1. Introduction

Due to their ability to keep finished-goods inventories low while providing reasonably responsive service to customers, assemble-to-order systems have become increasingly popular in recent years. Under such a system, a wide variety of finished products can be produced from a relatively small number of components. Perhaps the best-known example of this approach to manufacturing is that of Dell Computer. Dell does not stock fully assembled computers, but relies on inventories of components (monitors, processors, hard drives, etc.) which are stored both at its own facility (in very small quantities) and at nearby facilities operated by its suppliers. When a customer order arrives, these components can be quickly assembled into whatever specific product the customer wants, and the product can usually be delivered to the customer within 5 to 6 days (Magretta 1998). In the automobile industry, Renault, BMW and Volvo are pioneers in building cars based on direct orders from customers. To shorten leadtimes, key suppliers have set up plants close to the auto makers’ facilities, and some even connect their operations directly to the assembly plant with conveyors. Finally, a version of assemble-to-order is practiced by amazon.com and all other catalog and online retailers. Here a “product” corresponds to a particular combination of items in an order. When a customer order arrives, these items must be picked from “component” inventory and assembled before shipping.

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

In this paper, we explore the impact of decentralized decision making on the behavior of assemble-to-order systems. Specifically, we consider a system where the assembler buys three components (two product-specific and one common) from three independent suppliers and produces two end products to satisfy stochastic customer demands. The assembler sets wholesale prices paid to the suppliers, and the suppliers then decide how much component capacity to install. At the end of a single selling season, demands are observed, and the assembler makes production decisions based on the capacity constraints. For this setting, we analyze the assembler’s optimal pricing decision and the equilibrium capacities that the component suppliers choose in response to those prices.

To investigate the impact of decentralization, we also derive the optimal/equilibrium pricing/capacity decisions for two variations of the above system — one system under a centralized decision making and one where the common component is replaced by two product-dedicated components. By comparing the behavior in these three systems, we identify various types of inefficiencies that arise from decentralization and that are closely related to the multi-component, multi-product structures analyzed here.

There exists a significant body of research analyzing performance measures and inventory policies for centralized assemble-to-order systems. A complete list of work in this area can be found in the review of ATO system by Song and Zipkin (2003). Another related stream of work is the research on decentralized single-product assembly systems. Gerchak and Wang (2004) study a setting that is similar to the one considered here, but with only a single finished product. 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.

Other related work includes the study of capacity investment, component commonality and resource flexibility. Relevant research on component commonality includes 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 (2004) investigates the relationship between the commonality and
flexible capacity problems. For a recent review of work addressing game-theoretic capacity investment by multiple agents, see Van Mieghem (2003).

2. Model

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

![Figure 1: ATO System](image)

Components $A$, $B$ and $C$ are produced by suppliers $A$, $B$ and $C$ respectively, while a single assembler makes assembly decisions for the two products. Each supplier incurs a constant per-unit cost for each unit of capacity installed or reserved. We assume that supplier production costs (once capacity is installed) are zero. Demands for the two products are independent random variables. The market prices for the two products are exogenous and we assume, without loss of generality, that the price of product 2 is higher than that of product 1. We assume that both products have a positive profit margin, and that the demand distributions and all cost parameters are common knowledge.

All firms are geographically close and lead times for assembly are negligible, so that final products can be assembled to order after demand is observed. However, suppliers need
to plan in advance the amount of capacity they reserve for their component production activities. As a result, the sequence of events is as follows. First, the assembler acts as the Stackelberg leader by choosing the wholesale price it will pay to each supplier for each unit of component produced. Then, the suppliers simultaneously install or reserve their production capacities. Next, the assembler observes consumer demands, decides how many units of each product 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.

3. Main Results

We first study the behavior of the system under centralized control. Product 2 has the higher production priority due to its higher price. In this setting, we show that the optimal capacity investment strategy for the centralized system is to invest in all three components, and we identify a condition on the cost and revenue parameteris under which risk-pooling of the common component occurs (i.e., situations where the capacity of the common component is strictly less than the sum of the capacities of the two dedicated components).

Next, we analyze the equilibrium behavior of the decentralized system. We show that, despite its inherent complexity, the system is reasonably well behaved. Specifically, for any choice of wholesale prices by the assembler, a unique Pareto-optimal equilibrium always exists in the suppliers’ capacity game. Also, the assembler’s optimal wholesale prices lie in one of two regions, and these regions result in different behaviors in the subsequent capacity game – one region leads to risk pooling with respect to the common component, while the other does not. We also provide a partial characterization of the optimal assembler prices, and identify which price choices lead to risk-pooling of the common component.

We compare the behavior of the above systems both analytically and numerically. The comparison yields a number of interesting insights. First, consistent with what has often been observed in other decentralized systems, capacity stocking levels in the decentralized system are always lower than in the centralized system. In addition, this multi-product setting allows us to identify new sources of inefficiency resulting from decentralization. Since the final production priorities in the decentralized system are determined by each product’s profit margin – i.e., the market price minus the wholesale prices paid to the component suppliers – rather than the market price alone, we find that these priorities may switch
from what they are under centralized decision making. Also, risk pooling with respect to the common component occurs less frequently in the decentralized system. In particular, we find that (i) if risk pooling of the common component did not occur in the centralized system, then it did not occur in the decentralized system under the assembler’s optimal pricing; (ii) in most of the scenarios where risk pooling of the common component occurred in the optimal solution of the centralized system, risk pooling did not occur in the decentralized system under the assembler’s optimal pricing.

Our numerical study suggests a few possible explanations for this reduced frequency of risk pooling in the decentralized setting. First, the low component capacity levels associated with understocking inhibit risk pooling. When the capacities are low, the probability of consuming all units of product-dedicated components is high so there is little benefit from risk pooling of the common component. A second explanation is related to the shift in incentives that results from decentralization. In a centralized setting, the optimal level of risk pooling is determined by trading off the cost savings from reducing capacity of the common component against some loss in expected sales. In a decentralized setting, the supplier providing the common component faces a trade off of this type and may wish to risk pool, whereas the assembler experiences only the negative aspect of risk pooling – the possible loss of sales of products. Although risk pooling is less common in the decentralized system, our numerical study suggests that when it occurs, the component capacities in the decentralized system may be more imbalanced – implying a higher degree of risk pooling – than in the integrated system.

We finally explore the impact of decentralization on the value of having a common component. That is, does the flexibility represented by a common component (vs. two dedicated components) always make the assembler better off (or at least not worse off)? To address this question we consider a variation of the ATO system in which the two products use only dedicated components. We find that if the decentralized system with a common component exhibits no risk pooling with respect to the common component, then the assembler is always better off having dedicated components for the two products. Our numerical study suggests that a system with dedicated components may also be preferred by the assembler when risk pooling of the common component occurs in the system with a common component. So, the apparent flexibility provided by a common component may actually hurt the performance of the assembler in a decentralized system, even while it may help under centralized control.
References


