trade in 2007, up from \$105 billion in 2001 (OECD 2009). If including domestically produced and consumed products, the total magnitude could be several hundred billion dollars more (OECD 2008). The problem is no longer limited to prestigious and easy-to-manufacture products, such as designer clothing, branded sportswear, and fashion accessories. It affects nearly all product categories including items that have an impact on personal health and safety such as pharmaceuticals, food, drink, toys, medical equipment, and automotive parts (OECD 2008).

Counterfeits are broadly categorized into two types: non-deceptive and deceptive (Grossman and Shapiro 1988a). A *non-deceptive* counterfeit is the counterfeit a consumer can distinguish from the brand-name product at time of purchase. This type of counterfeits tends to be sold at a substantial discount through an unauthorized sales channel. For example, consumers can easily tell that $10 luxury watches sold by street vendors are counterfeit. On the contrary, a *deceptive* counterfeit is the counterfeit a consumer believes to be authentic at time of purchase even if it is, in fact, counterfeit. In order to deceive consumers, this type of counterfeit goods has to infiltrate licit supply chains; for example, fake auto parts were found in legitimate repair shops, counterfeit pharmaceutical products at chemists, food products on supermarket shelves (OECD 2008), and pirated software products sold by one of the largest re-sellers (Bass 2010). Solomon (2009) notes that counterfeit drugs make their way through the licit supply chain via a distributor who moves a product from a low-cost channel to a high-cost channel. Collusion between counterfeiters and licit supply chain members occurs due to a higher profit from selling counterfeits (Green and Smith 2002, Bass 2010). A deceptive counterfeit is usually sold at the price that is the same as or close to that of its branded product so as to deceive consumers. Although it appears to function properly at time of purchase, it lacks durability and often involves health and safety risks of consumers. Examples of deceptive counterfeits abound in both developing and developed countries. In China, after the luxury furniture sold by the licit retailer turned out to be deceptive counterfeits, customers (who were previously deceived) have posted details of how their products were shoddily made or reeking of foul-smelling lacquers (Barboza 2011). In the U.S., a licit distributor who bought counterfeit networking cards for $25 each sold them to the Marine Corps for $625 each after repackaging the cards to make them appear to be Cisco products (McKinley 2010); and a number of physicians bought a fake version of cancer drug Avastin for $1,995 per 400-milligram vial from a Canadian company (cheaper than $2,400 of authentic Avastin), and billed patients the full list price (Weaver 2012, Weaver et al. 2012).

Brand-name companies are spending millions of dollars in order to stop or at least to reduce the incidence of counterfeits. They hire full-time employees, invest in new technologies, and redesign their products to make counterfeiting more difficult (Balfour 2005). However, the anti-counterfeiting strategies found to be useful to one product may not work for another or can even unintentionally make counterfeits flourish more in the market. For example, Chinese shoe manufacturers successfully addressed their counterfeiting issues by improving the quality of their products (Qian 2008). This is the outcome of the competition in which high-quality authentic products defeat low-quality *non-deceptive* counterfeits. However, the same strategy backfired against a Scotch whisky company in the Thailand market (Green and Smith 2002). At the peak of the company's sales in 1988, 42% of its premium Scotch whisky sales was counterfeit; high quality made the products more popular and attracted more counterfeits. In this case, the counterfeits were sold as the genuine products and commanding the same price, i.e., sold as *deceptive* counterfeits. After the initial attempt to fight counterfeits by improving quality had failed, the company eventually succeeded in radically reducing the incidence of counterfeiting by establishing a system that monitors supply chains: the company focused on identifying members in its supply chain who were selling the counterfeits, facilitating seizure of counterfeits and punishing counterfeiters.

These contrasting results illustrate a need for anti-counterfeiting strategies that are tailored to specific products. Yet, due to the limited understanding of relations among the types of counterfeits and the effectiveness of anti-counterfeiting strategies, OECD (2008) calls for research that strengthens the analysis of counterfeiting and says:

"Assessing the factors driving production and consumption of counterfeit and pirated products can generate insights into the types of products that are most likely to be infringed, . . . , and lead to more efficient and effective [anti-counterfeiting] strategies."

This paper attempts to provide such an analysis by providing insights to the following questions: (Q1) What anti-counterfeiting strategies should a brand-name company use to improve her own profit? (Q2) What is the impact of anti-counterfeiting strategies on the profit of a counterfeiter? (Q3) What is the impact of counterfeits on consumer welfare? Do consumers also benefit from the strategies that are effective in combating counterfeits?

To answer these questions, we develop a normative model of licit and illicit supply chains, in which a brand-name company competes with her potential counterfeiter. The counterfeiter in our model is either non-deceptive or deceptive, and decides the level of functional quality and wholesale price of his goods after observing the quality and price of the brand-name product. Depending on his type, the counterfeiter faces different opportunities and risks. The *non-deceptive* counterfeiter competes directly with a brand-name company for price and quality. Thus the counterfeiter may have to invest in improving the quality of his goods, which will increase the risk of losing the investment in case of getting caught by the authorities. Conversely, the *deceptive* counterfeiter may not need to invest as much in improving the quality as non-deceptive counterfeits (as long as he can deceive consumers successfully at time of purchase), but he has to infiltrate a licit supply chain via a legitimate distributor who sources both brand-name and counterfeit products. The legitimate distributor then faces a trade-off between a greater profit margin and a risk of getting punished for selling counterfeits.

After finding the equilibrium decisions of the counterfeiter and the distributor, we evaluate the following anti-counterfeiting strategies of which the effectiveness depends on the subsequent reaction of the strategic counterfeiter: (i) quality strategy that alters the quality of brand-name products against a counterfeiter, (ii) pricing strategy that alters the price of brand-name products against a counterfeiter, (iii) marketing campaign that educates consumers about the dangers of counterfeits, and (iv) enforcement strategy that increases the chances to seize the production of counterfeits. Our analysis highlights that the optimal strategy of the brand-name company differs depending on whether she faces the non-deceptive or deceptive counterfeiter. Although it is ideal to see the strategies that increase the profit of the brand-name company be also effective in reducing the profit of the counterfeiter and benefit consumers, our analysis shows that this is not the case for most strategies. It is therefore imperative for industries and governments to understand the type of potential counterfeiters and to carefully consider a trade-off among different objectives in implementing an anti-counterfeiting strategy.

2 Literature Review

Traditional supply chain management research is focused on licit supply chains in which members of supply chains interact with each other by exchanging goods and services legally. In this era of globalization, supply chains are no longer confined within one country as more and more companies offshore and outsource their operations to less developed countries. However, this has created a frightening phenomenon: an ever-rising flood of counterfeit items coming into markets (Business Week 2005). This paper is intended to shed light on counterfeit problems in both licit and illicit supply chains and to analyze the effectiveness of anti-counterfeiting strategies.

The majority of studies on counterfeits are conceptual and descriptive. They provide frameworks for fighting counterfeiting usually based on case studies. For example, Olsen and Granzin (1992) emphasize the importance of dealers' cooperation for a manufacturer to implement a program to combat counterfeits. Jacobs et al. (2001) investigate a number of counterfeiting incidences, and propose various measures of fighting these illegal activities. Staake and Fleisch (2008) provide an extensive review of this literature.

Marketing researchers have conducted empirical studies on counterfeits. They mainly focus on the demand side of counterfeits, and try to answer questions such as why consumers purchase counterfeits and how to educate consumers not to purchase counterfeits. Eisend and Schuchert-Guler (2006) review this literature and conclude that further investigation is needed to develop a general framework that integrates the existing results consistently. Recently, using data from

Chinese shoe companies, Qian (2008) finds that brand-name companies tend to improve their product quality after the entry of non-deceptive counterfeiters.

There are only a handful of analytical studies that present prescriptive models of counterfeits. Grossman and Shapiro (1988a, 1988b) develop equilibrium models of trades between brand-name firms in a home country and low-quality producers in a foreign country. To sell their goods as counterfeits in the home market, foreign producers must pass the goods through the home-country border, hence facing the risk of confiscation. Grossman and Shapiro (1988a) analyze the consequences of deceptive counterfeits in a market where consumers cannot observe the quality of a product, and provide a welfare analysis of border inspection policy. Grossman and Shapiro (1988b) present a Cournot competition model between brand-name products and non-deceptive counterfeits given their exogenous quality levels. Because non-deceptive counterfeits can contribute positively to consumer welfare due to their lower price, the authors conclude that policies that discourage foreign counterfeiting need not improve welfare, which is consistent with our finding. Scandizzo (2001) views competition between brand-name firms and non-deceptive counterfeiters as a patent race over time. The author finds that counterfeits improve consumer welfare while reducing firms' profits, and that the more skewed the income distribution within the economy is towards the poor, the greater the welfare effect and the smaller the profit effect.

There have been growing interests in counterfeit research among operations researchers. Liu et al. (2005) study the decision of an inventory manager who can source both genuine and deceptive counterfeit products and sell them to consumers at one price. Sun et al. (2010) study a global firm's decision of outsourcing the production of its components to a foreign country. The firm faces a trade-off between lower labor cost and increased risk of imitation by a foreign firm. The authors find an optimal strategy in choosing the range of components to transfer. Zhang et al. (2012) analyze the case when a brand-name firm faces non-deceptive counterfeits. They show that a non-deceptive counterfeit lowers the price and profit of the brand-name product, and a brand-name firm has more incentive to improve her own quality rather than reducing that of a counterfeit. They also analyze a situation in which two brand-name products compete, which we do not consider in this paper.

We draw on and contribute to this stream of research by addressing the following important issues in counterfeiting problems:

(1) Strategic counterfeiters: The common assumption used in the literature is that the quality is fixed a priori. For example, Grossman and Shapiro (1988a, 1988b) assume that foreign producers always choose the lowest quality because they lack capital, resource, and technology for quality improvement and that there are no entry costs of counterfeiters. Today, thanks to outsourcing

and offshoring of numerous global firms, counterfeiters benefit greatly from increasingly easy access to modern production facilities (Staake and Fleisch 2008). Schmidle (2010) notes that today's counterfeiters come in varying levels of quality depending on their intended markets, and diversify their products and distribution channels to manage the risks involved in their criminal activities. In our model, a counterfeiter decides functional quality and wholesale price of his products by considering a trade-off between the benefit from stealing brand value and the risk of confiscation. Our analysis shows that the effectiveness of anti-counterfeiting strategies depends critically on the strategic response of a counterfeiter to those strategies.

(2) Licit and illicit supply chains: The previous analytical papers assume that a counterfeiter is capable of selling his counterfeits directly to consumers regardless of his type. Although this is quite possible for non-deceptive counterfeits, a *deceptive* counterfeiter has to infiltrate a licit supply chain; today, very few consumers would be deceived by the counterfeits sold by street vendors or unknown websites. We take into account this *fundamental* difference in supply chains of non-deceptive and deceptive counterfeits, and demonstrate that an effective strategy against a non-deceptive counterfeiter may not be effective against a deceptive counterfeiter.

(3) Consumer characteristics: As consumers learn more about counterfeit problems from the media, they become more aware of the presence of counterfeits, and some even become more proactive by taking into account the likelihood of receiving *deceptive* counterfeits unknowingly when they purchase branded products from licit distributors. Our survey (of which the details are presented in §3) indicates that the proportion of proactive consumers in the U.S. is substantially lower than that in China. Our analysis provides insights into how this characteristic of consumers affects the effectiveness of anti-counterfeiting strategies.

(4) Evaluation of anti-counterfeiting strategies: We evaluate the aforementioned strategies by examining their impacts on a brand-name company, a counterfeiter, and consumers. Our analysis complements the previous findings (discussed above) of Grossman and Shapiro (1988a, 1988b) and Zhang et al. (2012). Grossman and Shapiro (1988a, 1988b) provide welfare analyses of border inspection policies, which are similar to the enforcement policies we study. Today, however, the confiscation of counterfeits not only occurs on the border when trading goods between countries but also occurs within each country due to the growing demand within its domestic market; in addition, the seizure of equipment is another threat to counterfeiters (Staake and Fleisch 2008). For example, the Chinese authorities, long unconcerned about counterfeiting, have begun to take actions as Chinese companies create their own intellectual properties (Business Week 2005). Zhang et al. (2012) focus on analyzing the effect of altering the quality of a brand-name good or a *non-deceptive* counterfeit on the profit of a brand-name firm. In doing so, they consider neither potential seizure

of counterfeits and equipment, nor welfare implication of those strategies.

Finally, we note that a counterfeiter's decision of his distribution channel is analogous to that of a legitimate firm (e.g., Xu et al. 2010), although the benefit and risk associated with each channel of counterfeits are unique as described above. Also, a research question similar to counterfeiting arises in the literature of parallel importing (or gray market) and software piracy. Parallel importing is the practice of purchasing products in a lower-priced region and shipping them to a higher priced region (e.g., Ahmadi and Yang 2000, Hu et al. 2011). While the parallel imported goods are authentic but sold at a lower price, counterfeits are not authentic, possess lower quality, and are sold at a lower price for non-deceptive counterfeits or at the same price for deceptive counterfeits. Piracy differs from counterfeiting in that piracy refers to infringement of copyright. In our model, software piracy can be viewed as a special case of counterfeiting, in which counterfeit products have almost the same functional quality as authentic ones but their cost of development and production is very low. Some of our results can be extended to software piracy problems; for example, consumers could be better off without piracy protection, which is consistent with Conner and Rumelt (1991).

In summary, the literature considers only one type of counterfeits with fixed quality that are sold directly to consumers. In contrast, our model captures recent changes in counterfeiting supply and demand by noting the fundamental differences between non-deceptive and deceptive counterfeits in consumers' awareness and distribution channels, and by considering counterfeiters' strategic decisions regarding price and functional quality in a market with different consumer characteristics. Our analysis provides novel insights into the effectiveness of several anti-counterfeiting strategies.

3 Model

We consider a market served by a brand-name company ('she') and her potential counterfeiter ('he'). The type of the counterfeiter is either non-deceptive or deceptive. We use subscript $i = B$ to denote the brand-name product, $i = N$ to denote the non-deceptive counterfeit, and $i = D$ to denote the deceptive counterfeit. A consumer in this market purchases at most one unit of a product. In making a purchasing decision of product i, a consumer considers his/her utility $u_i = \theta\phi_i - p_i$, where θ represents his/her taste, ϕ_i represents the quality of the product a consumer *perceives* at time of purchase, and p_i represents the retail price of the product. All consumers prefer high quality for a given price, but a consumer with a higher θ is more willing to pay to obtain a high-quality product. We assume that θ is uniformly distributed over $[0, 1]$ and that the size of the market is one. A consumer purchases a product only if the utility from purchasing the product is nonnegative in which case he/she selects a product that provides the highest utility. This is the standard vertical differentiation model, which is also used by Qian (2008) and Zhang et al. (2012). We next present our model components that capture the unique aspects of counterfeiting.

Depending on the counterfeit type, the quality of product i a consumer *perceives* at time of purchase, ϕ_i, may differ from its *real* quality q_i. (Throughout this paper, unless mentioned specifically as the perceived quality, quality refers to real quality.) For the *non-deceptive* counterfeit as well as the brand-name product, consumers know what product they are purchasing, so the perceived quality of either product is the same as its real quality; i.e., $\phi_B = q_B$ and $\phi_N = q_N$. However, for the *deceptive* counterfeit, consumers cannot distinguish it from the brand-name product at time of purchase. There are two types of consumers. First, some consumers are not aware of counterfeits, or even if they are aware, they may consider the likelihood of purchasing counterfeits negligible at legitimate stores. They perceive the quality of any product in the market as q_B; i.e., $\phi_B = \phi_D = q_B$. On the other hand, other consumers may be "proactive" in the sense that they take into account the likelihood of receiving deceptive counterfeits unknowingly even when purchasing products from legitimate stores. Let $\xi_s \in [0, 1]$ denote their expectation about the fraction of deceptive counterfeits in the market. Then proactive consumers perceive the quality of a product in the market as a weighted average of the quality of the brand-name product and that of the deceptive counterfeit; i.e., $\phi_B = \phi_D = (1 - \xi_s)q_B + \xi_s q_D$. Let λ ($\in [0, 1]$) denote the fraction of proactive consumers in the market. In practice, λ may vary depending on the characteristic of the market. For example, in our survey of 166 consumers over 4 popular product categories for deceptive counterfeits (see Table 1), we have found that 51% of consumers in China are proactive, whereas only 4 % of consumers in the U.S. are proactive. The low value of λ in the U.S. reflects the view of Rockoff and Weaver (2012), who say: "Most Americans don't question the integrity of the drugs they rely on. They view drug counterfeiting, if they are aware of it at all, as a problem for developing countries."

Table 1. Consumer Survey Results in the U.S. and China

	U.S.		China	
	Aware	Proactive	Aware	Proactive
Alcohol	14%	4%	94%	56%
Car Parts	25%	4%	54%	34%
Medical Drugs	41%	5%	86%	51%
Food, Drinks	22%	5%	90%	63%
Average	26%	4%	81%	51%

(Note) Respondents are college students and faculty with ages from 18 to 50. The number of respondents is 86 in the U.S., and it is 80 in China. Two questions were asked in the questionnaire: (1) Are you aware of the sale of counterfeits in each of the above product categories; and (2) For each product category in which you are aware of the sale of counterfeits, do you take into account the risk of getting a counterfeit and therefore discount the value of the product when you purchase a brand-name product at a full price in a legal store? Those customers who answered yes to (1) are considered "Aware", and those customers who answered yes to both (1) and (2) are considered "proactive." The absolute numbers may be escalated because respondents may be reminded of counterfeits by the questionnaire. Our survey indicates that being "aware" of the existence of counterfeits differs from being "proactive." One may explain such difference from cognitive psychology (e.g., Bendoly et al. 2010, Goldsmith and Amir 2010, and references therein); for example, it may be due to a positive-outcome "bias" or "wishful thinking" caused by overestimating the probability of good things happening.

Since the counterfeit bears the trademark of the brand-name product, a consumer enjoys the brand image even when he/she purchases the counterfeit. Thus we may represent the quality of the counterfeit as $q_i = f_i + \beta q_B$ ($i = N$ or D), where f_i (> 0) is the functional quality of the counterfeit i and βq_B (where $\beta > 0$) is the brand value that the counterfeit steals from the brand-name product. Essentially, we assume that a product has two attributes: functionality and brand value as in multi-attribute models in marketing (e.g., see Lilien et al. 1992). Brand value reflects advertising investments on which a counterfeiter may get a free ride. The parameter β captures the following two factors. First, β captures a fraction of the brand value in the quality of the brand-name product, q_B. For example, this fraction may be high for luxury goods because a brand plays a significant role when consumers purchase such products, whereas it may be low for fast moving consumer goods (which are sold quickly at relatively low cost) because a brand is less of a concern to consumers for such goods. Second, β captures a discount factor of the original brand value for the counterfeit because the counterfeit draws only a part of the brand value from the brand-name product.[1] Following the literature, we assume that the quality of the brand-name product is superior to that of the counterfeit; i.e., $q_B > q_N$ and $q_B > q_D$.

Either type of counterfeiter i ($= N$ or D) makes two decisions sequentially to maximize his expected profit: functional quality f_i and wholesale price w_i to a distributor. We assume that the counterfeiter makes these decisions after observing the quality q_B and price p_B of a brand-name product because counterfeiters always enter a market following a brand-name company, often after the brand-name product becomes popular. Different types of counterfeiters use different distribution channels to sell their goods. The *non-deceptive* counterfeiter ($i = N$) distributes his goods through an *illicit distributor*, who then decides the retail price of the non-deceptive counterfeit to consumers, p_N. On the other hand, the *deceptive* counterfeiter ($i = D$) has to break into a licit supply chain by distributing his goods through a *licit distributor*, who then sells both brand-name products and deceptive counterfeits to consumers at the same price p_B. In this case, the licit distributor determines a proportion $s \in [0, 1]$ of the deceptive counterfeit among all products he

[1] For some counterfeits, as their functional quality increases, they might look more similar to branded products so that they can steal a higher fraction of the brand quality. This can be modeled by setting $\beta = \beta_1 + \beta_2 f_i$, where β_1 captures the characteristic of a product category (like β in our base model), and β_2 captures the property mentioned above. All our subsequent results hold under this alternative model.

sells to consumers. We next describe the details of our model for non-deceptive and deceptive counterfeits, respectively.

When a *non-deceptive* counterfeiter exists in the market and sells his products through the illicit distributor, consumers will choose between the brand-name product and the counterfeit. Both products carry the same brand, but have different qualities and prices. The competition between the non-deceptive counterfeiter and the brand-name company is analogous to duopoly in a vertically differentiated market, but it is not the same because the members of the illicit supply chain bear the risks associated with counterfeiting. The non-deceptive counterfeiter and the illicit distributor make their decisions in three sequential stages as follows. In stage 1, the non-deceptive counterfeiter chooses his functional quality $f_N \in [\underline{f}, \overline{f}]$, where $\overline{f} > \underline{f} \geq 0$, and makes initial investment to develop and produce goods having f_N. The upper bound $\overline{f}$ may represent the functional quality of the brand-name product. We assume $\overline{f} < (1-\beta)q_B$ such that $q_B > q_N$. The lower bound $\underline{f}$ may represent the minimum level of quality at which a product functions or appears to function properly. To produce counterfeits having f_N, the counterfeiter needs to invest $t_N f_N^2$ in acquiring technology and setting up production facilities, where $t_N > 0$. This implies that the development of a product with higher quality requires increasingly more investment. The unit production cost of the counterfeit is normalized to zero. After the investment takes place, however, there are some chances that the investment will be confiscated because it is illegal to produce counterfeits. Suppose this occurs with a probability $\gamma \in (0, 1)$. The parameter γ captures the monitoring efforts of the government and the brand-name company on counterfeit production. The potential loss of the investment is a major risk to the counterfeiter that deters him from making large investments to improve the functional quality of his products (OECD 2008). If the counterfeiter's investment is confiscated, the counterfeiter cannot sell his goods to the market. Otherwise, the game proceeds to stage 2 in which the non-deceptive counterfeiter decides his wholesale price w_N to the illicit distributor. For simplicity, we represent all distributors/retailers in the illicit supply chain as one illicit distributor. In stage 3, the illicit distributor decides the retail price of the non-deceptive counterfeit to consumers, p_N. The illicit distributor has to pay a penalty of l_N if getting caught by the authorities with probability α_N.

When a *deceptive* counterfeiter exists in the market and sells his products through the licit distributor, consumers cannot distinguish deceptive counterfeits from brand-name products. Like the non-deceptive counterfeiter, the deceptive counterfeiter determines his functional quality $f_D \in [\underline{f}, \overline{f}]$ in stage 1, while facing the risk of getting his investment $t_D f_D^2$ confiscated. In stage 2, the deceptive counterfeiter decides his wholesale price w_D to the licit distributor, who later sells the counterfeits to consumers at the same price p_B as the brand-name products. In stage 3, the licit distributor determines a proportion s of the deceptive counterfeit among all products he sells to consumers. We model the risk of the licit distributor selling deceptive counterfeits with a likelihood α_D of getting caught and a penalty l_D. Since α_D tends to increase with more counterfeits in the market, we set α_D equal to the fraction of deceptive counterfeits, s. In §7, we consider a more general case in which α_D is a function of f_D as well as s.

We make the following assumptions to simplify our analysis. First, we assume that the licit distributor does not make a profit from selling brand-name products, while it makes a positive profit from selling deceptive counterfeits. Our results continue to hold for any fixed margin of the licit distributor from selling brand-name products. In online appendix C, we also analyze the case where the licit distributor decides its profit margins endogenously. Second, we normalize $l_N = 0$, while having $l_D = l > 0$. In practice, a loss of an illicit distributor from potential seizure is much smaller than that of a licit distributor. Illicit distributors are usually street vendors or internet sites. Since their potential loss from seizure is small, they tend to close their stores temporarily when they get caught and then reopen the same stores or open new ones later. For example, vendors in the Xiang Yang market in Shanghai, China, which were once famous for their high-quality counterfeits but closed due to the government's massive campaigns in 2006, relocated to the Yatai Xinyang market that is now famous among tourists (Naumann 2009). In contrast, the punishment on the licit distributor for illegal distribution of deceptive counterfeits is very severe. For example, the Chinese court sentenced the distributor of fake pills to 17 years in prison, the nation's longest term for the crime (Bennett 2010) and the U.S. court sentenced a distributor who sold counterfeit networking cards to the military to 51 months in prison, the maximum term recommended by federal prosecutors (McKinley 2010). Third, for both types of counterfeits, we assume that the probability of counterfeits getting confiscated at the production level (γ) is independent of that at the distribution level (α_N or α_D). In practice, it is extremely difficult to trace back the source of counterfeits even after discovering their distributors. For example, counterfeit versions of cancer drugs Faslodex and Avastin were detected at the distribution level in the U.S., but their sources have not been determined while suspecting offshore production (Rockoff et al. 2012, Weaver et al. 2012). Table 2 summarizes the notation of major variables and parameters.

4 Equilibrium Analysis

In this section, we present our model formulation and equilibrium analysis. We use backward induction to derive subgame-perfect Nash equilibrium. In §4.1 we present equilibrium (denoted by superscript *) in a market with a non-deceptive counterfeiter. In §4.2 we present equilibrium (denoted by superscript **) in a market with a deceptive counterfeiter. All proofs are provided in online appendix A.

Table 2. Summary of Key Notation

Symbol	Definition
i	Brand-name product (= B), non-deceptive counterfeit (= N), deceptive counterfeit (= D)
θ	Taste of consumers; $\theta\tilde{}U[0,1]$
p_i	Retail price of product i to consumers
q_i	(Real) Quality of product i
f_i	Functional quality of counterfeit product i; $f_i \in [\underline{f}, \overline{f}]$
π_i	Expected profit from selling product i
w_i	Wholesale price of product i to a distributor
t_i	Cost parameter used in the cost of developing functional quality
β	Fraction of the quality of brand-name products that counterfeits steal; $\beta \in \left(0, 1 - \frac{\overline{f}}{q_B}\right)$
γ	Probability that a counterfeiter's investment will be confiscated; $\gamma \in (0,1)$
l	Loss of the licit distributor if getting caught for selling deceptive counterfeits; $l > 0$
λ	Fraction of proactive consumers in the market; $\lambda \in [0,1]$
s	Fraction of deceptive counterfeits among all products the licit distributor sells; $s \in [0,1]$

4.1 Non-Deceptive Counterfeits

Suppose the brand-name product and the non-deceptive counterfeit exist in the market. There are three segments of consumers: (i) consumers who value the quality of a product highly and purchase the brand-name product, (ii) consumers who value the quality less and purchase the non-deceptive counterfeit, and (iii) consumers who value the quality the least and do not purchase any product. The consumer who is indifferent between purchasing the brand-name product and the non-deceptive counterfeit has the taste $\widetilde{\theta} = \frac{p_B - p_N}{q_B - q_N} = \frac{p_B - p_N}{(1-\beta)q_B - f_N}$, which solves $\widetilde{\theta} q_N - p_N = \widetilde{\theta} q_B - p_B$. Similarly, the consumer who is indifferent between purchasing the non-deceptive counterfeit and not purchasing any product has the taste $\widehat{\theta} = \frac{p_N}{q_N} = \frac{p_N}{f_N + \beta q_B}$. Let m_i ($\in [0,1]$) denote the market share of product i (= B or N), and let m_0 denote the proportion of consumers who do not purchase any product, so that $m_B + m_N + m_0 = 1$. Then:

$$m_B = 1 - \widetilde{\theta} = 1 - \frac{p_B - p_N}{(1-\beta)q_B - f_N} \text{ and } m_N = \widetilde{\theta} - \widehat{\theta} = \frac{p_B - p_N}{(1-\beta)q_B - f_N} - \frac{p_N}{f_N + \beta q_B}. \quad (1)$$

In stage 3, the illicit distributor determines the retail price to consumers, p_N, by solving:

$$\max_{p_N} (p_N - w_N) m_N = (p_N - w_N) \left\{ \frac{p_B - p_N}{(1-\beta)q_B - f_N} - \frac{p_N}{f_N + \beta q_B} \right\}. \quad (2)$$

By noting that the profit of the illicit distributor in (2) is concave in p_N, one can easily obtain her optimal retail price $p_N^*(w_N, f_N) = \frac{(\beta q_B + f_N) p_B + q_B w_N}{2 q_B}$.

In stage 2, the non-deceptive counterfeiter determines his wholesale price w_N. By anticipating the best response of the illicit distributor, the non-deceptive counterfeiter chooses his optimal wholesale price that maximizes his expected profit given by:

$$\pi_N(w_N, f_N) = (1-\gamma)\left\{w_N\left(\frac{p_B - p_N^*}{q_B - q_N} - \frac{p_N^*}{q_N}\right) - t_N f_N^2\right\} - \gamma t_N f_N^2. \tag{3}$$

In (3), $(1-\gamma)$ represents the likelihood that the counterfeiter is able to sell his goods without being confiscated, and the next term in the bracket represents the profit of the counterfeiter in that case. The initial investment $t_N f_N^2$ is considered a sunk cost in (3). Note that whether the confiscation of investment occurs after stage 1 or stage 2 does not affect the counterfeiter's decisions. If confiscation occurs after some units are sold, $(1-\gamma)$ can be interpreted as the fraction of sales the counterfeiter has generated before confiscation. Since π_N is concave in w_N, we can easily obtain the optimal wholesale price w_N^* and the corresponding expected profit of the non-deceptive counterfeiter π_N^*, respectively, as follows:

$$w_N^*(f_N) = \frac{p_B(f_N + \beta q_B)}{2q_B} \text{ and } \pi_N^*(f_N) = \frac{p_B^2(1-\gamma)(f_N+\beta q_B)}{8q_B\left\{(1-\beta)q_B - f_N\right\}} - t_N f_N^2. \tag{4}$$

By substituting w_N^* into p_N^*, one can verify that the illicit distributor charges a lower price than that of the brand-name product; i.e., $p_N^* < p_B$.

In stage 1, the non-deceptive counterfeiter decides the functional quality f_N by considering his optimal wholesale price in stage 2 and the best response of the illicit distributor in stage 3. The counterfeiter solves $\max_{f_N \in [\underline{f}, \overline{f}]} \pi_N^*(f_N)$ by considering the following trade-off: a higher level of functional quality will draw more consumers, but it requires more investment, hence increasing a potential loss from seizure.

Lemma 1 *For any given (p_B, q_B), the optimal functional quality of non-deceptive counterfeits, f_N^*, is as follows: if $t_N < \frac{(1-\gamma)p_B^2}{4\left\{(1-\beta)q_B - \underline{f}\right\}^3}$ and $\pi_N^*(\underline{f}) \geq \pi_N^*(\overline{f})$, then $f_N^* = \underline{f}$, and otherwise f_N^* can be $\overline{f}$ or $f_N^* \in (\underline{f}, \overline{f})$ that satisfies $\frac{\partial \pi_N^*}{\partial f_N}|_{f_N = f_N^*} = 0$.*

Lemma 1 shows that the non-deceptive counterfeiter may not always choose the lowest quality in contrast to the common assumption used in the literature (e.g., Grossman and Shapiro 1988a,b). In the *past*, non-deceptive counterfeits with low functional quality such as brand-name costumes, footwear and accessories dominated a counterfeit market. Their functional quality is just enough for consumers to use them, but their durability and performance are substandard. Consumers who purchase such counterfeits are those who want to enjoy the snob appeal of brands, but do not want to pay the high price of genuine goods. However, in *today*'s counterfeit markets, counterfeiters come in varying levels of quality depending on their intended markets (Schmidle 2010). For example, some counterfeit electronic devices such as cell phones include appealing features which are not included even in authentic products. This is called Shan-Zhai phenomenon in China. According to Gartner, Shan-Zhai phones account for more than 20 percent of sales in China (Barboza 2009). These counterfeiters usually face the least pressure from local enforcement agencies and some are likely to turn into licit competitors once intellectual property rights become more strictly enforced (Staake and Fleisch 2008). Our result stated in Lemma 1 is consistent with this observation of today's counterfeit markets.

4.2 Deceptive Counterfeits

Suppose the brand-name product and the deceptive counterfeit exist in the market. In this case, both brand-name products and deceptive counterfeits are sold at price p_B. While proactive consumers with proportion λ perceive the quality of a product in the market as $(1-\xi_s)q_B + \xi_s q_D$, the rest of consumers perceive the quality of a product in the market as q_B. Similar to Grossman and Shapiro (1988a), we assume that the expectation of proactive consumers about the fraction of deceptive counterfeits in the market is rational and hence is equal, in equilibrium, to the actual fraction of counterfeits; i.e., $\xi_s = s$. This notion of rational expectations equilibrium is also used in the recent operations management literature (e.g., Su and Zhang 2008, Cachon and Swinney 2009).

Similar to §4.1, we can obtain the market share of the brand-name product and that of the deceptive counterfeit, respectively, as follows:

$$m_B = (1-s)(1-\bar{\theta}) \text{ and } m_D = s(1-\bar{\theta}), \tag{5}$$

where $1-\bar{\theta} \equiv 1 - \frac{\lambda p_B}{(1-s)q_B + s(f_D+\beta q_B)} - \frac{(1-\lambda)p_B}{q_B}$ represents the "aggregate demand" for both brand-name and counterfeit products at price p_B. Among those consumers who purchase products for p_B, a fraction s of them receives deceptive counterfeits unknowingly.

In stage 3, the licit distributor solves the following problem to determine s:

$$\max_{s\in[0,1]} s(1-s)(p_B - w_D)\left\{1 - \frac{\lambda p_B}{(1-s)q_B + s(f_D+\beta q_B)} - \frac{(1-\lambda)p_B}{q_B}\right\} - sl. \tag{6}$$

In (6), $(1-s)$ represents the likelihood that the distributor will not be detected for selling counterfeits and $s(p_B - w_D)\left\{1 - \frac{\lambda p_B}{(1-s)q_B + s(f_D+\beta q_B)} - \frac{(1-\lambda)p_B}{q_B}\right\}$ represents the distributor's profit in that case. Recall that the distributor's profit margin from selling brand-name products is assumed zero (see §3). The term '$-sl$' in (6) represents the expected loss from potential seizure. From (6), we can show that the profit of the distributor is strictly decreasing in s for $s \in [\frac{1}{2}-\epsilon, 1]$, where ϵ is a small and positive constant. Moreover, the profit given in (6) is concave in s for $s < \frac{1}{2}$. Thus, s^{**} is 0 or it satisfies the first order condition in $(0, 0.5)$. In the remainder of this paper, we only consider the latter case (i.e., $s^{**} \in (0, 0.5)$) because there will be no deceptive counterfeits in the market when the counterfeiter fails to break into the licit supply chain (i.e., $s^{**} = 0$).

In stage 2, the deceptive counterfeiter decides his wholesale price w_D to maximize his expected profit given by:

$$\pi_D(w_D, f_D) = (1-\gamma)\left[w_D s^{**}\left\{1 - \frac{\lambda p_B}{(1-s^{**})q_B + s^{**}(f_D + \beta q_B)} - \frac{(1-\lambda)p_B}{q_B}\right\} - t_D f_D^2\right] - \gamma t_D f_D^2. \tag{7}$$

By noting that π_D is continuous in $w_D \in [0, p_B]$, we know that the optimal wholesale price w_D^{**} always exists in $[0, p_B]$. In the case when $\lambda > 0$, the closed-form expressions for s^{**} and w_D^{**} do not exist. In the case when $\lambda = 0$, we can obtain from the first-order condition of (6) that $s^{**}(w_D, f_D) = \frac{1}{2} - \frac{l q_B}{2(p_B - w_D)(q_B - p_B)}$. By substituting s^{**} into (7) and solving $\max_{w_D \in [0, p_B]} \pi_D(w_D, f_D)$, we obtain w_D^{**} and the corresponding expected profit of the deceptive counterfeiter π_D^{**} as follows:

$$w_D^{**}(f_D) = p_B - \sqrt{\frac{l p_B}{1 - \frac{p_B}{q_B}}} \text{ and } \pi_D^{**}(f_D) = \frac{1}{2}(1-\gamma)\left\{\sqrt{p_B\left(1 - \frac{p_B}{q_B}\right)} - \sqrt{l}\right\}^2 - t_D f_D^2. \tag{8}$$

From (8), we can generate the following insights. First, as the risk of the licit distributor selling counterfeits increases with l, the deceptive counterfeiter has to reduce his price w_D^{**} to compensate for the increased risk, resulting in a decrease in his expected profit π_D^{**}. Second, π_D^{**} increases with $p_B\left(1 - \frac{p_B}{q_B}\right)$, which is the revenue of the brand-name company without counterfeits. This is because the deceptive counterfeit gets a free ride on the brand name of the genuine product.

In stage 1, the counterfeiter decides the functional quality f_D by considering his optimal wholesale price in stage 2 and the best response of the licit distributor in stage 3. The following lemma shows that the deceptive counterfeiter may choose a different level of functional quality depending on the fraction of proactive consumers in the market, λ.

Lemma 2 *For any given (p_B, q_B), when $\lambda = 0$, the optimal functional quality of deceptive counterfeits f_D^{**} is $\underline{f}$. When $\lambda > 0$, there exists $\bar{t}_D$ (> 0) such that if $t_D \geq \bar{t}_D$, $f_D^{**} = \underline{f}$, and otherwise f_D^{**} can be $\overline{f}$ or $f_D^* \in (\underline{f}, \overline{f})$ that satisfies $\frac{\partial \pi_D^{**}}{\partial f_D}|_{f_D = f_D^*} = 0$.*

In the market with no proactive consumers (i.e., $\lambda = 0$), as one would expect, the deceptive counterfeiter always chooses the lower bound $\underline{f}$ for his functional quality because improving quality does not increase counterfeit sales. In this case, although a counterfeit is visually identical to its brand-name product, its low quality may result in a substantial financial loss to consumers or even endanger their health and safety. Consequently, both counterfeiter and distributor often face considerable punishments if they get caught. Typical examples are food, beverage, agricultural products, pharmaceuticals, and automotive spare parts (OECD 2008, Staake and Fleisch 2008). In the market with proactive consumers (i.e., $\lambda > 0$), although consumers cannot distinguish the deceptive counterfeit from the brand-name product, the deceptive counterfeiter can still find it optimal to improve his functional quality above the minimum level $\underline{f}$. The reason is as follows. When f_D is improved, both aggregate demand for brand-name and counterfeit products, $1 - \frac{\lambda p_B}{(1-s^{**})q_B + s^{**}(f_D+\beta q_B)} - \frac{(1-\lambda)p_B}{q_B}$, and the fraction of deceptive counterfeits, s^{**}, are increased. Thus the marginal benefit of functional quality is positive. If the marginal benefit exceeds the marginal cost, then the deceptive counterfeiter will choose his functional quality f_D^{**} above $\underline{f}$. In practice, some deceptive counterfeits reveal different levels of functional quality; for example, fake gasoline with different levels of adulteration has been reported (Lee et al. 2011).

5 Anti-Counterfeiting Strategies: Quality and Price

Having analyzed the equilibrium decisions of counterfeiters and distributors in licit and illicit supply chains, we examine the effectiveness of anti-counterfeiting strategies: quality and pricing strategies in §5, and marketing and enforcement strategies in §6. We analyze each strategy separately in order to isolate its effect on firms' profits and consumer welfare. When a firm implements multiple strategies simultaneously, one needs to aggregate the effect of each strategy to evaluate the overall effect.

We examine the effectiveness of quality and pricing strategies against the *non-deceptive* counterfeiter in §5.1, and against the *deceptive* counterfeiter in §5.2; then, we compare them in §5.3. In each of §5.1 and §5.2, we proceed our analysis as follows. First, we examine whether the brand-name company should choose higher/lower quality or price than the case with no counterfeiter in order to maximize *her expected profit* against the counterfeiter. Let q_B^m and p_B^m denote the optimal quality and price of the brand-name product with no counterfeiter in the market, respectively. Similarly, let q_B^* and p_B^* (resp., q_B^{**} and p_B^{**}) denote the optimal quality and price of the brand-name product in the presence of the *non-deceptive* (resp., *deceptive*) counterfeiter, respectively. Second, knowing that such strategies of choosing q_B^* and p_B^* (resp., q_B^{**} and p_B^{**}) instead of q_B^m and p_B^m improve the expected profit of the brand-name company, we examine how those strategies affect the *expected profit of the non-deceptive (resp., deceptive) counterfeiter*. Finally, we investigate how those strategies affect *expected consumer welfare*, which is defined as follows. When only brand-name products exist in the market, we can define consumer welfare as $CS_B = \int_{\frac{p_B}{q_B}}^1 (\theta q_B - p_B) d\theta$. Similarly, using (1) and (5), we can define CS_N or CS_D as consumer welfare in the market where non-deceptive or deceptive counterfeits co-exist with brand-name products, respectively, as follows:

$$CS_N = \int_{\widehat{\theta}}^{\widetilde{\theta}} (\theta q_N - p_N)\, d\theta + \int_{\widetilde{\theta}}^1 (\theta q_B - p_B) d\theta; \tag{9}$$

$$CS_D = s \int_{\bar{\theta}}^1 (\theta q_D - p_B)\, d\theta + (1-s) \int_{\bar{\theta}}^1 (\theta q_B - p_B)\, d\theta. \tag{10}$$

In (9), the first term represents the surplus of those consumers who purchase the non-deceptive counterfeit and the second term represents the surplus of those consumers who purchase the brand-name product. In (10), the first term represents the surplus of those consumers who are cheated and receive the deceptive counterfeit although they pay the price of the brand-name product, and the second term represents the surplus of those consumers who purchase and receive the brand-name product. Considering the chances that counterfeits do not reach the market due to seizure, we can further define ECS_N or ECS_D as the *expected* consumer welfare when the counterfeiter is non-deceptive or deceptive, respectively, as follows:

$$ECS_N = (1-\gamma)CS_N + \gamma CS_B \text{ and } ECS_D = (1-\gamma)CS_D + \gamma CS_B. \tag{11}$$

Let ECS_N^* or ECS_D^{**} denote the corresponding expected consumer welfare in equilibrium. We can show that $ECS_D^{**} < CS_B < ECS_N^*$. Intuition from this result is as follows. When non-deceptive counterfeits exist in the market, a consumer has a cheap alternative to the brand-name product. In equilibrium, the non-deceptive counterfeiter sets his price and functional quality such that he offers a higher utility to those consumers who enjoy the brand value of the brand-name product but do not appreciate its high quality or cannot afford its high price. Therefore, non-deceptive counterfeits improve consumer welfare. In contrast, when deceptive counterfeits exist, some consumers are cheated to receive low-quality deceptive counterfeits, resulting in a welfare loss.[2]

5.1 Non-Deceptive Counterfeits

This subsection examines the brand-name company's anti-counterfeiting strategies against the *non-deceptive* counterfeiter. We first examine the brand-name company's *quality* strategy to combat the non-deceptive counterfeiter. In the following proposition, we present the results for the case when $f_N^* = \underline{f}$ or $\overline{f}$, since the exposition of our results is much simpler in this case, while presenting the results for the case when $f_N^* \in (\underline{f}, \overline{f})$ in online appendix A (which involve complex conditions for parts (a) and (c)).

Proposition 1 *Suppose* $f_N^* = \underline{f}$ *or* $\overline{f}$. *Then:*
(a) $q_B^* > q_B^m$ *if and only if* $\beta < 1 - \left\{ \frac{q_B^m - q_N^*(q_B^m)}{q_B^m} \right\}^2$.

[2]We do not consider the socio-economic effects of counterfeiting on corruption, criminal activities, employment, environment, innovation, tax revenues, and so on. If taking into account these indirect or long-term effects into account, then non-deceptive counterfeits may also decrease consumer welfare. Moreover, the anti-counterfeiting strategies that reduce the incidence of counterfeits (e.g., marketing campaigns that reduce β) can have more positive benefits by lessening these harmful effects. We can also examine the aggregate effect of anti-counterfeiting strategies on social welfare, which may be defined as $SW_i = ECS_i + \pi_B - \pi_i$ for $i = N$ or D, by combining the results of profits and consumer welfare.

(b) Suppose $q_B^* > q_B^m$ (resp., $q_B^* < q_B^m$). Then π_N^* is lower (resp., higher) at $q_B = q_B^*$ than at $q_B = q_B^m$.

(c) Suppose $q_B^* > q_B^m$ (resp., $q_B^* < q_B^m$). Then ECS_N^* is higher (resp., lower) at $q_B = q_B^*$ than at $q_B = q_B^m$ unless f_N^* is decreased from $\overline{f}$ at $q_B = q_B^m$ to $\underline{f}$ at $q_B = q_B^*$.

First, consider the case when the non-deceptive counterfeit draws an insignificant amount of brand value from the brand-name product (i.e., $\beta < 1 - \{q_B^m - q_N^*(q_B^m)\}^2/(q_B^m)^2$). In this case, Proposition 1(a) shows that the brand-name company should set her product quality higher than q_B^m. This strategy not only improves the expected profit of the brand-name company (as compared to choosing $q_B = q_B^m$), but also decreases the expected profit of the non-deceptive counterfeiter (Proposition 1(b)). In this case, even though the improved quality of the brand-name product also improves the quality of the non-deceptive counterfeit, the difference in quality between two competing products becomes larger because the counterfeit steals only a small part of the brand value (i.e., low β). Consequently, the non-deceptive counterfeiter will lose its quality competition against the brand-name company. This result may explain how the shoe manufacturers mentioned in §1 successfully addressed their counterfeiting issues by improving the quality of their products (Qian 2008). Finally, Proposition 1(c) shows that, although this strategy improves the expected profit of the brand-name company and reduces the expected profit of the non-deceptive counterfeiter, it does *not always* benefit consumers. The reason is as follows. This strategy can lead the non-deceptive counterfeiter to lower his functional quality as well as his wholesale price in order to compete better against brand-name products with improved quality. Although this reduces the market share of non-deceptive counterfeits, those consumers who purchase non-deceptive counterfeits can suffer from lower quality, resulting a welfare loss. For example, Figure 1 illustrates that ECS_N^* falls when q_B is increased from $q_B^m = 3.37$ to $q_B^* = 3.4$.

Next, consider the case when the non-deceptive counterfeit draws a significant amount of brand value from the brand-name product (i.e., $\beta > 1 - \{q_B^m - q_N^*(q_B^m)\}^2/(q_B^m)^2$). In this case, Proposition 1(a) shows that it is not cost-effective for the brand-name company to improve her product quality because the non-deceptive counterfeiter gets a free ride on the improved quality of the brand-name product. While this strategy improves the expected profit of the brand-name company, it can also help the non-deceptive counterfeiter earn higher expected profit inadvertently (Proposition 1(b)), and make consumers suffer from the poor quality of the product (Proposition 1(c)). Therefore, in this case, the brand-name company may not use this strategy to combat the non-deceptive counterfeiter, and if she does, she must take special case to curb counterfeits in the market.

The following proposition shows how the brand-name company can combat the non-deceptive counterfeiter through her *pricing* strategy.

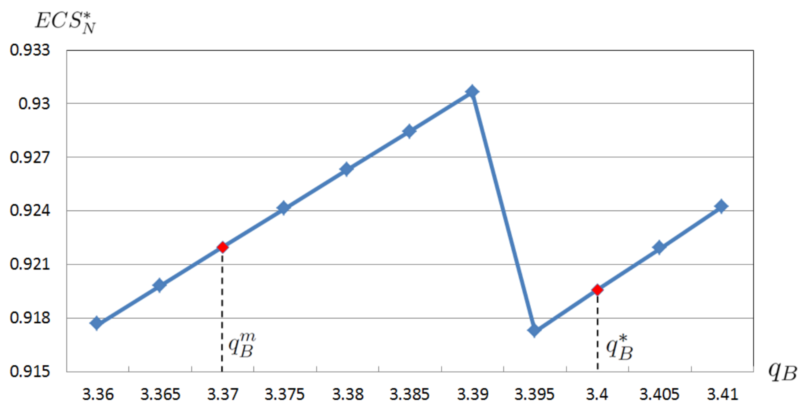

Figure 1: Expected consumer welfare as a function of q_B when non-deceptive counterfeits are in the market (Base parameters: $t = 0.01$, $p_B = 0.9$, $\beta = 0.1$, $\gamma = 0.58$, $c = 0.05$, $\underline{f} = 0.1$, and $\bar{f} = 2.5$)

Proposition 2 *(a) $p_B^* < p_B^m$ for all β.*
(b) π_N^* is lower at $p_B = p_B^*$ than at $p_B = p_B^m$ for all β.
(c) ECS_N^* is higher at $p_B = p_B^*$ than at $p_B = p_B^m$ unless f_N^* is decreased from $\overline{f}$ at $p_B = p_B^m$ to $\underline{f}$ at $p_B = p_B^*$ or $\frac{\partial f_N^*}{\partial p_B} > \kappa$ (where the expression of κ (> 0) is presented in the proof).

In contrast to the earlier quality strategy, Proposition 2(a) shows that for any β, it is always beneficial for the brand-name company to set her price p_B^* lower than p_B^m. This is because a lower price enables the brand-name company to compete better against non-deceptive counterfeits which are cheap alternatives of brand-name products. This strategy helps the brand-name company to gain more market share by inducing some consumers to switch from non-deceptive counterfeits to brand-name goods. As a result, this strategy also reduces the expected profit of the non-deceptive counterfeiter (Proposition 2(b)). We further find that the larger β is, the faster the expected profit of the non-deceptive counterfeiter will decrease. This is because the brand-name company relies more on price to compete with the non-deceptive counterfeiter when the quality levels of two products are not so distinguished due to the larger β. However, similar to the quality strategy, Proposition 2(c) shows that reducing p_B can hurt consumers by inducing the non-deceptive counterfeiter to reduce his quality level. This strategy has been used in practice; for example, the distributors of Hollywood films cut their DVD prices in Malaysia and Russia to combat rampant piracy (Whang 2001, Arvedlung 2004).

5.2 Deceptive Counterfeits

This subsection examines the brand-name company's anti-counterfeiting strategies against the *deceptive* counterfeiter. As we will show below, most effects of these strategies are monotonic when no proactive consumers exist in the market (i.e, $\lambda = 0$), whereas all effects of these strategies are *non-monotonic* when proactive consumers exist in the market (i.e., $\lambda > 0$). Thus, we first examine the former case analytically to establish monotonic results, and then conduct a numerical study for the latter case to show non-monotonicity. This approach will enable us to isolate the effect of λ, and explore dominant effects of anti-counterfeiting strategies when positive λ creates non-monotonic effects. Note that the results under $\lambda = 0$ also bear some practical relevance (asymptotically) because only a small fraction of consumers may be proactive in developed countries; for example, $\lambda = 0.04$ in the U.S. in our survey results shown in Table 1.

Let us first analyze the case when $\lambda = 0$. The following proposition shows, counter-intuitively, that by setting the *quality* level lower than the quality level with no counterfeiter in the market, the brand-name company can improve her expected profit, reduce the expected profit of the deceptive counterfeiter, and even improve expected consumer welfare.

Proposition 3 *Consider the market with $\lambda = 0$. In this market, the following results hold:*
(a) $q_B^{**} < q_B^m$.
(b) π_D^{**} is lower at $q_B = q_B^{**}$ than at $q_B = q_B^m$.
(c) ECS_D^{**} is higher at $q_B = q_B^{**}$ than at $q_B = q_B^m$ if $q_D^{**} < q_B - \frac{(1-p_B^2 q_B^{-2})\{1-(1-\gamma)s^{**}\}}{2(1-\gamma)}\{\frac{p_B^2}{q_B^3}s^{**} + \frac{1}{2}\left(1 - \frac{p_B^2}{q_B^2}\right)\frac{\partial s^{**}}{\partial q_B}\}^{-1}$ for $q_B \in [q_B^{**}, q_B^m]$.

Proposition 3(a) states that it is optimal for the brand-name company to set her quality q_B^{**} lower than q_B^m. Since consumers cannot distinguish deceptive counterfeits from brand-name products, this strategy reduces the perceived quality of any product in the market, and thus reduces the aggregate demand for both brand-name and counterfeit goods. However, the reduced aggregate demand discourages the licit distributor from taking the risk of selling deceptive counterfeits, hence resulting in a lower s^{**}. The result stated in Proposition 3(a) suggests that the latter (positive) effect dominates the former (negative) effect, so this strategy improves the expected profit of the brand-name company. This result highlights the importance of modeling the incentive of the licit distributor in this supply chain: Without the licit distributor, the positive effect of this strategy (i.e., lower s^{**}) would not exist and therefore the result opposite to Proposition 3(a) would be obtained. Since this strategy reduces both the aggregate demand and the proportion of deceptive counterfeits sold by the licit distributor, it will also reduce the expected profit of the deceptive counterfeiter (Proposition 3(b)). (More generally, we show in the proof that π_D^{**} is increasing in q_B for any λ.) Finally, contrary to our first intuition that lower quality will hurt consumers, Proposition 3(c) suggests that this strategy can improve consumer welfare. To understand this result, note that there are two opposing effects of having lower quality of brand-name products on consumer welfare: Consumers suffer from lower quality and fewer consumers buy products, but at the same time fewer consumers are deceived to buy low-quality counterfeits. Proposition 3(c) shows that when the quality of deceptive counterfeits is sufficiently low, the latter effect outweighs the former effect, benefiting consumers.

We next examine the effectiveness of the *pricing* strategy against the deceptive counterfeiter in the market with $\lambda = 0$.

Proposition 4 *Consider the market with $\lambda = 0$. In this market, the following results hold:*
(a) $p_B^{**} > p_B^m$.
(b) π_D^{**} can be higher or lower when $p_B = p_B^{**}$ than when $p_B = p_B^m$.
(c) ECS_D^{**} is higher at $p_B = p_B^{**}$ than at $p_B = p_B^m$ if $q_D^{**} < q_B - \frac{q_B - p_B}{1-\gamma}\{\frac{p_B}{q_B}s^{**} - \frac{1}{2}\left(1 + \frac{p_B}{q_B}\right)(q_B - p_B)\frac{\partial s^{**}}{\partial p_B}\}^{-1}$ for $p_B \in [p_B^m, p_B^{**}]$.

With no proactive consumers in the market, Proposition 4(a) states that the brand-name company can improve her expected profit by setting her price p_B^{**} higher than p_B^m (due to the reason similar to Proposition 3(a)). Unlike the earlier quality strategy, however, this pricing strategy has non-monotonic impact on the expected profit of the deceptive counterfeiter (Proposition 4(b)). To understand this result, note that there are two effects of raising p_B : (i) it reduces the aggregate demand for brand-name and counterfeit goods (i.e., $\frac{\partial}{\partial p_B}\left(1 - \frac{p_B}{q_B}\right) < 0$); and (ii) it increases the distributor's margin from selling deceptive counterfeits (i.e., $\frac{\partial}{\partial p_B}(p_B - w_D^{**}) = \frac{\partial}{\partial p_B}\sqrt{\frac{l p_B}{1 - \frac{p_B}{q_B}}} > 0$ from (8)). Because of the latter effect, the strategy of raising p_B does not always reduce the proportion s^{**} of deceptive counterfeits the licit distributor sells, nor does it always reduce the deceptive counterfeiter's market share m_D^{**} and her expected profit π_D^{**}. Therefore, in implementing this pricing strategy, a firm or the government should carefully consider these two counterbalancing effects of raising/reducing price. In practice, we observe both instances of raising or reducing prices: Newton et al. (2002) propose reducing drug prices to make counterfeiting less attractive by reducing the profit margins of fake drugs (i.e., opposite effect of (ii)), and Russia will raise vodka prices to put out of business makers of counterfeit alcohol (via effect (i)) although it will also affect licit companies (Reuters 2012). Finally, Proposition 4(c) suggests that this strategy can improve consumer welfare when the quality of deceptive counterfeits is sufficiently low. We can interpret this result similarly to Proposition 3(c).

Next, we analyze the case in which proactive consumers exist in the market (i.e., $\lambda > 0$). As we have mentioned at the beginning of this subsection, this additional factor causes all the effects of the anti-counterfeiting strategies to become non-monotonic. Specifically, the brand-name company's optimal quality q_B^{**} (resp., p_B^{**}), can be higher or lower than her quality q_B^m with no counterfeiter (resp., p_B^m); furthermore, the deceptive counterfeiter's expected profit π_D^{**} and the expected consumer welfare ECS_D^{**} are non-monotonic with a change of q_B or p_B. Because the closed-form expressions of s^{**}, w_D^{**} and f_D^{**} do not exist when $\lambda > 0$, no simple conditions can be derived analytically for monotonic results (see remarks on the proofs of Propositions 3 and 4 in online appendix A). For this reason, we conduct a numerical study to compare the results under $\lambda = 0$ with those under $\lambda > 0$, and explore dominant effects. The numerical experiments are conducted with the following settings: for $\lambda = 0$, 0.25 or 0.5, we constructed 1024 scenarios using the parameter values shown at the bottom of Table 3, so that they cover various possible scenarios and also satisfy positive s^{**} in equilibrium. We present a summary of the results in Table 3, which reads as follows: for example, when $\lambda = 0.5$, $q_B^{**} < q_B^m$ was observed in 97.3% of 1024 scenarios, and choosing q_B^{**} reduced π_D^{**} in 97.3% of 1024 scenarios and increased ECS_D^{**} in 5.3% as compared to choosing q_B^m.

Table 3. Effects of Quality and Pricing Strategies against Deceptive Counterfeits

	Effects of Choosing q_B^{**} vs. q_B^m			Effects of Choosing p_B^{**} vs. p_B^m		
	$q_B^{**} < q_B^m$	$\pi_D^{**} \downarrow$	$ECS_D^{**} \uparrow$	$p_B^{**} > p_B^m$	$\pi_D^{**} \downarrow$	$ECS_D^{**} \uparrow$
$\lambda = 0$	1	1	0.032	1	0.097	0.016
$\lambda = 0.25$	0.961	0.961	0.052	0.989	0.398	0.048
$\lambda = 0.5$	0.973	0.973	0.053	0.984	0.454	0.039

(Note) Each number in the table indicates a percent of scenarios for the corresponding effect. We used the following parameters: $t \in$\{0.005,0.01,0.015,0.02\}, $\beta \in$\{0.1,0.2,0.3,0.4\}, $\gamma \in$\{0.1,0.2,0.3,0.4\}, $l \in$\{0.005,0.01,0.015,0.02\}, $c \in$\{0.1,0.2,0.3,0.4\}, $\underline{f} = 0.1$, and $\overline{f} = (1-\beta) * q_B - 0.1$.

From Table 3, we can observe the following:

(1) The results obtained under $\lambda = 0$ continue to hold in most scenarios under $\lambda > 0$. However, in some scenarios, the brand-name company finds it optimal to set $q_B^{**} > q_B^m$ or $p_B^{**} < p_B^m$. We can explain this result as follows. First, recall from our discussions above that setting lower q_B^{**} or higher p_B^{**} reduces the aggregate demand for brand-name and counterfeit goods, and that the reduced aggregate demand discourages the licit distributor from taking the risk of selling counterfeits. Propositions 3(a) and 4(a) suggest that the latter (positive) effect always dominates the former (negative) effect when $\lambda = 0$. However, with proactive consumers in the market (i.e., $\lambda > 0$), the deceptive counterfeiter may improve his functional quality f_D^{**} in response to the reduced demand (see Lemma 2). This additional factor makes the licit distributor more willing to sell counterfeits, so that the positive effect does not always dominate the negative effect.

(2) In those scenarios where $q_B^{**} > q_B^m$, the strategy of setting higher q_B^{**} will increase the deceptive counterfeiter's expected profit π_D^{**} by making counterfeits flourish more in the market. This happens because the improved quality of the brand-name product results in an increase of the

aggregate demand of brand-name and counterfeit goods, which in turn incentivizes the licit distributor to procure more deceptive counterfeits. This may be the cause of the initial failure of the Scotch whisky company which improved her quality to combat deceptive counterfeits in the Thailand market (see §1). Also, from the table, we confirm that the expected profit of the deceptive counterfeiter is non-monotonic in p_B for any $\lambda \geq 0$, which can be explained similarly to Proposition 4(b).

(3) The expected consumer welfare ECS_D^{**} has increased in more scenarios in the market with $\lambda > 0$ than in the market with $\lambda = 0$. Similar to our explanation given in (1) above, this is because the counterfeiter may improve his functional quality f_D^{**} with proactive consumers. In general, for any $\lambda \in [0, 1]$, we show in online appendix A that if an anti-counterfeiting strategy improves the average product quality in the market, then it improves the expected consumer welfare.

(4) The number of scenarios in which π_D^{**} is decreased or ECS_D^{**} is increased is not necessarily monotonic in λ. For example, a change to q_B^{**} from q_B^m decreases π_D^{**} in all scenarios when $\lambda = 0$, in 96.1% of scenarios when $\lambda = 0.25$, and in 97.3% when $\lambda = 0.5$. This result indicates that anti-counterfeiting strategies are not necessarily more effective as more consumers are proactive. Similarly, we can show that more proactive consumers in the market does *not necessarily* benefit the brand-name company (i.e., π_B^{**} is non-monotonic in λ). The reason is as follows. Proactive consumers purchase products only when their expected utility is non-negative, considering the likelihood of receiving deceptive counterfeits unknowingly. As more consumers are proactive, therefore, a smaller number of consumers will purchase products. This reduced aggregate demand for products discourages the licit distributor from taking the risk of selling deceptive counterfeits. Thus, depending on which of the two effects (i.e., reduced aggregate demand and reduced s^{**}) dominates, the expected profit of the brand-name company as well as her market share may increase or decrease with λ.

5.3 Comparison: Non-Deceptive vs. Deceptive

We now compare the effect of each strategy against the *non-deceptive* counterfeiter with that against the *deceptive* counterfeiter. Using the results presented in §5.1 and §5.2, we summarize in Table 4 whether the brand-name company should choose higher/lower quality or price than the case with no counterfeiter in order to maximize her expected profit, and how such anti-counterfeiting strategies affect the expected profit of the counterfeiter and the expected consumer welfare. (If a dominant effect exists for a non-monotonic case, Table 4 reports only the dominant effect.)

Table 4. Effects of Anti-Counterfeiting Strategies: Non-Deceptive vs. Deceptive

Non-Deceptive Counterfeits			Deceptive Counterfeits		
Optimal Strategy	π_N^*	ECS_N^*	Optimal Strategy	π_D^{**}	ECS_D^{**}
$q_B^* > q_B^m$ (low β)	$\downarrow$	$\uparrow$	$q_B^{**} < q_B^m$	$\downarrow$	$\uparrow$ (low q_D^{**}) or $\downarrow$ (high q_D^{**})
$q_B^* < q_B^m$ (high β)	$\uparrow$	$\downarrow$			
$p_B^* < p_B^m$	$\downarrow$	$\uparrow$	$p_B^{**} > p_B^m$	$\updownarrow$	$\uparrow$ (low q_D^{**}) or $\downarrow$ (high q_D^{**})

From Table 4, we can draw the following insights:

(1) The optimal strategy of the brand-name company (that maximizes her expected profit) differs depending on whether she faces the non-deceptive or deceptive counterfeiter. For example, reducing price is optimal against the non-deceptive counterfeiter, whereas raising price is optimal against the deceptive counterfeiter.

(2) Even when the optimal strategy of the brand-name company is the same against both non-deceptive and deceptive counterfeiters, its impact on the counterfeiter's expected profit and the expected consumer welfare may not be the same. For example, when β is high, setting a lower quality level than the case with no counterfeiter improves the brand-name company's expected profit against either type of the counterfeiter. While this strategy is effective against the deceptive counterfeiter (i.e., reduces π_D^{**}), it does not work well against the non-deceptive counterfeiter (i.e., increases π_N^*). Moreover, its impact on the expected consumer welfare may not be the same across the two types of the counterfeiter, either.

(3) An ideal anti-counterfeiting strategy should improve the brand-name company's expected profit, reduce the counterfeiter's expected profit, and improve the expected consumer welfare. The pricing strategy is such an ideal strategy against the non-deceptive counterfeiter. For the other cases, a brand-name company or the government should carefully consider a trade-off among those three objectives in implementing an anti-counterfeiting strategy.

6 Anti-Counterfeiting Strategies: Marketing and Enforcement

In this section, we consider two other anti-counterfeiting strategies that are commonly used in practice. The first strategy we will consider is the marketing campaign that educates consumers about the adversity of counterfeit goods. For example, an electronic manufacturer may emphasize the fact that counterfeit electronics lack in safety features. This strategy helps reduce the brand value the counterfeit steals from the brand-name product, i.e., reduce β. The second strategy we will consider is the direct enforcement efforts to increase the chances to seize counterfeit products, γ. In executing these strategies, the brand-name company often collaborates with other organizations or the government. For example, French luxury goods association Comite Colbert launched a campaign (using playful slogans such as "real ladies don't like fake") in response to the threat of the counterfeit, and the French police raided the clandestine workshops making Hermes counterfeit accessories, of which the surveillance was part of an investigation into the international crime ring that robs many brands (Wellman 2012). Since the brand-name company does not have a full control of these parameters β and γ, we do not consider the brand-name company's optimal choices of these parameters; instead, we examine how reducing β or increasing γ will affect firms' expected profits and expected consumer welfare.

First, let us consider the market in which the brand-name company faces the *non-deceptive* counterfeiter. It is intuitive that both the marketing campaign and the enforcement strategy will improve the expected profit of the brand-name company and reduce the expected profit of the non-deceptive counterfeiter. However, we can show that both strategies hurt expected consumer welfare for the following reasons. The market campaign makes those consumers who purchase non-deceptive counterfeits enjoy the counterfeit brand less, resulting in a welfare loss. The enforcement strategy makes counterfeits less likely to reach the market, and hence it makes the non-deceptive counterfeiter more reluctant to invest in quality improvement. Therefore, consumers will suffer from less availability of non-deceptive counterfeits (which are cheaper substitutes for brand-name goods) as well as from their lower quality.

Next, we examine the effectiveness of two anti-counterfeiting strategies against the *deceptive* counterfeiter. The following proposition shows that the effectiveness of these strategies differs significantly from that against the non-deceptive counterfeiter.

Proposition 5 *For any given q_B and p_B,*
(a) (Marketing) When $\lambda = 0$, reducing β has no impact on π_B^{**} and π_D^{**}, whereas it reduces ECS_D^{**}. When $\lambda > 0$, reducing β decreases π_D^{**}, but it can increase or reduce π_B^{**} and ECS_D^{**}.
(b) (Enforcement) When $\lambda = 0$, increasing γ improves π_B^{**}, reduces π_D^{**}, and improves ECS_D^{**}. When $\lambda > 0$, increasing γ reduces π_D^{**}, but it can increase or reduce π_B^{**} and ECS_D^{**}.

Proposition 5(a) suggests that special care must be taken when implementing the marketing campaign against the deceptive counterfeiter. For the case when no consumers are proactive (i.e., $\lambda = 0$), the marketing campaign has no impact on the firms' expected profits because consumers do not take into account the possibility of receiving counterfeits unknowingly. This result is expected. On the other hand, proactive consumers correctly expect that they will derive less utilities when receiving deceptive counterfeits unknowingly. Thus, when $\lambda > 0$, the marketing campaign can reduce the expected profit of the deceptive counterfeiter by discouraging proactive consumers from purchasing products. However, it could backfire the brand-name company because proactive consumers reduce their consumption of brand-name products as well. For example, a large beverage company in Korea suffered from a sales drop of 15% after their counterfeiting problems were broadcasted in a TV program (Choi 2009). Finally, unlike the case when $\lambda = 0$, this strategy could improve expected consumer welfare when $\lambda > 0$ because a smaller number of proactive consumers purchase products and hence receive low-quality deceptive counterfeits.

Proposition 5(b) shows that when no proactive consumers exist in the market (i.e., $\lambda = 0$), the enforcement strategy works well against the deceptive counterfeiter. However, contrary to a common belief, this strategy may reduce the expected profit of the brand-name company and also hurt expected consumer welfare in the market where proactive consumers exist (i.e., $\lambda > 0$). This result can be explained as follows. Similar to the impact of this strategy on the non-deceptive counterfeiter (discussed above), by increasing the risk of counterfeiting, this strategy makes the deceptive counterfeiter reluctant to invest in quality improvement. While the lower quality of *non-deceptive* counterfeits helps the brand-name company regain its market share in quality competition, the lower quality of *deceptive* counterfeits reduces the perceived quality of products in the market with proactive consumers, hence reducing the aggregate demand for both brand-name goods and deceptive counterfeits. In this case, consumers also suffer from the lower quality of deceptive counterfeits although fewer consumers will receive deceptive counterfeits unknowingly. In online appendix B, we further study how different values of λ affect the effectiveness of these strategies.

7 Extension: Risk of Counterfeiting

In this section, we extend our base model to the case where the probability of counterfeits getting confiscated is a decreasing function of their functional quality. This is plausible in some situations because those consumers who have suffered due to the low quality of counterfeits can report them to the authorities, which may lead to the raid of counterfeit factories or distributors. For example, if the fake furniture mentioned in Barboza (2011) had functioned as well as its genuine furniture, a consumer might have not discovered that the furniture he/she has purchased is, in fact, counterfeit. Specifically, suppose that a counterfeiter will get caught by the authorities with the probability of $\gamma - \delta_1 f_i$ for $i = N$ or D, and that a licit distributor will get caught with the probability of $s - \delta_2 f_D$. We assume $\delta_1 > 0$ and $\delta_2 > 0$, so that the lower the quality of counterfeit goods, the higher the detection probabilities become.[3]

[3] We do not consider the case where the probability of the *illicit* distributor getting caught for selling *non-deceptive* counterfeits is decreasing with the quality of non-deceptive counterfeits. Such a case is unlikely in practice because consumers already know what they purchase. For example, a street vendor who sells \$10 fake watches is not more likely to get caught, as the quality of those watches gets worse. Furthermore, this probability does not affect our

When the *non-deceptive* counterfeiter exists in the market, it is easy to see that the price decisions of the illicit distributor and the counterfeiter in stages 3 and 2, respectively, are unchanged. In stage 1, the counterfeiter chooses his optimal functional quality f_N^* to maximize his expected profit, which is modified from (4) as follows:

$$\pi_N^*(f_N) = \frac{p_B^2(1-\gamma+\delta_1 f_N)(f_N+\beta q_B)}{8q_B\left\{(1-\beta)q_B - f_N\right\}} - t_N f_N^2.$$

Similar to Lemma 1, we can show that $f_N^* = \underline{f}$, $\overline{f}$ or $f_N^* \in (\underline{f}, \overline{f})$ that satisfies $\frac{\partial \pi_N^*}{\partial f_N}|_{f_N=f_N^*} = 0$, depending on the value of t_N and whether $\pi_N^*(\underline{f}) \geq \pi_N^*(\overline{f})$.

When the *deceptive* counterfeiter exists in the market, in stage 3, the licit distributor chooses its optimal fraction s^{**} of counterfeits by solving the following problem (which is modified from (6)):

$$\max_{s\in[0,1]} s\{1-s+\delta_2 f_D\}(p_B-w_D)\left\{1-\frac{\lambda p_B}{(1-s)q_B+sq_D}-\frac{(1-\lambda)p_B}{q_B}\right\}-(s-\delta_2 f_D)l. \qquad (12)$$

In stages 2 and 1, the counterfeiter decides w_D and f_D, respectively, to maximize his expected profit given by:

$$\pi_D(w_D,f_D) = w_D s^{**}(1-\gamma+\delta_1 f_D)\left\{1-\frac{\lambda p_B}{(1-s^{**})q_B+s^{**}(f_D+\beta q_B)}-\frac{(1-\lambda)p_B}{q_B}\right\}-t_D f_D^2.$$

When $\lambda = 0$, by following the procedure similar to that in the base model, we obtain the closed-form expressions of s^{**} and w^{**} as follows: $s^{**} = \frac{1+\delta_2 f_D}{2} - \frac{lq_B}{2(p_B-w_D)(q_B-p_B)}$ and $w_D^{**} = p_B - \sqrt{\frac{lp_B}{(1-\frac{p_B}{q_B})(1+\delta_2 f_D)}}$. In this case, unlike the base model (c.f. Lemma 2), $f_D^{**} > \underline{f}$ is possible even without proactive consumers. This is because high-quality counterfeits can induce the licit distributor to procure more counterfeits (i.e., increase s^{**}) by reducing the probability of the distributor getting caught. When $\lambda > 0$, similar to the base model, we can show the existence of s^{**}, w^{**} and f_D^{**}, but their closed-form expressions are not available.

Using the equilibrium analysis above, we show in the following corollary that the main results in the base model continue to hold in this extended model.

Corollary 1 *Suppose the probability of a counterfeiter getting caught is $\gamma - \delta_1 f_i$ for $i = N$ or D, and the probability of a licit distributor getting caught is $s - \delta_2 f_D$, where $\delta_1 > 0$ and $\delta_2 > 0$. Then:*
(a) Proposition 1 continues to hold.
(b) Propositions 2, 3 and 4 continue to hold except that the conditions in part (c) are different.
(c) Proposition 5 continues to hold except that increasing γ can increase or reduce π_B^{**} and ECS_D^{**} when $\lambda = 0$.

results due to our assumption that $l_N = 0$ (see §3).

Corollary 1 shows that a more general risk model in this section affects only the impact of enforcement strategy (that increases γ) on π_B^{**} and ECS_D^{**}. In the base model, Proposition 5 has shown that this strategy always improves π_B^{**} and ECS_D^{**} when $\lambda = 0$. However, Corollary 1(c) shows that this strategy can either increase or decrease π_B^{**} and ECS_D^{**} even when $\lambda = 0$. The intuition is as follows. In the base model, when $\lambda = 0$, the optimal functional quality f_D^{**} of the deceptive counterfeiter is always $\underline{f}$. However, as we have discussed above, $f_D^{**} > \underline{f}$ is possible in the extended model. In this case, as the investment for quality improvement becomes more risky with higher γ, the deceptive counterfeiter may find it optimal to reduce f_D^{**}. This in turn increases the risk of the licit distributor selling counterfeits (through $\delta_2 f_D$) as well as his own risk of getting caught (through $\delta_1 f_D$). As a result of these two opposing effects, we find that increasing γ can increase or decrease f_D^{**}. When f_D^{**} is increased, it will reduce the risk of the licit distributor selling deceptive counterfeits, hence increasing the fraction s^{**} of deceptive counterfeits; consequently, it could hurt the expected profit of the brand-name company, π_B^{**}. On the other hand, when f_D^{**} is decreased, consumers will suffer from the lower quality of deceptive counterfeits; thus, it could reduce ECS_D^{**}.

8 Concluding Remarks

Today counterfeit products are being produced and consumed in virtually all economies (OECD 2008). While easy-to-manufacture goods had dominated counterfeit supply until a decade ago, there has been an alarming expansion of product categories being infringed. As a result of outsourcing and offshoring, counterfeiters have easy access to modern technology and equipment, and they are capable of producing high-quality replicas. Consumers are not easily deceived by fake goods that are sold by vendors in open markets and unknown internet sites. These changing business conditions require industry and governments to enhance their understanding of the current and potential counterfeiters they may face and to develop strategies to limit their activities.

To aid the efforts of industry and governments to combat counterfeiting, we have developed a normative model of counterfeiting. Our model captures the recent changes in counterfeiting supply and demand that are not addressed in the previous literature. For example, the previous literature focuses on the pricing decision of a counterfeiter, assuming that the quality level of his goods is fixed, and he is capable of selling his goods, even deceptive ones, directly to consumers. In contrast, our model takes into account the strategic decisions of a counterfeiter regarding his price and functional quality; and the fundamental difference between non-deceptive and deceptive counterfeits in consumers' awareness, distribution channels, and penalty on illegal distribution. We have also considered the case when a fraction of consumers are proactive. Modeling these factors explicitly enables us to evaluate several anti-counterfeiting strategies against both types of counterfeiters, and to draw novel managerial insights.

Our analysis highlights that the strategies which are effective in combating the *non-deceptive* counterfeiter may not work well against the *deceptive* counterfeiter. Moreover, even if strategies help the brand-name company improve her expected profit, they may not be effective in limiting counterfeit activities, and they can even hurt consumers. For example:

- The strategy of improving the quality of brand-name products is effective in combating the *non-deceptive* counterfeiter only when the non-deceptive counterfeit steals an insignificant amount of brand value. This strategy may not be used in combating the non-deceptive counterfeiter in other situations or in combating the *deceptive* counterfeiter.
- The strategy of reducing the price of brand-name products is an ideal strategy against the *non-deceptive* counterfeiter. In contrast, when facing the *deceptive* counterfeiter, it can hurt the brand-name company's profit as well as consumer welfare, and also benefit the deceptive counterfeiter inadvertently.
- The marketing campaign and the enforcement strategy are effective in combating the *non-deceptive* counterfeiter, but they may not benefit the brand-name company or consumers when consumers are proactive toward *deceptive* counterfeits.

Therefore, industries and governments should understand the type of potential counterfeiters and the characteristics of consumers in order to design effective strategies to combat counterfeits. Without such understanding, anti-counterfeiting strategies could be ineffective and hurt consumer welfare.

Although our model captures the salient features of counterfeiting, we make several assumptions to maintain tractability. First, we do not consider the effect of positive or negative externality of counterfeits on brand-name products. For some product categories, counterfeits help to increase the size of user base of brand-name products, which refers to positive externality. A typical example is software piracy (Conner and Rumelt 1991). The negative externality of counterfeits refers to the negative impact of counterfeits on the value of a brand. More counterfeits in the market, less prestigious the brand becomes. Second, we assume that consumers are risk-neutral. In some situations, consumers show risk-prone or irrational behavior. For example, fraudsters use their phony pharmaceutical websites to take advantage of the recent swine-flu fears. Some consumers who are anxious for their children take risks of buying fake vaccines and bogus remedies from unknown websites (Taylor 2009). Behavioral research would help enrich our understanding of the risk attitudes of consumers. Third, our model does not capture the details of specific anti-counterfeiting technologies; e.g., technologies to authenticate products such as NanoInk (http://www.nanoink.net) and technologies to track and trace the movement of products through supply chains such as RFID. Yet, their broad use and success has been limited by a variety of factors, including the ability of counterfeiters to adopt or copy the technologies (OECD 2008). Our current model captures the role of these technologies to some degree: the former type of technologies is captured by the marginal cost of developing functional quality of a counterfeit product (i.e., with such technologies installed, a counterfeiter needs to spend more effort to copy authentic goods) and the latter type of technologies is captured by seizure rate (i.e., with RFID installed, the likelihood of seizing counterfeits increases). More detailed cost-benefit analysis of these technologies in specific industrial settings would be interesting future research.

References

Ahmadi, R., B. R. Yang. 2000. Parallel imports: challenges from unauthorized distribution channels. *Marketing Science* **19**(3) 279-294.

Arvedlung, E. 2004. Hollywood competes with the street in Russia; To combat rampant DVD piracy, U.S. film companies cut prices. *New York Times.* April 7, 2012.

Bass, D. 2010. Microsoft crosses swords with pirates. *Bloomberg Businessweek* July 26 – August 1, 2010.

Balfour, F. 2005. Fakes! *Business Week*, February 7, 2005.

Barboza, D. 2009. In China, knockoff cellphones are a hit. *New York Times*, April 27, 2009.

Barboza, D. 2011. Chinese upset over counterfeit furniture. *New York Times*, July 18, 2011.

Bendoly, E., R. Croson, P. Goncalves, K. Schultz. 2010. Bodies of knowledge for research in behavioral operations. *Production and Operations Management* **19**(4) 434-452.

Bennett, S. 2010. Pfizer: civil suits for drug counterfeiters. *Bloomberg Businessweek*, July 8, 2010.

Business Week. 2005. The counterfeit catastrophe. February 7, 2005.

Cachon, G. P., R. Swinney. 2009. Purchasing, pricing, and quick response in the presence of strategic consumers. *Management Science* **55**(3) 497–511.

Choi, J. 2009. Angry Vita500. *Joins*, May 8, 2009.

Conner, K., R. H. Rumelt. 1991. Software piracy: an analysis of protection strategies. *Management Science* **37**(2) 125-139.

Eisend, M., Schuchert-Guler P. 2006. Explaining counterfeit purchases: a review and preview. *Academy of Marketing Science Review.* Available at http://www.amsreview.org/articles/eisend12-2006.pdf.

Goldsmith, K., O. Amir. 2010. Can uncertainty improve promotions? *Journal of Marketing Research* **47**(6) 1070-1077.

Green, R. T., T. Smith. 2002. Executive insights: countering brand counterfeiters. *Journal of International Marketing* **10**(4) 89-106.

Grossman, G. M., C. Shapiro. 1988a. Counterfeit-product trade. *American Economic Review* **78**(1) 59-75.

Grossman, G. M., C. Shapiro. 1988b. Foreign counterfeiting of status goods. *Quarterly Journal of Economics* **103**(1) 79-100.

Hu, M., M. Pavlin, M. Shi. 2012. When gray markets have silver linings: all-unit discounts, gray markets and channel management. *Manufacturing & Service Operations Management,* Forthcoming.

Jacobs, L., A. C. Samli, T. Jedlik. 2001. The nightmare of international product piracy - exploring defensive strategies. *Industrial Marketing Management* **30**(6) 499-509.

Lee, J., S. Balakrishnan, J. Cho, S. Jeon, J. Kim. 2011. Detection of adulterated gasoline using colorimetric organic microfibers. *J. Mater. Chem* **21** 2648-2655.

Lilien, G. L., P. Kotler, K. S. Moorthy. 1992. *Marketing Models.* Prentice Hall, New Jersey.

Liu, K., J. A. Li, Y. Wu, K. K. Lai. 2005. Analysis of monitoring and limiting of commercial cheating: a newsvendor model. *Journal of the Operational Research Society* **56**(7) 844-854.

McKinley, J. 2010. Man sentenced for selling phony goods to military. *New York Times*, May, 6, 2010.

Naumann, S. 2009. Top 9 Shanghai markets. http://gochina.about.com/od/shoppinginshanghai/tp/SH_Markets.htm. Retrieved on November 21,2009.

Newton, P. N., N. J. White, J. A. Rozendaal, M. D. Green. 2002. Murder by fake drugs. *BMJ* **324** 800-801.

OECD (Organisation for Economic Co-operation and Development). 2008. *The Economic Impact of Counterfeiting and Piracy.* OECD Publishing.

OECD. 2009. *Magnitude of counterfeiting and piracy of tangible products - an update.* Available at http://www.oecd.org/dataoecd/57/27/44088872.pdf.

Olsen, J. E., K. L. Granzin. 1992. Gaining retailers' assistance in fighting counterfeiting: conceptualization and empirical test of a helping model. *Journal of Retailing* **68**(1) 90–109.

Qian, Y. 2008. Impacts of entry by counterfeiters. *Quarterly Journal of Economics* **123**(4) 1577-1609.

Reuters. 2012. Russia to hike minimum vodka prices by a third. December 26, 2012.

Rockoff, J. D., C. Weaver. 2012. Fake cancer drug found in the U.S. *Wall Street Journal.* February 15, 2012.

Rockoff, J. D., J. Whalen, C. Weaver. 2012. Fakes infiltrate injectable drugs. *Wall Street Journal.* February 16, 2012.

Scandizzo, S. 2001. Counterfeit goods and income inequality. Working paper, IMF.

Schmidle, N. 2010. Inside the knockoff-tennis-shoe factory. *New York Times*, August 19, 2010.

Solomon, S. E. 2009. On the trail of counterfeit drugs. *Business Week*, August 28, 2009.

Staake, T., E. Fleisch. 2008. *Countering Counterfeit Trade.* Springer-Verlag, Berlin, Germany.

Su, X., F. Zhang. 2008. Strategic customer behavior, commitment, and supply chain performance. *Management Science* **54**(10) 1759–1773.

Sun, J., L. Debo, S. Kekre, J. Xie. 2010. Component-based technology transfer: balancing cost saving and imitation risk. *Management Science* **56**(3) 536-552.

Taylor, M. 2009. Cybercrime capitalizes on swine-flu fears. *Wall Street Journal*. November 18, 2009.

Weaver, C., J. Whalen, B. Faucon. 2012. Drug distributor is tied to imports of fake Avastin. *Wall Street Journal*. March 7, 2012.

Weaver, C. 2012. U.S. fake-drug probe puts spotlight on role of doctors. *Wall Street Journal*. November 20, 2012.

Wellman, V. 2012. Multi-million-dollar counterfeit Birkin ring revealed to be run by Hermes STAFF as police bust global knock-off operation. *Mail Online*. June 15, 2012.

Whang, Y. 2001. Prices cut to combat piracy. *Hollywood Reporter – International Edition* **367**(15) 93.

Xu, Y., H. Gurnani, R. Desiraju. 2010. Strategic supply chain structure design for a proprietary component manufacturer. *Production and Operations Management* **19**(4) 371-389.

Zhang, J., L. J. Hong, R. Q. Zhang. 2012. Fighting strategies in a market with counterfeits. *Annals of Operations Research* **192**(1) 49-66.

Online Appendix

Appendix A. Proofs of Analytical Results

We use (A1) and (A2) to indicate the following assumptions we have made earlier: (A1) $q_B > q_N = f_N + \beta q_B$ and $q_B > q_D = f_D + \beta q_B$; (A2) $1 - \frac{p_B}{q_B} > 0$ and $1 - \frac{p_B - p_N}{q_B - q_N} > 0$ so that $m_B > 0$.

Proof of Lemma 1: From (4), we obtain $\frac{\partial^2 \pi_N^*}{\partial f_N^2} = \frac{(1-\gamma)p_B^2}{4\{(1-\beta)q_B - f_N\}^3} - 2t_N$, which is positive if $t_N < \frac{(1-\gamma)p_B^2}{8\{(1-\beta)q_B - f_N\}^3}$. Thus, if $t_N < \frac{(1-\gamma)p_B^2}{8\{(1-\beta)q_B - \underline{f}\}^3}$, π_N^* is convex in $f_N \in [\underline{f}, \overline{f}]$, so $f_N^* = \underline{f}$ when $\pi_N^*(\underline{f}) \geq \pi_N^*(\overline{f})$. Otherwise, f_N^* can be $\overline{f}$ or $f_N^* \in (\underline{f}, \overline{f})$ that satisfies the first order condition $\frac{\partial \pi_N^*}{\partial f_N}|_{f_N = f_N^*} = 0$. □

Remark A sufficient condition for $f_N^* > \underline{f}$ is $t_N < \frac{(1-\gamma)p_B^2}{16\underline{f}\{(1-\beta)q_B - \underline{f}\}^2}$, which can be obtained from $\frac{\partial \pi_N^*}{\partial f_N}|_{f_N = \underline{f}} = \frac{(1-\gamma)p_B^2}{8\{(1-\beta)q_B - \underline{f}\}^2} - 2t_N\underline{f} > 0$.

Proof of Lemma 2: When $\lambda = 0$, from (8), $\frac{\partial \pi_D^{**}}{\partial f_D} = -2t_D f_D < 0$, so $f_D^{**} = \underline{f}$. When $\lambda > 0$, we next show that $f_D^{**} = \underline{f}$ if $t_D \geq \bar{t}_D$. For any $f_D \in (\underline{f}, \overline{f}]$, $\pi_D^{**}(w_D^{**}(f_D), \underline{f}) \geq \pi_D^{**}(w_D^{**}(f_D), f_D)$ if $t_D \geq (1-\gamma)w_D^{**}(f_D)\{m_D^{**}(w_D^{**}(f_D), f_D) - m_D^{**}(w_D^{**}(f_D), \underline{f})\}(f_D^2 - \underline{f}^2)^{-1}$. Suppose $t_D \geq \bar{t}_D \equiv \max_{f_D \in (\underline{f}, \overline{f}]} (1-\gamma)w_D^{**}(f_D)\{m_D^{**}(w_D^{**}(f_D), f_D) - m_D^{**}(w_D^{**}(f_D), \underline{f})\}(f_D^2 - \underline{f}^2)^{-1}$. Then, for any $f_D \in (\underline{f}, \overline{f}]$,

$\pi_D^{**}(w_D^{**}(\underline{f}), \underline{f}) \geq \pi_D^{**}(w_D^{**}(f_D), \underline{f}) \geq \pi_D^{**}(w_D^{**}(f_D), f_D)$, where the first inequality is due to the optimality of $w_D^{**}(\underline{f})$ given $\underline{f}$, and the second inequality follows from $t_D \geq \bar{t}_D$. Therefore, $f_D^{**} = \underline{f}$.

In the rest of the proof, we show $\bar{t}_D > 0$ in two steps: we first show that s^{**} is increasing in f_D for given w_D, and then show that the market share of the deceptive counterfeiter, $m_D^{**} = s^{**}\left\{1 - \bar{\theta}(s^{**})\right\}$, is increasing in f_D for any given w_D. Then from the definition of $\bar{t}_D$, $\bar{t}_D > 0$. Let π_{LD} denote the expected profit of the licit distributor given in (6). Then

$\frac{\partial \pi_{LD}}{\partial s} = (1-2s)(p_B - w_D)\left\{1 - \bar{\theta}\right\} - s(1-s)(p_B - w_D)\frac{\partial \bar{\theta}}{\partial s} - l$, and

$\frac{\partial^2 \pi_{LD}}{\partial s \partial f_D} = (2-3s)(p_B - w_D)\frac{\lambda p_B s}{\{(1-s)q_B + s(f_D + \beta q_B)\}^2} + 2s(1-s)(p_B - w_D)\frac{\lambda p_B s(q_B - f_D - \beta q_B)}{\{(1-s)q_B + s(f_D + \beta q_B)\}^3}$,

where the first term is positive because we know from §4.2 that $s^{**} < 0.5$ and $w_D \leq p_B$, and the second term is also positive according to (A1). Therefore, $\frac{\partial \pi_{LD}}{\partial s}$ is increasing in f_D. Since s^{**} satisfies $\frac{\partial \pi_{LD}}{\partial s}|_{s=s^{**}} = 0$ due to the concavity of π_{LD} with respect to s, s^{**} is increasing in f_D.

Next, we show that m_D^{**} increases as f_D increases from f_{DL} to f_{DH} for given w_D. Suppose this does not hold. Then, π_{LD} satisfies the following:

$$\begin{aligned}
&\pi_{LD}(s^{**}(f_{DH}), f_{DH}) = s^{**}(f_{DH})(1 - s^{**}(f_{DH}))(p_B - w_D)\left\{1 - \bar{\theta}(s^{**}(f_{DH}), f_{DH})\right\} - s^{**}(f_{DH})l \\
&\leq s^{**}(f_{DL})(1 - s^{**}(f_{DH}))(p_B - w_D)\left\{1 - \bar{\theta}(s^{**}(f_{DL}), f_{DL})\right\} - s^{**}(f_{DH})l \\
&< s^{**}(f_{DL})(1 - s^{**}(f_{DH}))(p_B - w_D)\left\{1 - \bar{\theta}(s^{**}(f_{DL}), f_{DH})\right\} - s^{**}(f_{DH})l \\
&< s^{**}(f_{DL})(1 - s^{**}(f_{DL}))(p_B - w_D)\left\{1 - \bar{\theta}(s^{**}(f_{DL}), f_{DH})\right\} - s^{**}(f_{DL})l = \pi_{LD}(s^{**}(f_{DL}), f_{DH}),
\end{aligned}$$

where the first inequality follows from our premise, the second inequality follows from $\frac{\partial \bar{\theta}}{\partial f_D} = -\frac{\lambda p_B s}{\{(1-s)q_B + s(f_D + \beta q_B)\}^2} < 0$ for fixed s, and the last inequality follows from $\frac{\partial s^{**}}{\partial f_D} > 0$. However, this contradicts the condition that $s^{**}(f_{DH})$ maximizes the licit distributor's profit π_{LD} given f_{DH}. Therefore, m_D^{**} is increasing in f_D for given w_D, and $\bar{t}_D > 0$. $\square$

Remark A sufficient condition for $f_D^{**} > \underline{f}$ is $t_D < \underline{t}_D \equiv \max_{f_D \in (\underline{f}, \bar{f}]} (1-\gamma) w_D^{**}(\underline{f})\{m_D^{**}(w_D^{**}(\underline{f}), f_D) - m_D^{**}(w_D^{**}(\underline{f}), \underline{f})\}(f_D^2 - \underline{f}^2)^{-1}$. We show this by contradiction. Suppose $f_D^{**} = \underline{f}$ and define $f_{\max} = \arg\max_{f_D \in (\underline{f}, \bar{f}]} (1-\gamma) w_D^{**}(\underline{f})\{m_D^{**}(w_D^{**}(\underline{f}), f_D) - m_D^{**}(w_D^{**}(\underline{f}), \underline{f})\}(f_D^2 - \underline{f}^2)^{-1}$. Then $\pi_D^{**}(w_D^{**}(f_{\max}), f_{\max}) \geq \pi_D^{**}(w_D^{**}(\underline{f}), f_{\max}) > \pi_D^{**}(w_D^{**}(\underline{f}), \underline{f})$, where the first inequality is due to the optimality of $w_D^{**}(f_{\max})$ given $f_{\max}$, and the second inequality follows from $t_D < \underline{t}_D$. However, this contradicts our premise that $f_D^{**} = \underline{f}$. Therefore, $f_D^{**} > \underline{f}$ if $t_D < \underline{t}_D$.

Proof of Proposition 1: (a) The proof proceeds as follows: We first obtain q_B^m and q_B^*, and then derive the condition for $q_B^* > q_B^m$. When there is no counterfeiter, the expected profit of the brand-name company is given as follows:

$$\pi_B^m = (p_B - c)\left(1 - \frac{p_B}{q_B}\right) - t_B q_B^2, \tag{13}$$

where c (> 0) is the marginal cost of the brand-name product. From (13), $\frac{\partial^2 \pi_B^m}{\partial q_B^2} = -\frac{2(p_B - c)p_B}{q_B^3} -$

$2t_B < 0$, so we obtain $q_B^m = \frac{(p_B-c)p_B}{2t_B q_B^{m2}}$ from the first order condition. When the non-deceptive counterfeiter exists in the market, we obtain π_B^* after substituting p_N^* and w_N^* into m_B in (1) as follows:

$$\pi_B^* = (p_B - c)m_B - t_B q_B^2 = (p_B - c)\left\{1 - \frac{(1-\gamma)p_B}{4\{(1-\beta)q_B - f_N^*\}} - \frac{(3+\gamma)p_B}{4q_B}\right\} - t_B q_B^2. \quad (14)$$

From (14), when $f_N^* = \underline{f}$ or $\overline{f}$, $\frac{\partial^2 \pi_B^*}{\partial q_B^2} = -\frac{p_B-c}{2}\left(\frac{p_B(1-\gamma)(1-\beta)^2}{(q_B-q_N)^3} + \frac{p_B(3+\gamma)}{q_B^3}\right) - 2t_B < 0$ due to (A1). In this case, from the first order condition of (14), $q_B^* = \frac{p_B-c}{2t_B}\left\{\frac{(1-\gamma)(1-\beta)p_B}{4\{(1-\beta)q_B^* - f_N^*(q_B^*)\}^2} + \frac{(3+\gamma)p_B}{4q_B^{*2}}\right\}$.

We next show by contradiction that $q_B^* > q_B^m$ when $\beta < 1 - (\frac{q_B^m - q_N^*(q_B^m)}{q_B^m})^2$. Suppose $\beta < 1 - (\frac{q_B^m - q_N^*(q_B^m)}{q_B^m})^2$ and $q_B^* \leq q_B^m$. For $f_{NH} > f_{NL}$, from (4), we obtain $\frac{\partial \pi_N^*(f_{NH})}{\partial q_B} - \frac{\partial \pi_N^*(f_{NL})}{\partial q_B} = \frac{(\gamma-1)p_B^2(f_{NH}-f_{NL})(1-\beta)\{(1-\beta)q_B-(f_{NH}+f_{NL})/2\}}{4\{(1-\beta)q_B-f_{NH}\}^2\{(1-\beta)q_B-f_{NL}\}^2} < 0$ due to (A1), so f_N^* is decreasing in q_B. Then $q_B^* = \frac{p_B-c}{2t_B}\left\{\frac{(1-\gamma)(1-\beta)p_B}{4\{(1-\beta)q_B^* - f_N^*(q_B^*)\}^2} + \frac{(3+\gamma)p_B}{4q_B^{*2}}\right\} \geq \frac{p_B-c}{2t_B}\{\frac{(1-\gamma)(1-\beta)p_B}{4\{(1-\beta)q_B^m - f_N^*(q_B^m)\}^2} + \frac{(3+\gamma)p_B}{4q_B^{m2}}\} > \frac{(p_B-c)p_B}{2t_B q_B^{m2}} = q_B^m$, where the first inequality follows from $q_B^* \leq q_B^m$ and $f_N^*(q_B^*) \geq f_N^*(q_B^m)$, and the second inequality follows from $\beta < 1 - (\frac{q_B^m - q_N^*(q_B^m)}{q_B^m})^2$. Thus, there is a contradiction, so $q_B^* > q_B^m$ when $\beta < 1 - (\frac{q_B^m - q_N^*(q_B^m)}{q_B^m})^2$. The case in which $\beta \geq 1 - (\frac{q_B^m - q_N^*(q_B^m)}{q_B^m})^2$ can be shown similarly and is hence omitted.

(b) To establish the result in the proposition, it suffices to show that π_N^* is decreasing in q_B. The proof proceeds in two steps: We first show that π_N^* decreases in q_B for any given f_N, and then show that this result holds even when f_N^* changes with q_B. First, from (4), we obtain $\frac{\partial \pi_N^*}{\partial q_B} = \frac{(1-\gamma)p_B^2\left\{\beta q_B^2(\beta-1)+2q_B f_N(\beta-1)+f_N^2\right\}}{4q_B^2\{(1-\beta)q_B-f_N\}^2}$, which is negative by (A1) for any given f_N. Next, we consider the case in which f_N^* changes from f_{N1} to f_{N2} when q_B is increased from q_{BL} to q_{BH}. In this case, $\pi_N^*(f_{N1}, q_{BL}) \geq \pi_N^*(f_{N2}, q_{BL}) > \pi_N^*(f_{N2}, q_{BH})$, where the first inequality follows from $f_N^*(q_{BL}) = f_{N1}$ and the second inequality is due to $\frac{\partial \pi_N^*}{\partial q_B} < 0 \; \forall f_N$.

(c) We first prove that ECS_N^* is increasing in q_B for given f_N, and then prove that ECS_N^* decreases when f_N^* is decreased from f_{NH} to f_{NL} for any given q_B.

To prove that ECS_N^* is increasing in q_B, it suffices to show that $\frac{\partial CS_N}{\partial q_B} > 0$ for any given f_N because $\frac{\partial CS_B}{\partial q_B} > 0$ from the definition of CS_B. Now suppose that q_B is increased from q_{BL} to q_{BH}. Then q_N is also increased from q_{NL} to q_{NH} given f_N; $\widehat{\theta}$ is decreased from $\widehat{\theta}_L$ to $\widehat{\theta}_H$; and $\widetilde{\theta}$ is decreased from $\widetilde{\theta}_L$ to $\widetilde{\theta}_H$. Using $p_N^* = \frac{3p_B q_N}{4q_B}$, we can rewrite (9) and find $CS_N(q_{BH}) > \int_{\widehat{\theta}_L}^{\widetilde{\theta}_H}\left(\theta - \frac{3p_B}{4q_{BH}}\right)q_{NH}d\theta + \int_{\widetilde{\theta}_H}^{1}(\theta q_{BH} - p_B)d\theta > \int_{\widehat{\theta}_L}^{\widetilde{\theta}_H}\left(\theta - \frac{3p_B}{4q_{BH}}\right)q_{NH}d\theta + \int_{\widetilde{\theta}_H}^{\widetilde{\theta}_L}\left(\theta - \frac{3p_B}{4q_{BH}}\right)q_{NH}d\theta + \int_{\widetilde{\theta}_L}^{1}(\theta q_{BH} - p_B)d\theta > CS_N(q_{BL})$. The first inequality holds because $\widehat{\theta}_L > \widehat{\theta}_H$ and $\left(\theta - \frac{3p_B}{4q_{BH}}\right)q_{NH} > 0$ for $\theta \in (\widehat{\theta}_H, \widehat{\theta}_L)$. The second inequality holds because $\theta q_{BH} - p_B > \theta q_{NH} - p_N^*$ for $\theta \in (\widetilde{\theta}_H, \widetilde{\theta}_L)$. The third inequality follows from the fact that $q_{BH} > q_{BL}$ and $q_{NH} > q_{NL}$.

Next, suppose f_N^* is decreased from f_{NH} to f_{NL} for fixed q_B. Then $\widehat{\theta}$ remains the same, whereas $\widetilde{\theta}$ is decreased from $\widetilde{\theta}' \equiv \frac{p_B}{4\{(1-\beta)q_B - f_{NH}\}} + \frac{3p_B}{4q_B}$ to $\widetilde{\theta}'' \equiv \frac{p_B}{4\{(1-\beta)q_B - f_{NL}\}} + \frac{3p_B}{4q_B}$. Then,

$$ECS_N^*(f_{NH}) = (1-\gamma)\left\{\int_{\widehat{\theta}}^{\widetilde{\theta}'}\left(\theta - \frac{3p_B}{4q_B}\right)(f_{NH}+\beta q_B)d\theta + \int_{\widetilde{\theta}'}^{1}(\theta q_B - p_B)d\theta\right\} + \gamma\int_{\frac{p_B}{q_B}}^{1}(\theta q_B - p_B)d\theta$$
$$> (1-\gamma)\left\{\int_{\widehat{\theta}}^{\widetilde{\theta}''}\left(\theta - \frac{3p_B}{4q_B}\right)(f_{NH}+\beta q_B)d\theta + \int_{\widetilde{\theta}''}^{\widetilde{\theta}'}(\theta q_B - p_B)d\theta + \int_{\widetilde{\theta}'}^{1}(\theta q_B - p_B)d\theta\right\} + \gamma\int_{\frac{p_B}{q_B}}^{1}(\theta q_B - p_B)d\theta$$
$$> (1-\gamma)\left\{\int_{\widehat{\theta}}^{\widetilde{\theta}''}\left(\theta - \frac{3p_B}{4q_B}\right)(f_{NL}+\beta q_B)d\theta + \int_{\widetilde{\theta}''}^{1}(\theta q_B - p_B)d\theta\right\} + \gamma\int_{\frac{p_B}{q_B}}^{1}(\theta q_B - p_B)d\theta = ECS_N^*(f_{NL}).$$

In the above, the first inequality holds because $\left(\theta - \frac{3p_B}{4q_B}\right)(f_{NH}+\beta q_B) > \theta q_B - p_B$ for $\theta \in (\widetilde{\theta}'', \widetilde{\theta}')$, and the second inequality follows from $f_{NH} > f_{NL}$. $\square$

Remark When $f_N^* \in (\underline{f}, \overline{f})$, assuming $\frac{\partial^2 f_N^*}{\partial q_B^2} \geq 0$, we can still obtain $\frac{\partial^2 \pi_B^*}{\partial q_B^2} < 0$. From the first order condition of (14), $q_B^* = \frac{p_B - c}{2t_B}\{\frac{(1-\gamma)(1-\beta-\partial f_N^*/\partial q_B)p_B}{4\{(1-\beta)q_B^* - f_N^*(q_B^*)\}^2} + \frac{(3+\gamma)p_B}{4q_B^{*2}}\}$. The condition for $q_B^* > q_B^m$ then becomes $\frac{(1-\gamma)(1-\beta-\partial f_N^*/\partial q_B)}{4\{(1-\beta)q_B^* - f_N^*(q_B^*)\}^2} + \frac{3+\gamma}{4q_B^{*2}} - \frac{1}{q_B^{m2}} > 0$. Unfortunately, this condition cannot be simplified further to the form like $\beta < 1 - (\frac{q_B^m - q_N^*(q_B^m)}{q_B^m})^2$ because the closed-form expressions for f_N^* and $\frac{\partial f_N^*}{\partial q_B}$ are not available. The proof for (b) does not require $f_N^* = \underline{f}$ or $\overline{f}$, so it also holds for $f_N^* \in (\underline{f}, \overline{f})$. For (c), suppose $q_B^* > q_B^m$. From (11), $\frac{\partial ECS_N^*}{\partial q_B} = [q_B^2(1-\beta)\{16(1-\beta)q_B^2 + p_B^2(\beta(15+\gamma)-16)\} + f_N^*\{p_B^2(15+\gamma) - 16q_B^2\}\{2(1-\beta)q_B - f_N^*\} + p_B^2 q_B^2(1-\gamma)\partial f_N^*/\partial q_B]\left\{(1-\beta)q_B - f_N^*\right\}^{-2}$. Then ECS_N^* is decreasing in $q_B \in [q_B^m, q_B^*]$ so that ECS_N^* is lower at q_B^* than at q_B^m if $\frac{\partial f_N^*}{\partial q_B} < -\frac{q_B^2(1-\beta)\{16(1-\beta)q_B^2 + p_B^2(\beta(15+\gamma)-16)\} + f_N^*\{p_B^2(15+\gamma)-16q_B^2\}\{2(1-\beta)q_B - f_N^*\}}{p_B^2 q_B^2(1-\gamma)\left\{(1-\beta)q_B - f_N^*\right\}^2}$.

Proof of Proposition 2: (a) From (13), $\frac{\partial^2 \pi_B^m}{\partial p_B^2} = -\frac{2}{q_B} < 0$, so we obtain $p_B^m = \frac{q_B + c}{2}$ from the first order condition. Next, consider the market in which the non-deceptive counterfeiter exists. When $f_N^* = \underline{f}$ or $\overline{f}$, from (14), $\frac{\partial^2 \pi_B^*}{\partial p_B^2} = -\frac{1-\gamma}{2(q_B - q_N)} - \frac{3+\gamma}{2q_B} < 0$ and $p_B^* = \frac{q_B q_D^*(1-\gamma)}{2(-4q_B + 3q_D^* + \gamma q_D^*)} + \frac{q_B + c}{2}$; in this case, $p_B^* < p_B^m$ due to (A1). When $f_N^* \in (\underline{f}, \overline{f})$, we show $\frac{\partial \pi_B^*}{\partial p_B}|_{p_B = \frac{q_B + c}{2}} < 0$, which then results in $p_B^* < p_B^m$. From (14), we obtain $\frac{\partial \pi_B^*}{\partial p_B}|_{p_B = \frac{q_B + c}{2}} = \frac{(1-\gamma)\{4f_N^{*2} + 4(2\beta - 1)f_N^* q_B + 4(\beta - 1)\beta q_B^2 + (c - q_B)(c + q_B)\partial f_N^*/\partial p_B\}}{16\left\{(1-\beta)q_B - f_N^*\right\}^2}$, which is negative because: $4f_N^{*2} - 4(1-2\beta)f_N^* q_B - 4(1-\beta)\beta q_B^2 - \frac{\partial f_N^*}{\partial p_B}(q_B - c)(q_B + c) < 4f_N^{*2} - 4(f_N^* - \beta q_B)f_N^* - 4f_N^*\beta q_B - \frac{\partial f_N^*}{\partial p_B}(q_B - c)(q_B + c) = -\frac{\partial f_N^*}{\partial p_B}(q_B - c)(q_B + c) \leq 0$, where the first inequality is based on (A1) and the second inequality holds because $\frac{\partial f_N^*}{\partial p_B} \geq 0$ and $q_B > p_B \geq c$ by (A2).

(b) The proof is similar to that of Proposition 1(b), and is hence omitted.

(c) The proof for the case in which $f_N^* = \underline{f}$ or $\overline{f}$ is similar to that of Proposition 1(c). When $f_N^* \in (\underline{f}, \overline{f})$, from (11), $\frac{\partial ECS_N^*}{\partial p_B} = [-2\left\{(1-\beta)q_B - f_N^*\right\}\{16(1-\beta)q_B^2 + p_B q_B(\beta(15+\gamma)-16) + (p_B(15+\gamma) - 16q_B)f_N^*\} + p_B^2 q_B(1-\gamma)\partial f_N^*/\partial p_B]\left\{(1-\beta)q_B - f_N^*\right\}^{-2}$. Define $\kappa = \max_{p_B \in [p_B^*, p_B^m]} 2\left\{(1-\beta)q_B - f_N^*\right\}\{16(1-\beta)q_B^2 + p_B q_B(\beta(15+\gamma)-16) + (p_B(15+\gamma) - 16q_B)f_N^*\}p_B^{-2}q_B^{-1}(1-\gamma)^{-1}\left\{(1-\beta)q_B - f_N^*\right\}^{-2}$. Then ECS_N^* is increasing in $p_B \in [p_B^*, p_B^m]$ so that ECS_N^* is lower at p_B^* than at p_B^m if $\frac{\partial f_N^*}{\partial p_B} > \kappa$. $\square$

Proof of Proposition 3: (a) When the deceptive counterfeiter exists in the market with $\lambda = 0$, we obtain π_B^* after substituting s^{**} and w_D^* into m_B in (5) as follows:

$$\pi_B^{**} = (p_B - c)m_B - t_B q_B^2 = (p_B - c)\left[(1-\gamma)\left\{\frac{1}{2}(1-\frac{p_B}{q_B}) + \frac{1}{2}\sqrt{\frac{l(1-\frac{p_B}{q_B})}{p_B}}\right\} + \gamma(1-\frac{p_B}{q_B})\right] - t_B q_B^2. \tag{15}$$

From (15), $\frac{\partial^2 \pi_B^{**}}{\partial q_B^2} = (p_B - c)\{-\frac{p_B(1+\gamma)}{q_B^3} - \frac{(4q_B - 3p_B)(1-\gamma)}{8(q_B-p_B)q_B^3}\sqrt{\frac{l q_B p_B}{q_B - p_B}}\} - 2t_B < 0$ due to (A2), and $q_B^{**} = \frac{p_B - c}{2t_B}\left\{(1-\gamma)(\frac{p_B}{2q_B^{**2}} + \frac{1}{4q_B^{**2}}\sqrt{\frac{l p_B q_B^{**}}{q_B^{**} - p_B}}) + \frac{\gamma p_B}{q_B^{**2}}\right\}$ from the first order condition. By the same procedure in the proof of Proposition 1(a), we can prove by contradiction that $q_B^{**} < q_B^m$ if and only if $l < 4p_B(1 - \frac{p_B}{q_B^m})$. Since $s^{**} = \frac{1}{2} - \frac{1}{2}\sqrt{\frac{l}{p_B(1-\frac{p_B}{q_B})}} > 0$, we find that $l < p_B(1 - \frac{p_B}{q_B^m}) < 4p_B(1 - \frac{p_B}{q_B^m})$, so we always have $q_B^{**} < q_B^m$.

(b) To establish the result in the proposition, it suffices to show that π_D^{**} is increasing in q_B. When $\lambda = 0$, it is easy to see $\frac{\partial \pi_D^{**}}{\partial q_B} = \frac{\partial \pi_D^{**}}{\partial(1-\frac{p_B}{q_B})}\frac{p_B}{q_B^2} > 0\ \forall f_D$ from (8). Since $f_D^{**} = \underline{f}\ \forall q_B$, the result follows.

(c) When $\lambda = 0$, from (11), $\frac{\partial ECS_D^{**}}{\partial q_B} = -(1-\gamma)\left\{\frac{p_B^2}{q_B^3}s^{**} + \frac{1}{2}\left(1 - \frac{p_B^2}{q_B^2}\right)\frac{\partial s^{**}}{\partial q_B}\right\}(q_B - q_D^{**}) + \frac{1}{2}\left(1 - \frac{p_B^2}{q_B^2}\right)\{1 - (1-\gamma)s^{**}\}$. Using $\frac{\partial s^{**}}{\partial q_B} > 0$ and (A2), we obtain the condition given in Proposition 3(c). □

Remark When $\lambda > 0$, both $q_B^{**} < q_B^m$ and $q_B^{**} > q_B^m$ are possible as shown in Table 3. The condition for $q_B^{**} < q_B^m$ is $t_B > \frac{(p_B-c)}{2q_B^m}[\frac{\gamma p_B}{q_B^{m2}} - (1-\gamma)\{1 - \frac{(1-\lambda)p_B}{q_B^m} - \frac{\lambda p_B}{(1-s^{**}(q_B^m))q_B^m + s^{**}(q_B^m)(\beta q_B^m + f_D^{**}(q_B^m))}\}\frac{\partial s^{**}}{\partial q_B} + (1-\gamma)(1 - s^{**}(q_B^m))\{\frac{(1-\lambda)p_B}{q_B^{m2}} + \lambda p_B(1 + s^{**}(q_B^m)(\beta - 1 + \frac{\partial f_D^{**}}{\partial q_B}) + (f_D^{**}(q_B^m) + \beta q_B^m - q_B^m)\frac{\partial s^{**}}{\partial q_B})((1 - s^{**}(q_B^m))q_B^m + s^{**}(q_B^m)(\beta q_B^m + f_D^{**}(q_B^m)))^{-2}\}]$, which can be obtained from $\frac{\partial \pi_B^{**}}{\partial q_B}|_{q_B = q_B^m} < 0$. In this case, π_D^{**} is increasing in q_B as in the case when $\lambda = 0$ shown in the proof of Proposition 3(b). The proof follows the same procedure as in that of Lemma 2, so we provide a sketch of the proof here. For given w_D and f_D, we can show that s^{**} and m_D^{**} are increasing in q_B. Then when q_B is increased from q_{BL} to q_{BH}, the following inequalities hold in equilibrium: $\pi_D^{**}(w_D^{**}(q_{BH}), f_D^{**}(q_{BH}), q_{BH}) \geq \pi_D^{**}(w_D^{**}(q_{BL}), f_D^{**}(q_{BL}), q_{BH}) > \pi_D^{**}(w_D^{**}(q_{BL}), f_D^{**}(q_{BL}), q_{BL})$. Finally, when $\lambda > 0$, the condition for $\frac{\partial ECS_D^{**}}{\partial q_B} < 0$ given in Proposition 3(c) is modified to the following: $q_D^{**}\left\{2\bar{\theta}s^{**}\frac{\partial\bar{\theta}}{\partial q_B} + (1-\bar{\theta}^2)\frac{\partial s^{**}}{\partial q_B}\right\} < \frac{\gamma}{1-\gamma}\left(1 - \frac{p_B^2}{q_B^2}\right) + 2\left\{p_B - \bar{\theta}q_B(1-s^{**})\right\}\frac{\partial\bar{\theta}}{\partial q_B} + (1-\bar{\theta}^2)\left\{1 - \left(1 - \beta - \frac{\partial f_D^{**}}{\partial q_B}\right)s^{**} - q_B\frac{\partial s^{**}}{\partial q_B}\right\}$. Unfortunately, this condition cannot be simplified further since the closed-form expressions of s^{**} and f_D^{**} do not exist. The non-monotonicity of ECS_D^{**} is shown in Table 3.

Proof of Proposition 4: (a) From (15), $\frac{\partial^2 \pi_B^{**}}{\partial p_B^2} = -\frac{1+\gamma}{q_B} - \frac{1-\gamma}{2p_B^2}\sqrt{\frac{l q_B}{(q_B - p_B)p_B}}\left\{c + \frac{q_B(p_B - c)}{4(q_B - p_B)}\right\} < 0$, and $\frac{\partial \pi_B^{**}}{\partial p_B}|_{p_B = \frac{q_B + c}{2}} = \frac{c(1-\gamma)}{2(c+q_B)}\sqrt{\frac{l(q_B - c)}{q_B(c+q_B)}} > 0$ due to $q_B > p_B \geq c$ by (A2). Therefore, by the concavity of π_B^{**}, $p_B^{**} > p_B^m$.

(b) The non-monotonicity of π_D^{**} with respect to p_B is shown in Table 3.

(c) When $\lambda = 0$, from (11), $\frac{\partial ECS_D^{**}}{\partial p_B} = (1-\gamma)[\frac{s^{**}p_B}{q_B}\left(1 - \frac{q_D^{**}}{q_B}\right) + \left(1 - \frac{p_B}{q_B}\right)\{\frac{1}{2}\left(1 + \frac{p_B}{q_B}\right)\frac{\partial s^{**}}{\partial p_B}(q_D^{**} - q_B) - 1\}] + \gamma(\frac{p_B}{q_B} - 1)$. Using $\frac{\partial s^{**}}{\partial p_B} < 0$ and (A2), we obtain the condition given in Proposition 4(c).

□

Remark When $\lambda > 0$, both $p_B^{**} > p_B^m$ and $p_B^{**} < p_B^m$ are possible as shown in Table 3. The condition for $p_B^{**} > p_B^m$ is $\gamma(1-\frac{p_B^m}{q_B}) + (1-\gamma)(1-s^{**}(p_B^m))Y + (p_B - c)[-\frac{\gamma}{q_B} - (1-\gamma)Y\frac{\partial s^{**}}{\partial p_B} + (1-\gamma)(1-s^{**}(p_B^m))\{\frac{Y-1}{p_B^m} + \frac{\lambda p_B^m\{s^{**}(p_B^m)\frac{\partial f_D^{**}}{\partial p_B} + (f_D^{**}(p_B^m)+\beta q_B - q_B)\frac{\partial s^{**}}{\partial p_B}\}}{\{(1-s^{**}(p_B^m))q_B + s^{**}(p_B^m)(\beta q_B + f_D^{**}(p_B^m))\}^2}\}] > 0$, where $Y = 1 - \frac{(1-\lambda)p_B^m}{q_B} - \frac{\lambda p_B^m}{(1-s^{**}(p_B^m))q_B + s^{**}(p_B^m)(\beta q_B + f_D^{**}(p_B^m))}$. This can be obtained from $\frac{\partial \pi_B^{**}}{\partial p_B}|_{p_B = p_B^m} > 0$. The condition for $\frac{\partial ECS_D^{**}}{\partial p_B} > 0$ given in Proposition 4(c) is modified to the following: $q_D^{**}\left\{\overline{\theta}s^{**}\frac{\partial \overline{\theta}}{\partial p_B} - \frac{1-\overline{\theta}^2}{2}\frac{\partial s^{**}}{\partial p_B}\right\} < \frac{\gamma}{1-\gamma}\left(\frac{p_B}{q_B} - 1\right) + \left\{p_B - \overline{\theta}q_B(1-s^{**})\right\}\frac{\partial \overline{\theta}}{\partial p_B} + \frac{1-\overline{\theta}^2}{2}\left\{s^{**}\frac{\partial q_D^{**}}{\partial p_B} - q_B\frac{\partial s^{**}}{\partial p_B}\right\} - 1 + \overline{\theta}$. Unfortunately, this condition cannot be simplified further since the closed-form expressions of s^{**} and f_D^{**} do not exist. The non-monotonicity of ECS_D^{**} is established in Table 3.

Proof of Proposition 5: When $\lambda = 0$, we observe from (15) that π_B^{**} does not change with β, and that π_B^{**} is increasing in γ. When $\lambda > 0$, similar to the proof of Lemma 2, we can show that the aggregate demand, $(1-\bar{\theta})$, for the brand-name product and the deceptive counterfeit is increasing in β and decreasing in γ, and that the fraction of the brand-name product, $(1-s^{**})$, is decreasing in β and increasing in γ. The non-monotonicity of π_B^{**} is shown in our numerical experiments presented in online appendix B. The proofs for π_D^{**} and ECS_D^{**} are similar to those of Proposition 3(b)-(c), and hence are omitted. □

Proposition 6 *For any $\lambda \in [0,1]$, if any anti-counterfeiting strategy improves the average product quality in the market, $(1-s^{**})q_B + s^{**}(f_D^{**} + \beta q_B)$, then ECS_D^{**} increases.*

Proof of Proposition 6: Let us examine $\frac{\partial ECS_D^{**}}{\partial q_B}$. From the definition of CS_D, $\frac{\partial CS_D}{\partial q_B} > 0$ if and only if $\frac{\partial}{\partial q_B}\{(1-s^{**})q_B + s^{**}q_D\} > 0$. Since $\frac{\partial CS_B}{\partial q_B} > 0$, $\frac{\partial ECS_D^{**}}{\partial q_B} = (1-\gamma)\frac{\partial CS_D}{\partial q_B} + \gamma\frac{\partial CS_B}{\partial q_B} > 0$ if $\frac{\partial}{\partial q_B}\{(1-s^{**})q_B + s^{**}q_D\} > 0$. The results for the other parameters can be shown similarly. □

Proof of Corollary 1: (a) Suppose $f_N^* = \underline{f}$ or $\overline{f}$. We can obtain $\frac{\partial \pi_B^*}{\partial q_B}$ by replacing γ in the base model with $\gamma - \delta_1 f_N^*$. To show that Proposition 1(a) continues to hold, we need to prove that f_N^* is decreasing in q_B. For $f_{NH} > f_{NL}$, $\frac{\partial \pi_N^*(f_{NH})}{\partial q_B} - \frac{\partial \pi_N^*(f_{NL})}{\partial q_B}$ can be expressed as follows:

$$-\frac{p_B^2(1-\gamma+\delta_1 f_{NH})\{q_B(\beta q_B + f_{NH})(1-\beta) + f_{NH}(q_B - \beta q_B - f_{NH})\}}{8q_B^2\{(1-\beta)q_B - f_{NH}\}^2} + \frac{p_B^2(1-\gamma+\delta_1 f_{NL})\{q_B(\beta q_B + f_{NL})(1-\beta) + f_{NL}(q_B - \beta q_B - f_{NL})\}}{8q_B^2\{(1-\beta)q_B - f_{NL}\}^2}$$
$$< (1-\gamma+\delta_1 f_{NH})\left[-\frac{p_B^2\{q_B(\beta q_B + f_{NH})(1-\beta) + f_{NH}(q_B - \beta q_B - f_{NH})\}}{8q_B^2\{(1-\beta)q_B - f_{NH}\}^2} + \frac{p_B^2\{q_B(\beta q_B + f_{NL})(1-\beta) + f_{NL}(q_B - \beta q_B - f_{NL})\}}{8q_B^2\{(1-\beta)q_B - f_{NL}\}^2}\right]$$
$$= -\frac{(1-\gamma+\delta_1 f_{NH})p_B^2(f_{NH} - f_{NL})(1-\beta)\{(1-\beta)q_B - (f_{NH}+f_{NL})/2\}}{4\{(1-\beta)q_B - f_{NH}\}^2\{(1-\beta)q_B - f_{NL}\}^2},$$

which is negative due to (A1). Then, following the same procedure as in the proof of Proposition 1(a), we can show $q_B^* > q_B^m$ if and only if $\beta < 1 - \left\{\frac{q_B^m - q_N^*(q_B^m)}{q_B^m}\right\}^2$. For Proposition 1(b), when f_N is given, $\frac{\partial \pi_N^*}{\partial q_B} = \frac{(1-\gamma+\delta_1 f_N)p_B^2\left\{\beta q_B^2(\beta-1) + 2q_B f_N(\beta-1) + f_N^2\right\}}{4q_B^2\{(1-\beta)q_B - f_N\}^2} < 0$ due to (A1). When f_N^* changes from f_{N1} to f_{N2} as q_B is increased from q_{BL} to q_{BH}, $\pi_N^*(f_{N1}, q_{BL}) \geq \pi_N^*(f_{N2}, q_{BL}) > \pi_N^*(f_{N2}, q_{BH})$, where the first inequality follows from $f_N^*(q_{BL}) = f_{N1}$ and the second inequality is due to $\frac{\partial \pi_N^*}{\partial q_B} < 0\ \forall f_N$. For Proposition 1(c), to prove that ECS_N^* is increasing in q_B, it suffices to show that $\frac{\partial CS_N}{\partial q_B} > 0$.

Since $\frac{\partial CS_N}{\partial q_B}$ does not depend on γ when f_N is given, Proposition 1(c) continues to hold.

(b) We can show that the proof for Proposition 2 also applies to the case with $\gamma - \delta_1 f_N$ similarly to Proposition 1, except part (c) when $f_N^* \in (\underline{f}, \overline{f})$. With the extension, $\frac{\partial ECS_N^*}{\partial p_B} = \{(1-\beta)q_B - f_N^*\}^{-2}$ $[p_B^2 \frac{\partial f_N^*}{\partial p_B}\{-\delta_1 f_N^*(2\beta q_B - 2q_B + f_N^*) - q_B((\beta-1)\beta\delta_1 q_B + \gamma - 1)\} - 2((\beta-1)q_B + f_N^*)\{f_N^* p_B(\beta\delta_1 q_B - \gamma - 15 + \delta_1 f_N^* + \frac{16q_B}{p_B}) + q_B(16(\beta-1)q_B - p_B\beta(\gamma+15) + 16p_B)\}]$. Define $\kappa = \max_{p_B \in [p_B^*, p_B^m]} 2p_B^{-2}((\beta-1)q_B + f_N^*)\{f_N^* p_B(\beta\delta_1 q_B - \gamma - 15 + \delta_1 f_N^* + \frac{16q_B}{p_B}) + q_B(16(\beta-1)q_B - p_B\beta(\gamma+15) + 16p_B)\}\{-\delta_1 f_N^*(2\beta q_B - 2q_B + f_N^*) - q_B((\beta-1)\beta\delta_1 q_B + \gamma - 1)\}^{-1}$. Then ECS_N^* is increasing in $p_B \in [p_B^*, p_B^m]$ so that ECS_N^* is lower at p_B^* than at p_B^m if $\frac{\partial f_N^*}{\partial p_B} > \kappa$.

To show that Proposition 3(a) continues to hold, when $\lambda = 0$, we can prove by contradiction that $q_B^{**} < q_B^m$ if and only if $(1+\delta_2 f_D)l < 4p_B(1-\frac{p_B}{q_B^m})$ by the same procedure as in the proof of Proposition 1(a). Since $s^{**} = \frac{1+\delta_2 f_D}{2} - \frac{1}{2}\sqrt{\frac{l(1+\delta_2 f_D)}{p_B(1-\frac{p_B}{q_B})}} > 0$, we obtain $l < p_B(1-\frac{p_B}{q_B^m})(1+\delta_2 f_D) < \frac{4p_B}{1+\delta_2 f_D}(1-\frac{p_B}{q_B^m})$, so $q_B^{**} < q_B^m$ always holds. For Proposition 4(a), $\frac{\partial \pi_B^{**}}{\partial p_B}|_{p_B = \frac{q_B+c}{2}} = \frac{c(1-\gamma+\delta_1 f_D)}{2(c+q_B)}\sqrt{\frac{l(q_B-c)(1+\delta_2 f_D)}{q_B(c+q_B)}} > 0$, so $p_B^{**} > p_B^m$. For Propositions 3(b) and 4(b), we can show that, with the extension, s^{**} and m_D^{**} are increasing in q_B and decreasing in p_B for given w_D and f_D. Then, when q_B is increased from q_{BL} to q_{BH}, the following inequalities hold in equilibrium: $\pi_D^{**}(w_D^{**}(q_{BH}), f_D^{**}(q_{BH}), q_{BH}) \geq \pi_D^{**}(w_D^{**}(q_{BL}), f_D^{**}(q_{BL}), q_{BH}) > \pi_D^{**}(w_D^{**}(q_{BL}), f_D^{**}(q_{BL}), q_{BL})$. Similarly, when p_B is decreased from p_{BH} to p_{BL}, the following inequalities hold in equilibrium: $\pi_D^{**}(w_D^{**}(p_{BL}), f_D^{**}(p_{BL}), p_{BL}) \geq \pi_D^{**}(w_D^{**}(p_{BH}), f_D^{**}(p_{BH}), p_{BL}) > \pi_D^{**}(w_D^{**}(p_{BH}), f_D^{**}(p_{BH}), p_{BH})$. Propositions 3(c) and 4(c) can be shown similarly to Proposition 2(c).

(c) The proofs of π_D^{**} and ECS_D^{**} are similar to those of Proposition 3(b)-(c). When $\lambda = 0$, the non-monotonicity of π_B^{**} or ECS_D^{**} with respect to γ can be shown numerically as follows. Set $q_B = 1$, $p_B = 0.5$, $t_B = t_D = 0.01$, $c = 0.01$, $\beta = 0.1$, $l = 0.02$, and $\delta_1 = \delta_2 = 0.1$. As γ increases from 0.2 to 0.3, π_B^{**} increases from 0.174 to 0.181 and ECS_D^{**} increases from 0.070 to 0.076. As γ increases from 0.3 to 0.4, π_B^{**} decreases from 0.181 to 0.173 and ECS_D^{**} decreases from 0.070 to 0.069. $\square$

Appendix B. Numerical Experiments

This section contains our numerical study that examines the effectiveness of the marketing campaign and the enforcement strategy against the *deceptive* counterfeiter. Similar to the numerical study presented in §5.2, we have constructed 1024 scenarios for $\lambda = 0$, 0.25 or 0.5, using the parameter values shown in the bottom of Table 5, so that they cover various possible scenarios and also satisfy positive s^{**} in equilibrium. We computed the difference in firms' expected profits and expected consumer welfare associated with the adjacent values of β or γ for a fixed set of other parameter values. There are 3 increments of β or γ for a set of 256 possible values of (t_i, c, γ, l), so there are 768 scenarios for which we can examine the direction of changes with a decrease of β or an increase of γ. The results are summarized in Table 5, which reads as follows: for example, when $\lambda = 0.5$, reducing β increased π_B^{**} in 33.1% of 768 scenarios, decreased π_D^{**} in all scenarios, and increased ECS_D^{**} in 13.9% of 768 scenarios.

Table 5. Effects of Marketing Campaigns and Enforcement against Deceptive Counterfeits

	Effects of Reducing β			Effects of Increasing γ		
	$\pi_B^{**} \uparrow$	$\pi_D^{**} \downarrow$	$ECS_D^{**} \uparrow$	$\pi_B^{**} \uparrow$	$\pi_D^{**} \downarrow$	$ECS_D^{**} \uparrow$
$\lambda = 0$	no change	no change	0	1	1	1
$\lambda = 0.25$	0.374	1	0.260	1	1	0.952
$\lambda = 0.5$	0.331	1	0.139	0.990	1	0.927

(Note) Each number in the table indicates a percent of scenarios for the corresponding effect. We used the following parameters: $t \in \{0.005, 0.01, 0.015, 0.02\}$, $\beta \in \{0.1, 0.2, 0.3, 0.4\}$, $\gamma \in \{0.1, 0.2, 0.3, 0.4\}$, $l \in \{0.005, 0.01, 0.015, 0.02\}$, $c \in \{0.005, 0.01, 0.015, 0.02\}$, $\underline{f} = 0.1$, and $\overline{f} = (1-\beta) * q_B - 0.01$. (q_B, p_B) is fixed at (q_B^m, p_B^m).

Table 5 confirms the results stated in Proposition 5. In addition, similar to the quality and pricing strategies discussed in §5.2, these strategies are not necessarily more effective as more consumers are proactive with higher λ.

Appendix C. Extension: Price Decision of Licit Distributor

Suppose, instead of the brand-name company, the licit distributor decides the retail price of the brand-name product, p_B, when combating the *deceptive* counterfeiter. The brand-name company instead decides the wholesale price, w_B, to the licit distributor. The rest of the decisions remain the same as in the base model. The sequence of decisions is as follows: After observing the quality q_B and wholesale price w_B of the brand-name product, the *deceptive* counterfeiter decides his functional quality f_D and wholesale price w_D in stages 1 and 2, respectively. In stage 3, the licit distributor decides a fraction of deceptive counterfeits s, and then decides the retail price p_B. Note that the licit distributor can source products from the brand-name company at the wholesale price w_B, and/or from the deceptive counterfeiter at the wholesale price w_D, should counterfeit goods reach the market.

We analyze this model backwards. In stage 3, the licit distributor first solves the following problem to determine p_B:

$$\max_{p_B} \ (1-s)\{p_B - sw_D - (1-s)w_B\}\left\{1 - \frac{\lambda p_B}{(1-s)q_B + s(f_D + \beta q_B)} - \frac{(1-\lambda)p_B}{q_B}\right\} - sl. \tag{16}$$

One can verify from (16) that the profit of the licit distributor is concave in p_B. We obtain from the first-order condition the optimal retail price: $p_B^{**} = \frac{sw_D + (1-s)w_B}{2} + \frac{1}{2}\left(\frac{\lambda}{(1-s)q_B + s(f_D + \beta q_B)} + \frac{1-\lambda}{q_B}\right)^{-1}$.

Since $w_D < w_B$ and $f_D + \beta q_B < q_B$, the distributor charges a lower retail price p_B^{**} when the fraction of counterfeits s is larger or the fraction of proactive consumers λ is larger.

Next, to find the optimal fraction s^{**}, the licit distributor solves the following problem which is obtained by substituting p_B^{**} into (16):

$$\max_s \frac{1}{4}(1-s)\left\{\frac{\lambda}{(1-s)q_B + s(f_D+\beta q_B)} + \frac{1-\lambda}{q_B}\right\} \times \left[sw_D + (1-s)w_B - \left\{\frac{\lambda}{(1-s)q_B + s(f_D+\beta q_B)} + \frac{1-\lambda}{q_B}\right\}^{-1}\right]^2 - sl \quad (17)$$

Due to complexity, however, it is not possible to find the closed-form expression of s^{**} from (17). Although the part of our analysis in the base model does not rely on its closed-form expression, the impact of any anti-counterfeiting strategy on s^{**} becomes prohibitively complex to obtain any analytical result. Thus, we conduct extensive numerical experiments to examine the effects of the anti-counterfeiting strategies. We use the same set of parameters as in online appendix B. For each case with $\lambda = 0$ and $\lambda = 0.5$, there are 1024 scenarios in which we can investigate the anti-counterfeiting strategies that change quality or price from the case with no counterfeiter to the optimal levels. On the other hand, similar to Table 5, there are 768 scenarios for which we can examine the anti-counterfeiting strategies that reduce β or increase γ. We present a summary of the results in Table 6, which reads similarly to Table 5.

Table 6. Effects of Quality and Pricing Strategies in the Extended Model

	$\lambda = 0$			$\lambda = 0.5$		
	$\pi_B^{**} \uparrow$	$\pi_D^{**} \downarrow$	$ECS_D^{**} \uparrow$	$\pi_B^{**} \uparrow$	$\pi_D^{**} \downarrow$	$ECS_D^{**} \uparrow$
$q_B^m \to q_B^{**}$	1	0.320	0.672	1	0.998	0.647
$w_B^m \to w_B^{**}$	1	0.508	0.805	1	0.998	0.647
$\beta \downarrow$	no change	no change	0	0.135	1	0.135
$\gamma \uparrow$	1	1	1	0.779	1	0.798

Table 6 shows that the effects of anti-counterfeiting strategies remain directionally true in this extended model. For example, as the price changes from w_B^m to w_B^{**}, π_D^{**} and ECS_D^{**} can increase or decrease; this is consistent with Proposition 4. Also, as stated in Proposition 5, when $\lambda = 0$, reducing β has no impact on π_B^{**} and π_D^{**}, but reduces ECS_D^{**}, whereas increasing γ reduces π_D^{**} and increases π_B^{**} as well as ECS_D^{**}; when $\lambda > 0$, reducing β or increasing γ reduces π_D^{**}, but it can increase or reduce π_B^{**} and ECS_D^{**}. One notable exception is that when $\lambda = 0$, changing q_B^m to q_B^{**} can increase π_D^{**} although it always reduces π_D^{**} in the base model. This happens because of the additional lever (i.e., determining p_B as well as s) the licit distributor has in this extended model. In response to the change of q_B, the distributor can increase the aggregate demand for both brand-name and counterfeit goods by reducing the retail price p_B. As a result, we find that the distributor may increase or decrease the fraction of counterfeits, s^{**}, in response to this strategy, which thus creates a non-monotonic effect on π_D^{**}.