A Complexity Model for Assembly Supply Chains in the Presence of Product Variety and its Relationship to Cost

Hui Wang[1] Göker Aydın[2] S.Jack Hu[1,3]

[1] Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor

[2] Kelley School of Business, Indiana University, Bloomington

[3] Department of Mechanical Engineering, University of Michigan, Ann Arbor

Abstract

We propose a complexity measure for assembly supply chains, based on the concept of information entropy. This complexity measure takes into account factors such as the supply chain configuration, the level of variety offered by each node of the supply chain, and the demand ratios across all the variants offered by a node. We investigate the relationship between the complexity and the cost of an assembly supply chain. We first study the degree of consistency between the cost and complexity criteria when comparing assembly supply chains with the same configuration but different levels of product variety. We show that the cost and the complexity are equivalent under certain conditions, in the sense that both of them rank a given set of supply chains in the same order. Even when these conditions do not hold, our numerical study demonstrates that the cost and complexity criteria rank supply chains consistently in an overwhelming majority of cases. Furthermore, the inconsistencies occur mostly when the supply chains have very small cost differences. We then study how well the cost and complexity criteria agree when comparing assembly supply chains with the same level of product variety, but different configurations. In such cases, we show that the consistency between cost and complexity criteria is not very reliable. Overall, our results suggest that the complexity measure is a good proxy for the cost of an assembly supply chain when evaluating alternative levels of product variety that will be delivered by a given supply chain configuration, but not so when evaluating alternative supply chain configurations to deliver a given level of product variety.

1. Introduction

In many industries, previous decades brought about an explosion of product variety. For example, the number of distinct vehicle models offered in the US rose from 44 in 1969 to 168 in 2005 (Ward's Automotive Yearbook, 1970 & 2006). The number of styles of running shoes went from five in the early 70s to 285 in the late 90s (1998 Annual Report of the Federal Reserve Bank of Dallas). The growth of product variety brings with it many challenges. For instance, several studies suggest that high product variety has a negative effect on manufacturing performance such as increasing manufacturing complexity, lowering productivity and degrading quality (MacDuffie et al., 1996; Fisher and Ittner, 1999). Many manufacturing firms have adopted modular product designs to cope with the challenges posed by high variety. A modular design decomposes the product into several modules with standard interfaces, and high product variety is achieved through the combinational assembly of different modules. In doing so, modular designs support a large variety of end products while still maintaining high volume for each module, thereby creating economies of scale in production of modules and components (see, for example, Swaminathan, 2001).

One important trend in supply chain management, enabled by modular product designs, is the emergence of modular assembly supply chains. In a modular assembly supply chain, the manufacturer apportions the product into different modules, most of which are outsourced to and assembled by suppliers. The manufacturer does only the final assembly of a few modules. For instance, Volvo's S80 model is assembled from 23 different modules, delivered directly to the final assembly line by 17 different assembly units, 11 of which are operated by suppliers (Fredriksson, 2006).

In many instances, most notably when bringing a new product to the market, a manufacturing firm must make design decisions about how to modularize the product and what level of variety to offer for each module. Of course, such modularization and variety decisions influence the cost to be incurred across the entire supply chain. Thus, a manufacturing firm would benefit from getting a handle on the costs that correspond to different modularization and variety decisions. However, cost models require the estimation of many parameters, e.g., manufacturing costs, holding and shortage costs, transportation costs, production and transportation leadtimes. Furthermore, sophisticated cost models present many analytical challenges due to the network structure of an assembly supply chain, and even more so when product variety is taken into account. Hence, such cost models are difficult to develop, analyze and utilize when making modularization and variety decisions in assembly supply chains.

In this paper, we propose a new performance measure for assembly supply chains: complexity. This complexity measure is based on information entropy as applied to an assembly supply chain and recognizes the product variety to be offered by the manufacturer. We explore if, when and to what degree this complexity measure is consistent with cost. In particular, we illustrate the usefulness of this complexity measure by considering two different scenarios. First, we consider a manufacturer who has already settled on a given network of suppliers, but needs to compare different levels of product variety to be offered by the supply chain. We then analyze the agreement between the cost and complexity criteria when evaluating different levels of product variety. Second, we consider a manufacturer that has already decided to offer a given assortment of variants, but needs to choose whether to use a modular or non-modular assembly supply chain to deliver that assortment. Once again, we analyze the agreement between the cost and complexity criteria when choosing from a modular or non-modular supply chain.

When comparing supply chains with the same configuration but different levels of product variety, we show that the cost and the complexity are equivalent under certain conditions, in the sense that both of them rank a given set of supply chains in the same order. We also conduct an extensive numerical study to check how the agreement between cost and complexity is affected by different factors (e.g., number of echelons, number of suppliers, the distribution of consumer demand across variants offered by the manufacturer, etc.). Even when the sufficient conditions for equivalence do not hold, our numerical study demonstrates that the cost and complexity criteria rank supply chains consistently in an overwhelming majority of cases. Furthermore, the inconsistencies occur mostly when the supply chains have very small cost differences. In contrast, when comparing supply chains with the same level of product variety, but different configurations, we show that the consistency between the cost and complexity criteria is not very reliable. Overall, our results suggest that the complexity measure is a good proxy for the cost of an assembly supply chain when evaluating alternative levels of product variety that will be delivered by a given supply chain configuration, but not so when evaluating alternative supply chain configurations to deliver a given level of product variety.

The paper is organized as follows. In Section 2, we review the related literature. In Section 3, we introduce our complexity measure and we describe the cost model that will serve as a benchmark. We then study the relationship between cost and complexity as a function of variety (in Section 4) and as a function of supply chain configuration (in Section 5). Section 6 concludes the paper. All the proofs are included in Appendix B of the online supplement.

2. Literature Review

An *assembly system* is a tree-structured network, where each node produces a single item and has at most one successor. Early research on assembly systems seeks to characterize the optimal ordering policy at each node of the system to minimize the total system cost. Schmidt and Nahmias (1985) use a finite-horizon model to explore an assembly system with two components. Rosling (1989) studies the periodic review, infinite-horizon inventory problem and shows that the assembly system can be reduced to an equivalent serial system. Subsequent research has explored more elaborate assembly systems, for example, assembly systems in which not only the demand but also the supply may be random (e.g. Gurnani et al., 1996; Bollapragada et al., 2004), assembly systems in which inventory levels may be supplemented with returns (e.g. DeCroix and Zipkin, 2005) and decentralized assembly systems (e.g. Bernstein and DeCroix, 2006; Gerchak and Wang, 2004). In contrast to an assembly system, an *assemble-to-order system* is traditionally defined as a two-echelon system, in which multiple products are assembled from a given set of components. For a review of the research on assemble-to-order systems, see Song and Zipkin (2003). Assemble-to-order systems present one additional challenge that is not present in the single-product assembly system: allocating a component's inventory among several different products that use the component (hence, the need to restrict attention to two-echelon systems).

The assembly supply chain we consider in this paper has elements of both an assembly system (in that the supply chain has the same tree structure as the assembly system) and an assemble-to-order system (in that each node of the supply chain produces multiple variants of its output, each of which goes into multiple items in the downstream node). In a supply chain as complicated as the one considered in this paper, even when one limits attention to a given ordering policy, it is still very challenging to obtain an exact expression for inventory costs. As our main focus in this paper is to introduce a complexity measure, we do not attempt to develop an exact expression for the supply chain costs. To analyze the agreement between cost and complexity, we make simplifying assumptions to obtain only a rough estimate of the costs that will be incurred in the supply chain. In particular, we assume the nodes in the assembly supply chain can be decoupled, after which each node follows an order-up-to policy to satisfy a certain service level objective.

One application of the proposed complexity measure is to make product variety decisions for an assembly supply chain with a given network of suppliers. The operations literature has given scant attention to the management of product variety in a supply chain. Kurtulus and Toktay (2005) analyze if and when a retailer should delegate variety and pricing decisions to a manufacturer. Singh et al. (2005) examine the effect of supply chain structure, in particular, the effect of drop-shipping, on the optimal assortment. Aydin and Hausman (2009) consider the coordination of product variety decisions in a single-retailer, single-manufacturer supply chain. Similar to these papers, we use a demand model where the customer demand is allocated among many variants of a product. Given that we are modeling an assembly supply chain, the supply relationships in our model are significantly more complicated than in this earlier work. Instead of developing a detailed analysis of cost-minimizing or profit-maximizing variety levels, we focus on the use of complexity measure to evaluate alternative levels of product variety. In that sense, our work complements the earlier literature by showing the usefulness of complexity measure when choosing the level of variety.

Another application of the proposed complexity measure is to compare the efficiency of two different supply chain structures: modular versus non-modular assembly supply chains. Such supply chain configuration decisions received limited attention in the literature. Bernstein and DeCroix (2004) analyze a modular assembly supply chain to determine the optimal transfer prices between the manufacturers and subassemblers. They identify if and when a modular assembly supply chain is more beneficial than a non-modular assembly supply chain. Thomas and Warsing (2007) consider a service supply chain in which there is external demand not only for the modular end product, but also for the components that make up the end product. They evaluate the savings that could be obtained if the supply chain adjusted its inventory levels not only by placing orders with external vendors, but also by utilizing assembly and disassembly operations to shift the inventory from the end product to the components or vice versa. Taking an empirical approach, Randall and Ulrich (2001) use data from U.S. bicycle industry to examine the relationship among product variety, supply chain structure and system performance. They show that firms that match their supply chain structure to the type of variety they offer often outperform those that fail to make such choices, where the performance is measured based on cost and revenue analysis. Salvador et al. (2004) also use empirical data to explore how a firm's supply chain, defined as the whole of its supply, manufacturing and distribution networks, should be configured, when different degrees of customization are offered. In a similar vein, we compare modular and non-modular assembly supply chains, with the goal of illustrating the usefulness of the complexity measure.

With different applications in mind, several different definitions of complexity have been proposed by researchers. See, for example, Suh (2005) for a complexity definition particularly applicable in product design, and Cover and Thomas (1991) for a discussion of Kolmogorov complexity, which is a measure of computational resources needed to describe a string of text. A commonly-used complexity definition is based on the information entropy, proposed by Shannon (1948) in the context of communication systems. Shannon's information entropy is a measure of the uncertainty surrounding the outcome of a random experiment. The information entropy has been used to study complexity in several different areas, including communication networks, biology, and physics. Shannon's information entropy has been used to measure the complexity of manufacturing systems as well. Deshmukh et al. (1998) derive an information-theoretic entropy measure of complexity for a given combination and ratio of part types to be produced in a manufacturing system. Zhu et al. (2008) study the operator choice complexity in mixed model assembly lines and develop a methodology to find the optimal assembly sequence to minimize manufacturing complexity.

3. Model Description

Consider an assembly supply chain, where each node can have multiple suppliers, but a given node cannot be a supplier to multiple nodes. Suppose there are n nodes in the assembly supply chain. As a convention, we let node n be the final assembler. In keeping with our focus on product variety, we assume that each node in the most upstream echelon can produce a number of variants. A downstream node can potentially assemble any combination of the variants provided by its suppliers, and each combination counts as a distinct variant. See Figure 1 for an illustration. We assume that if variant v of node j is used when producing variant u of node i, then one unit of variant v goes into one unit of variant u. This assumption is merely to simplify the exposition.

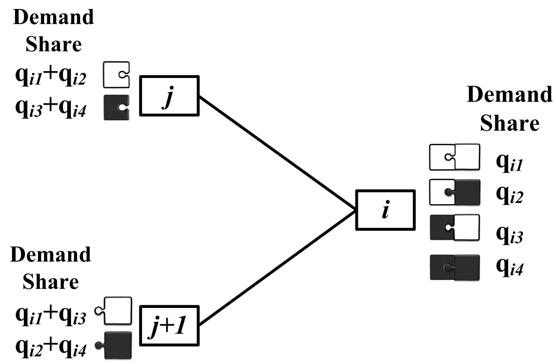

Figure 1: Relationship between the demand vectors of upstream and downstream nodes. In this figure, nodes j and $j{+}1$ are suppliers to node i. Each of node j and $j{+}1$ produces two variants, which results in node i producing four different variants of its product, where each variant corresponds to a distinct combination of the variants supplied by nodes j and $j+1$.

We define the following notation:

V_i : the number of variants that node i can produce, $i = 1, \ldots, n$.

S_i : the set of nodes that are suppliers to node i, $i = 1, \ldots, n$.

A_{ijv} : the set of variants produced at node i that use variant v from node j, where node j is a supplier to node i, i.e., $j \in S_i$.

Let q_{iv} denote the fraction of node-i demand that belongs to variant $v = 1, \ldots, V_i$. Hereafter, we refer to q_{iv} as the demand share of variant v at node i. In addition, define the vector $\boldsymbol{q}_i := (q_{i1}, q_{i2}, \ldots, q_{i,V_i})$, which captures the mix ratio of the variants produced by node i. Hereafter, we refer to $\boldsymbol{q}_i$ as the demand vector of node i. The final assembler's demand vector, $\boldsymbol{q}_n$, determines how the demand at all other upstream nodes is allocated among several variants produced at those nodes (see Figure 1). In particular, using the notation introduced so far, we have the following relationship between the demand share of variant v at node j, q_{jv}, and the demand vector of node i, $\boldsymbol{q}_i$, where node j is a supplier to node i, (i.e., $j \in S_i$):

$$q_{jv} = \sum_{u \in A_{ijv}} q_{iu}, j \in S_i. \tag{1}$$

3.1 Complexity Model of An Assembly Supply Chain

A measure of the assembly supply chain performance should take into account factors such as the supply chain's configuration, the number of variants produced at each node of the supply chain, and the demand vector of each node. In an effort to capture these factors, we use a complexity measure based on Shannon's information entropy. The information entropy of a random experiment is a measure of the uncertainty about the outcome of the random experiment (Shannon, 1948). According to Shannon's definition, the information entropy of a random experiment with r possible outcomes, whose probabilities of occurrence are $p_1, p_2, ..., p_r$, is

$$H = -\Upsilon \sum_{i=1}^{r} p_i \log_2 p_i, \tag{2}$$

where the positive constant Υ amounts to a scaling factor.

In what follows, we offer two different complexity measures for an assembly supply chain, *node-based* complexity and *arc-based* complexity, and we relate these complexity measures to Shannon's information entropy.

Node-based Complexity: Let H_N denote the node-based complexity of the supply chain and define it as

$$H_N = -\sum_{i=1}^{n} \sum_{v=1}^{V_i} \frac{q_{iv}}{n} \log_2 \frac{q_{iv}}{n}. \tag{3}$$

One possible interpretation of this complexity measure is the following. Suppose, for each node, we form a pool of variants produced by that node, where each variant is represented in a quantity proportional to its demand share. Consider the random experiment where we first pick a node at random, and we then pick one item from this node's pool. The probability of picking variant v of node i is q_{iv}/n (since we pick node i with probability $1/n$ and variant v of node i with probability q_{iv}). The information entropy of this random experiment is given by (3) and yields our node-based complexity measure. Loosely speaking, the node-based complexity indicates the level of uncertainty about what variant in the supply chain will be demanded next.

Arc-based Complexity: The node-based complexity has the attractive property that it takes into account the number of nodes in the supply chain as well as the split of demand among the variants. However, there is one aspect of the supply chain that is ignored by the previous definition. Two supply chains may have the same number of nodes, but they can be very different from one another due to the configuration of supply relationships among the nodes. See, for example, Figure 2 for an illustration. In order to better capture the information about supply chain configuration, we offer an alternative complexity measure, the arc-based complexity. Let L_i denote the number of suppliers of node i, i.e., $L_i \equiv |S_i|$. In order to have a complete picture of the flows into the supply chain, we also assume that there is a virtual supplier that is linked to each of the nodes in the most upstream echelon of the supply chain. Hence, $L_i = 1$ for all suppliers i in the most upstream echelon. Let K denote the number of arcs in this supply chain, including those that connect the virtual supplier with the suppliers in the most upstream echelon, i.e., $K = \sum_{i=1}^{n} L_i$. We let H_A denote the arc-based complexity and define it as follows:

$$H_A = -\sum_{i=1}^{n} \sum_{j \in S_i} \sum_{v=1}^{V_i} \frac{q_{iv}}{K} \log_2 \frac{q_{iv}}{K} \tag{4}$$

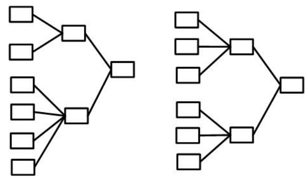

Figure 2: Two assembly supply chains with the same number of nodes, but distinct configurations. Node-based complexity ignores how the nodes are linked to form a certain configuration, which motivates the need for arc-based complexity.

This complexity measure can also be interpreted in the context of Shannon's information entropy. Suppose again, for each node, we form a pool of variants produced by that node, where each variant is represented in a quantity proportional to its demand share. Consider the random experiment where we first pick an arc of the supply chain at random, and we then pick one item from the pool of this arc's end-node. The probability of picking a certain arc a, which connects node i to a supplier node $j \in S_i$, and then picking variant v from the pool of node i is q_{iv}/K (since we pick arc a with probability $1/K$ and variant v of node i with probability q_{iv}). It can be shown that the information entropy of this random experiment is given by (4), which yields our arc-based complexity measure and the arc-based complexity defined by (4) can be simplified as follows (see Appendix A of the online supplement for a derivation):

$$H_A = -\sum_{i=1}^{n} \sum_{v=1}^{V_i} L_i \frac{q_{iv}}{K} \log_2 \frac{q_{iv}}{K}. \tag{5}$$

Loosely speaking, the arc-based complexity indicates the level of uncertainty about the next flow of material that will occur in the supply chain.

The complexity measures we propose are useful to the extent that they agree with a more direct performance measure like cost. In the next subsection, we describe a simple cost model for an assembly supply chain that will be used as a benchmark to study the effectiveness of node-based and arc-based complexity measures.

3.2 Cost Model of An Assembly Supply Chain

Several types of costs can be incurred in an assembly supply chain, including inventory costs, manufacturing costs and transportation costs. In this section, we describe a simple cost model for an assembly supply chain based on inventory costs only. This simple cost model will be used to analyze the effectiveness of the complexity measure in evaluating supply chain performance. That is, we will analyze whether more costly supply chains rank higher on the complexity scale as well. In the conclusion section, we discuss certain assumptions under which our results extend to include other types of costs such as manufacturing and transportation costs.

3.2.1 Overview of the Cost Model

We assume that each node in the supply chain keeps inventory of its own variants, but no inventory of the inputs from the suppliers. Hence, every time node i decides to replenish the inventory of one of its variants, it must first buy the necessary inputs and then assemble those. The assumption that the nodes do not hold any component inventory comes close to reality in environments where the supply chain partners, utilizing lean production principles, are located close to one another. At the other extreme, one could assume that all the nodes keep inventory of components, but no inventory of the finished goods. This assumption would be close to reality in environments where all supply chain partners assemble to order. All of the results in the next section continue to hold if we assume that only component inventories are held.

In an assembly system, the ordering decisions for an item tend to depend on the inventory levels for other items that it will be assembled with. This feature of an assembly system makes it challenging to characterize the optimal inventory decisions. In addition to this difficulty that is inherent in any assembly system, our assembly supply chain presents further challenges due to the presence of product variety. In our assembly supply chain, each item supplied by a node is an input to multiple variants at the downstream node. This one-to-many relationship brings up an important challenge: After receiving the delivery of an input, how should a node allocate the delivered quantity among many variants that must share the supply of this input? The allocation policy and inventory levels at upstream locations interact in complicated ways to determine the inventory costs, making it hard to analytically characterize the inventory cost of the supply chain.

In order to obtain a simple enough cost benchmark, we utilize a cost model that relies on two simplifying assumptions. First, we assume that each node uses an order-up-to policy for each variant so as to achieve a certain a service level objective for that variant. Such policies are practical and they are commonly used. Second, and perhaps more restrictively, we assume that the appropriate order-up-to levels can be determined by decoupling the assembly supply chain into a collection of standalone nodes. We utilize the decoupling assumption to create a tractable cost benchmark. The higher the service levels across the supply chain, the closer to reality the decoupling assumption is and the better the approximations obtained. Similar assumptions have been used by others when analyzing inventory costs in the presence of product variety. For example, to model a production environment with delayed differentiation, Lee and Tang (1997) consider a model where a series of common operations eventually fork into two distinct series of operations to produce two distinct products. In their model there exists buffer inventory between each pair of operations, and the buffer at each operation needs to satisfy a certain service level objective. Lee and Tang (1997) also make a decoupling assumption, essentially assuming that the buffer size at each operation can be determined independent of other operations. We next describe the details of our cost model.

3.2.2 Derivation of the Supply Chain Cost

We assume that each node uses a periodic-review, order-up-to policy in an infinite-horizon setting with leadtime. The order-up-to level for each variant at each node is chosen to satisfy what is commonly called a type-1 service level objective, that is, the probability of meeting the demand in full in a given period. For simplicity, we assume that all variants at all nodes have the same type-1 service level objective, denoted by α, where $0 < \alpha < 1$. It would not be difficult to extend the model so that the service level objectives differ across variants and/or nodes. The leadtime for replenishing the inventory of a variant covers the span of activities starting with the purchase of the inputs and ending with the assembly of the variant. Let l_i denote the leadtime to replenish the inventory of a variant at node i.

The timing of events is as follows: (1) At the beginning of a period, each node $i = 1, \ldots, n$ receives the units that it ordered l_i periods ago. (2) Each node then reviews the inventory positions of its variants (inventory on hand plus pipeline inventory that has been ordered but not yet received) and orders enough of each variant to bring its inventory position to the desired order-up-to level. (3) The demand for each variant at each node is realized. (4) After the demands are realized, node i incurs an overage cost of c_{io} per unit of leftover inventory and a unit underage cost of c_{iu} per unit of shortage. We assume that all the unmet demand is backordered. Notice our implicit assumption that the underage and overage costs are the same across all the variants of node i.

As for the demand model, suppose that the total per-period demand faced by the final assembler (node n) follows a Poisson distribution with rate λ. Given the final assembler's demand vector, denoted by $\boldsymbol{q}_n := (q_{n1}, q_{n2}, \cdots, q_{n,V_n})$, the demand for variant v of the final assembler also follows a Poisson distribution with rate λq_{nv}, $v = 1, 2, \cdots, V_n$. Furthermore, the per-period demands are independent across variants. Hence, using the normal approximation to Poisson random variables, we assume that the per-period demand for variant v of the final assembler is normal with mean and variance λq_{nv}, $v = 1, \ldots, V_n$, and the per-period demands are independent across variants. This approximation is valid when λ is large. Furthermore, for analytical convenience, we assume that the per-period demands of a variant are independent and identically distributed (i.i.d.) over periods. Similarly, for the remaining nodes 1 through $n-1$, we assume that the per-period demand for variant v of node j is normally distributed with a mean and variance λq_{jv}, $v = 1, \ldots, V_j$, where q_{jv} is given by (1). Notice that the final assembler's demand vector influences both the mean and variance of the demands at upper echelons, since the demand shares of variants at upper echelons are determined by the final assembler's demand vector. We assume that, at each of nodes

1 through $n-1$, the demands are independent across variants and, for a given variant, the demands are independent across periods.

Using standard arguments from inventory theory, the order-up-to level for variant v of node i is given by

$$y^*_{iv} = (l_i + 1)\lambda q_{iv} + z(\alpha)\sqrt{(l_i + 1)\lambda q_{iv}}, \tag{6}$$

where $z(\alpha)$ is α-fractile of the standard normal distribution, i.e., $z(\alpha)$ is such that $\Phi_N(z(\alpha)) = \alpha$ where $\Phi_N(\cdot)$ is the cumulative distribution function (c.d.f.) of the standard normal distribution. Furthermore, the expected per-period inventory cost for variant v of node i is given by

$$I_{iv} = \sqrt{(l_i + 1)\lambda q_{iv}}\Big[(c_{io} + c_{iu})\alpha z(\alpha) - c_{iu}z(\alpha) + (c_{io} + c_{iu})\phi_N(z(\alpha))\Big] \tag{7}$$

where $\phi_N(\cdot)$ is the probability density function (p.d.f.) of the standard normal distribution. For notational convenience, we let

$$C_i := (c_{io} + c_{iu})\alpha z(\alpha) - c_{iu}z(\alpha) + (c_{io} + c_{iu})\phi_N(z(\alpha)),$$

and we refer to C_i as the *cost coefficient* of node i. Then, the expected per-period inventory cost of the entire assembly supply chain is given by

$$I = \sum_{i=1}^{n}\sum_{v=1}^{V_i} I_{iv} = \sum_{i=1}^{n}\sum_{v=1}^{V_i} C_i\sqrt{(l_i + 1)\lambda q_{iv}}. \tag{8}$$

Thanks, in particular, to our decoupling assumption, this cost expression is simple. The simplicity of the cost expression, however, does not necessarily make it an easy tool for practical purposes. To obtain the supply chain costs through this expression, one would need to estimate the unit underage cost, the unit overage cost and the leadtime for each node of the supply chain. In contrast, the complexity measures we introduced earlier do not require the estimation of these parameters.

4. Complexity versus Cost When Choosing Variety

Consider a final assembler who is committed to working with a given network of suppliers and needs to decide what level of variety to offer through this already finalized supply chain. In this section, we analyze if, when and to what degree the cost and the complexity criteria lead to consistent recommendations for such a supply chain. Different variety decisions lead to different demand vectors. Hence, we ask the following question: Given a set of supply chains that share the same configuration but that differ from one another in the number of variants and demand vectors, does the complexity criterion rank these supply chains in the same order as the cost criterion would? First, we show that, under certain conditions, complexity and cost are equivalent. We then conduct an extensive numerical study to further investigate the agreement between the cost and complexity.

4.1 Conditions for Equivalence between Complexity and Cost

To help compare supply chains, we define some additional notation. Suppose we have m supply chains to compare. Let V_i^k be the number of variants offered by node $i = 1, \ldots, n$ in supply chain $k = 1, \ldots, m$. Let $\boldsymbol{q}_i^k$ denote the demand vector at node i of supply chain k. In addition, let I^k, H_N^k and H_A^k denote, respectively, the cost, node-based complexity and arc-based complexity of supply chain k. The following proposition states one condition under which the cost and complexity are equivalent.

Proposition 1. *Consider a set of m supply chains that have the same configuration but differ from one another in the number of variants and the demand vectors of the nodes. Suppose all m supply chains have the property that the demand shares are equal across all variants of the final assembler, i.e., $\boldsymbol{q}_{nj}^k = \frac{1}{V_n^k}$ for $j = 1, \ldots, V_n^k$. The ordering of these supply chains according to cost is the same as the ordering of them according to node-based complexity and arc-based complexity, i.e., $I^k > I^l$ iff $H_N^k > H_N^l$ and $I^k > I^l$ iff $H_A^k > H_A^l$.*

Proposition 1 implies that, under the condition that demand shares at the final assembler are evenly distributed, both the cost criterion and the complexity criterion will rank a given set of supply chains in the same order. In what follows, we show such equivalence holds under a less restrictive condition as well. Consider the case where the demand shares of all the variants offered by the final assembler are the same except for one *dominant* variant whose demand share is larger than all the others. In this case, the equivalence between the cost and the complexity continues to hold.

Proposition 2. *Consider a set of m supply chains that use the same configuration and offer the same set of variants, but differ from one another in the demand vectors of their nodes. Suppose all supply chains have the property that the demand shares of all variants of the final assembler are equal except for one dominant variant (indexed to be variant 1), whose demand share is larger than the demand shares of all the other variants, i.e., $\boldsymbol{q}_{n1}^k \geq \boldsymbol{q}_{n2}^k = \ldots = \boldsymbol{q}_{n,V_n^k}^k$ for $k = 1, \ldots, m$. The ordering of these supply chains according to cost is the same as the ordering of them according to node-based complexity and arc-based complexity, i.e., $I^k > I^l$ iff $H_N^k > H_N^l$ and $I^k > I^l$ iff $H_A^k > H_A^l$.*

To obtain another generalization on when the cost and the complexity are equivalent, we turn to majorization theory, which is useful in comparing how disordered two vectors are. In preparation for our next result, we first provide a formal definition of majorization. Let $\boldsymbol{x} = (x_1, x_2, ..., x_d)$ and $\boldsymbol{y} = (y_1, y_2, ..., y_d)$ be two d dimensional real vectors. Let $(x_{(1)}, x_{(2)}, ..., x_{(d)})$ indicate the vector obtained by sorting the entries of vector $\boldsymbol{x}$ in decreasing order. We say $\boldsymbol{y}$ majorizes $\boldsymbol{x}$, denoted as $\boldsymbol{y} \succ \boldsymbol{x}$, if the following two conditions are satisfied:

$$\sum_{i=1}^{k} x_{(i)} \leq \sum_{i=1}^{k} y_{(i)}, k = 1, 2, ..., d-1, \text{ and } \sum_{i=1}^{d} x_{(i)} = \sum_{i=1}^{d} y_{(i)}. \tag{9}$$

In our setting, majorization has a meaningful interpretation: If one demand vector majorizes another, the majorizing vector represents a more predictable demand pattern than the majorized vector. This interpretation has been used in the literature when analyzing the effect of variety on inventory costs; for example, van Ryzin and Mahajan (1999) interpret that a majorized demand vector represents a more fashionable product's demand, where consumer choice is less predictable and, therefore, more evenly scattered across variants. In our context, majorization provides a meaningful way to compare supply chains through their demand vectors. The next proposition utilizes the majorization theory to describe another scenario where the cost and complexity criteria are equivalent.

Proposition 3. *Consider a set of m supply chains that have the same configuration but differ from one another in the demand vectors of the final assemblers. Suppose the supply chains have the property that, for any given node in the uppermost echelon, the demand vector of supply chain 1 majorizes that of supply chain 2 which majorizes that of supply chain 3 and so on, i.e., $\boldsymbol{q}_i^1 \succ \boldsymbol{q}_i^2 \succ \ldots \succ \boldsymbol{q}_i^m$ for all nodes i in the uppermost echelon. The ordering of these supply chains according to cost is the same as the ordering of them according to node-based complexity and arc-based complexity, i.e., $I^k > I^l$ iff $H_N^k > H_N^l$ and $I^k > I^l$ iff $H_A^k > H_A^l$.*

One interpretation of Proposition 3 is the following: If a set of supply chains can be ordered according to the demand variability faced by the most-upstream suppliers, then the cost and complexity criteria will rank these supply chains in the same order. The underpinnings of this result are two intuitive observations: First, loosely speaking, the higher the demand variability faced by a supplier, the higher the cost incurred by the supplier and the higher the contribution of this supplier to the complexity of the supply chain. In that sense, higher cost and higher complexity go hand in hand as demand variability increases. Second, as we show in the proof of Proposition 3, if the uppermost echelons of supply chains can be ordered according to their demand variability, the same ordering cascades down the supply chain to lower echelons, which essentially means that the supply chains themselves can be ordered according to the demand variability they face. As a consequence, supply chains with higher demand variability end up being simultaneously more costly and more complex.

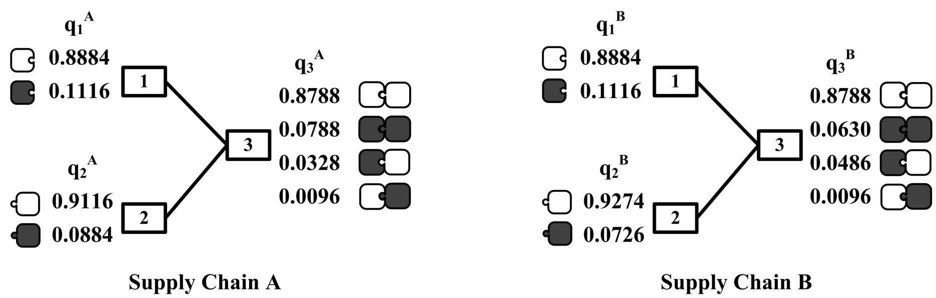

Figure 3: An example where complexity and cost are inconsistent. Here $\boldsymbol{q}_3^A \succ \boldsymbol{q}_3^B$, supply chain B has higher cost but with lower complexity. Both supply chain A and B have the same cost coefficients $C_1 = C_2 = 1$ and $C_3 = 8$

Given Proposition 3, one may wonder if a similar result holds when the demand vectors of the final assemblers are ordered according to the majorization criterion. In other words, is the following claim true?

Claim: Assume supply chains A and B have the same configuration. If the final assembler's demand vector in supply chain A majorizes the final assembler's demand vector in supply chain B, then the supply chain with the higher cost also has higher complexity.

Unfortunately, the above claim is not true. The condition that the demand vector of one final assembler majorizes that of another is not strong enough, because such an ordering does not propagate up the supply chain to upper echelons, making it impossible to say that one supply chain faces more demand variability than the other. Figure 3 shows a counterexample to the above claim. In this example, the demand vector of the final assembler in supply chain A majorizes the demand vector of the final assembler in supply chain B. Here, supply chain A has lower cost, but higher complexity.

The counterexample shows that there are instances where the supply chain with higher complexity could have lower cost, causing an inconsistency between how the cost and the complexity criteria rank supply chains. This is not surprising given that the cost and the complexity are two entirely different functions. Nonetheless, the results of this section show that the cost and complexity share a similar enough structure that they will rank a given set of supply chains consistently under certain conditions. In the next subsection, we conduct a numerical study to further investigate the the degree of agreement between cost and complexity.

4.2 Numerical Analysis of the Consistency between Complexity and Cost

We first investigate how five different factors affect the consistency between the complexity and the cost of a two-echelon supply chain. These five factors are: complexity definition, number of suppliers to the final assembler, underage and overage cost difference between two consecutive echelons, number of variants offered by suppliers in the uppermost echelon, and the evenness of the final assembler's demand vector. We first present our results for a two-echelon supply chain and subsequently analyze the effect of the number of echelons.

4.2.1 Design of Numerical Study

Consider two supply chains that differ from one another only in their demand vectors. (One can generate many such supply chain pairs by randomly generating a demand vector for each of the two final assemblers, which then determines the demand vectors of all the other nodes.) For a given pair of supply chains, if the supply chain with higher cost has higher complexity as well, we say cost and complexity are consistent. Otherwise, we say cost and complexity are inconsistent. By checking many pairs of supply chains, one can observe how common inconsistencies are and how the likelihood of an inconsistency depends on the five factors identified above. To perform a systematic analysis, we conduct a numerical experiment using a 2^5 full factorial design (where each of the five factors can take two values and all 32 combinations of factor values are considered).[1] Table 1 shows the levels allowed for each factor in our experiment, which we discuss next:

First, the complexity definition can be one of two types: Node-based or arc-based complexity.

Second, the number of suppliers to the final assembler is either two or three.

Third, we assume that all nodes in the uppermost echelon provide the same number of variants, which can be either two or three.

Fourth, we let the underage and overage cost difference between the two echelons to be high or low. We refer to this factor as cost disparity hereafter. To obtain high and low levels for cost disparity, we start by assuming that the final assembler's cost coefficient is at least as big as the sum of the suppliers' cost coefficients. Such an assumption is justified since the cost coefficient is based on underage and overage costs, and the final assembler puts together components from suppliers and adds further value to the product, thus inflating the overage and underage costs beyond those of the components from the suppliers. We say the cost disparity is low if the final assembler's cost coefficient is the sum of the suppliers' cost coefficients, and high if the final assembler's cost coefficient is twice the sum of the suppliers' cost coefficients.

The fifth and last factor in our experiment is the evenness of the final assembler's demand vector. To obtain two different degrees of evenness, we use the following procedure to generate the demand vectors. Given that the final assembler offers V_n variants, we draw V_n random numbers, denoted as R_v, $v = 1, 2, ..., V_n$, from the uniform distribution, $U(a-Kb, a+Kb)$ (where $a > Kb > 0$

[1] A *full two-level factorial design* is an experiment in which several factors, each of which can take two values, vary independently, and the experiment is conducted for all possible combinations of the levels across all such factors. It is used to study the effect of each factor on the response variable, as well as the effects of interactions between factors on the response variable (Wu and Hamada, 2000).

so that all the random numbers are positive). We then obtain the final assembler's demand vector $\boldsymbol{q}_n$ from random numbers R_v, $v = 1, 2, ..., V_n$ by letting $\boldsymbol{q}_{nv} := \frac{R_v}{\sum_{k=1}^{V_n} R_v}$. Given this procedure, observe that the choice of K influences how even the resulting demand vector will be. We let $K = 1$ to obtain a more even demand vector and $K = 8$ to obtain a less even demand vector.

	Levels	
Factor	-	+
1. Complexity definition	Arc-based Complexity	Node-based Complexity
2. Number of suppliers	2	3
3. Number of variants	2	3
4. Cost disparity between two echelons	$C_i = \sum_{j \in S_i} C_j$	$C_i = 2 \sum_{j \in S_i} C_j$
5. Evenness of the demand vector	K=8	K=1

Table 1: Design of Experiments. We use a two-level, full factorial experiment.

Under each combination of the five factors, $10,000$ pairs of supply chains are generated where each pair consists of two supply chains that differ from one another in terms of their demand vectors. For each supply chain pair, we determine whether cost and complexity are consistent. We then determine the inconsistency percentage among these 10,000 pairs. This entire experiment (10,000 pairs for each of 32 combinations) is then replicated three times.

In this numerical study we fix the leadtime parameters so that the leadtime to replenish the inventory of a variant at node i is L_i periods, where L_i is the number of suppliers to node i (equivalently, the number of inputs that are assembled by node i). Recall that the leadtime covers the span of activities starting with the purchase of the inputs and ending with the assembly of the variant. Therefore, one would expect that the larger the number of inputs that go into a variant, the longer the time it takes to replenish the variant's inventory. This effect is what we wish to capture in a simple way through our assumption that the replenishment leadtime for a variant is equal to the number of inputs that go into the variant.

4.2.2 Results of the Numerical Study

Table S-1 in Appendix C of the online supplement shows the inconsistency percentage, also called inconsistency rate, for each replication of each of the 32 combinations. Using this data, we determine the statistical significance of each of the five factors (at a 99% confidence level), reported in Table 2. The statistical analysis is based on ANOVA (analysis of variance). For a given factor, the *effect* column in Table 2 is the change in the average inconsistency rate when the factor's value changes from '-' to '+'. For example, the average inconsistency rate increases by 0.00635% when the number of suppliers increases from two to three. We next discuss the results for each of the five factors listed earlier.

First, the complexity definition has a statistically significant effect on the inconsistency rate.

Term	Effect	t-Statistics	p Value
Constant		130.94	< 0.0001
Complexity definition	0.00659	18.89	< 0.0001
Number of suppliers	0.00635	18.20	< 0.0001
Number of variants	0.00076	2.18	0.033
Cost disparity between two echelons	0.00252	7.22	< 0.0001
Uniformity of the demand vector	-0.02700	-77.35	< 0.0001

Table 2: Estimated Effects, t-Statistics and p-Value (%). A positive (negative) effect implies that when the factor's value changes from '-' to '+', the inconsistency rate increases (decreases).

Observe from Table 2 that the inconsistency rate is lower under arc-based complexity than under node-based complexity. In that sense, the arc-based complexity is a better performance measure of an assembly supply chain than node-based complexity. To see why this is the case, first recall our assumption that a node with a larger number suppliers faces a larger lead time, resulting in higher costs for those nodes. Hence, nodes with more suppliers contribute more to the supply chain cost. Now, notice that, under node-based complexity, given by (3), the contributions of all nodes to the supply chain complexity are weighted equally, whereas, under arc-based complexity, given by (5), node i's contribution to the supply chain complexity is weighted by the number of suppliers to that node, L_i. Because arc-based complexity gives more weight to nodes with larger number of suppliers, it comes closer to capturing the supply chain cost compared to node-based complexity.

Second, the number of suppliers is also statistically significant, and an increase in the number of suppliers results in an increase in the inconsistency rate. As the number of suppliers further increases, one would hope that the inconsistency rate does not grow indefinitely, but stabilizes instead. To resolve this question, we re-ran our experiment with two new values for the number of suppliers: four and five (as opposed to two and three suppliers in the original experiment). For this experiment where the number of suppliers can be either four or five, the inconsistency data and the results of the statistical analysis are summarized in Tables S-2 and S-3 in Appendix C of the online supplement. When the number of suppliers goes from four to five, the number of suppliers no longer has a statistically significant effect on the inconsistency rate. This is encouraging because it implies that there is a natural bound on how large the inconsistency rate can grow as the number of suppliers increases.

Third, the effect of the number of variants is not statistically significant.

Fourth, the cost disparity between two echelons is statistically significant and the higher the cost disparity, the higher the inconsistency rate. However, similar to the effect of the number of suppliers, the effect of the cost disparity also becomes statistically insignificant once the cost disparity becomes large enough. For an experiment where the cost coefficient of the final assembler

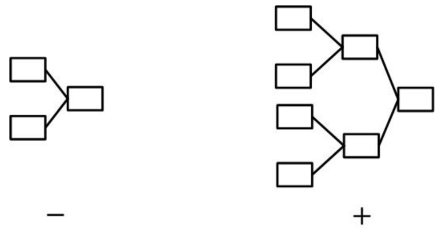

Figure 4: Two new levels of Factor 'B,' the number of echelons, which can be two (-) or three (+).

is either four or five times as large as the sum of the suppliers' cost coefficients (as opposed to the original experiment where the final assembler's cost coefficient was either equal to the sum of the suppliers' cost coefficients or twice as large), the inconsistency data and the results of the statistical analysis are summarized in Tables S-4 and S-5 in Appendix C of the online supplement.

Finally, the evenness of the final assembler's demand vector has a statistically significant effect on the inconsistency rate: the more even the demand vector, the smaller the inconsistency rate. To see the intuition behind this, recall from Proposition 1 that, if all variants of the final assembler have the same demand share, then cost and complexity are equivalent. The larger the evenness of the final assembler's demand vector, the closer one comes to the scenario of evenly-distributed demand, where cost and complexity are equivalent.

We have so far focused on a numerical experiment where supply chains have two echelons. To determine the effect of the number of echelons, we run a similar experiment with five factors. In this new experiment, we fix the number of suppliers for each node as two (thus, dropping the number of suppliers from the list of factors). Instead, we add the number of echelons as a new factor and we allow the number of echelons to be two or three, shown in Figure 4. Tables S-6 and S-7 in Appendix C of the online supplement show the inconsistency data of this experiment and the results of the significance test. Observe that the number of echelons increasing from two three has a statistically significant effect on the inconsistency rate. Thus, one should be more careful about applying the complexity measure to supply chains with larger number of echelons. However, it should be noted that the magnitude of increase in inconsistency rate is is so small that the number of echelons does not appear to be a major cause for concern: The inconsistency percentage increases by 0.198% when the number of echelons increases from two to three.

4.2.3 The Cost of Inconsistencies

The previous discussion shows that inconsistencies between cost and complexity occur rarely. Even though inconsistencies are rare, complexity may still be unreliable if it favors a supply chain that is much more costly than another. We next analyze all the cases where cost and complexity were inconsistent (among three replications of 10,000 examples for each of 32 different factor value

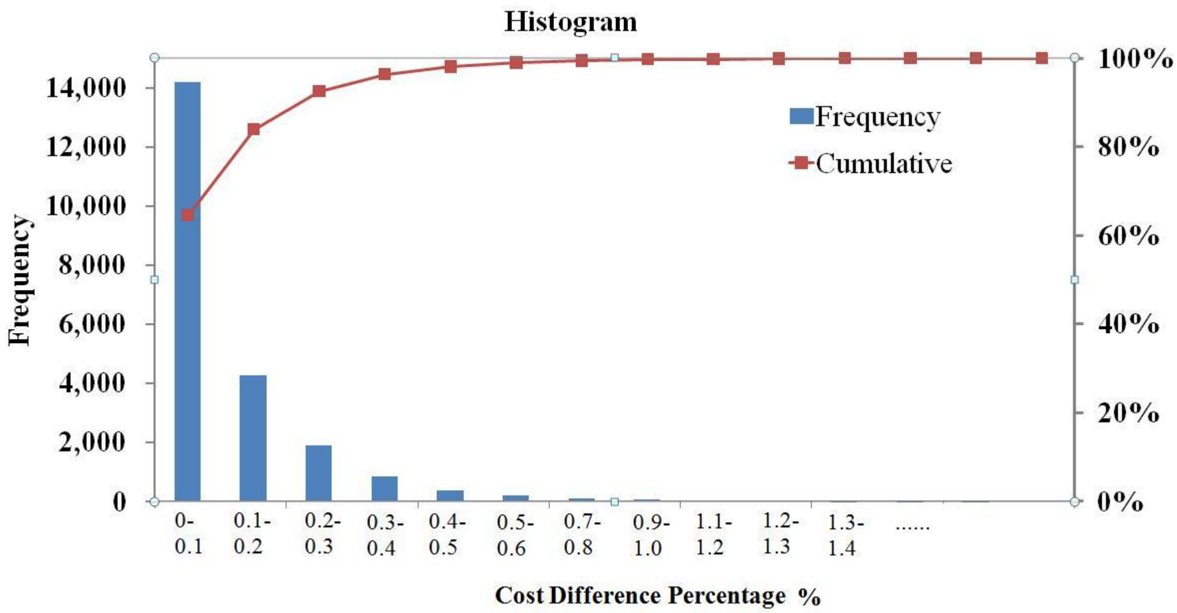

Figure 5: Histogram of cost difference. Considering all the problem instances where cost and complexity are inconsistent, more than 80% of those instances occur when the cost difference is less than 0.2%.

combinations, resulting in a total of 960,000 examples). For each example where cost and complexity were inconsistent, we check the cost difference between the two supply chains. The mean, median, minimum, maximum and standard deviation of these cost differences (expressed as a percentage of the less costly supply chain's cost) are shown in Table 3. Notice that the largest cost difference ever encountered is less than 1.21%. This provides good support for complexity as a measure of the supply chain performance, because what we see here is that when cost and complexity are inconsistent, the costs of the two supply chains are virtually the same. In addition, Figure 5 shows a frequency diagram for the cost differences (again, expressed as a percentage). Notice that the frequency diagram exhibits a long tail ending at the maximum cost difference of 1.21% and more than half of the inconsistencies occur when cost difference is less than 0.1%.

median	mean	standard deviation	max	min
0.0583	0.1012	0.1286	1.2089	4.68×10^{-8}

Table 3: Cost difference statistics (%).

4.2.4 A Modified Numerical Study

It is remarkable that cost and complexity disagree so rarely (2.29% of all examples), and when they do disagree, the cost difference is tiny as discussed above. One may wonder if such strong consistency is an artefact of our numerical set-up. In particular, in our experiments, when randomly generating the demand vector for a supply chain, we draw random numbers from a uniform distribution

$U(a - Kb, a + Kb)$. Because all the random numbers are coming from the same distribution, our numerical set-up might be introducing a bias toward more even demand vectors, which would in turn reduce the extent of inconsistencies. To remove this possible bias, we modify our numerical experiment by introducing deliberately large differences between the random numbers that are used to generate the demand vectors. To do so, when generating a set of V_n random numbers, denoted as R_v, $v = 1, 2, \cdots, V_n$, we draw R_v from a uniform distribution $U(a_v - Kb, a_v + Kb)$, where $a_v > Kb > 0$ and $a_v = e^v$, $v = 1, 2, \cdots, V_n$. Notice that the expected value of R_v is now a_v, which grows exponentially in v. This numerical set-up ensures that there will be big differences between R_v values, which will then lead to uneven demand vectors. Given this method of generating random demand vectors, we run another experiment with a 2^4 full factorial design, where the four factors are: complexity definition, number of suppliers, number of variants, cost disparity. The values of these factors are the same as those described in Table 2. We find that the inconsistency rate now increases to 7.05% of all examples. Once again, we focus on all the comparisons where the cost and complexity were inconsistent, and determine the cost difference between the two supply chains in each and every one of those cases. See Table 4 for the summary statistics of these cost differences. As expected, this new study, which deliberately creates uneven demand vectors, leads to larger cost differences. For example, the maximum cost difference in the event of inconsistency is now 10.97% and the mean cost difference is 1.47%. However, as the frequency diagram shows, the cost differences still show a long-tailed pattern and 60% of the observations show less than a 2% cost difference. In conclusion, even a deliberate attempt to induce uneven demand vectors does not result in a significant discrepancy between cost and complexity.

median	mean	standard deviation	max	min
1.113	1.469	1.314	10.966	8.07×10^{-5}

Table 4: Cost Difference Statistics (%)

5. Complexity versus Cost When Choosing a Configuration

In Section 4 we studied the relationship between cost and complexity by comparing supply chains that have the same configuration but different demand vectors. We now employ complexity to compare two supply chains that have the same demand vector but different supply chain configurations. In particular, we will compare a prototypical modular supply chain with a prototypical non-modular supply chain. See Figure 7 for a depiction of the modular and non-modular supply chain configurations. The non-modular supply chain is a two-echelon supply chain where four suppliers serve the final assembler, who then puts together the parts produced by the suppliers. In comparison, in the modular assembly supply chain, there is a mid-echelon with two suppliers, each

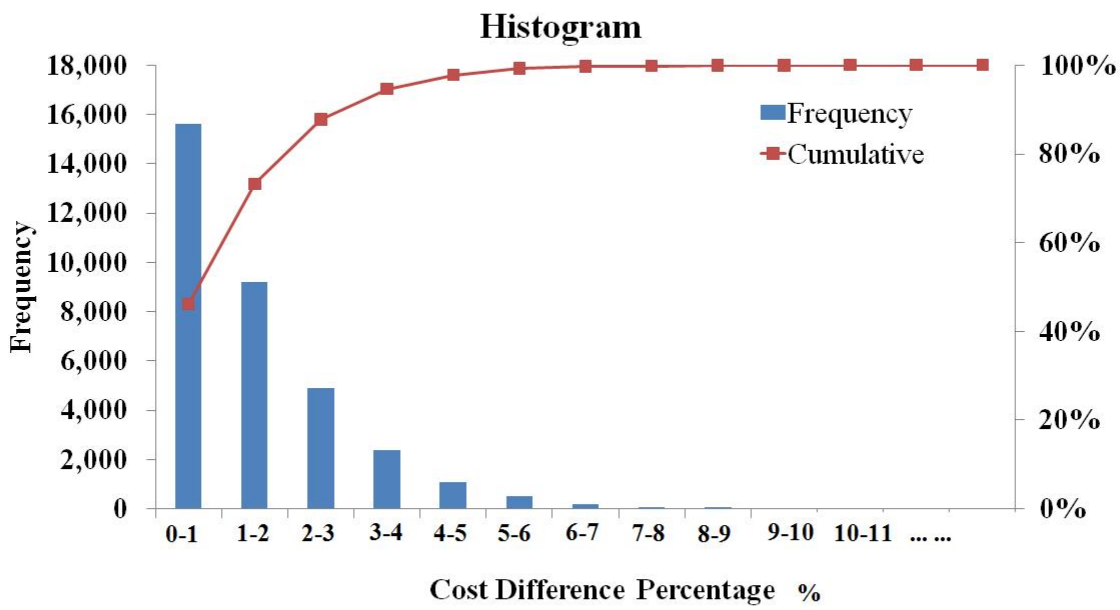

Figure 6: Histogram of cost differences in the modified numerical study. Considering all the problem instances where cost and complexity are inconsistent, more than 80% of those instances occur when the cost difference is less than 3%.

of which assembles parts produced by two suppliers in the upper echelon. The final assembler then receives modules from the mid-echelon suppliers and assembles these modules.

We assume that all the nodes in the uppermost echelon of both the modular and non-modular supply chains provide the same number of variants, denoted by γ. Consequently, the final assemblers in both supply chains offer the same number of variants, γ^4, and we assume that their demand vectors are the same, since our goal is to compare the configurations only.

As for the cost model, recall that the cost coefficient of node i, which depends only on the unit underage and overage costs and the service level, has been defined as

$$C_i := (c_{io} + c_{iu})\alpha z(\alpha) - c_{iu} z(\alpha) + (c_{io} + c_{iu})\phi_N(z(\alpha)).$$

Here, we slightly modify this notation. Let C_i^N denote the cost coefficient of node i in the non-modular supply chain and C_i^M the cost coefficient of node i in the modular supply chain. (See Figure 7 for the labeling of the nodes.) For ease of exposition, we assume that all nodes in the same echelon have the same cost coefficient, i.e., $C_1^N = C_2^N = C_3^N = C_4^N$, $C_1^M = C_2^M = C_3^M = C_4^M$ and $C_5^M = C_6^M$. Furthermore, to make a fair comparison between the two supply chains, we assume that the final assembler of both supply chains have the same cost coefficient, i.e., $C_7^M = C_7^N$, as do the suppliers in the most upstream echelon, i.e., $C_1^N = \ldots = C_4^N = C_1^M = \ldots = C_4^M$.

In this section we assume that the leadtime to replenish the inventory of a variant at node i is L_i periods, that is, the number of suppliers to node i or, equivalently, the number of inputs

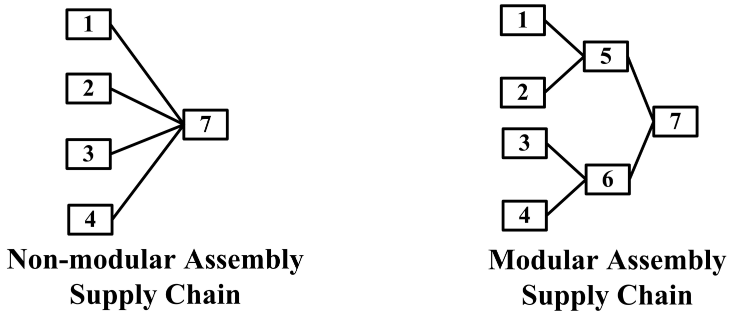

Figure 7: Prototypical configurations for non-modular vs. modular assembly supply chains

that are assembled by node i. This assumption is meant to reflect the fact that the larger the number of inputs that go into a variant, the longer the time it takes to assemble the variant. As a consequence of this assumption, the replenishment lead time of the final assembler is four periods in the non-modular assembly supply chain and two periods in the modular assembly supply chain. The leadtime reduction in the modular supply chain helps the final assembler reduce its inventory costs compared to the non-modular supply chain. On the flipside, however, there are two mid-echelon suppliers in the modular assembly supply chain, which do not exist in the non-modular assembly supply chain. These new suppliers inflate the total inventory cost of the modular assembly supply chain compared to the non-modular one. Hence, when using cost as the criterion, the trade-off in moving from a non-modular configuration to a modular one is the cost reduction achieved by the final assembler in the modular supply chain versus the additional costs created by two new suppliers.

Because the complexity criterion does not explicitly recognize the cost parameters, it is possible that the complexity criterion will lead to markedly different choices between two configurations. Nonetheless, the complexity criterion may be promising, because it leads to a similar trade-off as the cost criterion:[2] The complexity of each node is weighted by the number of links to that node, meaning that the final assembler's contribution to complexity is lower in the modular supply chain, but the modular supply chain's complexity is inflated by the addition of two new suppliers in the mid-echelon.

Given that the complexity criterion follows a similar trade-off as the cost criterion, but does not take into account any of the cost information, it is not clear whether the cost and complexity criteria will yield similar results. The next proposition sheds some light on this question.

Proposition 4. *Consider the modular and non-modular assembly supply chains shown in Figure 7 and assume that the final assemblers of the two chains have the same demand vector and the*

[2]We observed in the last section that arc-based complexity performs better than node-based complexity. Hence, in this section the complexity measure we use is the arc-based complexity.

demand shares are equal across all the variants of the final assembler. Then:

(*a*) ***According to complexity criterion:*** *If the number of variants offered by a node in the most upstream echelon,* γ*, is two or more, then the modular assembly supply chain is better. If* $\gamma = 1$*, then the non-modular assembly supply chain is better.*

(*b*) ***According to cost criterion:*** *There exists a threshold* t *such that if the number of variants offered by a node in the most upstream echelon,* γ*, is greater than or equal to* t*, then the modular assembly supply chain is better. If* $\gamma < t$*, then the non-modular assembly supply chain is better.*

The proposition shows that cost and complexity may lead to different choices between modular and non-modular configurations: The threshold t, which applies when the cost criterion is used, may be different from one, in which case the cost and complexity criteria may disagree. The upside of this proposition, however, is that the choice between non-modular and modular supply chains exhibits the same trend with respect to the number of variants, γ, regardless of whether cost or complexity is used: a larger number of variants favors the modular assembly supply chain.

Proposition 4 uses the assumption that the demand shares of all variants produced by the final assembler are equal. We next relax this assumption and analyze how the choice between two configurations depends on the demand share of a given variant under both the cost and complexity criteria. In particular, consider one of the variants produced by a node in the most upstream echelon, say, variant V_1 of node 1. Let us write the demand vector of node 1 as $(a_1(1-p), a_2(1-p), ..., a_{n-1}(1-p), p)$, where p is the demand share of variant V_1 at node 1 and the remaining demand at this node is shared arbitrarily by the other variants produced by the node. Proposition 5 describes how the choice between modular and non-modular configurations depends on p.

Proposition 5. *Consider the modular and non-modular assembly supply chains shown in Figure 7 and assume that the final assemblers of the two chains have the same demand vector. Let* $(a_1(1-p), a_2(1-p), ..., a_{n-1}(1-p), p)$ *be the demand vector of node 1. Then:*

(*a*) ***According to complexity criterion:*** *One of the following is true:*

(*i*) *The non-modular supply chain is better for all* $p \in [0, 1]$*, or*

(*ii*) *The modular supply chain is better for all* $p \in [0, 1]$*, or*

(*iii*) *There exist* p_1 *and* p_2 *such that* $0 \leq p_1 < p_2 \leq 1$ *and the non-modular supply chain is better for* $p \in [0, p_1)$*, the modular supply chain is better for* $p \in [p_1, p_2)$ *and the non-modular supply chain is better for* $p \in [p_2, 1]$*.*

(*b*) ***According to cost criterion:*** *One of the following is true:*

(*i*) *The non-modular supply chain is better for all* $p \in [0, 1]$*, or*

(*ii*) *The modular supply chain is better for all* $p \in [0, 1]$*, or* (*iii*) *There exist* $\widehat{p}_1$ *and* $\widehat{p}_2$ *such that* $0 \leq \widehat{p}_1 < \widehat{p}_2 \leq 1$ *and the non-modular supply chain is better for* $p \in [0, \widehat{p}_1)$*, the modular supply chain is better for* $p \in [\widehat{p}_1, \widehat{p}_2)$ *and the non-modular supply chain is better for* $p \in [\widehat{p}_2, 1]$*.*

The first observation from the proposition is that cost and complexity may lead to different decisions depending on how similar or dissimilar p_1 and p_2 are to $\widehat{p}_1$ and $\widehat{p}_2$. Nonetheless, the proposition shows that the structural behavior of the modular versus non-modular configuration choice is the same under both cost and complexity. The intuition behind the result is as follows. If p is close to zero, the implication is that there is very little demand for the final assembler's products that use this variant. Hence, it is almost as if those products do not exist, which is effectively equivalent to reducing the variety offered by the supply chain. Hence, the non-modular configuration is preferred. Likewise, when p is close to one, the demand shares of all other variants at supplier 1 decreases, thus resulting in a number of products with practically no demand at the final assembler. This again results in non-modular supply chain being preferred. However, when p is moderate, the variant does not depress the variety offered by the supply chain in any way, which motivates the use of modular configuration.

The previous propositions show that while cost and complexity may show similar behavior, they may lead to different configuration choices because the decisions depend on thresholds that may differ between cost and complexity. In particular, the thresholds for cost criterion depend on the cost coefficients of the nodes whereas complexity does not pay any attention to the cost coefficients. This negligence on complexity criterion's part may cause significant inconsistencies when it comes to configuration choice. To further understand the effect of cost coefficients on the consistency between cost and complexity, we ask the following question: What can we say about the effect of the final assembler's cost coefficient on the consistency between cost and complexity? The following proposition provides a partial answer to this question.

Proposition 6. *Consider the modular and non-modular assembly supply chains shown in Figure 7 and, for a given set of parameters, suppose that both cost and complexity criteria favor the non-modular supply chain. Then:*

(*a*) *If the final assembler's cost coefficient decreases, then cost and complexity continue to be consistent and both favor the non-modular supply chain.*

(*b*) *If the final assembler's cost coefficient increases beyond a certain threshold, then cost begins to favor the modular supply chain and complexity continues to favor the non-modular supply chain, thereby creating an inconsistency between the two criteria.*

One implication of the proposition is that whether cost and complexity agree depends very much on what the cost coefficients are. In our numerical examples, we have observed that the consistency is very sensitive to the cost coefficients. In fact, the consistency is so sensitive to the cost coefficient that cost and complexity may almost always agree for one set of cost coefficients and may almost never agree for another set of cost coefficients, regardless of where the other parameters are set such as number of variants and the demand vectors. Hence, one needs to be cautious when using complexity to make configuration choices.

6. Conclusion

In this paper we proposed a complexity measure for assembly supply chains, based on the concept of information entropy. This complexity measure takes into account factors such as the supply chain configuration, the level of variety offered by each node of the supply chain, and the demand split across all the variants offered by a node. We investigated the relationship between the complexity and the cost of an assembly supply chain. In particular, we showed that, when comparing assembly supply chains with the same configuration but different levels of product variety, the cost and the complexity are equivalent under certain conditions. Even when these conditions do not hold, our numerical study demonstrated that the cost and complexity criteria rank supply chains consistently in an overwhelming majority of cases. The agreement between the cost and complexity criteria was shown to be lesser when comparing assembly supply chains with the same level of product variety, but different configurations. Overall, we found that the complexity measure is a good proxy for the cost of an assembly supply chain when evaluating alternative levels of product variety that will be delivered by a given supply chain configuration, but not so when evaluating alternative supply chain configurations to deliver a given level of product variety.

In Section 4, we observed that complexity is a good proxy for cost when making variety decisions, but our analysis in that section focused on a cost model where only inventory costs were accounted for. It is not difficult to extend the results to the case where transportation costs are also accounted for. In fact, as long as one is willing to assume that the transportation cost along an arc of the supply chain depends only on the total volume that travels along that arc, but not on the volumes of specific variants, all the results of Section 4 continue to hold. In order to extend the results to the case where manufacturing costs are also included in the supply chain cost, it would be sufficient to assume that the expected per-period manufacturing cost for variant v of node i is increasing and concave in the demand share of the variant, q_{iv}. This assumption is not unreasonable: Under this assumption, given two nodes that differ only in terms of their demand vectors, the node that produces many medium-volume products will incur larger manufacturing costs compared to a node that produces a few high-volume products coupled with low-volume products. This outcome is reasonable because it captures the economies of scale that a node can enjoy by offering high-volume products.

Our analysis in Section 5 showed that, when using complexity to choose between modular and non-modular configurations, the consistency between cost and complexity is very sensitive to the cost disparity between two echelons. This happens, because the complexity criterion uses no cost information at all. One could of course improve the consistency between cost and complexity by trying to reflect in the complexity definition the cost disparities that exist among the nodes of the supply chain. For example, one may want to weigh more heavily the complexity contribution of nodes whose unit holding and shortage costs are higher. While this may be an attractive modification, it may also beat the purpose of using the complexity criterion in the first place, because one of the important advantages of the complexity criterion is to absolve the decision maker of the need to rely on cost data.

Acknowledgement

The authors gratefully acknowledge the financial support from the Engineering Research Center for Reconfigurable Manufacturing Systems of the National Science Foundation under Award Number EEC-9529125 and the National Science Foundation under Award Number CMMI-0825438, both at The University of Michigan.

References

Aydin, G., W. H. Hausman. 2009. The role of slotting fees in the coordination of assortment decisions. Forthcoming in Production and Operations Management.

Bernstein, F., G. A. DeCroix. 2004. Decentralized pricing and capacity decisions in a multi-tier system with modular assembly. *Management Science* **50**(9) 1293–1308.

Bernstein, F., G. A. DeCroix. 2006. Inventory policies in a decentralized assembly system. *Operations Research* **54**(2) 324–336.

Bollapragada, R., U. S. Rao, J. Zhang. 2004. Managing inventory and supply performance in assembly systems with random supply capacity and demand. *Management Science* **50**(12) 1729–1743.

Cover, Thomas M., Joy A. Thomas. 1991. *Elements of Information Theory*. 2nd ed. John Wiley & Sons.

DeCroix, G. A., P. H. Zipkin. 2005. Inventory management for an assembly system with product or component returns. *Management Science* **51**(8) 1250–1265.

Deshmukh, A. V., J. J. Talavage, M. M. Barash. 1998. Complexity in manufacturing systems, part 1: Analysis of static complexity. *IIE Transactions* **30**(7) 645–655.

Fisher, M. L., C. D. Ittner. 1999. The impact of product variety on automobile assembly operations: Empirical evidence and simulation. *Management Science* **45**(6) 771–786.

Fredriksson, P. 2006. Mechanisms and rationales for the coordination of a modular assembly system: the case of Volvo cars. *International Journal of Operations & Production Management* **26**(4) 350–370.

Gerchak, Y., Y. Wang. 2004. Revenue-sharing vs. wholesale-price contracts in assembly systems with random demand. *Production and Operations Management* **13**(1) 23–33.

Gurnani, H., R. Akella, J. Lehoczky. 1996. Optimal order policies in assembly systems with random demand and random supplier delivery. *IIE Transactions* **28**(11) 865–879.

Kurtulus, M., B Toktay. 2005. Category captainship: Outsourcing retail category management. Working paper, INSEAD.

Lee, H., C. S. Tang. 1997. Modeling the costs and benefits of delayed product differentiation. *Management Science* **43**(1) 40–53.

MacDuffie, J. P., K. Sethuraman, M. L. Fisher. 1996. Product variety and manufacturing performance: Evidence from the international automotive assembly plant study. *Management Science* **42**(3) 350–369.

Marshall, A. W., I. Olkin. 1979. *Inequalities: theory of majorization and its applications*. Academic Press: New York.

Randall, T., K. Ulrich. 2001. Product variety, supply chain structure, and firm performance: analysis of the U. S. bicycle industry. *Management Science* **47**(12) 1588–1604.

Rosling, K. 1989. Optimal inventory policies for assembly systems under random demand. *Operations Research* **37**(4) 565–579.

Salvador, F., M. Rungtusanatham, C. Forza. 2004. Supply-chain configurations for mass customization. *Production Planning and Control* **15**(4) 381–397.

Schmidt, C. P., S. Nahmias. 1985. Optimal policy for a two-stage assembly system under random demand. *Operations Research* **33**(5) 1130–1145.

Shannon, C. E. 1948. A mathematical theory of communication. *Bell System Technical Journal* **27** 379–423.

Singh, P., H. Groenevelt, N. Rudi. 2005. Product variety and supply chain structures. Working paper, University of Rochester.

Song, J., P. Zipkin. 2003. *Supply chain operations: Assemble-to-order systems, In Handbooks in Operations Research and Management Science, Vol. 11: Supply Chain Management*. Eds. A.G. de Kok and S.C. Graves, Elsevier, Amsterdam, Netherlands.

Suh, N. P. 2005. *Complexity Theory and Applications*. Oxford University Press, New York.

Swaminathan, J. M. 2001. Enabling customization using standardized operations. *California Management Review* **43**(3).

Thomas, D. J., D. P. Warsing. 2007. A periodic inventory model for stocking modular components. *Production and Operations Management* **16**(3) 343–359.

van Ryzin, G., S Mahajan. 1999. On the relationship between inventory costs and variety benefits in retail assortments. *Management Science* **45**(11) 1496–1509.

Wu, C.F.J., M. Hamada. 2000. *Experiments: Planning, Analysis, and Parameter Design Optimization*. John Wiley, Hoboken, NJ.

Zhu, X., S. J. Hu, Y. Koren, S. P. Marin. 2008. Modeling of manufacturing complexity in mixed-model assembly lines. *Journal of Manufacturing Science and Engineering* **130**(5).

Online Supplement

A Complexity Model for Assembly Supply Chains... by Wang, Aydın and Hu

Appendix A: Derivation of Arc-based Complexity

<u>Random Experiment:</u> *For each node in the supply chain, we form a pool of variants produced by that node, where each variant is represented in a quantity proportional to its demand share. Consider the random experiment where we first pick an arc of the supply chain at random, and we then pick one variant from the pool of this arc's end-node.*

Let L_i denote the number of suppliers connected to node i in the supply chain, $i = 1, 2, \ldots, n$. Recall that we let $L_i = 1$ for any node i in the uppermost echelon, corresponding to our convention that there is a virtual supplier that is linked to the nodes in the uppermost echelon. As discussed earlier, this convention allows us to capture the flows into the supply chain. Let us denote the virtual supplier as node 0. Then the total number of arcs in the supply chain equals to $K = \sum_{i=1}^{n} L_i$. Let R_{ji} represent the arc starting from node $j = 0, 1, \ldots, n-1$ and ending at node $i = 1, \ldots, n$. Then the probability of picking a certain arc R_{ji} and then picking variant v from the pool of end node i is $\frac{q_{iv}}{K}$ (since we pick arc R_{ji} with probability $\frac{1}{K}$ and variant v of node i with probability q_{iv}). Substituting probabilities of all possible outcomes back into Shannon's information entropy equation (2), the information entropy of the random experiment yields our arc-based complexity:

$$H_A = -\sum_{i=1}^{n} \sum_{j \in S_i} \sum_{v=1}^{V_i} \frac{q_{iv}}{K} \log_2 \frac{q_{iv}}{K} \tag{S-1}$$

For any node $i = 1, \ldots, n$,

$$-\sum_{j \in S_i} \sum_{v=1}^{V_i} \frac{q_{iv}}{K} \log_2 \frac{q_{iv}}{K} = -L_i \sum_{v=1}^{V_i} \frac{q_{iv}}{K} \log_2 \frac{q_{iv}}{K}.$$

Therefore, equation (S-1) can be rewritten as

$$H_A = -\sum_{i=1}^{n} \sum_{v=1}^{V_i} L_i \frac{q_{iv}}{K} \log_2 \frac{q_{iv}}{K} \tag{S-2}$$

Appendix B: Proofs

Throughout the appendix, we make frequent use of the majorization theory. Hence, we start with a definition of majorization.

Definition 1. *(Marshall and Olkin, 1979) For any real vector* $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \Re^n$*, let* $x_{(1)} \geq x_{(2)} \geq \ldots \geq x_{(n)}$ *denote the components of* $\boldsymbol{x}$ *in non-increasing order.*

B1.1 *For* $\boldsymbol{x}, \boldsymbol{y} \in \Re^n$*, we say* $\boldsymbol{x}$ *majorizes* $\boldsymbol{y}$*, written as* $\boldsymbol{x} \succ \boldsymbol{y}$ *if*

$$\sum_{i=1}^{k} x_{(i)} \geq \sum_{i=1}^{k} y_{(i)}, k = 1, \ldots, n-1 \text{ and } \sum_{i=1}^{n} x_{(i)} = \sum_{i=1}^{n} y_{(i)},$$

B1.2 $\boldsymbol{x} \succ \boldsymbol{y}$ *only if* $\sum g(x_i) \leq \sum g(y_i)$ *for all continuous concave functions* $g : \Re \to \Re$.

Proof of Proposition 1

Given the demand shares are equal across all variants of the final assembler (indexed as node n according to our convention), the demand vector of supply chain k's final assembler is $\boldsymbol{q}_n^k = \left(\frac{1}{V_n^k}, \frac{1}{V_n^k}, ..., \frac{1}{V_n^k}\right)$, where V_n^k is the number of variants offered by supply chain k's final assembler and $\boldsymbol{q}_n^k$ is a vector of dimension V_n^k. It now follows from (1) that all other nodes in supply chain k also have evenly-distributed demand vectors, i.e., $\boldsymbol{q}_i^k = \left(\frac{1}{V_i^k}, \frac{1}{V_i^k}, ..., \frac{1}{V_i^k}\right)$, $i = 1, 2, \cdots, n$, where $\boldsymbol{q}_i^k$ is a vector of dimension V_i^k.

First, consider two supply chains, indexed as 1 and 2, and suppose that supply chain 2's final assembler offers more variants than supply chain 1's, that is, $V_n^2 > V_n^1$. Then, each and every node in supply chain 2 offers at least as many variants as its counterpart in supply chain 1, that is, $V_i^2 \geq V_i^1$, $i = 1, 2 \ldots, n$. To enable the use of majorization arguments when comparing the demand vectors of these two supply chains, we define, for $i = 1, \ldots, n$, vectors $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$, both of dimension V_i^2, as follows:

$$\boldsymbol{x}_{iv} = \boldsymbol{q}_{iv}^1 = \frac{1}{V_i^1} \text{ for } v = 1, \ldots, V_i^1 \text{ and } \boldsymbol{x}_{iv} = 0 \text{ for } v = V_i^1 + 1, \ldots, V_i^2,$$
$$\boldsymbol{y}_{iv} = \boldsymbol{q}_{iv}^2 = \frac{1}{V_i^2} \text{ for } v = 1, \ldots, V_i^2.$$

Using the definitions of vectors $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$ and the definition of node-based complexity, given by (3), we can write[3]

$$H_N^1 = -\sum_{i=1}^{n} \sum_{v=1}^{V_i^2} \frac{\boldsymbol{x}_{iv}}{n} \log_2 \frac{\boldsymbol{x}_{iv}}{n},$$
$$H_N^2 = -\sum_{i=1}^{n} \sum_{v=1}^{V_i^2} \frac{\boldsymbol{y}_{iv}}{n} \log_2 \frac{\boldsymbol{y}_{iv}}{n}.$$

Now, observe that $-x \log_2 x$ is concave in x. In addition, notice that the vector $\boldsymbol{x}_i$ majorizes the vector $\boldsymbol{y}_i$ for $i = 1, \ldots, n$. Thus, we can apply the majorization theorem to obtain $H_N^2 \geq H_N^1$.

[3]Notice that if $\boldsymbol{x}_{iv} = 0$, then $\frac{\boldsymbol{x}_{iv}}{n} \log_2 \frac{\boldsymbol{x}_{iv}}{n} = 0$ as well.

Using the definition of arc-based complexity, given by (5), we can write

$$H_A^1 = -\sum_{i=1}^{n} L_i \sum_{v=1}^{V_i^2} \frac{\boldsymbol{x}_{iv}}{K} \log_2 \frac{\boldsymbol{x}_{iv}}{K},$$

$$H_A^2 = -\sum_{i=1}^{n} L_i \sum_{v=1}^{V_i^2} \frac{\boldsymbol{y}_{iv}}{K} \log_2 \frac{\boldsymbol{y}_{iv}}{K}.$$

Applying the majorization theorem yields $H_A^2 \geq H_A^1$.

Similarly, using the expression for supply chain cost I, given by (8), we can write

$$I^1 = \sum_{i=1}^{n} C_i \sqrt{\lambda(l_i+1)} \sum_{v=1}^{V_i^2} \sqrt{\boldsymbol{x}_{iv}},$$

$$I^2 = \sum_{i=1}^{n} C_i \sqrt{\lambda(l_i+1)} \sum_{v=1}^{V_i^2} \sqrt{\boldsymbol{y}_{iv}}.$$

Observing that $\sqrt{x}$ is concave in x and applying the majorization theorem, we obtain $I^2 \geq I^1$.

In summary, given two supply chains whose demand vectors are evenly distributed, arc-based complexity, node-based complexity and cost rank these two supply chains in the same order. Because the three orderings are the same when comparing an arbitrary pair of supply chains, they will be the same when comparing an arbitrary number of supply chains.

Proof of Proposition 2

Since all supply chains are identical in terms of the number of variants offered by each node, we drop the superscript k from the number of variants offered by node i of supply chain, and we write V_i instead of V_i^k. Consider two supply chains, where the demand vectors of the final assemblers in supply chains 1 and 2 are, respectively, $\boldsymbol{q}_n^1 = (q_{n1}^1, q_{n2}^1, ..., q_{n,V_n}^1)$ and $\boldsymbol{q}_n^2 = (q_{n1}^2, q_{n2}^2, ..., q_{n,V_n}^2)$. As required by the proposition, suppose that: (i) $q_{n1}^1 \geq q_{n2}^1 = \ldots = q_{n,V_n}^1$ and (ii) $q_{n1}^2 \geq q_{n2}^2 = \ldots = q_{n,V_n}^2$. Furthermore, without loss of generality, index the supply chains 1 and 2 so that: (iii) $q_{n1}^1 \geq q_{n1}^2$. Next, we will show that node-based complexity, arc-based complexity and cost rank these two supply chains in the same order. To do so, we first prove that properties (i) through (iii) are satisfied for nodes $i = 1, \ldots, n-1$ as well.

For a given node $i = 1, \ldots, n-1$, to see why (i) and (ii) hold, observe that one could divide the variants offered by the final assembler into V_i disjoint subsets, each containing V_n/V_i variants of the final assembler, and the demand shares of the variants in each subset add up to the demand share of a variant at node i. As a result, at node i of supply chain k, there must be $V_i - 1$ variants that all have the same demand share and the remaining variant's demand share is $q_{n1}^k + \frac{(1-q_{n1}^k)}{V_n-1}\left[(V_n/V_i) - 1\right]$, which is larger than the others. Let us index this variant with the larger share as variant 1. Now, for a given node $i = 1, \ldots, n-1$, to see why (iii) holds, observe that

$$\begin{aligned} q_{i1}^1 - q_{i1}^2 &= q_{n1}^1 + [(V_n/V_i) - 1] \frac{1 - q_{n1}^1}{V_n - 1} - q_{n1}^2 - [(V_n/V_i) - 1] \frac{1 - q_{n1}^2}{V_n - 1} \\ &= \left(q_{n1}^1 - q_{n1}^2\right) \left(1 - \frac{(V_n/V_i) - 1}{V_n - 1}\right) \\ &\geq 0 \end{aligned}$$

Now that we have shown properties (i) though (iii) hold for any node i, we can conclude that $\boldsymbol{q}_i^1 \succ \boldsymbol{q}_i^2$ for node $i = 1, \ldots, n$. Applying the majorization theorem, we obtain $H_N^2 \geq H_N^1$, $H_A^2 \geq H_A^1$, and $I^2 \geq I^1$. Since the three orderings are the same for any arbitrary pair, it follows that the three orderings are the same for an arbitrary number of supply chains.

Proof of Proposition 3:

For the purposes of this proof, we define Q_i to be the set of suppliers in the uppermost echelon whose variants are used in the module produced by node i. For a vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_a$ with dimensions, respectively, $m_1, m_2, \ldots, m_a$, define the operation $\Omega(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_a)$ as the sorted component-wise multiplication of vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_a$, that is, the following vector with dimension $\Pi_{j=1}^a m_j$:

$$(\boldsymbol{x}_{11}\boldsymbol{x}_{21} \cdots \boldsymbol{x}_{a1}, \boldsymbol{x}_{11}\boldsymbol{x}_{21} \cdots \boldsymbol{x}_{a2}, \ldots, \boldsymbol{x}_{1,m_1}\boldsymbol{x}_{2,m_2} \cdots \boldsymbol{x}_{a,m_a}),$$

sorted in descending order. Using these definitions, notice that the demand vector of any given node i can be written as $\boldsymbol{q}_i = \Omega_{j \in Q_i} \boldsymbol{q}_j$. This definition will be useful in the proof that follows.

We first prove the result for two supply chains. Suppose that there are a suppliers in the most upstream echelon in both supply chains, denoted as nodes $1, \ldots, a$. Assume that, as required by the proposition, supply chains 1 and 2 have the same configuration and the demand vector of each node in the uppermost echelon of supply chain 1 majorizes its counter-part in supply chain 2, $\boldsymbol{q}_s^1 \succ \boldsymbol{q}_s^2$, $s = 1, \ldots, a$. In this proof, we write the node-based complexity as a function of the demand vectors of the nodes in the uppermost echelon, that is, we write $H_N(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_a)$ instead of $H_N(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n)$. (Note that this reduction of the argument list is possible because $\boldsymbol{q}_{a+1}$ through $\boldsymbol{q}_n$ can be recovered from $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_a$, using the Ω-operation defined above.) We will prove that, if $\boldsymbol{q}_s^1 \succ \boldsymbol{q}_s^2, s = 1, \ldots, a$, then $H_N(\boldsymbol{q}_1^1, \ldots, \boldsymbol{q}_a^1) \leq H_N(\boldsymbol{q}_1^2, \ldots, \boldsymbol{q}_a^2)$, and $H_A(\boldsymbol{q}_1^1, \ldots, \boldsymbol{q}_a^1) \leq H_A(\boldsymbol{q}_1^2, \ldots, \boldsymbol{q}_a^2)$ and $I(\boldsymbol{q}_1^1, \ldots, \boldsymbol{q}_a^1) \leq I(\boldsymbol{q}_1^2, \ldots, \boldsymbol{q}_a^2)$. The proof is conducted in two steps.

Step 1: Suppose for now that the demand vectors of nodes 2 through a are the same in both supply chains, but node 1's demand vector in supply chain 1 majorizes the demand vector of node

1 in supply chain 2. That is, $\boldsymbol{q}_1^1 \succ \boldsymbol{q}_1^2$, $\boldsymbol{q}_2^1 = \boldsymbol{q}_2^2, \ldots, \boldsymbol{q}_a^1 = \boldsymbol{q}_a^2$. For notational convenience, let $\boldsymbol{q}_j = \boldsymbol{q}_j^1 = \boldsymbol{q}_j^2, j = 2, \ldots, a$. We next prove that $\boldsymbol{q}_i^1 \succ \boldsymbol{q}_i^2$, $i = a+1, \ldots, n$.

If the variants provided by node i do not use any variant from node 1, i.e, $1 \notin Q_i$, then the demand vectors of node i in both supply chains 1 and 2 are the same, i.e., $\boldsymbol{q}_i^1 = \boldsymbol{q}_i^2$ for $i = a+1, \ldots, n$ such that $1 \notin Q_i$. Hence, $\boldsymbol{q}_i^1 \succ \boldsymbol{q}_i^2$ holds trivially for $i = a+1, \ldots, n$ such that $1 \notin Q_i$.

Consider now the case where one or more variants of node i use the variants from node 1, i.e., $1 \in Q_i$. We can write $\boldsymbol{q}_i^1 = \Omega\left(\boldsymbol{q}_1^1, \Omega_{j \in Q_i/\{1\}} \boldsymbol{q}_j\right)$ and $\boldsymbol{q}_i^2 = \Omega\left(\boldsymbol{q}_1^2, \Omega_{j \in Q_i/\{1\}} \boldsymbol{q}_j\right)$. Let $A = \boldsymbol{q}_1^1$, $B = \boldsymbol{q}_1^2$ and $C = \Omega_{j \in Q_i/\{1\}} \boldsymbol{q}_j$ and apply Lemma 1 to conclude that $\boldsymbol{q}_i^1 \succ \boldsymbol{q}_i^2$.

Thus, $\boldsymbol{q}_i^1 \succ \boldsymbol{q}_i^2$ for $i = 1, \ldots, n$. It now follows from the majorization theorem that $H_N(\boldsymbol{q}_1^2, \boldsymbol{q}_2 \ldots, \boldsymbol{q}_a) \geq H_N(\boldsymbol{q}_1^1, \boldsymbol{q}_2, \ldots \boldsymbol{q}_a)$, $H_A(\boldsymbol{q}_1^2, \boldsymbol{q}_2 \ldots, \boldsymbol{q}_a) \geq H_A(\boldsymbol{q}_1^1, \boldsymbol{q}_2, \ldots \boldsymbol{q}_a)$ and $I(\boldsymbol{q}_1^2, \boldsymbol{q}_2 \ldots, \boldsymbol{q}_a) \geq I(\boldsymbol{q}_1^1, \boldsymbol{q}_2, \ldots \boldsymbol{q}_a)$.

<u>Step 2:</u> It follows from Step 1 that, if $\boldsymbol{q}_i^1 \succ \boldsymbol{q}_i^2$ for $i = 1, \ldots, a$, then

$$\begin{aligned}
H_N(\boldsymbol{q}_1^2, \boldsymbol{q}_2^2 \ldots, \boldsymbol{q}_{a-1}^2, \boldsymbol{q}_a^2) &\geq H_N(\boldsymbol{q}_1^1, \boldsymbol{q}_2^2 \ldots, \boldsymbol{q}_{a-1}^2, \boldsymbol{q}_a^2) \\
&\geq H_N(\boldsymbol{q}_1^1, \boldsymbol{q}_2^1 \ldots, \boldsymbol{q}_{a-1}^2, \boldsymbol{q}_a^2) \\
&\geq \ldots \geq H_N(\boldsymbol{q}_1^1, \boldsymbol{q}_2^1 \ldots, \boldsymbol{q}_{a-1}^1, \boldsymbol{q}_a^2) \\
&\geq H_N(\boldsymbol{q}_1^1, \boldsymbol{q}_2^1 \ldots, \boldsymbol{q}_{a-1}^1, \boldsymbol{q}_a^1)
\end{aligned}$$

Hence, $H_N(\boldsymbol{q}_1^2, \ldots, \boldsymbol{q}_a^2) \geq H_N(\boldsymbol{q}_1^1, \ldots, \boldsymbol{q}_a^1)$. Using the same line of arguments, we obtain

$$H_A(\boldsymbol{q}_1^2, \ldots, \boldsymbol{q}_a^2) \geq H_A(\boldsymbol{q}_1^1, \ldots, \boldsymbol{q}_a^1) \text{ and } I(\boldsymbol{q}_1^2, \ldots, \boldsymbol{q}_a^2) \geq I(\boldsymbol{q}_1^1, \ldots, \boldsymbol{q}_a^1).$$

In Steps 1 and 2, we have proved that, given two supply chains that satisfy the conditions of the proposition, arc-based complexity, node-based complexity and cost rank these two supply chains in the same order. Because the three orderings are the same when comparing an arbitrary pair of supply chains, they will be the same when comparing an arbitrary number of supply chains.

Proof of Proposition 4:

For the purposes of this proof, let H^{A-N} and H^{A-M} denote the arc-based complexity of the non-modular and modular supply chains, respectively. Likewise, let I^N and I^M denote the respective costs of the non-modular and modular supply chains. In addition, let $\boldsymbol{q}_i^N$ and $\boldsymbol{q}_i^M$ denote demand vector of node i under non-modular and modular supply chains, respectively. Here the numbering of nodes follows the same convention introduced in Figure 7. Since the demand shares are equal across all the variants of the final assembler, we get the following demand vectors:

$$\boldsymbol{q}_i^N = \boldsymbol{q}_i^M = \left(\frac{1}{\gamma}, \frac{1}{\gamma}, \ldots, \frac{1}{\gamma}\right)_{1 \times \gamma}, i = 1, 2, 3, 4 \tag{S-3}$$

$$\boldsymbol{q}_5^M = \boldsymbol{q}_6^M = \left(\frac{1}{\gamma^2}, \frac{1}{\gamma^2}, \ldots, \frac{1}{\gamma^2}\right)_{1\times\gamma^2} \tag{S-4}$$

$$\boldsymbol{q}_7^N = \boldsymbol{q}_7^M = \left(\frac{1}{\gamma^4}, \frac{1}{\gamma^4}, \ldots, \frac{1}{\gamma^4}\right)_{1\times\gamma^4}. \tag{S-5}$$

Proof of (a). Using the complexity criterion: Recall that the arc-based complexity of an assembly supply chain is $\sum_{i=1}^{n}\sum_{v=1}^{V_i} -L_i \frac{q_{iv}}{K}\log_2\frac{q_{iv}}{K}$. With some algebra, it can be checked that the arc-based complexity can also be written as

$$\log_2 K - \frac{1}{K}\sum_{i=1}^{n} L_i \sum_{v=1}^{V_i} q_{iv}\log_2 q_{iv}. \tag{S-6}$$

By substituting in (S-6) the expression for $\boldsymbol{q}_i^N$, $i = 1,2,3,4$ given by (S-3), and the expression for $\boldsymbol{q}_7^N$, given by (S-5), and letting $K = 8$, $L_i = 1$, $i = 1,\ldots,4$ and $L_7 = 4$, we obtain the following expression for the complexity of the non-modular assembly supply chain:

$$H^{A-N} = \frac{5}{2}\log_2\gamma + \log_2 8.$$

Similarly, by substituting in (S-6) the expression for $\boldsymbol{q}_i^M$, $i = 1,2,3,4$ given by (S-3), the expression for $\boldsymbol{q}_i^M$, $i = 5,6$ given by (S-4) and the expression for $\boldsymbol{q}_7^M$, given by (S-5), and letting $K = 10$, $L_i = 1$, $i = 1,\ldots,4$ and $L_5 = L_6 = L_7 = 2$, we obtain the following expression for the complexity of the modular assembly supply chain:

$$H^{A-M} = 2\log_2\gamma + \log_2 10.$$

The complexity difference between non-modular and modular supply chains is

$$H^{A-N} - H^{A-M} = \frac{1}{2}\log_2\gamma - 0.3219$$

When $\gamma = 1$, $H^{A-N} - H^{A-M} < 0$, and the non-modular assembly supply chain is better. When $\gamma \geq 2$, $H^{A-N} - H^{A-M} > 0$, and the modular assembly supply chain is better.

Proof of (b). Using the cost criterion: Since we assume the leadtime of node i is L_i, the number of inputs assembled at node i. Therefore the cost of an assembly supply chain is

$$I = \sum_{i=1}^{N}\sum_{v=1}^{V_i} I_{iv} = \sum_{i=1}^{n}\sum_{v=1}^{V_i} C_i\sqrt{(L_i+1)\lambda q_{iv}}. \tag{S-7}$$

Because we assume all the nodes in the same echelon have the same cost coefficient, we let, for notational convenience, $C_I = C_i^N = C_i^M$, $i = 1,2,3,4$ and $C_{II} = C_i^M$, $i = 5,6$ and $C_{III} = C_7^N = C_7^M$.

By substituting in (S-7) the expression for $\boldsymbol{q}_i^N$, $i = 1, 2, 3, 4$ given by (S-3), and the expression for $\boldsymbol{q}_7^N$, given by (S-5), we obtain the following expression for the cost of the non-modular assembly supply chain:

$$I^N = C_{III}\gamma^2\sqrt{5\lambda} + 4C_I\sqrt{2\lambda\gamma}.$$

Similarly, by substituting in (S-7) the expression for $\boldsymbol{q}_i^M$, $i = 1, 2, 3, 4$ given by (S-3), the expression for $\boldsymbol{q}_i^M$, $i = 5, 6$ given by (S-4) and the expression for $\boldsymbol{q}_7^M$, given by (S-5), we obtain the following expression for the cost of the modular assembly supply chain:

$$I^M = C_{III}\gamma^2\sqrt{3\lambda} + 2C_{II}\gamma\sqrt{3\lambda} + 4C_I\sqrt{2\lambda\gamma}.$$

The cost difference between non-modular and modular supply chain is,

$$I^N - I^M = -2\gamma C_{II}\sqrt{3} + \gamma^2 C_{III}(\sqrt{5} - \sqrt{3}).$$

Define $t := \frac{2\sqrt{3}C_{II}}{(\sqrt{5}-\sqrt{3})C_{III}}$. If $\gamma \leq t$, $I^N - I^M \leq 0$ and non-modular assembly supply chain is better. If $\gamma > t$, $I^N - I^M > 0$ and modular assembly supply chain is better.

Proof of Proposition 5:

For the purposes of this proof, let H^{A-N} and H^{A-M} denote the arc-based complexity of the non-modular and modular supply chains, respectively. Likewise, let I^N and I^M denote the respective costs of the non-modular and modular supply chains. In addition, let $\boldsymbol{q}_i^N$ and $\boldsymbol{q}_i^M$ denote demand vector of node i under non-modular and modular supply chains, respectively. Here the numbering of nodes follows the same convention introduced in Figure 7. For notational convenience, let $q_{iv} = q_{iv}^N = q_{iv}^M$, where $i = 1, 2, 3, 4, 7$, $v = 1, 2, \ldots, V_i$. Let $q_{5v} = q_{5v}^M, v = 1, 2, \ldots, V_5$ and $q_{6v} = q_{6v}^M, v = 1, 2, \ldots, V_6$.

Proof of (a). Using the complexity criterion: Recall that the arc-based complexity of an assembly supply chain is

$$-\sum_{i=1}^{n}\sum_{v=1}^{V_i} L_i \frac{q_{iv}}{K}\log_2\frac{q_{iv}}{K} = \log_2 K - \sum_{i=1}^{n}\frac{L_i}{K}\sum_{v=1}^{V_i} q_{iv}\log_2 q_{iv}. \tag{S-8}$$

By substituting $K = 8$, $L_i = 1$ for $i = 1, \ldots, 4$ and $L_7 = 4$ in (S-8), we obtain the following expression for the complexity of the non-modular assembly supply chain:

$$H^{A-N} = \log_2 8 - \frac{1}{8}\sum_{i=1}^{4}\sum_{v=1}^{\gamma} q_{iv}\log_2 q_{iv} - \frac{4}{8}\sum_{v=1}^{\gamma^4} q_{7v}\log_2 q_{7v}$$

Similarly, by substituting $K = 10$, $L_i = 1$ for $i = 1, \ldots, 4$, $L_5 = L_6 = L_7 = 2$ in (S-8), we obtain the following expression for the complexity of the modular assembly supply chain:

$$H^{A-M} = \log_2 10 - \frac{1}{10}\sum_{i=1}^{4}\sum_{v=1}^{\gamma} q_{iv}\log_2 q_{iv} - \frac{2}{10}\sum_{i=5}^{6}\sum_{v=1}^{\gamma^2} q_{iv}\log_2 q_{iv} - \frac{2}{10}\sum_{v=1}^{\gamma^4} q_{7v}\log_2 q_{7v}$$

Therefore, the complexity difference between non-modular and modular assembly supply chains is

$$H^{A-N} - H^{A-M} = \log_2\frac{8}{10} - \frac{1}{40}\sum_{i=1}^{4}\sum_{v=1}^{\gamma} q_{iv}\log_2 q_{iv} + \frac{2}{10}\sum_{i=5}^{6}\sum_{v=1}^{\gamma^2} q_{iv}\log_2 q_{iv} - \frac{3}{10}\sum_{v=1}^{\gamma^4} q_{7v}\log_2 q_{7v}$$

Notice that $\boldsymbol{q}_{iv}$s depend on p only for $i = 1, 5, 7$. Hence, the term

$$\kappa_1 := \log_2\frac{8}{10} - \frac{1}{40}\sum_{i=2}^{4}\sum_{v=1}^{\gamma} q_{iv}\log_2 q_{iv} + \frac{2}{10}\sum_{v=1}^{\gamma^2} q_{6v}\log_2 q_{6v}$$

in the difference $H^{A-N} - H^{A-M}$ does not depend on p, and we can rewrite $H^{A-N} - H^{A-M}$ as

$$H^{A-N} - H^{A-M} = -\frac{1}{40}\sum_{v=1}^{\gamma} q_{1v}\log_2 q_{1v} + \frac{2}{10}\sum_{v=1}^{\gamma^2} q_{5v}\log_2 q_{5v} - \frac{3}{10}\sum_{v=1}^{\gamma^4} q_{7v}\log_2 q_{7v} + \kappa_1 \tag{S-9}$$

We know the following about the demand vectors $\boldsymbol{q}_1$, $\boldsymbol{q}_5$, and $\boldsymbol{q}_7$:

$$\boldsymbol{q}_1 = (a_1(1-p), a_2(1-p), ..., a_{\gamma-1}(1-p), p)_{1\times\gamma}\text{ , where}\sum_{i=1}^{\gamma-1} a_i = 1$$

$$\boldsymbol{q}_5 = \big(a_1(1-p)\boldsymbol{q}_{21}, \ldots, a_{\gamma-1}(1-p)\boldsymbol{q}_{2\gamma}, p\boldsymbol{q}_{21}, \ldots, p\boldsymbol{q}_{2\gamma}\big)_{1\times\gamma^2}$$

$$\boldsymbol{q}_7 = \big(a_1(1-p)\boldsymbol{q}_{21}\boldsymbol{q}_{31}\boldsymbol{q}_{41}, \ldots, a_{\gamma-1}(1-p)\boldsymbol{q}_{2\gamma}\boldsymbol{q}_{3\gamma}\boldsymbol{q}_{4\gamma}, p\boldsymbol{q}_{21}\boldsymbol{q}_{31}\boldsymbol{q}_{41}, \ldots, p\boldsymbol{q}_{2\gamma}\boldsymbol{q}_{3\gamma}\boldsymbol{q}_{4\gamma}\big)_{1\times\gamma^4}$$

We substitute these demand vectors in (S-9) to obtain:

$$\begin{aligned} H^{A-N} - H^{A-M} =& \frac{1}{8}(p\sum_{i=1}^{\gamma-1} a_i\log_2 a_i - (1-p)\log_2(1-p) - p\log_2 p) + \kappa_1 \\ & - \frac{1}{10}\sum_{i=1}^{\gamma-1} a_i\log_2 a_i + \frac{2}{10}\sum_{j=1}^{\gamma}\boldsymbol{q}_{2j}\log_2\boldsymbol{q}_{2j} - \frac{3}{10}\sum_{j=1}^{\gamma}\sum_{k=1}^{\gamma}\sum_{l=1}^{\gamma}\boldsymbol{q}_{2j}\boldsymbol{q}_{3k}\boldsymbol{q}_{4l}\log_2\boldsymbol{q}_{2j}\boldsymbol{q}_{3k}\boldsymbol{q}_{4l} \end{aligned}$$

Notice that the term

$$\kappa_2 := \frac{1}{10}\sum_{i=1}^{\gamma-1} a_i\log_2 a_i + \frac{2}{10}\sum_{j=1}^{\gamma}\boldsymbol{q}_{2j}\log_2\boldsymbol{q}_{2j} - \frac{3}{10}\sum_{j=1}^{\gamma}\sum_{k=1}^{\gamma}\sum_{l=1}^{\gamma}\boldsymbol{q}_{2j}\boldsymbol{q}_{3k}\boldsymbol{q}_{4l}\log_2\boldsymbol{q}_{2j}\boldsymbol{q}_{3k}\boldsymbol{q}_{4l}$$

does not depend on p. We can rewrite the complexity difference between non-modular and modular assembly supply chains as:

$$H^{A-N} - H^{A-M} = \frac{1}{8}\left(p\sum_{i=1}^{\gamma-1} a_i\log_2 a_i - (1-p)\log_2(1-p) - p\log_2 p\right) + \kappa_1 + \kappa_2$$

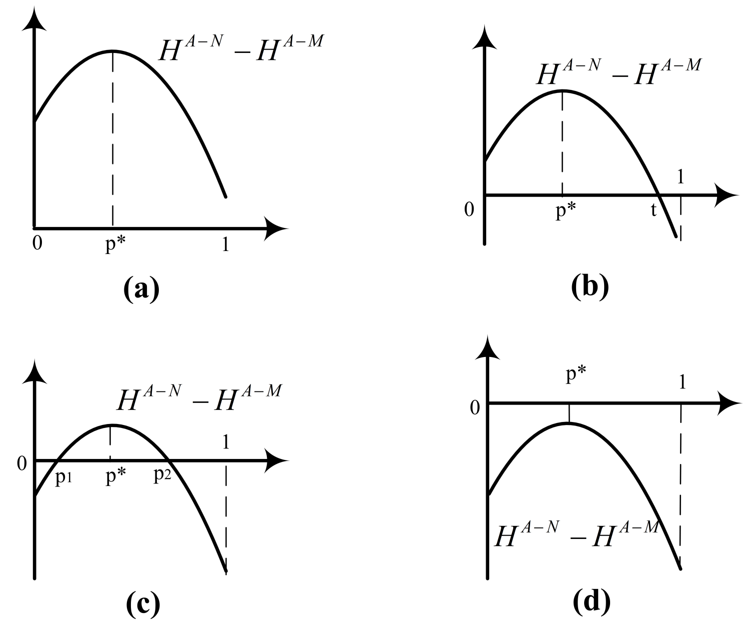

Figure 8: Four different scenarios of $H^{A-N} - H^{A-M}$

For ease of notation, let $R := \sum_{i=1}^{\gamma-1} a_i \log_2 a_i$ and $f(p) := \frac{1}{8}\left(pR - (1-p)\log_2(1-p) - p\log_2 p\right)$. With these definitions, note that $H^{A-N} - H^{A-M} = f(p) + \kappa_1 + \kappa_2$. It is not difficult to check that $f(0) = 0$ and $f(1) < 0$. Furthermore:

$$
\begin{aligned}
f'(p) &= \frac{1}{8}(R - \log_2 p + \log_2(1-p)), \\
f''(p) &= -\frac{1}{8p(1-p)\ln 2} < 0 \text{ for } p \in (0,1).
\end{aligned}
$$

Observe from above that $f(p)$ is concave in p. Define p^* so that $f'(p^*) = 0$. One can check that $p^* = \frac{2^R}{1+2^R}$. Because $R \leq 0$, we observe $p^* = \frac{2^R}{1+2^R} \leq 1/2$. Therefore, the function $f(p)$ reaches its highest point for some $p \leq 1/2$. This observation, combined with $f(0) = 0$ and $f(1) < 0$, implies that the difference $H^{A-N} - H^{A-M} = f(p) + \kappa_1 + \kappa_2$ is concave in p and reaches its highest point at some $p < 1$ and its lowest point at $p = 1$. Next, we use this information and we consider a series of cases to show all the possibilities about the sign of $H^{A-N} - H^{A-M}$ as a function of p.

Case 1. $f(1) + \kappa_1 + \kappa_2 \geq 0$: See Figure 8 (a) for an illustration of this case. Given that $H^{A-N} - H^{A-M}$ reaches its lowest point at $p = 1$, it follows that, in this case, $H^{A-N} - H^{A-M} \geq 0$ for all p. Therefore, the modular assembly supply chain is better for any p.

Case 2. $f(1) + \kappa_1 + \kappa_2 < 0$: In this case, we need to consider two subcases.

Case 2(a). $f(1) + \kappa_1 + \kappa_2 < 0$ *and* $f(0) + \kappa_1 + \kappa_2 \geq 0$: See Figure 8 (b) for an illustration of this case. In this case, the function $H^{A-N} - H^{A-M}$ starts out non-negative at $p = 0$ and ends up being negative at $p = 1$. Since the function $f(p)$ is strictly concave, there must exist $t \in [0,1]$ such that $H^{A-N} - H^{A-M} \geq 0$ for $p \leq t$ and $H^{A-N} - H^{A-M} < 0$ for $p > t$.

Case 2(b). $f(1) + \kappa_1 + \kappa_2 < 0$ and $f(0) + \kappa_1 + \kappa_2 < 0$: There are two further subcases to consider. If $f(p^*) + \kappa_1 + \kappa_2 \geq 0$, then there must exist p_1 and p_2 such that $H^{A-N} - H^{A-M}$ is non-negative for all $p \in [p_1, p_2]$ and negative elsewhere. (See Figure 8 (c) for an illustration of this case.) If $f(p^*) + \kappa_1 + \kappa_2 < 0$, then it must be that $H^{A-N} - H^{A-M}$ is negative for all p. (See Figure 8 (d) for an illustration of this case.)

Proof of (b). Using the cost criterion: We assume the leadtime of node i is L_i, the number of inputs assembled at node i. Then the cost of an assembly supply chain is

$$I = \sum_{i=1}^{n} \sum_{v=1}^{V_i} I_{iv} = \sum_{i=1}^{n} \sum_{v=1}^{V_i} C_i \sqrt{(L_i + 1)\lambda q_{iv}} \tag{S-10}$$

By substituting $L_i = 1$ for $i = 1, \ldots, 4$ and $L_7 = 4$ in (S-10), we obtain the cost of the non-modular assembly supply chain:

$$I^N = C_{III}\sqrt{5\lambda} \sum_{v=1}^{V_7} \sqrt{q_{7v}} + C_I \sqrt{2\lambda} \sum_{i=1}^{4} \sum_{v=1}^{V_i} \sqrt{q_{iv}}$$

Similarly, by substituting $L_i = 1$ for $i = 1, \ldots, 4$, $L_5 = L_6 = L_7 = 2$ in (S-10), we obtain the cost of the modular assembly supply chain:

$$I^M = C_{III}\sqrt{3\lambda} \sum_{l=1}^{V_7} \sqrt{q_{7l}} + C_{II}\sqrt{3\lambda} \sum_{i=5}^{5} \sum_{i=1}^{V_i} \sqrt{q_{iv}} + C_I \sqrt{2\lambda} \sum_{i=1}^{4} \sum_{v=1}^{V_i} \sqrt{q_{iv}}$$

Cost difference between non-modular and modular assembly supply chain is

$$I^N - I^M = C_{III}(\sqrt{5} - \sqrt{3})\sqrt{\lambda} \sum_{v=1}^{V_7} \sqrt{q_{7v}} - C_{II}\sqrt{3\lambda}\left(\sum_{v=1}^{V_5} \sqrt{q_{5v}} + \sum_{v=1}^{V_6} \sqrt{q_{6v}}\right) \tag{S-11}$$

Recall we have the following demand vectors:

$$\boldsymbol{q}_1 = (a_1(1-p), a_2(1-p), ..., a_{\gamma-1}(1-p), p)_{1\times\gamma}, \text{where} \sum_{i=1}^{\gamma-1} a_i = 1$$

$$\boldsymbol{q}_5 = \left(a_1(1-p)\boldsymbol{q}_{21}, \ldots, a_{\gamma-1}(1-p)\boldsymbol{q}_{2\gamma}, p\boldsymbol{q}_{21}, \ldots, p\boldsymbol{q}_{2\gamma}\right)_{1\times\gamma^2}$$

$$\boldsymbol{q}_7 = \left(a_1(1-p)\boldsymbol{q}_{21}\boldsymbol{q}_{31}\boldsymbol{q}_{41}, \ldots, a_{\gamma-1}(1-p)\boldsymbol{q}_{2\gamma}\boldsymbol{q}_{3\gamma}\boldsymbol{q}_{4\gamma}, p\boldsymbol{q}_{21}\boldsymbol{q}_{31}\boldsymbol{q}_{41}, \ldots, p\boldsymbol{q}_{2\gamma}\boldsymbol{q}_{3\gamma}\boldsymbol{q}_{4\gamma}\right)_{1\times\gamma^4}$$

By substituting the demand vector $\boldsymbol{q}_5$ and $\boldsymbol{q}_7$ back into (S-11), we can rewrite the cost difference between non-modular and modular assembly supply chains as

$$I^N - I^M = S \cdot R\sqrt{1-p} + R\sqrt{p} - T$$

where S, R and T, constants with respect to p, are given by

$$S = \sum_{i=1}^{\gamma-1} \sqrt{a_i},$$
$$R = C_{III}(\sqrt{5} - \sqrt{3})\sqrt{\lambda} \sum_{j=1}^{\gamma} \sum_{k=1}^{\gamma} \sum_{l=1}^{\gamma} \sqrt{\boldsymbol{q}_{2j}\boldsymbol{q}_{3k}\boldsymbol{q}_{4l}} - C_{II}\sqrt{3\lambda} \sum_{j=1}^{\gamma} \sqrt{\boldsymbol{q}_{2j}},$$
$$T = C_{II}\sqrt{3\lambda} \sum_{v=1}^{V_6} \sqrt{q_{6v}}.$$

Let $g(p) = S \cdot R\sqrt{1-p} + R\sqrt{p}$. With this definition, notice that $I^N - I^M = g(p) - T$. Taking derivatives of $G(p)$, we obtain:

$$g'(p) = -\frac{SR}{\sqrt{1-p}} + \frac{R}{\sqrt{p}}$$
$$g''(p) = -R\left(\frac{S}{4}(1-p)^{-3/2} + \frac{1}{4}p^{-3/2}\right)$$

Notice from above that $G(p)$ is concave if $R \geq 0$ and convex if $R < 0$. Hence, we divide the proof into two cases depending on whether R is negative or non-negative.

(i) $R \geq 0$:

If $R \geq 0$, then $g''(p) \leq 0$, and $g(p)$ is concave. Let p^* be the maximizer of $g(p)$. One can check that $p^* = \frac{R^2}{S^2+R^2}$ and $p^* \in (0,1)$. In addition, $g(0) = SR$ and $g(1) = R$. Because $R \geq 0$ and $S = \sum_{i=1}^{\gamma-1} \sqrt{a_i} \geq 1$, it follows that $g(0) = SR \geq g(1) = R$. Using these observations, we note that $g(p)$ is concave and reaches its highest point at $p^* < 1$ and its lowest point at $p = 1$. Keeping this in mind, we now consider a number of subcases, depending on the value of T relative to $R \leq SR \leq g(p^*)$.

Case 1. $T \leq R$: See Figure 9 (a) for an illustration of this case. The lowest point of $I^N - I^M$, which occurs at $p = 1$, is given by $g(1) - T = R - T \geq 0$. Hence, $I^N - I^M \geq 0$ for all $p \in [0,1]$, and the modular assembly supply chain is better for all $p \in [0,1]$.

Case 2. $R < T \leq SR$: See Figure 9 (b) for an illustration of this case. The function $I^N - I^M$ is equal to $SR - T \geq 0$ at $p = 0$ and is equal to $R - T < 0$ at $p = 1$. Since the function $g(p)$ is strictly concave, there must exist $t \in [0,1]$ such that $I^N - I^M \geq 0$ for $p \leq t$ and $I^N - I^M < 0$ for $p > t$.

Case 3. $R \leq SR < T \leq g(p^*)$: See Figure 9 (c) for an illustration of this case. The function $I^N - I^M$ is equal to $SR - T < 0$ at $p = 0$, reaches $g(p^*) - T \geq 0$ at its peak and is equal to $R - T \leq 0$. In other words, $I^N - I^M$ starts negative, becomes positive and then again negative. Given $g(p)$ is concave, there must exist p_1 and p_2 such that $I^N - I^M$ is non-negative for all $p \in [p_1, p_2]$ and negative elsewhere.

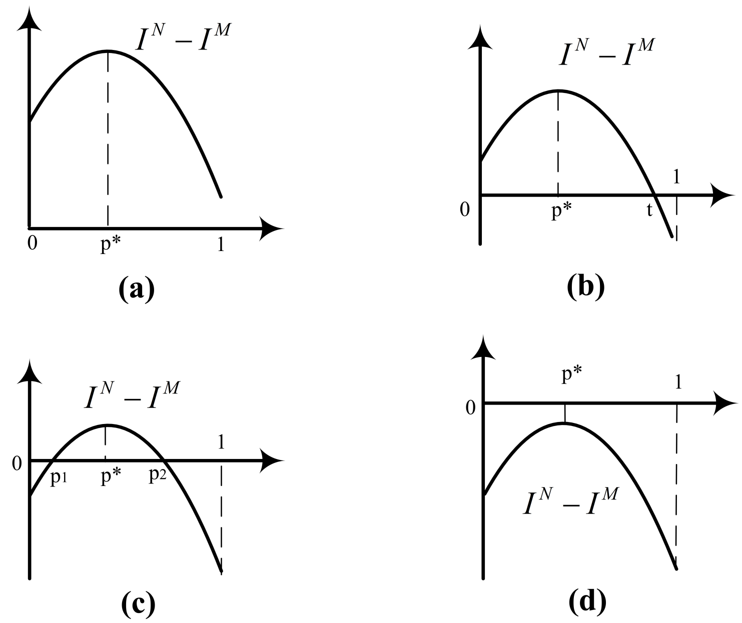

Figure 9: Four different scenarios of $I^N - I^M$

<u>*Case 4.* $R \le SR \le g(p^*) < T$:</u> See Figure 9 (d) for an illustration of this case. The function $I^N - I^M$ is equal to $g(p^*) - T < 0$ at its peak. Therefore, $I^N - I^M < 0$ for all p, which implies that the non-modular assembly supply chain is better for any p.

<u>**(ii)** $R < 0$</u>:

If $R < 0$, then $g''(p) > 0$, and $g(p)$ is convex. Let p^* be the minimizer of $g(p)$. One can check that $p^* = \frac{R^2}{S^2+R^2}$ and $p^* \in (0, 1)$. In addition, $g(0) = SR$ and $g(1) = R$. Because $R < 0$ and $S = \sum_{i=1}^{\gamma-1} \sqrt{a_i} \ge 1$, it follows that $g(0) = SR \le g(1) = R$. Using these observations, we note that $g(p)$ is convex and reaches its lowest point at $p^* < 1$ and its highest point at $p = 1$. Since $g(1) = R < 0$, then it follows that $g(p) < 0$ for all $p \in [0, 1]$. Hence, $I^N - I^M = g(p) - T < 0$ for all $p \in [0, 1]$ and the non-modular assembly chain has lower cost than modular assembly supply chain for any p.

Proof of Proposition 6:

For the purposes of this proof, let I^N and I^M denote the respective costs of the non-modular and modular supply chains. In addition, let $\boldsymbol{q}_i^N$ and $\boldsymbol{q}_i^M$ denote demand vector of node i under non-modular and modular supply chains, respectively. Here the numbering of nodes follows the same convention introduced in Figure 7. Notice that $\boldsymbol{q}_i^N = \boldsymbol{q}_i^M$ for node $i = 1, 2, 3, 4, 7$, because the supply chains are the same in terms of the final assembler's demand vector. For notational convenience, let $q_{iv} = q_{iv}^N = q_{iv}^M$, where $i = 1, 2, 3, 4, 7$, $v = 1, 2, \ldots, V_i$. Let $q_{5v} = q_{5v}^M, l = 1, 2, \ldots, V_5$

and $q_{6v} = q_{6v}^M, l = 1, 2, \ldots, V_6$. Because we assume all the nodes in the same echelon have the same cost coefficient, we let, for notational convenience, $C_I = C_i^N = C_i^M$, $i = 1, 2, 3, 4$ and $C_{II} = C_i^M$, $i = 5, 6$ and $C_{III} = C_7^N = C_7^M$. Recall that costs of an assembly supply chain as follows, if we assume the leadtime of node i is L_i, the number of inputs assembled at node i.

$$I = \sum_{i=1}^{n} \sum_{v=1}^{V_i} I_{iv} = \sum_{i=1}^{n} \sum_{v=1}^{V_i} C_i \sqrt{(L_i + 1)\lambda q_{iv}}$$

By substituting $L_i = 1$ for $i = 1, \ldots, 4$ and $L_7 = 4$ in the above equation, we obtain the cost of the non-modular assembly supply chain:

$$I^N = C_I \sqrt{2\lambda} \sum_{i=1}^{4} \sum_{v=1}^{V_i} \sqrt{q_{iv}} + C_{III} \sqrt{5\lambda} \sum_{v=1}^{V_7} \sqrt{q_{7v}}$$

Similarly, by substituting $L_i = 1$ for $i = 1, \ldots, 4$, $L_5 = L_6 = L_7 = 2$ in the above equation, we obtain the cost of the modular assembly supply chain:

$$I^M = C_I \sqrt{2\lambda} \sum_{i=1}^{4} \sum_{v=1}^{V_i} \sqrt{q_{iv}} + C_{II} \sqrt{3\lambda} \sum_{i=5}^{5} \sum_{i=1}^{V_i} \sqrt{q_{iv}} + C_{III} \sqrt{3\lambda} \sum_{v=1}^{V_7} \sqrt{q_{7v}}$$

Cost difference between non-modular and modular assembly supply chain is

$$I^N - I^M = -C_{II} \sqrt{3\lambda} \left(\sum_{v=1}^{V_5} \sqrt{q_{5v}} + \sum_{v=1}^{V_6} \sqrt{q_{6v}} \right) + C_{III} (\sqrt{5} - \sqrt{3}) \sqrt{\lambda} \sum_{v=1}^{V_7} \sqrt{q_{7v}} \qquad \text{(S-12)}$$

Observe from (S-12) that $I^N - I^M$ is an increasing function of C_{III}. Suppose that $I^N - I^M < 0$ at a given value of C_{III}, which means that the cost criterion favors the non-modular assembly supply chain. If we decrease the cost coefficient of the final assembler, C_{III}, we will continue to have $I^N - I^M < 0$ (because $I^N - I^M$ is a increasing function of C_{III}), so the cost criterion continues to favor the non-modular supply chain. Since the complexity of a supply chain does not change when C_{III} changes, part (a) of the proposition follows.

On the other hand, if we increase the cost coefficient of the final assembler, C_{III}, $I^N - I^M$ will eventually exceed zero. Hence, there is a threshold T such that $I^N - I^M \geq 0$ for when $C_{III} \geq T$, which means that the cost criterion favors the modular supply chain once C_{III} exceeds a threshold. The complexity again does not depend on C_{III}. Hence, part (b) follows.

Lemma 1. *Suppose m-dimensional vectors $A = (a_1, a_2, ..., a_m)$ and $B = (b_1, b_2, ..., b_m)$ and n-dimensional vector $C = (c_1, c_2, ..., c_n)$ are all sorted in descending order. If $A \succ B$, then $\Omega(A, C) \succ \Omega(B, C)$ where the operation $\Omega(\boldsymbol{x}, \boldsymbol{y})$ is the vector obtained by component-wise multiplication of vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, sorted in descending order.*

Proof of Lemma 1: Define $(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m)^{\downarrow}$ as the vector $(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m)$ sorted in descending order. We prove the result by induction on n, the dimension of the vector C. First, when $n = 1$, in which case $C = (c_1)$, we have:

$$\Omega(A, C) = (a_1c_1, a_2c_1, ..., a_mc_1), \tag{S-13}$$

$$\Omega(B, C) = (b_1c_1, b_2c_1, ..., b_mc_1). \tag{S-14}$$

Because $A = (a_1, a_2, ..., a_m) \succ B = (b_1, b_2, ..., b_m)$, it follows that

$$\Omega(A, C) = (a_1c_1, a_2c_1, ..., a_mc_1) \succ \Omega(B, C) = (b_1c_1, b_2c_1, ..., b_mc_1).$$

Hence, the result holds when $n = 1$. Suppose that if $n = k-1$, in which case $C = (c_1, c_2, \ldots, c_{k-1})$, the result holds. That is:

$$\begin{aligned} &\Omega(A, C) = (a_1c_1, ..., a_mc_1, a_1c_2, ..., a_mc_2, ..., a_1c_{k-1}, ..., a_mc_{k-1})^{\downarrow} \\ \succ\, &\Omega(B, C) = (b_1c_1, ..., b_mc_1, b_1c_2, ..., b_mc_2, ..., b_1c_{k-1}, ..., b_mc_{k-1})^{\downarrow}. \end{aligned}$$

In the remainder of the proof, define

$$\begin{aligned} T &= \left(t_1, t_2, ..., t_{m\cdot(k-1)}\right) := (a_1c_1, ..., a_mc_1, a_1c_2, ..., a_mc_2, ..., a_1c_{k-1}, ..., a_mc_{k-1})^{\downarrow} \\ S &= \left(s_1, s_2, ..., s_{m\cdot(k-1)}\right) := (b_1c_1, ..., b_mc_1, b_1c_2, ..., b_mc_2, ..., b_1c_{k-1}, ..., b_mc_{k-1})^{\downarrow} \end{aligned}$$

Notice that $T \succ S$ by the induction assumption. To complete the induction, we will prove that if $n = k$ and $C = (c_1, c_2, \ldots, c_k)$, then $\Omega(A, C) \succ \Omega(B, C)$. $\Omega(A, C)$ is obtained by inserting the numbers $a_1c_k, \ldots, a_mc_k$ in descending order into the vector T. $\Omega(B, C)$ is obtained similarly by inserting the numbers $b_1c_k, \ldots, b_mc_k$ in descending order into the vector S. To complete the induction, we use yet another induction, this time on the pairs of numbers inserted into vectors T and S. To clarify, we will first show that, after inserting the first pair of numbers, a_1c_k into T and b_1c_k into S, the resulting vectors $(T, a_1c_k)^{\downarrow}$ and $(S, b_1c_k)^{\downarrow}$ are such that $(T, a_1c_k)^{\downarrow} \succ (S, b_1c_k)^{\downarrow}$. We will then make the induction assumption that after inserting $g - 1$ pairs of numbers, we still have $(T, a_1c_k, \ldots, a_{g-1}c_k)^{\downarrow} \succ (S, b_1c_k, \ldots, b_{g-1}c_k)^{\downarrow}$. We will then show that, after inserting the g-th pair of numbers, a_gc_k and b_gc_k, we have $(T, a_1c_k, \ldots, a_gc_k)^{\downarrow} \succ (S, b_1c_k, \ldots, b_gc_k)^{\downarrow}$. This will conclude the proof of both the inner and outer induction, concluding the proof of Lemma 1.

Suppose the first pair of numbers to be inserted, a_1c_k and b_1c_k, are such that $t_i \geq a_1c_k \geq t_{i+1}$ and $s_j \geq b_1c_k \geq s_{j+1}$. Then, after inserting a_1c_k into T and b_1c_k into S, and sorting the vectors, we get

$$\begin{aligned} (T, a_1c_k)^{\downarrow} &= (t_1, t_2, ..., t_i, a_1c_k, t_{i+1}, \ldots, t_{m\cdot(k-1)}) \\ (S, b_1c_k)^{\downarrow} &= (s_1, s_2, ..., s_j, b_1c_k, s_{j+1}, \ldots, s_{m\cdot(k-1)}). \end{aligned}$$

We next show that $(T, a_1c_k)^{\downarrow} \succ (S, b_1c_k)^{\downarrow}$. Consider two cases: $i \leq j$ or $i > j$.

<u>*Case 1: $i \leq j$*</u>

To see why $(T, a_1c_k)^{\downarrow} \succ (S, b_1c_k)^{\downarrow}$, note that:

(i) For $z = 1, 2, ..., i$, we have $\sum_{l=1}^{z} t_l \geq \sum_{l=1}^{z} s_l$ since $T \succ S$.

(ii) For $z = i+1, i+2, ..., j$, we have $\sum_{l=1}^{z-1} t_l + a_1c_k \geq \sum_{l=1}^{z-1} t_l + t_z = \sum_{l=1}^{z} t_l \geq \sum_{l=1}^{z} s_l$, where the first inequality holds because (a) $a_1c_k \geq t_{i+1}$ by assumption and (b) $t_{i+1} \geq t_z$ since the vector T is sorted, and the second inequality holds because $T \succ S$.

(iii) For $z = j+1, j+2, ..., m(k-1)+1$, $\sum_{l=1}^{z-1} t_l + a_1c_k \geq \sum_{l=1}^{z-1} s_l + b_1c_k$ because (a) $T \succ S$ implies that $\sum_{l=1}^{z-1} t_l \geq \sum_{l=1}^{z-1} s_l$, and (b) $A \succ B$ implies that $a_1c_k \geq b_1c_k$.

<u>*Case 2: $i > j$*</u>

To see why $(T, a_1c_k)^{\downarrow} \succ (S, b_1c_k)^{\downarrow}$, note that:

(i) For $z = 1, 2, ..., j$, we have $\sum_{l=1}^{z} t_l \geq \sum_{l=1}^{z} s_l$ since $T \succ S$.

(ii) For $z = j+1, j+2, ..., i$, we have $\sum_{l=1}^{z} t_l = \sum_{l=1}^{z-1} t_l + t_z \geq \sum_{l=1}^{z-1} s_l + a_1c_k \geq \sum_{l=1}^{z-1} s_l + b_1c_k$. The first inequality holds because $T \succ S$ and $t_i \geq a_1c_k$, and the second inequality holds because $a_1c_k \geq b_1c_k$ from $A \succ B$.

(iii) For $z = i+1, i+2, ..., m(k-1)+1$, $\sum_{l=1}^{z-1} t_l + a_1c_k \geq \sum_{l=1}^{z-1} s_l + b_1c_k$ because (a) $T \succ S$ implies that $\sum_{l=1}^{z-1} t_l \geq \sum_{l=1}^{z-1} s_l$, and (b) $A \succ B$ implies that $a_1c_k \geq b_1c_k$.

Hence, we have shown that after inserting the first pair of numbers, a_1c_k and b_1c_k, it is true that $(T, a_1c_k)^{\downarrow} \succ (S, b_1c_k)^{\downarrow}$. We now make the induction assumption that, after inserting $a_{g-1}c_k$ and $b_{g-1}c_k$, it is true that $(T, a_1c_k, a_2c_k, \ldots, a_{g-1}c_k)^{\downarrow} \succ (S, b_1c_k, b_2c_k, \ldots, b_{g-1}c_k)^{\downarrow}$. For the remainder of the proof, define

$$T' = (t'_1, t'_2, \ldots, t'_{m\cdot(k-1)+(g-1)}) := (T, a_1c_k, a_2c_k, \ldots, a_{g-1}c_k)^{\downarrow}$$
$$S' = (s'_1, s'_2, \ldots, s'_{m\cdot(k-1)+(g-1)}) := (S, b_1c_k, b_2c_k, \ldots, b_{g-1}c_k)^{\downarrow}$$

Thus, the induction assumption can be written as $T' \succ S'$. To complete the induction, we prove that after inserting a_gc_k and b_gc_k into T' and S', it is true that

$$\begin{aligned}(T', a_gc_k)^{\downarrow} &= (T, a_1c_k, a_2c_k, \ldots, a_{g-1}c_k, a_gc_k)^{\downarrow}\\ \succ (S', b_gc_k)^{\downarrow} &= (S, b_1c_k, b_2c_k, \ldots, b_{g-1}c_k, b_gc_k)^{\downarrow}\end{aligned}$$

Suppose $t'_i \geq a_gc_k \geq t'_{i+1}$ and $s'_j \geq b_gc_k \geq s'_{j+1}$. Then:

$$(T', a_gc_k)^{\downarrow} = (T, a_1c_k, a_2c_k, \ldots, a_{g-1}c_k, a_gc_k)^{\downarrow} = \left(t'_1, t'_2, \ldots, t'_i, a_gc_k, t'_{i+1}, \ldots, t'_{m\cdot(k-1)+(g-1)}\right) \tag{S-15}$$

$$(S', b_gc_k)^{\downarrow} = (S, b_1c_k, b_2c_k, \ldots, b_{g-1}c_k, b_gc_k)^{\downarrow} = \left(s'_1, s'_2, \ldots, s'_j, b_gc_k, s'_{j+1}, \ldots, s'_{m\cdot(k-1)+(g-1)}\right) \tag{S-16}$$

We will consider two cases: $i \leq j$, and $i > j$.

Case 1: $i \leq j$

To see why $(T', a_g c_k)^{\downarrow} \succ (S', b_g c_k)^{\downarrow}$, note that:

(i) For $z = 1, 2, ..., i$, we have $\sum_{l=1}^{z} t'_l \geq \sum_{l=1}^{z} s'_l$ since $T' \succ S'$.

(ii) For $z = i+1, i+2, ..., j$, we have $\sum_{l=1}^{z-1} t'_l + a_g c_k \geq \sum_{l=1}^{z-1} t'_l + t'_z = \sum_{l=1}^{z} t'_l \geq \sum_{l=1}^{z} s'_l$, where the first inequality holds because (a) $a_g c_k \geq t'_{i+1}$ by assumption and (b) the vector T' is sorted and, hence, $t'_{i+1} \geq t'_z$, and the second inequality holds because $T' \succ S'$.

(iii) For $z = j+1, j+2, ..., m(k-1)+g$, the set $\{t'_1, \ldots, t'_{z-1}, a_g c_k\}$ can be divided into two disjoint subsets, $\{t_1, \ldots, t_{z-g}\}$ and $\{a_1 c_k, \ldots, a_g c_k\}$ (see (S-15)). Similarly, the set $\{s'_1, \ldots, s'_{z-1}, b_g c_k\}$ can also be divided into two disjoint subsets, $\{s_1, \ldots, s_{z-g}\}$ and $\{b_1 c_k, \ldots, b_g c_k\}$ (see (S-16)). Now, we obtain $\sum_{l=1}^{z-1} t'_l + a_g c_k = \sum_{l=1}^{z-g} t_l + \sum_{u=1}^{g} a_u c_k \geq \sum_{l=1}^{z-g} s_l + \sum_{u=1}^{g} b_u c_k = \sum_{l=1}^{z-1} s'_l + b_g c_k$, where the inequality holds because (a) $\sum_{l=1}^{z-g} t_l \geq \sum_{l=1}^{z-g} s_l$ by $T \succ S$ and (b) $\sum_{u=1}^{g} a_u c_k \geq \sum_{u=1}^{g} b_u c_k$ by $A \succ B$.

Case 2: $i > j$

To see why $(T', a_g c_k)^{\downarrow} \succ (S', b_g c_k)^{\downarrow}$, note that:

(i) For $z = 1, 2, ..., j$, we have $\sum_{l=1}^{z} t'_l \geq \sum_{l=1}^{z} s'_l$ since $T' \succ S'$.

(ii) For $z = j+1, j+2, ..., i$, the set $\{t'_1, \ldots, t'_z\}$ can be divided into two disjoint subsets, $\{t_1, \ldots, t_{z-y}\}$ and $\{a_1 c_k, \ldots, a_y c_k\}$, where $y \leq g$. Then,

$$\begin{aligned}
\sum_{l=1}^{z} t'_l &= \sum_{l=1}^{z-y} t_l + \sum_{u=1}^{y} a_u c_k \\
&= \sum_{l=1}^{z-g} t_l + \sum_{l=z-g}^{z-y} t_l + \sum_{u=1}^{y} a_u c_k \\
&\geq \sum_{l=1}^{z-g} t_l + \sum_{u=y+1}^{g} a_u c_k + \sum_{u=1}^{y} a_u c_k \\
&= \sum_{l=1}^{z-g} t_l + \sum_{u=1}^{g} a_u c_k
\end{aligned} \tag{S-17}$$

where the inequality holds because T' is sorted in descending order and $t_{z-y} \geq a_{y+1} c_k$. Now, the set $\{s'_1, \ldots, s'_{z-1}, b_g c_k\}$ can also be divided into two disjoint subsets, $\{s_1, \ldots, s_{z-g}\}$ and $\{b_1 c_k, \ldots, b_g c_k\}$ (see (S-16)). We now note that $\sum_{l=1}^{z} t'_l \geq \sum_{l=1}^{z-g} t_l + \sum_{u=1}^{g} a_u c_k \geq \sum_{l=1}^{z-g} s_l + \sum_{u=1}^{g} b_u c_k = \sum_{l=1}^{z-1} s'_l + b_g c_k$, where the first inequality follows from (S-17) and the second inequality holds because (a) $\sum_{l=1}^{z-g} t_l \geq \sum_{l=1}^{z-g} s_l$ by $T \succ S$ and (b) $\sum_{u=1}^{g} a_u c_k \geq \sum_{u=1}^{g} b_u c_k$ by $A \succ B$.

(iii) For $z = i+1, i+2, ..., m(k-1)+g$, the set $\{t'_1, \ldots, t'_{z-1}, a_g c_k\}$ can be divided into two disjoint subsets, $\{t_1, \ldots, t_{z-g}\}$ and $\{a_1 c_k, \ldots, a_g c_k\}$ (see (S-15)). Similarly, the set $\{s'_1, \ldots, s'_{z-1}, b_g c_k\}$ can also be divided into two disjoint subsets, $\{s_1, \ldots, s_{z-g}\}$ and $\{b_1 c_k, \ldots, b_g c_k\}$ (see (S-16)). Now, we obtain $\sum_{l=1}^{z-1} t'_l + a_g c_k = \sum_{l=1}^{z-g} t_l + \sum_{u=1}^{g} a_u c_k \geq \sum_{l=1}^{z-g} s_l + \sum_{u=1}^{g} b_u c_k = \sum_{l=1}^{z-1} s'_l + b_g c_k$, where the inequality holds because (a) $\sum_{l=1}^{z-g} t_l \geq \sum_{l=1}^{z-g} s_l$ by $T \succ S$ and (b) $\sum_{u=1}^{g} a_u c_k \geq \sum_{u=1}^{g} b_u c_k$ by $A \succ B$..

Thus, we have proven the inner induction, which then proves the outer induction and concludes the proof.

Appendix C: Tables

	Factor					Inconsistency Rate (%)		
Run	Complexity Definition	Num of Suppliers	Num of Variants	Cost Disparity	Demand Evenness	Rep 1	Rep 2	Rep 3
1	-	-	-	-	-	2.31	1.97	2.26
2	+	-	-	-	-	2.82	2.5	2.79
3	-	+	-	-	-	3.24	3.64	3.1
4	+	+	-	-	-	4.59	4.96	4.17
5	-	-	+	-	-	3.16	3.37	3.05
6	+	-	+	-	-	3.51	3.85	3.31
7	-	+	+	-	-	3.54	3.53	3.47
8	+	+	+	-	-	4.27	4.22	4
9	-	-	-	+	-	2.43	2.57	2.51
10	+	-	-	+	-	3.26	3.16	3.12
11	-	+	-	+	-	3.74	3.98	3.82
12	+	+	-	+	-	5.05	5.19	4.94
13	-	-	+	+	-	3.8	3.72	3.41
14	+	-	+	+	-	4.26	4.17	3.66
15	-	+	+	+	-	3.66	3.63	4.19
16	+	+	+	+	-	4.35	4.42	4.91
17	-	-	-	-	+	0.5	0.53	0.5
18	+	-	-	-	+	1.1	1.03	1.07
19	-	+	-	-	+	0.8	0.81	0.91
20	+	+	-	-	+	1.77	1.68	1.76
21	-	-	+	-	+	0.64	0.68	0.58
22	+	-	+	-	+	0.92	1.12	0.9
23	-	+	+	-	+	0.56	0.69	0.47
24	+	+	+	-	+	0.88	0.87	0.8
25	-	-	-	+	+	0.86	0.74	0.77
26	+	-	-	+	+	1.5	1.26	1.45
27	-	+	-	+	+	0.86	0.84	0.82
28	+	+	-	+	+	1.93	1.66	1.84
29	-	-	+	+	+	0.82	0.71	0.67
30	+	-	+	+	+	1.28	1.25	1.04
31	-	+	+	+	+	0.49	0.75	0.66
32	+	+	+	+	+	0.86	0.98	0.96

Table S-1: The percentage of inconsistencies for each of the 32 value combinations of the five factors shown in Table 1. The possible values for each factor is as shown in Table 1. For each combination, three replications are run.

	Factor					Inconsistency Rate (%)		
Run	Complexity Definition	Num of Suppliers	Num of Variants	Cost Disparity	Demand Evenness	Rep 1	Rep 2	Rep 3
1	-	-	-	-	-	3.78	3.75	3.61
2	+	-	-	-	-	5.19	5.19	5.11
3	-	+	-	-	-	3.87	3.7	4.14
4	+	+	-	-	-	5.82	5.42	5.93
5	-	-	+	-	-	3.71	3.67	3.81
6	+	-	+	-	-	4.56	4.27	4.7
7	-	+	+	-	-	3.67	3.72	3.58
8	+	+	+	-	-	4.55	4.66	4.38
9	-	-	-	+	-	4.36	4.01	3.95
10	+	-	-	+	-	5.66	5.42	5.52
11	-	+	-	+	-	3.92	3.66	4.07
12	+	+	-	+	-	5.71	5.46	5.92
13	-	-	+	+	-	3.76	3.48	3.78
14	+	-	+	+	-	4.69	4.34	4.44
15	-	+	+	+	-	3.45	3.65	3.8
16	+	+	+	+	-	4.33	4.46	4.78
17	-	-	-	-	+	0.86	0.62	0.82
18	+	-	-	-	+	1.8	1.47	1.75
19	-	+	-	-	+	0.56	0.54	0.57
20	+	+	-	-	+	1.42	1.36	1.3
21	-	-	+	-	+	0.57	0.61	0.57
22	+	-	+	-	+	0.77	0.69	0.61
23	-	+	+	-	+	0.51	0.57	0.6
24	+	+	+	-	+	0.62	0.75	0.8
25	-	-	-	+	+	0.76	0.79	0.69
26	+	-	-	+	+	1.54	1.72	1.53
27	-	+	-	+	+	0.58	0.67	0.63
28	+	+	-	+	+	1.33	1.31	1.38
29	-	-	+	+	+	0.53	0.63	0.64
30	+	-	+	+	+	0.83	0.85	0.9
31	-	+	+	+	+	0.65	0.49	0.57
32	+	+	+	+	+	0.8	0.73	0.68

Table S-2: The percentage of inconsistencies for each of the 32 value combinations of the five factors shown in Table 1. The possible values for each factor is as shown in Table 1, except that the number of suppliers is now four or five (as opposed two or three). For each combination, three replications are run. Notice that the number of suppliers is no longer a statistically significant effect.

Term	Effect	t-Statistics	p Value
Constant		186.28	< 0.0001
Complexity definition	0.865	30.53	< 0.0001
Number of suppliers	-0.026	-0.91	0.365
Number of variants	-0.520	-18.35	< 0.0001
Cost disparity between two echelons	0.048	1.71	0.093
Evenness of the demand vector	-3.530	-124.57	< 0.0001

Table S-3: Estimated Effects, t-Statistics and p-Value (%) of S-2. A positive (negative) effect implies that when the factor's value changes from '-' to '+', the inconsistency rate increases (decreases).

Run	Factor					Inconsistency Rate (%)		
	Complexity Definition	Num of Suppliers	Num of Variants	Cost Disparity	Demand Evenness	Rep 1	Rep 2	Rep 3
1	-	-	-	-	-	2.92	2.64	2.93
2	+	-	-	-	-	3.45	3.21	3.46
3	-	+	-	-	-	3.59	4.19	3.59
4	+	+	-	-	-	4.94	5.51	4.66
5	-	-	+	-	-	3.48	3.88	3.51
6	+	-	+	-	-	3.87	4.38	3.83
7	-	+	+	-	-	3.71	3.73	3.64
8	+	+	+	-	-	4.44	4.42	4.17
9	-	-	-	+	-	2.77	2.82	2.81
10	+	-	-	+	-	3.6	3.41	3.42
11	-	+	-	+	-	3.92	4.18	3.97
12	+	+	-	+	-	5.23	5.39	5.09
13	-	-	+	+	-	4.02	3.93	3.77
14	+	-	+	+	-	4.48	4.38	4.02
15	-	+	+	+	-	3.73	3.66	4.27
16	+	+	+	+	-	4.42	4.45	4.99
17	-	-	-	-	+	0.89	0.95	0.81
18	+	-	-	-	+	1.49	1.45	1.38
19	-	+	-	-	+	0.97	1.04	1.16
20	+	+	-	-	+	1.94	1.91	2.05
21	-	-	+	-	+	0.88	0.99	0.9
22	+	-	+	-	+	1.2	1.49	1.22
23	-	+	+	-	+	0.63	0.77	0.5
24	+	+	+	-	+	0.97	0.97	0.85
25	-	-	-	+	+	1.05	0.85	0.94
26	+	-	-	+	+	1.69	1.37	1.62
27	-	+	-	+	+	0.91	0.95	0.9
28	+	+	-	+	+	1.98	1.79	1.92
29	-	-	+	+	+	0.98	0.77	0.77
30	+	-	+	+	+	1.44	1.31	1.14
31	-	+	+	+	+	0.5	0.76	0.69
32	+	+	+	+	+	0.87	1.01	0.99

Table S-4: The percentage of inconsistencies for each of the 32 value combinations of the five factors shown in Table 1. The possible values for each factor is as shown in Table 1, except that the cost of the downstream echelon is now four or five times as large as the sum of cost coefficients of the nodes in the upstream echelon (as opposed to being the same as the sum or twice as large). For each combination, three replications are run. Notice that the cost disparity is no longer a statistically significant effect.

Term	Effect	t-Statistics	p Value
Constant		137.13	< 0.0001
Complexity definition	0.647	17.49	< 0.0001
Number of suppliers	0.382	10.33	< 0.0001
Number of variants	-0.082	-2.21	0.030
Cost disparity between two echelons	0.091	2.46	0.017
Evenness of the demand vector	-2.797	-75.62	< 0.0001

Table S-5: Estimated Effects, t-Statistics and p-Value (%) of S-4. A positive (negative) effect implies that when the factor's value changes from '-' to '+', the inconsistency rate increases (decreases).

	Factor					Inconsistency Rate (%)		
	Complexity	Num of	Num of	Cost	Demand	Rep	Rep	Rep
Run	Definition	Echelon	Variants	Disparity	Evenness	1	2	3
1	-	-	-	-	-	1.97	2.1	2.14
2	+	-	-	-	-	2.59	2.64	2.76
3	-	+	-	-	-	3.99	3.61	3.81
4	+	+	-	-	-	4.59	4.16	4.42
5	-	-	+	-	-	3.13	3.59	3.12
6	+	-	+	-	-	3.63	3.96	3.48
7	-	+	+	-	-	1.42	1.51	1.58
8	+	+	+	-	-	1.61	1.68	1.92
9	-	-	-	+	-	2.77	2.49	2.65
10	+	-	-	+	-	3.42	3.14	3.37
11	-	+	-	+	-	4.35	4.66	4.27
12	+	+	-	+	-	4.92	5.31	4.74
13	-	-	+	+	-	3.65	3.45	3.66
14	+	-	+	+	-	4.12	3.99	4.05
15	-	+	+	+	-	2.94	3.27	3.23
16	+	+	+	+	-	3.18	3.67	3.43
17	-	-	-	-	+	0.49	0.49	0.53
18	+	-	-	-	+	1.01	1.05	1.32
19	-	+	-	-	+	0.8	0.87	1
20	+	+	-	-	+	1.18	1.13	1.23
21	-	-	+	-	+	0.69	0.52	0.52
22	+	-	+	-	+	1.03	0.89	0.9
23	-	+	+	-	+	0.7	0.76	0.8
24	+	+	+	-	+	0.67	0.71	0.76
25	-	-	-	+	+	0.98	0.64	0.77
26	+	-	-	+	+	1.58	1.34	1.4
27	-	+	-	+	+	1.01	1.21	1.1
28	+	+	-	+	+	1.37	1.51	1.29
29	-	-	+	+	+	0.74	0.67	0.71
30	+	-	+	+	+	1.07	1.05	1.17
31	-	+	+	+	+	1.07	1.1	1.07
32	+	+	+	+	+	1.13	1.11	1.06

Table S-6: The percentage of inconsistencies for each of the 32 value combinations of the five factors shown in Table 1. The possible values for each factor is as shown in Table 1, except that the number of suppliers is fixed at two and the factor of the number of suppliers is replaced with number of echelons (number of echelons could either be two or three). Notice that number of echelons is a statistically significant effect.

Term	Effect	t-Statistics	p Value
Constant		143.02	< 0.0001
Complexity definition	0.399	13.40	< 0.0001
Number of echelons	0.198	6.64	< 0.0001
Number of variants	-0.333	-11.20	< 0.0001
Cost disparity between two echelons	0.530	17.79	< 0.0001
Evenness of the demand vector	-2.332	-78.35	< 0.0001

Table S-7: Estimated Effects, t-Statistics and p-Value (%) of S-6. A positive (negative) effect implies that when the factor's value changes from '-' to '+', the inconsistency rate increases (decreases).