

July 22, 2019

ESMT Working Paper 19-02

Beyond retail stores: Managing product proliferation along the supply chain

Işık Biçer, Rotterdam School of Management, Erasmus University

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

Tamer Boyacı, ESMT Berlin

Copyright 2019 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) 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

Rotterdam School of Management, Erasmus University, Rotterdam 3000 DR, The Netherlands, bicer@rsm.nl,

Florian Lückner

Cass Business School, City, University of London, London EC1Y, UK, ,

Tamer Boyacı

ESMT Berlin, Schlossplatz 1, 10178, Berlin, Germany, ,

Problem definition. Product proliferation occurs in supply chains to produce diverse product portfolios from a limited variety of raw materials. In such a setting, upstream operational decisions are often plagued with high demand uncertainty, whereas downstream decisions are exposed to high product variety. We compare two operational strategies (i.e., lead-time reduction and process redesign), both of which improve the flexibility of supply chains by delaying product proliferation. **Academic/Practical Relevance.** The well-known examples of Zara and Benetton illustrate the usefulness of these two operational strategies in managing product proliferation. We develop an integrated framework to evaluate the impact of changes in supply chain structure, cost and lead time of each echelon on profits. Utilizing this framework, we compare the relative value of the two strategies that have only been investigated in isolation in the literature. **Methodology.** We use the multiplicative martingale model of forecast evolution (m-MMFE) to characterize the demand-updating process, and develop a dynamic optimization model to determine the optimal order quantities at different echelons. We compare alternative operational strategies analytically as well as numerically using Markovian sampling. **Results.** 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 a complementary strategy of deferring high cost operations to later stages and then focusing on reducing lead times of those scheduled after the proliferation substantially increases profits. **Managerial Implications.** We develop a decision typology that points out 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. Managing a diverse product portfolio has a significant effect on increasing demand-supply mismatch costs (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 production (Atali and Özer 2012, Kouvelis and Tian 2014, Dong et al. 2018), processing (Boyabatlı et al. 2011, Boyabatlı 2015), and logistics (Gao et al. 2014) operations. 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

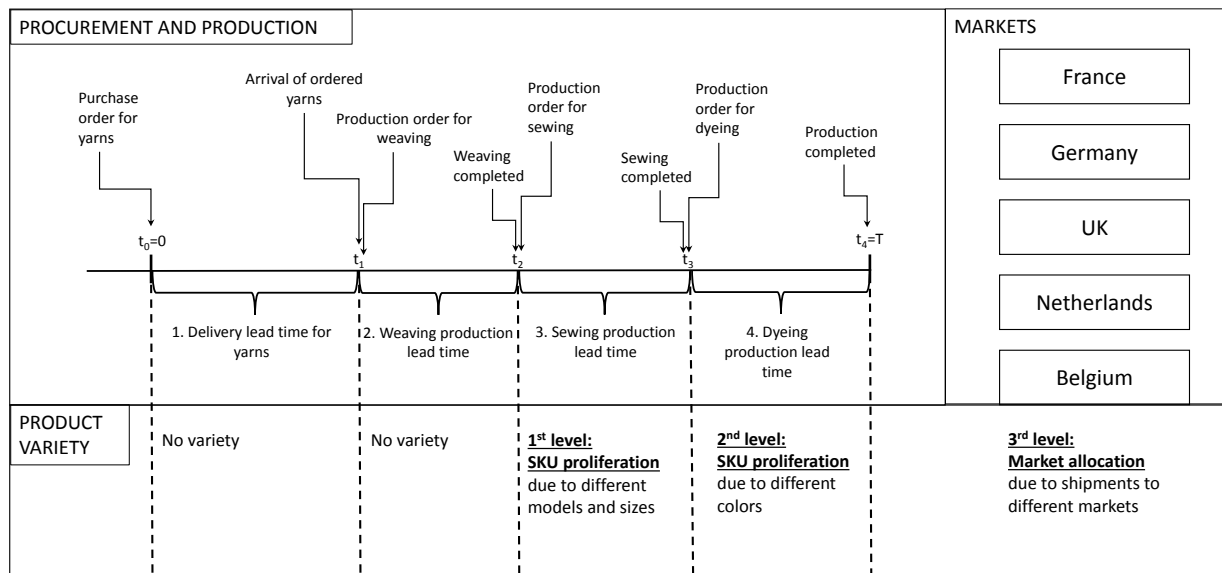


Figure 1 Product proliferation in the fashion-apparel industry

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.). As a matter of fact, the primary motivation for this paper stems from authors’ involvement with a leading global manufacturer of composites used mainly by tyre producers in the automotive industry.¹ 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

¹ A letter from this manufacturer with more than \$700 million annual revenue is attached to the cover letter.

MANUFACTURING IN THE PROCESS INDUSTRY					
<p>The diagram illustrates the manufacturing timeline from $t_0=0$ to $t_4=T$. Key milestones include: <ul style="list-style-type: none"> $t_0=0$: Production order for polymer materials. t_1: Production completed (polymer), Production order for fabrication. t_2: Production completed (first-level), Production order for fabrication. t_3: Production completed (second-level), Production order for dipping. $t_4=T$: Production completed. The timeline is divided into four lead time segments: <ol style="list-style-type: none"> 1. Polymer processing lead time 2. First-level fabrication lead time 3. Second-level fabrication lead time 4. Dipping lead time </p>					
PRODUCT VARIETY					
No variety		1st level: SKU proliferation due to different twisting	2nd level: SKU proliferation due to different weaving	3rd level: SKU proliferation due to different chemical blending and dipping durations	

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. Manufacturers can deploy inventory pooling for upstream orders to mitigate uncertainty, but this is not possible for downstream orders, post-proliferation. Due to high demand uncertainty at upstream echelons and high product variety at downstream echelons, mismatches between supply and demand are inevitable, which in turn have negative impact on long-term profits.

Delaying differentiation is an effective strategy for improving responsiveness in supply chains with proliferation (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 “postponement” strategy, many other companies followed Benetton’s lead (Parsons and Graves 2005, Viswanathan and Allampalli 2012, Kouvelis and Tian 2014). 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 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 Gallien 2010, Caro and Martínez-de Albéniz 2015).

Although both approaches aim to improve the responsiveness of supply chains, they have been investigated in isolation in the literature. In the absence of an integrated framework that models

supply chain structures with proliferation under an evolutionary demand environment, it is difficult to compare lead-time reduction with process redesign. For a holistic understanding, such a comparison needs to cover both the case where the firm has the process flexibility to adjust the sequence of operations and the case where it does not. We aim to address this gap in this paper. To this end, we first develop a framework which enables systematic evaluation of the impact of changes in the supply chain with respect to point(s) of proliferation, costs and lead times. We then utilize this framework to delineate the conditions that render process redesign and/or lead time reduction most valuable. We derive practical insights and recommendations to improve supply chain responsiveness in different settings.

Our paper makes three important contributions to the extant literature. First, from a modeling perspective, we develop an analytical framework for dynamically optimizing inventory/ordering quantities in (multi-product) supply chains with product proliferation. This framework extends existing inventory models in the literature that incorporates forecast evolution (Wang et al. 2012, Oh and Özer 2013, 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 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. As a result, the value of postponing the point of 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.

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 (delaying differentiation by swapping with a downstream operation may not be profitable for example). 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 paper is mainly related to the OM literature focusing on the value of supply chain responsiveness. Prominent papers in the literature have explored how the value of responsiveness is shaped by information asymmetry (Lutze and Özer 2008), strategic customer behaviour (Cachon and Swinney 2011), competition among buyers (Caro and Martínez-de Albéniz 2010), demand-forecast evolution (Wang et al. 2012, Oh and Özer 2013, Biçer and Seifert 2017, Biçer et al. 2018), multiple orders (Martínez-de Albéniz 2011, Song and Zipkin 2012, Calvo and Martínez-de Albéniz 2015, Cao and So 2016). A substantial amount of the work in the literature focuses on the fast fashion industry. We refer the reader to Caro and Martínez-de Albéniz (2015) for an overview of the fast-fashion business model, its industry outlook, and untapped problems.

Within this broad literature, from a modeling perspective, the closest works 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. Our model is more general because we optimize ordering decisions in a multi-echelon setting such that 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.

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 (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 framework incorporates

both the inventory pooling and improved forecast accuracy motivations for delaying differentiation in an integrative manner. The resulting model lends itself for systematic comparison of alternative operational strategies that have only been explored in isolation in the literature.

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 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.

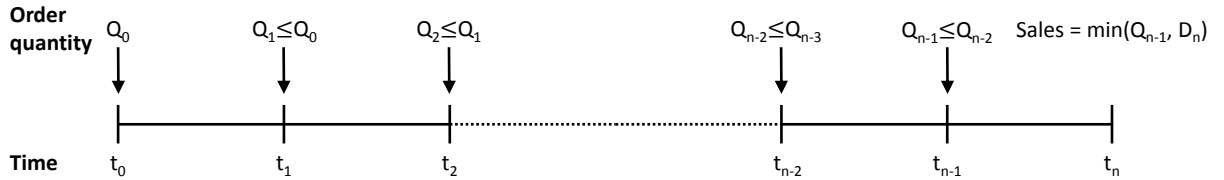


Figure 3 Timeline of ordering decisions for a single-product model without product proliferation

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, Oh and Özer 2013, Biçer et al. 2018). Let $(\Omega, \mathcal{F}, \mathbb{P})$ to denote a

filtered probability space on which demand forecasts follow the m-MMFE process. Demand process on $(\Omega, \mathcal{F}, \mathbb{P})$ is adapted by filtration $(\mathcal{F}_t)_{t \geq t_0} \subseteq \mathcal{F}$ such that:

- The actual demand value conditional on the demand forecast (i.e., $D_n|D_i$ for $i \in \{0, \dots, n-1\}$) is square integrable so that $D_n|D_i$ has a finite variance.
- $\mathcal{F}_{t_i} \subseteq \mathcal{F}_{t_{i+1}}$ for $i \in \{0, \dots, n-1\}$ such that forecasters do not lose information over time.
- $\mathcal{F} \subseteq \mathcal{F}_{t_n}$, implying that demand realized at time t_n is immediately known by the forecasters.

These properties are considered to be the features of a good forecast-updating practice, which are also satisfied by the m-MMFE (Oh and Özer 2013). We remark that our aim here is not to develop a new forecasting model but rather to model the evolution of demand forecasts for a manufacturer with a good forecasting practice.

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 ε_i 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:

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

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 σ 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.

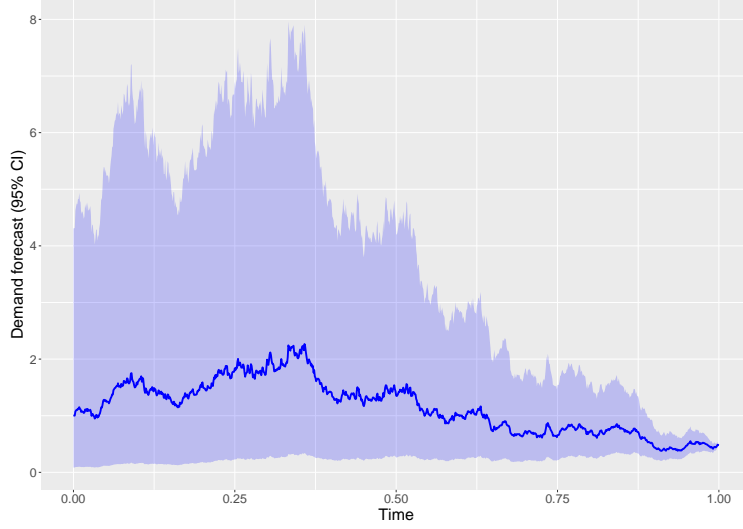


Figure 4 A sample path of the demand forecast according to the m-MMFE with 95% two-sided confidence intervals.

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 use κ_i to denote the accumulated unit cost from t_0 until t_{i+1} such that:

$$\kappa_i = \sum_{j=0}^i c_j, \quad \forall i \in \{1, \dots, n-1\}. \quad (3)$$

Thus, κ_{n-1} is the total cost to produce one unit of the end product.

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:

$$V_{n-1}(Q_{n-2}, D_{n-1}) = \max_{Q_{n-1} \leq Q_{n-2}} \left\{ \mathbb{E}_{D_n | D_{n-1}} \left[p \min(D_n, Q_{n-1}) \right] - \kappa_{n-1} Q_{n-1} \right\}. \quad (4)$$

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:

$$V_i(Q_{i-1}, D_i) = \max_{Q_i \leq Q_{i-1}} \left\{ \mathbb{E}_{D_{i+1}|D_i} \left[V_{i+1}(Q_i, D_{i+1}) - c_i \mathbb{E}_{Q_{i+1}|Q_i} [Q_i - Q_{i+1}] \right] \right\}. \quad (5)$$

The order quantity at t_0 is not constrained, so we set $Q_{-1} = +\infty$. Let the functions to be maximized in Equations (4) and (5) be denoted respectively as:

$$G_{n-1}(Q_{n-1}, D_{n-1}) = \mathbb{E}_{D_n|D_{n-1}} \left[p \min(D_n, Q_{n-1}) \right] - \kappa_{n-1} Q_{n-1}, \quad (6)$$

$$G_i(Q_i, D_i) = \mathbb{E}_{D_{i+1}|D_i} \left[V_{i+1}(Q_i, D_{i+1}) - c_i \mathbb{E}_{Q_{i+1}|Q_i} [Q_i - Q_{i+1}] \right], \quad (7)$$

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 (5) 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.²

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

$$q_i = \min(Q_{i-1}, Q_i^*), \quad (8)$$

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:

$$Q_i^* = \{Q_i \mid g_i(Q_i, D_i) = 0\}, \quad (9)$$

$$\begin{aligned} g_i(Q_i, D_i) = & p \Pr(D_n > Q_i, \mathbf{D}_{\{i+1, n-1\}} > \bar{\mathbf{D}}_{\{i+1, n-1\}}) - c_{n-1} \Pr(\mathbf{D}_{\{i+1, n-1\}} > \bar{\mathbf{D}}_{\{i+1, n-1\}}) \\ & - c_{n-2} \Pr(\mathbf{D}_{\{i+1, n-2\}} > \bar{\mathbf{D}}_{\{i+1, n-2\}}) - \dots - c_{i+1} \Pr(D_{i+1} > \bar{D}_{i+1}) - \kappa_i = 0 \end{aligned} \quad (10)$$

² The proofs of all results are presented in our on-line appendix.

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 (10) reduces to the newsvendor solution for $i = n-1$ such that:

$$g_{n-1}(Q_{n-1}, D_{n-1}) = pPr(D_n > Q_i) - \kappa_{n-1} = 0. \quad (11)$$

For $i < n-1$, the solution is still in spirit the newsvendor solution. The first term of the right-hand side of Equation (10) 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 (10) 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 costs from t_0 until t_{i+1} , which amounts to κ_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 Δt .*

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 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 Θ_i to denote the set of all components produced at echelon $i \in \{1, \dots, n\}$ at time t_i . We further

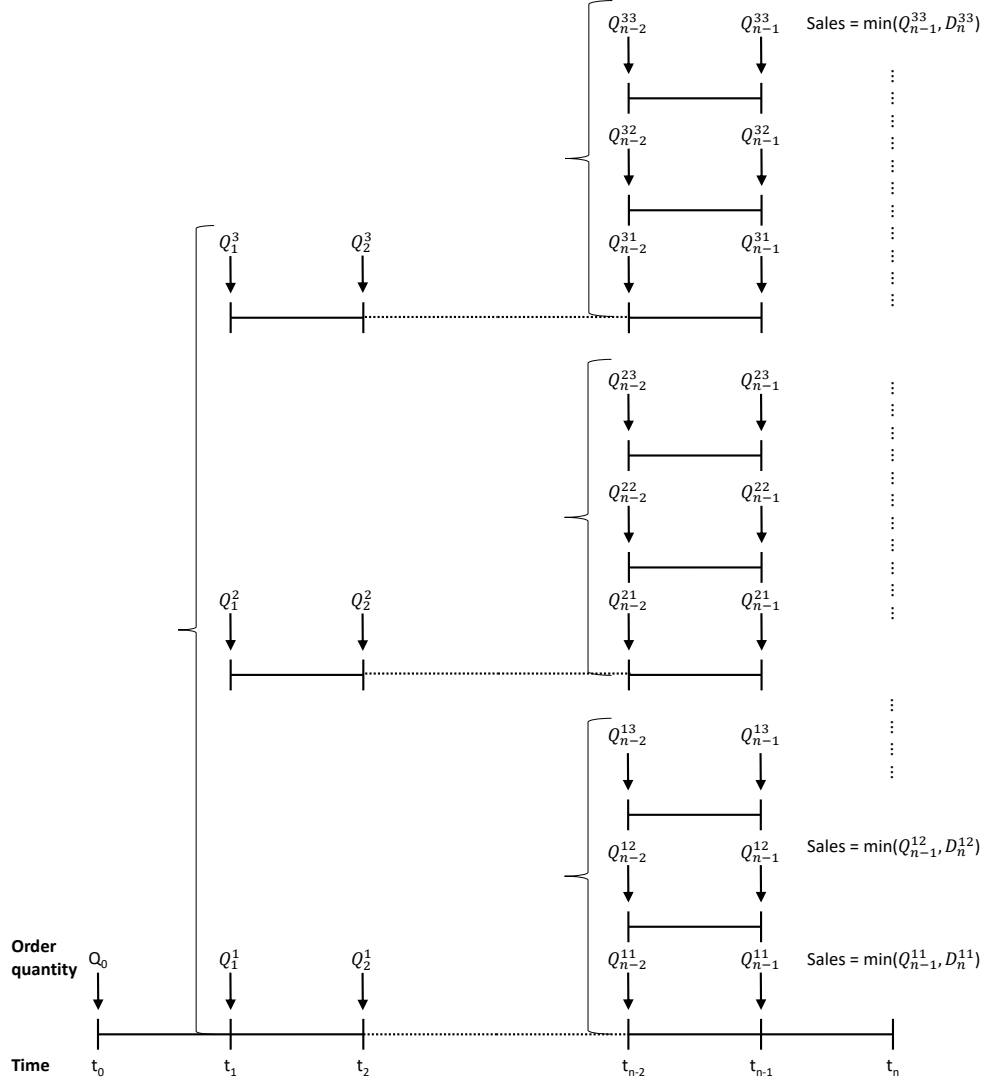


Figure 5 Ordering decisions in a supply chain with the product proliferation at the decision epochs t_1 and t_{n-2}

partition the set Θ_i into k pairwise disjoint subsets such as Θ_i^j for $j \in \{1, \dots, k\}$ and $k = |\Theta_{i-1}|$.

We define Θ_i^j as the set that contains all components that use the same upstream resource as their input. We then have by definition:

$$\Theta_i = \bigcup_{j \in \Theta_{i-1}} \Theta_i^j \quad \text{and} \quad \emptyset = \bigcap_{j \in \Theta_{i-1}} \Theta_i^j. \quad (12)$$

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 Θ_{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 Θ_{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:

$$\sum_{j \in \Theta_i^k} Q_i^j \leq Q_{i-1}^k. \quad (13)$$

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 Υ_i^k which represents the set of end products produced by using component k at echelon i . Therefore, Υ_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):

$$\begin{aligned} \text{Maximize} \quad & z = \sum_{j \in \Theta_{n-1}} p_j \mathbb{E} \left(\mathcal{W}_j(Q_{n-1}^j, D_n^j) \right) - \kappa_{n-1}^j Q_{n-1}^j \\ \text{subject to:} \quad & \end{aligned} \quad (14)$$

$$\sum_{j \in \Theta_{n-1}^k} Q_{n-1}^j \leq Q_{n-2}^k, \quad \forall k \in \Theta_{n-2}, \quad (15)$$

$$Q_{n-1}^j \geq 0, \quad \forall j \in \Theta_{n-1}, \quad (16)$$

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 (15) guarantees that the sum of order quantities of the items in a set Θ_{n-1}^k is less than the amount of their parent item k . In the appendix, we provide the solution for the mathematical problem (14)–(16). 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 Θ_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 Θ_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. The next proposition sheds light on how this benefit is shaped by operational costs.

PROPOSITION 4. *The value of postponing 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, postponing 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. Next, we investigate the impact of lead times.

PROPOSITION 5. *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. However, the model is not tractable enough to make direct analytical comparisons between alternative supply chain configurations with different points of proliferation, lead times and costs. For this objective, we resort to numerical experiments. In order to conduct these comparisons comprehensively and to limit any bias that may come from specific parameter settings, we set up a large scale numerical study based on Markovian sampling.

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 ρ 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: $t_4 - t_0 \sim U(0, 1)$. Then, we 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.

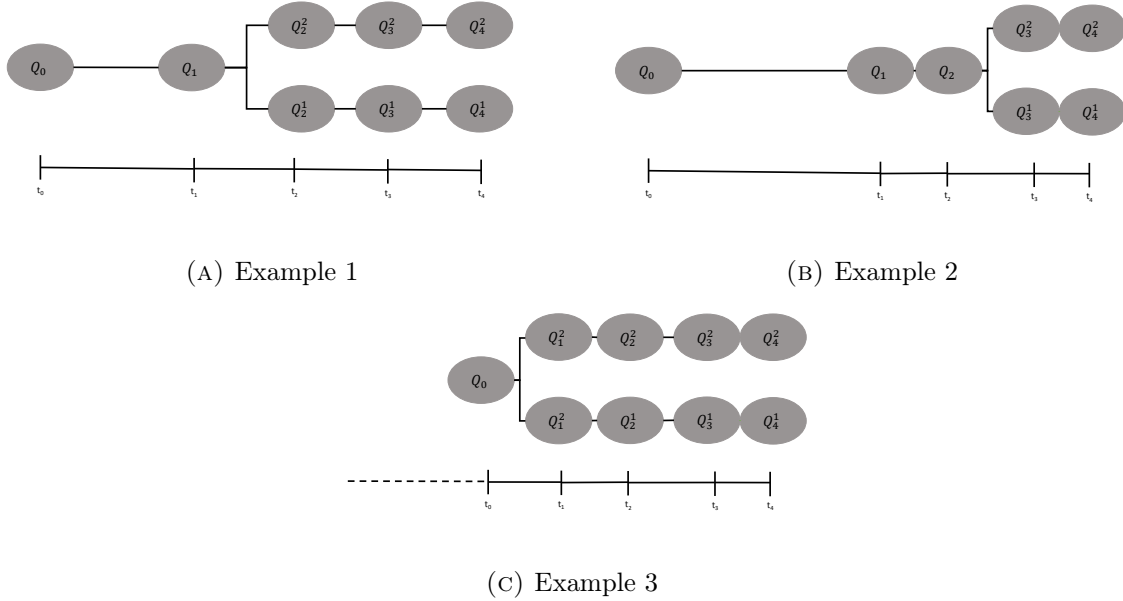


Figure 6 Some potential supply chain structures

We fix the selling price to one and salvage value to zero. Then, we sample the total production cost per unit from a uniform distribution: $\kappa_4 \sim U(0, 1)$. We randomly allocate the total production cost to the operations. After determining the input parameters, we implement a Markovian Sampling method to generate demand paths and 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:

$$\begin{aligned} Profit = & \beta_0 + \beta_1 c_1 + \beta_2 c_2 + \beta_3 c_3 + \beta_4 c_4 + \beta_5 \rho + \beta_6 CV + \beta_7 TimeofDiff + \beta_8 StepofDiff \\ & + \beta_9(t_1 - t_0) + \beta_{10}(t_2 - t_1) + \beta_{11}(t_3 - t_2) + \beta_{12}(t_4 - t_3). \end{aligned} \quad (17)$$

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 (17).

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

	Mean	SD	Min	Max	Pearson's Correlation										
					1	2	3	4	5	6	7	8	9	10	11
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. ρ	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. StepofDiff	2.48	1.12	1	4	-0.01	0.01	-0.01	0.00	0.00	-0.01	0.71				
9. $t_1 - t_0$	0.22	0.24	0.00	1.00	0.00	-0.01	-0.02	0.01	0.00	0.00	0.28	0.00			
10. $t_2 - t_1$	0.21	0.24	0.00	1.00	0.00	0.01	0.00	0.01	0.00	0.02	0.06	0.00	-0.29		
11. $t_3 - t_2$	0.22	0.24	0.00	1.00	0.00	0.00	-0.01	0.01	0.00	0.00	-0.17	0.00	-0.30	-0.29	
12. $t_4 - t_3$	0.22	0.24	0.00	1.00	-0.01	0.01	0.02	-0.02	0.00	-0.02	-0.38	0.00	-0.30	-0.29	-0.29

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

	Estimate	VIF
β_0 (Intercept)	2.0534208***	
β_1 (c_1)	-2.0029586***	1.027807
β_2 (c_2)	-1.9264279***	1.022989
β_3 (c_3)	-1.8938598***	1.023589
β_4 (c_4)	-1.8259906***	1.025135
β_5 (ρ)	-0.0209655***	1.000668
β_6 (CV)	-0.3968276***	1.001141
β_7 (TimeofDiff)	0.0295013***	4.597373
β_8 (StepofDiff)	-0.0097502***	3.289500
β_9 ($t_1 - t_0$)	-0.0737628***	2.884366
β_{10} ($t_2 - t_1$)	-0.1125864***	3.287310
β_{11} ($t_3 - t_2$)	-0.1356525***	4.024351
β_{12} ($t_4 - t_3$)	-0.1559704***	4.988999

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

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.

The numerical results substantiate our analytical results on costs and lead times. The coefficients of the cost parameters are such that $|\beta_1| > |\beta_2| > |\beta_3| > |\beta_4|$, 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 β_9 , β_{10} , β_{11} , and β_{12} 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, $|\beta_{12}| > |\beta_{11}| > |\beta_{10}| > |\beta_9|$.

Regarding the point of product proliferation, total profit increases as the timing of differentiation (*TimeofDiff*) is delayed, given that β_7 has a positive value. We also include another variable (i.e., *StepofDiff*) to indicate the rank of the echelon in which the proliferation occurs (i.e., equal to $i + 1$, where i is the subscript of t_i). The higher the value of its coefficient, the later the product proliferation. This variable is expected to be positively correlated with *TimeofDiff*, and indeed β_8 also has a positive value.³ 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, postponing 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 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 postponing 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 $0.2 * 15.6\% = 3.12\%$. Consistent with our analytical results (Propositions 3 and 5), 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.

³ 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.

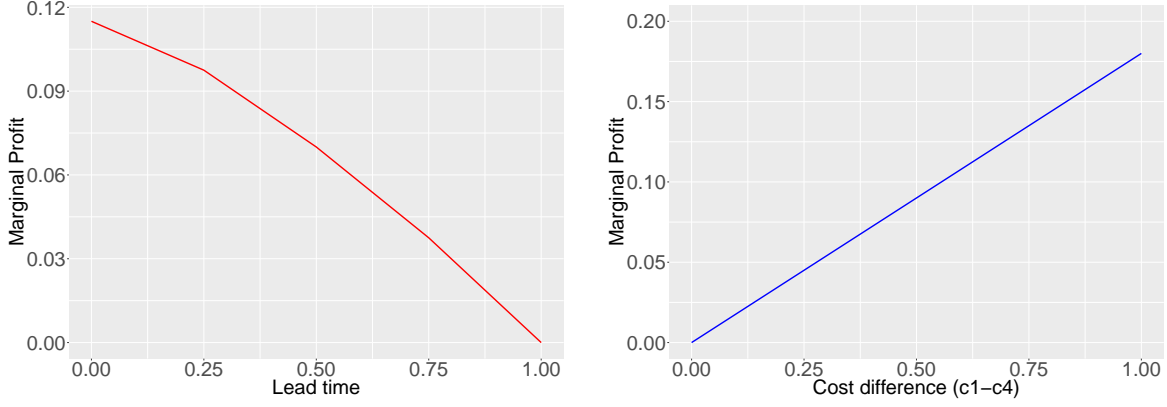


Figure 7 Value of lead-time reduction versus process redesign

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 fourth operation on the right panel (utilizing coefficients $\beta_1, \beta_4, \beta_9, \beta_{10}, \beta_{11}$, and β_{12}). 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 lead-time values for each operation the same; $t_4 - t_3 = t_3 - t_2 = t_2 - t_1 = t_1 - t_0 = 0.25$. 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.

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 postponing the point of proliferation is lead-time reduction. Propositions 3 and 5 along with the numerical findings $|\beta_{12}| > |\beta_{11}| > |\beta_{10}| > |\beta_9|$ 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 design, 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 depend 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 8.

In developing the decision typology in Figure 8, we make use of the classification in Ferdows et al. (2016) that categorizes manufacturers based on two dimensions, namely their product characteristics

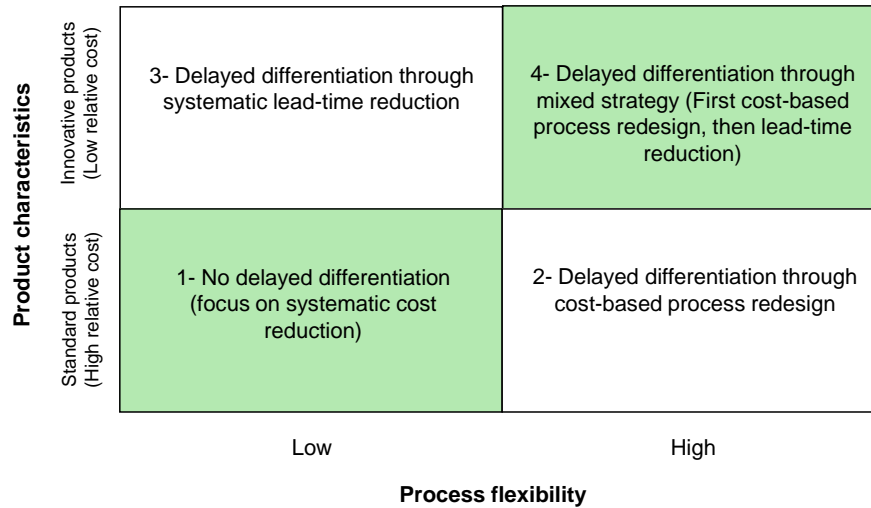


Figure 8 Decision typology based on product characteristics and process flexibility

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.⁴

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

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

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 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.⁵ This mixed strategy enables ASML to delay both the point of proliferation and high cost operations.

7. Concluding Remarks

The last decade's competitive business environment is marked by increasingly more discerning and demanding customers. Customers expect products to be more refined and customized to their needs, and they request them to be delivered faster and at lower prices. Building a responsive supply chain is a viable option to meet these expectations. Faced with enhanced proliferation in their supply chains, companies have been trying to improve responsiveness by delaying differentiation, either by redesigning processes (and hence supply chain structure) or by leadtime reduction. In this paper, we focus on the juxtaposition of the two approaches. To this end, we develop a framework that integrates the evolutionary dynamics of demand forecasts with ordering decisions in a supply chain such that product proliferation may occur at different stages. Utilizing this framework, we price the value of process redesign and lead-time reduction and compare their values depending on such factors as demand characteristics, cost and lead-time parameters. Distilling these results, we develop a typology that maps the most effective operational strategy for delaying differentiation depending on product/market characteristics and process flexibility.

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

⁵ See also https://staticwww.asml.com/doclib/investor/07_analyst_day_internet_031113.pdf

promised lead times would be an interesting avenue of future research that requires incorporation of lead-time quotation and product proliferation models.

Acknowledgments

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.
- Boyabath O (2015) Supply management in multiproduct firms with fixed proportions technology. *Management Science* 61(12):3013–3031.
- Boyabath O, Kleindorfer PR, Koontz SR (2011) Integrating long-term and short-term contracting in beef supply chains. *Management Science* 57(10):1771–1787.
- 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.
- Cachon GP, Swinney R (2011) The value of fast fashion: Quick response, enhanced design, and strategic consumer behavior. *Management Science* 57(4):778–795.
- Calvo E, Martínez-de Albéniz V (2015) Sourcing strategies and supplier incentives for short-life-cycle goods. *Management Science* 62(2):436–455.

- 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, Gallien J (2010) Inventory management of a fast-fashion retail network. *Operations Research* 58(2):257–273.
- Caro F, Martínez-de Albéniz V (2010) The impact of quick response in inventory-based competition. *Manufacturing & Service Operations Management* 12(3):409–429.
- 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).
- Dong L, Guo X, Turcic D (2018) Selling a product line through a retailer when demand is stochastic: Analysis of price-only contracts. *Manufacturing & Service Operations Management* URL <http://dx.doi.org/10.1287/msom.2018.0720>, [Online version ahead of print].
- Ferdows K, Vereecke A, De Meyer A (2016) Delaying the global production network into congruent sub-networks. *Journal of Operations Management* 41:63–74.
- Gao L, Thomas DJ, Freimer MB (2014) Optimal inventory control with retail pre-packs. *Production and Operations Management* 23(10):1761–1778.
- 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).
- 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.
- Lutze H, Özer Ö (2008) Promised lead-time contracts under asymmetric information. *Operations Research* 56(4):898–915.
- Martínez-de Albéniz V (2011) Using supplier portfolios to manage demand risk. Kouvelis P, Boyabatli O, Dong L, Li R, eds., *The Handbook of Integrated Risk Management in Global Supply Chains*, 425–445 (John Wiley and Sons Inc).
- Oh S, Özer Ö (2013) Mechanism design for capacity planning under dynamic evolutions of asymmetric demand forecasts. *Management Science* 59(4):987–1007.
- 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.

Online Appendix

Proof of Theorem 1

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

$$G_{n-1}(Q_{n-1}, D_{n-1}) = \mathbb{E}_{D_n|D_{n-1}} \left(p \min(D_n, Q_{n-1}) \right) - \kappa_{n-1} Q_{n-1}, \quad (18)$$

$$= (p - \kappa_{n-1}) Q_{n-1} - p \int_0^{Q_{n-1}} (Q_{n-1} - D_n) f(D_n | D_{n-1}) dD_n, \quad (19)$$

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

$$\frac{\partial G_{n-1}}{\partial Q_{n-1}} = p(1 - F(Q_{n-1} | D_{n-1})) - \kappa_{n-1} = 0. \quad (20)$$

With $p > \kappa_{n-1}$, $G_{n-1}(\cdot, D_{n-1})$ is a concave function for any given D_{n-1} . We then define an auxiliary function such that:

$$J_{n-1}(Q_{n-2}, D_{n-1}) = \max_{Q_{n-1} \leq Q_{n-2}} \left\{ G_{n-1}(Q_{n-1}, D_{n-1}) + \kappa_{n-2} Q_{n-1} \right\}. \quad (21)$$

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

$$J_{n-1}(Q_{n-2}, D_{n-1}) = \begin{cases} G_{n-1}(Q_{n-2}, D_{n-1}) + \kappa_{n-2} Q_{n-2} & \text{if } Q_{n-1}^* > Q_{n-2}, \\ G_{n-1}(Q_{n-1}^*, D_{n-1}) + \kappa_{n-2} Q_{n-1}^* & \text{if } Q_{n-1}^* \leq Q_{n-2}. \end{cases} \quad (22)$$

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

$$G_{n-2}(Q_{n-2}, D_{n-2}) = \mathbb{E}_{D_{n-1}|D_{n-2}} \left(J_{n-1}(Q_{n-2}, D_{n-1}) \right) - \kappa_{n-2} Q_{n-2}, \quad (23)$$

which is also concave because $J_{n-1}(\cdot, D_{n-1})$ 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:

$$Q_i^* = \arg \max_{Q_i} \{G_i(Q_i, D_i)\}, \quad \forall i \in \{0, 1, \dots, n\}. \quad (24)$$

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

$$J_{i+1}(Q_i, D_{i+1}) = \begin{cases} G_{i+1}(Q_i, D_{i+1}) + \kappa_i Q_i & \text{if } Q_{i+1}^* > Q_i, \\ G_{i+1}^*(D_{i+1}) + \kappa_i Q_{i+1}^* & \text{if } Q_{i+1}^* \leq Q_i, \end{cases} \quad (25)$$

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

$$G_i(Q_i, D_i) = \mathbb{E}_{D_{i+1}|D_i} [J_{i+1}(Q_i, D_{i+1})] - \kappa_i Q_i, \quad (26)$$

$$\begin{aligned} &= \int_{\overline{D}_{i+1}}^{+\infty} (G_{i+1}(Q_i, D_{i+1}) + \kappa_i Q_i) f(D_{i+1} | D_i) dD_{i+1} \\ &\quad + \int_0^{\overline{D}_{i+1}} (G_{i+1}^*(D_{i+1}) + \kappa_i Q_{i+1}^*) f(D_{i+1} | D_i) dD_{i+1} - \kappa_i Q_i, \end{aligned} \quad (27)$$

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:

$$\frac{\partial G_i}{\partial Q_i} = g_i(Q_i, D_i) = \int_{\bar{D}_{i+1}}^{+\infty} (g_{i+1}(Q_i, D_{i+1}) + \kappa_i) f(D_{i+1}|D_i) dD_{i+1} - \kappa_i = 0. \quad (28)$$

Using Equation (28), 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 (20). Combining Equations (20) and (28), the optimal value of Q_{n-2} can be calculated by:

$$\begin{aligned} g_{n-2}(Q_{n-2}, D_{n-2}) &= \int_{\bar{D}_{n-1}}^{+\infty} (pPr(D_n > Q_{n-2}|D_{n-1}) - c_{n-1}) f_{n-1}(D_{n-1}|D_{n-2}) dD_{n-1} - \kappa_{n-2}, \\ &= pPr(D_n > Q_{n-2}, D_{n-1} > \bar{D}_{n-1}) - c_{n-1}Pr(D_{n-1} > \bar{D}_{n-1}) \\ &\quad - \kappa_{n-2}. \end{aligned} \quad (29)$$

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

$$\begin{aligned} g_i(Q_i, D_i) &= pPr(D_n > Q_i, \mathbf{D}_{\{i+1, n-1\}} > \bar{\mathbf{D}}_{\{i+1, n-1\}}) - c_{n-1}Pr(\mathbf{D}_{\{i+1, n-1\}} > \bar{\mathbf{D}}_{\{i+1, n-1\}}) \\ &\quad - c_{n-2}Pr(\mathbf{D}_{\{i+1, n-2\}} > \bar{\mathbf{D}}_{\{i+1, n-2\}}) - \dots - c_{i+1}Pr(D_{i+1} > \bar{D}_{i+1}) - \kappa_i, \end{aligned} \quad (30)$$

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 Q_i^* value for $i \in \{0, \dots, n-1\}$ is found by Equation (10):

$$\begin{aligned} g_i(Q_i, D_i) &= pPr(D_n > Q_i, \mathbf{D}_{\{i+1, n-1\}} > \bar{\mathbf{D}}_{\{i+1, n-1\}}) - c_{n-1}Pr(\mathbf{D}_{\{i+1, n-1\}} > \bar{\mathbf{D}}_{\{i+1, n-1\}}) \\ &\quad - c_{n-2}Pr(\mathbf{D}_{\{i+1, n-2\}} > \bar{\mathbf{D}}_{\{i+1, n-2\}}) - \dots - c_{i+1}Pr(D_{i+1} > \bar{D}_{i+1}) - \kappa_i \\ &= 0. \end{aligned} \quad (31)$$

Suppose $Q_{i+1} = Q_{i+1}^* = Q_i = Q_i^*$. Then,

$$g_{i+1}(Q_{i+1}, D_i) - g_i(Q_i, D_i) = c_{i+1}Pr(D_{i+1} > \bar{D}_{i+1}) + \kappa_i - \kappa_{i+1}, \quad (32)$$

$$= -c_{i+1}(1 - Pr(D_{i+1} > \bar{D}_{i+1})) < 0. \quad (33)$$

For $Q_{i+1} = Q_i$, therefore, $g_{i+1}(Q_{i+1}, D_i) - g_i(Q_i, D_i)$ takes a negative value, failing to be equal to zero. So, the expression “ $g_{i+1}(Q_{i+1}, D_i) - g_i(Q_i, D_i) = 0$ ” is satisfied when $Q_{i+1} < Q_i$. Applying the same procedure to Q_{i+j+1} for a given D_{i+j} for $j \in \{1, \dots, n-i-2\}$ (by induction), we obtain $Q_0^* > \mathbb{E}[Q_1^*|D_0] > \dots > \mathbb{E}[Q_{n-1}^*|D_0]$. Combining this result with $q_i = \min(Q_{i-1}, Q_i^*)$, we get $q_0 > \mathbb{E}[q_1|D_0] > \dots > \mathbb{E}[q_{n-1}|D_0]$.

Proof of Proposition 2

Part A: For $j \in \{i, \dots, n-1\}$, suppose c_j is increased by Δc_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,

$$g_i(Q'_i, D_i) - g_i(Q_i, D_i) = g_i(Q'_i, D_i) = -\Delta c_j \Pr(\mathbf{D}_{\{i+1,j\}} > \bar{\mathbf{D}}_{\{i+1,j\}}). \quad (34)$$

Therefore, “ $g_i(Q'_i, D_i) = 0$ ” is satisfied for $Q'_i < Q_i$.

For $j \in \{0, \dots, i-1\}$, suppose again $Q'_i = Q_i$ and $g_i(Q_i, D_i) = 0$. And,

$$g_i(Q'_i, D_i) - g_i(Q_i, D_i) = g_i(Q'_i, D_i) = -\Delta c_j. \quad (35)$$

Likewise, “ $g_i(Q'_i, D_i) = 0$ ” is satisfied for $Q'_i < Q_i$. 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:

$$G_0(Q_0^*, D_0 | i = j) = \int_0^{Q_0^*} g_0(Q_0, D_0) dQ_0, \quad (36)$$

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 (10) yields the following result:

$$\begin{aligned} G_0(Q_0^*, D_0 | i = j+1) - G_0(Q_0^*, D_0 | i = j) &= - \int_0^{Q_0^*} (c_i - c_{fixed}) \Pr(\mathbf{D}_{\{1,i+1\}} > \bar{\mathbf{D}}_{\{1,i+1\}}) dQ_0 \\ &\quad + \int_0^{Q_0^*} (c_i - c_{fixed}) \Pr(\mathbf{D}_{\{1,i\}} > \bar{\mathbf{D}}_{\{1,i\}}) dQ_0. \end{aligned} \quad (37)$$

Equation (37) 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 Δt , the starting times for the first $i+1$ operations are updated as follows:

$$t_0 + \Delta t = t_1 + \Delta t = \dots = t_j + \Delta t = \dots = t_i + \Delta t. \quad (38)$$

Then, ordering decisions for the first $i + 1$ operations are made after a delay of Δt . 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 Δt , the starting times for the first $j + 1$ operations are likewise delayed for Δt . 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 (14)–(16) can be written as a large-scale LP model:

$$\begin{aligned} \text{Maximize}_{Q_{n-1}^j, \forall j \in \Theta_{n-1}} \quad & z = \sum_{j \in \Theta_{n-1}} \left(p_j \sum_{r \in \mathcal{S}} Pr(\mathcal{W}_j^r) \mathcal{W}_j^r - \kappa_{n-1}^j Q_{n-1}^j \right) \end{aligned} \quad (39)$$

subject to:

$$\mathcal{W}_j^r - Q_{n-1}^j \leq 0, \quad \forall j \in \Theta_{n-1}, \quad \forall r \in \mathcal{S}, \quad (40)$$

$$\mathcal{W}_j^r \leq D_{n,r}^j, \quad \forall j \in \Theta_{n-1}, \quad \forall r \in \mathcal{S}, \quad (41)$$

$$\sum_{j \in \Theta_{n-1}^k} Q_{n-1}^j \leq Q_{n-2}^k, \quad \forall k \in \Theta_{n-2}, \quad (42)$$

$$Q_{n-1}^j \geq 0, \quad \mathcal{W}_j^r \geq 0, \quad \forall j \in \Theta_{n-1}, \quad \forall r \in \mathcal{S}. \quad (43)$$

We remark that we add Constraints (40) and (41) to satisfy the condition $W_j = \min\{Q_{n-1}^j, D_n^j\}$ for the optimal solution. Then, the dual problem is:

$$\begin{aligned} \text{Minimize}_{\lambda_{j,r}, \beta_{j,r}, \gamma} \quad & w = \sum_{j \in \Theta_{n-1}} \sum_{r \in \mathcal{S}} \beta_{j,r} D_{n,r}^j + \sum_{k \in \Theta_{n-2}} \gamma_k Q_{n-2}^k \end{aligned} \quad (44)$$

subject to:

$$\lambda_{j,r} + \beta_{j,r} \geq Pr(\mathcal{W}_j^r) p_j, \quad \forall j \in \Theta_{n-1}, \quad r \in \mathcal{S}, \quad (45)$$

$$\sum_{r \in \mathcal{S}} \lambda_{j,r} - \gamma_k \leq \kappa_{n-1}^j, \quad \forall j \in \Theta_{n-1}^k, \quad k \in \Theta_{n-2}, \quad r \in \mathcal{S}, \quad (46)$$

$$\lambda_{j,r} \geq 0, \quad \beta_{j,r} \geq 0, \quad \gamma_k \geq 0, \quad \forall j \in \Theta_{n-1}^k, \quad k \in \Theta_{n-2}, \quad r \in \mathcal{S}. \quad (47)$$

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

Then, the constraint (46) is written as follows:

$$p_j Pr(D_n^j > Q_{n-1}^j) - \gamma_k \leq \kappa_{n-1}^j, \quad \forall j \in \Theta_{n-1}. \quad (48)$$

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

$$\gamma_k = (p_j - \kappa_{n-1}^j) - p_j Pr(D_n^j \leq Q_{n-1}^j) = g_{n-1}^j(Q_{n-1}^j, D_{n-1}^j), \quad \forall j \in \Theta_{n-1}^k. \quad (49)$$

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

$$w = \sum_{j \in \Theta_{n-1}} \left[(p_j - \kappa_{n-1}^j) Q_{n-1}^j - p_j \int_0^{Q_{n-1}^j} (Q_{n-1}^j - D_n^j) f^j(D_n^j) dD_n^j \right]. \quad (50)$$

Equation (50) 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 (49). Therefore, we have the following conditions of optimality:

$$\gamma_k = g_{n-1}^j(Q_{n-1}^j, D_{n-1}^j) \geq 0, \quad \forall j \in \Theta_{n-1}^k, \quad k \in \Theta_{n-2} \quad (51)$$

$$\sum_{j \in \Theta_{n-1}^k} Q_{n-1}^j \leq Q_{n-2}^k, \quad \forall k \in \Theta_{n-2}. \quad (52)$$

If the constraint (42) is not binding for a given $k \in \Theta_{n-2}$, the dual variable γ_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 Θ_{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 Θ_{n-1}^k .

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

$$\text{Maximize}_{Q_i^j, \forall j \in \Theta_i} z = \sum_{j \in \Theta_i} G_i^{\Upsilon_i^j}(Q_i^j, \mathbf{D}_i^{\Upsilon_i^j}) \quad (53)$$

subject to:

$$\sum_{j \in \Theta_i^k} Q_i^j \leq Q_{i-1}^k, \quad 0 \leq Q_i^j, \quad \forall k \in \Theta_{i-1}, \quad \forall j \in \Theta_i^k. \quad (54)$$

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 Υ_i^j , and $\mathbf{D}_i^{\Upsilon_i^j}$ is the vector of demand forecasts of the products in Υ_i^j at time t_i . We recall that Υ_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 (53)–(54) is formulated as:

$$\text{Minimize } w = \sum_{\lambda_j, \beta_j, \gamma_k} \gamma_k Q_{i-1}^k \quad (55)$$

subject to:

$$g_i^{\Upsilon_i^j}(Q_i^j, \mathbf{D}_i^{\Upsilon_i^j}) \leq \gamma_k, \quad \forall j \in \Theta_i^k, \quad \forall k \in \Theta_{i-1}, \quad (56)$$

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:

$$\gamma_k = g_i^{\Upsilon_i^j}(Q_i^j, \mathbf{D}_i^{\Upsilon_i^j}) \geq 0, \quad \forall j \in \Theta_i^k, \quad k \in \Theta_{i-1} \quad (57)$$

$$\sum_{j \in \Theta_i^k} Q_i^j \leq Q_{i-1}^k, \quad \forall k \in \Theta_{i-1}. \quad (58)$$

Following the steps similar to Appendix 7, we obtain the following expression for $t = t_{n-2}$:

$$\begin{aligned} G_{n-2}^{\Upsilon_{n-2}^j}(Q_{n-2}^j, \mathbf{D}_{n-2}^{\Upsilon_{n-2}^j}) &= \int_{\bar{D}_{n-1}^{\Upsilon_{n-2}^j}}^{+\infty} (G_{n-1}^{\Upsilon_{n-2}^j}(Q_{n-2}^j, D_{n-1}^{\Upsilon_{n-2}^j}) + \kappa_{n-2} Q_{n-2}^j) f(D_{n-1}^{\Upsilon_{n-2}^j} | D_{n-2}^{\Upsilon_{n-2}^j}) dD_{n-1}^{\Upsilon_{n-2}^j} \\ &\quad + \int_0^{\bar{D}_{n-1}^{\Upsilon_{n-2}^j}} (G_{n-1}^{*\Upsilon_{n-2}^j}(Q_{n-2}^j) + \kappa_{n-2}^j Q_{n-2}^j) f(D_{n-1}^{\Upsilon_{n-2}^j} | D_{n-2}^{\Upsilon_{n-2}^j}) dD_{n-1}^{\Upsilon_{n-2}^j} \\ &\quad - \kappa_{n-2}^j Q_{n-2}^j, \end{aligned} \quad (59)$$

where $D_{n-1}^{\Upsilon_{n-2}^j}$ is a random variable denoting the sum of demand forecasts of the items in the set Υ_{n-2}^j at $t = t_{n-1}$ (i.e., $\sum_{k \in \Upsilon_{n-2}^j} D_{n-1}^k$). $\bar{D}_{n-1}^{\Upsilon_{n-2}^j}$ 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,

$$\begin{aligned} g_{n-2}^{\Upsilon_{n-2}^j}(Q_{n-2}^j, \mathbf{D}_{n-2}^{\Upsilon_{n-2}^j}) &= \int_{\bar{D}_{n-1}^{\Upsilon_{n-2}^j}}^{+\infty} (g_{n-1}^{\Upsilon_{n-2}^j}(Q_{n-1}^j, D_{n-1}^{\Upsilon_{n-2}^j}) + \kappa_{n-2}^j) f(D_{n-1}^{\Upsilon_{n-2}^j} | D_{n-2}^{\Upsilon_{n-2}^j}) dD_{n-1}^{\Upsilon_{n-2}^j} \\ &\quad - \kappa_{n-2}^j, \end{aligned} \quad (60)$$

where $g_{n-1}^{\Upsilon_{n-2}^j}(Q_{n-1}^j, D_{n-1}^{\Upsilon_{n-2}^j}) = g_{n-1}^k(Q_{n-1}^k, D_{n-1}^k)$ for any $k \in \Upsilon_{n-2}^j$ as given by Equation (51). Using the last expression, we obtain the following result by induction:

$$g_i^{\Upsilon_i^j}(Q_i^j, \mathbf{D}_i^{\Upsilon_i^j}) = \int_{\bar{D}_{i+1}^{\Upsilon_i^j}}^{+\infty} (g_{i+1}^{\Upsilon_i^j}(Q_{i+1}^j, D_{i+1}^{\Upsilon_i^j}) + \kappa_i^j) f(D_{i+1}^{\Upsilon_i^j} | D_i^{\Upsilon_i^j}) dD_{i+1}^{\Upsilon_i^j} - \kappa_i^j = \gamma_k, \forall j \in \Theta_i^k. \quad (61)$$

If the constraint (54) for a given k is not binding, the dual variable γ_k becomes zero. In this case, the optimal solution reduces to the solution of $|\Theta_i^k|$ unconstrained problems using Equation (10). Let Q_i^{j*} denote the order quantity for $j \in \Theta_i^k$ and $k \in \Theta_{i-1}$ satisfying Equation (10) and \hat{Q}_i^j denote the order quantity satisfying Equation (61). Then, the optimal order quantity is found by a state-dependent base-stock policy:

$$q_i^j = \begin{cases} Q_i^{j*} & \text{if } \sum_{j \in \Theta_i^k} Q_i^{j*} < Q_{i-1}^k, \\ \hat{Q}_i^j & \text{if } \sum_{j \in \Theta_i^k} Q_i^{j*} \geq Q_{i-1}^k. \end{cases} \quad (62)$$

Proof of Proposition 4

We assume without loss of generality that product proliferation occurs once at time t_i along the supply chain. The mathematical model given by (53)–(54) is then rewritten as follows with an objective function of maximizing the expected profit at time t_i :

$$\text{Maximize}_{Q_i^j, \forall j \in \Theta_i} z = \sum_{j \in \Theta_i} G_i^j(Q_i^j, \mathbf{D}_i^{\Theta_i}) \quad (63)$$

subject to:

$$\sum_{j \in \Theta_i^k} Q_i^j \leq Q_{i-1}^k, \quad \forall k \in \Theta_{i-1}. \quad (64)$$

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

We first consider the case in which the constraint (64) is not binding. Then, the objective function (63) takes the following form:

$$z = \sum_{j \in \Theta_i} G_i^j(Q_i^j, \mathbf{D}_i^{\Theta_i}) = \sum_{j \in \Theta_i} G_i^j(Q_i^j, D_i^j) \quad (65)$$

Suppose $c_i^j = c_{i+1}^j = \dots = c_{n-1}^j = 0 \ \forall j \in \Theta_i$. For any $j \in \Theta_i$,

$$G_i^j(Q_i^{j*}, D_i^j) = G_i^j(Q_{i-1}^j, D_i^j), \quad (66)$$

such that optimal order quantity Q_i^{j*} is equal to what has been ordered in the previous period (i.e., Q_{i-1}^j).

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:

$$Q_{i-1}^j = Q_i^j = \dots = Q_{n-1}^j. \quad (67)$$

Thus, the expected profit is not affected by the ordering decision at t_i . Then,

$$G_i^j(Q_i^j, D_i^j) / \partial t_i = 0 \quad \rightarrow \quad \partial z / \partial t_i = 0. \quad (68)$$

Now, suppose that $c_i^j > 0$ while $c_{i+1}^j = \dots = c_{n-1}^j = 0 \ \forall j \in \Theta_i$. Then,

$$G_i^j(Q_i^{j*}, D_i^j) = (p_j - \kappa_i^j)Q_i^{j*} - p_j \int_0^{Q_i^{j*}} (Q_i^{j*} - D_n^j) f(D_n^j | D_i^j) dD_n^j, \quad (69)$$

$$= p_j \int_0^{Q_i^{j*}} D_n^j f(D_n^j | D_i^j) dD_n^j = p_j D_i^j \Phi\left(\frac{\ln(Q_i^{j*} / D_i^j)}{\sigma \sqrt{t_n - t_i}} - \sigma \sqrt{t_n - t_i} / 2\right). \quad (70)$$

The optimal order quantity Q_i^{j*} is found by:

$$Q_i^{j*} = D_i^j e^{-\sigma^2(t_n - t_i)/2 + \Phi^{-1}((p_j - \kappa_i^j)/p_j) \sigma \sqrt{t_n - t_i}}. \quad (71)$$

Plugging this derivation into Equation (70), we obtain:

$$G_i^j(Q_i^{j*}, D_i^j) = p_j D_i^j \Phi\left(\Phi^{-1}((p_j - \kappa_i^j)/p_j) - \sigma \sqrt{t_n - t_i}\right), \quad (72)$$

and

$$\partial G_i^j(Q_i^j, D_i^j) / \partial t_i > 0 \quad \rightarrow \quad \partial z / \partial t_i > 0. \quad (73)$$

Comparing Equations (68) and (73), we can state that the value of postponing the point of proliferation (i.e., $\partial z / \partial t_i$) increases as the costs of the operations after the proliferation point increase.

We now focus on the case in which the constraint (64) is binding. Suppose $c_i^j = c_{i+1}^j = \dots = c_{n-1}^j = 0 \ \forall j \in \Theta_i$. Then, the statements given by Equation (68) still hold true. When $c_i^j > 0$ and $c_{i+1}^j = \dots = c_{n-1}^j = 0 \ \forall j \in \Theta_i$, we obtain:

$$G_i^j(Q_i^j, D_i^j) = (p_j - \kappa_i^j)Q_i^j - p_j \int_0^{Q_i^j} (Q_i^j - D_n^j) f(D_n^j | D_i^j) dD_n^j, \quad (74)$$

$$= (p_j - \kappa_i^j)Q_i^j - p_j Q_i^j \Phi\left(\frac{\ln(Q_i^j / D_i^j)}{\sigma \sqrt{t_n - t_i}} + \sigma \sqrt{t_n - t_i} / 2\right) \\ + p_j D_i^j \Phi\left(\frac{\ln(Q_i^j / D_i^j)}{\sigma \sqrt{t_n - t_i}} - \sigma \sqrt{t_n - t_i} / 2\right). \quad (75)$$

Order quantity in Equation (75) is not the same as Equation (71). The fact that the constraint (64) is binding for the former equation makes the order quantity in (75) less than what can be found by Equation (71). Using Equation (75), we obtain:

$$\partial G_i^j(Q_i^j, D_i^j) / \partial t_i > 0 \quad \rightarrow \quad \partial z / \partial t_i > 0. \quad (76)$$

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

Proof of Proposition 5

The proof follows directly from Proposition 3.

Recent ESMT Working Papers

	ESMT No.
Marginality, dividends, and the value in games with externalities Frank Huettner, ESMT Berlin André Casajus, HHL Leipzig Graduate School of Management	19-01
Consumer choice under limited attention when alternatives have different information costs Frank Huettner, ESMT Berlin Tamer Boyacı, ESMT Berlin Yalçın Akçay, Melbourne Business School	16-04 (R3)
Opaque queues: Service systems with rationally inattentive customers Caner Canyakmaz, ESMT Berlin Tamer Boyacı, ESMT Berlin	18-04
The Coleman-Shapley-index: Being decisive within the coalition of the interested André Casajus, HHL Leipzig Graduate School of Management Frank Huettner, ESMT Berlin	18-03
Reverse privatization as a reaction to the competitive environment: Evidence from solid waste collection in Germany Juri Demuth, E.CA Economics Hans W. Friederiszick, ESMT Berlin and E.CA Economics Steffen Reinhold, E.CA Economics	18-02