MICHIGAN ROSS

Working Paper

Product Variety, Sourcing Complexity, and the Bottleneck of Coordination

Yue Maggie Zhou
Stephen M. Ross School of Business
University of Michigan

Xiang Wan
Fisher College of Business
Ohio State University

Ross School of Business Working Paper Series
Working Paper No. 1329
August 2016

PRODUCT VARIETY, SOURCING COMPLEXITY, AND THE BOTTLENECK OF COORDINATION

Yue Maggie Zhou*

Stephen M. Ross School of Business
University of Michigan
701 Tappan St., R4446
Ann Arbor, MI 48109-1234
ymz@umich.edu

Xiang Wan

Fisher College of Business
Ohio State University
Columbus, OH 43210
wan.207@osu.edu

August 10, 2016

*Corresponding author.

PRODUCT VARIETY, SOURCING COMPLEXITY, AND THE BOTTLENECK OF COORDINATION

ABSTRACT

This paper studies the coordination burden for firms that pursue variety as their main product strategy. We propose that product variety magnifies the tension between scale economy in production and scope economy in distribution, giving rise to complex intra-firm sourcing relationships. Sourcing complexity worsens coordination performance for the firm and poses a dilemma for organization design: A hierarchical structure with sourcing hubs reduces sourcing complexity but can create bottlenecks at the hubs. We empirically examine operations data for about 300 distribution centers within a major soft drink bottling company in 2010–2011. Results support our hypotheses, illuminating the source of complexity in multi-product firms and the challenge for organization design in managing product variety.

Key words: product variety, complexity, coordination, bottleneck, organization structure, stockouts

INTRODUCTION

Complex systems, or systems with a large number of interdependent relationships (Simon, 1962), have been a subject of intense managerial research over the last two decades. In particular, the business firm has been viewed as a complex system that transfers information, materials, and energy between tasks (Baldwin, 2008). The costs of coordinating such a complex system limit a firm's ability to diversify into related businesses (Zhou, 2011), and may partially offset the strategic benefits of product variety (MacDuffie, Sethuraman, & Fisher, 1996). Two gaps exist in this literature. First, despite the large number of studies on the consequences of complexity, the presence of complexity is mostly assumed rather than theorized (Puranam, Raveendran, & Knudsen, 2012). Second, the various conceptualizations of complexity and its costs are difficult to measure using empirical data. Most prior work keeps the concept at an abstract level using NK models and relies on computer simulations (Burton & Obel, 1980; Ethiraj, Levinthal, & Roy, 2008; Lenox, Rockart, & Lewin, 2006, 2007; Levinthal, 1997; Rivkin, 2000b; Rivkin & Siggelkow, 2007) or case studies (Siggelkow, 2001, 2002). The few exceptional econometric studies are limited to the industry (Schilling & Steensma, 2001), product (Hoetker, 2006), or transaction level (Puranam, 2001) rather than the firm level. For these reasons, identifying potential sources of complexity in real business firms will help us to pinpoint the loci of coordination, as well as to reevaluate strategies that cause complexity in the first place.

Against this background, our paper investigates the emergence of a particular type of complexity: the complexity that pertains to inter-unit sourcing networks within multi-product firms. We argue that the tradeoff between economies of scope and scale creates complex inter-unit sourcing relationships—what we call "sourcing complexity"—for firms that pursue variety as their main product strategy. On the one hand, pursuing economies of scope requires firms to distribute multiple varieties to the same customer through shared marketing and delivery channels. On the other hand, achieving economies of scale requires specializing production by variety at the plant level. For instance, a plant may produce one variety of cereal for markets across the Eastern seaboard as opposed to producing many varieties of cereal for New Jersey alone. Optimization *within* each stage along the value chain creates a coordination burden *between* stages: Specializing distribution by customer order to maximize scope economy while specializing production by variety to maximize scale economy causes sourcing complexity between production units (e.g., plants) and distribution units (e.g., distribution centers).

We then analyze the consequences of sourcing complexity for performance and organization design, respectively. We argue that sourcing complexity worsens performance by imposing extra coordination burdens on organization units. For example, plants and distribution centers have to coordinate for the order, execution, and confirmation of shipments, for the scheduling of production according to idiosyncratic demand at the distribution centers, and for the appraisal of product specifications. While some elements of these tasks can be centralized, each pair of units in a sourcing relationship has its own contingencies that need to be coordinated. Failure to communicate, to appraise interactions, or to correct decision errors will cause delays and a poor synchronization of tasks, such as a mismatch between supply and demand. A mismatch of this type results in a stockout, where a customer order is not completely fulfilled due to insufficient supply. In addition, sourcing complexity presents challenges for organization design. In particular, a hierarchical structure can reduce sourcing complexity by allowing plants to ship to a small number of sourcing hubs, which then consolidate shipments and forward them to DCs. However, a hierarchical structure also creates coordination burdens for the sourcing hubs, which may in turn cascade to downstream DCs. As an increasing number of shipments pass through hubs, those hubs can become bottlenecks, negatively affecting the performance of their downstream DCs.

We propose three hypotheses based on these arguments: (1) In the presence of economies of scale in production, product variety increases sourcing complexity; (2) Sourcing complexity worsens coordination performance; and (3) Controlling for sourcing complexity, DCs experience worse coordination performance when they source through hubs. We empirically test these hypotheses in the context of the soft drink industry, where the two dominant concentrate manufacturers (CMs), Coca-Cola and Pepsi, compete fiercely on product variety and service level, including low stockout rates. Bottling processes are both capital-intensive and highly specialized by variety, while sales are driven by variety assortment and direct-store delivery. As a result, the tradeoff between scale economy in production and scope economy in distribution is salient. We examine monthly operations data across about 300 DCs at a major soft drink bottling company (the Company) wholly owned by one of the CMs between 2010 and 2011. Detailed operations data allows us to study sourcing relationships and coordination performance at a granular level for each pair of DC and stock-keeping-unit (SKU)—the finest level of product variety (Fosfuri & Giarratana, 2009).

We find evidence supporting our hypotheses. First, we show that product variety increases sourcing complexity for each node in the network: DCs that carry more varieties also source from a larger number of units. Second, as DCs' sourcing complexity increases, their performance worsens (in the form of a higher stockout rate) for each SKU they carry. These results hold even when controlling for a host of other factors that might affect stockouts, including sales quantity, sales volatility, inventory, demand forecast, as well as seasonal, regional, and DC–SKU-pair fixed effects. We also find a spillover effect: When a DC sources from a larger number of units for a particular SKU, other SKUs carried by that DC also experience a higher stockout rate. Finally, a hierarchical structure increases coordination burden at hubs, such that hub DCs experience higher stockout rates than non-hub DCs, particularly when their downstream DCs suffer demand shocks. In addition, when controlling for sourcing complexity, DCs experience a higher stockout rate when they source through a hub. A similar spillover effect becomes apparent here, as well: When a DC sources through a hub for a particular SKU, other SKUs the DC carries (those not sourced through any hubs) experience a higher stockout rate. These results are robust to a number of alternative measures and subsamples that mitigate potential endogeneity.

The paper's main theoretical contribution is marrying several streams of work on complexity, organization design, and product variety. First, it complements recent studies showing that firms pursuing economies of scope face a coordination burden created by complexity, and confirms that complex interdependencies along a firm's value chain contribute to this coordination burden (Zhou, 2011). By explicating the relationship between product variety, sourcing complexity, and coordination burdens in firms, we also confirm that acting on strategic opportunities requires a dynamic capability for coordinating highly interdependent productive systems (Teece, Pisano, & Shuen, 1997), including the capability to design an effective organizational structure (Agarwal & Helfat, 2009; Gulati & Puranam, 2009), such as a sourcing network.

Secondly, the paper complements classic work on the role of hierarchical structure in alleviating coordination burdens in multi-product, multi-divisional firms (Chandler, 1962). In the Company that we study, top management focused on long-term strategies (carbonated vs. non-carbonated drinks, domestic vs. international markets); regional business units focused on quarterly promotion, demand forecast, and production scheduling of various product varieties; plants and DCs focused on weekly production, distribution, and the shipments of physical goods;

and hubs coordinated this complex sourcing network. Our theory about complexity and hierarchical structure can be generalized to most sourcing networks within firms that integrate both plants and distribution centers. We build on Chandler (1962) by extending recent efforts to conceptualize the locus of coordination and bottlenecks in complex systems (Baldwin, 2014). We show that while hubs help to reduce complexity, these loci of coordination may turn into organizational bottlenecks when the burden of coordination is too large.

Finally, while prior studies (mostly in the operations management literature) have related stockouts to demand forecast errors, inventory shortages, and scheduling difficulties arising from product variety (Fisher, 1997; Fisher & Ittner, 1999), these problems have been largely studied at the level of a single plant or production line. Studies about interactions between organizational units are rare (Ramdas, 2003). This paper fills that important gap.

THEORETICAL DEVELOPMENT

Product Variety, Economies of Scope vs. Scale, and Sourcing Complexity

Product variety offers the potential for economies of scope. On the demand side, if customers have demands for multiple varieties, and their demands for each variety are not perfectly correlated and vary stochastically over time, the volatility in aggregate demand for all varieties will be smaller than the summation of volatilities in the demand for each variety (e.g., Anupindi *et al.*, 2011; Cachon & Terwiesch, 2012). On the supply size, offering multiple varieties allows a firm to share production facilities (to some extent), marketing and distribution channels, brand reputation, and knowledge across a large number of products (Ramdas, 2003). For example, holding constant the total shelf space at retail customers' stores, consumers' preference for multiple varieties implies small delivery sizes, as well as frequent orders and deliveries for each variety. Frequent deliveries to a customer may disrupt the customer's business and cost the individual distributors more in total transportation. In contrast, aggregating a customer's orders for multiple varieties into a single order and delivering the order in its entirety by a single distributor reduces the number of stops and saves transportation costs. Handling a customer order in its entirety also allows sales and marketing staff to specialize, thereby servicing major customers more efficiently "with a single face." For these reasons, delivering a customer order with all the varieties included is often required in industries where direct-store delivery by manufacturers is the norm.

However, product variety hinders the potential for economies of scale, or the potential for the manufacturing plants to amortize fixed-cost inputs (equipment, process technology, training, and manufacturing overhead) and set-up time over a high volume of output. Figure 1 depicts the relationship between product variety, scale economy, and the internal sourcing network between plants and distribution centers. In the absence of economies of scale, each DC can have a dedicated plant making all of its varieties. With the presence of economies of scale, firms that produce only a small number of varieties can have a similar sourcing network. Such network has a one-to-one correspondence between plants and DCs, as shown in Figure 1a. A simple and decentralized etwork allows each plant and DC to assume full responsibility for customers in their geographic area.

Insert Figure 1 about here

If manufacturing processes exhibit economies of scale (as almost all do), increasing product variety will compromise these economies in several ways (Fisher and Ittner, 1999). As an assembly line processes an increasing number of varieties, learning accumulates more slowly because workers must alternate their focus between different varieties and apply a different job to each work-in-progress arriving at their station. Variation in job requests increases the probability of error, which increases the amount of downtime, minor on-line rework, and major off-line repairs, ultimately reducing labor productivity. In addition, sharing multiple varieties in the same production process increases equipment and overhead costs. For example, a greater variety requires more costly variety-handling systems (including information systems, parts presentation systems, and conveyor systems) and more coordination overhead. Workers need to be trained in skills to produce multiple varieties or to operate variety-handling machinery and software. Maintenance and technical staff need to spend time with tooling and retooling. Managers need to make frequent decisions about line scheduling and adjustments. These extra costs offset economies of scale and limit the number of varieties that each plant can process.

Thus, with the presence of economies of scale, firms offering a large number of varieties will specialize production based on variety to preserve scale economy at the plant level: each customer order will be “split” for production by specialized plants. Each DC will then source products from multiple plants, “reconfigure” varieties into the requested assortments, and deliver them according to the original customer order, thereby maintaining the economies of scope in distribution (Closs, Nyaga, & Voss, 2010). In the extreme case, where the products manufactured in each plant do not overlap (achieving the highest economies of scale in production), each DC will source products from all plants, as shown in Figure 1b. Therefore, for firms with great product variety, maximizing scale economy in production and scope economy in distribution creates a “complex system” of sourcing relationships among the plants and DCs.

More generally, a complex system is a system with a large number of interdependent relationships (Simon, 1962). Interdependencies may arise when the outcome of a task performed by an agent or organization unit depends on another task, often a task performed by another agent or organization unit (Milgrom & Roberts, 1990; Puranam *et al.*, 2012). For example, a firm can be viewed as a complex system of value-chain tasks interrelated through physical input–output feedback loops designed to transfer and transform information and materials (Baldwin, 2008; Porter, 1985a; Sturgeon, 2002). In fact, as summarized by Lenox, Rockart, and Lewin (2008), within a real firm, interdependence-induced complexity can exist along multiple dimensions and at multiple levels, such as the complexity in decisions regarding business and corporate strategies (Levinthal, 1997; Porter, 1997; Rivkin, 2000a; Siggelkow, 2002), in production technologies and managerial practices (Ichniowski, Shaw, & Prennushi, 1997; Milgrom & Roberts, 1990, 1995), and in product-design choices (Baldwin & Clark, 2000; Henderson & Clark, 1990). Most of these studies represent a complex system using a matrix or a network, with the total number of non-empty cells in the matrix or the total number of linkages in the network denoting the prevalence of interdependencies and, therefore, complexity (Baldwin & Clark, 2000; Kauffman, 1993; Levinthal, 1997).

In our context of the sourcing and shipment network, each DC's performance (e.g., stockouts, which affect customer satisfaction and future sales) depends on the sourcing plants' timely production and shipments to the DCs, and each plant's performance (e.g., production costs) depends on the DCs' accurate demand forecasts and timely communication of any demand shocks or product defects. The sourcing and shipment network can therefore be a complex system. The complexity of such a system can be measured using the number of linkages, or the number of sourcing and shipment relationships, in the system. We therefore propose the following hypothesis:

> *Hypothesis 1: In the presence of economies of scale in production, product variety increases sourcing complexity.*

Sourcing Complexity and Coordination Burden

In general, complexity can create a coordination burden in many ways. Because of complex interdependencies, organization units must engage in ongoing communication in order to understand the factors affecting one another's decisions and to track the decisions that are actually made, (e.g., decisions about planned production, shipments, and inventory holding), particularly when multiple equilibria exist (Arrow, 1974; Becker & Murphy, 1992). Organization units must also process more information about the interactions between decisions (Simon, 1955). This increased workload of communication and information processing will in turn create more opportunities for decision errors (Levinthal, 1997; Sutherland, 1980).

In developing our theory for the particular context of a sourcing network, we assume that every sourcing relationship creates a coordination burden. This is consistent with prior studies on inter-unit shipments. For example, Miller and Vollmann (1985) specified four types of costly "transactions" that are required for inter-unit shipments. Logistical transactions relate to the order, execution, and confirmation of shipments. Balancing transactions pertain to the scheduling of materials, labor, capacity, and production according to demand and customer orders. Quality transactions occur when units communicate to appraise specifications. Change transactions accommodate changes in engineering designs, schedules, specifications, and shipments. Managing these inter-unit transactions costs the units time and effort and contributes to their coordination burden.

While some elements of these inter-unit transactions can be centralized, each pair of sourcing relationships has its own contingencies that need to be coordinated. For example, in our empirical context, even though the centralized business units schedule a quarterly production plan for each plant, as well as a sourcing matrix among plants and DCs for each SKU based on the specialization, capacity, production costs, and location of each plant, plants and DCs have to coordinate among themselves to manage any real-time deviation from these quarterly plans. Therefore the coordination burden of a given unit is proportional to the number of shipment relationships it manages, and the coordination burden of the entire sourcing network is proportional to the total number of shipment relationships in, or the complexity of, the sourcing network. As an illustration, the network in Figure 1b is “more complex” than that in Figure 1a, each node (both plants and DCs) in Figure 1b has a higher coordination burden than each node in Figure 1a, and the network in Figure 1b has a higher aggregate coordination burden (with 25 sourcing relationships) than the network in Figure 1a (with five sourcing relationships).

The increased coordination burden in a complex (as opposed to a simple) network implies increased workload related to communicating and processing/appraising interdependencies, which creates more opportunities for decision error. Failure to communicate and appraise interactions, or to correct decision errors, will cause delays and a poor synchronization of tasks. These coordination failures in turn may increase the probability of a mismatch of supply and demand: a stockout. We therefore predict the following:

> *Hypothesis 2: Sourcing complexity increases the probability of coordination failures and thus worsens coordination performance.*

Hierarchical Structure and Bottlenecks of Coordination

A potential organizational response to complexity is to adopt a hierarchical structure to manage interdependencies among organization units (Zhou, 2013). A hierarchical sourcing structure is one in which a few hubs receive consolidated shipments from plants, and then forward them along to non-hub DCs. This hierarchical structure reduces the number of sourcing relationships and, thus the total coordination burden (according to H2). In an extreme case, a single centralized sourcing hub reduces the number of sourcing relationships to the number of plants plus the number of DCs. For example, a centralized sourcing structure would reduce the number of sourcing relationships from 25 in Figure 1b to 10 in Figure 2a.

Insert Figure 2 about here

However, even though a hierarchical structure reduces the overall complexity of the sourcing network, it places extra coordination burden on the hubs. As a central hub (Figure 2a), DC 3 now needs to coordinate five incoming shipment relationships with five plants and four outgoing shipment relationships with four other DCs, which is substantially more than the five incoming shipment relationships it needs to coordinate under a totally decentralized structure (Figure 1b). Therefore, as a locus of interdependent relationships between plants and DCs, a hub can become a bottleneck of coordination for its downstream DCs.

In a complex system, a bottleneck can be conceptualized as a component that obstructs a flow, thereby limiting the overall performance of a system (Baldwin, 2014). We argue that coordination bottlenecks arise at loci of coordination due to, on the one hand, a congestion of extensive interdependencies and, on the other hand, limits in the coordination capacity of each organizational unit. This is like a route network in the airline industry. A point-to-point system ensures maximum flexibility and timely travel between any pair of airports when the entire route network is relatively simple. As the number of routes increases and the route network gets more complex, a hub-and-spoke system reduces the total number of routes, easing coordination for the entire network and reducing average delay. At the same time, compared to direct flights, flights that pass through hubs are more likely to experience delays due to congestions at the hubs.

Figure 2b gives a schematic example. When part of the shipments from plants 2, 3, 4, and 5 are sent to a hub (DC3) instead of to individual DCs (e.g., DC3, DC4, DC5), the total number of relationships in the network reduces from 25 (Figure 1b) to 17. This reduction in complexity will improve average coordination, in keeping with Hypothesis 2. For example, plant 2 now only needs to coordinate with two DCs (DC1 and DC3)—instead of with the five DCs shown in Figure 1b—which will ease coordination for all DCs that require products made by plant 2. At the same time, even though DC1 and DC2 each maintain five sourcing relationships, DC1 coordinates directly with the plants for its own needs, whereas DC2 coordinates directly with four plants (plants 1, 3, 4, 5) and indirectly with plant 2 through a hub (DC3). Potential congestions at the hub will worsen coordination performance for DC2 but not for DC1, which does not source through any hub. In general, as a locus of coordination (e.g., a hub) approaches its coordination capacity, it becomes a bottleneck, and its subordinate units will experience a deterioration in performance. We therefore propose the following:

Hypothesis 3: Controlling for sourcing complexity, DCs experience worse coordination performance when they source through a hub.

THE SOFT DRINK INDUSTRY AND THE BOTTLING COMPANY

The soft-drink industry provides an excellent context for our study. First, competition in the industry is mainly focused on product variety and the quality of direct-store-delivery service (e.g., stockouts), the two key concepts of our theoretical analyses. Rapid development of new product variety is a dominant strategy in the industry. Coca-Cola and Pepsi have each produced thousands of SKUs. Coca-Cola, for instance, introduced twice as many new brands/flavors of its Coca-Cola soft drink in 2010 as in 2009 (Coca Cola Press Release, 2010). A low stockout rate is an important performance measure in the industry. A low stockout rate not only improves customer satisfaction; it also reduces the chance that customers switch to rival products. In addition to cola products, the CMs produce sparkling drinks, bottled and vitamin waters, sports drinks, fruit juices, teas, coffees, and many more. Secondly, the variety of soft drink products can be evaluated along relatively few dimensions, such as brand, content, flavor, and packaging (material and size), but with numerous options along each dimension, allowing for large but comparable variations across DCs and SKUs. Finally, the bottling process is capital-intensive and relies on high-speed production lines that are "interchangeable only for products of similar type and packages of similar size" (Yoffie & Kim, 2011: 2). In contrast, the distribution process is largely influenced by drop size. For example, bottlers in NYC are among the least profitable because of that city's notorious traffic, parking, and large population of small stores selling small quantities. As a result, bottlers always try to consolidate variety assortment along the same route and to the same customer, making coordination between production and distribution more salient.

Our data come from one of the largest bottlers. Like most of its peers, the Company employs a make-to-stock (as opposed to make-to-order) inventory system. Products are produced and stocked at a stable pace according to a forecast of future demand, i.e., before retailers place actual orders. Orders arrive in various forms and are entered into a centralized electronic ordering (e-order) system. Customers can order via call centers and large customers can enter their orders directly. In addition, sales representatives and truck drivers from the Company visit some of their retail customers regularly (often at the delivery time). They inspect their customers' inventory and sales records and use wireless handheld computers to place electronic orders for replenishment on behalf of their customers. The majority of these orders are replenished on a weekly or biweekly basis. Customers and sales representatives order based on the Company's product lists, national advertisement, and promotion deals, without knowledge of the actual inventory level at the DCs. The Company owns about 50 bottling plants and 264 DCs in the

United States. Our data cover its U.S. operations. The Company delivers its products on trucks to retailer stores both large (e.g., supermarkets) and small (e.g., convenience stores). Stockouts occur when a DC cannot deliver an entire order for an SKU to a given retail outlet. Unfilled demand is not backordered. New orders for the next period are placed based on the retail store's current inventory levels.

Regional business units design a quarterly production plan for each plant, as well as a sourcing matrix among plants and DCs for each SKU based on the specialization, capacity, production costs, and location of each plant. Plants and DCs mostly coordinate among themselves to manage any deviation from these quarterly plans. The Company tries to retain four weeks of forecasted demand in inventory at the beginning of every four-week period, though actual inventory levels can vary due to demand/production variations and capacity constraints. Each plant produces a certain array of varieties. Most plants ship to more than one DC, and most DCs get shipments from more than one plant. Plants do not see order information.

EMPIRICAL DESIGN

In order to test our hypotheses, we need operations data across different units (plants, as well as hub and non-hub distribution units (DCs)), including data about the sourcing relationships among these units, and data about the coordination performance of each product variety at each DC. With this data, we first test H1 at the DC level. We construct a sourcing network among the units and estimate the complexity of that network as a function of product variety. Given that the number of the units in the network is constant over the sample period, the total number of inter-unit sourcing relationships in the sourcing network is perfectly correlated with the number of sourcing relationships for each DC in the network. We therefore estimate the number of sourcing relationships for each DC as a function of the number of product varieties that the DC carried in each period. Next we evaluate H2 and H3 by estimating a key measure of coordination performance (e.g., stockouts) at the DC–product variety level. For H2, we estimate stockouts at a focal DC as a function of sourcing complexity the DC faces. For H3, we compare stockouts between SKUs that were sourced from a hub vs. those that were not sourced form a hub.

Sample and Variables

We obtained operations data for all 264 DCs from the third month of 2010 to the second month of 2011. Together the DCs delivered about 1,400 SKUs of products owned by the CM parent. An SKU is defined as a unique combination of brand, content, flavor, weight, container material-size, and package material-size. It captures the finest level of product variety (Fosfuri & Giarratana, 2009). Our data are weekly except for inventory level, which is available for each four-week period. Because of the frequency of the inventory data, and in order to save computation time, we aggregated the data to the period level. Our final sample contained about one million DC–SKU–period observations.

Product Variety$_{it}$ is measured using two sets of variables, one based on the number of brands carried by the DC and the other based on the number of SKUs. In order to avoid collinearity between number of brands and number of SKUs, we orthogonalized the SKU measure.

Sourcing Complexity$_{it}$, our main variable, is intended to capture the complexity, or the number of inter-unit sourcing relationships, in the sourcing network. We therefore used the number of units from which DC i received shipments for all SKUs in period t.

Stockout$_{sit}$, our coordination performance measure, is a dummy variable that takes the value of 1 if SKU s experienced at least one stockout at DC i in period t, and is 0 otherwise. Frequent stockouts result in customer dissatisfaction and ultimately hurt sales, profitability, and future demand; it is therefore a frequent subject and important performance measure in the product-variety literature (Anderson, Fitzsimons, & Simester, 2006; Musalem *et al.*, 2010). We measured stockouts using a dummy rather than a continuous variable because the distribution of stockout quantity is highly skewed. The average rate of a stockout is 27 percent (Table 1), which means that in 73 percent of cases stockout quantity was zero. In addition, the Company used the dummy variable rather than the magnitude of the stockout (usually small) as a performance measure and adjusted production and sourcing accordingly. We followed this practice and used the dummy variable. We ran a robustness check using the continuous measure; results were similar.

Being A Hub$_{sit}$ is a dummy variable that captures if DC *i* shipped SKU *s* to at least one other DC during period *t.* Based on the inter-unit shipment data, we constructed a sourcing network for each SKU in each period and identified all the hubs for each SKU.

Sourcing Through Hub$_{sit}$ is a dummy variable that captures if DC *i* received SKU *s* through a hub.

We included several control variables, all at the DC–SKU–period level, for factors that would affect stockouts based on standard textbooks in operations management (e.g., Anupindi et al., 2011; Cachon & Terwiesch, 2012). The Company standardizes the quantity of sales, inventory, and shipments of all products into cases. For example, 24 12-ounce cans are counted as one case. *Sales*$_{sit}$ is the quantity of sales in cases, log transformed. *Sales Volatility*$_{sit}$ is the standard deviation of weekly sales quantity within a period. Conditioned on the inventory level, the more volatile sales are, the higher the chance of stockouts. *Beginning Inventory*$_{sit}$ is the quantity of inventory in cases, log transformed, at the beginning of each period. The higher the inventory level is, the lower the chance of stockouts. *Demand Forecast*$_{sit}$ is the ratio between forecasted and actual sales. The higher the demand forecast is, the lower the likelihood of stockouts. *Shipment Quantity*$_{sit-1}$ is the shipment quantity in standard cases DC *i* receives during previous period, log transformed.

Table 1 provides descriptive statistics at the DC–SKU–period level. The table shows that the average stockout rate was 27 percent; supplementary statistics show that the average stockout rate was 25 percent for non-hub DCs and 46 percent for hub DCs. An average DC sold 31 brands and 569 product SKUs. An average DC received products from 4.3 other units.[1] In about 12 percent of the cases, the focal DC shipped the focal SKU to at least one other DC; supplementary statistics show that an average hub shipped to 24 other DCs. In about 53 percent of cases, the focal DC received the focal SKU through a hub. An average DC sold over 63 (=exp(4.18)) cases of each SKU it carried in every period, with an average volatility of 0.17. It carried about 28 days of sales as inventory for each SKU. On average, DCs tended to over-forecast: the forecasted demand was about 104 percent of actual sales. This suggests that stockouts were more likely due to day-to-day demand variations within a period rather than persistent under-forecasting throughout the period, placing the coordination burden primarily on the units that coordinate weekly shipments. On average, a DC received 11 (=exp(2.38)) cases of the focal SKU from other units.

Insert Table 1 about here

Table 2 presents the correlation matrix.

Insert Table 2 about here

Empirical Strategy

---

[1] Although unlikely, it is not impossible that in a given four-week period a DC did not source from any plant but sold from inventory. We found this happened in three percent of cases. We reran our analyses excluding these observations; our results remained the same.

We first estimated sourcing complexity across all SKUs at the DC level based on the following specification:

$$Sourcing\ Complexity_{it} = \alpha_0 + Region_i + Season_t + \alpha_1 Product\ Variety_{it} + A \cdot CV_{it} + \varepsilon_{it} \quad (1),$$

where *Sourcing Complexity*$_{it}$ and *Product Variety*$_{it}$ are as defined earlier. CV_{it} is a vector of control variables including sales, sales volatility, inventory, and demand forecast at the DC level. $Region_i$ are fixed effects for one of the five regions that DC *i* belong to. Units in different regions are relatively separate from units within the same region, though at times DCs also received shipments from outside their region to cover shortage within the region. We also replaced regional fixed effects with DC fixed effects in a robustness analysis to account for unobservable and time-invariant DC characteristics. $Season_t$ are season dummies capturing season-specific factors that may influence demand or supply. We clustered standard errors at the DC level to account for correlation within each DC. H1 predicts α_1>0.

We then estimated the probability of DC–SKU–level stockouts against DC-level sourcing complexity, according to the following specification:

$$Stockout_{ist} = \beta_0 + Region_i + Season_t + \beta_1 Product\ Variety_{it}$$
$$+\beta_2 Sourcing\ Complexity_{it} + \beta_3 Sourcing\ Throug\ Hub_{sit} + B \cdot CV_{ist} + \varepsilon_{ist} \quad (2),$$

where $Stockout_{ist}$, *Product Variety*$_{it}$, *Sourcing Complexity*$_{it}$, *Sourcing Through Hub*$_{sit}$, $Region_i$, and $Season_t$ are as defined earlier. CV_{ist} is a vector of control variables including sales, sales volatility, inventory, demand forecast, and shipment quantity received during last period at the DC–SKU level. To account for unobserved SKU-specific factors that might influence stockouts, we included the average stockout rate for the focal SKU across all DCs during the sample period. In a robustness regression, we also replaced regional fixed effects with DC-SKU pair fixed effects to remove time-invariant DC–SKU characteristics such as whether the SKU is a carbonated or non-carbonated product, or the years of experience DC *i* has in carrying SKU *s*. H2 predicts $\beta_2 > 0$, and H3 predicts $\beta_3 > 0$.

We took a few steps to address the issue of potential endogeneity, including selection, unobserved heterogeneity, and reverse causality. These are explained in detail in the results section.

RESULTS

Product Variety and Sourcing Complexity

Table 3 estimates sourcing complexity at the DC level. We start with a cross-sectional model with only the control variables and season dummies. Column (1) suggests that DCs with a higher level of aggregate sales (from all SKUs) and DCs with a higher level of inventory were associated with more sourcing complexity. On the other hand, DCs with more volatile aggregate sales were associated with less sourcing complexity.

Insert Table 3 about here

Column (2) introduces product variety, measured using number of brands and SKU, respectively. Consistent with H1, DCs that carried more varieties sourced from a larger number of units. Column (3) decomposes the time-varying brand and SKU counts into two components: $Product\ Variety_{it} = \overline{Product\ Variety}_i + (Product\ Variety_{it} - \overline{Product\ Variety}_i)$, where the first cross-sectional component represents the average number of brands or SKUs carried by DC *i* over the sample period, and the second intertemporal component represents the number of brands or SKUs above or below the average carried by DC *i* in period *t*. Results in Column (3) shows that a DC carrying one more brand of product than another DC would source from 0.213 more units (*p-value<0.01*) on average. The average number of brands a DC carried in our sample was 31, suggesting that an average DC sourced from more than six other units, a little higher than the summary statistics in Table 1. In addition, controlling for the number of brands, DCs carrying a larger number of SKUs on average also sourced from more other units. Furthermore, the coefficients to the intertemporal variables suggest that DCs systematically increased/decreased sourcing relationships when product variety increased or decreased.

Column (4) adds regional dummies to control for unobserved time-invariant heterogeneity across the five regions. Results are similar to that in Column (3), suggesting a low level of regional heterogeneity. Finally, to account for time-invariant unobserved heterogeneity across DCs that could contribute to variations in sourcing complexity, we added DC fixed effects in Column (5). As expected, the coefficients to product-variety variables are economically less significant with DC fixed effects, but they are still statistically significant, suggesting a robust correlation between product variety and sourcing complexity. H1 is supported.

Sourcing Complexity and Coordination

Table 4 estimates stockouts at the DC–SKU level for non-hub DCs; *p-values* for all point estimates were less than 0.01 unless otherwise noted. Column (1) contains only product variety, control variables, and season dummies, which are included in all models. Coefficients to the control variables are pretty stable across different models. As expected, an SKU was more likely to experience a stockout if it was sold by a DC that carried a greater variety of products, in larger quantities, and with greater volatility. On the other hand, the likelihood of stockouts was lower when a DC held more inventories, forecasted more sales relative to actual sales, and received more quantity from other units during the last period, for the SKU.

Insert Table 4 about here

Column (2) adds sourcing complexity, or the number of units from which the DC received shipments for all SKUs, as an independent variable. The coefficients show that sourcing complexity is positively associated with product variety.

Column (3) decomposes the time-varying variable of sourcing complexity into two components: $Sourcing\ Complexity_{it} = \overline{Sourcing\ Complexity}_i + (Sourcing\ Complexity_{it} - \overline{Sourcing\ Complexity_i})$, where the first cross-sectional component represents the average number of units that DC *i* sourced from over the sample period, most likely based on quarterly sourcing plans made by the regional business units; the second intertemporal component represents the number of units DC *i* sourced from in period *t*, representing variation in the current period. The coefficients show that sourcing complexity had a significantly positive impact on stockouts. A marginal effect calculation shows that SKUs carried by a DC that on average sourced from one unit more than another DC during the sample period (while keeping all variables at their mean values) would experience an increase in stockout rate by one percentage point compared to SKUs carried by the other DC (the average stockout rate was 27 percent for each period according to Table 1). In addition, during a period when a DC sourced from one more unit compared to the average number of units it sourced from, the SKUs it carried would experience an increase in stockout rate by another one percentage point. After controlling for sourcing complexity, the impact of product variety on stockouts became less significant both economically and statistically. H2 is supported.

A potential issue of endogeneity is selection. That is, SKUs that were more likely to experience stockouts happened to be carried by DCs that had greater sourcing complexity. To account for unobserved SKU-specific factors that might influence stockouts, we added in Column (4) the average stockout rate for the focal SKU across all DCs during the current period.

The results show that, controlling for their average rates of stockouts, SKUs carried by DCs with greater sourcing complexity were more likely to experience stockouts. Coefficients in Columns (3) and (4) are similar, suggesting that SKU-specific heterogeneity does not drive the results. Similarly, column (5) adds region fixed effects to account for unobserved heterogeneity across regions that might influence stockouts. Results are similar to column (4), suggesting that the effect was not driven by regional effects. Finally, column (6) adds DC–SKU pair fixed effects to remove all the time-invariant differences across DC–SKU pairs. As expected, the coefficients are less significant but continue to support H2.

Mechanisms. As a supplementary analysis to test the mechanisms, we estimated the number of truck loads each DC received every period, the period-average of weekly shipments as percentages of weekly forecasted sales, and the volatility in such percentages, respectively. We found that DCs exposed to greater sourcing complexity received their shipments for each SKU (1) in a larger number of separate truck loads (*p-value=0.055*); (2) with a statistically insignificant increase in aggregate shipment quantity relative to forecasted sales (*p-value=0.103*), and (3) with a greater volatility from week to week (*p-value=0.055*). These results are not presented here due to space limits but are available from the authors upon requests; they imply that sourcing complexity is associated with more-frequent deliveries, with smaller quantities per load and more volatile quantity, which cost more time and effort to coordinate and can lead to increased stockouts.

Hubs and Coordination Bottlenecks

We proposed in the theory section that the extra burden of coordinating both incoming and outgoing shipments at a hub will worsen performance at downstream DCs. To establish this mechanism, we ran a few tests in Table 5 to see if hubs indeed experienced an extra coordination burden. We first compared the performance of hub vs. non-hub DCs. A marginal calculation based on the coefficients in Column (1) shows that, controlling for (inward) sourcing complexity, SKUs carried by hub DCs experienced a six-percentage-point higher stockout rate than SKUs carried by non-hub DCs (p-value<0.001); in addition, SKUs carried by a hub DC that supplied to one more unit than another hub DC experienced a 0.2 percentage-point higher stockout rate than SKUs carried by the other DC (p-value<0.001). As a supplementary analysis, we calculated the predicted probability of stockouts for every value of outward linkages (i.e., the number of downstream DCs the focal DC supplied to, which ranged from 0 to 73 in our sample) based on the coefficients in Column (1), keeping the value of "Being a hub (1,0)" at 1 and the values of all other variables at their mean. The results are not presented due to space limit but are available from the authors upon request; they show that the predicted probability of stockouts did increase more than linearly with the number of outward linkages, providing some evidence of a limit in coordination capacity at the hubs.

Insert Table 5 about here

We then explored one mechanism that augmented the coordination burden at hubs: demand shocks at downstream DCs. We focused on the subsample of hub-DCs; that is, DCs that forwarded shipments to other DCs during the sample period. We measured downstream demand shocks using forecast bias and weather shocks, respectively, at the downstream DCs. For the regression in Column (2) we defined a hub as experiencing a positive (negative) downstream demand shock if any of its downstream DCs experienced sales that were more than 20 percent above (below) forecasted sales. Coefficients in Column (2) suggest that a positive demand shock at a downstream DC was associated with an increased stockout rate at the upstream hub (*p*-

value<0.001), although the effect would be partially offset if any other downstream DCs had a negative demand shock (*p-value=0.026*).

Column (3) uses an alternative measure of demand shocks based on weather conditions at the downstream DCs. We collected daily weather data for years 2005 to 2011 from the National Centers for Environmental Information[2] and recorded the median temperature for each four-week period in each DC's state. We then defined a hub as experiencing a positive (negative) downstream demand shock if any of its downstream DCs had a median temperature for the current period that was more than five percent above (below) the median temperature at the downstream DC's location during the same four-week period over the past five years. Coefficients in column (3) suggest that weather shocks at individual downstream DCs in this case did not have a significant impact on the stockout rate at the upstream hub, but if some downstream DCs experienced a positive shock and some experienced a negative shock, then the upstream DC would experience an increased stockout rate (p-values = 0.026).

In sum, the results in Table 5 suggest that reduced coordination performance at a hub was partly due to demand shocks at its downstream DCs. These downstream shocks reverberated upstream and caused stockouts at the upstream hubs. This is consistent with a "bullwhip" effect in which inventory volatility increasingly swings in response to shifts in downstream demand as one moves further up the supply chain, due to information friction between the neighboring stages (Forrester, 1961). The negative coefficient to the interaction term between the positive and negative downstream demand shocks in Column (2) supports the notion that stockouts at a hub could be lower as centralization allows pooling of imperfectly correlated downstream demand shocks. However, the economic magnitude of this coefficient relative to the coefficient of

[2] www.ncdc.noaa.gov/cdo-web/#t=secondTabLink.

positive demand shocks suggests that the pooling effect was not sufficient to offset the impact of a vertical coordination burden (bullwhip effect). The positive coefficient to the interaction term in Column (3) suggests that the pooling effect, if any, was not sufficient enough to offset the impact of a horizontal coordination burden at the hubs.

Performance Consequence of Sourcing through Coordination Bottlenecks

Having established that hubs were exposed to a greater coordination burden than non-hubs, we now turn to the performance consequence for non-hub DCs when they sourced from a hub, the focus of H3. Our findings are presented in Table 6. A marginal-effect calculation based on Column (1) shows that non-hub DCs sourcing through a hub had a stockout rate four percent higher than non-hub DCs that did not source through a hub (*p-value<0.001*). Column (2) adds DC–SKU pair fixed effects; results are similar. H3 is supported.

Insert Table 6 about here

Reverse Causality

We have addressed various issues of endogeneity, such as selection and unobserved heterogeneity by using regional and season fixed effects, average stockout rate per SKU in each period, as well as DC–SKU fixed effects. However, the endogeneity issue of reverse causality remains: DCs expecting more stockouts sourced from more units and/or from hubs rather than the other way around. While the possibility of reverse causality cannot be fully ruled out in our context, we adopted a few methods to alleviate the concern.

To mitigate reverse causality between sourcing complexity and stockouts, we measured sourcing complexity using the number of units from which the DC sourced *all* SKUs, not just the focal SKU. Furthermore, in Column (3) of Table 6 we separate sourcing complexity into sourcing complexity for the focal SKU and sourcing complexity for other SKUs—the number of units from which the focal DC sourced all other SKUs, *excluding* units from which the DC also sourced the focal SKU. The coefficients in Column (3) suggest that both types of sourcing complexity had a significant and positive impact on stockouts (*p-value<0.001*). These findings not only support H2 but also suggest a "spillover" effect of sourcing complexity: When a DC experienced sourcing complexity for a particular SKU, other SKUs carried by the DC (sourced from different units) also experienced a higher stockout rate. We also ran similar regressions with lagged value of these measures; results were similar.

Similarly, to mitigate reverse causality between sourcing through hubs and stockouts, we measured the variable *Sourcing Through Hub* based on whether the focal DC sourced through a hub for *any* SKUs that it carried, not just the focal SKU. Furthermore, a DC expecting more stockouts for an SKU would be more likely to source the SKU through a hub because the hub could reallocate quantity across DCs to avoid stockouts. To address this possibility, in Column (3) of Table 6 we separate sourcing through hubs into sourcing through hubs for the focal SKU and sourcing through hubs for any non-focal SKUs. The variable *Sourcing Through Hub* for any non-focal SKUs only turned to 1 when the hub forwarded other SKUs to multiple downstream DCs but the focal SKU only to the focal DC; in this case, there would be no benefit of risk pooling with, or quantity reallocation from, other DCs for the focal SKU at the focal DC. The results in Column (3) suggest that both sourcing through a hub for the focal SKU and sourcing through a hub for a non-focal SKU had a significant and positive impact on the focal SKU's stockout rate (*p-value<0.001*). This not only supports H3, it also suggests a "spillover" effect of sourcing through hubs: When a DC sourced a particular SKU from a hub, other SKUs carried by the DC (and not sourced through a hub) experienced an increased stockout rate.

In sum, results in tables 3–6 suggest that (1) greater product variety was associated with greater sourcing complexity; (2) sourcing complexity reduced coordination performance (increased stockouts); (3) hubs experienced greater stockout rates, partly due to coordination with downstream DCs; and (4) sourcing through hubs increased downstream DCs' stockout rate. These findings support our hypotheses.

CONCLUSIONS AND DISCUSSION AND CONCLUSION

This paper examined the coordination burden for firms that pursue variety as their main product strategy. In particular, we focused on complexity in internal sourcing relationships as a mechanism that may give rise to additional coordination burden. Our results, based on DCs within a major soft drink bottling company, confirmed that an increase in variety was associated with an increase in sourcing complexity, and that an increase in sourcing complexity was associated with worsened coordination performance. In addition, an intermediary hub in the sourcing network may become a bottleneck and negatively impact the performance of the units it coordinates.

Recent studies have identified task complexity as a source of coordination cost that results in diseconomies of scope (Zhou, 2011), but the micro-level organization processes that cultivate complexity have remained unexplored. This paper focused specifically on the tension between scale and scope economies to suggest that the pursuit of scale economy generates production rigidity, while pursuing downstream synergies through cross-selling creates organizational interdependencies and complexity. We also empirically explored product line extension—the purest form of firm scope expansion—to demonstrate that complexity-induced coordination burden may, indeed, reduce economies of scope.

These results also extend recent attempts to conceptualize the locus and limitation of coordination in complex task systems. As complexity increases, these loci of coordination turn into organizational bottlenecks due to limits on their coordination capacity. Organizations face a tradeoff in designing these hubs, which might reduce complexity in the overall network but become a bottleneck themselves due to local congestion. This further illustrates the point that economies of scope "may decline not because of exogenous opportunity constraints but because of the rising costs of coordinating interdependencies" (Zhou 2011: 625).

Our findings also have implications for the literature on product variety, a pivotal competition strategy (Adner & Levinthal, 2001; Caves & Ghemawat, 1992; Porter, 1985b). In particular, our findings illuminate the inherent tradeoffs between the strategic benefits and organizational costs of a variety proliferation strategy. By demonstrating the coordination burden arising from sourcing complexity associated with great variety, the paper echoes prior strategy research on the potential downside of product variety (Barnett & Freeman, 2001; Cottrell & Nault, 2004; Sorenson, 2000).

The paper has a couple of limitations that create opportunities for future research. First, it analyzes only one element of performance—coordination—instead of the overall net benefit of a product proliferation strategy. This was mainly due to data constraints: We do not have market share or profit information at the DC or SKU level. However, our objective was not to evaluate the net benefit or cost of a product proliferation strategy. Rather, it was to pinpoint a particular mechanism through which product variety might undermine operational efficiency. Secondly, our sample is limited to one company. Studying coordination across hundreds of DCs and thousands of SKUs within a single company offers several benefits. It eliminates unobserved firm heterogeneity that might confound coordination. In addition, it allows us to extract detailed operations data such as sales, inventory, forecasts, shipments, and stockouts, all of which are critical for our analysis of coordination. Still, future studies can explore across-firm heterogeneity when detailed operations data become available for a larger sample of firms.

In conclusion, this paper highlighted the intricate and important trade-off between scale and scope economies within firms that pursue variety as their primary product strategy. It provided empirical evidence that complex interdependencies between adjacent stages of the value chain, and the coordination burden that follows such complexity, may place significant limits on the implementation of a variety proliferation strategy and the corresponding organizational design.

References

Adner R, Levinthal D. 2001. Demand heterogeneity and technology evolution: Implications for product and process innovation. *Management Science* **47**(5): 611-662.

Agarwal R, Helfat CE. 2009. Strategic Renewal of Organizations. *Organization Science* **20**(2): 281-293.

Anderson ET, Fitzsimons GJ, Simester D. 2006. Measuring and Mitigating the Costs of Stockouts. *Management Science* **52**(11): 1751-1763.

Anupindi R, Chopra S, Deshmukh SD, Mieghem JAV, Zemel E. 2011. *Managing Business Process Flows* (3rd ed.). Prentice Hall.

Arrow KJ. 1974. *The Limits of Organization*. W. W. Norton & Company: New York.

Baldwin CY. 2008. Modularity, transactions, and the boundaries of firms: A synthesis. *Industrial and Corporate Change* **17**(1): 155-196.

Baldwin CY. 2014. Bottlenecks, Modules and Dynamic Architectural Capabilities. *HBR Working Paper 15-028*.

Baldwin CY, Clark KB. 2000. *Design Rules (Volume 1): The Power of Modularity*. MIT Press: Cambridge, MA.

Barnett WP, Freeman J. 2001. Too much of a good thing? Product proliferation and organizational failure. *Organization Science* **12**(5): 539-558.

Becker GS, Murphy KM. 1992. The division of labor, coordination costs, and knowledge. *The Quarterly Journal of Economics* **107**(4): 1137-1160.

Burton RM, Obel B. 1980. A computer simulation test of the M-form hypothesis. *Administrative Science Quarterly* **25**(3): 457-466.

Cachon G, Terwiesch C. 2012. *Matching Supply with Demand: An Introduction to Operations Management*. McGraw-Hill/Irwin.

Caves RE, Ghemawat P. 1992. Identifying mobility barriers. *Strategic Management Journal* **13**(1): 1-12.

Chandler ADJ. 1962. *Strategy and Structure: Chapters in the History of the American Industrial Enterprise*. The MIT Press: Cambridge, Massachusetts.

Closs DJ, Nyaga GN, Voss MD. 2010. The differential impact of product complexity, inventory level, and configuration capacity on unit and order fill rate performance. *Journal of Operations Management* **28**(1): 47-57.

Coca Cola Press Release. 2010. The Coca-Cola Company finalizes transaction with Coca-Cola enterprises.

Cottrell T, Nault BR. 2004. Product variety and firm survival in the microcomputer software industry. *Strategic Management Journal* **25**(10): 1005-1025.

Ethiraj SK, Levinthal D, Roy RR. 2008. The dual role of modularity: Innovation and imitation. *Management Science* **54**(5): 939-955.

Fisher ML. 1997. What is the right supply chain for your product? *Harvard Business Review* **75**(March-April): 105-117.

Fisher ML, Ittner CD. 1999. The impact of product variety on automobile assembly operations: Empirical evidence and simulation analysis. *Management Science* **45**(6): 771-786.

Forrester JW. 1961. *Industrial Dynamics*. MIT Press: Cambridge, Massachusetts.

Fosfuri A, Giarratana MS. 2009. Masters of war: Rivals' product innovation and new advertising in mature product markets. *Management Science* **55**(2): 181-191.

Gulati R, Puranam P. 2009. Renewal Through Reorganization: The Value of Inconsistencies Between Formal and Informal Organization. *Organization Science* **20**(2): 422-440.

Henderson R, Clark K. 1990. Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms. *Administrative Science Quarterly* **35**(1): 9-30.

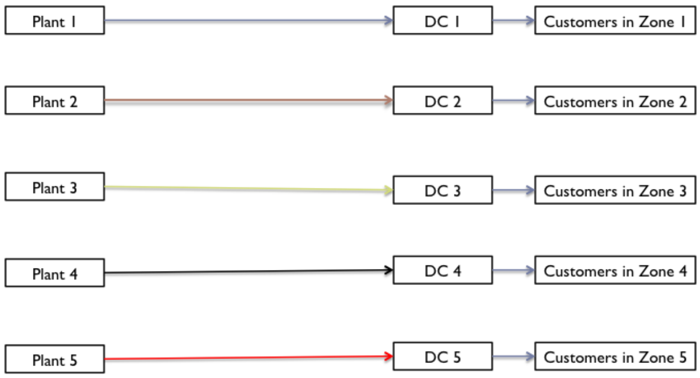
Hoetker G. 2006. Do modular products lead to modular organizations? *Strategic Management Journal* **27**(6): 501-518.

Ichniowski C, Shaw K, Prennushi G. 1997. The effects of human resource management practices on productivity: A study of steel finishing lines. *American Economic Review* **87**(3): 291-313.
Kauffman SA. 1993. *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press: USA.
Lenox MJ, Rockart SF, Lewin AY. 2006. Interdependency, competition, and the distribution of firm and industry profits. *Management Science* **52**(5): 757-772.
Lenox MJ, Rockart SF, Lewin AY. 2007. Interdependency, competition, and industry dynamics. *Management Science* **53**(4): 599-615.
Lenox MJ, Rockart SF, Lewin AY. 2008. Does Interdependency Affect Industry Profitability? An Empirical Test. *Working paper*.
Levinthal DA. 1997. Adaptation on Rugged Landscapes. *Management Science* **43**(7): 934-950.
MacDuffie JP, Sethuraman K, Fisher ML. 1996. Product variety and manufacturing performance: Evidence from the international automotive assembly plant Study. *Management Science* **42**(3): 350-369.
Milgrom P, Roberts J. 1990. The Economics of Modern Manufacturing: Technology, Strategy, and Organization. *The American Economic Review* **80**(3): 511-528.
Milgrom P, Roberts J. 1995. Complementarities and fit strategy, structure, and organizational change in manufacturing. *Journal of Accounting and Economics* **19**: 179-208.
Miller J, Vollmann TE. 1985. The hidden factory. *Harvard business review* **63**(5): 142-150.
Musalem A, Olivares M, Bradlow ET, Terwiesch C, Corsten D. 2010. Structural estimation of the effect of out-of-stocks. *Management Science* **56**(7): 1180-1197.
Porter ME. 1985a. *Competitive Advantage: Creating and Sustaining Superior Performance*. Free Press: New York.
Porter ME. 1985b. *Competitive Advantage: Creating and Sustaining Superior Performance*. Free Press: New York.
Porter ME. 1997. *COMPETITIVE STRATEGY*. MCB UP Ltd: New York.
Puranam P. 2001. A knowledge based perspective on the internal boundaries of the firm. The case of technology grafting acquisitions. *Working Paper*.
Puranam P, Raveendran M, Knudsen T. 2012. ORGANIZATION DESIGN: THE EPISTEMIC INTERDEPENDENCE PERSPECTIVE. *Academy of Management Review* **37**(3): 419-440.
Ramdas K. 2003. Managing product variety: An integrative review and research directions. *Production and Operations Management* **12**(1): 79-101.
Rivkin JW. 2000a. Imitation of Complex Strategies. *Management Science* **46**(6): 824-844.
Rivkin JW. 2000b. Imitation of Complex Strategies. *Management Science* **46**(6): 824.
Rivkin JW, Siggelkow N. 2007. Patterned interactions in complex systems: Implications for exploration. *Management Science* **53**(7): 1068-1085.
Schilling MA, Steensma HK. 2001. THE USE OF MODULAR ORGANIZATIONAL FORMS: AN INDUSTRY-LEVEL ANALYSIS. *Academy of Management Journal* **44**(6): 1149-1168.
Siggelkow N. 2001. Change in the Presence of Fit: The Rise, the Fall, and the Renaissance of Liz Claiborne. *The Academy of Management Journal* **44**(4): 838-857.
Siggelkow N. 2002. Evolution toward Fit. *Administrative Science Quarterly* **47**(1): 125-159.
Simon HA. 1955. A behavioral model of rational choice. *The Quarterly Journal of Economics* **69**(1): 99-118.
Simon HA. 1962. The architecture of complexity. *Proceedings of the American Philosophical Society* **106**: 467-482.
Sorenson O. 2000. Letting the market work for you: an evolutionary perspective on product strategy. *Strategic Management Journal* **21**(5): 577-592.
Sturgeon TJ. 2002. Modular production networks: a new American model of industrial organization. *Industrial and Corporate Change* **11**(3): 451-496.

Sutherland JW. 1980. A quasi-empirical mapping of optimal scale of enterprise. *Management Science* **26**(10): 963-981.
Teece DJ, Pisano G, Shuen A. 1997. Dynamic capabilities and strategic management. *Strategic Management Journal* **18**(7): 509-533.
Yoffie DB, Kim R. 2011. Cola wars continue: Coke and Pepsi in 2010. *Harvard Business School Case: 9-711-462*.
Zhou YM. 2011. Synergy, coordination costs, and diversification choices. *Strategic Management Journal* **32**(6): 624-639.
Zhou YM. 2013. Designing for complexity: Using divisions and hierarchy to manage complex tasks. *Organization Science* **24**(2): 339-355.

Figure 1 Product Variety and Inter-unit Sourcing

a. Order-based sourcing for firms with low-scale-economy products or a small product variety



b. Variety-based sourcing for firms with high-scale-economy products and a large product variety

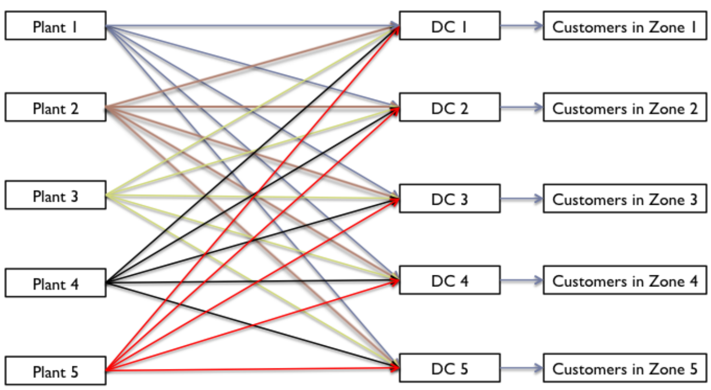
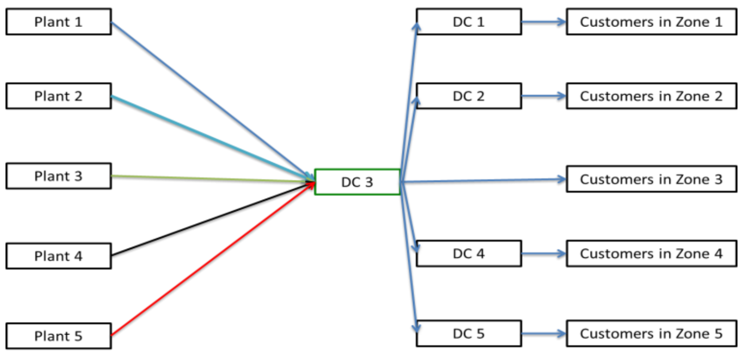


Figure 2 Sourcing through Hubs

a. Variety-based sourcing through a centralized hub



b. Variety-based sourcing through an intermediary hub

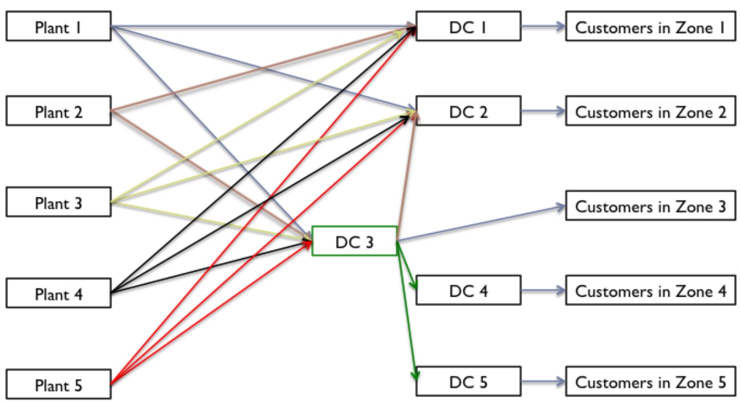


Table 1 Summary Statistics for Key Variables at DC-SKU Level

		Definition	Mean	SD	Min	Max
(1)	Stockout	The DC-SKU experienced at least one stockout during the period (1,0)	0.27	0.44	0	1
(2)	Product Variety –Brands	Number of brands of products carried by the DC	31.26	4.14	20	52
(3)	Product Variety –SKUs	Number of SKUs carried by the DC	568.86	87.38	172	916
(4)	Sourcing Complexity	Number of units from which the DC received shipments for all SKUs	4.30	3.07	0	17
(5)	Being A Hub (1,0)	The DC shipped to at least one other DC during the period (1,0)	0.12	0.32	0	1
(6)	Sourcing Through Hub (1,0)	The DC received shipments from a hub for any SKU during the period (1,0)	0.53	0.50	0	1
(7)	Sales	Log (Sales quantity in standard cases)	4.18	2.30	0	7.60
(8)	Sales Volatility	Coefficient of variation (CV) in sales, or the standard deviation of weekly sales in each period normalized by its mean	0.17	0.20	0	0.87
(9)	Beginning Inventory	Inventory at the beginning of the period as percentage of actual sales	27.51	34.20	0	152.25
(10)	Demand Forecast	Forecasted sales as percentage of actual sales	1.04	0.39	0.21	2
(11)	Shipment Quantity, lagged	Log (Shipment quantity in standard cases) during last period	2.38	3.16	0	12.32
(12)	Average SKU Stockout	Average stockout rate of the focal SKU across all DCs during current period	0.24	0.13	0	1

Table 2 Correlation Matrix

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1)	Stockout	1.00											
(2)	Product Variety –Brands	0.06	1.00										
(3)	Product Variety –SKUs	0.05	0.49	1.00									
(4)	Sourcing Complexity	0.13	0.22	0.43	1.00								
(5)	Being A Hub (1,0)	0.15	0.10	0.20	0.49	1.00							
(6)	Sourcing Through Hub (1,0)	0.08	-0.05	-0.11	-0.09	-0.17	1.00						
(7)	Sales	0.21	0.001	0.04	0.18	0.25	0.36	1.00					
(8)	Sales Volatility	0.12	-0.03	-0.05	-0.08	-0.06	-0.01	-0.12	1.00				
(9)	Beginning Inventory	0.002	0.02	0.05	0.10	0.19	-0.23	-0.01	0.31	1.00			
(10)	Demand Forecast	-0.11	0.02	0.01	0.004	0.01	-0.08	-0.08	0.05	0.04	1.00		
(11)	Shipment Quantity, lagged	0.05	-0.02	-0.05	-0.002	-0.11	0.63	0.50	-0.11	-0.23	0.03	1.00	
(12)	Average SKU Stockout	0.29	-0.02	-0.07	-0.03	0.04	0.19	0.35	0.14	0.06	-0.06	0.21	1.00

Table 3 Product Variety and Sourcing Complexity

DV=Sourcing Complexity	(1)		(2)		(3)		(4)		(5)	
Product Variety										
- Brands, total			0.233	(0.000)						
			[0.032]							
- Brands, average					0.213	(0.000)	0.236	(0.000)		
					[0.032]		[0.039]			
- Brands, above (below) average					0.195	(0.004)	0.191	(0.005)	0.036	(0.069)
					[0.067]		[0.068]		[0.020]	
- SKUs, orthogonalized, total			1.050	(0.000)						
			[0.153]							
- SKUs, orthogonalized, average					0.827	(0.000)	0.869	(0.000)		
					[0.137]		[0.245]			
- SKUs, orthogonalized, above (below) average,					0.467	(0.001)	0.425	(0.001)	0.072	(0.027)
					[0.114]		[0.123]		[0.032]	
Sales	1.247	(0.000)	1.524	(0.000)	1.505	(0.000)	1.518	(0.000)	0.457	(0.004)
	[0.265]		[0.266]		[0.262]		[0.248]		[0.158]	
Sales Volatility	-19.981	(0.000)	-7.098	(0.115)	-6.829	(0.124)	-7.042	(0.093)	-0.771	(0.322)
	[4.840]		[4.490]		[4.424]		[4.177]		[0.776]	
Beginning Inventory	0.119	(0.000)	0.097	(0.000)	0.095	(0.000)	0.087	(0.000)	-0.005	(0.481)
	[0.024]		[0.021]		[0.021]		[0.021]		[0.007]	
Demand Forecast	-0.840	(0.497)	-0.379	(1.073)	-0.221	(0.820)	0.013	(0.989)	0.028	(0.915)
	[1.235]		[1.073]		[0.974]		[0.930]		[0.262]	
Season FE	Yes		Yes		Yes		Yes		Yes	
Region FE	No		No		No		Yes		No	
DC FE	No		No		No		No		Yes	
Observations	3,430		3,430		3,430		3,430		3,430	
Adjusted R^2	0.40		0.47		0.47		0.50		0.91	

Robust standard errors clustered at DC level are included in square brackets for all columns. *p*-values are in parentheses. All tests are two-tailed.

Table 4 Sourcing Complexity and Stockouts at Non-hub DCs

DV= Stockout (1,0)	(1)	(2)	(3)	(4)	(5)	(6)
Sourcing Complexity						
- For all SKUs, total		0.055				
		[0.011]				
- For all SKUs, average			0.052	0.075	0.085	
			[0.012]	[0.013]	[0.015]	
- For all SKUs, above (below) average			0.067	0.070	0.056	0.011
			[0.021]	[0.024]	[0.020]	[0.004]
Product Variety - Brands	0.031	0.018	0.018	0.019	0.017	0.018
	[0.008]	[0.007]	[0.007]	[0.007]	[0.006]	[0.002]
Product Variety - SKUs, orthogonalized	0.036[a]	-0.029[b]	-0.026[c]	-0.012[d]	0.040[e]	-0.025[f]
	[0.026]	[0.025]	[0.025]	[0.027]	[0.040]	[0.010]
Sales	0.264	0.260	0.260	0.191	0.190	0.307
	[0.011]	[0.010]	[0.010]	[0.011]	[0.011]	[0.004]
Sales Volatility	2.281	2.306	2.306	1.972	1.976	2.241
	[0.044]	[0.044]	[0.044]	[0.043]	[0.042]	[0.018]
Beginning Inventory	-0.005	-0.005	-0.005	-0.006	-0.006	-0.007
	[0.000]	[0.0003]	[0.000]	[0.000]	[0.000]	[0.000]
Demand Forecast	-0.539	-0.547	-0.547	-0.515	-0.519	-0.543
	[0.024]	[0.023]	[0.023]	[0.023]	[0.022]	[0.008]
Shipment Quantity, lagged	-0.031	-0.033	-0.033	-0.043	-0.044	-0.053
	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]	[0.002]
Average SKU Stockout				4.435	4.432	5.485
				[0.053]	[0.053]	[0.033]
Season FE	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	No	No	No	No	Yes	No
DC-SKU FE	No	No	No	No	No	Yes
Observations	895,210	895,210	895,210	895,210	895,210	884,818
Pseudo R^2	0.08	0.08	0.08	0.12	0.12	0.15

Robust standard errors clustered at DC level are included in square brackets for all columns. *p*-values for all point estimates are less than 0.01 unless noted otherwise. All tests are two-tailed. [a] *p*-value=0.168. [b] *p*-value=0.240. [c] *p*-value=0.289. [d] *p*-value=0.658. [e] *p*-value=0.315. [f] *p*-value=0.016. [g] *p*-value=0.459. [h] *p*-value=0.916. [i] *p*-value=0.043. [i] *p*-value=0.572.

Table 5 Coordination Burden at Hubs

DV= Stockout (1,0)	Hub and non-hub DCs: relative performance (1)	Hub DCs: under- (over-) forecasted demand at downstream DCs (2)	Hub-DCs: heat (cold) Shock at downstream DCs (3)
Being a Hub (1,0)	0.342 (0.000)		
	[0.095]		
Number of DCs the hub supplied to	0.011 (0.012)	0.002 (0.617)	0.003 (0.561)
	[0.004]	[0.006]	[0.005]
Demand shock at receiving DCs			
Positive demand shock at any downstream DC (1, 0)		0.151 (0.000)	-0.029 (0.637)
		[0.030]	[0.061]
Negative demand shock at any downstream DC (1, 0)		0.025 (0.374)	-0.026 (0. 540)
		[0.028]	[0.043]
*Positive demand shock at any downstream DC (1, 0)*Negative demand shock at any downstream DC (1, 0)*		-0.097 (0.026)	0.210 (0.026)
		[0.040]	[0.094]
Sourcing Complexity - for all SKUs	0.038 (0.004)	0.015 (0.490)	0.014 (0.513)
	[0.013]	[0.021]	[0.022]
Product Variety - Brands	0.017 (0.009)	0.054 (0.047)	0.054 (0.043)
	[0.006]	[0.027]	[0.027]
Product Variety - SKUs, orthogonalized	-0.026 (0.473)	-0.101 (0.353)	-0.101 (0.358)
	[0.037]	[0.108]	[0.111]
Sales	0.253 (0.000)	0.239 (0.000)	0.238 (0.000)
	[0.009]	[0.016]	[0.016]
Sales Volatility	2.228 (0.000)	1.314 (0.000)	1.334 (0.000)
	[0.041]	[0.093]	[0.094]
Beginning Inventory	-0.005 (0.000)	-0.004 (0.000)	-0.004 (0.000)
	[0.003]	[0.003]	[0.004]
Demand Forecast	-0.528 (0.000)	-0.271 (0.000)	-0.312 (0.000)
	[0.022]	[0.044]	[0.045]
Shipment Quantity, lagged	-0.020 (0.000)	0.008 (0.133)	0.008 (0.121)
	[0.005]	[0.006]	[0.006]
Season FE	Yes	Yes	Yes
Region FE	Yes	Yes	Yes
Observations	1,011,792	116,582	116,582
Pseudo/Adjusted R^2	0.09	0.06	0.06

Robust standard errors clustered at DC level are included in square brackets for all columns. *p*-values are in parentheses. All tests are two-tailed.

Table 6 Performance of Non-hub DCs When They Source through Hubs

DV=Stockout (1,0)	**(1)**	**(2)**	**(3)**
Sourcing Through Hub (1,0)			
For any SKU	0.261	0.433	
	[0.025]	[0.010]	
For focal SKU			0.193
			[0.046]
For any non-focal SKU			0.301
			[0.049]
Sourcing Complexity			
For all SKU	0.080	0.010	
	[0.013]	[0.004]	
For focal SKU			0.079
			[0.013]
For any non-focal SKU			0.080
			[0.015]
Product Variety - Brands	0.018	0.018	0.018
	[0.007]	[0.002]	[0.007]
Product Variety - SKUs, orthogonalized	0.041[a]	-0.023[b]	0.041[d]
	[0.041]	[0.010]	[0.041]
Sales	0.186	0.250	0.189
	[0.012]	[0.004]	[0.011]
Sales Volatility	1.946	2.257	1.936
	[0.042]	[0.018]	[0.042]
Beginning Inventory	-0.005	-0.005	-0.005
	[0.0003]	[0.0001]	[0.0003]
Demand Forecast	-0.490	-0.521	-0.491
	[0.021]	[0.008]	[0.021]
Shipment Quantity, lagged	-0.063	-0.044	-0.061
	[0.005]	[0.002]	[0.005]
Average SKU Stockout	4.412	5.486	4.426
	[0.053]	[0.033]	[0.053]
Season FE	Yes	Yes	Yes
Region FE	Yes	Yes	Yes
DC-SKU FE	No	Yes	No
Observations	895,210	884,818	895,210
Pseudo/Adjusted R^2	0.13	0.11	0.13

Robust standard errors clustered at DC level are included in square brackets for all columns. *p*-values for all point estimates are less than 0.01 unless noted otherwise. All tests are two-tailed. [a] *p*-value=0.307. [b] *p*-value=0.024. [c] *p*-value=0.231. [d] *p*-value=0.313.