Incentives and Flexibility in a Decentralized Multi-Product Assemble-to-Order System

Fernando Bernstein  Gregory A. DeCroix  Yulan Wang

The Fuqua School of Business, Duke University, Durham, NC 27708, USA
December 1, 2004

In this paper, we explore the impact of decentralized decision making on the behavior of assemble-to-order systems. Specifically, we consider a system where the assembler buys three components (two product-specific and one common) from three independent suppliers and produces two end products to satisfy stochastic customer demands. The assembler sets wholesale prices paid to the suppliers, and the suppliers then decide how much component capacity to install. At the end of a single selling season, demands are observed, and the assembler makes production decisions based on the capacity constraints. We prove that, for any choice of wholesale prices, there exists a unique Pareto-optimal equilibrium in the suppliers’ capacity game. We show that the assembler’s optimal wholesale prices lie in one of two regions – one leads to component risk pooling and one does not. We also derive the optimal/equilibrium capacity/pricing decisions for two other variations of the above system: one with a centralized decision maker and one where the common component is replaced by two product-specific components. By comparing behavior in the three systems, we identify and explore various types of inefficiencies arising from decentralization. Consistent with previous studies of other supply-chain settings, we find that decentralization leads to understocking in terms of component capacities. In addition, however, we identify other inefficiencies that are directly related to the multi-component, multi-product structure analyzed here. Specifically, risk pooling of the common component is less frequent under decentralized decision making. Also, the apparent flexibility of a common component may actually hurt performance in a decentralized system whereas it can only help in a centralized system.
1 Introduction

Due to their ability to keep finished-goods inventories low while providing reasonably responsive service to customers, assemble-to-order systems have become increasingly popular in recent years. Under such a system, a wide variety of finished products can be produced from a relatively small number of components. Perhaps the best-known example of this approach to manufacturing is that of Dell Computer. Dell does not stock fully assembled computers, but relies on inventories of components (monitors, processors, hard drives, etc.) which are stored both at its own facility (in very small quantities) and at nearby facilities operated by its suppliers. When a customer order arrives, these components can be quickly assembled into whatever specific product the customer wants, and the product can usually be delivered to the customer within 5 to 6 days (Magretta 1998). Several other computer manufacturers are also following this type of assemble-to-order strategy. This approach has been applied successfully in many other industries as well. For example, Cisco Systems, builds approximately 65% of products using assemble-to-order, and aims to ship a customer order within two weeks of receipt (Carbone 2000). In the automobile industry, Renault, BMW and Volvo are pioneers in building cars based on direct orders from customers. To shorten leadtimes, key suppliers have set up plants close to the auto makers’ facilities, and some even connect their operations directly to the assembly plant with conveyors. By the end of 2002, Renault was producing 55% of its output using assemble-to-order techniques (Kochan 2003). BMW has reduced their delivery leadtime to 10 days, and a customer order can be changed as few as 6 days before final assembly (Weernink 2002). Finally, a version of assemble-to-order is practiced by amazon.com and all other catalog and online retailers. Here a “product” corresponds to a particular combination of items in an order. When a customer order arrives, these items must be picked from “component” inventory and assembled before shipping.

In addition to their fundamental inventory/production strategy (stocking only components and producing finished goods only once an order is received), most assemble-to-order systems in practice also involve some level of decentralization. This can be seen in the above examples in both the computer and automobile industries. The majority of the component inventories that Dell relies on for its assembly process are actually held by its suppliers. Similarly, in the automobile industry, component production capacity required to feed the final assembly plant is typically owned and controlled by the suppliers. Given recent trends
towards outsourcing, modular assembly, etc., this decentralization aspect may become even more important in the years ahead.

While there has been significant research into the operation of assemble-to-order systems, very little attention has been paid to the impact of decentralized decision making in such systems. This paper focuses on that impact.

We analyze a decentralized assemble-to-order (ATO) system consisting of two finished products and three components (with one component dedicated to each product and one common component shared by them). Component capacity decisions are made by individual component suppliers and final production decisions are made by an assembler. Prior to a single selling season, the assembler sets wholesale prices for the components and then the suppliers simultaneously select component production capacities. Once demand for the finished products has been observed, the assembler decides how much of each to produce, subject to the availability of component capacities. Demands for the finished products are stochastic and independent, and unfilled demands are lost. Suppliers are paid only for the components actually used to produce finished products. We show that for any choice of wholesale prices by the assembler, a unique Pareto-optimal equilibrium always exists in the suppliers’ capacity game. We also provide a partial characterization of the optimal assembler prices, and identify which price choices lead to risk-pooling of the common component (i.e., situations where the capacity of the common component is strictly less than the sum of the capacities of the two dedicated components).

To address the impact of decentralization, we also analyze a version of this system under a centralized decision maker, and we compare behavior of the two systems both analytically and numerically. This comparison yields a number of interesting insights. First, consistent with what has often been observed in other decentralized systems, capacity stocking levels in the decentralized system are always lower than in the centralized system. In addition, this multi-product setting allows us to identify a new source of inefficiency resulting from decentralization – risk pooling is less common in the decentralized system. We also explore the impact of decentralization on the value of having a common component. To address this question we construct variations of both the centralized and decentralized models where the common component is split into two dedicated components, each provided by a different supplier. While in a centralized system a common component always yields (weakly) higher profits, this is not the case for the decentralized system. In the decentralized system, the apparent flexibility of a common component can actually reduce profits, even in cases when
having the common component results in risk pooling.

There exists a significant body of research analyzing performance measures and inventory policies for centralized assemble-to-order systems. Some early work in this area includes analysis of the repair-kit problem by Smith et al. (1980) and Graves (1982). Lu and Song (2003) formulate a customer-order level cost-minimization model to determine the joint optimal base-stock levels in the multi-product ATO system and compare it with the single-item cost minimization model. Lu et al. (2003) treat the ATO system as a set of queues driven by a common, multiclass batch Poisson input. They derive the first two moments of the joint queue-length distribution, and investigate the effect of demand and leadtime variability on the system performance. See the review of ATO system by Song and Zipkin (2003) for a complete list of work in this area.

Another related stream of work is the research on decentralized single-product assembly systems. Wang and Gerchak (2003) study a setting that is similar to the one considered here, but with only a single finished product. Under two different leadership assumptions ("assembler-as-leader" and "supplier-as-leader") they derive expressions for equilibrium supplier capacities and analyze the effect of system structure and parameters on performance of the system. Tomlin (2003) studies a similar system and explores the use of share-the-gain contracts to increase supplier capacities. Bernstein and DeCroix (2004a) study the issue of modular assembly in a multi-tier assembly system, where some of the assembly work is done by a middle tier of subassemblers. They characterize equilibrium capacities and prices in that setting and explore the best way for the assembler to structure the modular assembly system. They also show that modular assembly is only beneficial to the assembler if subassemblers can perform assembly work less expensively.

Other related work includes the study of capacity investment, component commonality and resource flexibility. Relevant research on component commonality includes Collier (1982), Baker et al. (1986), Gerchak et al. (1988) and Gerchak and Henig (1989). In settings similar to ours, these papers explore the benefits of component commonality under centralized decision making. Van Mieghem (1998) studies the issue of resource flexibility in a one-period, two-product setting, with one dedicated resource for each product and one flexible resource that can be used for either product. Netessine et al. (2002) explore the implications of flexibility in a service environment. Finally, Van Mieghem (2004) investigates the relationship between the commonality and flexible capacity problems. For a recent review of work addressing game-theoretic capacity investment by multiple agents, see Van
Mieghem (2003).

The rest of the paper is organized as follows. Section 2 introduces the basic model and notation. The behavior of the decentralized assemble-to-order system is characterized in Section 3, while in Section 4 we analyze the system under centralized decision making. In Section 5 we compare behavior in the two systems and derive managerial insights regarding the impacts of decentralization. Section 6 provides some concluding remarks.

# Model

Consider an assemble-to-order (ATO) system that produces two products (labeled 1 and 2) using three components (labeled $A$, $B$, and $C$). Product 1 consists of one unit each of components $A$ and $B$, while product 2 consists of one unit each of components $B$ and $C$. (More general component quantity requirements can be reduced to this case by rescaling the problem parameters.) As a result, we say that components $A$ and $C$ are dedicated to products 1 and 2, respectively, while component $B$ is common to the two products. Figure 1 below illustrates the system.

![Figure 1: ATO System](image)

Components $A$, $B$ and $C$ are produced by suppliers $A$, $B$ and $C$ respectively, while a single assembler makes assembly decisions for the two products. Customer demand for product $j$ ($j = 1, 2$) during a single selling season, denoted by $D_j$, is a stochastic random variable with a continuous distribution represented by the cumulative distribution function (cdf) $F_j(\cdot)$ and density function $f_j(\cdot)$. Assume that $f_j(x) > 0$ for all $x \geq 0$ and $f_j(x) = 0$ otherwise, and also that demands for the two products are independent. We denote by $F_{12}(\cdot)$ the cdf
of aggregate consumer demand for products 1 and 2, i.e., \( D_1 + D_2 \). Since demands are independent, \( F_{12}(\cdot) = F_1 * F_2(\cdot) \), where * indicates convolution. Define \( \tilde{F}_j(\cdot) = 1 - F_j(\cdot) \).

Let \( p_j \) be the market price for each product net of the assembly costs incurred by the assembler, so unit assembly costs are not included separately. These prices are exogenous. Each suppliers \( i \) incurs a cost \( c_i \) for each unit of capacity installed or reserved. Assume that supplier production costs (once capacity is installed) are zero. (The analysis can be extended to the case of positive production costs.) We assume that both products have a positive profit margin, i.e., \( p_1 > c_A + c_B \) and \( p_2 > c_B + c_C \), and that the demand distributions and all cost parameters are common knowledge.

Assume that all firms are geographically close and that lead times for assembly are negligible, so that final products can be assembled to order after demand is observed. However, suppliers need to plan in advance the amount of capacity they reserve for their component production activities. As a result, the sequence of events is as follows. First, the assembler acts as the Stackelberg leader by choosing the wholesale price \( w_i \) it will pay to each supplier \( i \) for each unit of component \( i \) produced. Then, the suppliers simultaneously install or reserve their production capacities \( Q_i \). (For technical reasons, assume that the feasible set for \( Q_i \) is \([0, \overline{Q}_i] \) for some large finite \( \overline{Q}_i \).) Next, the assembler observes consumer demands \( D_j \), decides how many units \( y_j \) of each product \( j \) to assemble subject to the suppliers’ capacity constraints, and places the corresponding orders with the suppliers. Any unsatisfied demands are lost. Finally, all costs and revenues are incurred.

Below is a summary of notation used in the paper. (The superscript \( T \) indicates the transpose.)

\[
c_i = \text{supplier } i \text{'s unit capacity cost for component production, } i = A, B, C; \\
c = (c_A, c_B, c_C)^T; \\
w_i = \text{assembler’s per-unit wholesale price for supplier } i = A, B, C; \\
w = (w_A, w_B, w_C)^T; \\
Q_i = \text{capacity selected by supplier } i, i = A, B, C; \\
Q = (Q_A, Q_B, Q_C)^T; \\
Q_{-i} = \text{the vector } Q \text{ with the component corresponding to player } i \text{ removed};
\]
\[ p_j = \text{product } j\text{'s market price, } j = 1, 2; \]
\[ p = (p_1, p_2)^T; \]
\[ D_j = \text{demand for product } j, j = 1, 2; \]
\[ D = (D_1, D_2)^T; \]
\[ y_j = \text{quantity assembled of product } j, j = 1, 2; \]
\[ y = (y_1, y_2)^T. \]

3 Equilibrium Analysis

In this section, we analyze the equilibrium behavior of the ATO system by considering the events in reverse order. First, we investigate the assembler’s optimal production decision once the wholesale prices and component capacities have been set and demand has been realized. In what follows, it will be convenient to explicitly express the dependence of these finished-product production quantities on the component capacities and realized demand by writing the production vector as \( y(Q, D) = (y_1(Q, D), y_2(Q, D))^T \). Next, we explore the capacity game, in which suppliers simultaneously select capacity levels \( Q_i \) given the wholesale price vector \( w \) and anticipating the resulting assembly decision \( y(Q, D) \). Finally, we consider the optimization problem in which the assembler selects the vector of wholesale prices \( w \) that maximizes its profit in anticipation of the suppliers’ equilibrium capacity vector \( Q \) and its own optimal finished product assembly decision.

3.1 Assembler’s Production Decision

When analyzing the assembler’s optimal production decision \( y(Q, D) \), we assume (without loss of optimality) that the vector of wholesale prices \( w \) is such that both finished products earn a positive margin for the assembler, and (without loss of generality) that product 2 has the (weakly) higher margin, i.e., \( 0 < p_1 - w_A - w_B \leq p_2 - w_B - w_C \). The optimal production vector that maximizes the supplier’s profit \( \Pi_0 \) is defined as the solution to the following linear program:
Max $\Pi_0 = (p_1-w_A-w_B)y_1 + (p_2-w_B-w_C)y_2$

subject to $y_1 \leq Q_A$

$y_1 + y_2 \leq Q_B$

$y_2 \leq Q_C$

$0 \leq y_1 \leq D_1$

$0 \leq y_2 \leq D_2$

The only resource shared by the two products is component $B$. Since each product uses the same amount of this resource and product 2 is more profitable, it is optimal for the assembler to give priority to product 2 when choosing production quantities.\footnote{Without loss of optimality, we assume that the assembler gives priority to product 2 even when profit margins are equal.} That is, the assembler satisfies as much demand for product 2 as possible given capacities of components $B$ and $C$, and then uses any remaining capacity of component $B$ along with capacity of component $A$ to satisfy as much demand for product 1 as possible. As a result, we obtain the following closed-form expressions for the optimal production quantities given component capacities $Q$ and demands $D$:

$$y_1(Q,D) = \min\{Q_A, Q_B - \min\{Q_B, Q_C, D_2\}, D_1\}, \quad (1)$$

$$y_2(Q,D) = \min\{Q_B, Q_C, D_2\}. \quad (2)$$

3.2 Suppliers’ Capacity Game

Faced with a wholesale price set by the assembler, and anticipating the assembler’s production decision in the last stage, each supplier chooses the capacity that maximizes its expected profit given the capacities chosen by the other suppliers. Given vectors $Q$ and $D$, the total number of units $s_i(Q,D)$ sold by supplier $i$ is given by

$$s_A(Q,D) = y_1(Q,D)$$

$$s_B(Q,D) = y_1(Q,D) + y_2(Q,D)$$

$$= \min\{Q_A + \min\{Q_B, Q_C, D_2\}, Q_B, D_1 + \min\{Q_B, Q_C, D_2\}\}$$

$$= \min\{Q_B, \min\{Q_A, D_1\} + \min\{Q_C, D_2\}\}$$

$$s_C(Q,D) = y_2(Q,D).$$
Supplier $i$’s expected profit function is then given by

$$\Pi_i(Q_i|Q_{-i}) = w_i E_D[s_i(Q, D)] - c_i Q_i.$$  

The following result establishes some key properties of the suppliers’ capacity game.

**Proposition 1.** Given a vector $w$ of wholesale prices, each supplier $i$’s expected profit function is concave in $Q_i$. As a result, there exists at least one Nash Equilibrium capacity vector $Q^*(w)$ in the suppliers’ capacity game.

In the remainder of this section we focus on characterizing such equilibria.

In decentralized single-product assembly systems, the economic complementarities associated with the product structure (i.e., a component is only valuable if matched with the other components) often cause associated supplier capacity games to be supermodular. (See, e.g., Wang and Gerchak 2003, Bernstein and DeCroix 2004a, Bernstein and DeCroix 2004b.) This is not the case in the current setting, however, due to the existence of multiple products and the shared component $B$. To see why this is true, consider how a supplier’s optimal capacity choice would be affected by changes in the other suppliers’ capacities. Pairwise complementarity does exist between supplier B’s capacity choice and the capacity choice of either supplier $A$ or supplier $C$. All else being equal, supplier $A$ (or supplier $C$) would want to (weakly) increase its capacity in response to an increase in supplier $B$’s capacity (and vice versa), since there would now be more components available to match with its own. However, capacities of components $A$ and $C$ act as economic substitutes – a higher level of one leads to a lower level of the other – since they must compete for the limited capacity of component $B$. This observation is closely related to a result in Zipkin (2003), which shows that linear programs like the assembler’s production problem are generally not supermodular in the right-hand sides of the constraints when there are more than two constraints.

Although the suppliers’ capacity game is not supermodular in the original decision variables, we will see later that the structures of the suppliers’ best-response functions are such that some degree of complementarity among the capacity choices does exist. We will exploit this fact to help identify, compare and compute equilibria in the game.

As a first step in characterizing each supplier’s best-response function, define supplier $i$’s isolated optimal capacity with respect to demand for product $j$ as

$$\hat{Q}_i^j = F_{j-1}^{-1} \left( \frac{c_i}{w_i} \right).$$
In addition, for supplier B, define its isolated optimal capacity with respect to aggregate demand for products 1 and 2 as

\[ \hat{Q}^{12}_{B} = \bar{F}_{12}^{-1} \left( \frac{c_B}{w_B} \right) . \]

(Note that \( \hat{Q}^{1}_{A} \), \( \hat{Q}^{12}_{B} \), and \( \hat{Q}^{2}_{C} \) correspond to the optimal newsvendor quantities for suppliers A, B, and C, respectively, when the other suppliers have ample production capacity.) We now explore each supplier’s best-response function in detail.

**Supplier C’s Best-Response Function**

Consider supplier C’s profit function,

\[ \Pi_C(Q_C|Q_A, Q_B) = w_C E_D [s_C(Q, D)] - c_C Q_C = w_C E_D [\min(Q_C, Q_B, D_2)] - c_C Q_C. \tag{3} \]

If \( Q_B \geq \hat{Q}^{2}_{C} \), clearly supplier C’s optimal capacity is \( \hat{Q}^{2}_{C} \). If instead \( Q_B < \hat{Q}^{2}_{C} \), then \( \Pi_C \) is increasing on \( Q_C < Q_B \) and decreasing (linearly with a slope of \(-c_C\)) on \( Q_C > Q_B \), so the optimal capacity is \( Q_C = Q_B \). As a result, supplier C’s best-response function is

\[ r_C(Q_A, Q_B) = r_C(Q_B) = \min(\hat{Q}^{2}_{C}, Q_B). \tag{4} \]

(Note that supplier C’s best-response function is independent of \( Q_A \). This is due to the fact that product 2 has higher production priority than product 1.)

**Supplier B’s Best-Response Function**

Consider now supplier B’s profit function,

\[ \Pi_B(Q_B|Q_A, Q_C) = w_B E_D [s_B(Q, D)] - c_B Q_B = w_B E_D [\min(Q_B, \min(D_1, Q_A) + \min(D_2, Q_C))] - c_B Q_B. \tag{5} \]

From (5) it is clear that supplier B’s optimal capacity choice satisfies \( Q_B \leq Q_A + Q_C \), since any higher level would result in wasted capacity for component B. Establishing differentiability of supplier B’s expected sales requires careful treatment of some technical details. In the appendix we show that \( E_D [s_B(\cdot, D)] \) is differentiable for \( Q_B < Q_A + Q_C \). If

\[ \frac{\partial E_D [s_B(Q, D)]}{\partial Q_B} \geq \frac{c_B}{w_B} \]

10
for all $Q_B < Q_A + Q_C$, then supplier $B$’s optimal capacity choice is $Q_B = Q_A + Q_C$. Otherwise, the optimal capacity choice satisfies $Q_B < Q_A + Q_C$, and is characterized by

$$\frac{\partial E_D [s_B(Q, D)]}{\partial Q_B} = \frac{c_B}{w_B}. \quad (6)$$

In the appendix, we also show that

$$\frac{\partial E_D [s_B(Q, D)]}{\partial Q_B} = P_D (\{D : Q_B \leq \min(D_1, Q_A) + \min(D_2, Q_C)\}).$$

The set $\{D : Q_B \leq \min(D_1, Q_A) + \min(D_2, Q_C)\}$ corresponds to region I in Figure 2. Then, (6) can be written as

$$1 - F_1(Q_B - Q_C) - F_2(Q_B - Q_A)\bar{F}_1(Q_A) - \int_{Q_B - Q_C}^{Q_A} F_2(Q_B - x) dF_1(x) = \frac{c_B}{w_B}, \quad (7)$$

where the first two terms correspond to the probability that the demand vector $D$ falls in the union of regions I, II, and III in the figure, while the second and third terms correspond to the probability that $D$ falls in regions II and III, respectively. If there is a $Q_B < Q_A + Q_C$ that solves (7), then we call that solution $r_{BAC}(Q_A, Q_C)$. Note that the left-hand side of (7) can be interpreted as the probability that supplier $B$ stocks out, so (7) is a two-dimensional analogue of the optimality condition for the classical newsvendor problem.

![Figure 2: Supplier B’s First Order Condition](image)

We can now characterize supplier $B$’s best-response function.
Proposition 2. The best-response function for supplier B is given by

\[ r_B(Q_A, Q_C) = \begin{cases} 
Q_A + Q_C, & \text{if } \bar{F}_1(Q_A) \bar{F}_2(Q_C) \geq \frac{c_B}{w_B}; \\
r_{BAC}(Q_A, Q_C), & \text{if } \bar{F}_1(Q_A) \bar{F}_2(Q_C) < \frac{c_B}{w_B}. 
\end{cases} \]

In addition, if \( \bar{F}_1(Q_A) \bar{F}_2(Q_C) < \frac{c_B}{w_B} \), then

\[ 0 < \frac{\partial r_{BAC}}{\partial Q_A} < 1 \quad \text{and} \quad 0 < \frac{\partial r_{BAC}}{\partial Q_C} < 1. \]

As can be seen from the expression for \( r_B(Q_A, Q_C) \), for very low values of \( Q_A \) and \( Q_C \), supplier B does not benefit from risk pooling. If \( Q_A \) or \( Q_C \) is increased, \( Q_B \) increases by the same amount up to the point where \( \bar{F}_1(Q_A) \bar{F}_2(Q_C) = \frac{c_B}{w_B} \). Further increases in \( Q_A \) or \( Q_C \) are only partially matched by supplier B – i.e., supplier B now benefits from risk pooling.

Supplier A’s Best-Response Function

Finally, consider supplier A’s profit function

\[ \Pi_A(Q_A|Q_B, Q_C) = w_A E_D [s_A(Q, D)] - c_A Q_A \\
= w_A E_D [\min (Q_A, Q_B - \min(Q_B, Q_C, D_2), D_1)] - c_A Q_A. \] (8)

From (8) it is clear that supplier A’s optimal capacity choice satisfies \( Q_A \leq Q_B \), since any higher level would result in wasted capacity for component A. Now fix a pair \((Q_B, Q_C)\), and assume that \( Q_C \leq Q_B \). (Under supplier C’s best response, this will always hold.) We first obtain the derivative of the expected profit function for supplier A.

Lemma 1. If \( Q_C \leq Q_B \), then \( \Pi_A \) is differentiable for \( Q_A \neq Q_B - Q_C \), and

\[ \frac{\partial \Pi_A}{\partial Q_A} = \begin{cases} 
w_A \bar{F}_1(Q_A) - c_A, & \text{if } Q_A < Q_B - Q_C \\
E_D [s_A(Q, D)] - \bar{F}_1(Q_A) - c_A, & \text{if } Q_A > Q_B - Q_C 
\end{cases} \] (9)

Note from (9) that \( \partial \Pi_A / \partial Q_A \) is strictly decreasing in \( Q_A \), with a downward (discontinuous) jump at \( Q_A = Q_B - Q_C \). Thus, if we can find a capacity \( Q_A \) that satisfies \( \partial \Pi_A / \partial Q_A = 0 \), then that is supplier A’s best response. However, such a point may not always exist.

Let \( r_A(Q_B, Q_C) \) be supplier A’s best-response function. Recall that \( \hat{Q}_A^1 \) solves \( w_A \bar{F}_1(Q_A) - c_A = 0 \), so if \( \hat{Q}_A^1 < Q_B - Q_C \) (or, equivalently, \( \frac{c_A}{w_A} > \bar{F}_1(Q_B - Q_C) \)), \( r_A(Q_B, Q_C) = \hat{Q}_A^1 \). Suppose instead that \( \hat{Q}_A^1 \geq Q_B - Q_C \). Then, \( \partial \Pi_A / \partial Q_A > 0 \) for all \( Q_A < Q_B - Q_C \), so
\( r_A(Q_B, Q_C) \geq Q_B - Q_C \). Now let \( r_{AB}(Q_B) \) be the solution to \( w_A F_2(Q_B - Q_A) \hat{F}_1(Q_A) - c_A = 0 \). (Such a solution exists if and only if \( Q_B \geq \hat{Q}_B \), where

\[
\hat{Q}_B = F_2^{-1} \left( \frac{c_A}{w_A} \right).
\]

Note that \( r_{AB}(\hat{Q}_B) = 0 \). If \( r_{AB}(Q_B) \) exists and \( r_{AB}(Q_B) > Q_B - Q_C \) (or, equivalently, \( \frac{c_A}{w_A} < F_2(Q_B) \hat{F}_1(Q_B - Q_C) \)), then \( r_A(Q_B, Q_C) = r_{AB}(Q_B) \). Otherwise, \( \partial \Pi_A / \partial Q_A < 0 \) for \( Q_A > Q_B - Q_C \), in which case \( r_A(Q_B, Q_C) = Q_B - Q_C \). In summary,

\[
r_A(Q_B, Q_C) = \begin{cases} 
\hat{Q}_A, & \text{if } \frac{c_A}{w_A} > F_1(Q_B - Q_C) \\
Q_B - Q_C, & \text{if } \frac{c_A}{w_A} < F_2(Q_B) \hat{F}_1(Q_B - Q_C) \\
r_{AB}(Q_B), & \text{otherwise.}
\end{cases}
\]  

(10)

Recall that the best-response function for supplier \( C \) is independent of \( Q_A \). As a result, one need only consider the best response of supplier \( A \) to the pair \( (Q_B, r_C(Q_B)) \), i.e., \( r_A(Q_B, r_C(Q_B)) \). In addition, note that \( r_C(Q_B) = \min(\hat{Q}_C^2, Q_B) \leq Q_B \). The expression in (10) is then valid replacing \( Q_C = r_C(Q_B) \). This composite best-response function is characterized in Proposition 3 and illustrated in Figure 3.

**Proposition 3.** If \( \frac{c_A}{w_A} + \frac{c_C}{w_C} \geq 1 \), or \( \hat{Q}_C^2 \leq \hat{Q}_B \), we have that

\[
r_A(Q_B, r_C(Q_B)) = \begin{cases} 
\hat{Q}_A^1, & \text{if } Q_B > \hat{Q}_A^1 + \hat{Q}_C^2 \\
Q_B - \hat{Q}_C^2, & \text{if } \hat{Q}_C^2 < Q_B \leq \hat{Q}_A^1 + \hat{Q}_C^2 \\
0, & \text{if } Q_B \leq \hat{Q}_C^2.
\end{cases}
\]

If \( \frac{c_A}{w_A} + \frac{c_C}{w_C} < 1 \), or \( \hat{Q}_C^2 > \hat{Q}_B \), we have that

\[
r_A(Q_B, r_C(Q_B)) = \begin{cases} 
\hat{Q}_A^1, & \text{if } Q_B > \hat{Q}_A^1 + \hat{Q}_C^2 \\
\hat{Q}_B - \hat{Q}_C^2, & \text{if } \hat{Q}_C^2 + F_1^{-1} \left( \frac{c_A}{w_A} \frac{w_C}{w_C - c_C} \right) \leq Q_B \leq \hat{Q}_A^1 + \hat{Q}_C^2 \\
r_{AB}(Q_B), & \text{if } Q_B \leq Q_B < \hat{Q}_C^2 + F_1^{-1} \left( \frac{c_A}{w_A} \frac{w_C}{w_C - c_C} \right) \\
0, & \text{if } Q_B < \hat{Q}_B.
\end{cases}
\]

In either case, \( r_A(Q_B, r_C(Q_B)) \) is non-decreasing in \( Q_B \).

The last statement in Proposition 3 is of particular interest. From (10) note that \( r_A(Q_B, Q_C) \) is actually non-increasing in \( Q_C \). By embedding supplier \( C \)'s best response in supplier \( A \)'s best response function, however, we obtain a version of component complementarity that will be useful when identifying and comparing capacity equilibria.
Equilibrium Analysis

Proposition 1 established the fact that, for any vector $w$ of wholesale prices, there exists at least one Nash equilibrium in the suppliers’ capacity game. Providing a detailed characterization of Nash equilibria in this game is complicated for two reasons. First, for any given vector of wholesale prices there always exist multiple equilibria. Second, for different wholesale prices different types of equilibria can arise. Despite this complexity, we are able to completely characterize the possible Nash equilibria in this game. In addition, we show that the game always has a unique Pareto-optimal Nash equilibrium, and we identify different forms that that equilibrium can take depending on relationships among the suppliers’ capacity costs and the wholesale prices. A formal statement of these results appears in Theorem 1 below.

The approach used to identify and compare all equilibria that can arise in each of the cases is based on a systematic consideration of all possible values of supplier $B$’s capacity $Q_B$. For each value of $Q_B$, we first evaluate supplier $C$’s best response $r_C(Q_B)$ and then supplier $A$’s best response $r_A(Q_B, r_C(Q_B))$. If the original $Q_B$ is in turn supplier $B$’s best response
to the other suppliers’ responses, then \((r_A(Q_B, r_C(Q_B)), Q_B, r_C(Q_B))\) is a Nash equilibrium. Otherwise no Nash equilibrium exists in which supplier \(B\) chooses that particular capacity \(Q_B\). After identifying all Nash equilibria, we compare the resulting profits for the three suppliers to identify preferences among the equilibria. This comparison is made possible by the complementarity among the capacity levels – i.e., the fact that \(r_C(Q_B)\) and \(r_A(Q_B, r_C(Q_B))\) are both non-decreasing in \(Q_B\).

The values of the three fractiles \(1 - \frac{c_A}{w_A}, \frac{c_B}{w_B}\), and \(\frac{c_C}{w_C}\), or equivalently, the three capacity quantities \(\hat{Q}_B, \hat{Q}_B^2\), and \(\hat{Q}_C^2\), play a key role in the characterization. We start by considering values of \(Q_B\) satisfying \(0 \leq Q_B \leq \min(\hat{Q}_B, \hat{Q}_B^2, \hat{Q}_C^2)\). In this case \(r_C(Q_B) = Q_B\) and \(r_A(Q_B, r_C(Q_B)) = 0\) (see (4) and Proposition 3). In addition, since \(Q_B \leq \hat{Q}_B^2\), Proposition 2 implies that \(r_B(0, Q_B) = 0 + Q_B = Q_B\), so that any vector \((0, Q_B, Q_B)\) with \(Q_B\) in the specified range is a Nash equilibrium. Next we compare these equilibria based on supplier profits. Clearly supplier \(A\) is indifferent since it earns zero profit in each case. Supplier \(C\)'s profit is increasing in \(Q_B\) in this range since \(Q_B \leq \hat{Q}_C^2\), so increasing \(Q_B\) is analogous to loosening an upper bound constraint in a newsvendor problem for supplier \(C\). As a result, supplier \(C\) prefers \(Q_B = \min(\hat{Q}_C^2, \hat{Q}_B^2, \hat{Q}_B)\). Finally, supplier \(B\)'s profit under any such equilibrium is \(\pi_B = w_B E[D_2][\min(Q_B, D_2)] - c_B Q_B\), which is concave and reaches its maximum at \(Q_B = \hat{Q}_B^2\). Thus supplier \(B\) also prefers \(Q_B = \min(\hat{Q}_C^2, \hat{Q}_B^2, \hat{Q}_B)\), so \((0, Q_B, Q_B)\) with \(Q_B = \min(\hat{Q}_C^2, \hat{Q}_B^2, \hat{Q}_B)\) is Pareto best among these equilibria.

For \(Q_B > \min(\hat{Q}_C^2, \hat{Q}_B^2, \hat{Q}_B)\) the types of equilibria that can occur and the form of the Pareto-optimal equilibrium depend on the relative values of the three fractiles (capacity quantities). A characterization of the resulting Pareto-optimal equilibrium for each possible case is given in Theorem 1. To facilitate this characterization for the cases with \(\hat{Q}_B = \min(\hat{Q}_B^2, \hat{Q}_B^2, \hat{Q}_C^2)\), we define the function

\[
I(Q_B) \equiv F_1(Q_B - r_C(Q_B)) + \int_{Q_B - r_C(Q_B)}^{r_{AB}(Q_B)} F_2(Q_B - x) dF_1(x),
\]

on the range \(\hat{Q}_B < Q_B < \hat{Q}_C^2 + \frac{F^{-1}_1\left(\frac{c_{AWC}}{w_A(w_C - c_C)}\right)}{F_1^{-1}\left(\frac{c_{AWC}}{w_A(w_C - c_C)}\right)}\). It is easy to verify that \(I(\hat{Q}_B) = 0\), \(I\left(\hat{Q}_C^2 + \frac{F^{-1}_1\left(\frac{c_{AWC}}{w_A(w_C - c_C)}\right)}{F_1^{-1}\left(\frac{c_{AWC}}{w_A(w_C - c_C)}\right)}\right) = 1 - \frac{c_{AWC}}{w_A(w_C - c_C)}\), and that \(I(\cdot)\) is increasing. (Note that when \(Q_A = r_{AB}(Q_B)\) and \(Q_C = r_C(Q_B)\), (7) can be written as \(I(Q_B) = 1 - \frac{c_A}{w_A} - \frac{c_B}{w_B}\).)

**Theorem 1.** For any vector \(w\) of wholesale prices, there exists a unique Pareto-optimal equilibrium \(Q^*(w)\) in the suppliers’ capacity game. This equilibrium can take one of the fol-
lowing forms, depending on the relationships among fractiles based on the suppliers’ capacity costs and the wholesale prices.

(i) If \( \max\{\frac{c_A}{w_C}, 1 - \frac{c_B}{w_A}\} \leq \frac{c_B}{w_B} \), then \((0, \hat{Q}_B^2, \hat{Q}_B^2)\) is the Pareto-optimal equilibrium.

(ii) If \( \max\{\frac{c_B}{w_B}, 1 - \frac{c_A}{w_A}\} \leq \frac{c_C}{w_C} \), then

\[
\left( \min\left( \frac{c_B w_C}{w_B c_C}, \hat{Q}_A^1 \right), \min\left( \frac{c_B w_C}{w_B c_C}, \hat{Q}_A^1 \right) + \hat{Q}_C^2, \hat{Q}_C^2 \right)
\]

is the Pareto-optimal equilibrium.

(iii) If \( \frac{c_A}{w_C} \leq \frac{c_B}{w_B} < 1 - \frac{c_A}{w_A} \), then the Pareto-optimal equilibrium is \((r_{AB}(Q_B), Q_B, r_C(Q_B))\), for the unique \(Q_B\) in the range

\[
\hat{Q}_B^2 < Q_B < \hat{Q}_C^2 + \hat{F}_1^{-1}\left( \frac{c_A}{w_A (w_C - c_C)} \right)
\]

satisfying \(I(Q_B) = 1 - \frac{c_A}{w_A} - \frac{c_B}{w_B}\). That equilibrium \(Q^*\) satisfies \(Q_B^* < Q_A^* + Q_C^*\).

(iv) If \( \frac{c_B}{w_B} < \frac{c_C}{w_C} < 1 - \frac{c_A}{w_A} \), then there are two possibilities:

(a) If \( \frac{c_B}{w_B} + \frac{c_A}{w_A} > \frac{c_A}{w_A (w_C - c_C)} \), then the Pareto-optimal equilibrium is \((r_{AB}(Q_B), Q_B, r_C(Q_B))\), for the unique \(Q_B\) in the range

\[
\hat{Q}_B < Q_B < \hat{Q}_C^2 + \hat{F}_1^{-1}\left( \frac{c_A}{w_A (w_C - c_C)} \right)
\]

satisfying \(I(Q_B) = 1 - \frac{c_A}{w_A} - \frac{c_B}{w_B}\). That equilibrium \(Q^*\) satisfies \(Q_B^* < Q_A^* + Q_C^*\).

(b) If \( \frac{c_B}{w_B} + \frac{c_A}{w_A} < \frac{c_A}{w_A (w_C - c_C)} \), then the Pareto-optimal equilibrium is

\[
\left( \min\left( \frac{c_B w_C}{w_B c_C}, \hat{Q}_A^1 \right), \min\left( \frac{c_B w_C}{w_B c_C}, \hat{Q}_A^1 \right) + \hat{Q}_C^2, \hat{Q}_C^2 \right)
\]

In addition to providing a characterization of the Pareto-optimal equilibrium for any possible relationship among costs and wholesale prices, the preceding result also illustrates the variety of different behaviors that can arise in this system. To facilitate discussion of these behaviors, we interpret the conditions in (i) - (iv) of Theorem 1 as defining regions in the wholesale-price space. If the parameters lie in region (i), then the equilibrium capacities include zero capacity for component A (resulting in zero assembly of product 1), so the system reduces to a single-product assembly setting. In the other cases, positive capacity is installed for all three components at equilibrium. However, regions (iii) and (iv)(a) exhibit risk pooling with respect to the common component (i.e., \(Q_B^* < Q_A^* + Q_C^*\)) while regions (ii) and (iv)(b) do not (i.e., \(Q_B^* = Q_A^* + Q_C^*\)).
3.3  Assembler’s Pricing Decision

Anticipating the suppliers’ (unique Pareto-optimal) equilibrium capacity vector \(Q^*(w)\) in response to a given wholesale price vector \(w\), the assembler selects \(w\) to maximize its own expected profit. In this section, we provide a partial characterization of the assembler’s optimal wholesale prices. We show that, under optimal assembler pricing, all or part of several of the regions in Theorem 1 can be ignored, thus significantly reducing the set of wholesale prices the assembler needs to consider. By reducing and combining regions, we define two new regions and show that the assembler’s optimal wholesale price vector lies in one of these, with prices in one region resulting in risk-pooling behavior in the suppliers’ capacity game and the other resulting in non-risk-pooling behavior.

Consider the four wholesale-price regions identified in Theorem 1 (i) - (iv). It is first easy to verify that \(Q^*(w)\) is continuous across the four regions. Now for any vector \(w\) in region (i), the assembler could decrease \(w_C\) until \(\frac{c_C}{w_C} = \frac{c_B}{w_B}\) and \(\hat{Q}^2_B = \hat{Q}^2_C\). This would increase the assembler’s margin without affecting the suppliers’ equilibrium capacity choices, so it would increase the assembler’s profit. A vector \(w\) in region (i) satisfying \(\frac{c_C}{w_C} = \frac{c_B}{w_B}\) also lies in region (ii) (i.e., on the boundary of regions (i) and (ii)). The assembler can thus ignore region (i) when selecting optimal prices. For any vector \(w\) in region (ii) such that \(\bar{F}^{-1}_1\left(\frac{c_B w_C}{w_B c_C}\right) \neq \hat{Q}^1_A\), the assembler could decrease either \(w_B\) or \(w_A\) (while remaining in the same region) until \(\bar{F}^{-1}_1\left(\frac{c_B w_C}{w_B c_C}\right) = \hat{Q}^1_A\), i.e., until \(\frac{c_B}{w_B} w_C = \frac{c_A}{w_A}\). Since this increases the assembler’s margin without affecting the suppliers’ equilibrium capacity choices, this again increases the assembler’s profit, and thus reduces region (ii) to one in which \(w\) satisfies \(1 - \frac{c_C}{w_C} \leq \frac{c_A}{w_A} = \frac{c_B}{w_B} w_C\). A similar argument applies to region (iv)(b) if \(\frac{c_B}{w_B} w_C < \frac{c_A}{w_A}\), i.e., in this case, the assembler decreases \(w_B\). On the other hand, if \(\frac{c_B}{w_B} w_C > \frac{c_A}{w_A}\), then as in region (ii) a decrease in \(w_A\) increases the assembler’s margin without affecting the suppliers’ capacity equilibrium, as long as the change in \(w_A\) keeps the wholesale price vector within the region defined in (iv)(b), i.e., as long as \(\frac{c_C}{w_C} \leq 1 - \frac{c_A}{w_A}\). If \(\frac{c_B}{w_B} w_C \leq 1 - \frac{c_C}{w_C}\), then \(w_A\) can be decreased all the way until \(\frac{c_A}{w_A} = \frac{c_B}{w_B} w_C \leq 1 - \frac{c_C}{w_C}\) while still staying within that region. If instead \(1 - \frac{c_C}{w_C} < \frac{c_B}{w_B} w_C\), then the assembler can profitably decrease \(w_A\) until \(\frac{c_A}{w_A} = 1 - \frac{c_C}{w_C} < \frac{c_B}{w_B} w_C\) while staying within region (iv)(b). Further decreases in \(w_A\) move the price vector into region (ii), so the argument for that region applies and the assembler can profitably reduce \(w_A\) until \(\frac{c_A}{w_A} = \frac{c_B}{w_B} w_C\). Combining these observations, among wholesale
prices in regions (i), (ii) or (iv)(b), the assembler needs only consider vectors \( w \) satisfying
\[
\frac{c_A}{w_A} = \frac{c_B}{w_B} \frac{w_C}{C_C},
\]
and any such vector results in equilibrium supplier capacities \((\hat{Q}_A, \hat{Q}_A + \hat{Q}_C, \hat{Q}_C)\) – i.e., non-risk-pooling behavior by the suppliers.

For wholesale prices in region (iv)(a), if \( Q_C^* = Q_B^* < \hat{Q}_C^2 \) then the assembler could increase its profit by reducing \( w_C \) until \( \hat{Q}_C^2 = Q_B^* \), thus increasing its margin without affecting the suppliers’ equilibrium capacity levels. Note that after such a change, \( \hat{Q}_C^2 = Q_B^* > \hat{Q}_B \), which implies \( \frac{c_C}{w_C} < 1 - \frac{c_A}{w_A} \). Also,
\[
1 - \frac{c_A}{w_A} \frac{w_C}{w_A(w_C - c_C)} = I \left( \hat{Q}_C^2 + \hat{F}_1^{-1} \left( \frac{c_A}{w_A} \frac{w_C}{w_A(w_C - c_C)} \right) \right) > I(\hat{Q}_C^2) = I(Q_B^*) = 1 - \frac{c_A}{w_A} - \frac{c_B}{w_B},
\]
which implies \( \frac{c_B}{w_B} + \frac{c_A}{w_A} > \frac{c_A}{w_A} \frac{w_C}{w_A(w_C - c_C)} \). So the new wholesale price vector remains in region (iv)(a). Thus, this region can be reduced by adding the condition \( I(\hat{Q}_C^2) \leq 1 - \frac{c_A}{w_A} - \frac{c_B}{w_B} \).

The resulting capacity equilibrium takes the form \( (r_{AB}(Q_B), Q_B, \hat{Q}_C^2) \), for the unique \( Q_B \) satisfying \( I(Q_B) = 1 - \frac{c_A}{w_A} - \frac{c_B}{w_B} \). A similar argument applies to wholesale prices in region (iii). As a result, that region can also be reduced by adding the condition \( I(\hat{Q}_C^2) \leq 1 - \frac{c_A}{w_A} - \frac{c_B}{w_B} \), and the resulting equilibrium also takes the form \( (r_{AB}(Q_B), Q_B, \hat{Q}_C^2) \). In addition, the conditions \( \frac{c_C}{w_C} \leq \frac{c_B}{w_B} < 1 - \frac{c_A}{w_A} \) in region (iii) imply the inequality \( \frac{c_B}{w_B} > \frac{c_A}{w_A} \frac{c_C}{w_A(w_C - c_C)} \). Combining regions (iii) and (iv)(a), we then have that the assembler needs only consider wholesale price vectors \( w \) that satisfy the following three conditions:
\[
\text{max} \left\{ \frac{c_C}{w_C}, \frac{c_B}{w_B} \right\} < 1 - \frac{c_A}{w_A}; \quad \frac{c_B}{w_B} > \frac{c_A}{w_A} \frac{c_C}{w_A(w_C - c_C)}; \quad I(\hat{Q}_C^2) \leq 1 - \frac{c_A}{w_A} - \frac{c_B}{w_B}.
\]

Any such prices result in the equilibrium capacity vector \( Q^* = \left( r_{AB}(Q_B), Q_B, \hat{Q}_C^2 \right) \), where \( Q_B \) is the unique capacity value satisfying \( I(Q_B) = 1 - \frac{c_A}{w_A} - \frac{c_B}{w_B} \), and this equilibrium exhibits risk-pooling behavior on the part of the suppliers – i.e., \( Q_B^* < Q_A^* + Q_C^* \).

In summary, solving the assembler’s pricing problem requires a pair of numerical searches – one over \( w \) satisfying (11) and one over \( w \) satisfying (12). The assembler can then compare profits associated with the best solution from each region and choose the better of the two. Note that in the process of this search the assembler may consider some vectors \( w \) such that the assumption on product margins no longer holds (i.e., \( p_1 - w_A - w_B > p_2 - w_B - w_C \)). For such \( w \), we simply “flip” the system by relabeling products 1 and 2 and components \( A \) and \( C \) so that the assumption holds once again before continuing with the analysis of the suppliers’ capacity game.
Interestingly, it is always optimal for the assembler to set prices such that supplier C’s equilibrium capacity $Q^*_C$ is equal to that supplier’s isolated optimal capacity $\hat{Q}^2_C$. This type of behavior occurs for all suppliers in a single-product decentralized assembly system (see Gerchak and Wang 2004). In contrast, in our setting, this occurs for supplier A only under non-risk-pooling outcomes, while it never occurs for supplier B. This difference in behavior is a direct result of the presence of a common component in the multi-product structure analyzed here.

4 Centralized System

In order to explore the impacts of decentralizing decision making in an assemble-to-order system, it is necessary to understand the behavior of the system under centralized control – i.e., where a single central planner controls all of the suppliers and the assembler. We analyze such a system in this section. All of the relevant assumptions from the previous section continue to apply, plus we assume without loss of generality that $p_1 \leq p_2$. Since there is no need for wholesale prices in this system, the central planner’s problem can be modeled as a two-stage stochastic program. Working backwards, the second stage occurs after the capacity vector $Q$ has been chosen and the demand vector $D$ observed. At this point the planner chooses the production vector $y$ that solves the following production mix problem.

$$\begin{align*}
\text{Max} \quad & \Pi(Q,D) = p_1 y_1 + p_2 y_2 \\
\text{subject to} \quad & y_1 \leq Q_A \quad (13) \\
& y_1 + y_2 \leq Q_B \quad (14) \\
& y_2 \leq Q_C \quad (15) \\
& 0 \leq y_1 \leq D_1 \\
& 0 \leq y_2 \leq D_2
\end{align*}$$

In the first stage, prior to observing demand the planner chooses the capacity vector $Q$ to maximize expected profit

$$V(Q) = E(\Pi(Q,D)) - c^T Q,$$

in anticipation of possible demand outcomes and the subsequent optimal production decision.

Variations on arguments used in the previous section can be used to show that the optimal capacity vector satisfies $Q_A \leq Q_B$, $Q_C \leq Q_B$, and $Q_B \leq Q_A + Q_C$. Given any
set of capacities satisfying those conditions, and any demand realization, we can obtain a closed-form expression for the optimal production quantities in the above linear program, as well as the optimal dual variables \( \lambda \) associated with the capacity constraints (13)-(15). (Similar to the decentralized system, if \( p_1 = p_2 \) we assume, without loss of optimality, that the decision maker gives priority to product 2.) In fact, the demand space \( \mathbb{R}_{\geq 0}^{2} \) can be partitioned into five regions, where all points in a given region have common expressions for their optimal primal and dual variables. The boundaries of the regions, as well as the corresponding optimal variables, are as follows:

\[
\Omega_0 = \{ Q : D_1 \leq Q_A, D_2 \leq Q_C, D_1 + D_2 \leq Q_B \}, \quad y = \begin{pmatrix} D_1 \\ D_2 \end{pmatrix}, \quad \lambda = 0
\]

\[
\Omega_1 = \{ Q : D_1 \leq Q_B - Q_C, D_2 \geq Q_C \}, \quad y = \begin{pmatrix} D_1 \\ Q_C \end{pmatrix}, \quad \lambda = \begin{pmatrix} 0 \\ 0 \\ p_2 \end{pmatrix}
\]

\[
\Omega_2 = \{ Q : D_1 \geq Q_B - Q_C, D_2 \geq Q_C \}, \quad y = \begin{pmatrix} Q_B - Q_C \\ Q_C \end{pmatrix}, \quad \lambda = \begin{pmatrix} 0 \\ p_1 \\ p_2 - p_1 \end{pmatrix}
\]

\[
\Omega_3 = \{ Q : D_1 \geq Q_B - Q_C, Q_B - Q_A \leq D_2 \leq Q_C, D_1 + D_2 \geq Q_B \}, \quad y = \begin{pmatrix} Q_B - D_2 \\ D_2 \end{pmatrix}, \quad \lambda = \begin{pmatrix} 0 \\ p_1 \\ 0 \end{pmatrix}
\]

\[
\Omega_4 = \{ D_1 \geq Q_A, D_2 \leq Q_B - Q_A \}, \quad y = \begin{pmatrix} Q_A \\ D_2 \end{pmatrix}, \quad \lambda = \begin{pmatrix} p_1 \\ 0 \\ 0 \end{pmatrix}
\]

In the remaining analysis of the capacity choice problem, we explicitly include the constraints \( Q_A \leq Q_B, Q_C \leq Q_B, \) and \( Q_B \leq Q_A + Q_C \). By linear programming theory \( V(Q) \) is concave and so the Karush-Kuhn-Tucker conditions are necessary and sufficient for optimality. The following result formally characterizes the optimal solution to the capacity choice problem.

**Proposition 4.** A capacity investment vector \( Q^0 = (Q_A^0, Q_B^0, Q_C^0)^T \in \mathbb{R}_{\geq 0}^{3} \) is optimal if and only if there exists a \( v = (v_A, v_B, v_C)^T \in \mathbb{R}_{\geq 0}^{3} \) and a \( \bar{\mu} \) such that

\[
0^T P_D(\Omega_1(Q^0)) + \begin{pmatrix} 0 \\ 0 \\ p_1 \end{pmatrix} P_D(\Omega_2(Q^0)) + \begin{pmatrix} 0 \\ p_1 \\ 0 \end{pmatrix} P_D(\Omega_3(Q^0))
\]

\[
+ \begin{pmatrix} p_1 \\ 0 \\ 0 \end{pmatrix} P_D(\Omega_4(Q^0)) = c - v + \bar{\mu}
\]

\[
v^T Q^0 = 0, \quad \mu_1(Q_A^0 + Q_C^0 - Q_B^0) = 0, \quad \mu_2(Q_B^0 - Q_C^0) = 0, \quad \mu_3(Q_B^0 - Q_A^0) = 0.
\]
The first-order conditions in Proposition 4 can be expressed as

\[ F_2(Q_B^0 - Q_A^0) \tilde{F}_1(Q_A^0) = \frac{c_A - v_A - \mu_1 + \mu_3}{p_1}, \quad (17) \]

\[ 1 - F_1(Q_B^0 - Q_C^0) - F_2(Q_B^0 - Q_A^0) \tilde{F}_1(Q_A^0) - \int_{Q_B^0 - Q_C^0}^{Q_A^0} F_2(Q_B^0 - x) dF_1(x) = \frac{c_B - v_B + \mu_1 - \mu_2 - \mu_3}{p_1}, \quad (18) \]

\[ \bar{F}_2(Q_C^0) \left( 1 - \frac{p_1}{p_2} \tilde{F}_1(Q_B^0 - Q_C^0) \right) = \frac{c_C - v_C - \mu_1 + \mu_2}{p_2}. \quad (19) \]

plus the complementary slackness conditions. The following result establishes some properties of the centralized optimal solution.

**Theorem 2.** The optimal capacity investment strategy for the centralized system has the following properties:

(i) It is always optimal to invest in all three components and \( Q^0 \) solves (16) with \( v = 0 \).

(ii) Under the optimal capacity vector \( Q^0 \), \( Q_A^0 < Q_B^0 \) always holds, and \( \mu_3 = 0 \).

(iii) The optimal capacity satisfies the boundary condition \( Q_B^0 = Q_A^0 + Q_C^0 \), with \( Q_A^0 = F_1^{-1}(\frac{c_A + c_B}{p_1}) \) and \( Q_C^0 = \bar{F}_2^{-1}(\frac{c_B + c_C}{p_2}) \), if and only if

\[ \frac{(c_A + c_B)}{p_1} \frac{(c_B + c_C)}{p_2} \geq \frac{c_B}{p_1}. \quad (20) \]

Otherwise, \( Q_B^0 < Q_A^0 + Q_C^0 \).

The non-risk-pooling condition (20) is related to the optimality condition for a simple newsvendor problem. Recall that such a condition prescribes a quantity \( Q \) such that the probability of stocking out is equal to the ratio (marginal cost of increasing \( Q \))/(marginal revenue obtained from the sale of one more unit). (See, for example, the expressions for \( Q_A^0 \) and \( Q_C^0 \) in part (iii) of the theorem.) Now fix \( Q_A = Q_A^0 \) and \( Q_C = Q_C^0 \), and consider choosing \( Q_B = Q_A^0 + Q_C^0 \). Given that capacity choice, the probability of stocking out of component \( B \) is equal to \( P_D(D_1 \geq Q_A^0, D_2 \geq Q_C^0) = F_1(Q_A^0)F_2(Q_C^0) \), which equals the left-hand side of (20). Also, the marginal cost of increasing \( Q_B \) is equal to \( c_B \), while the marginal revenue obtained from the last unit of capacity added is equal to \( p_1 \) (since product 1 receives lower priority). Thus the ratio of the two is equal to the right-hand side of (20). If (20) holds, then it is economically attractive to add that final unit of component \( B \) capacity - i.e., to
forgo risk pooling - while if it does not hold, it is better to keep component $B$ capacity lower and to take advantage of risk pooling.

5 Impacts of Decentralization

In this section, we compare various aspects of the ATO systems under centralized and decentralized decision making. Some of these comparisons are derived from analytical results, while others are based on observations from a numerical study.

The numerical study examines a large number of specific problems generated by varying the capacity cost parameters and finished product demand distributions. In all scenarios studied, the demands for finished products are two independent random variables with Normal distributions truncated at zero to avoid negative demand realizations. (Specifically, starting with a Normal distribution with cdf $G_j(\cdot)$ with mean and standard deviation as described below, demand has a distribution with cdf $F_j(x) = 0$ for $x < 0$ and $F_j(x) = (G_j(x) - G_j(0))/(1 - G_j(0))$ for $x \geq 0$.) Product prices for all cases are $p_1 = 14$ and $p_2 = 15$. We consider seven sets of 12 scenarios each corresponding to $c_B = 1, \ldots, 12$, while $c_A = c_C = 1$ in all scenarios. The parameters for the Normal distributions are given in the table below.

<table>
<thead>
<tr>
<th>Set</th>
<th>$\mu_1$</th>
<th>$\sigma_1$</th>
<th>$\mu_2$</th>
<th>$\sigma_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20</td>
<td>6</td>
<td>20</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>12</td>
<td>20</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>3</td>
<td>30</td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>3</td>
<td>30</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>30</td>
<td>9</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>30</td>
<td>12</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>30</td>
<td>15</td>
<td>10</td>
<td>3</td>
</tr>
</tbody>
</table>

A final set of eleven scenarios differs from the fourth set only in that $c_A = c_C = 2$, thus limiting $c_B = 1, \ldots, 11$ to maintain the profitability of both products.

5.1 Capacity Levels

When the suppliers independently select their capacity levels, the assembler needs to set the wholesale prices so that the suppliers earn a positive margin for each component assembled into an end product while, at the same time, ensuring that it is profitable to assemble and
sell those finished products to consumers. It is well known that this double marginalization effect usually leads to lower capacity levels than are optimal under centralized control. (See, for example, Lariviere 1998 and Gerchak and Wang 2004.) Not surprisingly, this type of inefficiency is also present in our ATO system. Indeed, in all numerical scenarios, the decentralized system led to lower component capacity levels than in the centralized system.

We can show analytically that this is always the case when no risk pooling of the common component occurs in both the centralized and the decentralized systems. In the decentralized system, when the optimal assembler wholesale prices result in no risk pooling, the equilibrium component capacity levels are given by

\[ Q^*_C = \bar{F}_2^{-1} \left( \frac{c_C}{w_C} \right), \quad Q^*_A = \bar{F}_1^{-1} \left( \frac{c_A}{w_A} \right), \quad \text{and} \quad Q^*_B = Q^*_A + Q^*_C. \]

In the no-risk-pooling case, the centralized optimal capacity levels are

\[ Q^*_A = \bar{F}_1^{-1} \left( \frac{c_A + c_B}{p_1} \right), \quad Q^*_C = \bar{F}_2^{-1} \left( \frac{c_B + c_C}{p_2} \right), \quad \text{and} \quad Q^*_B = Q^*_A + Q^*_C. \]

As a result, \( Q^*_A > Q^*_A \) if and only if \( \frac{c_A}{w_A} > \frac{c_A + c_B}{p_1} \). The latter inequality is equivalent to \( p_1 - w_A - w_B > 0 \) and \( w_C \geq c_C \). Similarly, \( Q^*_C < Q^*_C \), and thus \( Q^*_B < Q^*_B \) as well.

5.2 Risk Pooling of Common Component B

As mentioned above, the tendency of decentralized systems to understock has been reported in prior research on simpler systems. In this section and the next, we explore two new types of inefficiencies that only arise in systems with the multi-product, common-component structure analyzed here. This section addresses decentralization’s effect on the incidence of risk pooling of the common component.

As shown in Theorem 2(iii), we can characterize those settings in which risk pooling with respect to component B capacity is optimal in the centralized system. The next result provides an alternative version of that characterization, which focuses on the effect of \( c_B \) on risk pooling. (The result follows from (20) and simple algebra, so we omit its proof.)

**Corollary 1.** In the centralized system:

(i) If \( (p_2 - c_C - c_A)^2 - 4c_AC < 0 \), it is optimal to set the capacity level for component B equal to the sum of those of components A and C (no risk pooling), for any value of \( c_B \).

(ii) If \( (p_2 - c_C - c_A)^2 - 4c_AC \geq 0 \), then there exist \( 0 < c^*_B \leq c^*_{\text{max}} < p_2 - c_C \) such that risk pooling of component B occurs for \( c_B \) in the interval \([c^*_B, c^*_{\text{max}}]\).

Corollary 1 suggests that, in the centralized system, risk pooling of common component B is either never optimal (i.e., regardless of the value of \( c_B \)) or it arises for intermediate values
of $c_B$ only. That is, risk pooling is less likely to be optimal for low values of $c_B$, since in those cases it is reasonably inexpensive to match the aggregate capacity levels of components $A$ and $C$. Also, risk pooling is less likely for relatively high values of $c_B$, since in those cases the optimal capacity level for the common component tends to be small relative to the distributions of demands for finished products, and so the same holds for the capacity levels of the dedicated components, since each of those is an economic complement to component $B$. For relatively low capacity levels, the benefit of matching the aggregate capacity levels of the dedicated components outweighs the capacity investment costs for component $B$.

Based on the numerical study, similar conclusions can be made for the decentralized system. That is, depending on the problem parameters, either risk pooling of the common component did not occur for any value of $c_B$, or risk pooling was optimal for values of $c_B$ in an interval of intermediate $c_B$ values. However, in all numerical cases this interval was within (and narrower than) $[c_B^{\text{min}}, c_B^{\text{max}}]$. Thus, we can make the following observations, based on the numerical findings.

**Observation 1.** Risk pooling of the common component $B$ is less common in the decentralized system than in the centralized system:

(i) If risk pooling of the common component did not occur in the centralized system, then it did not occur in the decentralized system under the assembler's optimal pricing.

(ii) In 87% of scenarios where risk pooling of the common component occurred in the optimal solution of the centralized system, risk pooling did not occur in the decentralized system under the assembler's optimal pricing.

Our numerical study suggests a few possible explanations for this reduced frequency of risk pooling in the decentralized setting. First, the low component capacity levels associated with understocking inhibit risk pooling. When capacities are low, the probability of consuming all units of components $A$ and $C$ is high so there is little benefit from risk pooling of component $B$. A second explanation is related to the shift in incentives that results from decentralization. In a centralized setting risk pooling is often desirable, and the optimal level of risk pooling is determined by trading off the cost savings from reducing capacity against some loss in expected sales. In a decentralized setting, supplier $B$ faces a trade off of this type and may wish to risk pool, whereas the assembler experiences only the negative aspect of risk pooling – the possible loss of sales of product 1, since this product has the lower priority. One way that the assembler could achieve some cost savings to compensate
for this possible drop in sales would be to reduce the price paid to supplier A, and thus that supplier’s capacity. This may prompt supplier B to further reduce its own capacity, thus exacerbating the understocking. Of course, this approach by the assembler also involves a trade off – the cost savings from reducing supplier A’s price versus an additional potential reduction in sales of product 1 – and thus risk pooling will still occur in some decentralized settings.

The numerical study also allows us to obtain insights into what factors tend to favor risk pooling in the decentralized system. By comparing sets 1 and 2, 3 and 4, and 5 through 7, we consistently find that risk pooling is more common when demand variability is higher. By comparing sets 1, 3 and 5, we can assess the impact of demand asymmetry on risk pooling (in all of these scenarios, both products have the same demand coefficient of variation). We find that risk pooling is more likely to occur in set 3, which corresponds to low demand for product 1 and high demand for product 2, since in this case the impact of risk pooling on sales of product 1 is lower. Lastly, by comparing the final set of eleven scenarios with the fourth set, we find that higher values of \( c_A \) and \( c_C \) are associated with less frequent risk pooling. These higher parameter values usually lead to lower capacities for suppliers A and C so, similar to the understocking argument, there is little benefit from risk pooling.

Finally, we explore the degree of risk pooling when it occurs in either the decentralized or the centralized systems. To that end, we utilize the capacity imbalance factor

\[
\gamma = \frac{Q_A + Q_C - Q_B}{Q_B}
\]

as a measure of the degree of capacity imbalance or risk pooling (see Van Mieghem 2003). Note that \( \gamma = 0 \) corresponds to the case where there is no risk pooling of the common component. The numerical study suggests that the capacity imbalance factor can be higher in the decentralized system than in the centralized system, even though for many values of the parameter \( c_B \) the imbalance factor is zero in the decentralized system and positive in the centralized system. That is, risk pooling is less common in the decentralized system, but when it occurs, the component capacities in the decentralized system may be more imbalanced – implying a higher degree of risk pooling – than in the integrated system. This is illustrated in Figure 4 below, which corresponds to the scenarios with \( \mu_1 = 10, \sigma_1 = 3, \mu_2 = 30, \sigma_2 = 9 \). Note, for example, that for \( c_B = 5 \), the imbalance factor in the decentralized system is 23.7%, while in the centralized system it is 13.9%.
5.3 Component Commonality

The previous section assumed the presence of a common component and focused on whether or not the system made use of it through risk pooling. This section asks a slightly different question: Does the flexibility represented by a common component (vs. two dedicated components) always make the assembler better off (or at least not worse off)? To explore this question, consider a variation of the ATO system in which the two products use only dedicated components. That is, product 1 (product 2) is obtained by assembling a unit of component \( A \) (component \( C \)) and a unit of component \( B_1 \) (component \( B_2 \)). Components \( B_1 \) and \( B_2 \) may be identical and simply reserved for use in specific product configurations, or they may be slightly different but perform similar or equal functions in the final product. In the latter case, we can think of the common component \( B \) as a re-engineered version of components \( B_1 \) and \( B_2 \) that can be used for both final products.

The two components \( B_1 \) and \( B_2 \) are sourced from different suppliers. This allows the assembler to offer two (possibly different) wholesale prices for these components. We denote by \( Q_{Bi} \) the capacity of component \( Bi \) for product \( i, i = 1, 2 \), and by \( w_{Bi} \) the corresponding wholesale prices paid by the assembler. In addition, we let \( c_{Bi} \) be the unit capacity cost for production of component \( i \), and we assume that \( c_{Bi} = c_B, i = 1, 2 \).

The two subsystems (corresponding to products 1 and 2) are independent of each other, and each one represents a single-product assembly system. The equilibrium capacity choices and optimal assembler wholesale prices for these systems under centralized and decentralized control have already been studied, see e.g., Gerchak and Wang (2004). For reference here, we summarize the relevant results for such systems in the following proposition. The proof

![Figure 4: Degree of Capacity Imbalance](image)

Figure 4: Degree of Capacity Imbalance
follows from Propositions 3 and 4 in Gerchak and Wang (2004).

**Proposition 5.** Consider the two single-product assembly systems described above, in which product 1 is assembled from components A and B and product 2 from components B2 and C. Under decentralized decision making:

(i) For any given vector of wholesale prices \( w = (w_A, w_{B1}, w_{B2}, w_C) \), the Pareto optimal capacity equilibrium is 
\[
Q_A^* = Q_{B1}^* = \min \left\{ \bar{F}_1^{-1} \left( \frac{c_A}{w_A} \right), \bar{F}_1^{-1} \left( \frac{c_{B1}}{w_{B1}} \right) \right\}
\]
for product 1, and 
\[
Q_C^* = Q_{B2}^* = \min \left\{ \bar{F}_2^{-1} \left( \frac{c_{B2}}{w_{B2}} \right), \bar{F}_2^{-1} \left( \frac{c_C}{w_C} \right) \right\}
\]
for product 2.

(ii) It is optimal for the assembler to set wholesale prices so that 
\[
\frac{c_A}{w_A} = \frac{c_{B1}}{w_{B1}} \quad \text{and} \quad \frac{c_{B2}}{w_{B2}} = \frac{c_C}{w_C}.
\]
This allows us to express the assembler’s profit as a function of the capacities \( Q_A = Q_{B1} \) and \( Q_C = Q_{B2} \). Furthermore, if demand distributions have increasing failure rates, the assembler’s profit function is concave in \( Q_A \) and \( Q_C \).

In the centralized system:

(iii) The optimal capacity levels are given by 
\[
Q_A^0 = Q_{B1}^0 = \bar{F}_1^{-1} \left( \frac{(c_A + c_B)}{p_1} \right)
\]
and 
\[
Q_C^0 = Q_{B2}^0 = \bar{F}_2^{-1} \left( \frac{(c_B + c_C)}{p_2} \right).
\]

In comparing the systems with and without a common component, note first that under centralized control, component commonality is always preferred in terms of total system profit. Indeed, \( Q_A = Q_A^0 \), \( Q_B = Q_{B1}^0 + Q_{B2}^0 \), \( Q_C = Q_C^0 \) is always feasible in the system with a common component and leads to the same profit as in the system with two dedicated components. Thus, the optimal solution with a common component yields at least as much profit. In fact, if no risk pooling is optimal in the centralized setting, the two systems yield identical solutions and profits, while if risk pooling is optimal, the common component system yields strictly higher profit.

Under decentralized decision making, we now investigate how a system with a common component compares to one with two dedicated components in terms of profitability for the assembler. Interestingly, if the system with a common component exhibits no risk pooling with respect to the common component (i.e., \( Q_B = Q_A + Q_C \)), then the assembler is always better off having two dedicated capacities for the two products. To see this, let \( w^*_A, w^*_B \) and \( w^*_C \) be the optimal wholesale prices in the system with a common component \( B \). From Section 3.3, we have that

\[
\frac{c_A}{w^*_A} \frac{c_C}{w^*_C} = \frac{c_B}{w^*_B} \quad (21)
\]
Now for the system with two dedicated components $B_1$ and $B_2$, consider the vector of wholesale prices $(w_A, w_{B_1}, w_{B_2}, w_C) = (w_A^*, c_B w_A^*/c_A, c_B w_C^*/c_C, w_C^*)$. This choice of wholesale prices leads to the same equilibrium capacity levels for components $A$ and $C$, and to equilibrium capacity levels for components $B_1$ and $B_2$ whose sum is equal to the equilibrium capacity level for component $B$ in the system with a common component. However, from (21), we have that $c_C/w_C^* = w_{B_i}/w_B^*$ and $c_A/w_A^* = w_{B_2}/w_B^*$, implying that $w_{B_i} < w_B^*$ for $i = 1, 2$. That is, the assembler can induce the same capacity equilibrium in the system with dedicated components with lower wholesale prices for components $B_1$ and $B_2$. Thus, the assembler’s profit is higher in the system with dedicated components.

Our numerical study suggests that a system with dedicated components may also be preferred by the assembler when risk pooling of component $B$ occurs in the system with a common component. In such cases, the cost to the assembler of lower capacity of component $B$ (due to risk pooling) may outweigh the flexibility benefit of the common component. This trade off is not present in the centralized system since all costs and benefits are internalized.

6 Conclusions

In this paper we analyzed a decentralized assemble-to-order system in which independent suppliers control component capacity decisions while the assembler decides what component prices to offer and how many of each finished product to produce. We showed that, despite its inherent complexity, the system is reasonably well behaved. Specifically, for any wholesale prices set by the assembler there exists a unique Pareto-optimal equilibrium in the suppliers’ capacity game. Also, the assembler’s optimal wholesale prices lie in one of two regions, and these regions result in different behaviors in the subsequent capacity game – one region leads to risk pooling with respect to the common component, while the other does not. In order to explore the impact of decentralization on the behavior of assemble-to-order systems, we also analyzed optimal/equilibrium capacity/pricing decisions for two other variations of our basic system: one with a centralized decision maker and one where all components are product-specific. By comparing behavior in the three systems, we identified several types of inefficiencies that can arise from decentralization. Similar to other decentralized supply chain settings studied previously, we found that decentralization leads to understocking in terms of component capacities. In addition to this, however, we identified new types of inefficiencies that are more directly related to the multi-component, multi-product setting
studied here. First, we found that risk pooling with respect to the common component occurs less frequently in the decentralized system. Second, we found that the apparent flexibility provided by a common component may actually hurt performance in a decentralized system whereas it can only help in a centralized system.

The system studied here is a simple one, with only two finished products and three components. However, the results appear to extend to more complex systems in some special cases. For example, the analysis can be extended to the case in which each of the two products has multiple dedicated components associated with it. In addition, the basic analysis here would appear to extend to systems with more than two products that share one common component. While realistic systems are usually more complex, the results for this simple model provide valuable insights into the forces at work when assemble-to-order systems are combined with decentralized decision making.

References


Appendix: Proofs

Proof of Proposition 1. For given $Q$ and $D$, $s_i(Q, D) = \min\{Q, Q^0_i(Q-\epsilon, D)\}$, where $Q^0_A(Q_B, Q_C, D) = \min(D_1, Q_B - \min(Q_B, Q_D, D_2))$, $Q^0_B(Q_A, Q_C, D) = \min(D_1, Q_A) + \min(D_2, Q_C)$, and $Q^0_C(Q_B, D) = \min(D_2, Q_B)$. Thus, $\Pi_i(Q_i|Q-\epsilon, D)$ is piecewise linear and concave, since

$$\Pi_i(Q_i|Q-\epsilon, D) = \begin{cases} (w_i - c_i)Q_i, & \text{for } 0 \leq Q_i \leq Q^0_i(D, Q-\epsilon) \\ w_iQ^0_i(D, Q-\epsilon) - c_iQ_i, & \text{for } Q_i > Q^0_i(D, Q-\epsilon) \end{cases}$$

(22)

This implies that $\Pi_i(Q_i|Q-\epsilon) = E_D[\Pi_i(Q_i|Q-\epsilon, D)]$ is also concave. The existence of a Nash equilibrium then follows from Theorem 1.2 in Fudenberg and Tirole (1991).

Differentiability of the expected sales functions

The following result can be used to establish the differentiability of supplier $i$’s expected sales with respect to that supplier’s capacity choice.

Lemma 2. Let $S_i(Q) = E_D[\min(Q_i, g(Q-\epsilon, D))]$, where $Q \in \mathbb{R}^{m}_{\geq 0}$, $D$ is an $n$-dimensional non-negative random variable with joint probability distribution $P_D$, and $g : \mathbb{R}^{m-1}_{\geq 0} \times \mathbb{R}^n_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ is continuous. Let $Q$ be fixed. If $Q_i$ is such that $P_D(\{D : Q_i = g(Q-\epsilon, D)\}) = 0$, then $S_i(Q_i, Q-\epsilon)$ is differentiable at $Q_i$ and

$$\frac{\partial S_i}{\partial Q_i} = P_D(\{D : Q_i \leq g(Q-\epsilon, D)\}).$$

Proof. Define the sets

$$\Delta_i = \{D : Q_i \leq g(D, Q-\epsilon)\},$$

$$\Delta_i(\epsilon) = \{D : Q_i + \epsilon \leq g(D, Q-\epsilon)\},$$

$$\Delta'_i(\epsilon) = \{D : Q_i \leq g(D, Q-\epsilon) < Q_i + \epsilon\} = \Delta_i \setminus \Delta_i(\epsilon).$$

Note that $\Delta_i(0) = \Delta_i$,

$$\tilde{\Delta}_i(\epsilon) = \Delta'_i(\epsilon) \cup \Delta_i \text{ and } \Delta'_i(\epsilon) \cap \tilde{\Delta}_i = \emptyset,$$

(23)

where $\tilde{A}$ is the complement of the set $A$,

$$\Delta_i = \Delta_i(\epsilon) \cup \Delta'_i(\epsilon) \text{ and } \Delta_i(\epsilon) \cap \Delta'_i(\epsilon) = \emptyset.$$  

(24)
We intend to compute
\[
\frac{\partial S_i}{\partial Q_i} = \lim_{\epsilon \to 0} \frac{S(Q_i + \epsilon, Q_{-i}) - S(Q_i)}{\epsilon}.
\]
Consider the limit as \(\epsilon \to 0^+\). (The proof for the left limit is similar.) For a set \(A\), define \(1_A(x) = 1\) if \(x \in A\) and \(1_A(x) = 0\) if \(x \not\in A\). Note that
\[
\frac{1}{\epsilon} [S(Q_i + \epsilon, Q_{-i}) - S(Q_i)] =
\frac{1}{\epsilon} \left[ E_D [(Q_i + \epsilon) 1_{\Delta_i}] + E_D [g(Q_{-i}, D) 1_{\Delta_i}] - E_D [Q_i 1_{\Delta_i}] - E_D [g(Q_{-i}, D) 1_{\Delta_i}] \right] =
\frac{1}{\epsilon} \left[ (Q_i + \epsilon) P_D(\Delta_i) + E_D [g(Q_{-i}, D) 1_{\Delta_i}] - Q_i P_D(\Delta_i) - Q_i P_D(\Delta_i') \right] =
P_D(\Delta_i) + \frac{1}{\epsilon} E_D [(g(Q_{-i}, D) - Q_i) 1_{\Delta_i'}],
\]
where the second equality follows from (23) and (24). Note that \(\Delta_i' = \{D : 0 \leq g(D, Q_{-i}) - Q_i < \epsilon\}\), which implies that
\[
0 \leq \frac{1}{\epsilon} E [(g(D, Q_{-i}) - Q_i) 1_{\Delta_i'}] \leq P_D(\Delta_i').
\]
Also, note that
\[
\lim_{\epsilon \to 0^+} P_D(\Delta_i') = \cap_{\epsilon > 0} \{D : 0 \leq g(D, Q_{-i}) - Q_i < \epsilon\} = P_D(\{D : Q_i = g(D, Q_{-i})\}),
\]
and
\[
\lim_{\epsilon \to 0^+} P_D(\Delta_i) = \cap_{\epsilon > 0} \{D : Q_i + \epsilon \leq g(D, Q_{-i})\} = P_D(\Delta_i).
\]
If \(P_D(\{D : Q_i = g(D, Q_{-i})\}) = 0\), then (25) and (26) imply that
\[
\lim_{\epsilon \to 0^+} \frac{S(Q_i + \epsilon, Q_{-i}) - S(Q_i)}{\epsilon} = P_D(\Delta_i).
\]
Supplier \(C\)'s expected sales function is simple enough that Lemma 2 is not required (though it does apply). To compute the derivative of supplier \(B\)'s expected sales function, note that
\[
E_D [s_B(Q_B, Q_{-B}, D)] = E_D [\min(Q_B, g(Q_{-B}, D))],
\]
where \(g(Q_{-B}, D) = \min(D_1, Q_A) + \min(D_2, Q_C)\). The condition of Lemma 2 is verified for \(Q_B < Q_A + Q_C\), noting that \(D\) is a non-negative random variable, since
\[
P_D(\{D : Q_B = g(Q_{-B}, D)\}) = P_D(\{D : Q_B = D_1 + D_2, Q_B - Q_C \leq D_1 \leq Q_A\})
+ P_D(\{D : D_2 = Q_B - Q_A, D_1 > Q_A\})
+ P_D(\{D : D_1 = Q_B - Q_C, D_2 > Q_C\}) = 0.
\]
Differentiability of supplier $A$’s expected sales function is addressed in Lemma 1.

**Proof of Proposition 2.** Fix the values of $Q_A$ and $Q_C$. It is easy to verify that the left-hand side of (7) is decreasing in $Q_B$, and that as $Q_B$ approaches $Q_A + Q_C$, it approaches $\bar{F}_1(Q_A)\bar{F}_2(Q_C)$. As a result, supplier $B$’s optimal capacity choice is $Q_B = Q_A + Q_C$ if and only if $\bar{F}_1(Q_A)\bar{F}_2(Q_C) \geq \frac{c_B}{w_B}$. For values of $Q_A$ and $Q_C$ with $\bar{F}_1(Q_A)\bar{F}_2(Q_C) < \frac{c_B}{w_B}$, supplier $B$’s best response is given by the solution to (7).

Note that, leaving $Q_A$ fixed,

$$\frac{\partial r_{BAC}}{\partial Q_C} = \frac{f_1(Q_B - Q_C)\bar{F}_2(Q_C)}{f_2(Q_B - Q_A)\bar{F}_1(Q_A) + f_1(Q_B - Q_C)\bar{F}_2(Q_C) + \int_{Q_B - Q_C}^{Q_A} f_2(Q_B - x)dF_1(x)}, \quad (27)$$

which implies that $0 < \frac{\partial r_{BAC}}{\partial Q_C} < 1$. Similarly, when $Q_C$ is fixed, we have that

$$\frac{\partial r_{BAC}}{\partial Q_A} = \frac{f_2(Q_B - Q_A)\bar{F}_1(Q_A)}{f_2(Q_B - Q_A)\bar{F}_1(Q_A) + f_1(Q_B - Q_C)\bar{F}_2(Q_C) + \int_{Q_B - Q_C}^{Q_A} f_2(Q_B - x)dF_1(x)}, \quad (28)$$

implying that $0 < \frac{\partial r_{BAC}}{\partial Q_C} < 1$. \[\blacksquare\]

**Proof of Lemma 1.** First, for $Q_A < Q_B - Q_C$, supplier $A$’s profit function is given by $\Pi_A = w_A E_D (\min(Q_A, D_1)) - c_A Q_A$, so the first half of (8) follows immediately. For $Q_A > Q_B - Q_C$, we have from (8) that

$$\frac{\partial \Pi_A}{\partial Q_A} = w_A \frac{\partial E_D[s_A(Q, D)]}{\partial Q_A} - c_A.$$

Note that $E_D[s_A(Q, D)] = E_D [\min(Q_A, g(Q_B, Q_C, D))]$, where

$$g(Q_B, Q_C, D) = \min (Q_B - \min(Q_B, Q_C, D_2), D_1) = \min (Q_B - \min(Q_C, D_2), D_1),$$

since $Q_C \leq Q_B$. The condition of Lemma 2 is verified for $Q_A \leq Q_B$ and $Q_A > Q_B - Q_C$, since

$$P_D (\{D : Q_A = g(Q_B, Q_C, D)\}) = P_D (\{D : Q_A = D_1, D_2 \leq Q_B - Q_A\}) + P_D (\{D : Q_A = Q_B - D_2, D_1 \geq Q_A\}) = 0.$$Then,

$$\frac{\partial E_D[s_A(Q, D)]}{\partial Q_A} = P_D (Q_A \leq \min(Q_B - \min(Q_C, D_2), D_1))$$

$$= P_D (Q_A \leq Q_B - \min(Q_C, D_2)) \bar{F}_1(Q_A)$$

$$= P_D (Q_A \leq Q_B - D_2 \text{ and } D_2 \leq Q_C) \bar{F}_1(Q_A)$$

$$= P_D (D_2 \leq Q_B - Q_A) \bar{F}_1(Q_A).$$
Proof of Proposition 3. First, if \( \frac{c_A}{w_A} + \frac{c_C}{w_C} \geq 1 \), we have that \( Q_B \geq \hat{Q}_C^2 \). For \( Q_B \leq \hat{Q}_C^2 \), \( r_C(Q_B) = Q_B \), which implies that \( F_2(r_C(Q_B)) \hat{F}_1(Q_B - r_C(Q_B)) = F_2(Q_B) \leq \frac{c_A}{w_A} \) since \( Q_B \leq \hat{Q}_B \), and \( \frac{c_A}{w_A} < 1 = \hat{F}_1(Q_B - r_C(Q_B)) \). As a result, for any \( Q_B \leq \hat{Q}_C^2 \), \( r_A(Q_B, r_C(Q_B)) = Q_B - r_C(Q_B) \). For \( \hat{Q}_A^2 < Q_B \leq \hat{Q}_A^1 + \hat{Q}_C^2 \), \( r_C(Q_B) = \hat{Q}_C^2 \), which implies that \( F_2(r_C(Q_B)) \hat{F}_1(Q_B - r_C(Q_B)) = \left( 1 - \frac{c_C}{w_C} \right) \hat{F}_1(Q_B - \hat{Q}_C^2) < 1 - \frac{c_C}{w_C} \leq \frac{c_A}{w_A} \), and \( \hat{F}_1(Q_B - r_C(Q_B)) = \hat{F}_1(Q_B - \hat{Q}_C^2) \geq \hat{F}_1(\hat{Q}_A^1) = \frac{c_A}{w_A} \). As a result, \( r_A(Q_B, r_C(Q_B)) = Q_B - \hat{Q}_C^2 \). For \( Q_B > \hat{Q}_A^1 + \hat{Q}_C^2 \), \( \hat{F}_1(Q_B - \hat{Q}_C^2) < \hat{F}_1(\hat{Q}_A^1) = c_A/w_A \), so \( r_A(Q_B, r_C(Q_B)) = \hat{Q}_A^1 \).

Next, if \( \frac{c_A}{w_A} + \frac{c_C}{w_C} < 1 \), we have that \( \hat{Q}_B < \hat{Q}_C^2 \). Similar to above, we have that for \( Q_B < \hat{Q}_B \), \( r_C(Q_B) = Q_B \) and \( r_A(Q_B, r_C(Q_B)) = Q_B - r_C(Q_B) \). If \( \hat{Q}_B \leq Q_B < \hat{Q}_C^2 \), then \( r_C(Q_B) = Q_B \) and \( \frac{c_A}{w_A} \leq F_2(Q_B) \hat{F}_1(Q_B - r_C(Q_B)) = F_2(Q_B) \), so that \( r_A(Q_B, r_C(Q_B)) = \hat{Q}_A^2 \). Similarly, if \( \hat{Q}_C^2 \leq Q_B < \hat{Q}_B \), then \( \frac{c_A}{w_A} \leq F_2(Q_B) \hat{F}_1(Q_B - \hat{Q}_C^2) \geq \hat{F}_1(\hat{Q}_A^1) = \frac{c_A}{w_A} \), so that \( r_A(Q_B, r_C(Q_B)) = \hat{Q}_A^2 \). Finally, if \( \hat{Q}_B^2 + F_1^{-1}(\frac{c_A}{w_A} \frac{w_C}{w_C - c_C}) \leq Q_B \leq \hat{Q}_A^1 + \hat{Q}_C^2 \), then \( r_C(Q_B) = \hat{Q}_B^2 \), \( \frac{c_A}{w_A} \geq F_2(\hat{Q}_B^2) \hat{F}_1(Q_B - \hat{Q}_B^2) \), and \( \frac{c_A}{w_A} \leq \hat{F}_1(Q_B - \hat{Q}_B^2) \), and as a result \( r_A(Q_B, r_C(Q_B)) = Q_B - \hat{Q}_B^2 \). Again, as above, \( r_A(Q_B, r_C(Q_B)) = \hat{Q}_A^1 \) for \( Q_B > \hat{Q}_A^1 + \hat{Q}_C^2 \).

Finally, note that

\[
0 < \frac{\partial r_{AB}(Q_B)}{\partial Q_B} = \frac{f_2(Q_B - Q_A) \hat{F}_1(Q_A)}{f_2(Q_B - Q_A) \hat{F}_1(Q_A) + F_2(Q_B - Q_A) f_1(Q_A)} < 1,
\]

which implies that \( r_{AB}(Q_B) > 0 \) for all \( Q_B > \hat{Q}_B \). Also,

\[
r_{AB} \left( \hat{Q}_C^2 + \frac{c_A}{w_A} \frac{w_C}{w_C - c_C} \right) = \frac{c_A}{w_A} \frac{w_C}{w_C - c_C}.
\]

Then, \( r_A(Q_B, r_C(Q_B)) \) is non-decreasing in \( Q_B \). ■

Proof of Theorem 1. The range \( 0 \leq Q_B \leq \min(\hat{Q}_B, \hat{Q}_B^2, \hat{Q}_C^2) \) was analyzed in the discussion preceding the theorem statement in the text. For values of \( Q_B > \hat{Q}_A^1 + \hat{Q}_C^2 \), we have that \( r_C(Q_B) = \hat{Q}_C^2 \) and \( r_A(Q_B, r_C(Q_B)) = \hat{Q}_A^1 \) (see Proposition 3), so that there cannot be an equilibrium in that range since \( r_B(Q_A, Q_C) \leq Q_A + Q_C \). Thus, we only need to analyze the range \( \min(\hat{Q}_B, \hat{Q}_B^2, \hat{Q}_C^2) < Q_B \leq \hat{Q}_A^1 + \hat{Q}_C^2 \), and we do so by considering four different cases with respect to the fractiles \( 1 - \frac{c_A}{w_A}, \frac{c_B}{w_B} \) and \( \frac{c_C}{w_C} \).

(i) First suppose \( \frac{c_C}{w_C} < 1 - \frac{c_A}{w_A} \leq \frac{c_B}{w_B} \), which is equivalent to \( \hat{Q}_B^2 \leq Q_B < \hat{Q}_C^2 \). For \( \hat{Q}_B^2 < Q_B < \hat{Q}_B \), \( r_C(Q_B) = Q_B \) and \( r_A(Q_B, r_C(Q_B)) = 0 \), but \( r_B(0, Q_B) = Q_B \) if and only if
\( \hat{F}_1(0) \hat{F}_2(Q_B) = \hat{F}_2(Q_B) \geq \frac{c_B}{w_B} \) if and only if \( Q_B \leq \hat{Q}_B^2 \). Therefore, there is no equilibrium with \( \hat{Q}_B^2 < Q_B \leq \hat{Q}_B \). For \( \hat{Q}_B \leq Q_B \leq \hat{Q}_C^2 \), \( r_C(Q_B) = Q_B \) and \( r_A(Q_B, r_C(Q_B)) = r_{AB}(Q_B) > 0 \). Then, in order for such a \( Q_B \) to be part of an equilibrium, we must have 
\[ Q_B = r_B(r_{AB}(Q_B), Q_B) = r_{BAC}(r_{AB}(Q_B), Q_B), \]
where the last equality follows since \( Q_B > \hat{Q}_B^2 \) implies \( \hat{F}_1(r_{AB}(Q_B)) \hat{F}_2(Q_B) \leq \hat{F}_2(Q_B) < \hat{F}_2(\hat{Q}_B^2) = \frac{c_B}{w_B}. \) Then, from (7), we would need 
\[ 1 - F_2(Q_B - r_{AB}(Q_B)) \hat{F}_1(r_{AB}(Q_B)) - \frac{c_B}{w_B} > 0. \]
The latter cannot hold since \( 1 - \frac{c_A}{w_A} \leq \frac{c_B}{w_B} \), and by the definition of \( r_{AB}(Q_B) \), \( F_2(Q_B - r_{AB}(Q_B)) \hat{F}_1(r_{AB}(Q_B)) = \frac{c_A}{w_A}. \) Thus, there is no equilibrium in this range either. For \( \hat{Q}_B^2 < Q_B < \hat{Q}_C^2 + \hat{F}_1^{-1}\left( \frac{c_B}{w_A w_C - c_C} \right), r_C(Q_B) = \hat{Q}_C^2 \), but otherwise the preceding argument applies here as well, so there is no equilibrium in this range.

For \( \hat{Q}_C^2 + \hat{F}_1^{-1}\left( \frac{c_B}{w_A w_C - c_C} \right) \leq Q_B \leq \hat{Q}_A^2 + \hat{Q}_C^2 \), \( r_C(Q_B) = \hat{Q}_C^2 \) and \( r_A(Q_B, r_C(Q_B)) = Q_B - \hat{Q}_C^2 \), so any equilibrium must be of the form \( (Q_B - \hat{Q}_C^2, Q_B, \hat{Q}_C^2) \) and from Proposition 2 we must have 
\[ \hat{F}_1(Q_B - \hat{Q}_C^2) \hat{F}_2(\hat{Q}_C^2) = \hat{F}_1(Q_B - \hat{Q}_C^2) \frac{c_C}{w_B} \leq \frac{c_B}{w_B}. \]
But \( \hat{F}_1(Q_B - \hat{Q}_C^2) \hat{F}_2(\hat{Q}_C^2) < \hat{F}_2(\hat{Q}_C^2) = \frac{c_B}{w_B} \), where the last inequality follows since \( Q_B < \hat{Q}_C^2 \).

Suppose instead \( 1 - \frac{c_A}{w_A} \leq \frac{c_C}{w_C} \leq \frac{c_B}{w_B} \), which is equivalent to \( \hat{Q}_B^2 \leq \hat{Q}_C^2 \leq \hat{Q}_B \). For \( \hat{Q}_B^2 < Q_B \leq \hat{Q}_C^2, \ r_C(Q_B) = Q_B \) and \( r_A(Q_B, r_C(Q_B)) = 0 \). At the same time, \( r_B(0, Q_B) = Q_B \) if and only if \( \hat{F}_1(0) \hat{F}_2(Q_B) = \hat{F}_2(Q_B) \geq \frac{c_B}{w_B} \), but this cannot hold since \( Q_B > \hat{Q}_B^2 \), so there is no equilibrium in this range. For \( \hat{Q}_B^2 < Q_B \leq \hat{Q}_A^2 + \hat{Q}_C^2 \), \( r_C(Q_B) = \hat{Q}_C^2 \) and \( r_A(Q_B, r_C(Q_B)) = \hat{Q}_C^2 \), but again \( \hat{F}_1(Q_B - \hat{Q}_C^2) \hat{F}_2(\hat{Q}_C^2) < \frac{c_B}{w_B} \), which implies that there is no equilibrium in this range.

Thus, for max \( \left\{ \frac{c_C}{w_C}, 1 - \frac{c_A}{w_A} \right\} \leq \frac{c_B}{w_B} \), \( (0, \hat{Q}_B^2, \hat{Q}_B^2) \) is the Pareto-optimal equilibrium.

(ii) Suppose now that \( 1 - \frac{c_A}{w_A} \leq \frac{c_B}{w_B} < \frac{c_C}{w_C} \), which is equivalent to \( \hat{Q}_C^2 < \hat{Q}_B^2 \leq \hat{Q}_B \). For
\[ \hat{Q}_C^2 < Q_B \leq \hat{Q}_B^2 + \min\left( \hat{F}_1^{-1}\left( \frac{c_B}{w_B c_C} \right), \hat{Q}_A^1 \right), \]
\( r_C(Q_B) = \hat{Q}_C^2 \) and \( r_A(Q_B, r_C(Q_B)) = Q_B - \hat{Q}_C^2 \). At the same time, \( r_B(Q_B - \hat{Q}_C^2, \hat{Q}_C^2) = Q_B \) if and only if \( \frac{c_B}{w_B} \leq \hat{F}_1(Q_B - \hat{Q}_C^2) \hat{F}_2(\hat{Q}_C^2) = \hat{F}_1(Q_B - \hat{Q}_C^2) \frac{c_C}{w_B} \) which is equivalent to \( Q_B \leq \hat{Q}_B^2 + \hat{F}_1^{-1}\left( \frac{c_B}{w_B c_B} \right) \). As a result, for all \( Q_B \) in the range (31) the vectors \( (Q_B - \hat{Q}_C^2, Q_B, \hat{Q}_C^2) \) are equilibria. If \( \hat{F}_1^{-1}\left( \frac{c_B}{w_B c_C} \right) < \hat{Q}_A^1 \), then we also need to consider the range \( \hat{Q}_C^2 + \hat{F}_1^{-1}\left( \frac{c_B}{w_B c_C} \right) < Q_B \leq \hat{Q}_A^1 + \hat{Q}_C^2 \). However, there is no equilibrium in this range since \( r_C(Q_B) = \hat{Q}_C^2 \) and \( r_A(Q_B, \hat{Q}_C^2) = Q_B - \hat{Q}_C^2 \) but \( r_B(Q_B - \hat{Q}_C^2, \hat{Q}_C^2) = r_{BAC}(Q_B - \hat{Q}_C^2, \hat{Q}_C^2) < Q_B \) by Proposition 2 and the definition of \( r_{BAC} \). Then, we need to compare equilibria of the form \( (Q_B - \hat{Q}_C^2, Q_B, \hat{Q}_C^2) \) for \( Q_B \) in the range
\[ \hat{Q}_C^2 < Q_B \leq \hat{Q}_B^2 + \min\left( \hat{F}_1^{-1}\left( \frac{c_B}{w_B c_C} \right), \hat{Q}_A^1 \right). \]
Supplier $C$ is clearly indifferent. For supplier $A$, note that its profit for any equilibrium of this form is $\pi_A = w_A E_{D_1} [\min(Q, D_1)] - c_A Q$ for $Q = Q_B - \hat{Q}_C^2$ and $Q_B$ in the range (32). Thus, $0 \leq Q \leq \min \left( F_1^{-1} \left( \frac{c_B w_C}{w_B c_C} \right), \hat{Q}_A^1 \right)$, and since $\pi_A$ is concave and reaches its maximum at $\hat{Q}_A^1$, $Q_B = \hat{Q}_C^2 + \min \left( F_1^{-1} \left( \frac{c_B w_C}{w_B c_C} \right), \hat{Q}_A^1 \right)$ is preferred by supplier $A$. Similarly, supplier $B$’s profit in this range is $\pi_B = w_B E_{D_2} \left[ \min(D_1, Q_B - \hat{Q}_C^2) + \min(D_2, \hat{Q}_C^2) \right] - c_B Q_B$, which is concave and reaches its maximum at $Q_B = \hat{Q}_C^2 + \hat{Q}_B^1$. Since $\hat{Q}_B^1 = F_1^{-1} \left( \frac{c_B w_C}{w_B c_C} \right) \geq F_1^{-1} \left( \frac{c_B w_C}{w_B c_C} \right)$, supplier $B$ prefers the same equilibrium.

If $\frac{c_B}{w_B} < 1 - \frac{c_A}{w_A} \leq \frac{c_C}{w_C}$, which is equivalent to $\hat{Q}_C^2 < \bar{Q}_B < \hat{Q}_B^2$, the analysis is exactly the same as above.

Thus, for $\max \left\{ \frac{c_B}{w_B}, 1 - \frac{c_A}{w_A} \right\} \leq \frac{c_C}{w_C}$,

\[
\left( \min \left( F_1^{-1} \left( \frac{c_B w_C}{w_B c_C} \right), \hat{Q}_A^1 \right), \min \left( F_1^{-1} \left( \frac{c_B w_C}{w_B c_C} \right), \hat{Q}_A^1 \right) + \hat{Q}_C^2, \hat{Q}_C^2 \right)
\]

is the unique Pareto optimal equilibrium.

(iii) Suppose that $\frac{c_C}{w_C} \leq \frac{c_B}{w_B} < 1 - \frac{c_A}{w_A}$, which is equivalent to $\bar{Q}_B < \hat{Q}_B^2 \leq \hat{Q}_C^2$. Note that for $Q_B$ in the range $\bar{Q}_B < Q_B < \hat{Q}_C^2$ we have $r_A(Q_B, r_C(Q_B)) + r_C(Q_B) = r_{AB}(Q_B) + Q_B > Q_B$. Similarly, for $Q_B$ in the range $\hat{Q}_C^2 < Q_B < \bar{Q}_B < \hat{Q}_B^2 + F_1^{-1} \left( \frac{c_A w_C}{w_A (w_C - c_C)} \right)$ we have $r_A(Q_B, r_C(Q_B)) + r_C(Q_B) = r_{AB}(Q_B) + \hat{Q}_C^2 > Q_B$, since $r_{AB}(Q_B) + \hat{Q}_C^2 = Q_B$ when $Q_B = \hat{Q}_C^2 + \bar{Q}_B^1 \left( \frac{c_A w_C}{w_A (w_C - c_C)} \right)$ and $\partial Q_{AB}/\partial Q_B < 1$ for $\bar{Q}_B < Q_B < \hat{Q}_B^2 + F_1^{-1} \left( \frac{c_A w_C}{w_A (w_C - c_C)} \right)$. As a result, for $Q_B$ in the range $\bar{Q}_B < Q_B < \hat{Q}_C^2 + F_1^{-1} \left( \frac{c_A w_C}{w_A (w_C - c_C)} \right)$, there cannot be an equilibrium with $Q_B = Q_A + Q_C$. So, any equilibrium in this range must have $Q_B = r_B(r_A(Q_B), r_C(Q_B)) = r_{BAC}(r_A(Q_B), r_C(Q_B))$, or equivalently,

\[
I(Q_B) = 1 - \frac{c_A}{w_A} - \frac{c_B}{w_B}.
\]

(33)

It is easy to verify that $I(\cdot)$ is strictly increasing for $Q_B$ in this range, that

\[
I \left( \hat{Q}_B^2 \right) \leq F_2 \left( \hat{Q}_B^2 \right) F_1 \left( r_{AB} \left( \hat{Q}_B^2 \right) \right) < \left( 1 - \frac{c_B}{w_B} \right) \left( 1 - \frac{c_A w_C}{w_A (w_C - c_C)} \right) \leq 1 - \frac{c_A}{w_A} - \frac{c_B}{w_B},
\]

and that

\[
I \left( \hat{Q}_C^2 + \bar{Q}_B^1 \left( \frac{c_A w_C}{w_A (w_C - c_C)} \right) \right) = 1 - \frac{c_A w_C}{w_A (w_C - c_C)} > 1 - \frac{c_A}{w_A} - \frac{c_B}{w_B},
\]

since $\frac{c_C}{w_C} \leq \frac{c_B}{w_B} < 1 - \frac{c_A}{w_A}$. So there is no solution to (33), and thus no equilibrium with $\bar{Q}_B < Q_B \leq \hat{Q}_B^2$. However, there exists a unique solution $Q_B$ to (33) in the range $\hat{Q}_B^2 < Q_B < \hat{Q}_C^2 + F_1^{-1} \left( \frac{c_A w_C}{w_A (w_C - c_C)} \right)$, and this corresponds to the unique equilibrium

\[
(r_{AB}(Q_B^*), r_{BAC}(r_{AB}(Q_B^*), r_C(Q_B^*)), r_C(Q_B^*))
\]

37
in that range. For \( \hat{Q}_C^2 + F_1^{-1}\left(\frac{c_A w_C}{w_A(w_C - c_C)}\right) \leq Q_B \leq \hat{Q}_A^1 + \hat{Q}_C^2 \), the analysis is as in (i), and there is no equilibrium in that range. Then, we need to compare \((0, \bar{Q}_B, \bar{Q}_B)\) with \((r_{AB}(Q_B^*), r_{BAC}(Q_B^*), r_C(Q_B^*)), r_C(Q_B^*))\). Supplier A is better off under the latter equilibrium, since it earns a positive profit. Since \(Q_B \geq Q_C\) in both cases, supplier C’s expected profit is \(\Pi_C = w_C E_{D_2}[\min(Q_C, D_2)] - c_C Q_C\), and is concave and maximized at \(\bar{Q}_C^2\). Since \(\bar{Q}_B < \bar{Q}_B^2 \leq \min(Q_B^*, \bar{Q}_C^2) = r_C(Q_B^*) \leq \bar{Q}_C^2\), supplier C also prefers the latter equilibrium. For supplier B, its profit is increasing in \(Q_A\) and \(Q_C\). Then, since \(r_C(Q_B^*) \geq \bar{Q}_B\), \(\Pi_B(0, Q_B, Q_B) \geq \Pi_B(r_{AB}(Q_B^*), \bar{Q}_B, r_C(Q_B^*)) \leq \Pi_B(r_{AB}(Q_B^*), r_{BAC}(r_{AB}(Q_B^*), r_C(Q_B^*)), r_C(Q_B^*))\), where the last inequality follows since \(r_{BAC}(r_{AB}(Q_B^*), r_C(Q_B^*))\) is supplier B’s best response to the other suppliers’ capacities. Thus, \((r_{AB}(Q_B^*), r_{BAC}(r_{AB}(Q_B^*), r_C(Q_B^*)), r_C(Q_B^*))\) is the unique Pareto-optimal equilibrium in this region.

(iv) Finally, suppose that \(\frac{c_A}{w_B} \leq \frac{c_C}{w_C} < 1 - \frac{c_A}{w_A}\), which is equivalent to \(\bar{Q}_B < \hat{Q}_C^2 \leq Q_B^2\). We prove parts (a) and (b) together by considering a sequence of ranges of \(Q_B\) between \(\bar{Q}_B\) and \(\hat{Q}_A^1 + \hat{Q}_C^2\). Consider first the range

\[
\hat{Q}_B < Q_B < \hat{Q}_C^2 + F_1^{-1}\left(\frac{c_A w_C}{w_A(w_C - c_C)}\right). \tag{34}
\]

As in (iii), there is a unique Nash equilibrium of the form \((r_{AB}(Q_B^*), r_{BAC}(Q_B^*), r_C(Q_B^*)), r_C(Q_B^*))\), if there is a \(Q_B^*\) in the range (34) satisfying \(I(Q_B) = 1 - \frac{c_A}{w_A} - \frac{c_B}{w_B}\). Otherwise, there is no equilibrium in this range.

If

\[
\frac{c_A w_C}{w_A(w_C - c_C)} < \frac{c_A}{w_A} + \frac{c_B}{w_B}, \tag{35}
\]

then, again as in (iii), such a solution exists. If (35) does not hold, then there is no equilibrium for \(Q_B\) in the range (34), since in that case \(I(Q_B) < 1 - \frac{c_A}{w_A} - \frac{c_B}{w_B}\) for all \(Q_B < \hat{Q}_C^2 + F_1^{-1}\left(\frac{c_A w_C}{w_A(w_C - c_C)}\right)\).

For the range

\[
\hat{Q}_C^2 + F_1^{-1}\left(\frac{c_A w_C}{w_A(w_C - c_C)}\right) \leq Q_B \leq \hat{Q}_A^1 + \hat{Q}_C^2, \tag{36}
\]

\(r_A(Q_B, r_C(Q_B)) = Q_B - \hat{Q}_C^2\) and \(r_C(Q_B) = \hat{Q}_C^2\). So any equilibrium must be of the form \((Q_B - \hat{Q}_C^2, Q_B, \hat{Q}_C^2)\). From Proposition 2, a capacity \(Q_B\) in (36) leads to such an equilibrium if and only if

\[
\hat{F}_1(Q_B - \hat{Q}_C^2) \hat{F}_2(\hat{Q}_C^2) = \hat{F}_1(Q_B - \hat{Q}_C^2) \frac{c_C}{w_C} \geq \frac{c_B}{w_B},
\]

38
or, equivalently, \( Q_B \leq \hat{Q}_C^2 + \bar{F}_1^{-1}\left(\frac{c_B \, w_C}{w_B \, c_C}\right) \). It is easy to verify that (35) is equivalent to
\[
\frac{c_{AWC}}{w_A(w_C - c_C)} < \frac{c_B \, w_C}{w_B \, c_C},
\]
and so if (35) holds no such \( Q_B \) exists in the range (36). However, if (35) does not hold, then each \( Q_B \) in the non-empty interval
\[
\hat{Q}_C^2 + \bar{F}_1^{-1}\left(\frac{c_A w_C}{w_A(w_C - c_C)}\right) \leq Q_B \leq \hat{Q}_C^2 + \min\left\{ \hat{Q}_A^1, \bar{F}_1^{-1}\left(\frac{c_B \, w_C}{w_B \, c_C}\right) \right\}
\]
leads to such an equilibrium. Furthermore, no equilibrium exists in the range \( \hat{Q}_C^2 + \min\left\{ \hat{Q}_A^1, \bar{F}_1^{-1}\left(\frac{c_B \, w_C}{w_B \, c_C}\right) \right\} < Q_B \leq \hat{Q}_A^1 + \hat{Q}_C^2 \).

If (35) holds then, following a similar argument as in (iii), the equilibrium \((r_{AB}(Q_B^*), r_{BAC}(Q_B^*), r_C(Q_B^*), r_C(Q_B^*))\) identified above is the unique Pareto optimal equilibrium in the wholesale price region (iv)(a). If (35) does not hold, then following a similar argument as in (ii),
\[
\left( \min \left( \bar{F}_1^{-1}\left(\frac{c_B \, w_C}{w_B \, c_C}\right), \hat{Q}_A^1 \right) \right), \min \left( \bar{F}_1^{-1}\left(\frac{c_B \, w_C}{w_B \, c_C}\right), \hat{Q}_A^1 \right) + \hat{Q}_C^2, \hat{Q}_C^2
\]
is the unique Pareto optimal equilibrium in the wholesale price region (iv)(b).

**Proof of Proposition 4.** We define \( v = (v_A, v_B, v_C)^T \) as the dual variable corresponding to \( Q \geq 0 \), \( \mu_1 \) as the dual variable corresponding to \( Q_B - (Q_A + Q_C) \leq 0 \), \( \mu_2 \) as the dual variable corresponding to \( Q_B - Q_A \geq 0 \). We also define
\[
L(Q) = \mu_1(Q_A + Q_C - Q_B) + \mu_2(Q_B - Q_C) + \mu_3(Q_B - Q_A),
\]
and note that \(-\nabla_Q L(Q) = (\mu_3 - \mu_1, \mu_1 - \mu_2 - \mu_3, \mu_2 - \mu_1)^T\). The first-order conditions can then be written as \( \nabla_Q E(\Pi(Q, D)) = c - v - \nabla_Q L(Q) = c - v + \bar{\mu} \). Harrison and Van Mieghem (1998) show that differentiation and expectation can be interchanged so that \( \nabla_Q E(\Pi(Q, D)) = E_D(\lambda(Q, D)) \). In addition, since the dual variables are constant in each of the regions \( \Omega_i, i = 1, ..., 5 \), \( E_D(\lambda(Q, D)) \) can be expressed as the left-hand side of (16).

**Proof of Theorem 2.** (i) Clearly, \( Q = 0 \) can never be optimal since the central planner can always choose a capacity vectors \( Q = (\epsilon, \epsilon, 0) \) or \( (0, \epsilon, \epsilon) \), with \( \epsilon \) small enough, and earn a positive profit. Similarly, \( Q_B = 0 \) or \( Q_A = Q_C = 0 \) are never optimal. Also, if only two capacity levels are positive, clearly it would be optimal to set them equal. First assume
\[Q^0 = (0, Q, Q)\] with \(Q > 0\). In this case, \(v_B = v_C = \mu_3 = 0\), so (17)-(19) become:

\[
F_2(Q) = \frac{c_A - v_A - \mu_1}{p_1},
\]

\[
\bar{F}_2(Q) = \frac{c_B + \mu_1 - \mu_2}{p_1},
\]

\[
F_2(Q) \left(1 - \frac{p_1}{p_2}\right) = \frac{c_C - v_C - \mu_1 + \mu_2}{p_2}.
\]

Substituting (37) into (38) yields \(p_1 - c_A - c_B = -v_A - \mu_2 \leq 0\) which contradicts our assumption that \(p_1 > c_A + c_B\). Next, assume \(Q^0 = (Q, Q, 0)\) with \(Q > 0\). In this case, \(v_A = v_B = \mu_2 = 0\) and (17)-(19) become:

\[
0 = \frac{c_A - \mu_1 + \mu_3}{p_1},
\]

\[
F_1(Q) = \frac{c_B + \mu_1 - \mu_3}{p_1},
\]

\[
1 - \frac{p_1}{p_2} \bar{F}_1(Q) = \frac{c_C - v_C - \mu_1}{p_2}.
\]

Substituting (39) into (40) yields \(p_2 - c_B - c_C = -v_C - \mu_3 \leq 0\) which contradicts our assumption that \(p_2 > c_B + c_C\). Thus, the centralized system will have \(Q^0 > 0\).

(ii) If the vector \(Q^0 = (Q, Q, Q)\) with \(Q > 0\) and \(Q_C > 0\) was optimal for the centralized system, we would then have that \(v_A = v_B = v_C = \mu_1 = 0\), and (17) would become 0 = \(\frac{c_A + \mu_3}{p_1}\). However, \(c_A = -\mu_3 \leq 0\) contradicts the fact that \(c_A > 0\). So for the optimal \(Q^0\), \(Q_A > Q_B\).

(iii) Assume that \(Q^0_B = Q^0_A + Q^0_C\). Then, \(v_A = v_B = v_C = \mu_2 = \mu_3 = 0\), and (17)-(19) become

\[
F_2(Q^0_C) \bar{F}_1(Q^0_A) = \frac{c_A - \mu_1}{p_1},
\]

\[
\bar{F}_1(Q^0_A) \bar{F}_2(Q^0_C) = \frac{c_B + \mu_1}{p_1},
\]

\[
\bar{F}_2(Q^0_C) \left(1 - \frac{p_1}{p_2} \bar{F}_1(Q^0_A)\right) = \frac{c_C - \mu_1}{p_2}.
\]

Solving (41)-(43) yields \(\bar{F}_1(Q^0_A) = \frac{c_A + \mu_3}{p_1}\), \(F_2(Q^0_C) = \frac{(c_A - \mu_1)(c_B + c_C)}{p_2(c_B + \mu_1)}\) and \(\bar{F}_2(Q^0_C) = \frac{c_B + c_C}{p_2}\). The fact that \(F_2(Q^0_C) + \bar{F}_2(Q^0_C) = 1\) implies that \(\mu_1 = \frac{(c_A + c_B)(c_A + c_B) - p_2 c_B}{p_2} \geq 0\), which is equivalent to \(\frac{(c_A + c_B)(c_B + c_C)}{p_2 c_B} \geq 1\).

Conversely, assume now that \(\frac{(c_A + c_B)(c_B + c_C)}{p_2 c_B} \geq 1\), but \(Q^0_B < Q^0_A + Q^0_C\). Replacing the values of \(c_A, c_B\) and \(c_C\) from (16) into (20), rearranging the terms, and using the fact that
\(v_A = v_B = v_C = \mu_1 = \mu_3 = 0\), we obtain the following inequality

\[
\mu_2 (p_1 P_D(\Omega_3) + p_2 P_D(\Omega_1 + \Omega_2)) \geq \\
p_2 p_1 (P_D(\Omega_2 + \Omega_3) - P_D(\Omega_1 + \Omega_2) P_D(\Omega_2 + \Omega_3 + \Omega_4)) + \mu_2 p_2.
\]

(To simplify notation, we write \(\Omega_i(Q^0)\) as \(\Omega_i\) for the remainder of this proof.) We first show that

\[
p_2^2 P_D(\Omega_3) P_D(\Omega_2 + \Omega_3 + \Omega_4) < p_2 p_1 (P_D(\Omega_2 + \Omega_3) - P_D(\Omega_1 + \Omega_2) P_D(\Omega_2 + \Omega_3 + \Omega_4)). \tag{44}
\]

Indeed, we have that \(p_1 \leq p_2\) and \(P_D(\Omega_2 + \Omega_3 + \Omega_4) < 1\), which imply that the left hand side of (44) is less than \(p_1 p_2 P_D(\Omega_3)\). Also, it is immediate to verify that \(P_D(\Omega_1 + \Omega_2) P_D(\Omega_2 + \Omega_3 + \Omega_4) < \bar{F}_1(Q^0_B - Q^0_C) \bar{F}_2(Q^0_C) = P_D(\Omega_2)\). This implies that \(P_D(\Omega_2 + \Omega_3) - P_D(\Omega_1 + \Omega_2) P_D(\Omega_2 + \Omega_3 + \Omega_4) > P_D(\Omega_2 + \Omega_3) - P_D(\Omega_2) > P_D(\Omega_3)\), which shows that the inequality in (44) is valid. Finally, note that

\[
\mu_2 (p_1 P_D(\Omega_3) + p_2 P_D(\Omega_1 + \Omega_2)) \leq \mu_2 p_2 (P_D(\Omega_1 + \Omega_2 + \Omega_3)) \leq \mu_2 p_2.
\]

Thus, we arrive at a contradiction. \(\blacksquare\)