

Biçer, Isık; Lücker, Florian; Boyacı, Tamer

**Working Paper**

## Beyond retail stores: Managing product proliferation along the supply chain

ESMT Working Paper, No. 19-02 (R1)

**Provided in Cooperation with:**

ESMT European School of Management and Technology, Berlin

*Suggested Citation:* Biçer, Isık; Lücker, Florian; Boyacı, Tamer (2020) : Beyond retail stores: Managing product proliferation along the supply chain, ESMT Working Paper, No. 19-02 (R1), European School of Management and Technology (ESMT), Berlin,  
<https://nbn-resolving.de/urn:nbn:de:101:1-2020021113491256101164>

This Version is available at:

<https://hdl.handle.net/10419/214163>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

February 10, 2020

**ESMT Working Paper 19-02 (R1)**

# Beyond retail stores: Managing product proliferation along the supply chain

Işık Biçer, Schulich School of Business, York University

Florian Lücker, Cass Business School, City, University of London

Tamer Boyacı, ESMT Berlin

Revised version

Copyright 2020 by ESMT European School of Management and Technology GmbH, Berlin, Germany, <https://esmt.berlin/>.

All rights reserved. This document may be distributed for free – electronically or in print – in the same formats as it is available on the website of the ESMT ([www.esmt.org](http://www.esmt.org)) for non-commercial purposes. It is not allowed to produce any derivatives of it without the written permission of ESMT.

Find more ESMT working papers at [ESMT faculty publications](#), [SSRN](#), [RePEc](#), and [EconStor](#).

# Beyond Retail Stores: Managing Product Proliferation Along the Supply Chain

Işık Biçer

Schulich School of Business, York University, 111 Ian MacDonald Blvd, M3J1P3, Toronto, ON, Canada,  
bicer@schulich.yorku.ca,

Florian Lücker

Cass Business School, City, University of London, London EC1Y, UK, florian.lucker@city.ac.uk,

Tamer Boyacı

ESMT Berlin, Schlossplatz 1, 10178, Berlin, Germany, tamer.boyaci@esmt.org,

Product proliferation occurs in supply chains to produce diverse products from a limited variety of raw materials. In such a setting, manufacturers can establish market responsiveness and/or cost efficiency in alternative ways. Delaying the point of the proliferation helps manufacturers improve the responsiveness by postponing the ordering decisions of the final products until partial or full resolution of demand uncertainty. This strategy can be implemented in two different approaches: (1) redesigning the operations such that the point of proliferation is swapped with a downstream operation and (2) reducing the lead times. To establish cost efficiency, manufacturers can systematically reduce the costs of all operations or postpone the high-cost operations. We consider a multi-echelon and multi-product newsvendor problem with demand forecast evolution to analyze the value of each operational lever of the responsiveness and the efficiency. We use the multiplicative martingale model of forecast evolution to characterize the demand-updating process, and develop a dynamic optimization model to determine the optimal order quantities at different echelons. We show that reducing the lead time of a downstream operation is more beneficial to manufacturers than reducing the lead time of an upstream operation by the same amount, whereas reducing the costs of upstream operations is more favorable than reducing the costs of downstream operations. We also indicate that delaying the proliferation may cause a loss of profit even if it can be achieved with no additional cost. We develop a decision typology that shows effective operational strategies depending on product/market characteristics and process flexibility.

*Key words:* Product proliferation; lead-time reduction; process redesign; delayed differentiation.

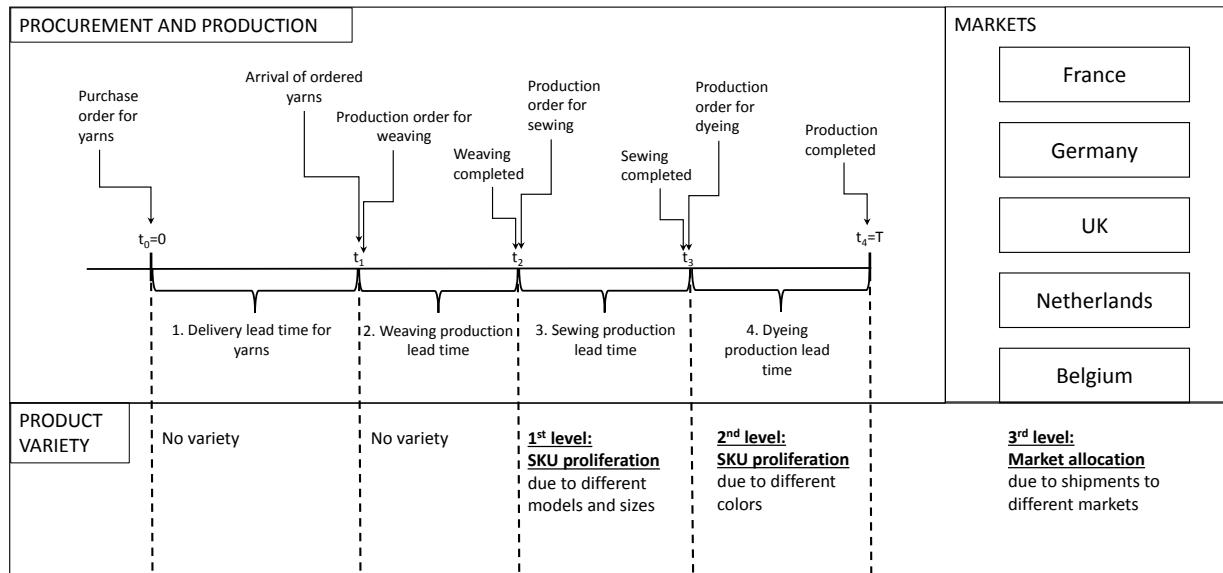
---

## 1. Introduction

Digital transformation in the retail industry (e.g., omni-channel retailing, recommendation systems and user-oriented product development using social media) has led to an increase in demand for niche items in almost all product categories (Brynjolfsson et al. 2011). Retailers now carry more diverse product portfolios than in past decades in both online and physical stores. The expansion of product portfolios has a negative impact on supply-demand mismatches in the retail industry (Rajagopalan 2013). Arguably, the challenges associated with diverse product portfolios are not only limited to downstream sales channels (retailers, online channels), but start with upstream operations (Atalı and Özer 2012). In fact, it is not uncommon that manufacturers attempt to fulfill customer demand for broad product lines by *using the same upstream resources and differentiating products over time* as they get close to markets. This strategy helps them to benefit from economies of scale for upstream resources and to postpone product differentiation until acquisition of more accurate market demand forecasts.

Fashion apparel is perhaps the most celebrated industry where product proliferation is prominent and has profound impact on profitability. Figure 1 depicts the supply chain structure for a typical fashion-apparel manufacturer serving multiple markets. Global manufacturers like Zara, H&M, and Uniqlo sell a variety of clothes in each selling season, which are produced by the same textile but sewn and colored differently. After a design team responsible for a product line develops new designs to be sold in the next season, yarns selected by the design team are ordered. Production occurs sequentially involving the weaving, sewing, and dyeing processes. First, yarns are transformed into textile by the weaving process. Then, the textile is sewn into different models and sizes. Finally, the items are dyed into different colors to complete the production. Product proliferation occurs sequentially, in three stages. The first occurs after the sewing process, the second occurs after the dyeing process, and the third occurs when the products are labeled and shipped.

The examples of product proliferation are not limited to the fashion-apparel industry. In the consumer packaged-goods industry, a limited variety of ingredients (e.g., milk, fruits, and yogurt



**Figure 1** Product proliferation in the fashion-apparel industry

bacteria) are used to make a wide variety of products in bulk (e.g., raw milk, raw yogurt, yogurt drink, flavoured milk and yogurt). The products are filled in different-sized containers and then sold in the market. Therefore, product proliferation first occurs during production, and again during the filling process. It is also common in the process industry, where manufacturers differentiate products along the supply chain to fulfill the customer demand for alternative “recipes” (each recipe corresponds to a certain product specification that determines product performance along different dimensions such as thermal resistance, elasticity, etc.). In Figure 2, we present an example that we observe in a leading global manufacturer of composites used mainly by tyre producers in the automotive industry. Demand for the manufacturer is both volatile and seasonal due to the seasonality of the tyre sales. The manufacturer first processes some chemicals with polypropylene to produce polymer materials. These materials are first shaped through a twisting operation and then go through a second-level fabrication process of weaving. Finally, the weaved products enter a chemical blending process in which they are dipped to chemical liquids to bring the products to the right level of thermal resistance and elasticity. Although the variety of polymer materials are limited, there exists a high variety of end products due to product proliferation in the last three stages.

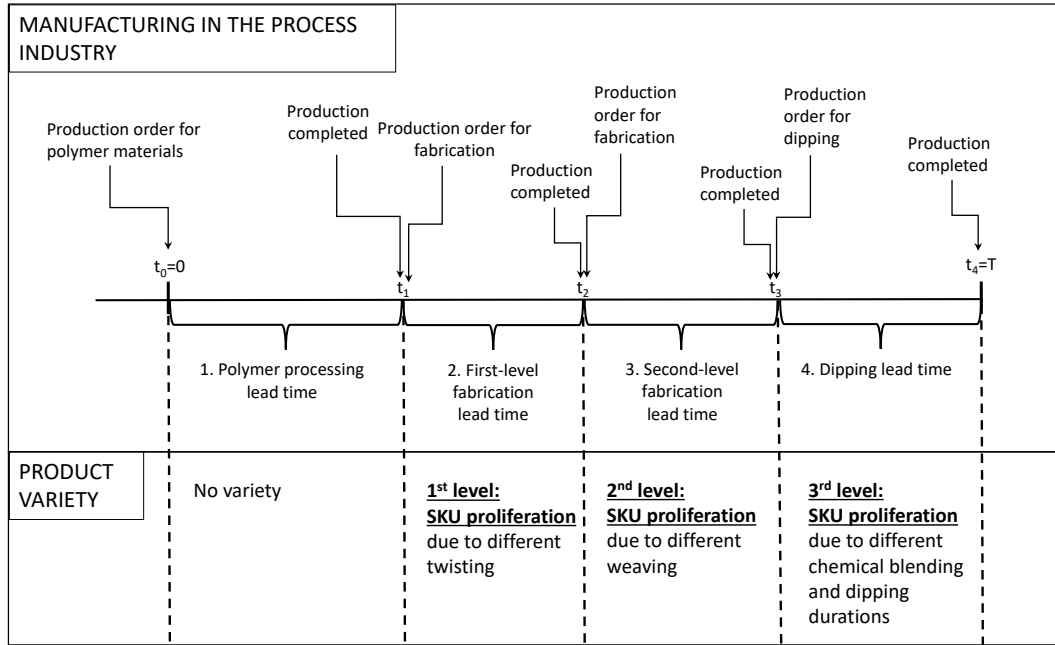


Figure 2 Product proliferation in our process industry example

Manufacturers operating in such settings are often exposed to high demand uncertainty for upstream production orders. As the production moves forward, demand uncertainty is partially resolved due to additional valuable demand information collected from the market. For downstream production orders, however, manufacturers are exposed to high product variety. Trading off the cost efficiency against the operational responsiveness, Fisher (1997) indicates that physically efficient supply chains are better aligned with the products with low demand uncertainty, whereas market-responsive supply chains are better aligned with the products with high demand uncertainty. Due to evolutionary risk structure in a product proliferation model, the utilization of both market-responsive and cost-efficient strategies may improve the profits depending on the supply chain structure along with the cost, the demand, and the lead-time parameters.

*Delaying differentiation* is an effective strategy for improving responsiveness in supply chains in which product proliferation occurs (the terms “differentiation” and “proliferation” are used interchangeably). It enables manufacturers to take advantage of inventory pooling at upstream echelons, while ensuring that the proliferation at downstream occurs with more accurate demand information. There are two practical approaches to operationalizing delayed differentiation, both of which have

been widely popularized by their implementation in the fashion apparel industry. The first approach is to *redesign the processes* so that the operations that cause proliferation are deferred to a later stage in the supply chain. Benetton, the Italian clothing company, is the first firm which successfully implemented this approach and reversed the order of dyeing and knitting operations (Heskett and Signorelli 1989, Lee and Tang 1997). Traditionally Benetton spun and dyed the yarns first and then knitted the colored yarns. In 1972, the company began dyeing clothes rather than yarns to postpone the costly dyeing operation. This allowed Benetton to postpone product differentiation until it could observe accurate market demand information, leading to higher profits due to the decrease in supply-demand mismatches. Given the success of this approach, many other companies followed the Benetton's lead (Parsons and Graves 2005, Viswanathan and Allampalli 2012, Kouvelis and Tian 2014).<sup>1</sup>

The second approach is to *reduce lead times* for each operation in the supply chain. Zara, the Spanish fashion apparel company, followed this strategy and became the market leader in 2008 (Ghemawat and Nueno 2006). Demand forecasts are often plagued with high uncertainty when lead times tend to be long. Reducing lead times allows manufacturers to postpone the point of proliferation and actual ordering decisions closer to market demand, making it possible to place production orders based on more accurate demand forecasts. This in turn leads to a decrease in supply-demand mismatches (Caro and Martínez-de Albéniz 2015).

To establish cost efficiency, manufacturers can systematically reduce the costs of all the operations along the supply chain or postpone the high-cost operations to a later stage. The former helps them reduce the unit product cost, whereas the latter makes it possible to avoid unnecessary overutilization of expensive resources. The value of each operational lever of the responsiveness and

<sup>1</sup> We remark that process redesign does not necessarily require swapping of operations; it may also be achieved by changing the way operations are performed (and associated costs). In the case of our leading process industry example, it is possible to meet the technical specifications requested by a customer by changing either technical grades used during the fabrication process or the chemical recipes used during the blending process. The latter enables the postponement of the point of proliferation, but increases production costs.

the efficiency depends on the supply chain structure, the cost and the lead-time values of each operation along the supply chain. Our objective in this paper is to quantify their costs and benefits. To this end, we consider a multi-echelon and multi-product newsvendor model with demand forecast evolution. We make three important contributions to the extant literature. First, from a modeling perspective, we develop an analytical framework for dynamically optimizing inventory/ordering quantities in a multi-echelon and multi-product newsvendor setting with demand forecast evolution. This framework extends existing inventory models in the literature that incorporates forecast evolution (Wang et al. 2012, Biçer and Seifert 2017) to multi-product and multi-echelon settings. Our framework takes the supply chain structure along with lead times and cost values for each echelon as inputs, incorporates the evolution of demand forecasts using the multiplicative martingale model of forecast evolution (m-MMFE), and optimizes the ordering decisions at each echelon. We characterize the optimal strategy and investigate the effects on the profits of salient parameters—costs, lead times, proliferation points.

Second, utilizing this framework, we analytically demonstrate the critical impact of cost changes, lead-time reduction, and postponement points on optimal inventory levels and consequent profits. For example, we establish that reducing the lead time of a downstream operation is more beneficial to manufacturers than reducing the lead time of an upstream operation by the same amount, whereas reducing the costs of upstream operations is more favorable than reducing the costs for downstream operations. The value of delaying the proliferation increases when the downstream operations (post-proliferation) are more costly. We substantiate our descriptive insights with a comprehensive numerical study based on Markovian sampling and find that delaying the proliferation may cause a loss of profit for manufacturers if its implementation requires swapping a high-cost downstream operation with a low-cost upstream operation. This result is counter-intuitive to the conventional wisdom that the postponement strategy would always help reducing the cost of mismatches between supply and demand (please see Zinn (2019) for a historical review of the evolution of the postponement research).



Third, we translate the descriptive results into prescriptive insights for practicing managers, in particular with respect to the implementation of delayed differentiation. On one hand, we provide normative support for redesigning the process when the operation that is causing proliferation is also more costly. In Benetton’s case, dyeing operation is costlier than knitting, and it also causes a high degree of product differentiation. Hence, postponing dyeing to after knitting (by swapping the order of operations) clearly improves the profits. On the other hand, if the operation causing product proliferation is less costly than the other operations, it is not clear how to redesign the process. We show that in such cases delaying differentiation may lead to profit losses even when it can be achieved with no additional cost. Our results indicate that a complementary strategy of deferring high-cost operations to later stages and then focusing on reducing lead times of those scheduled after the proliferation would effectively endow manufacturers with the desired benefits. We also consider the scenario where process sequence cannot be altered, and delineate conditions under which it makes sense to prioritize lead-time reduction over cost reduction, and vice versa. Going a step further, we synthesize our prescriptions and map them into a typology that points out to the most appropriate strategy based on product/market characteristics and process flexibility.

## 2. Literature review

Our research has natural connections with the works that study postponement strategies for delaying product differentiation. One stream within this literature focuses on the design of supply chain structures (Johnson and Anderson 2000, Lee and Tang 1997, 1998), capacity investments (Kouvelis and Tian 2014), and inventory levels at the decoupling points (i.e., vanilla boxes) (Swaminathan and Tayur 1998, Paul et al. 2015). Common to these papers is that demand is assumed to be random without an evolutionary form, so the benefits of postponement are only attributed to inventory pooling—benefits due to improved forecast accuracy are not incorporated. Another stream focuses precisely on demand evolution. In particular, Aviv and Federgruen (2001a,b) analyze the value of order postponement in a multi-period inventory setting where sales occur in each period and demand forecasts are updated in a Bayesian manner. In a similar vein, Atalı and Özer (2012) develop a two-stage production model with product differentiation occurring at the beginning of the second stage

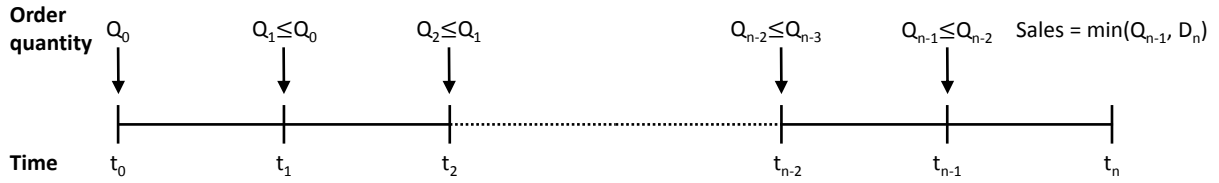
under a Markov-modulated demand model. They show that the value of postponement increases with higher operational flexibility (as measured by difference in minimum and maximum production limits). Our contribution to this literature is the development of a multi-echelon and multi-product newsvendor model with demand forecast evolution. Utilizing this model, we quantify the impact of supply chain structure, the cost and the lead-time values on profits in a product-proliferation setting. Thus, our results shed light into how to employ the two operational levers of delaying the proliferation to improve the profits.

Our paper is also connected to the OM literature that focuses on the multi-ordering inventory models with demand forecast evolution. The closest papers within this literature are Wang et al. (2012) and Biçer and Seifert (2017) because they also develop integrated dynamic inventory models with the martingale model of forecast evolution. Wang et al. (2012) model a newsvendor with multiple ordering opportunities and increasing costs over time, and characterize optimal base-stock levels. Biçer and Seifert (2017) extend Wang et al. by including capacity limitations and allowing for multiple products. In both papers, the ordering decisions are made only for the end products, not for the components or the raw materials at the upstream echelons. We contribute to the extant literature such that we optimize ordering decisions in a multi-echelon setting in which the order quantity of a given operation determines the capacity for the immediate downstream echelon. We also consider the possibility of product proliferation to occur at any echelon in the supply chain. For the same reasons, our model differs from single-item inventory models with evolving demands and multiple ordering opportunities. Song and Zipkin (2012) study such a setting where order quantities can be updated downwards (after paying the cost) as new demand information arrives. Cao and So (2016) consider an assembler ordering from two suppliers (effectively two ordering decisions) with demand forecasts updated over time.

### 3. Model preliminaries

Consider a supply chain with  $(n + 1)$  echelons, where the most downstream echelon  $n$  is closest to the customer and echelon 0 is the farthest from the customer. Echelon  $i + 1$  is considered to be the

downstream and echelon  $i - 1$  the upstream of echelon  $i$ . Supply chain activities occur sequentially such that the operation at echelon  $i$  uses the output of echelon  $i - 1$  as input and transforms it into output. The output of echelon  $i$  is then used as input for echelon  $i + 1$ . Without loss of generality, we assume that one unit of input is transferred into one unit of output. The manufacturer has to make  $n$  ordering decisions at the time epoch  $t_i$  for  $i \in \{0, 1, \dots, n - 1\}$ . Hence, there is a positive lead time at each echelon;  $t_{i+1} - t_i > 0$  for  $i \in \{0, 1, \dots, n - 1\}$  and expediting is not allowed. For ease of exposition, suppose for now that there is a single final product, and let  $Q_i$  denote the order quantity at echelon  $i$ . The order quantity  $Q_i$  for  $i \in \{1, \dots, n - 1\}$  is constrained by the order quantity at the previous echelon (i.e.,  $Q_i \leq Q_{i-1}$ ), while the first order quantity  $Q_0$  is unrestricted. We use  $D_i$  to denote the demand forecast at time  $t_i$  for  $i \in \{0, \dots, n\}$ , with the end demand forecast  $D_n$  representing the actual market demand. The timeline of ordering decisions for this single-product model without any product proliferation is depicted in Figure 3.



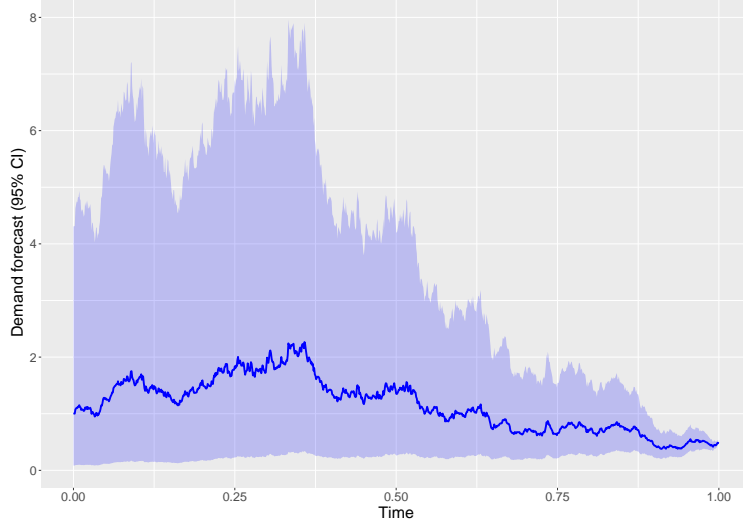
**Figure 3** Timeline of ordering decisions for a single-product model without product proliferation

We model the evolution of demand forecasts  $D_i$  from  $t_0$  to  $t_n$  according to the multiplicative martingale model (m-MMFE), which is known to fit very well to empirical data of demand-forecast updates (Heath and Jackson 1994, Wang et al. 2012, Biçer et al. 2018). According to the m-MMFE, the demand forecasts at  $t = t_i$  for  $i \in \{1, \dots, n\}$  are given by  $D_i = D_0 \exp(\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_i)$ , where  $\varepsilon_t$  follows a normal distribution:

$$\varepsilon_i \sim \mathcal{N}(-\sigma^2(t_i - t_{i-1})/2, \sigma\sqrt{t_i - t_{i-1}}), \quad \forall i \in \{1, \dots, n\}. \quad (1)$$

Therefore, the end demand conditional on the demand forecast at  $t_i$  follows a lognormal distribution:

$$\ln(D_n)|D_i \sim \mathcal{N}(\ln(D_i) - \sigma^2(t_n - t_i)/2, \sigma\sqrt{t_n - t_i}), \quad \forall i \in \{0, \dots, n - 1\}. \quad (2)$$



**Figure 4** A sample path of the demand forecast according to the m-MMFE with 95% two-sided confidence intervals.

In Figure 4 we present an example of the evolution of demand forecasts according to the m-MMFE. We simulate a random path assuming that the initial demand forecast is scaled to one and the  $\sigma$  value is set to one. The forecast evolves from  $t_0 = 0$  until  $t_n = 1$ . The solid curve represents the mean forecast, and the shaded area shows the 95% confidence interval. As the time approaches to the realization of market demand ( $t \rightarrow 1$ ), the forecast accuracy increases significantly as indicated by a reduction of the distance between the upper and the lower bounds of the confidence interval.

The following sequence of events occur at each decision epoch  $t_i$  for  $i \in \{0, \dots, n-1\}$ : i) manufacturer observes the demand forecast  $D_i$ ; ii) the order quantity of the previous operation  $Q_{i-1}$  is reviewed; iii) the order quantity  $Q_i$  is determined, and the manufacturer incurs an operational cost  $c_i Q_i$ . In what follows, we formulate the manufacturer's optimization problem and derive its solution. We do this first for the single-product case and then move on to the most general scenario with product proliferation.

#### 4. Single product model

Consider the single-product model shown in Figure 3, where the final product is sold in a single market. The product is processed from raw materials through a sequence of operations, and sold in the market at a price of  $p$  per unit. We assume that there is no salvage value for the excess inventory.

Thus, a revenue of  $p \min(D_n, Q_{n-1})$  is collected at time  $t_n$ . Let  $c_i$  denote the cost of processing the  $i^{th}$  operation per unit input. This includes all the cost elements such as labor, utility, material, and other operational costs that the manufacturer incurs only from  $t_i$  until  $t_{i+1}$ .

We formulate the manufacturer's optimization problem as a dynamic program (DP). At each decision epoch  $t_i$ , the manufacturer observes the state, which consists of the available supply  $Q_{i-1}$  at the upstream echelon and demand forecast  $D_i$ , and then determines the ordering quantity  $Q_i$  that maximizes expected profits. For the last decision epoch  $t_{n-1}$ , the ordering decision is a constrained newsvendor problem:

$$V_{n-1}(Q_{n-2}, D_{n-1}) = \max_{Q_{n-1} \leq Q_{n-2}} \left\{ \mathbb{E}_{D_n | D_{n-1}} \left[ p \min(D_n, Q_{n-1}) \right] - c_{n-1} Q_{n-1} \right\}. \quad (3)$$

The ordering decisions at the previous decision epochs (i.e.,  $\forall i \in \{0, \dots, n-2\}$ ) can be determined dynamically according to the following Bellman equation:

$$V_i(Q_{i-1}, D_i) = \max_{Q_i \leq Q_{i-1}} \left\{ \mathbb{E}_{D_{i+1} | D_i} \left[ V_{i+1}(Q_i, D_{i+1}) - c_i Q_i \right] \right\}. \quad (4)$$

The order quantity at  $t_0$  is not constrained, so we set  $Q_{-1} = +\infty$ . Let the functions to be maximized in Equations (3) and (4) be denoted respectively as:

$$G_{n-1}(Q_{n-1}, D_{n-1}) = \mathbb{E}_{D_n | D_{n-1}} \left[ p \min(D_n, Q_{n-1}) \right] - c_{n-1} Q_{n-1}, \quad (5)$$

$$G_i(Q_i, D_i) = \mathbb{E}_{D_{i+1} | D_i} \left[ V_{i+1}(Q_i, D_{i+1}) - c_i Q_i \right], \quad (6)$$

with  $g_i(Q_i, D_i) = \partial G_i(Q_i, D_i) / \partial Q_i$ .

Observe that the optimal value of  $Q_i$  in Equation (4) depends on the demand forecasts in all future decision epochs. We define a new parameter  $\bar{D}_j$  for  $j \in \{i+1, \dots, n-1\}$  to represent the critical demand forecast values at time  $t_j$ . If  $D_j \geq \bar{D}_j$  for all  $j \in \{i+1, \dots, n-1\}$ , the optimal order quantities in all the remaining decision epochs become equal to  $Q_i$ . If  $D_j < \bar{D}_j$  for  $j > i$ , the optimal value of  $Q_j$  becomes less than  $Q_i$ . Therefore,  $\bar{D}_j$  values for  $j \in \{i+1, \dots, n-1\}$  determine the lower bounds for demand forecasts that make optimal order quantity at time  $t_j$  equal to  $Q_i$ . Solving the DP model by backward induction, we characterize the optimal ordering policy at each decision epoch, which is presented in the next theorem.<sup>2</sup>

<sup>2</sup> The proofs of all results are presented in our on-line appendix.

THEOREM 1. *The optimal order quantity, denoted by  $q_i$  for  $i \in \{0, \dots, n-1\}$ , satisfies:*

$$q_i = \min(Q_{i-1}, Q_i^*), \quad (7)$$

where  $Q_i^*$  is the optimal order quantity for the unconstrained problem (without “ $Q_i \leq Q_{i-1}$ ”), which is found by the following expressions:

$$Q_i^* = \{Q_i \mid g_i(Q_i, D_i) = 0\}, \quad (8)$$

$$\begin{aligned} g_i(Q_i, D_i) = & pPr(D_n > Q_i, \mathbf{D}_{\{i+1, n-1\}} > \bar{\mathbf{D}}_{\{i+1, n-1\}}) - c_{n-1}Pr(\mathbf{D}_{\{i+1, n-1\}} > \bar{\mathbf{D}}_{\{i+1, n-1\}}) \\ & - c_{n-2}Pr(\mathbf{D}_{\{i+1, n-2\}} > \bar{\mathbf{D}}_{\{i+1, n-2\}}) - \dots - c_{i+1}Pr(D_{i+1} > \bar{D}_{i+1}) - c_i = 0 \end{aligned} \quad (9)$$

with  $\mathbf{D}_{\{i+1, n-1\}}$  denoting the vector of demand forecasts from  $i+1$  to  $n-1$  and  $\bar{\mathbf{D}}_{\{i+1, n-1\}} = (\bar{D}_{i+1}, \dots, \bar{D}_{n-1})$  denoting the vector of critical demand forecasts from  $i+1$  to  $n-1$ .

It can be easily verified that equation (9) reduces to the newsvendor solution for  $i = n-1$  such that:

$$g_{n-1}(Q_{n-1}, D_{n-1}) = pPr(D_n > Q_i) - c_{n-1} = 0. \quad (10)$$

For  $i < n-1$ , the solution is still in spirit the newsvendor solution. The first term of the right-hand side of Equation (9) gives the expected value of the marginal revenue generated by ordering one additional unit when  $(Q_i - 1)$  units are already ordered. The marginal revenue not only depends on the final demand realization  $D_n$  but also on the updated demand forecasts at the remaining decision epochs. Even when  $D_n > Q_i$ , the marginal revenue may be zero if the manufacturer decides to reduce the order quantity in any of the subsequent production stages. The remaining terms of the right-hand side of Equation (9) give the expected value of the marginal cost of ordering one additional unit when  $(Q_i - 1)$  units are already ordered. When the  $Q_i^{th}$  unit is ordered, the manufacturer incurs the cost  $c_i$ . If the demand forecast at the next decision epoch exceeds the critical value (i.e.,  $D_{i+1} \geq \bar{D}_{i+1}$ ), the manufacturer orders  $Q_i$  units at  $t_{i+1}$  and incurs an additional cost of  $c_{i+1}$  per unit and so forth.

PROPOSITION 1. *Optimal order quantity in an upstream echelon is always higher than the expected (optimal) order quantity in a downstream echelon such that  $q_0 > \mathbb{E}[q_1 | D_0] > \dots > \mathbb{E}[q_{n-1} | D_0]$ .*

Proposition 1 states that the interdependency between order quantities (due to supply constraints) and the accumulating cost structure induce the manufacturer to order in large quantities for the upstream operations even though the manufacturer expects the final order quantity to be lower. Next, we present the impact of cost parameters on optimal order quantities and the expected profit.

**PROPOSITION 2.** *A- Let  $\mathbf{q} = \{q_0, q_1, \dots, q_{n-1}\}$  be the vector of optimal order quantities at each decision epoch. If  $c_j$  for  $j \in \{0, \dots, n-1\}$  increases, the optimal order quantities are updated such that  $\mathbf{q}' = \{q'_0, q'_1, \dots, q'_{n-1}\}$ , where  $q'_i$  is statistically smaller than  $q_i$  (i.e.,  $q'_i \prec q_i$ )  $\forall i \in \{0, \dots, n-1\}$ .*

*B- Let  $c_0 = c_1 = \dots = c_{i-1} = c_{i+1} = \dots = c_{n-1} = c_{fixed}$  and  $c_i > c_{fixed}$ . Then, swapping the operation  $i$  with any operation from the set  $\{i+1, \dots, n-1\}$  increases the total expected profit.*

Part A of Proposition 2 describes how the order quantities are affected by an increase in the cost of any operation. If the cost of an operation increases, order quantities at all decision epochs decrease. Part B shows how the sequence of the operations should be redesigned depending on the operational costs. By postponing an operation with a higher cost later than the other operations, the manufacturer increases its profits. Swapping the high cost operation with a downstream lower cost operation increases the upstream order quantity and hence the available supply (upper bounds) for the downstream operation. An increase in the upper bounds for the downstream quantities provides the manufacturer with additional flexibility to adjust order quantities according to updated demand forecasts, leading to higher profits. This result is in line with Lee and Tang (1997) and Cao and So (2016). Lee and Tang (1997) state that redesigning the production processes such that high value-added and short operations take place later than low value-added and long operations leads to higher profits. Cao and So (2016) find that a manufacturer can generate high profits if a supplier with a long lead time supplies a low-value component, whereas another supplier with the short lead time supplies a high-value component. Part B of Proposition 2 establishes effectively the same result for a more general setting. We now turn our attention to the impact of lead times.

PROPOSITION 3. *Reducing the lead time of operation  $i$  for  $i \in \{0, \dots, n-1\}$  by an amount of  $\Delta t \leq t_{i+1} - t_i$  increases expected profit more than what can be achieved by reducing the lead time of operation  $j < i$  by the same amount of  $\Delta t$ .*

This proposition states that reducing the lead time of a downstream operation is more beneficial to the manufacturer than reducing the lead time of an upstream operation by the same amount.

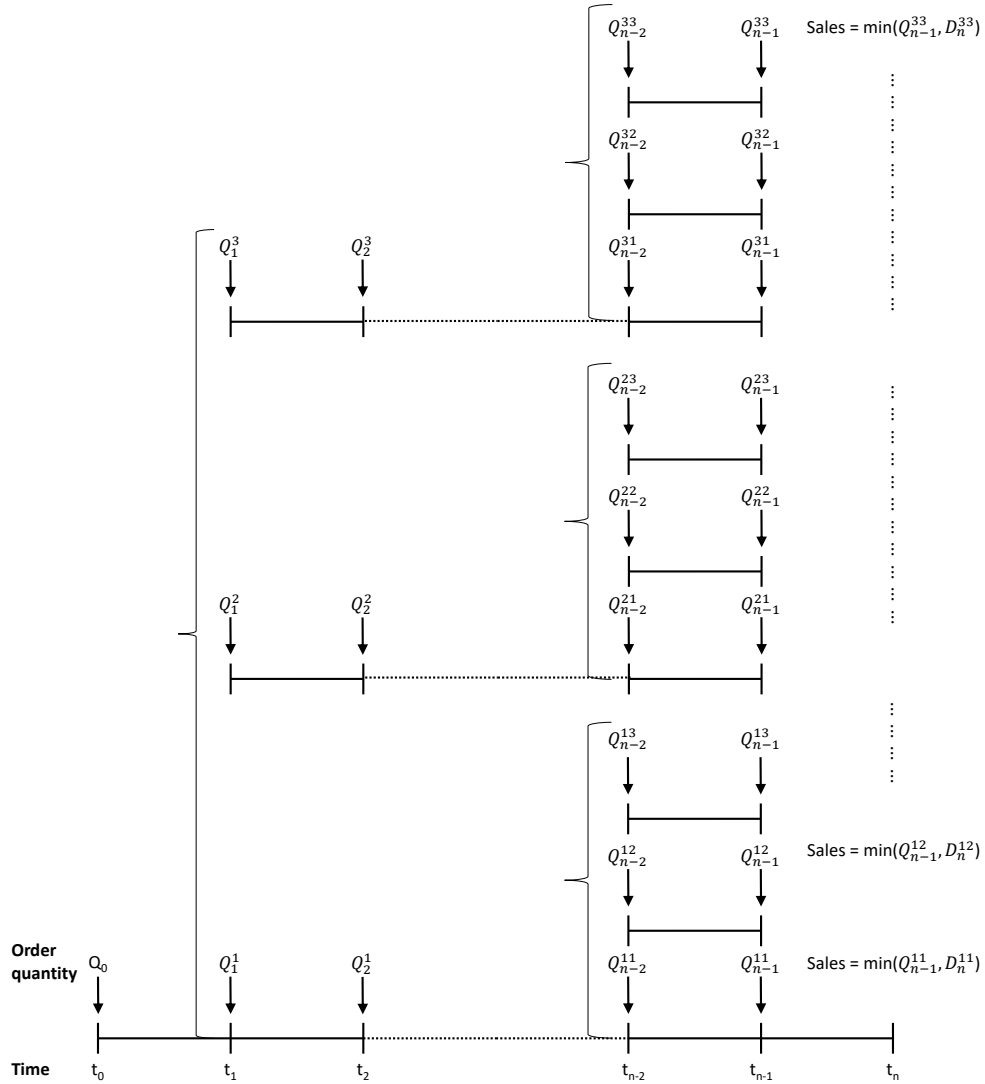
The analytical results given by Propositions 1–3 provide useful insights and clear guidance on how a manufacturer should implement process-redesign and lead-time-reduction practices. Even in the absence of product proliferation at any echelon, manufacturers can still increase the profits by redesigning their processes to postpone high-cost operations. When a manufacturer aims to reduce its operational costs, it should first focus on the upstream operations and then move sequentially downstream. However, the manufacturer should start from the downstream operations and then move upstream if the objective is to reduce the lead time.

## 5. Product proliferation model

We now extend the single-product model to the multi-product case where the raw materials or semi-finished products can be transformed into a variety of products. Product proliferation is allowed at any decision epoch. Given the resulting supply chain structure with the proliferation points, we determine the optimal order quantities at each decision epoch.

To facilitate model development, in Figure 5 we present an example where product proliferation occurs at two epochs:  $t_1$  and  $t_{n-2}$ . We use  $Q_i^j$  to denote the order quantity placed for component  $j$  at time  $t_i$ . We use a unique code to label the component  $j$  at  $t_i$ . The code is a sequence of single digits, and the length of the code gives how often product proliferation occurs from  $t_0$  until  $t_i$ . In our example in Figure 5, at  $t_1$  three different products are ordered, each taking a different digit number. The second proliferation occurs at  $t_{n-2}$ , where the inventory of each product is allocated to produce three differentiated products, amounting to nine SKUs available in the market. Thus, a new digit is added to the product code at  $t_{n-2}$ . Suppose, for example, a fashion-apparel manufacturer selling a product line to different markets uses a three-digit product code (e.g., 361). The first digit





**Figure 5** Ordering decisions in a supply chain with the product proliferation at the decision epochs  $t_1$  and  $t_{n-2}$

represents the size (e.g., small, medium, or large). The second digit represents the color. The third digit denotes the market. The three-digit code means that product proliferation occurs three times along the supply chain (one for size, one for color, and the last for different markets).

The primary challenge in solving the product proliferation problem lies with the need to link the demand dynamics to the ordering constraints. For each ordering decision, it is necessary to consolidate the demand updates of different end products and then allocate the limited supply available from the previous operation to process different semi-finished or end products. We define *two different sets and their subsets* to formalize the problem. To capture the resource constraints, we

use  $\Theta_i$  to denote the set of all components produced at echelon  $i \in \{1, \dots, n\}$  at time  $t_i$ . We further partition the set  $\Theta_i$  into  $k$  pairwise disjoint subsets such as  $\Theta_i^j$  for  $j \in \{1, \dots, k\}$  and  $k = |\Theta_{i-1}|$ . We define  $\Theta_i^j$  as the set that contains all components that use the same upstream resource as their input. We then have by definition:

$$\Theta_i = \bigcup_{j \in \Theta_{i-1}} \Theta_i^j \quad \text{and} \quad \emptyset = \bigcap_{j \in \Theta_{i-1}} \Theta_i^j. \quad (11)$$

Recalling our example in Figure 5,  $\Theta_{n-1} = \{11, 12, 13, 21, 22, 23, 31, 32, 33\}$ . There are nine ordering decisions in the previous period (i.e.,  $t = t_{n-2}$ ), and therefore the set  $\Theta_{n-1}$  is partitioned into nine subsets such that:

$$\Theta_{n-1} = \Theta_{n-1}^{11} \cup \Theta_{n-1}^{12} \cup \Theta_{n-1}^{13} \cup \Theta_{n-1}^{21} \cup \Theta_{n-1}^{22} \cup \Theta_{n-1}^{23} \cup \Theta_{n-1}^{31} \cup \Theta_{n-1}^{32} \cup \Theta_{n-1}^{33},$$

where  $\Theta_{n-1}^{11} = \{11\}$ ,  $\Theta_{n-1}^{12} = \{12\}$ ,  $\Theta_{n-1}^{13} = \{13\}$ ,  $\Theta_{n-1}^{21} = \{21\}$ ,  $\Theta_{n-1}^{22} = \{22\}$ ,  $\Theta_{n-1}^{23} = \{23\}$ ,  $\Theta_{n-1}^{31} = \{31\}$ ,  $\Theta_{n-1}^{32} = \{32\}$ , and  $\Theta_{n-1}^{33} = \{33\}$ . Likewise, at  $t = t_{n-2}$ ,  $\Theta_{n-2} = \{11, 12, 13, 21, 22, 23, 31, 32, 33\}$ .

There are three ordering decisions in the previous period (i.e.,  $t = t_{n-3}$ ) so  $\Theta_{n-2}$  is partitioned into three subsets:  $\Theta_{n-2}^1 = \{11, 12, 13\}$ ,  $\Theta_{n-2}^2 = \{21, 22, 23\}$ , and  $\Theta_{n-2}^3 = \{31, 32, 33\}$ .

With these sets defined, we can write down the ordering constraints between echelons. That is, the sum of the order quantities for the products that use the same input cannot be larger than the order quantity of the input at the immediate upstream echelon. In mathematical terms:

$$\sum_{j \in \Theta_i^k} Q_i^j \leq Q_{i-1}^k. \quad (12)$$

Returning back to the Figure 5 example, the order quantity constraints at  $t_{n-1}$  are  $Q_{n-1}^j \leq Q_{n-2}^j$  for each  $j \in \Theta_{n-1}$ . At  $t_{n-2}$ , we have three ordering constraints:

$$Q_{n-2}^{11} + Q_{n-2}^{12} + Q_{n-2}^{13} \leq Q_{n-3}^1,$$

$$Q_{n-2}^{21} + Q_{n-2}^{22} + Q_{n-2}^{23} \leq Q_{n-3}^2,$$

$$Q_{n-2}^{31} + Q_{n-2}^{32} + Q_{n-2}^{33} \leq Q_{n-3}^3.$$

We can then formalize the other order quantity constraints at  $t_i$  as  $Q_i^j \leq Q_{i-1}^j$  for  $i \in \{2, \dots, n-3\}$  and  $j \in \Theta_i$ . Finally, at  $t = t_1$ —that is, when the first proliferation occurs—we have  $Q_1^1 + Q_1^2 + Q_1^3 \leq Q_0$ .

We also define another set  $\Upsilon_i^k$  which represents the set of end products produced by using component  $k$  at echelon  $i$ . Therefore,  $\Upsilon_i^k$  includes the end products (sold in the markets), whose availability depends on the order quantity decision of  $Q_i^k$ . In Figure 5, for example, the quantity  $Q_1^1$  has a direct influence on the ordering decisions of the end products:  $Q_{n-1}^{11}$ ,  $Q_{n-1}^{12}$ , and  $Q_{n-1}^{13}$ . Therefore,  $\Upsilon_1^1 = \{11, 12, 13\}$ . The set  $\Upsilon_0 = \Theta_{n-1}$  since the quantity  $Q_0$  has direct influence on the final inventory of all end products. Let  $p_j$  for  $j \in \Theta_n$  denote the price of the products sold in the market.

To determine the maximum expected profit at  $t_{n-1}$ , we write the following stochastic programming (SP) model (Shapiro et al. 2009, Ch. 1):

$$\begin{aligned} \text{Maximize}_{Q_{n-1}^j, \forall j \in \Theta_{n-1}} \quad & z = \sum_{j \in \Theta_{n-1}} p_j \mathbb{E} \left( \mathcal{W}_j(Q_{n-1}^j, D_n^j) \right) - c_{n-1}^j Q_{n-1}^j \\ \text{subject to:} \end{aligned} \quad (13)$$

$$\sum_{j \in \Theta_{n-1}^k} Q_{n-1}^j \leq Q_{n-2}^k, \quad \forall k \in \Theta_{n-2}, \quad (14)$$

$$Q_{n-1}^j \geq 0, \quad \forall j \in \Theta_{n-1}, \quad (15)$$

where  $\mathcal{W}_j(Q_{n-1}^j, D_n^j) = \min\{Q_{n-1}^j, D_n^j\}$  denotes the sales and  $D_n^j$  is a random variable. Constraint (14) guarantees that the sum of order quantities of the items in a set  $\Theta_{n-1}^k$  is less than the amount of their parent item  $k$ . In the appendix, we provide the solution for the mathematical problem (13)–(15). Specifically, we transform the SP model into a linear programming (LP) model as demonstrated by Shapiro et al. (2009, Ch. 1–3). By analyzing the LP model and its dual, we partition the demand space and determine the shadow prices (see Van Mieghem (1998) for a similar method to solve an SP problem). We then proceed backwards in a similar fashion, using induction, and determine the optimal ordering policy for upstream echelons. The optimal policy is satisfied when all products in a set  $\Theta_i^k$  (for all  $k$  and  $i$  values) have the same marginal value of ordering one additional unit. If the quantity  $Q_{i-1}^k$  is highly restrictive, the marginal value for all products in the set  $\Theta_i^k$  would

have a positive value. If the quantity  $Q_{i-1}^k$  is excessive, the marginal value would become zero. This analysis reveals the structure of the optimal policy as well.

**THEOREM 2.** *The optimal ordering policy for all the items in each decision epoch is a resource-constrained, state-dependent base-stock policy, which depends on the evolution of demand forecasts and processing costs.*

With the characterization of the optimal policy at hand, we can use our framework to analyze the impact of point of proliferation, costs, and lead times. Clearly, everything else remaining the same, delaying differentiation (moving any point  $t_i$  with proliferation forward) is beneficial to the firm.

**PROPOSITION 4.** *The value of delaying the point of product proliferation increases as the costs of downstream operations taking place after the point of proliferation increase.*

Proposition 4 has important implications. It first underpins delaying differentiation by swapping costly operations that cause proliferation with downstream less costly operations. As documented by the Benetton's case, delaying the point of proliferation and the costly dyeing operation were both achieved by only swapping two operations. If such an improvement achieved by a single change is not possible, redesigning processes such that costly operations scheduled before the proliferation are swapped with less costly post-proliferation operations should precede any attempt to reduce lead times and postpone the proliferation point.

**PROPOSITION 5.** *There exists a threshold value for the cost of the operation that causes the proliferation such that swapping the point of proliferation with a more costly downstream operation causes a profit loss if the actual cost of the (proliferation) operation is less than the threshold value.*

Proposition 5 highlights the important factors that may render the postponement strategy harmful for manufacturers. If the product differentiation is achieved by carrying out a cheap operation, delaying the differentiation by swapping it with a more costly downstream operation may harm the profit. Therefore, practitioners should pay attention to the costs of operations before making any postponement decision. At the end of Section 6.1, we present an example that shows the breakeven cost difference justifying the value of postponement. Next, we investigate the impact of lead times.

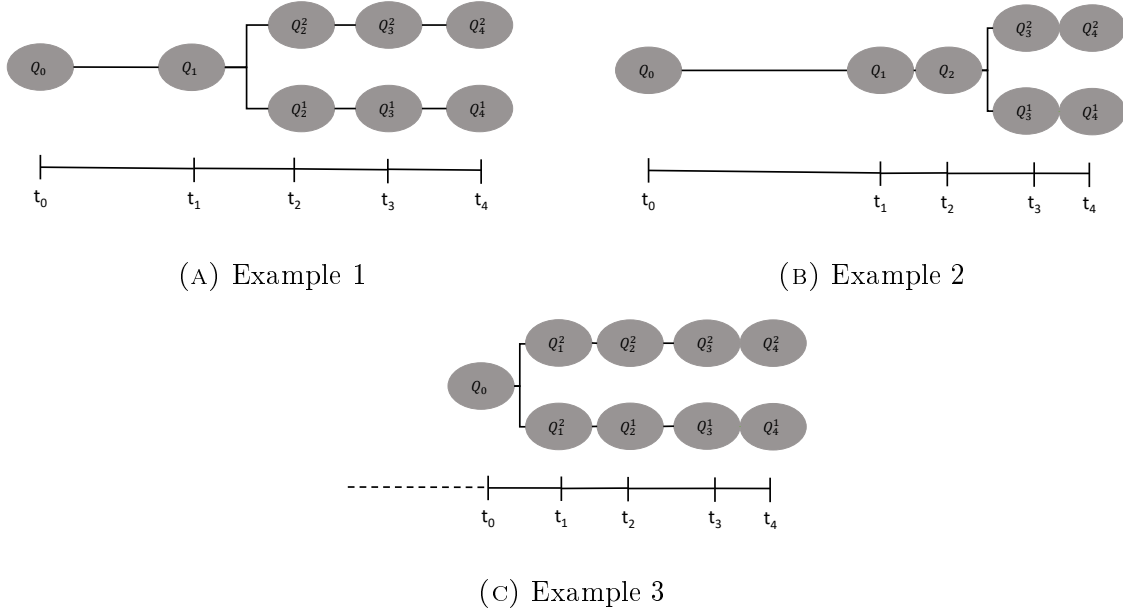
PROPOSITION 6. *Suppose product proliferation occurs once along the supply chain at time  $t_i$ . Reducing the lead time of operation  $j$  for  $j \in \{i, \dots, n-1\}$  by an amount of  $\Delta t_j \leq t_{j+1} - t_j$  increases the expected profit more than what can be achieved by reducing the lead time of operation  $j$  for  $j \in \{0, \dots, i-1\}$ .*

This proposition extends the results of Proposition 3 to the multi-product setting. It demonstrates that giving priority to downstream operations in lead-time reduction is more effective than upstream operations for manufacturers also when there is product proliferation in the supply chain.

## 6. Numerical analysis

The analytical results derived in the previous two sections provide valuable guidelines in terms of implementing delayed differentiation. Proposition 5 shows that the delayed differentiation through process redesign may harm the profits, which we believe to be counter-intuitive. To better understand the dynamics of process redesign and to complement our analytical results, we resort to numerical experiments. We set up a large scale numerical study based on Markovian sampling to limit any bias that may come from specific parameter settings.

We consider a setting with five echelons (i.e.,  $n = 4$ ; the first ordering decision is made at  $t_0$  and market demand is observed at  $t_4$ ) and two end products. We allow the demands of the two products to be correlated, and sample the correlation parameter  $\rho$  from a uniform distribution,  $\rho \sim U(-1, 1)$ . We assume that both products have the same value of the coefficient of variation (CV), and sample the CV value at each iteration from a uniform distribution,  $CV \sim U(0, 1)$ . Then, we calculate the volatility parameter for the m-MMFE using the formula:  $\sigma = \sqrt{\ln(CV^2 + 1)}$ . We normalize the forecasting horizon to one,  $t_4 = 1$ . We randomly sample the total time length ( $t_4 - t_0$ ) from a uniform distribution:  $t_4 - t_0 \sim U(0, 1)$ . We randomly allocate the total time length to different operations so that the lead times  $t_1 - t_0$ ,  $t_2 - t_1$ ,  $t_3 - t_2$ , and  $t_4 - t_3$  are randomly determined at each iteration. The product proliferation occurs at only one of those echelons, which is randomly selected. Figure 6 shows a small sample of three different supply chain structures that can be generated based on our random sampling procedure.



**Figure 6** Some potential supply chain structures

We fix the selling price to one and salvage value to zero. We sample the total production cost per unit from a uniform distribution of  $U(0,1)$ . We randomly allocate the total production cost to the operations. After determining the input parameters, we generate demand paths and calculate the realized profits. We set the demand forecast at time  $t=0$  to one. Given that the m-MMFE gives unbiased estimates, the expected revenue (i.e.,  $p \times D_{t=0}$ ) is equal to one. Therefore, the results of our statistical analysis can be interpreted in terms of *percentages of the expected revenue*.

For a given set of input parameters, we iteratively sample the demand forecast and optimize the order quantity at each decision epoch. Demand updating occurs according to the m-MMFE process, for which we use the conditional distribution given by Equation (2). Once the demand forecast for a decision epoch is generated, the optimal order quantity is determined using Theorem 2. Then, the “reward” values at each stage are calculated, yielding the total profit. We sample the input parameters 10000 times. Then, we generate ten random demand paths for each set of input parameters, obtaining 100000 data points for our analysis.

To analyze the results, we setup a linear regression model as follows:

$$\begin{aligned} Profit = & \beta_0 + \beta_1 c_1 + \beta_2 c_2 + \beta_3 c_3 + \beta_4 c_4 + \beta_5 \rho + \beta_6 CV + \beta_7 TimeofDiff \\ & + \beta_8 (t_1 - t_0) + \beta_9 (t_2 - t_1) + \beta_{10} (t_3 - t_2) + \beta_{11} (t_4 - t_3). \end{aligned} \quad (16)$$

The dependent variable is the total profit generated. We select the cost parameters, demand parameters (correlation between demand for the products, coefficient of variation), the time and point of proliferation, and the lead time for each operation as independent variables.

### 6.1. Analysis of the results

The descriptive statistics of the dependent and independent variables are given in Table 1. The estimates for the coefficients and the variance inflation factors (VIFs) are given in Table 2. The estimates are found to be robust after separately testing the underlying assumptions of the linear model (16).

	Pearson's Correlation													
	Mean	SD	Min	Max	1	2	3	4	5	6	7	8	9	10
Profit (DV)	0.78	0.65	-1.75	11.23										
1. $c_1$	0.13	0.17	0.00	0.99										
2. $c_2$	0.13	0.17	0.00	0.98	-0.08									
3. $c_3$	0.12	0.16	0.00	0.98	-0.09	-0.07								
4. $c_4$	0.12	0.17	0.00	0.97	-0.09	-0.08	-0.07							
5. $\rho$	0.00	0.57	-1.00	1.00	0.00	0.00	0.00	-0.01						
6. CV	0.49	0.29	0.00	1.00	0.01	0.01	0.00	0.01	-0.02					
7. TimeofDiff	0.45	0.35	0.00	1.00	0.00	0.00	-0.01	0.01	0.00	0.00				
8. $t_1 - t_0$	0.22	0.24	0.00	1.00	0.00	-0.01	-0.02	0.01	0.00	0.00	0.28			
9. $t_2 - t_1$	0.21	0.24	0.00	1.00	0.00	0.01	0.00	0.01	0.00	0.02	0.06	-0.29		
10. $t_3 - t_2$	0.22	0.24	0.00	1.00	0.00	0.00	-0.01	0.01	0.00	0.00	-0.17	-0.30	-0.29	
11. $t_4 - t_3$	0.22	0.24	0.00	1.00	-0.01	0.01	0.02	-0.02	0.00	-0.02	-0.38	-0.30	-0.29	-0.29

**Table 1 Means, Standard Deviations, Minima, Maxima, and Correlations**

The intuitive effects of demand uncertainty and correlation are readily observed in our numerical study. Specifically, the coefficient of the correlation parameter (i.e.,  $\beta_5 = -0.0209655$ ) is negative, meaning that total profit increases as the demand for products becomes more negatively correlated. This result is due to the benefits of inventory pooling in the stages before the point of product proliferation, in a way that the value of pooling inventory increases as demand for the products gets more negatively correlated. The coefficient of the CV is also negative (i.e.,  $\beta_6 = -0.3968276$ ), which is aligned with our expectations since an increase in demand uncertainty leads to higher supply-demand mismatches and lower profits.

	Estimate	VIF
$\beta_0$ (Intercept)	2.0534208***	
$\beta_1$ ( $c_1$ )	-2.0029586***	1.027807
$\beta_2$ ( $c_2$ )	-1.9264279***	1.022989
$\beta_3$ ( $c_3$ )	-1.8938598***	1.023589
$\beta_4$ ( $c_4$ )	-1.8259906***	1.025135
$\beta_5$ ( $\rho$ )	-0.0209655***	1.000668
$\beta_6$ ( $CV$ )	-0.3968276***	1.001141
$\beta_7$ ( <i>TimeofDiff</i> )	0.0295013***	4.597373
$\beta_8$ ( $t_1 - t_0$ )	-0.0737628***	2.884366
$\beta_9$ ( $t_2 - t_1$ )	-0.1125864***	3.287310
$\beta_{10}$ ( $t_3 - t_2$ )	-0.1356525***	4.024351
$\beta_{11}$ ( $t_4 - t_3$ )	-0.1559704***	4.988999

**Table 2** Summary statistics. Residual standard error: 0.3194 on 99986 degrees of freedom. R-squared: 0.7579.

The numerical results substantiate our analytical results on costs and lead times. The coefficients of the cost parameters satisfy the relationship:  $|\beta_1| > |\beta_2| > |\beta_3| > |\beta_4|$ , indicating that the impact of increasing the cost of an operation is more pronounced for the operations scheduled earlier. This result is in line with Part B of Proposition 2 and Proposition 4. In a similar vein, the values of  $\beta_8$ ,  $\beta_9$ ,  $\beta_{10}$ , and  $\beta_{11}$  are all negative, indicating that reducing the time to complete any operation has a positive impact on the profit. Furthermore, consistent with Proposition 3, the value of reducing lead times is higher for operations scheduled later—that is,  $|\beta_{11}| > |\beta_{10}| > |\beta_9| > |\beta_8|$ .

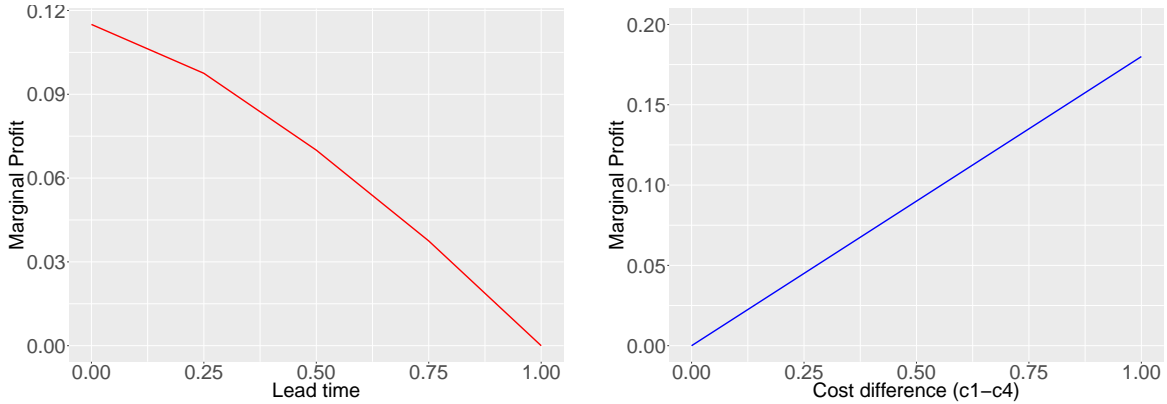
Regarding the point of product proliferation, total profit increases as the timing of differentiation (*TimeofDiff*) is delayed, given that  $\beta_7$  has a positive value.<sup>3</sup> Interestingly, however, the value of delaying the point of the proliferation is relatively low compared to process redesign after excluding the positive impact of reducing lead times. For example, delaying the point of the proliferation fully from the beginning of the forecast horizon  $t = 0$  to the end  $t = 1$  leads to only a 2.9% increase in profit. This is considerably lower than the benefits of swapping an upstream high-cost operation with a downstream low-cost operation for example. This emphasizes the fact that critical benefits of delayed differentiation is generated not so much from simply pushing the proliferation point closer

<sup>3</sup> We remark that there is no serious multicollinearity problem in our problem. This can be verified by examining the VIF values. It is generally accepted that VIF values above 5 or 10 indicates a multicollinearity problem, see for example James et al. (2013, pp.101–102). In our model all independent variables have VIF values less than 5.



to the market, but rather from the associated process redesign that accrue cost benefits and/or the lead-time reduction that comes along with it. To substantiate this, we note that in Table 2,  $\beta_4 - \beta_1 \gg \beta_7$ , implying that postponing high-cost operations creates more value than delaying the point of proliferation.

Next, we compare the effects of lead-time reduction with process redesign. We find that reducing the lead time of the last operation leads to an increase in the profits by 15.60% of absolute change in the lead time. If, for example, the lead time is reduced by 0.2, it helps to increase profits by  $0.2 \times 15.6\% = 3.12\%$ . Consistent with our analytical results (Propositions 3 and 6), this percentage reduces to 13.56%, 11.26%, and 7.38% for the third, second, and first operations, respectively. Although reducing lead times improves profits, its impact may be less than the value generated by swapping a high-cost operation with a low-cost downstream operation. For example, swapping the first operation with the last one changes the profit on average by  $2.003 - 1.826 = 17.7\%$  of the cost difference.



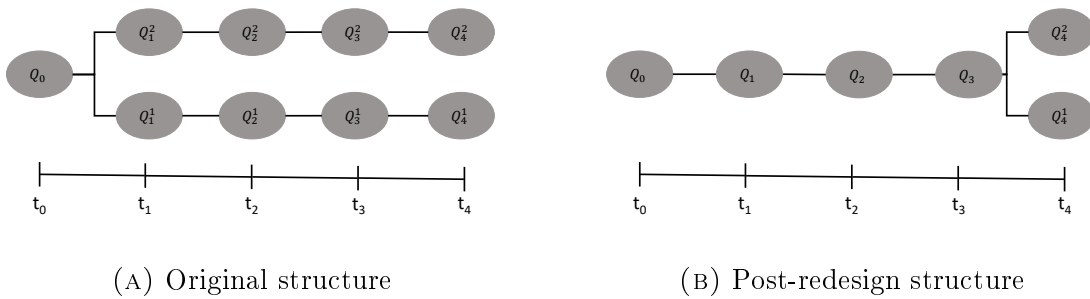
**Figure 7** Value of lead-time reduction versus process redesign

In Figure 7 we demonstrate the profit increase that can be achieved by lead-time reduction on the left panel and that by swapping the first operation with the forth operation on the right panel (utilizing coefficients  $\beta_1, \beta_4, \beta_8, \beta_9, \beta_{10}$ , and  $\beta_{11}$ ). To generate the left panel, we reduce the lead time from 100% to a new value given by the x-axis. Initially, we take the same lead-time values for each operation such that  $t_4 - t_3 = t_3 - t_2 = t_2 - t_1 = t_1 - t_0 = 0.25$ . We then reduce the lead time

starting from the most downstream operation going upstream. We find that the maximum marginal profit achieved through lead-time reduction is around 0.12. To generate the right panel, we varied the cost difference  $c_1 - c_4$  from zero to one and computed the marginal profit achieved by swapping the first and fourth operations. We find that maximum marginal profit achieved by this swapping of operations is 0.18.

In our numerical setting, we normalize total cost to one such that  $c_1 + c_2 + c_3 + c_4 = 1$ . Thus, having  $c_1 - c_4 > 0.8$  is not practically possible for most manufacturers. When  $c_1 - c_4$  is relatively high (e.g.,  $c_1 - c_4 > 0.40$ ), swapping the first and last operations helps increase the profit by more than 0.075. To achieve the same profit increase, the lead time has to be reduced by 0.55 (from one to 0.45). Evidently, redesigning the supply chain by swapping the operations has more potential to improve profits than lead-time reduction when  $c_1 - c_4$  is relatively high. If  $c_1 - c_4$  is relatively low (e.g.,  $c_1 - c_4 < 0.05$ ), swapping the operations does not have a significant positive impact on the profit. In this case, lead-time reduction is more beneficial to manufacturers than swapping the operations.

**Exemplifying the Risk of Postponement Strategy:** Now we focus on the case in which  $c_1 < c_4$ , and the proliferation occurs at the first echelon. The lead time for each operation is equal to 0.25. To postpone the proliferation point we consider swapping the first and the forth operations. In Figure 8, we present the original and post-redesign structures of the production system.



**Figure 8** The risky postponement strategy

Using  $\beta_1$ ,  $\beta_4$ , and  $\beta_7$ , we can show that the post-redesign expected profit becomes less than that for the original case when the following inequality holds:

$$Profit_B < Profit_A.$$

Plugging the parameter values to Equation (16) yields:

$$-2.002c_4 - 1.826c_1 + 0.029 \times 0.75 < -2.002c_1 - 1.826c_4 + 0.029 \times 0,$$

$$c_1 - c_4 < -0.123.$$

If the cost difference between the forth and the first operations is larger than 12.3% of the total cost (i.e.,  $c_4 - c_1 > 0.123$ ), swapping the operations to postpone the proliferation point is not justified. If  $c_4 - c_1 < 0.123$ , the benefits of delaying the proliferation compensate for the negative effect of pre-scheduling the high-cost operation on profits. Therefore, practitioners should be cautious of the potential negative impact of the postponement strategy on profits if the implementation of such a strategy requires pre-scheduling a high-cost activity.

## 6.2. Practical Implications and Insights

Our analytical results, combined with the evidence obtained from the preceding numerical study, offer significant insights regarding operational strategies that can be employed to implement delayed differentiation and improve bottom-line performance of supply chains with product proliferation. We now synthesize these strategies and translate them into managerial prescriptions that describe the most suitable conditions for implementing them.

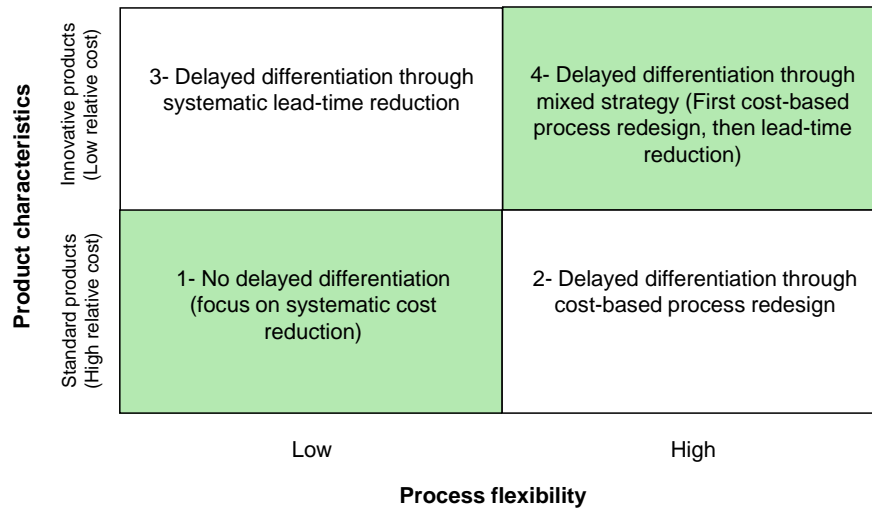
Propositions 2 and 4 together with the estimated coefficients  $|\beta_1| > |\beta_2| > |\beta_3| > |\beta_4|$  in our numerical study confirm that unit cost reductions at an upstream echelon is more beneficial than a unit cost reduction at a downstream echelon. Therefore, if changing the sequence of operations is not possible, manufacturers can still improve profits by systematically reducing costs, starting from upstream operations and then moving downstream. We call this strategy *systematic cost reduction*. Since upstream operations are often related to procurement of raw materials or subassemblies, systematic cost reduction calls for prioritizing the improvement of the procurement efficiency via

consolidating purchasing orders, creating purchase bundles, using low-cost substitutes, or other policies (Paranikas et al. 2015). Clearly, in the absence of any flexibility to alter the process, cost reductions do not delay differentiation. The only option that makes it possible to benefit from delaying the point of proliferation is lead-time reduction. Propositions 3 and 6 along with the numerical findings  $|\beta_{11}| > |\beta_{10}| > |\beta_9| > |\beta_8|$  corroborate that manufacturers should try to reduce first the lead time of downstream operations and then move upstream in the supply chain. We call this strategy *systematic lead-time reduction*.

When there is some flexibility in adjusting the process, Propositions 2 and 4 and Figure 7 affirm that manufactures can improve profits through a *cost-based process redesign* strategy, which effectively postpones high-cost operations to later stages, ideally post-proliferation. Under cost-based process redesign manufacturers can strategically increase profits without necessarily squeezing suppliers. If a costly operation also causes a high degree of proliferation, profit increase due to cost-based process redesign is magnified. As exemplified before, this can be achieved by swapping such a costly operation with a less costly downstream operation. If this is not possible, our results underscore the value of conducting cost-based process design before lead-time reduction efforts. To augment a cost-based redesign strategy with lead-time reduction, which we refer to as *mixed* strategy, manufacturers first conduct cost-based process redesign, and then reduce lead times, starting from downstream, post-proliferation operations and moving upwards in the supply chain.

The strategic value that can be harvested by manufacturers from adopting these four strategies depends naturally on characteristics of both the industry they operate in as well as the markets their products serve. Manufacturers not having the process flexibility to conduct re-sequencing or major process changes is indicative of relatively mature industries where manufacturing processes are standardized and widely adopted within the industry. Manufacturers in such industries have to rely on established templates for producing their products. Manufacturers using propriety processes to produce their products, however, may have the flexibility to change the sequence of their operations and redesign their processes based on cost and lead time parameters. Such cost-based process

redesign efforts will be most effective when manufacturing costs constitute a significant portion of revenues. As our results highlight, cost-based process redesign is particularly efficient when there is a large difference in production costs between adjacent echelons. These conditions are more likely to hold for manufacturers selling standard, more commoditized products. Such products have low gross margins, and total production cost constitutes a significant percentage of the total revenue. Consistent with the Pareto principle, it is common in practice that around 80% of total cost can be attributed to 20% of all the activities. Thus, it is more likely to expect a sizeable cost difference between the operations. Manufacturers selling innovative products on the other hand often have high gross margins because they can command higher prices for their products. When the gross margin is relatively high, total cost constitutes a small amount of the revenues, and it becomes more critical to complement cost-driven efforts with lead-time reduction. Amalgamating these insights, we derive the typology depicted in Figure 9.



**Figure 9** Decision typology based on product characteristics and process flexibility

In developing the decision typology in Figure 9, we make use of the classification in Ferdows et al. (2016) that categorizes manufacturers based on two dimensions, namely their product characteristics and process flexibility. We then align the resulting four quadrants with the most effective strategy for delaying differentiation and improving profits, as identified by our preceding analysis.

*1-Bottom-left quadrant (Systematic cost reduction):* These manufacturers produce commodity-like products using industry-standard production methods. Some business units of chemical companies (e.g., DuPont, BASF, etc) producing commodity-like products fall into this category. Given that the products are sold at a low profit margin, costs represent a significant portion of revenues. Cost-based process redesign is not possible for these manufacturers since the processes they employ are highly standardized. Accordingly, delaying differentiation is not a real option to cope with product proliferation. Systematic cost reduction is the only viable strategy for improving the bottom line. Since the cost of raw materials may constitute up to 80% of total revenue in such industries, it is not so uncommon that manufacturers try to reduce upstream costs by pressurizing their suppliers.<sup>4</sup>

*2-Bottom-right quadrant (Cost-based process redesign):* Some manufacturers excel in process flexibility while producing standard products. Ferdows et al. (2016) give an example of a US-based steel manufacturer producing steel rolls that built this flexibility through some advanced processes. Manufacturers in this category can adopt cost-based process design to cope with product proliferation. This was certainly the case for Benetton at the time when it resequenced its operations to postpone the costly dyeing operations for its highly standardized sweaters. Our leading process industry manufacturer also falls into this category. Even though the end products are customized to customer needs, the product is a standard input for tyres and costs take up a large fraction of revenues. In sharp contrast to the Benetton case, for the process industry manufacturer, the primary point of proliferation occurs at a low cost operation, namely weaving. Furthermore, swapping this weaving with downstream blending is not technically possible. Nevertheless, it is possible to limit variety introduced at the weaving but achieve target specifications through more sophisticated blending. This decreases weaving costs, makes blending more costly, but effectively delays proliferation. Our results advocate the adoption of this cost-based process redesign, which enables the manufacturer to take more advantage of upstream inventory pooling and improved forecast accuracy downstream.

*3-Top-left quadrant (Systematic lead-time reduction):* Manufacturers with strong brands, such as some fashion-apparel manufacturers and pharmaceutical companies fall into this group (Ferdows

<sup>4</sup> <https://www.mckinsey.com/industries/chemicals/our-insights/pursuing-purchasing-excellence-in-chemicals>

et al. 2016). Although standard processes are used in production, innovative/fashionable nature of the product and the brand value allow premium pricing and generate higher margins. As standardized processes leave little room for restructuring, lead time is the only lever for managing proliferation. Like Zara, manufacturers in this category should systematically reduce lead times to delay differentiation and thereby improve responsiveness and profits.

*4-Top-right quadrant (Mixed strategy):* Manufacturers with proprietary products and processes, such as Intel, can differentiate themselves through both product design and process technology (Ferdows et al. 2016). Their products are sold in the market at a high margin which makes lead-time reduction appealing. With process flexibility, redesigning the processes to postpone high-cost activities may also be possible, which amplifies the value generated by reducing lead times. Manufacturers in this category are ideally suited for following the mixed strategy of coupling cost-based process redesign with lead-time reduction. ASML, a Dutch company producing modular lithography systems for semiconductor manufacturers, has implemented this strategy as part of its value-sourcing initiative (van Rooy 2010). The company postponed the operations that required expensive components to a later stage in production, and reduced their sourcing lead times by paying the suppliers premiums.<sup>5</sup> This mixed strategy enables ASML to delay both the point of proliferation and high cost operations.

## 7. Concluding Remarks

In this paper, we develop an analytical model to quantify the impact of supply chain structure along with the cost, the demand, and the lead-time parameter on the profits in a multi-echelon and multi-product newsvendor model with the product proliferation occurring at pre-specified echelons. In such a setting, decision makers can improve the profits by establishing the responsiveness and/or the cost efficiency. Delaying the proliferation helps decision makers establish the responsiveness, which can be implemented through the lead-time reduction or the process redesign. Establishing the cost efficiency is also possible through the systematic cost reduction or by postponing the high-cost

<sup>5</sup> See also [https://staticwww.asml.com/doclib/investor/07\\_analyst\\_day\\_internet\\_031113.pdf](https://staticwww.asml.com/doclib/investor/07_analyst_day_internet_031113.pdf)

operations until partial or full resolution of demand uncertainty. Utilizing our analytical framework, we develop a decision typology that shows effective strategies depending on the product and the process characteristics.

Our model inherently assumes a make-to-stock supply chain with positive lead times for production stages but zero promised lead time for customers (i.e., maximum length of time in which a customer order is guaranteed to be delivered). When a customer is willing to wait, the manufacturer can quote a positive promised lead time at a discounted price and follow a combination of make-to-order and make-to-stock policies—that is, creating a decoupling point in the supply chain. Reducing lead time in this context would possibly help companies delay differentiation after the decoupling point, so product proliferation takes place after getting firm customer orders, completely eliminating inventory risk at downstream echelons. We believe that the trade-off between completely eliminating the downstream inventory risk and profit losses due to offering price discounts for longer promised lead times would be an interesting avenue of future research that requires incorporation of lead-time quotation and product proliferation models.

## References

- Atalı A, Özer Ö (2012) Stochastic multi-item inventory systems with markov-modulated demands and production quantity requirements. *Probability in the Engineering and Informational Sciences* 26(2):263–293.
- Aviv Y, Federgruen A (2001a) Design for postponement: A comprehensive characterization of its benefits under unknown demand distributions. *Operations Research* 49(4):578–598.
- Aviv Y, Federgruen A (2001b) Capacitated multi-item inventory systems with random and seasonally fluctuating demands: Implications for postponement strategies. *Management Science* 47(4):512–531.
- Biçer I, Hagspiel V, de Treville S (2018) Valuing supply-chain responsiveness under demand jumps. *Journal of Operations Management* 61:46–67.
- Biçer I, Seifert RW (2017) Optimal dynamic order scheduling under capacity constraints given demand-forecast evolution. *Production and Operations Management* 26(12):2266–2286.



- Brynjolfsson E, Hu Y, Simester D (2011) Goodbye pareto principle, hello long tail: The effect of search costs on the concentration of product sales. *Management Science* 57(8):1373–1386.
- Cao J, So KC (2016) The value of demand forecast updates in managing component procurement for assembly systems. *IIE Transactions* 48(12):1198–1216.
- Caro F, Martínez-de Albéniz V (2015) Fast fashion: Business model overview and research opportunities. Agrawal N, Smith SA, eds., *Retail Supply Chain Management: Quantitative Models and Empirical Studies*, 237–264 (Boston, MA, US: Springer).
- Ferdows K, Vereecke A, De Meyer A (2016) Delaying the global production network into congruent sub-networks. *Journal of Operations Management* 41:63–74.
- Fisher ML (1997) What is the right supply chain for your product? *Harvard Business Review* 75:105–117.
- Ghemawat P, Nueno JL (2006) Zara: Fast Fashion, Harvard Business School Case: 9-703-497.
- Heath DC, Jackson PL (1994) Modeling the evolution of demand forecasts with application to safety stock analysis in production/distribution systems. *IIE Transactions* 26(3):17–30.
- Heskett JL, Signorelli S (1989) Benetton (A), Harvard Business School Case: 9-685-014.
- James G, Witten D, Hastie T, Tibshirani R (2013) *An Introduction to Statistical Learning with Applications in R* (Springer, New York, US).
- Johnson ME, Anderson E (2000) Postponement strategies for channel derivatives. *The International Journal of Logistics Management* 11(1):19–36.
- Kouvelis P, Tian Z (2014) Flexible capacity investments and product mix: Optimal decisions and value of postponement options. *Production and Operations Management* 23(5):861–876.
- Lee HL, Tang CS (1997) Modelling the costs and benefits of delayed product differentiation. *Management Science* 43(1):40–53.
- Lee HL, Tang CS (1998) Variability reduction through operations reversal. *Management Science* 44(2):162–172.
- Paranikas P, Whiteford GP, Tevelson B, Belz D (2015) How to negotiate with powerful suppliers. *Harvard Business Review* 93(7):90–96.

- Parsons JCW, Graves SC (2005) Reebok NFL replica jerseys: A case for postponement, MIT Sloan School of Management, SCM Case.
- Paul A, Tan YR, Vakharia AJ (2015) Inventory planning for a modular product family. *Production and Operations Management* 24(7):1033–1053.
- Rajagopalan S (2013) Impact of variety and distribution system characteristics on inventory levels at US retailers. *Manufacturing & Service Operations Management* 15(2):191–204.
- Shapiro A, Dentcheva D, Ruszczyński A (2009) *Lectures on Stochastic Programming: Modeling and Theory* (SIAM:Society for Industrial and Applied Mathematics, Philadelphia, PA 19104-2688, USA).
- Song JS, Zipkin PH (2012) Newsvendor problems with sequentially revealed demand information. *Naval Research Logistics* 59(8):601–612.
- Swaminathan JM, Tayur SR (1998) Managing broader product lines through delayed differentiation using Vanilla Boxes. *Management Science* 44(12-part-2):S161–S172.
- Van Mieghem JA (1998) Investment strategies for flexible resources. *Management Science* 44(8):1071–1078.
- van Rooy JC (2010) *Logistical information sharing and collaboration between ASML and its supply chain*. Master’s thesis, Eindhoven University of Technology, URL <https://pure.tue.nl/ws/portalfiles/portal/46993635/691329-1.pdf>.
- Viswanathan S, Allampalli DG (2012) Four star industries Singapore–matching supply with demand, Nanyang Business School Case: ABCC-2009-005.
- Wang T, Atasu A, Kurtulus M (2012) Multiordering newsvendor with dynamic forecast evolution. *Manufacturing and Service Operations Management* 14(3):472–484.
- Zinn W (2019) A historical review of postponement research. *Journal of Business Logistics* 40(1):66–72.

## Online Appendix

### Proof of Theorem 1

At  $t = t_{n-1}$ , the expected profit can be formalized as a newsvendor problem:

$$G_{n-1}(Q_{n-1}, D_{n-1}) = \mathbb{E}_{D_n|D_{n-1}} \left( p \min(D_n, Q_{n-1}) \right) - c_{n-1} Q_{n-1}, \quad (17)$$

$$= (p - c_{n-1}) Q_{n-1} - p \int_0^{Q_{n-1}} (Q_{n-1} - D_n) f(D_n | D_{n-1}) \partial D_n, \quad (18)$$

where  $f(\cdot|\cdot)$  and  $F(\cdot|\cdot)$  denote conditional demand density and distribution functions, respectively. The optimal order quantity is obtained by:

$$\frac{\partial G_{n-1}}{\partial Q_{n-1}} = p(1 - F(Q_{n-1} | D_{n-1})) - c_{n-1} = 0. \quad (19)$$

With  $p > c_{n-1}$ ,  $G_{n-1}(\cdot, D_{n-1})$  is a concave function for any given  $D_{n-1}$ . Then,

$$V_{n-1}(Q_{n-2}, D_{n-1}) = \max_{Q_{n-1} \leq Q_{n-2}} \left\{ G_{n-1}(Q_{n-1}, D_{n-1}) \right\}. \quad (20)$$

For  $Q_{n-1}^* = \{Q_{n-1} | \partial G_{n-1} / \partial Q_{n-1} = 0\}$ ,

$$V_{n-1}(Q_{n-2}, D_{n-1}) = \begin{cases} G_{n-1}(Q_{n-2}, D_{n-1}) & \text{if } Q_{n-1}^* > Q_{n-2}, \\ G_{n-1}(Q_{n-1}^*, D_{n-1}) & \text{if } Q_{n-1}^* \leq Q_{n-2}. \end{cases} \quad (21)$$

$V_{n-1}(\cdot, D_{n-1})$  is a non-decreasing concave function due to the concavity of  $G_{n-1}(\cdot, D_{n-1})$ . Then,

$$G_{n-2}(Q_{n-2}, D_{n-2}) = \mathbb{E}_{D_{n-1}|D_{n-2}} \left( V_{n-1}(Q_{n-2}, D_{n-1}) \right) - c_{n-2} Q_{n-2}, \quad (22)$$

which is also concave because  $V_{n-1}(\cdot, D_{n-1})$  is concave. Then,  $G_i(\cdot, D_i)$  is a concave function (by induction)

for  $i \in \{0, 1, \dots, n-2\}$ , and the optimal policy is:

$$Q_i^* = \arg \max_{Q_i} \{G_i(Q_i, D_i)\}, \quad \forall i \in \{0, 1, \dots, n\}. \quad (23)$$

Suppose in period  $i+1$  for  $i \in \{0, 1, \dots, n-2\}$ ,

$$V_{i+1}(Q_i, D_{i+1}) = \begin{cases} G_{i+1}(Q_i, D_{i+1}) & \text{if } Q_{i+1}^* > Q_i, \\ G_{i+1}^*(D_{i+1}) & \text{if } Q_{i+1}^* \leq Q_i, \end{cases} \quad (24)$$

where  $G_{i+1}^*(D_{i+1}) = G_{i+1}(Q_{i+1}^*, D_{i+1})$ . Then,

$$G_i(Q_i, D_i) = \mathbb{E}_{D_{i+1}|D_i} [V_{i+1}(Q_i, D_{i+1})] - c_i Q_i, \quad (25)$$

$$= \int_{\bar{D}_{i+1}}^{+\infty} G_{i+1}(Q_i, D_{i+1}) f(D_{i+1} | D_i) \partial D_{i+1} + \int_0^{\bar{D}_{i+1}} G_{i+1}^*(D_{i+1}) f(D_{i+1} | D_i) \partial D_{i+1} - c_i Q_i, \quad (26)$$

where  $\bar{D}_{i+1}$  is the value of demand forecast at time  $i+1$  that makes the optimal order quantity equal to that of the previous period (i.e.,  $Q_{i+1}^* = Q_i$ ). Taking the first derivative, we obtain the following result:

$$\frac{\partial G_i}{\partial Q_i} = g_i(Q_i, D_i) = \int_{\bar{D}_{i+1}}^{+\infty} g_{i+1}(Q_i, D_{i+1}) f(D_{i+1}|D_i) \partial D_{i+1} - c_i = 0. \quad (27)$$

Using Equation (27), the optimal value of  $Q_i$  for  $i \in \{1, \dots, n-2\}$  can be found by backward induction.

The optimal value of  $Q_{n-1}$  is given by Equation (19). Combining Equations (19) and (27), the optimal value of  $Q_{n-2}$  can be calculated by:

$$\begin{aligned} g_{n-2}(Q_{n-2}, D_{n-2}) &= \int_{\bar{D}_{n-1}}^{+\infty} (pPr(D_n > Q_{n-2}|D_{n-1}) - c_{n-1}) f_{n-1}(D_{n-1}|D_{n-2}) \partial D_{n-1} - c_{n-2}, \\ &= pPr(D_n > Q_{n-2}, D_{n-1} > \bar{D}_{n-1}) - c_{n-1}Pr(D_{n-1} > \bar{D}_{n-1}) - c_{n-2}. \end{aligned} \quad (28)$$

By induction, we obtain for  $i \in \{0, 1, \dots, n-1\}$  the following result:

$$\begin{aligned} g_i(Q_i, D_i) &= pPr(D_n > Q_i, \mathbf{D}_{\{i+1, n-1\}} > \bar{\mathbf{D}}_{\{i+1, n-1\}}) - c_{n-1}Pr(\mathbf{D}_{\{i+1, n-1\}} > \bar{\mathbf{D}}_{\{i+1, n-1\}}) \\ &\quad - c_{n-2}Pr(\mathbf{D}_{\{i+1, n-2\}} > \bar{\mathbf{D}}_{\{i+1, n-2\}}) - \dots - c_{i+1}Pr(D_{i+1} > \bar{D}_{i+1}) - c_i, \end{aligned} \quad (29)$$

where  $\mathbf{D}_{\{i+1, n-1\}}$  is a vector denoting demand forecasts from period  $i+1$  to  $n-1$ . Then, the optimal order quantity in each period can be found by  $q_i = \min(Q_{i-1}, Q_i^*)$  such that  $Q_i^* = \{Q_i | g_i(Q_i, D_i) = 0\}$ .

### Proof of Proposition 1

The proof is straightforward from the final result of the proof of Theorem 1:  $q_i = \min(Q_{i-1}, Q_i^*)$ . If the demand forecast in period  $i$  turns out to be high, the order quantity is constrained by  $q_{i-1}$ . Otherwise,  $q_i < q_{i-1}$ . Therefore,  $q_0 > \mathbb{E}[q_1|D_0] > \dots > \mathbb{E}[q_{n-1}|D_0]$ .

### Proof of Proposition 2

**Part A:** For  $j \in \{i, \dots, n-1\}$ , suppose  $c_j$  is increased by  $\Delta c_j$ , making the cost of processing the  $j^{th}$  operation equal to  $c_j + \Delta c_j$ . Suppose  $Q'_i = Q_i$  and  $g_i(Q_i, D_i) = 0$ . Then,

$$g_i(Q'_i, D_i) - g_i(Q_i, D_i) = g_i(Q'_i, D_i) = -\Delta c_j Pr(\mathbf{D}_{\{i+1, j\}} > \bar{\mathbf{D}}_{\{i+1, j\}}). \quad (30)$$

If  $Q'_i = Q_i$ , the term  $g_i(Q'_i, D_i)$  fails to be equal to zero after the cost increase, meaning that setting  $Q'_i = Q_i$  does not optimize the ordering decision anymore. It follows from Equation (30) that the ordering decision after the cost increase is optimized for  $Q'_i < Q_i$  such that  $g_i(Q'_i, D_i) = 0$  for  $Q'_i < Q_i$ . Thus,  $q'_i \prec q_i$  for  $i \in \{0, \dots, j\}$ , meaning that an increase in the cost of an operation leads to a reduction of order quantities

at upstream echelons. Because the downstream order quantities are constrained by the upstream ones such that  $Q_0 \geq Q_1 \geq \dots \geq Q_{n-1}$ , such a reduction of upstream order quantities has a cascading impact of reducing also the downstream order quantities. Thus,  $q'_i \prec q_i$  for  $i \in \{0, \dots, n-1\}$ .

**Part B:** Given  $c_0 = \dots = c_{i-1} = c_{i+1} = \dots = c_{n-1} = c_{fixed} < c_i$ , the expected profit at  $t = t_0$  is written as follows:

$$G_0(Q_0^*, D_0 | i = j) = \int_0^{Q_0^*} g_0(Q_0, D_0) dQ_0, \quad (31)$$

where  $Q_0^*$  is the optimal order quantity at  $t_0$  when  $i = j$  for  $j \in \{0, \dots, n-2\}$ . Thus,  $G_0(Q_0^*, D_0 | i = j+1)$  gives a lower bound for the expected profit for  $i = j+1$ . Combining this expression with Equation (9) yields the following result:

$$\begin{aligned} G_0(Q_0^*, D_0 | i = j+1) - G_0(Q_0^*, D_0 | i = j) = & - \int_0^{Q_0^*} (c_i - c_{fixed}) Pr(\mathbf{D}_{\{1, i+1\}} > \bar{\mathbf{D}}_{\{1, i+1\}}) dQ_0 \\ & + \int_0^{Q_0^*} (c_i - c_{fixed}) Pr(\mathbf{D}_{\{1, i\}} > \bar{\mathbf{D}}_{\{1, i\}}) dQ_0. \end{aligned} \quad (32)$$

Equation (32) has always a non-negative value because  $Pr(\mathbf{D}_{\{1, i\}} > \bar{\mathbf{D}}_{\{1, i\}}) \geq Pr(\mathbf{D}_{\{1, i+1\}} > \bar{\mathbf{D}}_{\{1, i+1\}})$ . Thus, swapping a high-cost operation with the downstream adjacent one leads to higher profits. It is straightforward by induction that swapping the operation  $i$  with any operation from the set  $\{i+1, \dots, n-1\}$  increases the profit.

### Proof of Proposition 3

If the lead time of operation  $i$  is reduced by  $\Delta t$ , the starting times for the first  $i+1$  operations are updated as follows:

$$t_0 + \Delta t = t_1 + \Delta t = \dots = t_j + \Delta t = \dots = t_i + \Delta t. \quad (33)$$

Then, ordering decisions for the first  $i+1$  operations are made after a delay of  $\Delta t$ . Delaying the ordering decisions leads to improved demand accuracy for the first  $i+1$  decisions as given by Equation (2), therefore increasing the expected profit.

If the lead time of operation  $j$  is reduced by  $\Delta t$ , the starting times for the first  $j+1$  operations are likewise delayed for  $\Delta t$ . Reducing the lead-time of  $i$ , compared to that of  $j$  for  $i > j$ , makes it possible to delay additionally the decision epochs for  $j+1, j+2, \dots, i$ , resulting in a profit increase that is higher than what can be achieved by reducing the lead time of  $j$ .

## Proof of Theorem 2

Let  $D_{n,r}^j \geq 0$  denote a realization of  $D_n^j$  such that  $r \in \mathcal{S}$ , where  $\mathcal{S} = \{1, 2, \dots\}$  is defined as a large finite set of positive integers. The values of  $D_{n,r}^j \forall r \in \mathcal{S}$  constitute the set of all possible demand realizations. We use  $\mathcal{W}_j^r$  to denote the sales value for a demand realization of  $D_{n,r}^j$  such that  $\mathcal{W}_j^r = \mathcal{W}_j(Q_{n-1}^j, D_{n,r}^j)$ . Then, the SP model (13)–(15) can be written as a large-scale LP model:

$$\begin{aligned} \text{Maximize } z = & \sum_{j \in \Theta_{n-1}} \left( p_j \sum_{r \in \mathcal{S}} \text{Pr}(\mathcal{W}_j^r) \mathcal{W}_j^r - c_{n-1}^j Q_{n-1}^j \right) \end{aligned} \quad (34)$$

subject to:

$$\mathcal{W}_j^r - Q_{n-1}^j \leq 0, \quad \forall j \in \Theta_{n-1}, \quad \forall r \in \mathcal{S}, \quad (35)$$

$$\mathcal{W}_j^r \leq D_{n,r}^j, \quad \forall j \in \Theta_{n-1}, \quad \forall r \in \mathcal{S}, \quad (36)$$

$$\sum_{j \in \Theta_{n-1}^k} Q_{n-1}^j \leq Q_{n-2}^k, \quad \forall k \in \Theta_{n-2}, \quad (37)$$

$$Q_{n-1}^j \geq 0, \quad \mathcal{W}_j^r \geq 0, \quad \forall j \in \Theta_{n-1}, \quad \forall r \in \mathcal{S}. \quad (38)$$

We remark that we add Constraints (35) and (36) to satisfy the condition  $W_j = \min\{Q_{n-1}^j, D_n^j\}$  for the optimal solution. Then, the dual problem is:

$$\begin{aligned} \text{Minimize } w = & \sum_{j \in \Theta_{n-1}} \sum_{r \in \mathcal{S}} \beta_{j,r} D_{n,r}^j + \sum_{k \in \Theta_{n-2}} \gamma_k Q_{n-2}^k \end{aligned} \quad (39)$$

subject to:

$$\lambda_{j,r} + \beta_{j,r} \geq \text{Pr}(\mathcal{W}_j^r) p_j, \quad \forall j \in \Theta_{n-1}, \quad r \in \mathcal{S}, \quad (40)$$

$$\sum_{r \in \mathcal{S}} \lambda_{j,r} - \gamma_k \leq c_{n-1}^j, \quad \forall j \in \Theta_{n-1}^k, \quad k \in \Theta_{n-2}, \quad r \in \mathcal{S}, \quad (41)$$

$$\lambda_{j,r} \geq 0, \quad \beta_{j,r} \geq 0, \quad \gamma_k \geq 0, \quad \forall j \in \Theta_{n-1}^k, \quad k \in \Theta_{n-2}, \quad r \in \mathcal{S}. \quad (42)$$

The values of  $\lambda_{j,r}$  and  $\beta_{j,r}$  for each  $j \in \Theta_{n-1}$  are found by the parametric analysis:

1.  $\lambda_{j,r} = 0$  and  $\beta_{j,r} = \text{Pr}(\mathcal{W}_j^r) p_j$  for  $j \in \Theta_{n-1}$  when  $D_{n,r}^j \leq Q_{n-1}^j$ .
2. Likewise,  $\lambda_{j,r} = \text{Pr}(\mathcal{W}_j^r) p_j$  and  $\beta_{j,r} = 0$  for  $j \in \Theta_{n-1}$  when  $D_{n,r}^j > Q_{n-1}^j$ .

Then, the constraint (41) is written as follows:

$$p_j \text{Pr}(D_n^j > Q_{n-1}^j) - \gamma_k \leq c_{n-1}^j, \quad \forall j \in \Theta_{n-1}. \quad (43)$$

We set a value for the dual variable  $\gamma_k$  for  $k \in \Theta_{n-2}$  such that:

$$\gamma_k = (p_j - c_{n-1}^j) - p_j \text{Pr}(D_n^j \leq Q_{n-1}^j) = g_{n-1}^j(Q_{n-1}^j, D_{n-1}^j), \quad \forall j \in \Theta_{n-1}^k. \quad (44)$$

Then, the objective function value for the dual problem becomes:

$$w = \sum_{j \in \Theta_{n-1}} \left[ (p_j - c_{n-1}^j) Q_{n-1}^j - p_j \int_0^{Q_{n-1}^j} (Q_{n-1}^j - D_n^j) f^j(D_n^j) \partial D_n^j \right]. \quad (45)$$

Equation (45) is equivalent to the solution of the primal problem for the feasible  $Q_{n-1}^j$  values. It follows from the strong theorem of duality that the optimal solution satisfies Equation (44). Therefore, we have the following conditions of optimality:

$$\gamma_k = g_{n-1}^j(Q_{n-1}^j, D_{n-1}^j) \geq 0, \quad \forall j \in \Theta_{n-1}^k, \quad k \in \Theta_{n-2} \quad (46)$$

$$\sum_{j \in \Theta_{n-1}^k} Q_{n-1}^j \leq Q_{n-2}^k, \quad \forall k \in \Theta_{n-2}. \quad (47)$$

If the constraint (37) is not binding for a given  $k \in \Theta_{n-2}$ , the dual variable  $\gamma_k$  becomes zero. In this case, the optimal solution reduces to the solution of  $|\Theta_{n-1}^k|$  independent newsvendor problems in the last period—that is, the order quantity for each product in the set  $\Theta_{n-1}^k$  can be found solving an unconstrained newsvendor problem. Otherwise, the optimal solution exists at the point where the marginal value of producing one unit is the same for all products in the set  $\Theta_{n-1}^k$ .

In period  $i \in \{1, 2, \dots, n-2\}$ , the optimization problem is written as follows:

$$\text{Maximize } z = \sum_{j \in \Theta_i} G_i^{\Upsilon_i^j}(Q_i^j, \mathbf{D}_i^{\Upsilon_i^j}) \quad (48)$$

subject to:

$$\sum_{j \in \Theta_i^k} Q_i^j \leq Q_{i-1}^k, \quad 0 \leq Q_i^j, \quad \forall k \in \Theta_{i-1}, \quad \forall j \in \Theta_i^k. \quad (49)$$

The term  $G_i^{\Upsilon_i^j}(Q_i^j, \mathbf{D}_i^{\Upsilon_i^j})$  is the total expected profit generated from all the products in the set  $\Upsilon_i^j$ , and  $\mathbf{D}_i^{\Upsilon_i^j}$  is the vector of demand forecasts of the products in  $\Upsilon_i^j$  at time  $t_i$ . We recall that  $\Upsilon_i^j$  is the set of end products sold in the market whose availability depends on the order quantity decision of  $Q_i^j$ . We will discuss the derivation of  $G_i^{\Upsilon_i^j}(Q_i^j, \mathbf{D}_i^{\Upsilon_i^j})$  in detail below.

The dual problem (48)–(49) is formulated as:

$$\text{Minimize } w = \sum_{k \in \Theta_{i-1}} \gamma_k Q_{i-1}^k \quad (50)$$

subject to:

$$g_i^{\Upsilon_i^j}(Q_i^j, \mathbf{D}_i^{\Upsilon_i^j}) \leq \gamma_k, \quad \forall j \in \Theta_i^k, \quad \forall k \in \Theta_{i-1}, \quad (51)$$

with  $\partial G_i^{\Upsilon_i^j}(Q_i^j, \mathbf{D}_i^{\Upsilon_i^j}) / \partial Q_i^j = g_i^{\Upsilon_i^j}(Q_i^j, \mathbf{D}_i^{\Upsilon_i^j})$ . Then, the optimal solution in each period  $i \in \{1, \dots, n-2\}$  satisfies the following equations:

$$\gamma_k = g_i^{\Upsilon_i^j}(Q_i^j, \mathbf{D}_i^{\Upsilon_i^j}) \geq 0, \quad \forall j \in \Theta_i^k, \quad k \in \Theta_{i-1} \quad (52)$$

$$\sum_{j \in \Theta_i^k} Q_i^j \leq Q_{i-1}^k, \quad \forall k \in \Theta_{i-1}. \quad (53)$$

Following the steps similar to the proof of Theorem 1, we obtain the following expression for  $t = t_{n-2}$ :

$$\begin{aligned} G_{n-2}^{\Upsilon_{n-2}^j}(Q_{n-2}^j, \mathbf{D}_{n-2}^{\Upsilon_{n-2}^j}) &= \int_{\overline{D}_{n-1}^{\Upsilon_{n-2}^j}}^{+\infty} G_{n-1}^{\Upsilon_{n-2}^j}(Q_{n-2}^j, D_{n-1}^{\Upsilon_{n-2}^j}) f(D_{n-1}^{\Upsilon_{n-2}^j} | D_{n-2}^{\Upsilon_{n-2}^j}) \partial D_{n-1}^{\Upsilon_{n-2}^j} \\ &\quad + \int_0^{\overline{D}_{n-1}^{\Upsilon_{n-2}^j}} G_{n-1}^{*\Upsilon_{n-2}^j}(Q_{n-2}^j) f(D_{n-1}^{\Upsilon_{n-2}^j} | D_{n-2}^{\Upsilon_{n-2}^j}) \partial D_{n-1}^{\Upsilon_{n-2}^j} - c_{n-2}^j Q_{n-2}^j, \end{aligned} \quad (54)$$

where  $D_{n-1}^{\Upsilon_{n-2}^j}$  is a random variable denoting the sum of demand forecasts of the items in the set  $\Upsilon_{n-2}^j$  at  $t = t_{n-1}$  (i.e.,  $\sum_{k \in \Upsilon_{n-2}^j} D_{n-1}^k$ ).  $\overline{D}_{n-1}^{\Upsilon_{n-2}^j}$  is the value of demand forecast at  $t = t_{n-1}$  that makes the optimal order quantity at  $t_{n-1}$  equal to that of the previous period (i.e.,  $Q_{n-2}^j$ ). Then,

$$g_{n-2}^{\Upsilon_{n-2}^j}(Q_{n-2}^j, \mathbf{D}_{n-2}^{\Upsilon_{n-2}^j}) = \int_{\overline{D}_{n-1}^{\Upsilon_{n-2}^j}}^{+\infty} g_{n-1}^{\Upsilon_{n-2}^j}(Q_{n-2}^j, D_{n-1}^{\Upsilon_{n-2}^j}) f(D_{n-1}^{\Upsilon_{n-2}^j} | D_{n-2}^{\Upsilon_{n-2}^j}) \partial D_{n-1}^{\Upsilon_{n-2}^j} - c_{n-2}^j, \quad (55)$$

where  $g_{n-1}^{\Upsilon_{n-2}^j}(Q_{n-2}^j, D_{n-1}^{\Upsilon_{n-2}^j}) = g_{n-1}^k(Q_{n-2}^j, D_{n-1}^k)$  for any  $k \in \Upsilon_{n-2}^j$  as given by Equation (46). Using the last expression, we obtain the following result by induction:

$$g_i^{\Upsilon_i^j}(Q_i^j, \mathbf{D}_i^{\Upsilon_i^j}) = \int_{\overline{D}_{i+1}^{\Upsilon_i^j}}^{+\infty} g_{i+1}^{\Upsilon_i^j}(Q_{i+1}^j, D_{i+1}^{\Upsilon_i^j}) f(D_{i+1}^{\Upsilon_i^j} | D_i^{\Upsilon_i^j}) \partial D_{i+1}^{\Upsilon_i^j} - c_i^j = \gamma_k, \quad \forall j \in \Theta_i^k. \quad (56)$$

If the constraint (49) for a given  $k$  is not binding, the dual variable  $\gamma_k$  becomes zero. In this case, the optimal solution reduces to the solution of  $|\Theta_i^k|$  unconstrained problems using Equation (9). Let  $Q_i^{j*}$  denote the order quantity for  $j \in \Theta_i^k$  and  $k \in \Theta_{i-1}$  satisfying Equation (9) and  $\widehat{Q}_i^j$  denote the order quantity satisfying Equation (56). Then, the optimal order quantity is found by a state-dependent base-stock policy:

$$q_i^j = \begin{cases} Q_i^{j*} & \text{if } \sum_{j \in \Theta_i^k} Q_i^{j*} < Q_{i-1}^k, \\ \widehat{Q}_i^j & \text{if } \sum_{j \in \Theta_i^k} Q_i^{j*} \geq Q_{i-1}^k. \end{cases} \quad (57)$$



### Proof of Proposition 4

Suppose product proliferation occurs once at time  $t_i$  along the supply chain. The mathematical model given by (48)–(49) is then rewritten as follows with an objective function of maximizing the expected profit at time  $t_i$ :

$$\begin{aligned} \text{Maximize } z = & \sum_{j \in \Theta_i} G_i^j(Q_i^j, \mathbf{D}_i^{\Theta_i}) \\ \text{subject to: } & \end{aligned} \quad (58)$$

$$\sum_{j \in \Theta_i^k} Q_i^j \leq Q_{i-1}^k, \quad \forall k \in \Theta_{i-1}. \quad (59)$$

Then, the value of postponing the point of the proliferation is calculated by  $\partial z / \partial t_i$ .

As the next step, we fix  $c_i^j = c_{i+1}^j = \dots = c_{n-1}^j = 0 \ \forall j \in \Theta_i$  and analyze the impact of incrementally increasing  $c_i^j$  on the expected profit. Since  $c_i^j = c_{i+1}^j = \dots = c_{n-1}^j = 0$ , manufacturer orders the maximum amount in all the remaining periods (i.e.,  $\{i, i+1, \dots, n-1\}$ ) such that:

$$Q_{i-1}^j = Q_i^j = \dots = Q_{n-1}^j. \quad (60)$$

And,

$$G_i^j(Q_i^{j*}, D_i^j) = G_i^j(Q_{i-1}^j, D_i^j) \quad \forall j \in \Theta_i, \quad (61)$$

Given that the order quantity at  $t_i$  is set equal to  $Q_{i-1}$ , the expected profit is not affected by the ordering decision and we have the following relationship:

$$\partial G_i^j(Q_i^j, D_i^j) / \partial t_i = 0 \quad \rightarrow \quad \partial z / \partial t_i = 0. \quad (62)$$

Having set  $c_i^j = c_{i+1}^j = \dots = c_{n-1}^j = 0 \ \forall j \in \Theta_i$  implies that the constraint (59) is binding. Now, we slightly increase  $c_i^j$  such that  $c_i^j > 0$  and  $c_{i+1}^j = \dots = c_{n-1}^j = 0 \ \forall j \in \Theta_i$ . Then, we obtain:

$$G_i^j(Q_i^j, D_i^j) = (p_j - c_i^j)Q_i^j - p_j \int_0^{Q_i^j} (Q_i^j - D_n^j) f(D_n^j | D_i^j) \partial D_n^j, \quad (63)$$

$$\begin{aligned} &= (p_j - c_i^j)Q_i^j - p_j Q_i^j \Phi\left(\frac{\ln(Q_i^j / D_i^j)}{\sigma \sqrt{t_n - t_i}} + \sigma \sqrt{t_n - t_i} / 2\right) \\ &\quad + p_j D_i^j \Phi\left(\frac{\ln(Q_i^j / D_i^j)}{\sigma \sqrt{t_n - t_i}} - \sigma \sqrt{t_n - t_i} / 2\right). \end{aligned} \quad (64)$$

From Equation (64), we obtain:

$$\partial G_i^j(Q_i^j, D_i^j) / \partial t_i > 0 \quad \rightarrow \quad \partial z / \partial t_i > 0. \quad (65)$$

Therefore,  $\partial z / \partial t_i$  increases as  $c_i$  increases, which completes the proof of Proposition 4.

### Proof of Proposition 5

The proof has a similar flavor to the proof of Proposition 4. Without loss of generality we consider a simple structure with two ordering decisions taking place at  $t_0$  and  $t_1$ , and sales occur at  $t_2$ . Suppose product proliferation occurs once at time  $t_1$ . Then, the expected profit at time  $t_1$  is:

$$\text{Maximize}_{Q_1^j, \forall j \in \Theta_1} z = \sum_{j \in \Theta_1} G_1^j(Q_1^j, \mathbf{D}_1^{\Theta_1}) \quad (66)$$

subject to:

$$\sum_{j \in \Theta_1^k} Q_1^j \leq Q_0. \quad (67)$$

The  $Q_0$  term has a finite value for  $c_0 > 0$ . Suppose we reduce  $c_1^j$  to zero  $\forall j \in \Theta_1$  and then swap the second operation with the first one such that the proliferation now occurs at  $t_0$ . Then, the expected profit at time  $t_1$  becomes:

$$\text{Maximize}_{Q_1^j, \forall j \in \Theta_1} z = \sum_{j \in \Theta_1} G_1^j(Q_1^j, \mathbf{D}_1^{\Theta_1}) \quad (68)$$

subject to:

$$Q_1^j \leq Q_0^j, \quad \forall j \in \Theta_1. \quad (69)$$

Given that the cost of first operation after the swap is zero, we have  $Q_0^j = \infty \forall j \in \Theta_0$ . The problem (68)–(69) reduces to the unconstrained version of the problem (66)–(67). Therefore, the expected profit at time  $t_1$  increases after the swap (i.e., expediting the proliferation).

If  $c_0 = c_1^j \forall j \in \Theta_1$  in the initial case, then swapping the operations leads to a decrease in profits. It results from the fact that such a swapping of the operations prevents the manufacturer from getting the benefits of inventory pooling at time  $t_0$  when the costs of the operations are the same. Therefore, a threshold value exists between 0 and  $c_1$ .

### Proof of Proposition 6

The proof follows directly from Proposition 3.

## Recent ESMT Working Papers

	ESMT No.
<b>The impact of EU cartel policy reforms on the timing of settlements in private follow-on damages disputes: An empirical assessment of cases from 2001 to 2015</b>	19-03
<p>Hans W. Friederiszick, ESMT Berlin and E.CA Economics</p> <p>Linda Gratz, E.CA Economics</p> <p>Michael Rauber, E.CA Economics</p>	
<b>Marginality, dividends, and the value in games with externalities</b>	19-01
<p>Frank Huettnner, ESMT Berlin</p> <p>André Casajus, HHL Leipzig Graduate School of Management</p>	
<b>Consumer choice under limited attention when alternatives have different information costs</b>	16-04 (R3)
<p>Frank Huettnner, ESMT Berlin</p> <p>Tamer Boyaci, ESMT Berlin</p> <p>Yalçın Akçay, Melbourne Business School</p>	
<b>Opaque queues: Service systems with rationally inattentive customers</b>	18-04
<p>Caner Canyakmaz, ESMT Berlin</p> <p>Tamer Boyaci, ESMT Berlin</p>	
<b>The Coleman-Shapley-index: Being decisive within the coalition of the interested</b>	18-03
<p>André Casajus, HHL Leipzig Graduate School of Management</p> <p>Frank Huettnner, ESMT Berlin</p>	