Incentives and Commonality in a Decentralized Multi-Product Assembly System

Fernando Bernstein Gregory A. DeCroix Yulan Wang

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

Forthcoming in *Operations Research*

In this paper, we explore the impact of decentralized decision making on the behavior of multi-product assembly systems. Specifically, we consider a system where three components (two product-specific and one common) are used to produce two end products to satisfy stochastic customer demands. We study the system under both centralized and decentralized decision making. In the decentralized system we prove that, for any set 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 capacity imbalance and one does not. We use these results to derive insights regarding the inefficiencies that decentralization can cause in such systems. In particular, several of our findings indicate that outsourcing the management of component supplies may inhibit the use of operational hedging approaches for managing uncertainty.

1 Introduction

Over the years, firms producing multiple finished products have explored a variety of ways to provide responsive service to customers while keeping inventories as low as possible. One approach widely adopted in recent years has been the assemble-to-order approach, with Dell Computer being perhaps the best known example. By using many of its components across multiple product lines, Dell can offer tremendous product variety while holding inventory of only a limited number of components. This use of component commonality is just one example of how operational flexibility can be leveraged to improve supply chain responsiveness while avoiding excessive inventories. This approach and others based on the same principle have received significant attention in the academic literature in recent years, and have been successfully implemented in a number of companies. Some examples include operational hedging using flexible production capacity (e.g., Seagate Technologies – see Van Mieghem 1998b) and postponement (e.g., Hewlett-Packard – see Kopczak and Lee 1996).

In addition to their inventory/production strategy (achieving flexibility by stocking only components and producing finished goods only once an order is received), most assemble-to-order systems also involve some level of decentralization. For example, most of the component inventories that Dell uses in 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. Shortages in the supply of components can have a serious negative impact on the performance of supply chains. In 2000, basic memory chips and some high-frequency transistors used in cellular-phone manufacturing were in short supply. Building capacity for some of these components can take up to 18 months, making it difficult for end-product assemblers to bring new products to market in a timely fashion. (See Hilsenrath 2000.)

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, particularly the impact of decentralization on the use of commonality and hedging strategies in such systems.

We analyze an 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 must be made prior to a single selling season, while finished-product production decisions are made after observing demands during the season. Demands for the finished products are stochastic, and unfilled demands are lost. We study this system under both centralized and decentralized decision making. In the decentralized version, the assembler first sets wholesale prices it will pay for each unit of component it purchases. After observing those prices, three independent suppliers simultaneously choose how much capacity to install or reserve for each of their components. Then the assembler observes demand, selects the finished product production quantities, and places the corresponding orders with the suppliers. In the centralized version, all decisions are made by a single decision maker.

For the centralized system we present a characterization of the unique optimal solution. For the decentralized system, we show that for any choice of wholesale prices by the assembler, there exists a unique Pareto-optimal equilibrium in the suppliers' capacity game. We show that without loss of optimality the assembler can restrict attention to two price regions, one of which leads to capacity imbalance in the capacity game (i.e., situations where the common component is strictly less than the sum of the capacities of the two dedicated components, a version of operational hedging), and one of which leads to balanced capacities.

To explore the impact of decentralization on the ATO system, we compare behavior in the centralized and decentralized systems both analytically and numerically, leading to a number of interesting insights. Similar to many other supply chain settings (e.g., Lariviere and Porteus 2001, Wang and Gerchak 2003, Bernstein and DeCroix 2006), we find that decentralization leads to understocking in terms of component capacities. However, we also identify new types of inefficiencies that are specifically related to the multi-component, multi-product setting studied here. We show that decentralization can lead to one of the products being dropped, thus reducing the breadth of product offerings relative to the centralized system. We also find that capacity imbalance occurs less frequently in the decentralized system, and that its presence in that system depends on the marginal distributions of end-product demands, whereas it does not in the centralized system. In addition, we demonstrate that in some situations the wholesale prices in the decentralized system can alter the assembler's profit margins so that end-product production priorities are reversed from those in the centralized system. Finally, by comparing the decentralized system to one where the common component is replaced by two dedicated components, we find that the flexibility provided by a common component may actually hurt the assembler's performance in a decentralized system (even in the presence of capacity imbalance), whereas it can only improve performance in a centralized system when common and dedicated components have equal costs. By choosing dedicated components, the assembler avoids an incentive problem that arises with a common component – i.e., underinvestment in component capacity due to competition for the common component. Our results regarding the (in)frequency of capacity imbalance and the (un)attractiveness of component commonality suggest that decentralizing decision making (e.g., by outsourcing) may reduce the extent to which the supply chain takes advantage of these approaches for managing uncertainty. In addition to identifying inefficiencies arising from decentralization, we also explore how various factors affect system behavior. We find that capacity imbalance is less likely when demand variability is low, demand correlation is high, or dedicated component capacity costs are high. We also find that reversed production priorities are more likely to occur when the lower-margin product has significantly higher mean demand.

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 (2005) 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 systems 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. 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 (2004) 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. Zhang et al. (2005) study a multi-product, multi-component system, but their model and the focus of their analysis are somewhat different. They assume that the capacity of the shared component (or resource) is exogenous, and they identify wholesale-price only contracts that coordinate the supply chain while still providing all firms positive profits.

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 (1998a) studies the issue of resource flexibility in a two-product setting with one dedicated resource for each product and one flexible resource that can be used for either product. Harrison and Van Mieghem (1999) introduce the notions of operational hedging and capacity imbalance in the context of a system with multiple resources and stochastic demand. Van Mieghem and Rudi (2002) consider centralized newsvendor networks that allow for multiple products and multiple processing and storage points. They show that for certain networks the single-period solution extends to a dynamic setting. Netessine et al. (2002) explore the implications of flexibility in a service environment. Van Mieghem (2004) investigates the relationship between the commonality and flexible capacity problems. Finally, Van Mieghem (2006) studies the role of flexibility in risk-averse centralized newsvendor networks. 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. In Section 3 we analyze the system under centralized decision making, while the behavior of the decentralized assemble-to-order system is characterized in Section 4. 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. Appendix A contains a discussion of system behavior under a restricted wholesale pricing scheme (markup pricing), while Appendix B contains all proofs.

2 Model

Consider an ATO system that produces two products using three components. 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, components A and C are *dedicated* to products 1 and 2, respectively, while component B is *common* to the two products. This system is an example of a newsvendor network (see Van Mieghem and Rudi 2002). Figure 1 below illustrates the system.

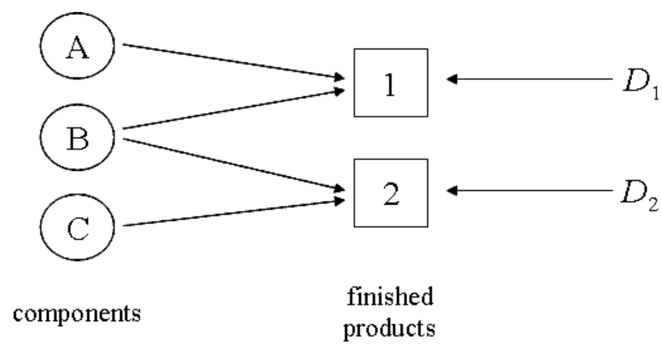


Figure 1: ATO system

Let D_j denote random demand for product j ($j = 1, 2$) during a single selling season, and assume that (D_1, D_2) has a continuous distribution with joint density $f(\cdot, \cdot)$. Let $f_j(\cdot)$ and $F_j(\cdot)$ be the marginal density and cumulative distribution function (cdf), respectively, for D_j, and let $\bar{F}_j(\cdot) = 1 - F_j(\cdot)$. Assume for ease of exposition that $f_j(x) > 0$ for all $x \geq 0$ and $f_j(x) = 0$ otherwise. (All the results can be extended to distributions with bounded support.) Let ρ denote the coefficient of correlation between D_1 and D_2. Also, let $P(H)$ denote the probability of an event H and $E[X]$ the expectation of a random variable X, with respect to the joint demand distribution.

Let p_j be the exogenous market price for each product net of the assembly costs. We assume, without loss of generality, that $p_1 \leq p_2$. Installing a unit of production capacity for component i

costs c_i. Assume that component production costs (once capacity is installed) are zero, and that both products have a positive profit margin, i.e., $p_1 > c_A + c_B$ and $p_2 > c_B + c_C$.

Under decentralized decision making, component capacity decisions are made by three independent suppliers corresponding to components A, B and C, and those suppliers incur the component capacity costs. Finished product assembly decisions are made by a single assembler. The demand distributions and all cost parameters are common knowledge. The sequence of events for this version of the model 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 $\left[0, \overline{Q}_i\right]$ 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. In the centralized version of the model, all decisions are made, and all costs and revenues are incurred, by a single central planner. That model follows the same sequence of events, except that the wholesale price setting step is omitted.

Let c, w, and Q, denote the (column) vectors of unit capacity costs, wholesale prices, and capacity levels, respectively, for the three components. Also, let Q_{-i} be the vector Q with the ith component removed. Finally, let p, D, and y, denote the (column) vectors of prices, demands, and production quantities, respectively, for the two products.

3 Centralized System

In order to explore the impacts of decentralizing decision making in an assemble-to-order system, we first examine the behavior of the system under centralized control. 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 to maximize $\Pi(Q, D) = p_1 y_1 + p_2 y_2$ subject to $Ay \leq Q$ and $y \leq Q$, where

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

(In what follows, it will be convenient to explicitly express the dependence of the finished-product production quantities on the component capacities and realized demands by writing the production vector as $y(Q,D) = (y_1(Q,D), y_2(Q,D))^T$.) This problem has a simple solution, which is reported in Harrison and Van Mieghem (1999). Because the only resource shared by the two products is component B, and each product uses the same amount of this resource, it is optimal to give the more profitable product (product 2) priority access to component B when choosing production quantities.[1] As a result, the optimal production quantities are:

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

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

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. Harrison and Van Mieghem (1999) derive first-order optimality conditions for maximizing $V(Q)$ subject to $Q \geq 0$, and show that at optimality the dedicated component capacities will not exceed the capacity of the common component ($Q_A \leq Q_B$ and $Q_C \leq Q_B$), and the common component capacity will not exceed the sum of the dedicated capacities ($Q_B \leq Q_A + Q_C$). Following similar arguments, we can obtain a characterization of the unique optimal solution in our setting, which we denote Q^0. The following result establishes some properties of the centralized optimal solution.

Theorem 1. *The optimal capacity investment strategy for the centralized system has the following properties:*

(i) It is always optimal to invest in all three components, that is $Q^0 > 0$.

(ii) Under the optimal capacity vector Q^0, $Q^0_A < Q^0_B$ always holds.

(iii) The optimal capacity satisfies the boundary condition $Q^0_B = Q^0_A + Q^0_C$, with $Q^0_A = \bar{F}_1^{-1}(\frac{c_A+c_B}{p_1})$ and $Q^0_C = \bar{F}_2^{-1}(\frac{c_B+c_C}{p_2})$, if and only if

$$P(D_1 \geq Q^0_A, D_2 \geq Q^0_C) \geq \frac{c_B}{p_1}. \tag{3}$$

Otherwise, $Q^0_B < Q^0_A + Q^0_C$.

The capacity balance condition (3) 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

[1] If $p_1 = p_2$ we assume, without loss of optimality, that the decision maker gives priority to product 2.

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 $P(D_1 \geq Q_A^0, D_2 \geq Q_C^0)$, which is the left-hand side of (3). 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 (3). If (3) holds, then it is economically attractive to add that final unit of component B capacity - i.e., to balance the capacities - while if it does not hold, it is better to keep component B capacity lower yielding capacity imbalance (e.g., to take advantage of risk pooling).

The following corollary of Theorem 1 exhibits the necessary and sufficient condition for $Q_B^0 = Q_A^0 + Q_C^0$ for the special cases of independent demands, and perfect negative and positive correlations. As in Van Mieghem (1998a), we model perfect positive demand correlation by $P(D_1 = D_2) = 1$, and perfect negative demand correlation by $P(D_1 + D_2 = K) = 1$ for some $K > 0$.[2]

Corollary 1. *The following are three special cases of the necessary and sufficient condition for balanced capacities in the optimal centralized solution:*

(*a*) *Perfectly negatively correlated demands:* $\frac{c_A}{p_1} + \frac{c_B + c_C}{p_2} \geq 1$

(*b*) *Independent demands:* $\frac{c_A + c_B}{p_1} \frac{c_B + c_C}{p_2} \geq \frac{c_B}{p_1}$

(*c*) *Perfectly positively correlated demands:* $\frac{c_B + c_C}{p_2} \geq \frac{c_B}{p_1}$

From the above conditions we see that if balanced capacities are optimal for perfect negative correlation, then they are also optimal for independent demands, and if optimal for independent demands they are optimal for perfect positive correlation. The latter can be seen immediately by comparing the left-hand sides of the conditions in (b) and (c) and recalling that $p_1 > c_A + c_B$ for profitability of product 1. The former can be seen by writing the conditions in (a) and (b) as $(c_B + c_C)/p_2 \geq 1 - c_A/p_1$ and $(c_B + c_C)/p_2 \geq 1 - c_A/(c_A + c_B)$, respectively, and again using the profitability condition for product 1. This makes intuitive sense – as demands for the two products become more highly correlated there is less opportunity to make use of the flexibility represented by unbalanced capacities. For the special case when demands (D_1, D_2) come from a bivariate Normal distribution, this observation can be extended to general correlations. In that case, the condition in Theorem 1(iii) depends on the demand distribution only through the correlation, and a higher

[2] Similar results are obtained when $D_2 = \alpha_1 D_1 + \alpha_2$, with $\alpha_1 > 0$ for perfect positive correlation and $\alpha_1 < 0$ for perfect negative correlation.

correlation makes balanced capacities more likely. (Van Mieghem 2004 makes a similar observation in a somewhat different setting – where in addition to the common component B, one can also install capacity of dedicated versions of that component. His observations relate to the impact of the demand distributions and correlation on the choice to install positive capacity of the common component, and his condition characterizing that choice is similar to that in (3). Van Mieghem 2006 also shows that in a setting like ours, which he refers to as a "Serial Network," when demands follow a bivariate Normal distribution, capacity imbalance is independent of the marginal demand distributions and is decreasing in demand correlation.)

What is somewhat surprising is the fact that even with perfectly negative demand correlation, there are cases when it is optimal to have balanced capacities. As the condition in (a) indicates, this can occur when either or both products have low profit margins. In such cases, stocking levels for the dedicated component(s) associated with the low-margin product(s) would naturally be low, and therefore the probability that demands for both products exceed the capacities of the associated dedicated components would be high. As a result, the benefit of potentially capturing all of that demand exceeds the cost savings from reducing the capacity of component B.

4 Decentralized System

In this section, we analyze the equilibrium behavior of the decentralized 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. 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.

4.1 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., $0 < p_1 - w_A - w_B \leq p_2 - w_B - w_C$. This implies that it is optimal for the assembler to give priority to product 2 when choosing production quantities.[3] (This assumption is purely for notational convenience. If wholesale prices are such that the margins are reversed, then all results in this section hold with products 1 and 2 and components A and C reversed.)

The assembler's optimal production vector $y(Q, D)$ is obtained by maximizing $\Pi_0 = (p_1 - w_A - w_B)y_1 + (p_2 - w_B - w_C)y_2$ subject to $Ay \leq Q$ and $0 \leq y \leq D$. Since this has the same form as the second-stage problem for the centralized system (including the priorities of the products), the optimal assembler production quantities are also given by (1) and (2).

4.2 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 sold by each supplier is given by $s_A(Q, D) = y_1(Q, D)$, $s_B(Q, D) = y_1(Q, D) + y_2(Q, D) = \min\{Q_B, \min\{Q_A, D_1\} + \min\{Q_C, D_2\}\}$, and $s_C(Q, D) = y_2(Q, D)$. Supplier i's expected profit function is then given by $E\Pi_i(Q_i|Q_{-i}) = w_i E[s_i(Q, D)] - c_i Q_i$. The following result establishes some key properties of the suppliers' capacity game.

Proposition 1. *Given a vector w of wholesale prices satisfying $0 < p_1 - w_A - w_B \leq p_2 - w_B - w_C$, 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.

We begin by characterizing each supplier's best-response function. As a first step, define supplier i's *isolated optimal capacity* with respect to demand for product j as $\hat{Q}_i^j = \bar{F}_j^{-1}(c_i/w_i)$. (Note that $\hat{Q}_A^1$ and $\hat{Q}_C^2$ correspond to the optimal newsvendor quantities for suppliers A and C, respectively, when the other suppliers have ample production capacity.) In addition, we define two functions that play a role in supplier A's and supplier B's best-response functions, respectively. First, let $r_{AB}(Q_B)$ be the solution to $w_A P(D_1 \geq Q_A, D_2 \leq Q_B - Q_A) - c_A = 0$. Such a solution exists if and only if $Q_B \geq \bar{Q}_B$, where $\bar{Q}_B = F_2^{-1}(c_A/w_A)$. (Note that $r_{AB}(\bar{Q}_B) = 0$.) Next, for given values of Q_A and Q_C, consider the following equation in Q_B:

$$\bar{F}_1(Q_B - Q_C) - P(D_1 \geq Q_A, D_2 \leq Q_B - Q_A) - P(Q_B - Q_C \leq D_1 \leq Q_A, D_1 + D_2 \leq Q_B) = \frac{c_B}{w_B}. \quad (4)$$

[3] Without loss of optimality, we assume that the assembler gives priority to product 2 even when profit margins are equal.

If there is a $Q_B < Q_A + Q_C$ that solves (4), then we call that solution $r_{BAC}(Q_A, Q_C)$. Note that the left-hand side of (4) can be interpreted as the probability that supplier B stocks out, so (4) is a two-dimensional analogue of the optimality condition for the classical newsvendor problem.

Theorem 2. *Consider a vector of wholesale prices $w = (w_A, w_B, w_C)$ satisfying $0 < p_1 - w_A - w_B \leq p_2 - w_B - w_C$. The supplier's best-response functions are characterized as follows.*

(*a*) **Supplier** A**.** *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). Supplier A's best-response function is given by*

$$r_A(Q_B, Q_C) = \begin{cases} \hat{Q}_A^1, & \text{if } \ \bar{F}_1(Q_B - Q_C) < \frac{c_A}{w_A} \\ r_{AB}(Q_B), & \text{if } \ P(D_1 \geq Q_B - Q_C, D_2 \leq Q_C) > \frac{c_A}{w_A} \\ Q_B - Q_C, & \text{otherwise.} \end{cases} \tag{5}$$

In addition, $r_A(Q_B, Q_C) \leq Q_B$.

(*b*) **Supplier** B**.** *Fix a pair (Q_A, Q_C). Supplier B's best-response function is given by*

$$r_B(Q_A, Q_C) = \begin{cases} Q_A + Q_C, & \text{if } P(D_1 \geq Q_A, D_2 \geq Q_C) \geq \frac{c_B}{w_B}; \\ r_{BAC}(Q_A, Q_C), & \text{if } P(D_1 \geq Q_A, D_2 \geq Q_C) < \frac{c_B}{w_B}. \end{cases} \tag{6}$$

In addition, if $P(D_1 \geq Q_A, D_2 \geq Q_C) < \frac{c_B}{w_B}$, then

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

(*c*) **Supplier** C**.** *Fix a pair (Q_A, Q_B). Supplier C's best-response function is given by*

$$r_C(Q_A, Q_B) = r_C(Q_B) = \min(\hat{Q}_C^2, Q_B) \leq Q_B. \tag{7}$$

The expression for $r_B(Q_A, Q_C)$ in (6) provides some insights into the impact of Q_A and Q_C on whether supplier B chooses capacity imbalance. For very low values of Q_A and Q_C, supplier B does not prefer capacity imbalance, since demands for both products are likely to exceed the available capacities of the dedicated components. If Q_A or Q_C is increased, Q_B increases by the same amount (preserving balanced capacities) up to the point where $P(D_1 \geq Q_A, D_2 \geq Q_C) = \frac{c_B}{w_B}$. Further increases in Q_A or Q_C are only partially matched by supplier B – i.e., supplier B now prefers capacity imbalance, since the cost of exactly matching an increase in capacity of either dedicated component exceeds the benefit of potentially higher sales.

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 2004, Bernstein and DeCroix 2006.) The best-response functions in (5)-(7) above indicate that pairwise complementarity between component B and each dedicated component exists in the current setting. This is not surprising – when considering just the two components used to make a single product, one would expect the same kind of complementarity as in single-product settings. However, supplier A's best-response function $r_A(Q_B, Q_C)$ is actually non-increasing in Q_C. This is due to the presence of multiple products and the shared component B. If supplier C increases its capacity, more units of component B may be used to produce product 2 (which has higher priority), leaving fewer units available to make product 1. This may cause supplier A to select a lower capacity level for component A – i.e., capacities of components A and C act as economic substitutes, so the suppliers' capacity game is not supermodular. 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.

The lack of supermodularity poses a technical challenge with respect to identifying, computing, and comparing equilibria in the capacity game. However, it is possible in this case to exploit other aspects of the problem structure to establish a version of complementarity that can be used to achieve similar ends. Specifically, note that the best-response function for supplier C is independent of Q_A. (Since product 2 has priority, supplier C never has to worry about "losing" some of component B to product 1 as the result of choices by supplier 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 (5) is then valid replacing $Q_C = r_C(Q_B)$. This composite best-response function is characterized in Proposition 2 and illustrated in Figure 2.

Proposition 2. *Consider a vector of wholesale prices* $w = (w_A, w_B, w_C)$ *satisfying* $0 < p_1 - w_A - w_B \leq p_2 - w_B - w_C$.
If $\frac{c_A}{w_A} + \frac{c_C}{w_C} \geq 1$ *(or equivalently* $\hat{Q}_C^2 \leq \bar{Q}_B$*) we have that*

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

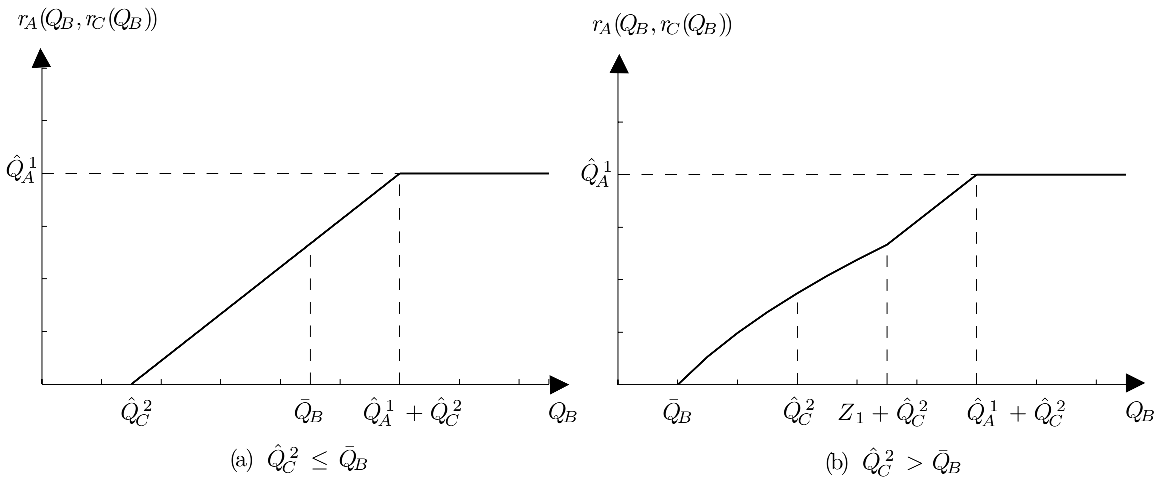


Figure 2: Supplier A's best response

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

$$r_A(Q_B, r_C(Q_B)) = \begin{cases} \hat{Q}_A^1, & \text{if } Q_B > \hat{Q}_A^1 + \hat{Q}_C^2 \\ Q_B - \hat{Q}_C^2, & \text{if } Z_1 + \hat{Q}_C^2 \leq Q_B \leq \hat{Q}_A^1 + \hat{Q}_C^2 \\ r_{AB}(Q_B), & \text{if } \bar{Q}_B \leq Q_B < Z_1 + \hat{Q}_C^2 \\ 0, & \text{if } Q_B < \bar{Q}_B, \end{cases}$$

where $Z_1 \leq \hat{Q}_A^1$ *is the unique solution to* $P(D_1 \geq Z_1, D_2 \leq \hat{Q}_C^2) = \frac{c_A}{w_A}$.[4]
In either case, $r_A(Q_B, r_C(Q_B))$ *is non-decreasing in* Q_B*.*

The last statement in Proposition 2 is of particular interest. By embedding supplier C's best response in supplier A's best-response function, we obtain a version of component complementarity that will be useful when identifying and comparing capacity equilibria.

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 equilibrium can take depending on relationships among the suppliers' capacity costs and the wholesale prices. A formal statement of these results appears in Theorem 3 below.

[4]This solution exists as long as $\frac{c_C}{w_C} \leq 1 - \frac{c_A}{w_A}$.

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 $\bar{Q}_B$, $\hat{Q}_B^2$ and $\hat{Q}_C^2$, play a key role in the characterization of the Pareto-optimal equilibrium. In order to state that characterization, we need two additional definitions related to these fractiles. First, for cases with $\bar{Q}_B = \min(\bar{Q}_B, \hat{Q}_B^2, \hat{Q}_C^2)$, we define the function

$$I(Q_B) \equiv F_1(Q_B - r_C(Q_B)) + \int_{Q_B - r_C(Q_B)}^{r_{AB}(Q_B)} \int_0^{Q_B - x_1} f(x_1, x_2) dx_1 dx_2,$$

on the range $\bar{Q}_B < Q_B < \hat{Q}_C^2 + Z_1$. (Note that when $Q_A = r_{AB}(Q_B)$ and $Q_C = r_C(Q_B)$, (4) can be written as $I(Q_B) = 1 - \frac{c_A}{w_A} - \frac{c_B}{w_B}$.) It is easy to verify that $I(\bar{Q}_B) = 0$, $I\left(\hat{Q}_C^2 + Z_1\right) = F_1(Z_1)$, and that $I(\cdot)$ is increasing. Second, we define Z_2 as the unique solution to $P(D_1 \geq Z_2, D_2 \geq \hat{Q}_C^2) = \frac{c_B}{w_B}$. This solution exists as long as $\frac{c_B}{w_B} \leq \frac{c_C}{w_C}$. Note that Z_2 is increasing in w_B and that $Z_2 = 0$ when $\frac{c_B}{w_B} = \frac{c_C}{w_C}$. When $\frac{c_B}{w_B} > \frac{c_C}{w_C}$, we define $Z_2 = 0$.

Theorem 3. *For any vector of wholesale prices $w = (w_A, w_B, w_C)$ satisfying $0 < p_1 - w_A - w_B \leq p_2 - w_B - w_C$, there exists a unique Pareto-optimal equilibrium $Q^*(w)$ in the suppliers' capacity game. This equilibrium takes one of the following forms, depending on the relationships among fractiles based on the suppliers' capacity costs and the wholesale prices.*

(i) If $\max\{\frac{c_C}{w_C}, 1 - \frac{c_A}{w_A}\} \leq \frac{c_B}{w_B}$, 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 $\left(\min\left(Z_2, \hat{Q}_A^1\right), \min\left(Z_2, \hat{Q}_A^1\right) + \hat{Q}_C^2, \hat{Q}_C^2\right)$ is the Pareto-optimal equilibrium.

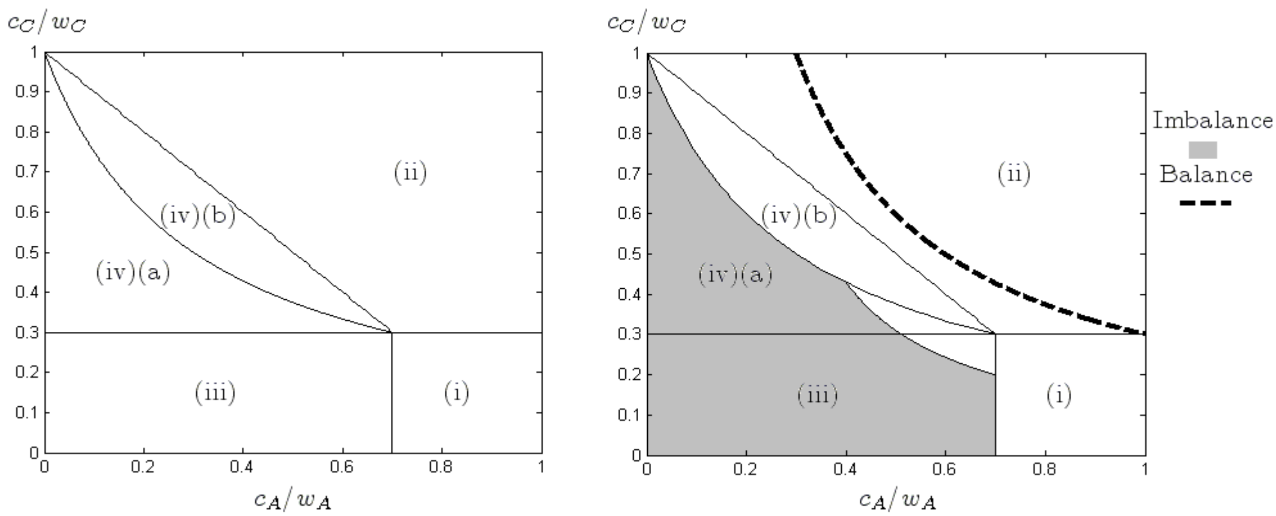
(iii) If $\frac{c_C}{w_C} \leq \frac{c_B}{w_B} < 1 - \frac{c_A}{w_A}$, then the Pareto-optimal equilibrium is $(r_{AB}(Q_B), Q_B, r_C(Q_B))$, for the unique Q_B in the range $\hat{Q}_B^2 < Q_B < \hat{Q}_C^2 + Z_1$ satisfying $I(Q_B) = 1 - \frac{c_A}{w_A} - \frac{c_B}{w_B}$. That equilibrium Q^* satisfies $Q_B^* \leq Q_A^* + Q_C^*$.[5]

[5]This inequality is strict, yielding capacity imbalance, unless $\frac{c_C}{w_C} = \frac{c_B}{w_B}$ and $P(D_1 \geq Z_1, D_2 \geq \hat{Q}_C^2) = P(D_2 \geq \hat{Q}_C^2)$. The latter equality can only hold when D_1 and D_2 are highly positively correlated.

(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* $Z_1 > Z_2$*, then the Pareto-optimal equilibrium is* $(r_{AB}(Q_B), Q_B,\ r_C(Q_B))$*, for the unique* Q_B *in the range* $\bar{Q}_B < Q_B < \hat{Q}_C^2 + Z_1$ *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* $Z_1 \leq Z_2$*, then the Pareto-optimal equilibrium is* $\left(\min\left(Z_2, \hat{Q}_A^1\right), \min\left(Z_2, \hat{Q}_A^1\right) + \hat{Q}_C^2, \hat{Q}_C^2\right)$.



(a) Supplier Equilibrium Outcomes

(b) Assembler's Optimal Pricing

Figure 3: Equilibrium outcomes for fixed c_B/w_B and independent demands.

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 behaviors that can arise. To facilitate discussion of these behaviors, we interpret the conditions in (i) - (iv) of Theorem 3 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. Because this is never optimal under centralized decision making, we see that decentralizing decision making can reduce the breadth of product offerings. In the other cases, positive capacity is installed for all three components. However, regions (iii) and (iv)(a) lead to capacity imbalance[6] (i.e., $Q_B^* < Q_A^* + Q_C^*$) while regions (ii) and (iv)(b) do not (i.e., $Q_B^* = Q_A^* + Q_C^*$). Figure 3(a) depicts the wholesale-price regions identified in Theorem 3 (for fixed c_B/w_B and independent demands). Note that in regions (iii) and (iv)(a), the fractile for at least one dedicated component is small relative to that

[6]With the exception noted in Footnote 6.

of component B, implying an incentive to build a large amount of that dedicated capacity and a greater preference for imbalance by supplier B.

4.3 Assembler's Pricing Decision

Following common practice in the literature (e.g., Gerchak and Wang 2004, Bernstein and DeCroix 2004), we assume that, given any vector of wholesale prices, the suppliers will select the unique Pareto-optimal equilibrium capacity vector $Q^*(w)$. Anticipating the suppliers' capacity response, 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 regions in Theorem 3 can be ignored, thus significantly reducing the set of prices the assembler needs to consider. By reducing and combining regions, we define two new regions and show that an optimal assembler wholesale price vector lies in one of these, with prices in one region leading to capacity imbalance and in the other leading to balanced capacities.

Proposition 3 below is stated under the assumption that the optimal wholesale prices yield production priorities matching the centralized case – i.e., $p_1 - w_A - w_B \leq p_2 - w_B - w_C$. In seeking optimal prices the assembler may need to consider prices such that the priorities are reversed – i.e., $p_1 - w_A - w_B > p_2 - w_B - w_C$. For such prices, the behavior in the component capacity game is as described in the previous section after switching components A and C and products 1 and 2. If the optimal wholesale prices result in reversed priorities, then Proposition 3 holds after making that same switch. Notice that such a change in production priorities represents another type of inefficiency arising from decentralized decision making – the system would now give production priority to a product that has a lower profit margin (from the entire supply chain's perspective). We explore what kind of settings are likely to result in such an outcome in Section 5.

Proposition 3. *A vector $w = (w_A, w_B, w_C)$ of optimal assembler wholesale prices satisfying $0 < p_1 - w_A - w_B \leq p_2 - w_B - w_C$, lies in one of the following two regions. Either,*

$$\frac{c_B}{w_B} \leq \frac{c_C}{w_C}; \qquad \bar{F}_1(Z_2) = \frac{c_A}{w_A}, \tag{8}$$

and any such wholesale-price vector results in equilibrium supplier capacities $(\hat{Q}_A^1, \hat{Q}_A^1 + \hat{Q}_C^2, \hat{Q}_C^2)$ – i.e., balanced capacities. Or,

$$\max\left\{\frac{c_C}{w_C}, \frac{c_B}{w_B}\right\} < 1 - \frac{c_A}{w_A}; \qquad Z_1 > Z_2; \qquad I(\hat{Q}_C^2) \leq 1 - \frac{c_A}{w_A} - \frac{c_B}{w_B}, \tag{9}$$

and 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* $I(Q_B) = 1 - \frac{c_A}{w_A} - \frac{c_B}{w_B}$*. This equilibrium exhibits capacity imbalance – i.e.,* $Q_B^* < Q_A^* + Q_C^*$*.*[7]

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 $\hat{Q}_C^2$. 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 outcomes with balanced capacities, 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.

In summary, solving the assembler's pricing problem requires a pair of numerical searches – one over w satisfying (8) and one over w satisfying (9). These two regions are depicted in Figure 3(b) (for fixed c_B/w_B and independent demands), with capacity balance occurring on the dashed curve and capacity imbalance occurring in the shaded region. The assembler can then compare profits associated with the best solution from each region and choose the better of the two. If some vector w results in reversed priorities, we simply "flip" the system by relabeling products 1 and 2 and components A and C before continuing with the analysis of the suppliers' capacity game.

When demands are perfectly positively or perfectly negatively correlated, the conditions in Proposition 3 can be simplified. Moreover, we can derive closed-form expressions for the equilibria arising in the case of capacity imbalance.

Corollary 2. *Consider again the case with* $0 < p_1 - w_A - w_B \leq p_2 - w_B - w_C$*. Suppose that when demands are perfectly negatively correlated,* $P(D_1 + D_2 = K) = 1$ *for some* $K > 0$*, while when demands are perfectly positively correlated,* $P(D_1 = D_2) = 1$*. Then,*

(*a*) *For perfectly negatively correlated demands, the optimal assembler wholesale prices lie in one of the following two regions. Either,*

$$1 - \frac{c_A}{w_A} + \frac{c_B}{w_B} = \frac{c_C}{w_C}$$

and any such wholesale-price vector results in equilibrium supplier capacities $(\hat{Q}_A^1, \hat{Q}_A^1 + \hat{Q}_C^2, \hat{Q}_C^2)$*. Or,*

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

[7] In the special case where the wholesale prices satisfy (9), and also $\frac{c_C}{w_C} = \frac{c_B}{w_B}$ and $P(D_1 \geq Z_1, D_2 \geq \hat{Q}_C^2) = P(D_2 \geq \hat{Q}_C^2)$, the resulting Pareto optimal capacity equilibrium is balanced, and given by $Q^* = (Z_1, Z_1 + \hat{Q}_C^2, \hat{Q}_C^2)$. See Footnote 6 and the proof of Theorem 3.

and any such prices result in an equilibrium that exhibits capacity imbalance. If $\frac{c_A}{w_A}+\frac{c_B}{w_B}+\frac{c_C}{w_C}>1$*, then* $Q_A^*=\overline{F}_1^{-1}\left(\frac{c_A}{w_A}\right)+\overline{F}_1^{-1}\left(\frac{c_A}{w_A}+\frac{c_B}{w_B}\right)-F_1^{-1}\left(\frac{c_C}{w_C}\right)$*,* $Q_B^*=K+\overline{F}_1^{-1}\left(\frac{c_A}{w_A}+\frac{c_B}{w_B}\right)-F_1^{-1}\left(\frac{c_C}{w_C}\right)$*, and* $Q_C^*=\hat{Q}_C^2=K-F_1^{-1}\left(\frac{c_C}{w_C}\right)$*. On the other hand, if* $\frac{c_A}{w_A}+\frac{c_B}{w_B}+\frac{c_C}{w_C}\leq 1$*, then* $Q^*=(\hat{Q}_A^1,K,\hat{Q}_C^2)$*.*

(*b*) *For perfectly positively correlated demands, let* F *be the common distribution of* D_1 *and* D_2*. The optimal assembler wholesale prices lie in one of the following two regions. Either,*

$$\frac{c_B}{w_B}=\frac{c_A}{w_A}\leq\frac{c_C}{w_C}$$

and any such wholesale-price vector results in equilibrium supplier capacities $Q_A^*=\overline{F}^{-1}\left(\frac{c_A}{w_A}\right)$*,* $Q_B^*=\overline{F}^{-1}\left(\frac{c_A}{w_A}\right)+\overline{F}^{-1}\left(\frac{c_C}{w_C}\right)$*, and* $Q_C^*=\overline{F}^{-1}\left(\frac{c_C}{w_C}\right)$*. Or,*

$$\frac{c_C}{w_C}\leq\frac{c_B}{w_B}<1-\frac{c_A}{w_A};\qquad \overline{F}^{-1}\left(\frac{c_C}{w_C}\right)\leq\overline{F}^{-1}\left(\frac{c_B}{w_B}\right)+\overline{F}^{-1}\left(\frac{c_A}{w_A}+\frac{c_B}{w_B}\right)$$

and any such prices result in the equilibrium with $Q_A^*=\overline{F}^{-1}\left(\frac{c_A}{w_A}+\frac{c_B}{w_B}\right)$*,* $Q_B^*=\overline{F}^{-1}\left(\frac{c_B}{w_B}\right)+\overline{F}^{-1}\left(\frac{c_A}{w_A}+\frac{c_B}{w_B}\right)$*, and* $Q_C^*=\overline{F}^{-1}\left(\frac{c_C}{w_C}\right)$*.*

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.

Our basic 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 have 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 $p_2=16$. We consider all combinations of the following parameters.

(c_A,c_B,c_C) : (1,2,1); (1,5,1); (1,8,1); (1,11,1); (5,2,5); (5,5,5); (5,8,5)
(μ_1,μ_2) : (10,30); (15,25); (20,20); (25,15); (30,10)
(σ_1,σ_2) : (3,3); (9,9)

In addition, in order to explore the impact of c_B on system behavior, for the case of $c_A = c_C = 1$, $(\mu_1, \mu_2) = (10, 30)$ and $\sigma_1 = \sigma_2 = 9$, we also consider $1 \leq c_B \leq 12$. These combinations yield a total of 78 scenarios. For each scenario we compute the optimal assembler prices and resulting equilibrium capacities under three different assumptions regarding demand correlations: perfect negative correlation, independence, and perfect positive correlation. To explore certain specific questions, in some cases we also consider intermediate correlation values.

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.) In addition, Netessine and Zhang (2005) show that complementarity of players' actions – which exists for the suppliers in this setting (from Theorem 2 and Proposition 2) – can exacerbate the effect of double marginalization. Not surprisingly, these types of inefficiencies are 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 both the centralized and decentralized systems yield balanced capacities. In the decentralized system, when optimal assembler wholesale prices result in balanced capacities, the equilibrium component capacity levels are given by $Q_A^* = \bar{F}_1^{-1}\left(\frac{c_A}{w_A}\right)$, $Q_C^* = \bar{F}_2^{-1}\left(\frac{c_C}{w_C}\right)$, and $Q_B^* = Q_A^* + Q_C^*$. In the centralized system, if balanced capacities are optimal, then those capacity levels are $Q_A^0 = \bar{F}_1^{-1}\left(\frac{c_A+c_B}{p_1}\right)$, $Q_C^0 = \bar{F}_2^{-1}\left(\frac{c_B+c_C}{p_2}\right)$, and $Q_B^0 = Q_A^0 + Q_C^0$. As a result, $Q_A^0 > Q_A^*$ if and only if $\frac{c_A}{w_A} > \frac{c_A+c_B}{p_1}$. The latter inequality is equivalent to $p_1 - w_A > \frac{c_B w_A}{c_A} = \frac{c_B}{\bar{F}_1(Z_2)}$, where the last equality follows from (8). From the definition of Z_2, we have that $\bar{F}_1(Z_2) \geq \frac{c_B}{w_B}$. Then, the result follows because $p_1 - w_A - w_B > 0$. Similarly, $Q_C^* < Q_C^0$ if and only if $p_2 - w_C > \frac{c_B w_C}{c_C}$, which holds because $\frac{c_B}{w_B} \leq \frac{c_C}{w_C}$ from (8) and $p_2 - w_B - w_C > 0$. Thus, $Q_B^* < Q_B^0$ follows as well.

5.2 Production Priorities

While double marginalization and the effects of complementarity have been studied before, our multi-product setting allows us to identify new sources of inefficiency resulting from decentralization. To begin with, we find that the products' production priorities in the decentralized system may not match those that are optimal for the supply chain as a whole. From the supply chain's perspective, the item with the higher market price should receive priority when allocating component B in the final production stage. Under decentralization, however, priority is determined by each product's profit margin from the assembler's perspective – the difference between the market price and the wholesale prices paid to the component suppliers. In the numerical study, reversed production priorities (i.e., product 1 having higher margin than product 2) arose in some settings where demand was asymmetric and higher for product 1 (both for independent and correlated demands). For example, when $c_A = c_C = 1$, $c_B = 5$, $\mu_1 = 30, \mu_2 = 10, \sigma_1 = \sigma_2 = 9$, and demands are independent, we find that the assembler's optimal prices yield $p_1 - w_A - w_B = 6.66 > 0.04 = p_2 - w_B - w_C$. To test this pattern further, we ran additional experiments based on this example using more extreme demand asymmetries. Specifically, we fixed mean total demand equal to 40, but allowed mean demand for product 1 to range from 0.5 to 39.5. (In order to avoid truncation from completely distorting the demand distributions, we fixed the coefficient of variation for both products' demands at 0.3, rather than fixing the standard deviations.) Consistent with the observations from our basic study, we found that production priorities were reversed under optimal assembler prices for the cases with largest demand of product 1.

To understand why reversed priorities would be more likely to occur when demand for product 1 is larger, consider the impact of the priority scheme on demands for the components. When product 2 has priority, the effective demand for product 1 (and thus the demand distribution seen by supplier A) is not the observed demand for product 1. Instead it is equal to the minimum of that and the leftover supply of component B after demand for product 2 is satisfied – i.e., the demand distribution is truncated. This reduces the amount of capacity supplier A is willing to install. To induce supplier A to install more, the assembler could offer a higher wholesale price, but doing so would reduce its margin on product 1 even further. As an alternative, the assembler could adjust prices so that product 1 has the higher margin (and thus production priority). Of course, this would shift the truncation effect to product 2. However, when demand for product 1 is substantially larger than for product 2, the impact of this truncation on product 2 demand would be smaller (in absolute terms) than it would be on product 1. In such situations, the dynamics of the decentralized setting may make it desirable for the assembler to cause production priorities to be reversed from what they are under centralized decision making.

5.3 Capacity Imbalance

In this section, we explore another new type of inefficiency that only arises in systems with the multi-product, common-component structure analyzed here – the impact of decentralization on the incidence of capacity imbalance. We begin by making the following observations, based on the numerical findings.

Observation 1. *Capacity imbalance is* less common *in the decentralized system than in the centralized system:*

(*i*) *If capacity imbalance did not occur for a specific case in the centralized system, then it did not occur for that case in the decentralized system under optimal assembler pricing.*

(*ii*) *Among cases when capacity imbalance occurred in the optimal solution of the centralized system, capacities were balanced in the decentralized system under optimal assembler pricing in 81% of the cases with perfectly negative demand correlation, 83% of cases with independent demands, and 100% of cases with perfectly positive demand correlation.*

Our numerical study suggests that this reduced frequency of capacity imbalance in the decentralized setting is driven by a combination of forces. The lower component capacity levels arising in this setting cause the probability of consuming all units of components A and C to be high, so there is little risk-pooling benefit to be derived from capacity imbalance. Imbalance is further inhibited by the shift in incentives that results from decentralization. In a centralized setting, the optimal degree of capacity imbalance 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 in some cases may wish to create capacity imbalance, whereas the assembler experiences only the negative aspect of unbalanced capacities – 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 and reducing the potential benefits from capacity imbalance. 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 capacity imbalance will still occur in some decentralized settings.

The numerical study also allows us to obtain insights into what factors tend to favor capacity imbalance in the decentralized system. We first explore the impact of different attributes of the demand distributions, and then the impact of the component cost parameters, on the occurence of capacity imbalance.

First, by comparing sets with $\sigma_1 = \sigma_2 = 3$ to those with $\sigma_1 = \sigma_2 = 9$, we find that capacity imbalance is less likely when demand variability is low. In fact, we now show that, for demand distributions with sufficiently low coefficients of variation, capacity imbalance does not occur under decentralized decision making.

Proposition 4. *Suppose that demand for product i is given by $D_i = \delta + X_i$, where (X_1, X_2) has joint density function $f(\cdot,\cdot)$ and $\delta \geq 0$, $i = 1, 2$. Let w^* and Q^* be an optimal wholesale price vector and the corresponding capacity equilibrium when $\delta = 0$. Then the optimal prices with $\delta > 0$ satisfy $w^*_A(\delta) + 2w^*_B(\delta) + w^*_C(\delta) < w^*_A + 2w^*_B + w^*_C$. In addition, there exists a $\overline{\delta}$ such that for all $\delta > \overline{\delta}$, the optimal assembler wholesale prices result in balanced capacities in the decentralized system.*

A comparison of results across different (μ_1, μ_2) pairs suggests that demand asymmetry can also have an impact on capacity imbalance. In particular, we find that capacity imbalance is more likely when mean demand for the higher margin product is larger (and thus mean demand for the lower margin product is smaller).[8] This behavior is somewhat surprising, since when taken to the limit, demand asymmetry results in a system with just a single product, which automatically has balanced capacities. One possible explanation for this impact of demand asymmetry is related to the earlier discussion of why capacity imbalance is less common in decentralized systems. To compensate for the loss in expected sales resulting from capacity imbalance, the assembler may choose to reduce the price it pays to supplier A, thus reducing that supplier's capacity level and moving the system closer to balanced capacities. When mean demand for product 1 is low, supplier A's capacity may already be low, which would limit the assembler's ability to pursue this approach.

Recall that the incidence of capacity imbalance is independent of the marginal demand distributions in the centralized system under some specific demand correlation assumptions (independence, perfect positive, and negative correlation), and under any specific demand correlation for bivariate Normal demands. Interestingly, using Theorem 3 one can show that for fixed wholesale prices the same holds in the decentralized system. However, when the assembler is allowed to optimally choose wholesale prices, the incidence of imbalance generally does depend on the marginal demand distributions – i.e., that dependence is not due to decentralization, per se, but to the pricing decision that comes with decentralization. (For example, for independent demands, capacity imbalance arises under the assembler's optimal prices when $c_A = c_C = 1$, $c_B = 5$, $\mu_1 = 10$, $\mu_2 = 30$ and

[8]Typically, this means that μ_2 is larger and μ_1 smaller. In a few cases, though, capacity imbalance is associated with high values of μ_1 and low values of μ_2. However, in all such cases, the optimal wholesale prices result in flipped production priorities as discussed in Section 5.2 – i.e., μ_1 (μ_2) is associated with the higher-margin (lower-margin) product.

$\sigma_1 = \sigma_2 = 9$, while it does not for the same cost values when $\mu_1 = 25, \mu_2 = 15$ and $\sigma_1 = \sigma_2 = 9$. In the centralized system, capacity imbalance occurrs in both cases. For $\rho = -0.5$, and $c_A = c_C = 1$, $c_B = 5$, capacity imbalance arises under the assembler's optimal wholesale prices when $\mu_1 = 10$, $\mu_2 = 30$, $\sigma_1 = 3$, and $\sigma_2 = 9$, while it does not for $\mu_1 = \mu_2 = 20$ and $\sigma_1 = \sigma_2 = 3$. Again, in the centralized system, capacity imbalance occurrs in both cases.) To understand this finding, recall that for wholesale prices leading to capacity imbalance, the effective demand experienced by supplier A (assuming product 1 has lower priority) is truncated by the leftover supply of component B after demand for product 2 is satisfied. This creates an incentive problem – the assembler must offer a higher wholesale price to induce supplier A to add capacity. This truncation, and the associated incentive problem, can be avoided if the assembler sets prices to yield balanced capacities. Since the degree to which supplier A's demand distribution is affected by the truncation (and thus the magnitude of the incentive problem) generally depends on both marginal demand distributions, the same is true of the assembler's choice (through pricing) of balanced or unbalanced capacities.

The final demand distribution attribute we consider is the degree to which demands for the two products are correlated. The results of the numerical study indicate that as demand correlation increases, capacity imbalance occurs less frequently. More specifically, the set of parameters for which capacity imbalance occurs is nested – i.e., if $\rho_1 < \rho_2$ and a set of parameters yields capacity imbalance for a correlation of ρ_2, then those parameters also yield capacity imbalance for ρ_1. This pattern is illustrated in Figure 4 for a specific numerical example (with $c_A = c_C = 1$, $\mu_1 = 10$, $\mu_2 = 30$, and $\sigma_1 = \sigma_2 = 9$) and various values of c_B.[9] Notice that this relationship between correlation and capacity imbalance in the decentralized system is consistent with the behavior characterized in Corollary 1 for the centralized system, and the intuition here is the same.

We now investigate the impact of the component cost parameters on capacity imbalance, beginning with the cost c_B of component B. First, notice that the conditions in Corollary 1 can be rewritten to show that, if *capacity imbalance occurs* in the centralized system, then it occurs for low values of c_B (i.e., $c_B < p_2(1 - c_A/w_A - c_C/w_C)$) if demands are perfectly negatively correlated, for intermediate values of c_B (i.e., $c_B^2 - (p_2 - c_A - c_C)c_B + c_A c_C < 0$) if demands are independent, and for high values of c_B (i.e., $c_B > p_1 c_C/(p_2 - p_1)$) if demands are perfectly positively correlated. Our numerical study reveals the same pattern of behavior in the decentralized system – this is illustrated in Figure 4. (Figure 4 also shows numerical results for intermediate correlations, illustrating how the range of c_B that leads to capacity imbalance shifts as ρ changes.) In addition, consistent with

[9]The capacity balance/imbalance regions characterized in Corollary 1(c) and Corollary 2(b) are valid for perfect positive correlation with $D_2 = D_1 + K$ for any K.

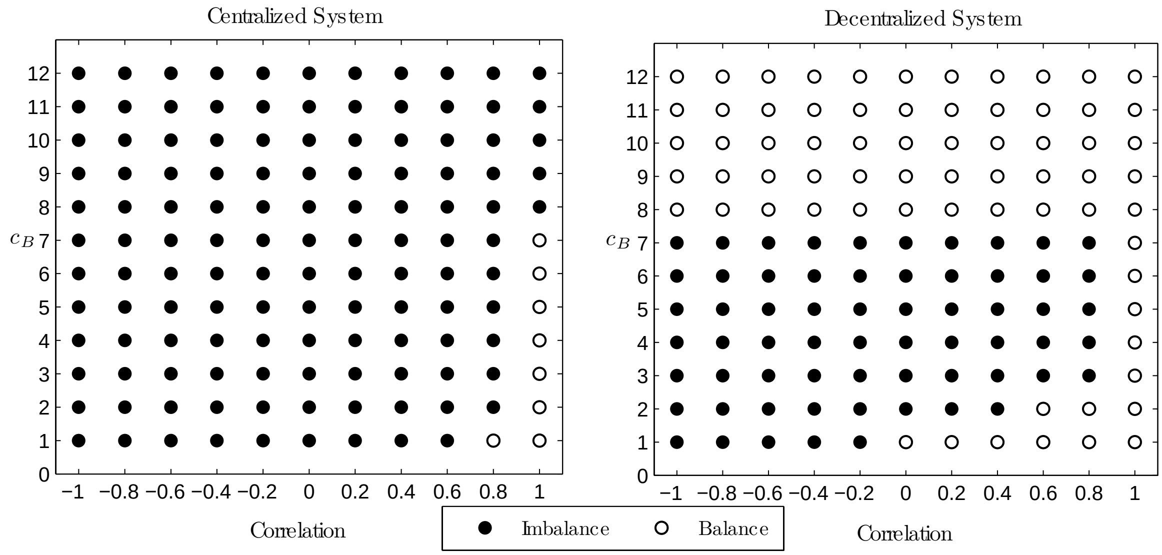


Figure 4: Incidence of capacity imbalance: c_B vs. correlation

Observation 1, the interval of c_B values leading to capacity imbalance in the decentralized system was always a subset of the corresponding interval for the centralized system.

One might expect that, when component B is expensive, capacity imbalance would always be desirable, since choosing $Q_B < Q_A + Q_C$ would yield high component B cost savings. Note, however, that this does not hold for independent or negatively correlated demands. To see the intuition behind this result, we consider the impact of different values of c_B in a little more detail. A higher c_B does give an incentive to choose a lower value for Q_B (either directly, in the centralized model, or via a higher value of w_B in the decentralized model), which corresponds to a lower fractile of the distribution of $D_1 + D_2$. Due to the complementarity among the components, this in turn leads to lower values of Q_A and Q_C, as those capacities adjust to lower fractiles of the marginal demand distributions that are appropriate given the new Q_B (and also possibly adjusted values of w chosen by the assembler). The sizes of the changes in the capacity values determine whether these shifts move the system toward balanced capacities (if $Q_A + Q_C$ decreases faster than Q_B) or unbalanced capacities (if the reverse). When product demands are highly negatively correlated, a small decrease in Q_B corresponds to a large decrease in the associated fractile of the distribution of $D_1 + D_2$. However, making the associated adjustments in the fractiles of the marginal distributions tends to require relatively large decreases in Q_A and Q_C. As a result, as c_B becomes large, it tends to move the system toward balanced capacities. When product demands are highly positively correlated, the decreases in Q_A and Q_C required to achieve any particular shifts in the fractiles of the marginal distributions are clearly the same as for the case of negative correlation (or any other correlation). However, a given decrease in the fractile of the $D_1 + D_2$ distribution tends to require a relatively larger decrease in Q_B compared with the case of negative correlation. Thus as c_B becomes large, the resulting capacity shifts tend to move the system toward unbalanced capacities. The case of independent demands falls in between, behaving more like the positive correlation case when c_B is low, and more like the negative correlation case when c_B is high – apparently the nature of c_B's effect in this case depends on which tail of the distribution the appropriate fractile falls into.

For the dedicated component costs, comparing the scenarios with $c_A = c_C = 1$ to those with $c_A = c_C = 5$, we consistently find that higher values of c_A and c_C are associated with less frequent occurrence of capacity imbalance (across all three correlations considered). 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 capacity imbalance. Figure 5 provides a more detailed illustration of the impact of c_A and c_C on capacity balance for a specific example with independent demands ($p_1 = 14$, $p_2 = 16$, $c_B = 5$, $\mu_1 = 10$, $\mu_2 = 30$ and $\sigma_1 = \sigma_2 = 9$). Once again, consistent with Observation 1, the capacity imbalance region is smaller for the decentralized system.

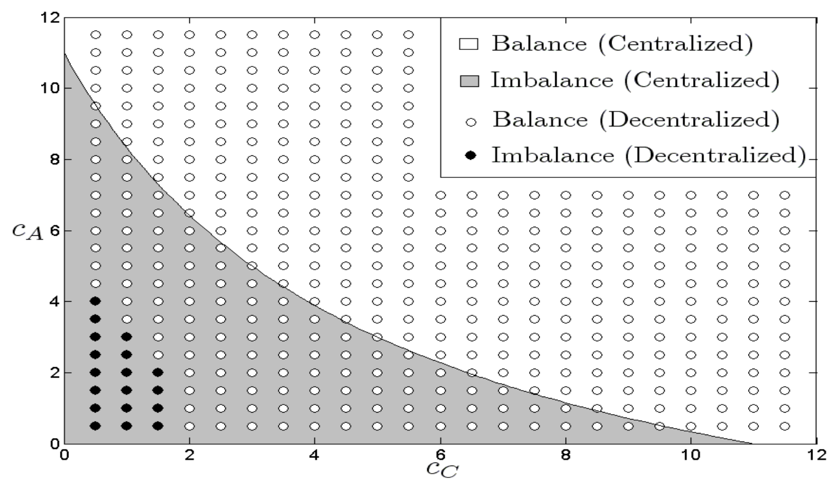


Figure 5: Impact of cost of dedicated components on capacity imbalance

While the preceding discussion has focused on the presence or absence of capacity imbalance, we now explore the degree of capacity imbalance when it does occur in either the decentralized or the centralized systems. To that end, we utilize the *capacity imbalance factor*

$$\gamma = \frac{Q_A + Q_C - Q_B}{Q_B}$$

(introduced in Van Mieghem 2003) as a measure of the degree of capacity imbalance. In particular, $\gamma = 0$ corresponds to the case where the capacities are balanced. Consistent with the earlier discussion of the impact of demand correlation on the incidence of capacity imbalance, we find that the capacity imbalance factor is decreasing in demand correlation. Our 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, capacity imbalance is less common in the decentralized system, but when it occurs, the component capacities in the decentralized system may be relatively more unbalanced than in the integrated system. This is illustrated in Figure 6 below, which corresponds to independent demands, $c_A = c_C = 1$, $\mu_1 = 10$, $\mu_2 = 30$ and $\sigma_1 = \sigma_2 = 9$. (We also graph the optimal wholesale prices that arise in the decentralized system.) Note, for example, that for $c_B = 3$, the imbalance factor in the decentralized system is 24.3%, while in the centralized system it is 22.0%.[10] To understand why the capacity imbalance factor can be higher in the decentralized system, recall that decentralization results in the costs and benefits of capacity imbalance being split between the assembler and supplier B. While the assembler does not receive any of the benefits, supplier B experiences the full benefits (reduced capacity costs) and only part of the costs (reduced revenues), since that supplier earns only part of the profit margin on products sold. As a result, supplier B has a greater preference for imbalance than a centralized decision maker. So while the assembler generally wants to set prices to avoid capacity imbalance, supplier B's preference for imbalance may be so strong in some cases that distorting prices enough to avoid it would not be optimal for the assembler, and in those cases the degree of imbalance may be greater than in the centralized system.

5.4 Choice of Flexibility

The previous section assumed the presence of a common component and focused on whether or not the decentralized system made use of it through capacity imbalance. This section asks a slightly different question: Does the flexibility represented by the very existence of a common component (vs. two dedicated components) always make the assembler in the decentralized system better off (or at least not worse off)? Just as we have assumed that the assembler is the leader (price setter) in the decentralized setting, we examine this question from the assembler's perspective, since the examples that motivated this work involve large, powerful assemblers who would be able to dictate and enforce the type of components they are being supplied.

To explore this question, consider a variation of the decentralized ATO system in which the two

[10] The capacity imbalance factor in the decentralized system is not continuous since the optimal assembler wholesale prices jump from (to) the capacity-balance region (8) to (from) the capacity-imbalance region (9), identified in Proposition 3. (See Figure 3b.)

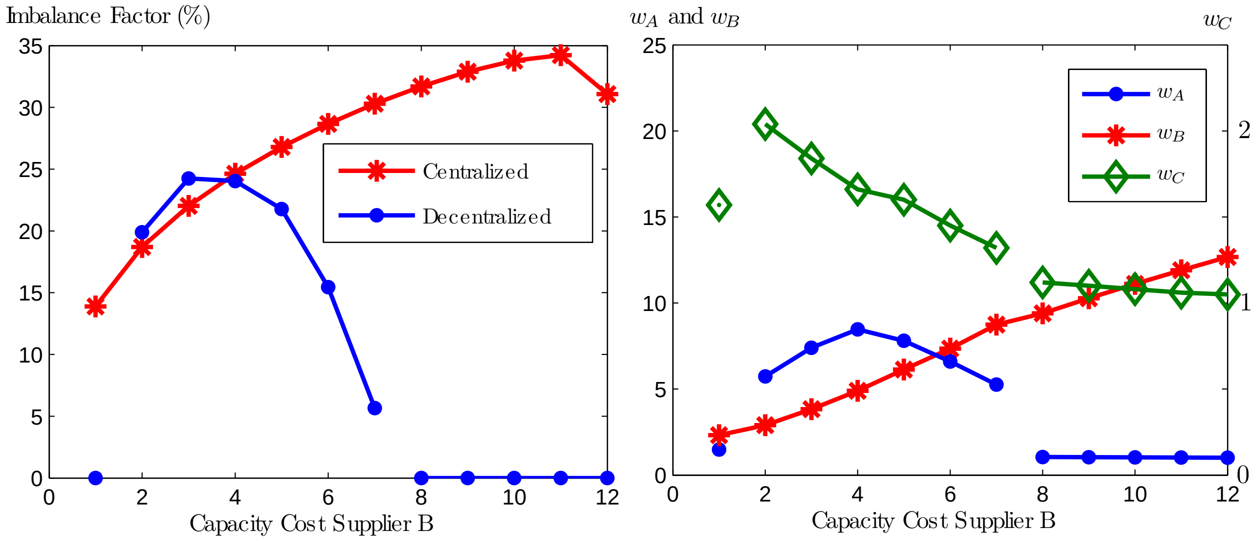


Figure 6: Degree of capacity imbalance / optimal decentralized wholesale prices

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 $B1$ (component $B2$). We denote by Q_{Bi} the capacity of component Bi for product i, and let c_{Bi} be the unit capacity cost for production of component i, $i = 1, 2$. In some cases, components that only need to work in a single finished product may be less costly than a common component that must work in both, resulting in $c_{Bi} < c_B, i = 1, 2$, while in other cases there may be no cost advantage, so that $c_{Bi} = c_B, i = 1, 2$ – we allow either possibility. In the decentralized system, the two dedicated components are sourced from the same supplier B at (possibly different) wholesale prices w_{Bi}, $i = 1, 2$.[11]

The system with dedicated components decomposes into two single-product assembly systems. Results regarding the centralized solution, and the equilibrium capacity levels and optimal assembler wholesale prices in the decentralized system are already known. The following corollary, which follows from Propositions 3 and 4 in Gerchak and Wang (2004), summarizes these results.

Corollary 3. (i) *In the centralized system, the optimal capacity levels are given by* $Q_A^0 = Q_{B1}^0 = \bar{F}_1^{-1}\left((c_A + c_{B1})/p_1\right)$ *and* $Q_C^0 = Q_{B2}^0 = \bar{F}_2^{-1}\left((c_{B2} + c_C)/p_2\right)$.

(ii) *In the decentralized system, for any given vector of wholesale prices* $w = (w_A, w_{B1}, w_{B2}, w_C)$, *the Pareto optimal capacity equilibrium is* $Q_A^* = Q_{B1}^* = \min\left\{\bar{F}_1^{-1}\left(c_A/w_A\right), \bar{F}_1^{-1}\left(c_{B1}/w_{B1}\right)\right\}$ *for product 1, and* $Q_C^* = Q_{B2}^* = \min\left\{\bar{F}_2^{-1}\left(c_{B2}/w_{B2}\right), \bar{F}_2^{-1}\left(c_C/w_C\right)\right\}$ *for product 2. Furthermore, it is optimal for the assembler to set wholesale prices so that* $c_A/w_A = c_{B1}/w_{B1}$ *and* $c_{B2}/w_{B2} = c_C/w_C$.

[11]The findings in this section continue to hold if the assembler is restricted to offering the same wholesale price for components $B1$ and $B2$, or if the assembler purchases components $B1$ and $B2$ from different suppliers.

If the marginal demand distributions have increasing failure rates, the assembler's profit function is unimodal in w_A and w_C.

Consider first the centralized system. If $c_{B1} = c_{B2} = c_B$, then a common component is always (weakly) preferred, but if $\min(c_{B1}, c_{B2}) < c_B$, then dedicated components may be preferred. Also, for values of c_B that lead to capacity imbalance in our original system, there is a unique threshold cost value strictly lower than c_B such that the system with a common component (dedicated components) is preferred if $c_{B1} = c_{B2}$ is higher (lower) than that threshold value. (To see this, note that if $c_{B1} = c_{B2} = 0$, then the system with dedicated components is clearly preferred, while the system with a common component is preferred when $c_{B1} = c_{B2} = c_B$. In addition, the centralized optimal profit in the system with dedicated components is decreasing in their costs.) This threshold arises from trading off the cost savings of using dedicated components with the savings associated with capacity imbalance.

We next investigate how a decentralized system with a common component compares to one with two dedicated components in terms of profitability for the assembler. Interestingly, as we show next, if in the system with a common component capacities are balanced, then the assembler is always better off having two dedicated components for the two products.

Proposition 5. *Let $c_{B1}, c_{B2} \leq c_B$. If optimal wholesale prices w^* in the system with a common component lead to an equilibrium with balanced capacities (i.e., $Q_B^* = Q_A^* + Q_C^*$), then the assembler can improve its profit by employing dedicated components $B1$ and $B2$ for each of the finished products (as opposed to a common component B). That is, in the system with dedicated components, the assembler can induce the same equilibrium capacity levels at lower wholesale prices.*

It is worth highlighting that the above result holds even when there is no unit cost advantage from using dedicated components $B1$ and $B2$, i.e., when $c_{B1} = c_{B2} = c_B$. (In that case, the assembler is strictly better off with dedicated components for any $\rho < 1$, and is indifferent in the special case of $\rho = 1$.) In other words, it establishes conditions under which the assembler would prefer not to have the flexibility of a common component, even when there is no direct cost advantage to that choice. Note that this contrasts with behavior in a system with a centralized decision maker, where a common component is always (weakly) preferred when dedicated and common components have equal costs. One possible explanation for this result is related to an earlier observation related to production priorities. When a common component is used, recall that the effective demand experienced by supplier A (assuming product 1 has lower priority) is truncated by the leftover supply of component B after demand for product 2 is satisfied. This reduces supplier

A's incentive to install capacity. When dedicated components are used, this truncation does not occur, so the effective demand supplier A experiences is stochastically larger. In other words, using dedicated components helps the assembler address an incentive problem without using prices.

Recall that, in our numerical study, balanced capacities arise (in the common component model) as the equilibrium outcome in all of the cases where balanced capacities are optimal for the centralized system, and in a large majority of the cases where capacity imbalance is optimal for the centralized system. In light of these results, Proposition 5 appears to apply in most scenarios. Our numerical study indicates that, for the remaining scenarios – i.e., those where capacity imbalance arises in the decentralized system with a common component – dedicated components may still be preferred by the assembler in some cases even without lower costs for the dedicated components, while in other cases a cost discount must be present for dedicated components to be preferred. To explore this issue further, we expanded the numerical study by considering two additional sets of experiments. In both, $p_1 = 14, p_2 = 16, c_A = c_C = 0.1, \mu_1 = 10$, and $\mu_2 = 30$ (we chose these values to maximize the occurrence and degree of capacity imbalance). In the first set, $c_B = 4$, and we systematically increase $\sigma_1 = \sigma_2$ from 2 to 15. In the second set, $\sigma_1 = \sigma_2 = 6$, while c_B varies from 0.5 to 12. We ran these experiments for independent demands, and perfect negative and positive correlations. For each instance, we calculated the assembler's profit in the system with a common component and its profit in the system with two dedicated components for multiple values of $c_{B1} = c_{B2} \leq c_B$. In all cases, we computed the threshold cost value such that if $c_{B1} = c_{B2}$ falls below the threshold, the assembler prefers the system with two dedicated components. In some cases, the threshold equals c_B implying that the assembler prefers the system with dedicated components even if producing the common component is no more expensive. When the threshold is lower than c_B, the assembler needs a certain amount of cost savings to shift its preference from the system with a common component to that with two dedicated components. Consistent with the observations regarding the occurrence of capacity imbalance, in both sets of experiments the percentage cost savings represented by the threshold values are lower for independent demands than for perfect negative correlation, and in fact, for perfect positive correlation all thresholds equal c_B. Figure 7 below exhibits these percentage cost savings for both sets of experiments, for the cases of independent and perfectly negatively correlated demands.

Note from Figure 7 that the threshold equals c_B (leading to 0% necessary cost savings) for very high and very low values of c_B in the case of independent demands, and for very high values of c_B in the case of perfect negative correlation. These patterns are similar to the relationship between c_B and the incidence of capacity imbalance under independent and perfectly negatively correlated

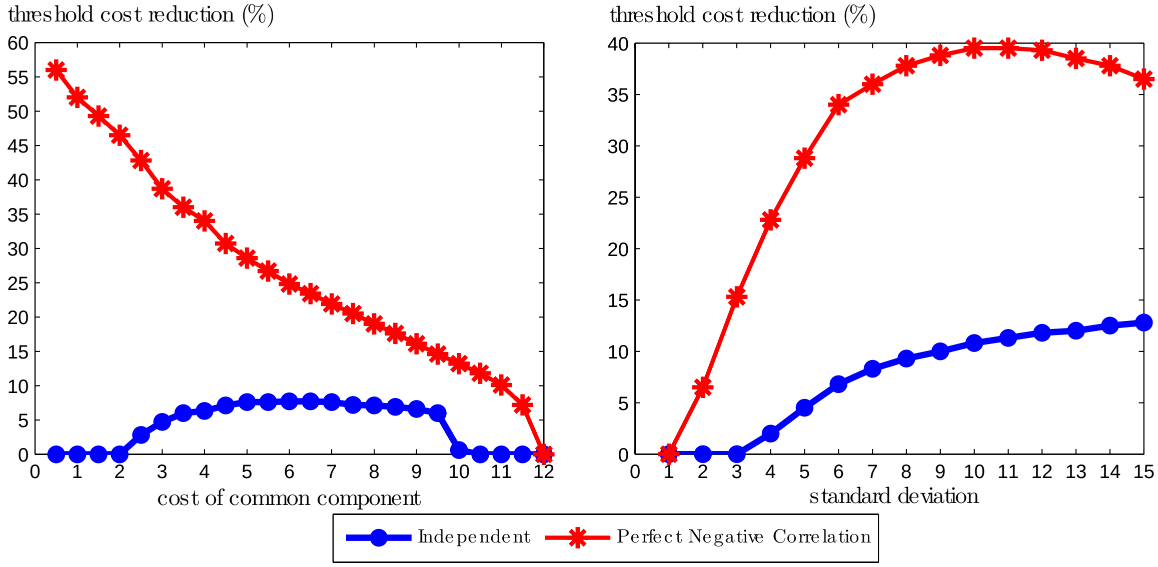


Figure 7: Cost reduction yielding preference for dedicated components

demands. Specifically, in those cases where capacity imbalance is less likely, the system with dedicated components is more profitable to the assembler even without a component cost advantage. On the other hand, when capacity imbalance is more likely, the component cost advantage needs to be larger to outperform the benefits of a system with a common component. A similar pattern arises as $\sigma_1 = \sigma_2$ changes – i.e., scenarios with low demand variability, where capacity imbalance is less likely, require smaller discounts for dedicated components to be preferred.

Finally, for the same two sets of experiments and independent demands, we computed the thresholds corresponding to the centralized system. These thresholds always resulted in higher percentage cost savings than those in the decentralized system. This makes sense given that when the capacity levels are unbalanced, the percentage cost savings represented by the threshold values are always positive in the centralized system, whereas they are zero in most cases under decentralized control.

6 Conclusions

In this paper we explored the impact of decentralization on the behavior of a multi-component, multi-product assemble-to-order system by analyzing and comparing centralized and decentralized versions of the system. Despite the inherent complexity of the decentralized system, we showed that it 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 capacity imbalance, while the other does not. By comparing behavior in the decentralized system to that of the centralized system, we obtained several insights regarding the impact of decentralization on the ATO system. 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 showed that decentralization can lead to one of the products being dropped, thus reducing the breadth of product offerings relative to the centralized system. We also found that capacity imbalance occurs less frequently in the decentralized system. In the decentralized system the presence of capacity imbalance depends on the marginal distributions of end-product demands, while in the centralized system this is not the case. In particular, our results suggested that low demand variance makes capacity imbalance less likely, while high demand for the higher-margin product (relative to the other product) makes capacity imbalance more likely. In addition, we demonstrated that in some situations the wholesale prices in the decentralized system can alter the assembler's profit margins so that the priority for allocating the shared component between products is reversed from what it would be in the centralized system. Finally, by comparing the decentralized system to one where the common component was replaced by two dedicated components, we found that the apparent flexibility provided by a common component may actually hurt performance for the assembler in a decentralized system, even when dedicated and common components have equal costs, whereas in that case a common component can only improve performance in a centralized system. Our results regarding the impact of decentralization on the likelihood of capacity imbalance and the attractiveness of component commonality provide a cautionary note for firms considering outsourcing management of their component supplies – doing so may reduce the extent to which the supply chain takes advantage of these operational hedging approaches for managing uncertainty.

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 product has multiple dedicated components. In addition, the basic analysis here would appear to extend to systems with more than two products that share one common component. Finally, if the assembler must choose a capacity level for assembly operations, then we would expect that capacity to be equal to the common component capacity in equilibrium, resulting in overall (weakly) lower capacities. 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.

Acknowledgments: The authors are grateful to the Senior Editor, Associate Editor, and the referees for helpful comments that have helped to improve the paper.

References

Baker, K., M. Magazine and H. Nuttle (1986), "The Effect of Commonality on Safety Stock in a Simple Inventory Model," *Management Sci.*, 32, 982-988.

Bernstein, F. and G. DeCroix (2004), "Decentralized Pricing and Capacity Decisions in a Multitier System with Modular Assembly," *Management Sci.*, 50, 1293-1309.

Bernstein, F. and G. DeCroix (2006), "Inventory Policies for Decentralized Assembly Systems," *Oper. Res.*, 54, 2, 324-336.

Collier, D.A. (1982), "Aggregate Safety Stocks and Component Part Commonality," *Management Sci.*, 28, 1296-1303.

Fudenberg, D. and J. Tirole (1991), *Game Theory*, The MIT Press, Cambridge, MA.

Gerchak, Y., M. Magazine and A. Gamble (1988), "Component commonality with service level requirements," *Management Sci.*, 34, 753-760.

Gerchak, Y. and M. Henig (1989), "Component Commonality in Assemble-to-Order Systems: Models and Properties," *Naval Logistic Research*, 36, 61-68.

Gerchak, Y. and Y. Wang (2004), "Revenue-Sharing vs. Wholesale-Price Contracts in Assembly Systems with Random Demand," *Production and Operations Management*, 13, 23-33.

Glasserman, P. (1994), Perturbation Analysis of Production Networks. D. Yao, ed. *Stochastic Modeling and Analysis of Manufacturing Systems*. Springer-Verlag, New York.

Graves, S. (1982), "A Multi-Item Inventory Model with a Job Completion Criterion," *Management Sci.*, 28, 1334-1336.

Harrison, J. M., and J. Van Mieghem (1999), "Multi-Resource Investment Strategies: Operational Hedging under Demand Uncertainty," *European Journal of Operations Research*, 113, 17-29.

Hilsenrath, J. (2000), "Parts Shortages Hamper Electronics Makers," *Wall Street Journal*, July 7.

Kopczak, L. and H. Lee (1996), "Hewlett-Packard: DeskJet Printer Supply Chain (A and B)," Stanford University Teaching Case.

Lariviere, M. (1998), "Supply Chain Contracting and Coordination with Stochastic Demand," in *Quantitative Models for Supply Chain Management*, S. Tayur, R. Ganeshan and M. Magazine (Eds), Kluwer Academic publishers, Dordrecht, Netherlands.

Lariviere, M. and E. Porteus (2001), "Selling to the Newsvendor: An Analysis of Price-Only Contracts," *Manufacturing & Service Oper. Management*, 3, 4, 293-305.

Lu, Y., J. Song and D. Yao (2003), "Order Fill Rate, Leadtime Variability, and Advance Demand Information in an Assemble-to-Order System," *Oper. Res.*, 51, 292-308.

Lu, Y., and J. Song (2005), "Order-Based Cost Optimization in Assemble-to-Order Systems," *Oper. Res.*, 53, 151-169.

Netessine, S., G. Dobson and R. Shumsky (2002), "Flexible Service Capacity: Optimal Investment and the Impact of Demand Correlation," *Oper. Res.*, 50, 375-388.

Netessine, S. and F. Zhang (2005), "Positive vs. Negative Externalities in Inventory Management: Implications for Supply Chain Design," *Manufacturing Service Oper. Management*, 7, 1, 58-73.

Smith, S., J. Chambers and E. Shlifer (1980), "Optimal Inventories Based on Job Completion Rate for Repairs Requiring Multiple Items," *Management Sci.*, 26, 849-852.

Song, J. and P. Zipkin (2003), "Supply Chain Operations: Assemble-to-Order Systems," *Handbooks in Operations Research and Management Science*, Vol. 11, *Supply Chain Management: Design, Coordination and Operations*, A. de Kok and S. Graves (Eds), Elsevier, Amsterdam, Netherlands.

Tomlin, B. (2003), "Capacity Investment in Supply Chains: Sharing the Gain Rather than Sharing the Pain," *Manufacturing Service Oper. Management*, 5, 317-333.

Van Mieghem, J. (1998a), "Investment Strategies for Flexible Resources," *Management Sci.*, 44, 1071-1078.

Van Mieghem, J. (1998b), "Seagate Technologies: Operational Hedging," Kellogg School of Management Teaching Case.

Van Mieghem, J. (2003), "Capacity Management, Investment, and Hedging: Review and Recent Developments," *Manufacturing Service Oper. Management*, 5, 269-302.

Van Mieghem, J. (2004), "Commonality Strategies: Value Drivers and Equivalence with Flexible Capacity and Inventory Substitution," *Management Sci.*, 50, 419-424.

Van Mieghem, J. (2006), "Risk-Averse Newsvendor Networks: Resource Flexibility, Sharing, and Hedging," Working Paper, Northwestern University.

Van Mieghem, J. and N. Rudi (2002), "Newsvendor Networks: Inventory Management and Capacity Investment with Discretionary Activities," *Manufacturing Service Oper. Management*, 4, 313-335.

Wang, Y. and Y. Gerchak (2003), "Capacity Games in Assembly Systems with Uncertain Demands," *Manufacturing Service Oper. Management*, 5, 252-267.

Zhang, X., J. Ou and S. Gilbert (2005), "Coordination of Stocking Decisions in an Assemble to Order Environment," Working Paper, National University of Singapore.

Zipkin, P. (2003), "Some Modularity Properties of Linear Programs," Working paper, Duke University.

Appendix A—Restricted Pricing Policy

In order to obtain some additional insights into the assembler's pricing decision and its impact on capacity imbalance, we explore a restricted version of the model. Specifically, we consider a setting where the assembler is restricted to choosing wholesale prices that are equal to a constant mark-up over the suppliers' costs. Our choice of this particular restriction was motivated by two factors. First, this type of pricing structure is intuitively appealing and would be easy to implement in practice. Second, constant mark-up pricing is actually optimal in a single-product decentralized assembly system, where the assembler selects wholesale prices so that the cost/wholesale-price fractiles are equal across all suppliers to ensure balanced capacity/quantity decisions (see Gerchak and Wang 2004).

Consider wholesale prices of the form $w_i = (1+m)c_i$, $i = A, B, C$, for any $0 \leq m \leq \min\left\{\frac{p_1}{c_A+c_B}, \frac{p_2}{c_B+c_C}\right\} - 1$ (where the upper bound preserves profitability of both products for the assembler). Note that under this restricted pricing policy, $Z_2 = 0$. Following the regions defined in Theorem 3, we can identify the Pareto optimal equilibrium for each value of m. At low mark-up levels, i.e., when $0 \leq m \leq 1$, only one product is produced. For high mark-up levels, i.e., when $1 < m \leq \min\left\{\frac{p_1}{c_A+c_B}, \frac{p_2}{c_B+c_C}\right\} - 1$, we have that $Z_1 > 0$. Then, all equilibrium capacities are positive and the wholesale prices can only lie in regions (iii) or (iv)(a) in Theorem 3. (Note that the interval is empty unless $p_1 > 2(c_A + c_B)$ and $p_2 > 2(c_B + c_C)$.) The following result provides sufficient conditions based on the system parameters that indicate which outcome – production of just one product, or capacity imbalance – will result from the assembler's optimal mark-up level.

Proposition 6 *Assume that the marginal distributions F_1 and F_2 have increasing failure rates. If either $p_1 \leq 2(c_A+c_B)$ or $p_2 \leq 2(c_B+c_C)$, then only one product is produced. If $p_1 > 2(c_A+c_B)$ and $p_2 > 2(c_B+c_C)$, then there exist constants $M_1 \geq 1$ and $M_2 > 1$, independent of all cost and revenue parameters but dependent on the distributions of demand for products 1 and 2, respectively, such that there is capacity imbalance under the assembler's optimal choice of mark-up if $p_1 > (M_1+1)(c_A+c_B)$ and $p_2 > (M_2 + 1)(c_B + c_C)$.*[12]

Proposition 6 provides a partial characterization of the market price and component cost parameters that lead to capacity imbalance under decentralized control with mark-up pricing. This result has some interesting implications. First, notice that, in contrast with single-product assembly systems, the optimal wholesale prices in the multi-product setting are not, in general, of the

[12] Capacity imbalance arises with the exception noted in Footnote 6 for highly positively correlated demands.

constant mark-up form. In particular, mark-up pricing cannot, in general, yield equilibrium supplier capacities that are positive for all suppliers and also balanced, which can arise under optimal pricing. Also, note that the sufficient condition for capacity imbalance depends not only on the market prices and component costs, but also on the demand distributions. For the special case of Normal demands, one can show that M_1 and M_2 are decreasing in the variance of demand for products 1 and 2, respectively. In other words, more variable demand makes capacity imbalance more likely. Comparing the sufficient conditions in Proposition 6 with equation (3) suggests that the set of parameters leading to capacity imbalance in the decentralized system under mark-up pricing lies within the corresponding set of parameters for the centralized system, except for some extreme cases. This is due to the fact that, for moderate demand coefficients of variation (which lead to high values of M_1 and/or M_2), the conditions $p_1 > (M_1+1)(c_A+c_B)$ and $p_2 > (M_2+1)(c_B+c_C)$ require one or both products to have very high profit margins. For the case of independent demands, note that in order for the condition in Corollary 1(b) and the condition $p_2 > (M_2 + 1)(c_B + c_C)$ to be both satisfied requires $c_B < c_A/M_2$, i.e., the common component must also represent a small fraction of total component cost. Although the pricing framework addressed in Proposition 6 is not optimal, we observe similar behaviors in our numerical exploration of optimal pricing in the Section 5.

Appendix B—Proofs

Proof of Theorem 1. We first present a characterization of Q^0. We follow the analysis on pages 23-25 in Harrison and Van Mieghem (1999) by defining the sets $\Omega_i(Q)$, $i = 0, ..., 4$, adapted to our setting by defining $\gamma = 1$, $x = A$, $y = C$ and $z = B$, and assigning product 2 the higher priority (which implies that $\Omega_i(Q)$ in their setting becomes $\Omega_{5-i}(Q)$ in our setting, $i = 2, 3$). Then, Q^0 satisfies

$$\begin{pmatrix} 0 \\ 0 \\ p_2 \end{pmatrix} P(\Omega_1^0) + \begin{pmatrix} 0 \\ p_1 \\ p_2 - p_1 \end{pmatrix} P(\Omega_2^0) + \begin{pmatrix} 0 \\ p_1 \\ 0 \end{pmatrix} P(\Omega_3^0) + \begin{pmatrix} p_1 \\ 0 \\ 0 \end{pmatrix} P(\Omega_4^0) = c - v + \tilde{\eta}, \tag{B.1}$$

plus the complementary slackness conditions $v^T Q^0 = 0$, $\eta_1(Q_A^0 + Q_C^0 - Q_B^0) = 0$, $\eta_2(Q_B^0 - Q_C^0) = 0$, $\eta_3(Q_B^0 - Q_A^0) = 0$, for some vectors $v = (v_A, v_B, v_C)^T \in \Re^3_{\geq 0}$ and $\tilde{\eta} \overset{def}{=} (\eta_3 - \eta_1, \eta_1 - \eta_2 - \eta_3, \eta_2 - \eta_1)^T$ with $\eta_1, \eta_2, \eta_3 \geq 0$, and where $\Omega_i^0 = \Omega_i(Q^0)$ (see equation 7 on page 25 in Harrison and Van Mieghem 1999). The equations in (B.1) are:

$$P(D_1 \geq Q_A^0, D_2 \leq Q_B^0 - Q_A^0) = \frac{c_A - v_A - \eta_1 + \eta_3}{p_1}, \tag{B.2}$$

$$P(D_1 \geq Q_B^0 - Q_C^0, D_2 \geq Q_C^0) + P(Q_B^0 - Q_A^0 \leq D_2 \leq Q_C^0, D_1 + D_2 \geq Q_B^0) = \frac{c_B - v_B + \eta_1 - \eta_2 - \eta_3}{p_1}, \tag{B.3}$$

$$\bar{F}_2(Q_C^0) - \frac{p_1}{p_2} P(D_1 \geq Q_B^0 - Q_C^0, D_2 \geq Q_C^0) = \frac{c_C - v_C - \eta_1 + \eta_2}{p_2}. \tag{B.4}$$

(i) Clearly, $Q = 0$ can never be optimal since the central planner can always choose a capacity vector $Q = (\epsilon, \epsilon, 0)$ or $(0, \epsilon, \epsilon)$, with ϵ 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, $v_B = v_C = \eta_3 = 0$, so (B.2)-(B.4) become:

$$F_2(Q) = \frac{c_A - v_A - \eta_1}{p_1}, \tag{B.5}$$

$$\bar{F}_2(Q) = \frac{c_B + \eta_1 - \eta_2}{p_1}, \tag{B.6}$$

$$\bar{F}_2(Q)\left(1 - \frac{p_1}{p_2}\right) = \frac{c_C - \eta_1 + \eta_2}{p_2}.$$

Substituting (B.5) into (B.6) yields $p_1 - c_A - c_B = -v_A - \eta_2 \leq 0$ which contradicts our assumption that $p_1 > c_A + c_B$. Next, assume $Q^0 = (Q, Q, 0)$ with $Q > 0$. In this case, $v_A = v_B = \eta_2 = 0$ and

(B.2)-(B.4) become:

$$0 = \frac{c_A - \eta_1 + \eta_3}{p_1},$$

$$\bar{F}_1(Q) = \frac{c_B + \eta_1 - \eta_3}{p_1}, \tag{B.7}$$

$$1 - \frac{p_1}{p_2}\bar{F}_1(Q) = \frac{c_C - v_C - \eta_1}{p_2}. \tag{B.8}$$

Substituting (B.7) into (B.8) yields $p_2 - c_B - c_C = -v_C - \eta_3 \leq 0$ which contradicts our assumption that $p_2 > c_B + c_C$. Thus, the centralized system will have $Q^0 > 0$.

(ii) Recall that $Q_A^0 \leq Q_B^0$, and suppose that $Q_A^0 = Q_B^0$. Because $Q_C^0 > 0$, we have that $\eta_1 = 0$, and $v = 0$ from (i). Then, from (B.2) we would have that $0 = c_A + \eta_3$. However, $c_A = -\eta_3 \leq 0$ contradicts the fact that $c_A > 0$. So for the optimal Q^0, $Q_A^0 < Q_B^0$.

(iii) Assume that $Q_B^0 = Q_A^0 + Q_C^0$. Then, $v_A = v_B = v_C = \eta_2 = \eta_3 = 0$, and (B.2)-(B.4) become

$$P(D_1 \geq Q_A^0, D_2 \leq Q_C^0) = \frac{c_A - \eta_1}{p_1}, \tag{B.9}$$

$$P(D_1 \geq Q_A^0, D_2 \geq Q_C^0) = \frac{c_B + \eta_1}{p_1}, \tag{B.10}$$

$$\bar{F}_2(Q_C^0) - \frac{p_1}{p_2}P(D_1 \geq Q_A^0, D_2 \geq Q_C^0) = \frac{c_C - \eta_1}{p_2}. \tag{B.11}$$

Solving (B.9)-(B.11) yields $\bar{F}_1(Q_A^0) = \frac{c_A + c_B}{p_1}$ and $\bar{F}_2(Q_C^0) = \frac{c_B + c_C}{p_2}$. Because $\eta_1 \geq 0$, (B.10) implies the that $P(D_1 \geq Q_A^0, D_2 \geq Q_C^0) \geq \frac{c_B}{p_1}$.

Conversely, assume now that $P(D_1 \geq Q_A^0, D_2 \geq Q_C^0) \geq c_B/p_1$, but $Q_B^0 < Q_A^0 + Q_C^0$. Then from (B.3) and using the fact that $v_A = v_B = v_C = \eta_1 = \eta_3 = 0$, we obtain the following inequality

$$P(\Omega_2) + P(\Omega_3) = \frac{c_B - \eta_2}{p_1}, \tag{B.12}$$

where $P(\Omega_2) = P(D_1 \geq Q_B^0 - Q_C^0, D_2 \geq Q_C^0)$ and $P(\Omega_3) = P(Q_B^0 - Q_A^0 \leq D_2 \leq Q_C^0, D_1 + D_2 \geq Q_B^0)$ (note that $Q_B^0 < Q_A^0 + Q_C^0$ implies that $P_D(\Omega_3) > 0$). Since $\eta_2 \geq 0$ and $Q_B^0 < Q_A^0 + Q_C^0$,

$$P(\Omega_2) + P(\Omega_3) > P(\Omega_2) \geq P(D_1 \geq Q_A^0, D_2 \geq Q_C^0) \geq \frac{c_B}{p_1} \geq \frac{c_B - \eta_2}{p_1}.$$

This contradicts (B.12). Thus, we must have $Q_B^0 = Q_A^0 + Q_C^0$. ■

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

$$\begin{aligned}\Pi_i(Q_i|Q_{-i}, D) &= \begin{cases}(w_i - c_i)Q_i, & \text{for } 0 \leq Q_i \leq Q_i^0(D, Q_{-i}) \\ w_i Q_i^0(D, Q_{-i}) - c_i Q_i, & \text{for } Q_i > Q_i^0(D, Q_{-i})\end{cases} \\ &= w_i \min\left(Q_i, Q_i^0(D, Q_{-i})\right) - c_i Q_i. \end{aligned} \tag{B.13}$$

This implies that $E\left[\Pi_i(Q_i|Q_{-i}, D)\right]$ 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 1. *Let $S_i(Q) = E\left[\min(Q_i, g(Q_{-i}, D))\right]$, where $Q \in \Re^m_{\geq 0}$, D is an n-dimensional non-negative random variable with joint probability distribution P, and $g : \Re^{m-1}_{\geq 0} \times \Re^n_{\geq 0} \to \Re_{\geq 0}$ is continuous. Let Q_{-i} be fixed. If $P\left(\{D : Q_i = g(Q_{-i}, D)\}\right) = 0$, then $S_i(Q_i, Q_{-i})$ is differentiable at Q_i and*

$$\frac{\partial S_i}{\partial Q_i} = P\left(\{D : Q_i \leq g(Q_{-i}, D)\}\right).$$

Proof. The function $\min(Q_i, g(Q_{-i}, D))$ is differentiable for all $Q_i \neq g(Q_{-i}, D)$. If $P\left(\{D : Q_i = g(Q_{-i}, D)\}\right) = 0$, then $\min(Q_i, g(Q_{-i}, D))$ is a.s. differentiable at Q_i. Since the derivative of $\min(Q_i, g(Q_{-i}, D))$ with respect to Q_i is bounded for $Q_i \neq g(Q_{-i}, D)$, then the result follows from Theorem 1.2. in Glasserman (1994). ■

Proof of Theorem 2. (a) **Supplier A's Best-Response Function.** Consider supplier A's profit function

$$\begin{aligned} E\Pi_A(Q_A|Q_B, Q_C) &= w_A E\left[s_A(Q, D)\right] - c_A Q_A \\ &= w_A E\left[\min\left(Q_A, Q_B - \min(Q_B, Q_C, D_2), D_1\right)\right] - c_A Q_A. \end{aligned} \tag{B.14}$$

From (B.14) 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 $Q_C \leq Q_B$. We first obtain the derivative of the expected profit function for supplier A. If $Q_C \leq Q_B$, then $E\Pi_A$ is differentiable for $Q_A \neq Q_B - Q_C$, and

$$\frac{\partial E\Pi_A}{\partial Q_A} = \begin{cases} w_A \bar{F}_1(Q_A) - c_A, & \text{if } \; Q_A < Q_B - Q_C \\ w_A P(D_1 \geq Q_A, D_2 \leq Q_B - Q_A) - c_A, & \text{if } \; Q_A > Q_B - Q_C \end{cases} \tag{B.15}$$

Indeed, for $Q_A < Q_B - Q_C$, supplier A's profit function is given by $E\Pi_A = w_A E\left(\min(Q_A, D_1)\right) - c_A Q_A$, so the first half of (B.14) follows immediately. For $Q_A > Q_B - Q_C$, we have from (B.14) that

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

Note that $E[s_A(Q,D)] = E[\min(Q_A, g(Q_B, Q_C, D))]$, 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 1 is verified for $Q_A \leq Q_B$ and $Q_A > Q_B - Q_C$, since $P(\{D : Q_A = g(Q_B, Q_C, D)\}) = P(\{D : Q_A = D_1, D_2 \leq Q_B - Q_A\}) +$
$P(\{D : Q_A = Q_B - D_2, D_1 \geq Q_A\}) = 0$. Then,

$$\begin{aligned}
\frac{\partial E[s_A(Q,D)]}{\partial Q_A} &= P(Q_A \leq \min(Q_B - \min(Q_C, D_2), D_1)) \\
&= P(Q_A \leq Q_B - \min(Q_C, D_2), D_1 \geq Q_A) \\
&= P(Q_A \leq Q_B - D_2, D_2 \leq Q_C, D_1 \geq Q_A) \\
&= P(D_2 \leq Q_B - Q_A, D_1 \geq Q_A),
\end{aligned}$$

where the last two equalities follow since $Q_A > Q_B - Q_C$.

Note from (B.15) that $\partial E\Pi_A/\partial Q_A$ is strictly decreasing in Q_A, with a downward jump at $Q_A = Q_B - Q_C$. Thus, if we can find a capacity Q_A that satisfies $\partial E\Pi_A/\partial Q_A = 0$, 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, $\frac{c_A}{w_A} > \bar{F}_1(Q_B - Q_C)$), then $r_A(Q_B, Q_C) = \hat{Q}_A^1$. Suppose instead that $\hat{Q}_A^1 \geq Q_B - Q_C$. Then, $\partial E\Pi_A/\partial Q_A > 0$ for all $Q_A < Q_B - Q_C$, so $r_A(Q_B, Q_C) \geq Q_B - Q_C$. Now If $r_{AB}(Q_B)$ exists and $r_{AB}(Q_B) > Q_B - Q_C$ (or, equivalently, $\frac{c_A}{w_A} < P(D_1 \geq Q_B - Q_C, D_2 \leq Q_C)$), then $r_A(Q_B, Q_C) = r_{AB}(Q_B)$. Otherwise, $\partial E\Pi_A/\partial Q_A < 0$ for $Q_A > Q_B - Q_C$, in which case $r_A(Q_B, Q_C) = Q_B - Q_C$.

(b) **Supplier B's Best-Response Function.** Consider now supplier B's profit function,

$$\begin{aligned}
E\Pi_B(Q_B|Q_A, Q_C) &= w_B E[s_B(Q,D)] - c_B Q_B \\
&= w_B E[\min(Q_B, \min(D_1, Q_A) + \min(D_2, Q_C))] - c_B Q_B. \qquad \text{(B.16)}
\end{aligned}$$

From (B.16) 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. To compute the derivative of supplier B's expected sales function, note that $E[s_B(Q_B, Q_{-B}, D)] = E[\min(Q_B, g(Q_{-B}, D))]$, where $g(Q_{-B}, D) = \min(D_1, Q_A) + \min(D_2, Q_C)$. The condition of Lemma 1 is verified for $Q_B < Q_A + Q_C$. If $\frac{\partial E[s_B(Q,D)]}{\partial Q_B} \geq \frac{c_B}{w_B}$ 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

$$\frac{\partial E[s_B(Q,D)]}{\partial Q_B} = P(\{D : Q_B \leq \min(D_1, Q_A) + \min(D_2, Q_C)\}) = \frac{c_B}{w_B}. \qquad \text{(B.17)}$$

The set $\{D : Q_B \leq \min(D_1, Q_A) + \min(D_2, Q_C)\}$ corresponds to region **I** in Figure 6. Then, (B.17) can be written as in (4), 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 third and fourth terms correspond to the probability that D falls in regions **II** and **III**, respectively.

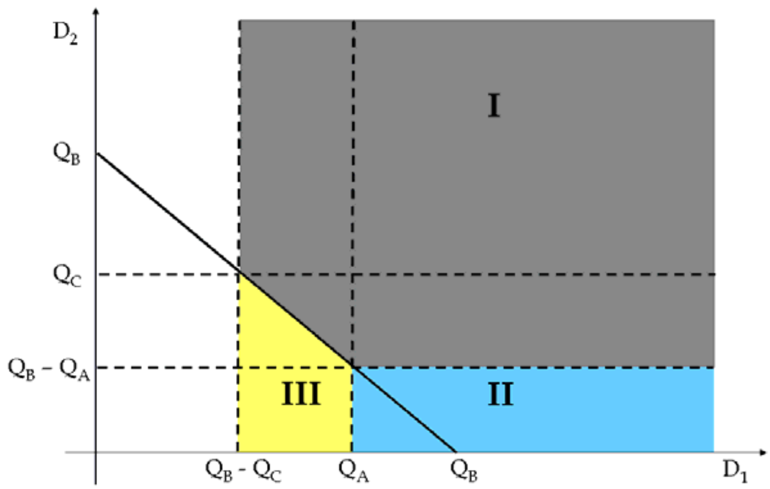


Figure 8: Supplier B's First Order Condition

We can now characterize supplier B's best-response function. Fix the values of Q_A and Q_C. It is easy to verify that the left-hand side of (4) is decreasing in Q_B, and that as Q_B approaches $Q_A + Q_C$, it approaches $P(D_1 \geq Q_A, D_2 \geq Q_C)$. As a result, supplier B's optimal capacity choice is $Q_B = Q_A + Q_C$ if and only if $P(D_1 \geq Q_A, D_2 \geq Q_C) \geq \frac{c_B}{w_B}$. For values of Q_A and Q_C with $P(D_1 \geq Q_A, D_2 \geq Q_C) < \frac{c_B}{w_B}$, supplier B's best response is given by the solution to (4).

Note that, leaving Q_A fixed,

$$0 < \frac{\partial r_{BAC}}{\partial Q_C} = \frac{\int_{Q_C}^{\infty} f(Q_B - Q_C, x)dx}{\int_{Q_C}^{\infty} f(Q_B - Q_C, x)dx + \int_{Q_A}^{\infty} f(x, Q_B - Q_A)dx + \int_{Q_B - Q_C}^{Q_A} f(x, Q_B - x)dx} < 1.$$

Similarly, when Q_C is fixed, we have that

$$0 < \frac{\partial r_{BAC}}{\partial Q_A} = \frac{\int_{Q_A}^{\infty} f(x, Q_B - Q_A)dx}{\int_{Q_C}^{\infty} f(Q_B - Q_C, x)dx + \int_{Q_A}^{\infty} f(x, Q_B - Q_A)dx + \int_{Q_B - Q_C}^{Q_A} f(x, Q_B - x)dx} < 1.$$

(c) **Supplier C's Best-Response Function.** Consider supplier C's profit function, $E\Pi_C(Q_C|Q_A, Q_B) = w_C E[s_C(Q, D)] - c_C Q_C = w_C E[\min(Q_C, Q_B, D_2)] - c_C Q_C$. If $Q_B \geq \hat{Q}_C^2$, clearly supplier C's optimal capacity is $\hat{Q}_C^2$. If instead $Q_B < \hat{Q}_C^2$, then $E\Pi_C$ 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 $r_C(Q_A, Q_B) = r_C(Q_B) = \min(\hat{Q}_C^2, Q_B)$. ■

Proof of Proposition 2. First, if $\frac{c_A}{w_A} + \frac{c_C}{w_C} \geq 1$, we have that $\bar{Q}_B \geq \hat{Q}_C^2$. For $Q_B \leq \hat{Q}_C^2$, $r_C(Q_B) = Q_B$, which implies that $P(D_1 \geq Q_B - r_C(Q_B), D_2 \leq r_C(Q_B)) = F_2(Q_B) \leq \frac{c_A}{w_A}$ since $Q_B \leq \bar{Q}_B$, and $\frac{c_A}{w_A} < 1 = \bar{F}_1(Q_B - r_C(Q_B))$. As a result, for any $Q_B \leq \hat{Q}_C^2$, $r_A(Q_B, r_C(Q_B)) = Q_B - r_C(Q_B) = 0$. For $\hat{Q}_C^2 < Q_B \leq \hat{Q}_A^1 + \hat{Q}_C^2$, $r_C(Q_B) = \hat{Q}_C^2$ which implies that $P(D_1 \geq Q_B - r_C(Q_B), D_2 \leq r_C(Q_B)) < 1 - \frac{c_C}{w_C} \leq \frac{c_A}{w_A}$, and $\bar{F}_1(Q_B - r_C(Q_B)) = \bar{F}_1(Q_B - \hat{Q}_C^2) \geq \bar{F}_1(\hat{Q}_A^1) = \frac{c_A}{w_A}$. As a result, $r_A(Q_B, r_C(Q_B)) = Q_B - \hat{Q}_C^2$. For $Q_B > \hat{Q}_A^1 + \hat{Q}_C^2$, $\bar{F}_1(Q_B - \hat{Q}_C^2) < \bar{F}_1(\hat{Q}_A^1) = c_A/w_A$, 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 $\bar{Q}_B < \hat{Q}_C^2$. Similar to above, we have that for $Q_B < \bar{Q}_B$, $r_C(Q_B) = Q_B$ and $r_A(Q_B, r_C(Q_B)) = Q_B - r_C(Q_B) = 0$. If $\bar{Q}_B \leq Q_B < \hat{Q}_C^2$, then $r_C(Q_B) = Q_B$ and $\frac{c_A}{w_A} \leq P(D_1 \geq Q_B - r_C(Q_B), D_2 \leq r_C(Q_B)) = F_2(Q_B)$, so that $r_A(Q_B, r_C(Q_B)) = r_{AB}(Q_B)$. If $\hat{Q}_C^2 \leq Q_B < \hat{Q}_C^2 + Z_1$, then $r_C(Q_B) = \hat{Q}_C^2$ and $\frac{c_A}{w_A} < P(D_1 \geq Q_B - \hat{Q}_C^2, D_2 \leq \hat{Q}_C^2)$, since $Q_B - \hat{Q}_C^2 < Z_1$, so that $r_A(Q_B, r_C(Q_B)) = r_{AB}(Q_B)$. Finally, if $Z_1 + \hat{Q}_C^2 \leq Q_B \leq \hat{Q}_A^1 + \hat{Q}_C^2$, then $r_C(Q_B) = \hat{Q}_C^2$, $\frac{c_A}{w_A} \geq P(D_1 \geq Q_B - \hat{Q}_C^2, D_2 \leq \hat{Q}_C^2)$ and $\frac{c_A}{w_A} \leq \bar{F}_1(Q_B - \hat{Q}_C^2)$, and as a result $r_A(Q_B, r_C(Q_B)) = Q_B - \hat{Q}_C^2$. 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

$$0 < \frac{\partial r_{AB}(Q_B)}{\partial Q_B} = \frac{\int_{Q_A}^{\infty} f(x, Q_B - Q_A)dx}{\int_{Q_A}^{\infty} f(x, Q_B - Q_A)dx + \int_0^{Q_B - Q_A} f(Q_A, x)dx} < 1, \tag{B.18}$$

which implies that $r_{AB}(Q_B) > 0$ for all $Q_B > \bar{Q}_B$. Also, $r_{AB}\left(\hat{Q}_C^2 + Z_1\right) = Z_1$. Then, $r_A(Q_B, r_C(Q_B))$ is non-decreasing in Q_B. ■

Proof of Theorem 3. We begin by considering the range $0 \leq Q_B \leq \min(\bar{Q}_B, \hat{Q}_B^2, \hat{Q}_C^2)$. In this case $r_C(Q_B) = Q_B$ and $r_A(Q_B, r_C(Q_B)) = 0$ (see (7) and Proposition 2). In addition, since $Q_B \leq \hat{Q}_B^2$, (6) 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 $Q_B = \min(\hat{Q}_C^2, \hat{Q}_B^2, \bar{Q}_B)$. Finally, supplier B's profit under any such equilibrium is $E\Pi_B = w_B E[\min(Q_B, D_2)] - c_B Q_B$, which is concave and reaches its maximum at $Q_B = \hat{Q}_B^2$. Thus supplier B also prefers $Q_B = \min(\hat{Q}_C^2, \hat{Q}_B^2, \bar{Q}_B)$, so $(0, Q_B, Q_B)$ with $Q_B = \min(\hat{Q}_C^2, \hat{Q}_B^2, \bar{Q}_B)$ is Pareto best among these equilibria.

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 2), 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 $\min(\bar{Q}_B, \hat{Q}_B^2, \hat{Q}_C^2) < Q_B \le \hat{Q}_A^1 + \hat{Q}_C^2$, 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 $\frac{c_C}{w_C} < 1 - \frac{c_A}{w_A} \le \frac{c_B}{w_B}$, which is equivalent to $\hat{Q}_B^2 \le \bar{Q}_B < \hat{Q}_C^2$. For $\hat{Q}_B^2 < Q_B < \bar{Q}_B$, $r_C(Q_B) = Q_B$ and $r_A(Q_B, r_C(Q_B)) = 0$, but $r_B(0, Q_B) = Q_B$ if and only if $\bar{F}_2(Q_B) \ge \frac{c_B}{w_B}$ if and only if $Q_B \le \hat{Q}_B^2$. Therefore, there is no equilibrium with $\hat{Q}_B^2 < Q_B \le \bar{Q}_B$. For $\bar{Q}_B \le Q_B \le \hat{Q}_C^2$, $r_C(Q_B) = Q_B$ 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 $P(D_1 \ge r_{AB}(Q_B), D_2 \ge Q_B) \le \bar{F}_2(Q_B) < \bar{F}_2(\hat{Q}_B^2) = \frac{c_B}{w_B}$. Then, from (4), we would need $1 - P(D_1 \ge r_{AB}(Q_B), D_2 \le Q_B - r_{AB}(Q_B)) - \frac{c_B}{w_B} > 0$. The latter cannot hold since $1 - \frac{c_A}{w_A} \le \frac{c_B}{w_B}$, and by the definition of $r_{AB}(Q_B)$, $P(D_1 \ge r_{AB}(Q_B), D_2 \le Q_B - r_{AB}(Q_B)) = \frac{c_A}{w_A}$. Thus, there is no equilibrium in this range either. For $\hat{Q}_C^2 < Q_B < \hat{Q}_C^2 + Z_1$, $r_C(Q_B) = \hat{Q}_C^2$, but otherwise the preceding argument applies here as well, so there is no equilibrium in this range. For $\hat{Q}_C^2 + Z_1 \le Q_B \le \hat{Q}_A^1 + \hat{Q}_C^2$, $r_C(Q_B) = \hat{Q}_C^2$ 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 3 we must have $P(D_1 \ge Q_B - \hat{Q}_C^2, D_2 \ge \hat{Q}_C^2) \ge \frac{c_B}{w_B}$. But $P(D_1 \ge Q_B - \hat{Q}_C^2, D_2 \ge \hat{Q}_C^2) < \bar{F}_2(\hat{Q}_C^2) < \bar{F}_2(\hat{Q}_B^2) = \frac{c_B}{w_B}$, where the last inequality follows since $\hat{Q}_B^2 < \hat{Q}_C^2$.
Suppose instead $1 - \frac{c_A}{w_A} \le \frac{c_C}{w_C} \le \frac{c_B}{w_B}$, which is equivalent to $\hat{Q}_B^2 \le \hat{Q}_C^2 \le \bar{Q}_B$. For $\hat{Q}_B^2 < Q_B \le \hat{Q}_C^2$, $r_C(Q_B) = Q_B$ 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 $\bar{F}_2(Q_B) \ge \frac{c_B}{w_B}$, but this cannot hold since $Q_B > \hat{Q}_B^2$, so there is no equilibrium in this range. For $\hat{Q}_C^2 < Q_B \le \hat{Q}_A^1 + \hat{Q}_C^2$, $r_C(Q_B) = \hat{Q}_C^2$ and $r_A(Q_B, r_C(Q_B)) = Q_B - \hat{Q}_C^2$, but again $P(D_1 \ge Q_B - \hat{Q}_C^2, D_2 \ge \hat{Q}_C^2) < \frac{c_B}{w_B}$, which implies that there is no equilibrium in this range.
Thus, for $\max\left\{\frac{c_C}{w_C}, 1 - \frac{c_A}{w_A}\right\} \le \frac{c_B}{w_B}$, $(0, \hat{Q}_B^2, \hat{Q}_B^2)$ is the Pareto-optimal equilibrium.

(ii) Suppose now that $1 - \frac{c_A}{w_A} \le \frac{c_B}{w_B} < \frac{c_C}{w_C}$, which is equivalent to $\hat{Q}_C^2 < \hat{Q}_B^2 \le \bar{Q}_B$. For

$$\hat{Q}_C^2 < Q_B \le \hat{Q}_C^2 + \min\left(Z_2, \hat{Q}_A^1\right), \tag{B.19}$$

$r_C(Q_B) = \hat{Q}_C^2$ 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 $\frac{c_B}{w_B} \le P(D_1 \ge Q_B - \hat{Q}_C^2, D_2 \ge \hat{Q}_C^2)$ which is equivalent to $Q_B \le \hat{Q}_C^2 + Z_2$. As a result, for all Q_B in the range (B.19) the vectors $(Q_B - \hat{Q}_C^2, Q_B, \hat{Q}_C^2)$ are equilibria. If $Z_2 < \hat{Q}_A^1$, then we also need to consider the range $\hat{Q}_C^2 + Z_2 < Q_B \le \hat{Q}_A^1 + \hat{Q}_C^2$. 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

$$\hat{Q}_C^2 \le Q_B \le \hat{Q}_C^2 + \min\left(Z_2, \hat{Q}_A^1\right). \tag{B.20}$$

Supplier C is clearly indifferent. For supplier A, note that its profit for any equilibrium of this form is $E\Pi_A = w_A E[\min(Q, D_1)] - c_A Q$ for $Q = Q_B - \hat{Q}_C^2$ and Q_B in the range (B.20). Thus, $0 \le Q \le \min\left(Z_2, \hat{Q}_A^1\right)$, and since π_A is concave and reaches its maximum at $\hat{Q}_A^1$, $Q_B = \hat{Q}_C^2 + \min\left(Z_2, \hat{Q}_A^1\right)$ is preferred by supplier A. Similarly, supplier B's profit in this range is $E\Pi_B = w_B E[\min(D_1, Q_B - \hat{Q}_C^2) + \min(D_2, \hat{Q}_C^2)] - c_B Q_B$, which is concave and reaches its maximum at $Q_B = \hat{Q}_C^2 + \hat{Q}_B^1$. Note that $\hat{Q}_B^1 = \bar{F}_1^{-1}\left(\frac{c_B}{w_B}\right) \ge Z_2$, since $P(D_1 \ge \hat{Q}_B^1, D_2 \ge \hat{Q}_C^2) \le \bar{F}_1(\hat{Q}_B^1) = \frac{c_B}{w_B}$. Then, supplier B prefers the same equilibrium.

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

Thus, for $\max\left\{\frac{c_B}{w_B}, 1 - \frac{c_A}{w_A}\right\} \le \frac{c_C}{w_C}$, $\left(\min\left(Z_2, \hat{Q}_A^1\right), \min\left(Z_2, \hat{Q}_A^1\right) + \hat{Q}_C^2, \hat{Q}_C^2\right)$ is the unique Pareto optimal equilibrium.

(iii) Suppose that $\frac{c_C}{w_C} \le \frac{c_B}{w_B} < 1 - \frac{c_A}{w_A}$, which is equivalent to $\bar{Q}_B < \hat{Q}_B^2 \le \hat{Q}_C^2$. 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 $\hat{Q}_C^2 \le Q_B < \hat{Q}_C^2 + Z_1$ 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 $Q_B = \hat{Q}_C^2 + Z_1$ and $\partial r_{AB}/\partial Q_B < 1$ for $\bar{Q}_B < Q_B < \hat{Q}_C^2 + Z_1$. As a result, for Q_B in the range $\bar{Q}_B < Q_B < \hat{Q}_C^2 + Z_1$, 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,

$$I(Q_B) = 1 - \frac{c_A}{w_A} - \frac{c_B}{w_B}. \tag{B.21}$$

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

$$I\left(\hat{Q}_B^2\right) \le \int_0^{r_{AB}(\hat{Q}_B^2)} \int_0^{\hat{Q}_B^2} f(x_1, x_2) dx_1 dx_2 = 1 - \frac{c_B}{w_B} - \int_{r_{AB}(\hat{Q}_B^2)}^{\infty} \int_0^{\hat{Q}_B^2} f(x_1, x_2) dx_1 dx_2 <$$

$$1 - \frac{c_B}{w_B} - \int_{r_{AB}(\hat{Q}_B^2)}^{\infty} \int_0^{\hat{Q}_B^2 - r_{AB}(\hat{Q}_B^2)} f(x_1, x_2) dx_1 dx_2 = 1 - \frac{c_A}{w_A} - \frac{c_B}{w_B},$$

and $I\left(\hat{Q}_C^2 + Z_1\right) = F_1(Z_1) \ge 1 - \frac{c_A}{w_A} - \frac{c_B}{w_B}$, since $\frac{c_A}{w_A} = P(D_1 \ge Z_1, D_2 \le \hat{Q}_C^2) = \bar{F}_1(Z_1) - P(D_1 \ge Z_1, D_2 \ge \hat{Q}_C^2) \ge \bar{F}_1(Z_1) - \bar{F}_2(\hat{Q}_C^2) \ge \bar{F}_1(Z_1) - \bar{F}_2(\hat{Q}_B^2) = \bar{F}_1(Z_1) - \frac{c_B}{w_B}$, where the last inequality follows from $\hat{Q}_C^2 \ge \hat{Q}_B^2$. So there is no solution to (B.21), and thus no equilibrium with $\bar{Q}_B < Q_B \le$

$\hat{Q}_B^2$. However, there exists a unique solution Q_B^* to (B.21) in the range $\hat{Q}_B^2 < Q_B \leq \hat{Q}_C^2 + Z_1$, and this corresponds to the unique equilibrium

$$(r_{AB}(Q_B^*), r_{BAC}(r_{AB}(Q_B^*), r_C(Q_B^*)), r_C(Q_B^*))$$

in that range. For $\hat{Q}_C^2 + Z_1 < Q_B \leq \hat{Q}_A^1 + \hat{Q}_C^2$ 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 $(r_{AB}(Q_B^*), r_{BAC}(r_{AB}(Q_B^*), r_C(Q_B^*)), r_C(Q_B^*))$. 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 $E\Pi_C = w_C E[\min(Q_C, D_2)] - c_C Q_C$, and is concave and maximized at $\hat{Q}_C^2$. Since $\bar{Q}_B < \hat{Q}_B^2 \leq \min(Q_B^*, \hat{Q}_C^2) = r_C(Q_B^*) \leq \hat{Q}_C^2$, supplier C also prefers the latter equilibrium. For supplier B, its profit is increasing in Q_A and Q_C. Then, since $r_C(Q_B^*) \geq \bar{Q}_B$, $E\Pi_B(0, \bar{Q}_B, \bar{Q}_B) \leq E\Pi_B(r_{AB}(Q_B^*), \bar{Q}_B, r_C(Q_B^*)) \leq E\Pi_B(r_{AB}(Q_B^*), r_{BAC}(r_{AB}(Q_B^*), r_C(Q_B^*)), r_C(Q_B^*))$, 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. If $\frac{c_B}{w_B} = \frac{c_C}{w_C}$ and $P(D_1 \geq Z_1, D_2 \leq \hat{Q}_C^2) = P(D_2 \leq \hat{Q}_C^2)$, then $I\left(\hat{Q}_C^2 + Z_1\right) = 1 - \frac{c_A}{w_A} - \frac{c_B}{w_B}$ and the Pareto optimal equilibrium is $(Z_1, \hat{Q}_C^2 + Z_1, \hat{Q}_C^2)$. Otherwise, the Pareto optimal capacity equilibrium is unbalanced.

(iv) Finally, suppose that $\frac{c_B}{w_B} \leq \frac{c_C}{w_C} < 1 - \frac{c_A}{w_A}$, which is equivalent to $\bar{Q}_B < \hat{Q}_C^2 \leq \hat{Q}_B^2$. 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

$$\bar{Q}_B < Q_B < Z_1 + \hat{Q}_C^2. \tag{B.22}$$

As in (iii), there is a unique Nash equilibrium of the form $(r_{AB}(Q_B^*), r_{BAC}(r_{AB}(Q_B^*), r_C(Q_B^*)), r_C(Q_B^*))$, if there is a Q_B^* in the range (B.22) satisfying $I(Q_B) = 1 - \frac{c_A}{w_A} - \frac{c_B}{w_B}$. Otherwise, there is no equilibrium in this range. Since $I(\bar{Q}_B) = 0$, a solution to $I(Q_B) = 1 - \frac{c_A}{w_A} - \frac{c_B}{w_B}$ in this range exists if and only if $I(\hat{Q}_C^2 + Z_1) = F_1(Z_1) > 1 - \frac{c_A}{w_A} - \frac{c_B}{w_B}$.
For the range

$$Z_1 + \hat{Q}_C^2 \leq Q_B \leq \hat{Q}_A^1 + \hat{Q}_C^2, \tag{B.23}$$

$r_A(Q_B, r_C(Q_B)) = Q_B - \hat{Q}_C^2$ 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 (6), a capacity Q_B in (B.23) leads to such an equilibrium if and only if $P(D_1 \geq Q_B - \hat{Q}_C^2, D_2 \geq \hat{Q}_C^2) \geq \frac{c_B}{w_B} = P(D_1 \geq Z_2, D_2 \geq \hat{Q}_C^2)$, or, equivalently, $Q_B \leq Z_2 + \hat{Q}_C^2$. Also, note that $\bar{F}_1(Z_1) = P(D_1 \geq Z_1, D_2 \leq \hat{Q}_C^2) + P(D_1 \geq Z_1, D_2 \geq \hat{Q}_C^2) = \frac{c_A}{w_A} + P(D_1 \geq Z_1, D_2 \geq \hat{Q}_C^2)$. Then, $F_1(Z_1) > 1 - \frac{c_A}{w_A} - \frac{c_B}{w_B}$ if and only if $P(D_1 \geq Z_1, D_2 \geq \hat{Q}_C^2) < \frac{c_B}{w_B} = P(D_1 \geq Z_2, D_2 \geq \hat{Q}_C^2)$ if and only if $Z_1 > Z_2$. In summary, if $Z_1 > Z_2$, then, following a similar argument as in (iii), the equilibrium $(r_{AB}(Q_B^*), r_{BAC}(r_{AB}(Q_B^*), r_C(Q_B^*)), r_C(Q_B^*))$ identified above is the unique Pareto optimal equilibrium in the wholesale price region (iv)(a). On the other hand, if $Z_1 \leq Z_2$, then following a similar argument as in (ii), $\left(\min\left(Z_2, \hat{Q}_A^1\right), \min\left(Z_2, \hat{Q}_A^1\right) + \hat{Q}_C^2, \hat{Q}_C^2\right)$ is the unique Pareto optimal equilibrium in the wholesale price region (iv)(b). ■

Proof of Proposition 3. Consider the four wholesale-price regions identified in Theorem 3 (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 $\hat{Q}_B^2 = \hat{Q}_C^2$. 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)). Thus, the optimal prices will never fall in the interior of region (i). For any vector w in region (ii) such that $Z_2 \neq \hat{Q}_A^1$, the assembler could decrease either w_B or w_A (while remaining in the same region) until $Z_2 = \hat{Q}_A^1$, i.e., until $\bar{F}_1(Z_2) = \frac{c_A}{w_A}$. Since this increases the assembler's margin without affecting the suppliers' equilibrium capacity choices, this again (weakly) increases the assembler's profit, and thus reduces region (ii) to one in which w satisfies $1 - \frac{c_C}{w_C} \leq \frac{c_A}{w_A} = \bar{F}_1(Z_2)$. A similar argument applies to region (iv)(b). If $Z_2 > \hat{Q}_A^1$, then the assembler can decrease w_B until equality holds, while remaining in region (iv)(b). On the other hand, if $Z_2 < \hat{Q}_A^1$, 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 $\bar{F}_1(Z_2) \leq 1 - \frac{c_C}{w_C}$, then w_A can be decreased all the way until $\frac{c_A}{w_A} = \bar{F}_1(Z_2) \leq 1 - \frac{c_C}{w_C}$ while still staying within that region. If instead $1 - \frac{c_C}{w_C} < \bar{F}_1(Z_2)$, then the assembler can profitably decrease w_A until $\frac{c_A}{w_A} = 1 - \frac{c_C}{w_C} < \bar{F}_1(Z_2)$ 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's profit (weakly) increases by reducing w_A until $\frac{c_A}{w_A} = \bar{F}_1(Z_2)$. Combining these observations, among wholesale prices in regions (i), (ii) or (iv)(b), the assembler needs only consider vectors w satisfying (8) and any such vector results in equilibrium supplier capacities $(\hat{Q}_A^1, \hat{Q}_A^1 + \hat{Q}_C^2, \hat{Q}_C^2)$.

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, $\hat{Q}_C^2 = Q_B^* > \bar{Q}_B$, which implies

$\frac{c_C}{w_C} < 1 - \frac{c_A}{w_A}$. Also, $F_1(Z_1) = I\left(\hat{Q}_C^2 + Z_1\right) > I(\hat{Q}_C^2) = I(Q_B^*) = 1 - \frac{c_A}{w_A} - \frac{c_B}{w_B}$, which implies $Z_1 > Z_2$. 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 $\left(r_{AB}(Q_B), Q_B, \hat{Q}_C^2\right)$, 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 $\left(r_{AB}(Q_B), Q_B, \hat{Q}_C^2\right)$. 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 $F_1(Z_1) > 1 - \frac{c_A}{w_A} - \frac{c_B}{w_B}$ or, equivalently, $Z_1 > Z_2$; see the proof of Theorem 2(iii). Combining regions (iii) and (iv)(a), we then have that the assembler needs only consider wholesale price vectors w that satisfy the three conditions in (9). 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 $I(Q_B) = 1 - \frac{c_A}{w_A} - \frac{c_B}{w_B}$.
■

Proof of Corollary 2. The result follows from Proposition 3, by noting that under perfect positive correlation, $Z_1 = \bar{F}^{-1}\left(\frac{c_A}{w_A} + \frac{c_C}{w_C}\right)$ if $\frac{c_A}{w_A} + \frac{c_C}{w_C} \leq 1$, and $Z_2 = \hat{Q}_B$ if $\frac{c_C}{w_C} \geq \frac{c_B}{w_B}$ and $Z_2 = 0$ otherwise. Under perfect negative correlation $F_1(x) = \bar{F}_2(K - x)$, and

$$Z_1 = \hat{Q}_A^1 \text{ if } \frac{c_A}{w_A} + \frac{c_C}{w_C} < 1, \text{ and } Z_2 = F_1^{-1}\left(\frac{c_C}{w_C} - \frac{c_B}{w_B}\right) \text{ if } \frac{c_B}{w_B} \leq \frac{c_C}{w_C}. \blacksquare$$

Proof of Proposition 4. Note that the joint density of $D_i = \delta + X_i$, $i = 1, 2$, is given by $f(x_1 - \delta, x_2 - \delta)$. For any fixed wholesale price with associated equilibrium (Q_A^*, Q_B^*, Q_C^*) (for $\delta = 0$), it is immediate to verify that the Pareto optimal capacity equilibrium for $\delta > 0$ is $(Q_A^* + \delta, Q_B^* + 2\delta, Q_C^* + \delta)$. The assembler's profit for this capacity equilibrium is given by $E\Pi_0^\delta(w) = E\Pi_0(w) + \delta(p_1 - w_A - w_B + p_2 - w_B - w_C)$. Now let w^* and $w^*(\delta)$ be optimal assembler's wholesale prices for $\delta = 0$ and $\delta > 0$, respectively. Note that w^* is not optimal for $\delta > 0$, since, for example, $\frac{\partial E\Pi_0^\delta(w^*)}{\partial w_A} = \frac{\partial E\Pi_0(w^*)}{\partial w_A} - \delta = -\delta < 0$. Then, $E\Pi_0(w^*) - \delta(w_A^* + 2w_B^* + w_C^*) < E\Pi_0(w^*(\delta)) - \delta(w_A^*(\delta) + 2w_B^*(\delta) + w_C^*(\delta)) < E\Pi_0(w^*) - \delta(w_A^*(\delta) + 2w_B^*(\delta) + w_C^*(\delta))$, where the first inequality follows from the optimality of $w^*(\delta)$ for $E\Pi_0^\delta$ and the second one from the optimality of w^* for $E\Pi_0$. Thus, $c_A + 2c_B + c_C \leq w_A^*(\delta) + 2w_B^*(\delta) + w_C^*(\delta) < w_A^* + 2w_B^* + w_C^*$.

To prove the last part, suppose first that $p_1 - w_A^*(\delta) - w_B^*(\delta) \leq p_2 - w_B^*(\delta) - w_C^*(\delta)$. Note that an optimal assembler wholesale price vector $w^*(\delta)$ will satisfy one of the conditions (8) or (9) stated in Proposition 3. Suppose that $w^*(\delta)$ satisfies (9). Then, $w_C^*(\delta) > c_C$ (otherwise, the first two inequalities in (9) would not hold). That means that the assembler can slightly reduce $w_C^*(\delta)$, leaving the other two wholesale prices unchanged and maintaining the higher priority for product

2, and still fall within the region described by condition (9) (note that $\hat{Q}_C^2$ is increasing in w_C and $I(Q)$ is increasing in Q). In the remainder of the proof, we replace $w_i^*(\delta)$ by w_i to simplify notation. Consider now the effect of the slight reduction in w_C, satisfying (9), on the assembler's expected profit $E\Pi_0^\delta$. To that end, for small $\epsilon > 0$, note that

$$\frac{E\Pi_0^\delta(w_A, w_B, w_C - \epsilon) - E\Pi_0^\delta(w)}{\epsilon} = (p_1 - w_A - w_B)\frac{Ey_1(w_A, w_B, w_C - \epsilon) - Ey_1(w)}{\epsilon}$$
$$+ \frac{(p_2 - w_B - w_C + \epsilon)Ey_2(w_A, w_B, w_C - \epsilon) - (p_2 - w_B - w_C)Ey_2(w)}{\epsilon} + \delta.$$

As w_C decreases to $w_C - \epsilon$, the equilibrium value for component C decreases by $\bar{F}_2^{-1}\left(\frac{c_C}{w_C}\right) - \bar{F}_2^{-1}\left(\frac{c_C}{w_C - \epsilon}\right)$, and one can show that the equilibrium values for components A and B also decrease, but by a smaller amount than component C. Then, from (1), we have $Ey_1(w_A, w_B, w_C - \epsilon) - Ey_1(w) \geq \bar{F}_2^{-1}\left(\frac{c_C}{w_C - \epsilon}\right) - \bar{F}_2^{-1}\left(\frac{c_C}{w_C}\right)$. Thus, using the Mean Value Theorem, we have that

$$\frac{E\Pi_0^\delta(w_A, w_B, w_C - \epsilon) - E\Pi_0^\delta(w)}{\epsilon} \geq -\frac{p_1 - w_A - w_B}{f_2\left(\bar{F}_2^{-1}\left(\frac{c_C}{\tilde{w}_C}\right)\right)}\frac{c_C}{\tilde{w}_C^2} + E\min\left\{\bar{F}_2^{-1}\left(\frac{c_C}{\tilde{w}_C}\right), X_2\right\}$$

$$-\frac{p_2 - w_B - \tilde{w}_C}{f_2\left(\bar{F}_2^{-1}\left(\frac{c_C}{\tilde{w}_C}\right)\right)}\frac{c_C^2}{\tilde{w}_C^3} + \delta \geq -\frac{p_1 - c_A - c_B}{c_C f_2\left(\bar{F}_2^{-1}\left(\frac{c_C}{\tilde{w}_C}\right)\right)} - \frac{p_2 - c_B - c_C}{c_C f_2\left(\bar{F}_2^{-1}\left(\frac{c_C}{\tilde{w}_C}\right)\right)} + \delta, \qquad \text{(B.24)}$$

where $c_C < w_C - \epsilon < \tilde{w}_C < w_C$, and since $w_A \geq c_A$ and $w_B \geq c_B$. Because $\tilde{w}_C < p_2$, $0 < \bar{F}_2^{-1}\left(\frac{c_C}{\tilde{w}_C}\right) < \bar{F}_2^{-1}\left(\frac{c_C}{p_2}\right)$ and, as $f_2 > 0$ and continuous, $f_2\left(\bar{F}_2^{-1}\left(\frac{c_C}{\tilde{w}_C}\right)\right)$ is bounded below by a constant independent of δ. We then have that $E\Pi_0^\delta(w_A, w_B, w_C - \epsilon) - E\Pi_0^\delta(w) \geq \epsilon(-\kappa + \delta)$, for a constant κ independent of δ. Therefore, if δ is sufficiently large (say, $\delta > \kappa$), we have that the assembler benefits from a small reduction in $w_C^*(\delta)$ and so $w^*(\delta)$ cannot be optimal. Finally, a similarly large δ can be constructed if product 1 has the higher priority, and $\bar{\delta}$ can be defined large enough to ensure the result for either product's production priority. ■

Proof of Proposition 5. Let w_A^*, w_B^* and w_C^* be optimal wholesale prices in the system with a common component B. From Proposition 3, we have that

$$\bar{F}_1(Z_2^*) = \frac{c_A}{w_A^*}, \qquad \text{(B.25)}$$

where $\frac{c_B}{w_B^*} = P\left(D_1 \geq Z_2^*, D_2 \geq \bar{F}_2^{-1}(c_C/w_C^*)\right) \leq \bar{F}_1(Z_2^*)$. Now for the system with two dedicated components $B1$ and $B2$, consider the vector of wholesale prices $(w_A, w_{B1}, w_{B2}, w_C) = (w_A^*, c_{B1}w_A^*/c_A, c_{B2}w_C^*/c_C, w_C^*)$. This choice of wholesale prices leads to the same equilibrium capacity levels for components A and C, and to equilibrium capacity levels for components $B1$ and

$B2$ whose sum is equal to the equilibrium capacity level for component B in the system with a common component. However, from (B.25), we have that $w_{B1} = \frac{c_{B1}w_A^*}{c_A} \leq \frac{c_B w_A^*}{c_A} = \frac{c_B}{F_1(Z_2^*)} \leq w_B^*$. In addition, from (8), $w_{B2} = \frac{c_{B2}w_C^*}{c_C} \leq \frac{c_B w_C^*}{c_C} \leq w_B^*$. Thus, $w_{Bi} \leq w_B^*$ 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 $B1$ and $B2$. Thus, the assembler's profit is higher in the system with dedicated components. ■

Proof of Proposition 6. Note from Theorem 2 that expected sales of both products are positive only if $m > 1$. Then, capacity imbalance can result from the assembler's optimal pricing only if $p_1 > 2(c_A + c_B)$ and $p_2 > 2(c_B + c_C)$. Assume now that these two inequalities hold. Because $p_2 \geq p_1$, we have that product 2 has the higher priority for any feasible mark-up m if

$$\frac{p_2}{c_B + c_C} \geq \frac{p_1}{c_A + c_B}. \tag{B.26}$$

If the reverse inequality holds (which can only happen if $c_A \leq c_C$), then product 2 has the higher priority for $0 \leq m \leq \frac{p_2 - p_1}{c_C - c_A} - 1$ and product 1 has the higher priority for $\frac{p_2-p_1}{c_C-c_A} - 1 < m \leq \frac{p_2}{c_B+c_C} - 1$. Note that expected sales of the product i with the higher priority are $E \min\left\{\overline{F}_i^{-1}\left(\frac{1}{1+m}\right), D_i\right\}$, regardless of whether the other product is produced or not. We next consider two possible cases – one where (B.26) holds and one where it does not.

First, if (B.26) holds, then product 2 has the higher priority for all m and the assembler's expected profit is given by $E\Pi_0(m) = (p_1 - (1+m)(c_A + c_B))Ey_1 + (p_2 - (1+m)(c_B + c_C))Ey_2$, where $Ey_1 = 0$ for $m \leq 1$. The assembler's optimal mark-up results in capacity imbalance (i.e., the assembler selects $m > 1$) if the derivative of $(p_2 - (1+m)(c_B + c_C))Ey_2$ with respect to m, evaluated at $m = 1$, is strictly positive. (Note that $(p_2 - (1+m)(c_B + c_C))Ey_2$ is unimodal in m.) This condition reduces to

$$-(c_B + c_C)E \min\left\{\overline{F}_2^{-1}(1/2), D_2\right\} + \frac{1}{8}(p_2 - 2(c_B + c_C))\frac{1}{f_2\left(\overline{F}_2^{-1}(1/2)\right)} > 0.$$

The result then follows by defining $M_1 = 1$ and $M_2 = 8f_2\left(\overline{F}_2^{-1}(1/2)\right)E \min\left\{\overline{F}_2^{-1}(1/2), D_2\right\} + 1$. (Because $f_2(\cdot)$ is continuous and $f_2(x) > 0$ for all $x \geq 0$, we have that $F_2(x)$ is continuous and strictly increasing. Then, $1 = \overline{F}_2(0) > 1/2$ implies that $0 < \overline{F}_2^{-1}(1/2)$, so that $M_2 > 1$.)

Next, if (B.26) does not hold, then product 2 has the higher priority only for m below $\frac{p_2-p_1}{c_C-c_A} - 1$. Note that $E\Pi_0(m)$ is continuous in m, since it is continuous within each interval where only one of the products has the higher priority, and for $m = \frac{p_2-p_1}{c_C-c_A} - 1$ both products have equal margins and the assembler's profit is the same regardless of which product is assigned the higher priority. We now further consider two cases. If $\frac{p_2 - p_1}{c_C - c_A} - 1 > 1$, then the analysis is the same as the one for the case where (B.26) holds. Finally, consider the case where $\frac{p_2 - p_1}{c_C - c_A} - 1 \leq 1$. Defining $M_i = 8 f_i \left(\overline{F}_i^{-1} (1/2) \right) E \min \left\{ \overline{F}_i^{-1} (1/2), D_i \right\} + 1$, we have that $M_i > 1$ implies that $E \min \left\{ \overline{F}_i^{-1} \left(\frac{1}{1+m} \right), D_i \right\}$ has a striclty positive derivative at $m = 1$, for $i = 1, 2$. Then, for $0 \leq m \leq \frac{p_2 - p_1}{c_C - c_A} - 1$, only product 2 is produced and the assembler's profit is increasing in this interval (since $M_2 > 1$). For $\frac{p_2 - p_1}{c_C - c_A} - 1 < m \leq 1$, only product 1 is produced and the assembler's profit is still increasing in this interval (since $M_1 > 1$). Then, the assembler's profit is increasing for $0 \leq m \leq 1$ since $E\Pi_0(m)$ is continuous. Thus, $M_1 > 1$ implies that the assembler's optimal mark-up is strictly higher than 1, concluding the result. ■