Cost Reduction Through Operations Reversal

Ki Ling Cheung
Department of Information Systems, Business Statistics, and Operations Management, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, imcheung@ust.hk

Jing-Sheng Song
Fuqua School of Business, Duke University, Durham, NC 27708, jssong@duke.edu

Yue Zhang
Fuqua School of Business, Duke University, Durham, NC 27708, yueyue.zhang@duke.edu

In some manufacturing and service processes, several stages must be performed, but there is some freedom in the ordering of stages. Operations reversal means switching the order of two stages. Several authors have studied the benefits of operations reversal, focusing on the reduction of a certain variable's variance or a related measure. This paper focuses instead on cost. We construct a model with the standard objective of minimizing the long-run average inventory-related cost. First, by using stochastic orders, we identify conditions under which operations reversal reduces cost. We find that in some cases the variability and cost objectives agree on when operations reversal is beneficial, but in other cases they disagree. In particular, when demands are multinomially distributed, variability reduction may be accompanied by cost increase. We show that, to guarantee a lower cost, we need certain properties on the aggregated demand at the choice-level (such as demands for sweaters of the same color). Finally, we examine the effects of cost parameters and lead times on operations reversal under the cost measure.

Key words: Restructuring, Process Design, Operations Sequencing

1. Introduction

A manufacturing process of a product often consists of many stages, with a certain feature being added to the product at each stage. As consumers demand increasingly high product variety, which means more choices for each feature of a product, manufacturers need to develop production capabilities to maintain customer satisfaction without jeopardizing supply chain costs. Operations reversal, which reverses two consecutive stages of the original manufacturing process, is one innovative approach some manufacturers employ to achieve this goal. The potential influence of operations reversal is multidimensional. The intermediate products, which are the unfinished products that have just finished the former stage but not started the latter one, vary under different production sequences as the feature added to them switches. When the demands for each end product are uncertain, a properly chosen sequence of production may benefit the supply chain by better matching the production volumes with demands. One well documented example is a practice at Benetton (Harvard Business School (1986)), in which the company reversed the dyeing and knitting stages in the traditional production process to improve operating efficiency and customer service. (See Harvard Business School (1986) and references therein for many other examples, including the sequencing of hardware and software installations for personal computers, the order of adding different features for deskjet printers at Hewlett-Packard, and the sequencing of various assembly stages in fountain pen production.) Other examples include sequencing cutting operations associated with a CNC machine (Bard and Feo (1989)), delaying the integration of an expensive component in auto industry (Schraner and Hausman (1997)), and sequencing of machining and heating operations in the mechanical industry (Shi et al. (2014)).

Noticing that Benetton only applied operations reversal to 20% of its woolen production process, Lee and Tang (1998) formulate a two-stage model with random demand to explore conditions under which operations reversal is beneficial. They use the total variability, measured by the sum of the variances of the production volumes, as the performance measure. Kapuscinski and Tayur (1999) examine the same issue but define the total variability as the sum of the standard deviations of the production volumes. Jain and Paul (2001) extend these studies to allow multiple market segments and random customer preferences. One key insight from these studies is that, the more distinctive feature (i.e., the feature that has more imbalanced consumer choice probabilities) should be processed first.

However, as mentioned by Lee and Tang (1998), a major limitation of the above modeling approach is the "use of the total variability as a performance measure. Future studies can include the explicit modeling of the cost consequences of variability directly." Indeed, even though a larger variability, whether demand- or supply-oriented, is often associated with a higher supply chain cost, Ridder et al. (1998) point out that this intuition is not necessarily correct. Given that the relationship between variability and cost implications of operations reversal has remained unclear, the goal of the present paper is to examine the problem from a cost perspective and compare the results with those based on the total variability measure.

More specifically, we construct a model with the standard objective of minimizing the long-run average inventory-related cost, based on the same manufacturing process as that of Lee and Tang (1998) and Kapuscinski and Tayur (1999). The standard inventory-cost objective is a simple, stylized measure of performance, and so are the standard variability measures. But it is worthwhile to study such simplified formulations to gain insights which more accurate but complex objectives would obscure.

We show that under multivariate normal demand, the total variability measure used by Kapuscinski and Tayur (1999) is equivalent to our cost measure. This is consistent with the classic single-item, single-stage inventory model with normal demand, where the optimal cost is proportional to the standard deviation of demand. However, for a general demand distribution, this is not always true. Indeed, we have examples to demonstrate that re-sequencing the production stages according to the criterion of total variability, as adopted by Lee and Tang (1998) and Kapuscinski and Tayur (1999), may even increase cost instead of reducing it in certain cases.

Using ideas from the theory of stochastic order, we introduce a new set of sufficient conditions under which operations reversal reduces cost. While the previous works demonstrate that the more "distinctive" feature should be processed first, our sufficient conditions imply that the feature with less variable demand (in the sense of convex order and dilation order) for each choice should be processed early in the production. Interestingly, the total variability criteria proposed by the previous studies are also implied by our sufficient conditions, even though these criteria are not sufficient conditions on their own. That is, our condition for reversal to improve performance is stronger than lower *total* variability.

While our focus is on multiple products, several researchers have studied optimal operations sequencing of single-product production systems. For example, Schraner and Hausman (1997) investigate the effect of holding costs and production lead times of different production stages. Shi et al. (2014) examine the effect of yield loss. Both of these studies also use the total system cost as a measure of effectiveness. Our paper can thus be considered complementary to these works. Of course, it would be ideal to combine all these features – different production costs and lead times, yield loss, multiple products – in one single model, but such a model presents tremendous challenge to analytical tractability even without considering production sequencing issues. Nonetheless, we take an initial step in examining the effect of different cost parameters and general lead times in our multiproduct model by further restricting the end product demands to be independent and identically distributed. We hope our analysis here can help facilitate future research efforts in developing more comprehensive models.

Finally, it is worth mentioning that our use of stochastic comparison techniques was inspired by and expands the efforts in the applications of these tools in the supply chain management literature. Gerchak and Mossman (1992) utilize variability order to study the effect of demand randomness on optimal inventory levels and the expected costs in the newsvendor model. Ridder et al. (1998) employ stochastic dominance relations to characterize demand variabilities in the newsvendor problem. Song (1994a) adopts stochastic comparison methods to study the effect of lead time and demand uncertainties in a base-stock system with the objective of long-run average cost. Song (1994b) conducts a similar analysis under the criteria of infinite-horizon expected total discounted cost. Song and Yao (2002) and Lu et al. (2003) investigate the influence of lead-time and demand variabilities in assemble-to-order systems. Iyer and Jain (2003) compare both the size and the variability of lead-time distributions for retailers in a manufacturer-retailer system. Gerchak and He (2003) utilize mean-preserving transformation and convex order to study the benefits of risk pooling for the newsvendor problem. Gupta and Cooper (2005) study the relationship between yield rate uncertainty and expected profit. Jemaï and Karaesmen (2005) explore inter-demand time variabilities in a make-to-stock queue. Corbett and Rajaram (2006) study the pooling effect for general multivariate demand distributions. Using multivariate stochastic comparision, they find the value of pooling to be larger when faced with less positively correlated demand. Zhang and Cheung (2008) derive the optimal retailer replenishment sequences in a distribution system with one supplier and multiple retailers via convex order and stochastic order. Song et al. (2010) analyze the effect of lead time and demand uncertainties in an (r, q) inventory system. More recently, Federgruen and Wang (2013) derive conditions for monotonicity of optimal policy parameters for inventory systems with (r, q) or (r, nq) policy, Huang et al. (2015) examine the sourcing strategy under random demand surges.

While most of the studies in the literature such as the ones mentioned above work with end product demands, our problem requires working with intermediate product demands, which are aggregated from end product demands. This is because our model involves multiple end products in a two-stage system and allows the option to reverse the two stages, as will be demonstrated in detail in Section 2.1. Working with intermediate product demand complicates the analysis substantially. For instance, few distributions preserve under convolution, and the intermediate product demands in the reversed system are intrinsically connected with those in the original system, excluding many results of stochastic comparison to be adopted without obtaining trivial results. The online Appendix B provides more details on this point.

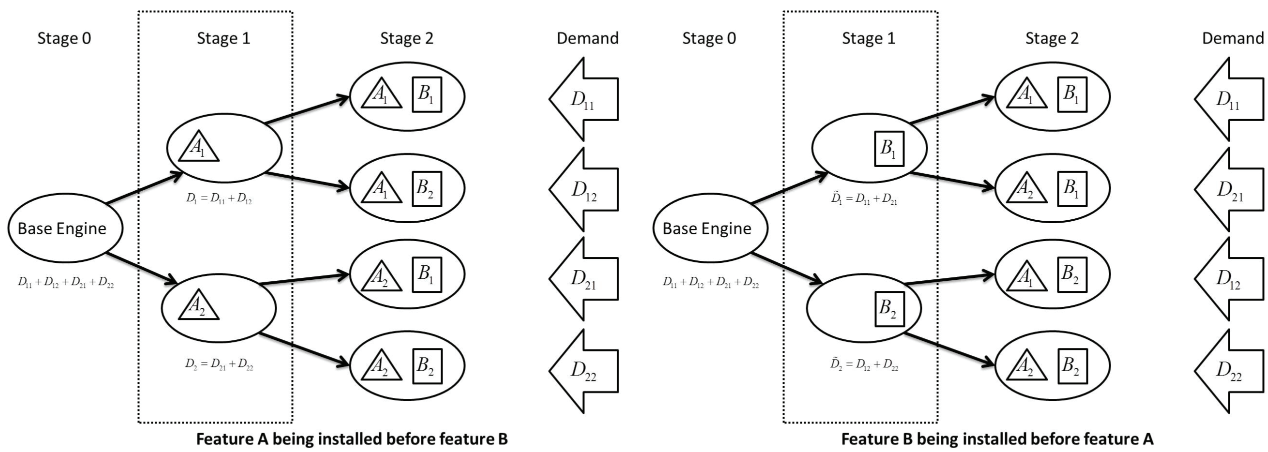
The rest of this paper is organized as follows. Section 2 introduces our cost model and summarizes previous works. Section 3 develops our results and relates them to those based on the variability measures. Section 4 examines the effect of different cost parameters and general lead times on reversal decisions (The online Appendix C provides detailed derivations of the inventory variables). Section 5 concludes our study. All proofs are provided in Appendix A.

2. Model and Previous Works

2.1. Model and Assumptions

Consider a periodic-review manufacturing process for a product (e.g., a sweater) with two features – feature A (e.g., style) and feature B (e.g., color). The process consists of two steps, each of which involves adding a separate feature to the base module through some processing procedure, such as knitting or dyeing. Suppose there are m choices for feature A (e.g., turtleneck, v-neck) and

n choices for feature B (e.g., red, blue), then there are mn variants of the end product. Denote $M = \{1, ..., m\}$, and $N = \{1, ..., n\}$. We rank the m (n) choices of feature A (B) in an arbitrary order labeled from 1 to m (n). Our objective is to determine whether we should process feature A or feature B in the first stage of the process. The two corresponding systems are denoted as System AB and System BA. If feature A (B) is processed first, there are m (n) different intermediate products with feature A (B) only. For simplicity, we use A_i (B_j) to denote the intermediate product with the ith (jth) choice of feature A (B). Similarly, A_iB_j denotes the end product with the ith choice of feature A and jth choice of feature B. The process charts of the two systems are displayed in Figure 1.



(a) System AB (b) System BA

Figure 1 Process Charts for the Two Systems under Different Processing Sequences

We assume the demands for each end product are i.i.d. over time, and let D_{ij} denote the one-period demand of product A_iB_j. However, demands across different end products in the same period can be correlated. In System AB, let $D_i = \sum_{j\in N} D_{ij}$ denote the one-period demand for A_i (e.g., the aggregate one-period demand for turtleneck sweaters). Similarly, $\tilde{D}_j = \sum_{i\in M} D_{ij}$ is the one-period demand for B_j (e.g., the aggregate one-period demand for red sweaters) in System BA. Denote

$$D = \sum_{i\in M} D_i = \sum_{j\in N} \tilde{D}_j = \sum_{i\in M, j\in N} D_{ij}$$

the aggregate end-product demand. Let

$$E(D) = \mu, \qquad V(D) = \sigma^2.$$

Here and throughout the paper, for any random variable Z, we use $E(Z)$, $V(Z)$, and $SD(Z)$ to denote its mean, variance, and standard deviation. Also, $Cov(Z, Z')$ denotes the covariance of Z and Z'.

Similar to Lee and Tang (1998), we assume the processing time at each stage is one period. Also, each stage follows a periodic-review, base-stock policy. (The inventory policy parameters and other related notation are summarized in Table 1.) Thus, we have a pull system and the production volume for each stage equals the demand of the previous period at that stage. Moreover, the system operates in an uncapacitated manner and no backlog is allowed. As implicitly implied by Lee and Tang (1998) and similar to Lee et al. (2000), in each period, if there is a shortage, whether at the level of intermediate product or end product, the manufacturer can expedite the production at a higher cost (eg. working overtime, hiring temporary workers, leasing extra machines, etc.). Alternatively, it can borrow products from other sources immediately to fill the demand at a higher cost, and returns those products when they become available eventually. We call this extra cost the penalty cost for inventory shortage.

There is no production setup cost, but there is an inventory holding cost at all stages. Assume there are always abundant base modules available for processing (at the beginning of stage 1, termed stage 0). We shall neglect the holding cost associated with the base module in our analysis since operations reversal does not affect the total amount of base modules used up in the manufacturing process. To focus exclusively on the effect of demand variability on operations reversal and facilitate comparison with the existing results, we assume the holding and penalty costs are stage dependent but feature independent, that is, the cost of adding any feature in a given stage is considered the same. We also assume they are homogeneous across different choices of a given feature within each stage , i.e., choice independent. Denote by h^k and p^k the unit per-period holding and penalty costs at stage k, $k = 1, 2$, respectively. We adopt the long-run average cost as the performance measure.

It is well known that for serial systems without setup costs, a base-stock policy is optimal (see Clark and Scarf (1960)). According to Theorem 6 in Van Houtum et al. (1996), a base-stock policy is optimal for the two-stage divergent system with independent demands under a balance assumption. Note that our system is a two-stage divergent system and the balance assumption is automatically satisfied due to the assumption that the manufacturer can borrow items and return them later. Hence a base-stock policy is optimal here when the demands for different end products are independent of each other. When demands are dependent, however, the optimal policy remains unknown. Nonetheless, base-stock policies are prevalent in many supply chains that involve multiple products with no setup costs, including serial systems, distribution systems, and assemble-to-order systems. Any analysis developed upon this type of policy would shed light on how to manage realistic supply chains. Thus, we assume a base-stock policy.

For stage $k=1,2$, define

$$R^k(x,y)=h^k(x-y)^+ + p^k(y-x)^+. \tag{1}$$

Note that (1) is the classic form of a newsvendor function, where x means the base-stock level, and y is the product demand. Clearly, $R^k(x,y)$ is convex in y for any given x. The long-run average costs for system AB and system BA for given base-stock policies can be expressed respectively as

$$C_{AB}(\mathbf{s};\mathbf{S})=\sum_{i\in M}E(R^1(s_i,D_i))+\sum_{i\in M,j\in N}E(R^2(S_{ij},D_{ij})), \tag{2}$$

and

$$C_{BA}(\tilde{\mathbf{s}};\tilde{\mathbf{S}})=\sum_{j\in N}E(R^1(\tilde{s}_j,\tilde{D}_j))+\sum_{i\in M,j\in N}E(R^2(\tilde{S}_{ij},D_{ij})), \tag{3}$$

where s_i and $\tilde{s}_j$ are the order up to levels for intermediate products A_i and B_j, S_{ij} and $\tilde{S}_{ij}$ are the order up to levels for end product A_iB_j in Systems AB and BA, D_i and $\tilde{D}_j$ are the one-period demands for intermediate products A_i and B_j, D_{ij} is the one-period demand for end product A_iB_j, all of which can be found in Table 1.

	System AB	System BA
End Product Demand	$D_{ij},\, i \in M,\, j \in N$	$D_{ij},\, i \in M,\, j \in N$
Base-Stock Level	$S_{ij},\, i \in M,\, j \in N$	$\tilde{S}_{ij},\, i \in M,\, j \in N$
Base-Stock Level Vector	$\mathbf{S}$	$\tilde{\mathbf{S}}$
Intermediate Product Demand	$D_i = \sum_{j \in N} D_{ij}$	$\tilde{D}_j = \sum_{i \in M} D_{ij}$
Distribution/Density	F_i/f_i	$\tilde{F}_j/\tilde{f}_j$
Base-Stock Level	s_i	$\tilde{s}_j$
Base-Stock Level Vector	$\mathbf{s} = \{s_1, ..., s_m\}$	$\tilde{\mathbf{s}} = \{\tilde{s}_1, ..., \tilde{s}_n\}$
Sum of Variances	$V_{AB} = \sum_{i \in M} V(D_i)$	$V_{BA} = \sum_{j \in N} V(\tilde{D}_j)$
Sum of Standard Deviations	$SD_{AB} = \sum_{i \in M} SD(D_i)$	$SD_{BA} = \sum_{j \in N} SD(\tilde{D}_j)$
Long-Run Average Cost	$C_{AB}(\mathbf{s}; \mathbf{S})$	$C_{BA}(\tilde{\mathbf{s}}; \tilde{\mathbf{S}})$
Optimal Base-Stock Levels	$\{\mathbf{s}^*, \mathbf{S}^*\}$	$\{\tilde{\mathbf{s}}^*, \tilde{\mathbf{S}}^*\}$

Table 1 Notations for System AB and System BA

The system costs in (2) and (3) are both composed of two sums. The first sum exclusively related to intermediate products, while the second sum is exclusively related to end products. Optimizing the cost function is equivalent to optimizing the two sums separately. Given $\mathbf{S} = \tilde{\mathbf{S}}$, the second sum is the same for the two systems, implying that the optimal solutions for this part are also the same. Therefore, $\mathbf{S}^* = \tilde{\mathbf{S}}^*$. Denote the optimal long-run average costs of system AB and system BA as

$$C^*_{AB} \equiv C_{AB}(\mathbf{s}^*; \mathbf{S}^*), \quad C^*_{BA} \equiv C_{BA}(\tilde{\mathbf{s}}^*; \tilde{\mathbf{S}}^*). \tag{4}$$

The difference between the two optimal costs is

$$\Delta C^* \equiv C^*_{AB} - C^*_{BA} = \sum_{i \in M} E(R^1(s^*_i, D_i)) - \sum_{j \in N} E(R^1(\tilde{s}^*_j, \tilde{D}_j)). \tag{5}$$

Since our subsequent analysis only involves the intermediate stage (stage 1), we suppress the superscripts from now on and replace h^1, p^1 and R^1 with h, p and R.

With the cost measure defined above, we say System AB is optimal (better than System BA, or A should be processed first) if

$$C^*_{AB} < C^*_{BA} \quad \text{or} \quad \Delta C^* < 0. \tag{6}$$

2.2. Previous Works

Before proceeding further, we briefly summarize the important results from three relevant papers using our notation.

2.2.1. Lee and Tang (1998) consider the same basic system as ours, but with the main focus on the case $m = n = 2$; that is, there are two features (e.g., style and color) and two choices for each feature (e.g., turtleneck and v-neck for style, red and blue for color). They assume end-product demands are multinomially distributed, and analyze the impact of operations reversal on the variabilities of production volumes. The authors argue that operations reversal only affects the variabilities in the intermediate stage, and use the sum of the variances of production volumes in that stage, i.e., V_{AB} for System AB and V_{BA} for System BA, as the performance measure. In other words, System AB is optimal if

$$V_{AB} < V_{BA}. \tag{7}$$

In terms of the sweater example, this says the sum of the variance of demand for turtleneck sweaters and that for v-neck sweaters is smaller than the sum of the variance of demand for red sweaters and that for blue sweaters. Let α (β) be the probability that a customer chooses the first choice of A (B) (e.g., turtleneck (red)). Assuming the aggregate demand of all end products satisfies

$$\mu > \sigma^2, \tag{8}$$

a key result of the paper is that a necessary and sufficient condition for (7) is

$$|\alpha - 0.5| > |\beta - 0.5|. \tag{9}$$

That is, System AB is optimal if feature A has more “distinctive” or “imbalanced” choice probabilities than feature B. In the sweater context, this means that, if turtleneck is either more popular or less popular than v-neck while red and blue are of similar popularity, then it is optimal to process the style feature first.

2.2.2. Kapuscinski and Tayur (1999) consider the same model as Lee and Tang (1998) and prove that condition (8) is not needed if the performance measure is replaced by the sum of the standard deviations of the production volumes in the intermediate stage, i.e., SD_{AB} for system AB and SD_{BA} for system BA. In other words, System AB is optimal if

$$SD_{AB} < SD_{BA}, \tag{10}$$

which holds if and only if (9) holds.

2.2.3. Jain and Paul (2001) also focus on the case of $m = n = 2$ but generalize the demand model of the previous two studies by incorporating characteristics of fashion-good markets, namely, market fragmentation and unpredictable customer preferences. To depict market fragmentation, the authors assume there are T independent market segments; within each segment τ, the demand follows a multinomial distribution with market size $\mathcal{N}_\tau$ and choice probabilities $\alpha_\tau(\beta_\tau)$ for the first choice of feature A (B) as in Lee and Tang (1998) and Kapuscinski and Tayur (1999). To capture the unpredictable customer preferences, the authors further assume $\alpha_\tau(\beta_\tau)$ to be random. Using this framework, the authors prove that if (i) the coefficient of variation ratios of all choice probabilities are greater than one, (ii) $V(\alpha_\tau) \leq V(\beta_\tau)$ for $\tau = 1, 2, ..., T$, then for any positive δ, there exists a positive integer Z_δ such that (iii) as long as the number of segments T is greater than Z_δ, $SD_{AB} + \delta < SD_{BA}$. Clearly, under conditions (i)-(iii), (10) holds, so System AB is optimal under the performance measure defined in Kapuscinski and Tayur (1999). In addition, the authors show these conditions also imply (7), hence System AB is optimal under the performance measure defined in Lee and Tang (1998) as well.

These works do not directly consider the cost implications of operations reversal.

3. Main Results

In this section, we identify conditions under which operations reversal is beneficial under our cost measure.

3.1. Necessary and Sufficient Condition

Working directly from the optimal cost comparison in (6), we can obtain the following necessary and sufficient condition for System AB to be optimal. Let

$$w = \frac{p}{h+p}.$$

PROPOSITION 1. *System AB is optimal under the cost measure if and only if*

$$\sum_{i=1}^{m} \int_0^{s_i^*} x f_i(x) dx \geq \sum_{j=1}^{n} \int_0^{\tilde{s}_j^*} x \tilde{f}_j(x) dx, \tag{11}$$

or equivalently,

$$\sum_{i=1}^{m} \left(E\left(s_i^* - D_i\right)^+ - w s_i^* \right) \leq \sum_{j=1}^{n} \left(E\left(\tilde{s}_j^* - \tilde{D}_j\right)^+ - w \tilde{s}_j^* \right), \tag{12}$$

or

$$\sum_{i=1}^{m} \left(E\left(D_i - s_i^*\right)^+ + (1-w) s_i^* \right) \leq \sum_{j=1}^{n} \left(E\left(\tilde{D}_j - s_j^*\right)^+ + (1-w) \tilde{s}_j^* \right). \tag{13}$$

Unfortunately, none of (11), (12) or (13) is intuitive and simple enough. In addition, they all depend on the knowledge of the optimal policy parameters. In the following, we present an example in which the optimal base-stock levels can be explicitly derived, so condition (11) can be expressed by the system parameters.

Example 1. Assume $m = n = 2$, $h = 1$, and $p = 5$. Suppose the demands for the four kinds of end products have independent uniform distributions: $D_{ij} \sim U[0, u_{ij}]$, $i, j = 1, 2$, where u_{ij} satisfy $u_{22} \geq 3 \max\{u_{12}, u_{21}\}$ and $\min\{u_{12}, u_{21}\} \geq 3u_{11}$. Under these conditions, the optimal base-stock levels are: $s_i^* = (3u_{i1} + 5u_{i2})/6$, $\tilde{s}_j^* = (3u_{1j} + 5u_{2j})/6$. With simple algebra we can show (11) is equivalent to

$$(u_{21} - u_{12})\left(25 + \frac{3u_{11}(u_{11} - 10)}{u_{12}u_{21}} + \frac{3(u_{12} + u_{21} - 10)}{u_{22}}\right) \geq 0. \tag{14}$$

If we further assume $u_{11} > 10$, then (14) can be simplified to $u_{21} \geq u_{12}$.

Note that in Example 1, the demands for different end products are assumed to be independent of each other, which implies $V_{AB} = \sum_{i,j} V(D_{ij}) = V_{BA}$, therefore Systems AB and BA are indifferent under the variability measure (7). However, this is no longer true under the cost measure (6).

While the conditions in Example 1 involve all system parameters, we next show that when demand is multivariate normal, the condition in Proposition 1 requires only partial demand parameters.

Corollary 1. *Suppose the demands follow a multivariate normal distribution. Then (12) is equivalent to*

$$\sum_{i=1}^{m} SD(D_i) \leq \sum_{j=1}^{n} SD(\tilde{D}_j), \qquad or \qquad SD_{AB} \leq SD_{BA}. \tag{15}$$

Thus, in this special case, the result reduces to (10). (This is consistent with the classic single-item inventory theory under normal demand; see, for example, p. 216 in Zipkin (2000)). In other words, for multivariate normal demand, the idea of minimizing total variability proposed by Kapuscinski and Tayur (1999) is equivalent to inventory cost minimization.

However, when demand is not multivariate normal, we do not have this equivalence. Below we present a counterexample to demonstrate that a smaller total variability (measured by either the total variance or the total standard deviation) could be accompanied by a higher inventory cost.

3.1.1. A Counterexample

Example 2. Consider the same demand structure as in Lee and Tang (1998) with $m = n = 2$. Suppose $\{D_{ij}\}$ are multinomially distributed with parameters $(\mathcal{N};\theta_{11},\theta_{12},\theta_{21},\theta_{22}) = (10;0.33,0.27,0.24,0.16)$, where θ_{ij} is the probability the customer will purchase A_iB_j. It can be verified that $V(D_1)=2.4$, $V(D_2)=2.4$, hence $V_{AB}=4.8$ and $SD_{AB}=3.0984$. Also, $V(\tilde{D}_1)=2.451$, $V(\tilde{D}_2)=2.451$, hence $V_{BA}=4.902$ and $SD_{BA}=3.1311$. Thus, in this example, we have $V_{AB}<V_{BA}$ and $SD_{AB}<SD_{BA}$. Therefore, whether we follow the measure of (7) or (10), System AB is optimal. Indeed, if we further assume independent choice probabilities α and β as in Lee and Tang (1998), then the above parameters imply $\alpha=0.6$ and $\beta=0.57$. Thus (9) holds.

However, assuming $h=1$ and $p=5$, we obtain $\Delta C^*=4.7249-4.5734>0$, so $C^*_{AB}>C^*_{BA}$. Thus, under the cost measure, System BA is optimal. This shows the cost measure and total variability measures are not equivalent.

3.2. Sufficient Condition

As shown above, in general, condition (11) is difficult to check because it includes the optimal policy parameters. Also, it does not provide intuition on what kind of demand characteristics can possibly lead to such a property. An exception is the multivariate normal demand case, for which it is sufficient to check the standard deviations of the intermediate product demands. Inspired by this special case, in this subsection, we aim to develop sufficient conditions that depend only on demand characteristics, not on the optimal policy parameters.

The following notion play important roles in our analysis.

Definition 1 (Convex Order). For two random variables Z_1 and Z_2, $Z_1 \le_{cx} Z_2$ if and only if $E(f(Z_1)) \le E(f(Z_2))$ for any convex function f.

The convex order is quite common. Examples include many familiar demand distributions, see, e.g., Table 1.1 of Müller and Stoyan (2002). The following facts will be useful:

Lemma 1 **(Müller and Stoyan (2002))**. *(a) $Z_1 \le_{cx} Z_2$ implies $E(Z_1)=E(Z_2)$ and $V(Z_1) \le V(Z_2)$. (b) Suppose both Z_1 and Z_2 are from one of the following families — Normal, Lognormal, Beta, Gamma, Weibull and Uniform. If (i) $E(Z_1)=E(Z_2)$ and (ii) $V(Z_1)\le V(Z_2)$, then $Z_1 \le_{cx} Z_2$.*

When the distributions of the two random variables fall into any the aforementioned families, Lemma 1(b) indicates that convex order can be identified by simply checking their means and variances.

Lemma 2 **(Müller and Stoyan (2002))**. *Let $F(\cdot)$ and $G(\cdot)$ be the cumulative distributions of Z_1 and Z_2, and $f(\cdot)$ and $g(\cdot)$ be the densities. If $E(Z_1) = E(Z_2)$, then $Z_1 \leq_{cx} Z_2$ provided any one of the following conditions holds:*

(a) g crosses f twice, first time from above, and the second time from below;

(b) G crosses F once from above.

Lemma 2 provides a sufficient condition to indicate convex ordering, which can be checked by simply examining the number of crossings of the distributions or density functions.

In addition, the following easier-to-check orders imply the convex order: (i) mean preserving spread order ($\leq_{mps}$); (ii) dangerous order ($\leq_D$) with equal means; (iii) dispersive order ($\leq_{disp}$) with equal means (see Müller and Stoyan (2002) for details). Thus, all relevant results in our paper still hold when the convex order is replaced by any of these stronger orders.

We begin with the special case when features A and B have the same number of choices.

Proposition 2. *Assume $M = N = \{1, ..., m\}$. System AB is optimal if*

$$D_i \leq_{cx} \tilde{D}_i, \quad i = 1, ..., m. \tag{16}$$

Proposition 2 states that a feature should be sequenced first in production if each of its choice has less variable demand in the sense of convex order than that of the other feature. In the sweater context, for instance, this means that if we are more certain about the aggregate demands of turtleneck and v-neck sweaters (i.e., the demand distributions are more concentrated around their means) than the aggregate demands for red and blue sweaters, respectively, then style should proceed color in production sequencing. Recall that we assume an arbitrary order for feature A's and feature B's choices. As long as there exists a particular order for the choices of each of feature A and feature B such that (16) holds, the proposition applies.

Note that from Lemma 1 (a) condition (16) implies

$$V(D_i) \le V(\tilde{D}_i), \quad i = 1, ..., m. \tag{17}$$

That is, the variance of the production volume of *each* intermediate product in System AB is lower than that in System BA. Because (17) implies (7), the sufficient condition (16) is stronger than the criteria of total variability. While (7) measures the total variance of the production volumes at the feature level, (16) goes deeper into the choice level.

For multivariate normal demand, from Lemma 1, condition (16) is equivalent to $E(D_i) = E(\tilde{D}_i)$ and $SD(D_i) \le SD(\tilde{D}_i)$ for all $i \in M$, which is stronger than (15).

To relate to the sufficient conditions developed in the literature, let us consider multinomial demand adopted by Lee and Tang (1998), Kapuscinski and Tayur (1999) and Jain and Paul (2001). Assume $m = n = 2$, and the end product demands are multinomially distributed with parameters $(\mathcal{N}; \theta_{11}, \theta_{12}, \theta_{21}, \theta_{22})$, where the market size $\mathcal{N}$ is random. Then a sufficient condition for ensuring a lower total variance is given by (9), where the choice probabilities α and β satisfy

$$\alpha = \theta_{11} + \theta_{12}, \qquad \beta = \theta_{11} + \theta_{21}. \tag{18}$$

However, our sufficient condition requires $D_{11} + D_{12} \le_{cx} D_{11} + D_{21}$, and $D_{21} + D_{22} \le_{cx} D_{12} + D_{22}$. These imply $\alpha = \beta$. Thus, these two sets of sufficient conditions do not overlap.

A numerical example for Proposition 2 is given below.

Example 3. Consider multivariate normally distributed end product demands with $m = n = 2$. Suppose $\{D_{ij}\}$ is distributed with mean vector $\mu = \{50, 60, 60, 70\}$ and covariance matrix

$$\Sigma = \begin{pmatrix} 6 & -3 & -1.5 & 0 \\ -3 & 8 & 0 & -1.5 \\ -1.5 & 0 & 8 & -3 \\ 0 & -1.5 & -3 & 6 \end{pmatrix}.$$

Assume $h = 1$ and $p = 5$. Then it is easy to verify that $E(D_1) = 110 = E(\tilde{D}_1)$, $E(D_2) = 130 = E(\tilde{D}_2)$, and $Var(D_1) = 8 < 11 = Var(\tilde{D}_1)$, $Var(D_2) = 8 < 11 = Var(\tilde{D}_2)$. By the properties of multivariate normal distribution, D_i and $\tilde{D}_j$ are normally distributed. Referring to Müller and Stoyan (2002), when both densities of X and Y are normal with $E(X) = E(Y)$ and $Var(X) \leq Var(Y)$, then $X \leq_{cx} Y$. Therefore, $D_1 \leq_{cx} \tilde{D}_1$, $D_2 \leq_{cx} \tilde{D}_2$ and Proposition 2 applies. Indeed it can be verified that $\Delta C^* = 8.4802 - 9.9439 < 0$, i.e., feature A should be processed first.

Next, to gain more intuitive insights from Proposition 2, we examine the following three settings.

1. (Additive end-product demand) Suppose the end product demands all have the same mean but different variabilities, and hence the different feature choices only affect the variabilities of the end product demands. Further assume that the effect of feature A's choice and that of feature B's choice are additive. Specifically, let

$$D_{ij} = \mu + \epsilon_i + \tilde{\epsilon}_j, \tag{19}$$

where μ is a nonnegative constant, and ϵ_i and $\tilde{\epsilon}_j$, $i, j \in M$, are random variables taking values in the interval $[-\frac{\mu}{2}, \frac{\mu}{2}]$ with zero mean and standard deviation σ. Here, ϵ_i ($\tilde{\epsilon}_j$) can be viewed as the variability factor associated with the ith (jth) choice of feature A (B). Further assume ϵ_i and $\tilde{\epsilon}_j$ are independent of each other for any i and j, but ϵ_i ($\tilde{\epsilon}_j$) can be dependent across i (j). Then (16) can be written as $m\mu + m\epsilon_i + \sum_{j=1}^m \tilde{\epsilon}_j \leq_{cx} m\mu + \sum_{j=1}^m \epsilon_j + m\tilde{\epsilon}_i$, which is equivalent to

$$m\epsilon_i + \sum_{j=1}^m \tilde{\epsilon}_j \leq_{cx} \sum_{j=1}^m \epsilon_j + m\tilde{\epsilon}_i. \tag{20}$$

Note that (20) implies that

$$V\left(\sum_{j=1}^m \tilde{\epsilon}_j\right) \leq V\left(\sum_{j=1}^m \epsilon_j\right). \tag{21}$$

Take $m = 2$, for instance, (21) reduces to

$$Cov(\tilde{\epsilon}_1, \tilde{\epsilon}_2) \leq Cov(\epsilon_1, \epsilon_2). \tag{22}$$

Therefore Proposition 2 also implies that the feature (in this case, feature A) with more positively correlated (or less negatively correlated) choice factors should be processed first. This way, the less positively correlated (or more negatively correlated) choices of the other feature (in this case, feature B) at the end-product level help reduce the demand variability at the intermediate stage.

This result remains true for general m. In the sweater context, the different choices of a feature can be negatively correlated. For instance, if red is the color in fashion of the season, then typically fewer consumers would purchase blue sweaters. On the other hand, the different choices of a feature can also be positively correlated. For example, the sweaters are usually introduced and advertised in collections. When one collection is welcomed by the consumers, the sales of different styles (turtleneck and v-neck sweaters) in the collection are all boosted. Based on the extent of negative or positive correlation, the feature with less negatively correlated (or more positively correlated) choices should be given priority. Suppose the company's historical data indicates that it is always the case that one color choice is extremely popular while the other color very unpopular (the popular color could vary from period to period), while the demands for turtleneck and v-neck styles are either both high or both low, then the style feature should be processed first in order to utilize the negative correlation of color choices in the aggregate intermediate demands.

2. (Multiplicative end-product demand) Again consider the case of same end product demand mean. But assume alternatively that the effect of feature A's choice and that of feature B's choice are multiplicative. Specifically, let

$$D_{ij} = \mu + \epsilon_i \tilde{\epsilon}_j, \tag{23}$$

where μ is a nonnegative constant, and ϵ_i and $\tilde{\epsilon}_j$, $i, j \in M$, are random variables taking values in the interval $[-\sqrt{\mu}, \sqrt{\mu}]$ with zero mean and standard deviation σ. Everything else is the same as in the additive setting above. Then (16) can be written as $m\mu + \sum_{j=1}^{m} \epsilon_i \tilde{\epsilon}_j \leq_{cx} m\mu + \sum_{j=1}^{m} \epsilon_j \tilde{\epsilon}_i$, which is equivalent to

$$\sum_{j=1}^{m} \epsilon_i \tilde{\epsilon}_j \leq_{cx} \sum_{j=1}^{m} \epsilon_j \tilde{\epsilon}_i. \tag{24}$$

Note that (24) implies (21). Thus Proposition 2 implies that, in this setting too, it is optimal to first process the feature with more positively correlated (or less negatively correlated) choice factors.

3. (Mean-preserving intermediate product demand) Noting that mean-preserving spread order implies convex order, we present an example below where the intermediate product demands in

System AB and System BA actually satisfy mean-preserving spread order. Consider the following end product demands with $m = n = 2$:

$$D_{11} = \alpha_1 X + (1-\alpha_2)\mu_X, \quad D_{12} = (\alpha_2 - \alpha_1)\mu_X,$$

$$D_{21} = (\alpha_2 - \alpha_1)X, \quad D_{22} = \gamma + \frac{\alpha_1 - \alpha_2}{\beta_1 - \beta_2}(\beta_1 \mu_X - \beta_2 X),$$

with X being a nonnegative random variable, $0 \leq \alpha_i,\ \beta_i \leq 1,\ \gamma > 0,\ \mu_X \equiv E(X)$ and assume that $\alpha_1 \leq \alpha_2$ and $\beta_1 \leq \beta_2$.

Let $Y = \gamma + \frac{\alpha_1 - \alpha_2}{\beta_1 - \beta_2}(\mu_X - X)$. Note that $\mu_Y \equiv E(Y) = \gamma$. Then the intermediate products are:

$$D_1 = D_{11} + D_{12} = \alpha_1 X + (1-\alpha_1)\mu_X, \quad \tilde{D}_1 = D_{11} + D_{21} = \alpha_2 X + (1-\alpha_2)\mu_X,$$

$$D_2 = D_{21} + D_{22} = \beta_1 Y + (1-\beta_1)\mu_Y, \quad \tilde{D}_2 = D_{12} + D_{22} = \beta_2 Y + (1-\beta_2)\mu_Y.$$

Therefore $D_i \leq_{mps} \tilde{D}_i$, which further imply $D_i \leq_{cx} \tilde{D}_i$, $i = 1, 2$.

From the construction of the end product demands, we see that D_{12} is constant, D_{21} and D_{22} are negatively correlated (due to the opposite signs of the coefficients of X), D_{11} and D_{21} are positively correlated (with randomness both solely from X with positive coefficients). As a result, we obtain consistent insights from the above three settings.

3.3. Generalization of Sufficient Condition

In this subsection we aim to extend the sufficient condition (16) to more general settings. Note that (16) implicitly requires the means of the corresponding intermediate product demand pair (D_i and $\tilde{D}_i$) to be equal. Our first endeavor is to allow the intermediate product means are unequal by utilizing the following stochastic order:

Definition 2 (Dilation Order). For two random variables Z_1 and Z_2, $Z_1 \leq_{dil} Z_2$ if and only if $Z_1 - E(Z_1) \leq_{cx} Z_2 - E(Z_2)$.

The notion of dilation order was introduced by Hickey (1986). The only application in the inventory literature that we are aware of is Huang et al. (2015), who use this order to numerically examine the impact of surge demand duration on sourcing strategies. The dilation order is location independent: if $Z_1 \leq_{dil} Z_2$, we also have $Z_1 + c \leq_{dil} Z_2$ and $Z_1 \leq_{dil} Z_2 + c$ for any constant c. The following orders also imply the dilation order: the dispersive ($\leq_{disp}$) and right spread ($\leq_{RS}$) orders (Shaked and Shanthikumar (1994)), the location independent riskier order ($\leq_{lir}$, Fagiuoli et al. (1999)), and mean residual life order ($\leq_{mrl}$) under certain conditions (Belzunce et al. (1997)). Thus, the relevant results in this paper still hold when the dilation order is replaced by any of these stronger orders.

Note that the dilation order is defined with the aid of convex order, although it does not require the two random variables to have equal means. Therefore the easier-to-check techniques we provided for convex order in §3.2 can be applied to check $Z_1 - E(Z_1) \leq_{cx} Z_2 - E(Z_2)$, and hence to verify the dilation order. In addition, Belzunce et al. (2005) develop a family of tests for the dilation order, and thus provide a guidance for how to verify this order using empirical data.

The generalized sufficient condition is given by Proposition 3.

Proposition 3. *Assume $M = N = \{1, ..., m\}$. System AB is optimal if*

$$D_i \leq_{dil} \tilde{D}_i, \;\; i = 1, ..., m. \tag{25}$$

Note that then $E(D_i) = E(\tilde{D}_i)$, (25) is equivalent to (16).

Thus, to ensure System AB to be optimal, $E(D_i) = E(\tilde{D}_i)$ is not necessary, what we need is that the centered variables $D_i - E(D_i)$ and $\tilde{D}_i - E(\tilde{D}_i)$ satisfy the convex order. Since the centered variables are simply shifted by their means from the original variables (intermediate product demands), the intuitions and insights we have obtained for (16) all carry over to (25). We further illustrate Proposition 3 by revisiting Example 3 in the following example.

Example 4. Consider the numerical setting in Example 3. If we change the mean of D_{12} from 60 to 80, i.e., $\mu = \{50, 80, 60, 70\}$, then the intermediate product demand means are: $E(D_1) = 130$, $E(\tilde{D}_1) = 110$, $E(D_2) = 130$ and $E(\tilde{D}_2) = 150$. Consequently, we no longer have $E(D_i) = E(\tilde{D}_i), i = 1, 2$ as in Example 3. However, since the covariance matrix remains the same as before, $Var(D_i) < Var(\tilde{D}_i), i = 1, 2$ still hold. Therefore, by the same argument as in Example 3, we have $D_i - E(D_i) \leq_{cx} \tilde{D}_i - E(\tilde{D}_i), \; i = 1, 2$, so Proposition 3 applies. Indeed it can be verified that $\Delta C^* = -1.4637 < 0$, i.e., feature A should be processed first.

So far our analysis covers the case of $m = n$. Our second endeavor in generalizing (16) is to consider $m \neq n$. For simplicity, we begin with the special case of $n = m+1$.

PROPOSITION 4. *Suppose that* $M = \{1, 2, ..., m\}, N = \{1, 2, ..., m+1\}$. *System AB is optimal if*

$$D_i \leq_{cx} \tilde{D}_i, \quad i = 1, ..., m-1, \quad \textit{and} \quad D_m \leq_{cx} \tilde{D}_m + \tilde{D}_{m+1}. \tag{26}$$

Note that in this proposition, when $m \neq n$, we can no longer require the intermediate production volume for each choice of feature A to be less variable than that of feature B. Take $m = 2$, for example, we replace $D_2 \leq_{cx} \tilde{D}_2$ with $D_2 \leq_{cx} \tilde{D}_2 + \tilde{D}_3$, which means the intermediate production volume for the second choice of feature A is less variable (in the convex order sense) than the combined intermediate production volumes of both the second and third choices of feature B.

Proposition 4 shows that feature A should be processed first if it has less variety and each choice is less variable than feature B. In the following, we demonstrate that when feature A is less variable but has more variety, e.g., $m = 3$, $n = 2$ with

$$D_1 \leq_{cx} \tilde{D}_1, \quad D_2 + D_3 \leq_{cx} \tilde{D}_2, \tag{27}$$

System AB may not have lower cost.

EXAMPLE 5. Suppose $M = \{1, 2, 3\}, N = \{1, 2\}$, and $\{D_{ij}, i = 1, 2, 3, j = 1, 2\}$ are multivariate normally distributed with mean vector $\mu = \{50, 130, 60, 80, 70, 90\}$ and covariance matrix

$$\Sigma = \begin{pmatrix} 6 & -3 & -1 & 0 & -1 & 0 \\ -3 & 8 & 0 & -1 & 0 & -1 \\ -1 & 0 & 6 & -3 & -1 & 0 \\ 0 & -1 & -3 & 8 & 0 & -3 \\ -1 & 0 & -1 & 0 & 7 & -3 \\ 0 & -1 & 0 & -3 & -3 & 8 \end{pmatrix}.$$

Assume $h = 1$ and $p = 5$. Then $E(D_1) = 180 = E(\tilde{D}_1)$, $E(D_2 + D_3) = 300 = E(\tilde{D}_2)$, and $V(D_1) = 8 < 13 = V(\tilde{D}_1)$, $V(D_2 + D_3) = 9 < 14 = V(\tilde{D}_2)$. By the properties of multivariate normal distribution,

D_i, $\tilde{D}_j$ and D_2+D_3 are normally distributed. By Müller and Stoyan (2002), we have (27). It is easy to calculate that $\Delta C^* = 12.9775 - 11.0142 > 0$, i.e., feature B should be processed first.

Note that in Example 5, $V_{AB} = 25 \le 27 = V_{BA}$, hence System BA has a larger total variability in the intermediate stage. But it is still optimal to process feature B first, which has one less choice than feature A. This illustrates that an extra choice can have a larger impact on operating cost than a larger variability.

For multivariate normal demand, from Lemma 1, condition (26) is equivalent to $E(D_i) = E(\tilde{D}_i)$, $SD(D_i) = SD(\tilde{D}_i),\ i = 1, ..., m-1$, and $E(D_m) = E(\tilde{D}_m + \tilde{D}_{m+1})$, $SD(D_m) = SD(\tilde{D}_m + \tilde{D}_{m+1})$, which implies (15).

In general, we can extend Proposition 4 to the general case when $m \neq n$.

Proposition 5. *Suppose $M = \{1,2,...,m\}$, and $N = \{1,2,...,n\}$ with $m < n$. System AB is optimal if there exist disjoint sets K_i, $i = 1,...,m$ such that (i) $\bigcup_{i=1}^{m} K_i \subset N$ and (ii) $D_i \le_{cx} \sum_{j\in K_i} \tilde{D}_j$.*

Given feature A with less choices than feature B, and the demand for each choice of feature A being less variable than the demand for some subset of feature B's choices, Proposition 5 demonstrates that priority of production should be assigned to feature A for sure.

4. Effects of Cost Parameters and Lead Times

So far we have restricted the cost parameters and lead times to be feature independent. In this section we shall relax these conditions and examine its influence on operations reversal.

4.1. Different Cost Parameters in Stage 1

In this subsection, we allow the unit per-period holding cost and penalty cost at Stage 1 to vary under different production sequences. Specifically, we denote the unit per-period holding cost and penalty cost at Stage 1 for System AB and System BA as h^1_{AB}, p^1_{AB} and h^1_{BA}, p^1_{BA}, respectively. Further define

$$R^1_{AB}(x,y) = h^1_{AB}(x-y)^+ + p^1_{AB}(y-x)^+,$$

and

$$R^1_{BA}(x,y) = h^1_{BA}(x-y)^+ + p^1_{BA}(y-x)^+.$$

Therefore we have

$$\Delta C^* \equiv C^*_{AB} - C^*_{BA} = \sum_{i\in M} E(R^1_{AB}(s^*_i, D_i)) - \sum_{j\in N} E(R^1_{BA}(\tilde{s}^*_j, \tilde{D}_j)). \tag{28}$$

Define

$$\gamma_{AB} = h^1_{AB}/p^1_{AB}, \text{ and } \gamma_{BA} = h^1_{BA}/p^1_{BA}.$$

Note that $s^*_i = F^{-1}_{D_i}\left(\frac{1}{1+\gamma_{AB}}\right)$ and $\tilde{s}^*_j = F^{-1}_{\tilde{D}_j}\left(\frac{1}{1+\gamma_{BA}}\right)$. Due to the complexity in the convoluted distribution of the intermdiate product demand, we make some simplification assumptions for tractability. Assume that the end product demands are independent and identically distributed. Consider the case of $m = n$. Then the intermediate product demands D_i and $\tilde{D}_j$ all have the same distribution, i.e., $D_i, D_j \sim D_0$, where D_0 is a random variable with the common distribution. Hence, s^*_i and s^*_j are independent of i and j. With these assumptions, we obtain

Proposition 6. *Assume that D_{ij} are independent and identically distributed, and that $M = N = \{1, ..., m\}$. System AB is optimal if and only if the following condition holds:*

$$\begin{aligned} &\eta\left(\gamma_{AB} E\left(F^{-1}_{D_0}\left(\frac{1}{1+\gamma_{AB}}\right) - D_0\right)^+ + E\left(D_0 - F^{-1}_{D_0}\left(\frac{1}{1+\gamma_{AB}}\right)\right)^+\right) \\ &\le \gamma_{BA} E\left(F^{-1}_{D_0}\left(\frac{1}{1+\gamma_{BA}}\right) - D_0\right)^+ + E\left(D_0 - F^{-1}_{D_0}\left(\frac{1}{1+\gamma_{BA}}\right)\right)^+, \end{aligned} \tag{29}$$

where $\eta = p^1_{AB}/p^1_{BA}$.

When the end demands are normally distributed, the sufficient and necessary condition can be further simplified as in the following corollary.

Corollary 2. *Assume that D_{ij} are independent and identically distributed, and that $M = N = \{1, ..., m\}$. Further assume that D_0 is normally distributed with parameter (μ_0, σ_0). Let Z denote a standard normal random variable with distribution function Φ, and $z^*_{AB} = \Phi^{-1}\left(\frac{1}{1+\gamma_{AB}}\right)$, $z^*_{BA} = \Phi^{-1}\left(\frac{1}{1+\gamma_{BA}}\right)$. Then System AB is optimal if and only if*

$$\eta\left(\gamma_{AB} E\left((z^*_{AB} - Z)^+\right) + E\left((Z - z^*_{AB})^+\right)\right) \le \gamma_{BA} E\left((z^*_{BA} - Z)^+\right) + E\left((Z - z^*_{BA})^+\right). \tag{30}$$

Proposition 6 provides the sufficient and necessary condition that the holding costs and penalty costs of both systems in Stage 1 need to satisfy for System AB to be optimal. The condition is complex and not intuitive. In the next proposition, we show that if System AB's cost parameters are lower than those for System BA, then System AB is indeed optimal.

Proposition 7. *Assume that D_{ij} are independent and identically distributed, and that $M = N = \{1, ..., m\}$. System AB is optimal if $h^1_{AB} \leq h^1_{BA}$ and $p^1_{AB} \leq p^1_{BA}$.*

Note that when $\gamma_{AB} = \gamma_{BA}$, the condition (29) given by Proposition 6 becomes $\eta \leq 1$, which also imply $h^1_{AB} \leq h^1_{BA}$ and $p^1_{AB} \leq p^1_{BA}$.

When end product demands are not independent and identically distributed, and the choices of features A and B are different, the relationship between cost parameters and the optimal production sequence become even more complex. The multiple factors involved are convoluted and quite difficult to be analyzed. Alternatively, we present numerical examples to demonstrate that, with same cost parameters, the optimal sequence could vary with demand structures.

Example 6. Consider the demand structure as in Lee and Tang (1998) with $m = n = 2$. Suppose $\{D_{ij}\}$ are multinomially distributed with parameters $(\mathcal{N}; \theta_{11}, \theta_{12}, \theta_{21}, \theta_{22}) = (10; 0.33, 0.27, 0.24, 0.16)$, where θ_{ij} is the probability the customer will purchase A_iB_j. Assume $h_{AB} = 1$, $p_{AB} = 5$, and $h_{BA} = 1.1$, $p_{BA} = 5.1$. It can be verified that $\Delta C^* = 4.5734 - 5.0196 < 0$, so $C^*_{AB} < C^*_{BA}$, i.e., System AB is optimal.

If we vary the demand parameters to be $(\mathcal{N}; \theta_{11}, \theta_{12}, \theta_{21}, \theta_{22}) = (10; 0.2, 0.4, 0.1, 0.3)$. It can be verified that $\Delta C^* = 4.7249 - 4.6025 > 0$, so $C^*_{AB} > C^*_{BA}$, i.e., System BA is optimal.

Example 7. Consider the demand structure as in Lee and Tang (1998) with $m = 2, n = 3$. Suppose $\{D_{ij}\}$ are multinomially distributed with parameters $(\mathcal{N}; \theta_{11}, \theta_{12}, \theta_{13}, \theta_{21}, \theta_{22}, \theta_{23}) = (10; 0.3, 0.3, 0.2, 0.1, 0.05, 0.05)$, where θ_{ij} is the probability the customer will purchase A_iB_j. Assume $h_{AB} = 2$, $p_{AB} = 6$, and $h_{BA} = 1$, $p_{BA} = 5$. It can be verified that $\Delta C^* = 6.1469 - 6.7794 < 0$, so $C^*_{AB} < C^*_{BA}$, i.e., System AB is optimal.

If we change the demand parameters to be $(\mathcal{N};\theta_{11},\theta_{12},\theta_{13},\theta_{21},\theta_{22},\theta_{23}) = (10;0.2,0.1,0.1,0.1,0.2,0.3)$. It can be verified that $\Delta C^* = 7.6382 - 6.9299 > 0$, so $C^*_{AB} > C^*_{BA}$, i.e., System BA is optimal.

4.2. Different Production Lead Times for Features A and B

In this subsection, we return to the original model with identical cost parameters for the two systems, and relax the assumption that the production lead times of both features to be one. In other words, we allow the lead times of adding feature A and feature B to be greater than one. Let L_A and L_B to be the lead time for adding features A and B, respectively. Assume that these lead times do not depend on the sequence of production.

For stages 1 and 2, we define the relevant inventory variables first. As we have assumed in §2.1 that the production system has no capacity constraints and no backlog is allowed, the inventory on order is equivalent to inventory in transit. For intermediate product A_i, at time t, $I_{A_i}(t)$ denotes the on hand inventory level, $ST_{A_i}(t)$ is the shortage, $IN_{Ai}(t)$ denotes the net inventory level (on hand inventory minus shortage), $IT_{A_i}(t)$ is the inventory in transit to stage 1, $ITP_{A_i}(t)$ is the inventory in transit position (inventory in transit plus net inventory). The variables for intermediate product B_j and end product A_iB_j are similarly defined. All notations are summarized in Table 2. The detailed expressions and derivations for these variables can be found in Appendix C.

Assume general holding and penalty costs h^i and p^i with $i = 0,1,2$, corresponding to stages 0, 1 and 2, irrespective of processing sequence. We obtain the long run average cost of system AB as follows.

$$
\begin{aligned}
& C_{AB}(\mathbf{s},\mathbf{S}) \\
= {} & E\left(h^0 \sum_{i\in M} IT_{A_i}(t,1)\right) + E\left(h^1\left(\sum_{i\in M} I_{A_i}(t,1) + \sum_{i\in M, j\in N} IT_{A_iB_j}(t,2)\right) + p^1 \sum_{i\in M} ST_{A_i}(t,1)\right) \\
& + E\left(h^2 \sum_{i\in M, j\in N} I_{A_iB_j}(t,2) + p^2 \sum_{i\in M, j\in N} ST_{A_iB_j}(t,2)\right) \\
= {} & h^0 E\left(\sum_{i\in M} D_i(L_A)\right) + h^1 E\left(\sum_{i\in M}(s_i - D_i(L_A))^+ + \sum_{i\in M, j\in N} D_{ij}(L_B)\right) \\
& + p^1 E\left(\sum_{i\in M}(D_i(L_A) - s_i)^+\right) + E\left(h^2 \sum_{i\in M, j\in N}(S_{ij} - D_{ij}(L_B))^+ + p^2 \sum_{i\in M, j\in N}(D_{ij}(L_B) - S_{ij})^+\right)
\end{aligned}
$$

	Stage 1	Stage 2
	System AB/System BA	
Lead time	L_A	L_B
On hand inventory	$I_{A_i}(t,1)/I_{B_j}(t,1)$	$I_{A_iB_j}(t,2)$
Shortage	$ST_{A_i}(t,1)/ST_{B_j}(t,1)$	$ST_{A_iB_j}(t,2)$
Net inventory	$IN_{A_i}(t,1)/IN_{B_j}(t,1)$	$IN_{A_iB_j}(t,2)$
Inventory in transit	$IT_{A_i}(t,1)/IT_{B_j}(t,1)$	$IT_{A_iB_j}(t,2)$
Inventory in transit position	$ITP_{A_i}(t,1)/ITP_{B_j}(t,1)$	$ITP_{A_iB_j}(t,2)$

Table 2 Inventory Notations for the Two Stages in System AB

Following a similar process, we can define notations and derive the long run average cost for system BA.

$$
\begin{aligned}
& C_{BA}(\tilde{\mathbf{s}},\tilde{\mathbf{S}}) \\
&= E\left(\sum_{j\in N} h^0(IT_{B_j}(t,1)\right) + E\left(h^1\left(\sum_{j\in N} I_{B_j}(t,1) + \sum_{i\in M, j\in N} IT_{A_iB_j}(t,2)\right) + p^1 \sum_{j\in N} ST_{B_j}(t,1)\right) \\
&\quad + E\left(h^2 \sum_{i\in M, j\in N} I_{A_iB_j}(t,2) + p^2 \sum_{i\in M, j\in N} ST_{A_iB_j}(t,2)\right) \\
&= h^0 E\left(\sum_{j\in N} \tilde{D}_j(L_B)\right) + h^1 E\left(\sum_{j\in N}\left(\tilde{s}_j - \tilde{D}_j(L_B)\right)^+ + \sum_{i\in M, j\in N} D_{ij}(L_A)\right) \\
&\quad + p^1 E\left(\sum_{j\in N}\left(\tilde{D}_j(L_B) - \tilde{s}_j\right)^+\right) + E\left(h^2 \sum_{i\in M, j\in N}\left(\tilde{S}_{ij} - D_{ij}(L_A)\right)^+ + p^2 \sum_{i\in M, j\in N}\left(D_{ij}(L_A) - \tilde{S}_{ij}\right)^+\right)
\end{aligned}
$$

When $L_A = L_B = 1$, it is easy to verify that, $C_{AB}(\mathbf{s},\mathbf{S})$ and $C_{BA}(\tilde{\mathbf{s}},\tilde{\mathbf{S}})$ reduce to (2) and (3).

For general L_A and L_B, as in Section 4.1, we again assume that the end product demands are independent and identically distributed, i.e., $D_{ij} \sim D$ for $i \in M$ and $j \in N$. Then $D_{ij}(L_A) \sim D(L_A)$, $D_{ij}(L_B) \sim D(L_B)$, $D_i(L_A) \sim D(nL_A)$, and $\tilde{D}_j(L_B) \sim D(mL_B)$, where $D(L)$ denotes the cumulative demand during a lead time L (or the lead time demand) given the single-period demand D.

The long-run average system costs then become:

$$C_{AB}(\mathbf{s},\mathbf{S})$$

$$= h^0 mE\left(D(nL_A)\right) + h^1 mnE(D(L_B)) + \sum_{i\in M} E\left(h^1\left(s_i - D(nL_A)\right)^+ + p^1\left(D(nL_A) - s_i\right)^+\right)$$
$$+ \sum_{i\in M, j\in N} E\left(h^2\left(S_{ij} - D(L_B)\right)^+ + p^2\left(D(L_B) - S_{ij}\right)^+\right),$$

and

$$C_{BA}(\tilde{\mathbf{s}}, \tilde{\mathbf{S}})$$
$$= h^0 nE\left(D(mL_B)\right) + h^1 mnE(D(L_A)) + \sum_{j\in N} E\left(h^1\left(\left(\tilde{s}_j - D(mL_B)\right)^+\right) + p^1\left(D(mL_B) - \tilde{s}_j\right)^+\right)$$
$$+ \sum_{i\in M, j\in N} E\left(h^2\left(\tilde{S}_{ij} - D(L_A)\right)^+ + p^2\left(D(L_A) - \tilde{S}_{ij}\right)^+\right)$$

Then it is easy to derive that $s_i^* = F^{-1}_{D(nL_A)}(w^1)$, $\tilde{s}_j^* = F^{-1}_{D(mL_B)}(w^1)$, $S_{ij}^* = F^{-1}_{D(L_B)}(w^2)$, and $\tilde{S}_{ij}^* = F^{-1}_{D(L_A)}(w^2)$, where $w^1 \equiv \frac{p^1}{h^1+p^1}$ and $w^2 \equiv \frac{p^2}{h^2+p^2}$.

By using the above notations, it is easy to derive that,

$$\begin{aligned}\Delta C^* \equiv C_{AB}^* - C_{BA}^* &= mnE(D)(L_B - L_A)(h^1 - h^0) \\ &\quad + mE\left(R^1\left(F^{-1}_{D(nL_A)}(w^1), D(nL_A)\right)\right) + mnE\left(R^2\left(F^{-1}_{D(L_B)}(w^2), D(L_B)\right)\right) \\ &\quad - nE\left(R^1\left(F^{-1}_{D(mL_B)}(w^1), D(mL_B)\right)\right) - mnE\left(R^2\left(F^{-1}_{D(L_A)}(w^2), D(L_A)\right)\right).\end{aligned} \tag{31}$$

When features A and B have the same number of choices but require different lead times to process, we can obtain a sufficient and necessary condition for System AB to be optimal.

Proposition 8. *Assume $M = N = \{1, ..., m\}$ and D_{ij} are independent and identically distributed. If $L_A > L_B$, then the sequence AB is optimal if and only if*

$$\Delta h \Delta L \geq \frac{1}{mE(D)} E(\Delta R(L_A) - \Delta R(L_B)), \tag{32}$$

where $\Delta h \equiv h^1 - h^0$, $\Delta L \equiv L_A - L_B$, and $\Delta R(L) \equiv R^1\left(F^{-1}_{D(mL)}(w^1), D(mL)\right) - mR^2\left(F^{-1}_{D(L)}(w^2), D(L)\right)$.

Proposition 8 demonstrates that, the feature with the longer production lead time should be processed earlier only if it can lead to sufficient pipeline inventory cost reduction. On the other hand, when $E(\Delta R(L_A) - \Delta R(L_B)) \le 0$, then (32) is automatically satisfied, and the feature with the longer production lead time should always be prioritized. In the sweater context, knitting is typically more time-consuming than dyeing. By knitting first and then dyeing, cost is lower since the pipeline inventory spends more time at the earlier stage (in the form of bleached yarns before knitting completes) than at the later stage (in the form of knitted yarns before dyeing completes).

When the end demands are further assumed to be normally distributed, the sufficient and necessary condition provided in Proposition 8 can be simplified as follows.

Corollary 3. *Assume that D_{ij} are independent and identically distributed, and that $M = N = \{1, ..., m\}$, and that $L_A > L_B$. Further assume that D is normally distributed with parameter (μ, σ). Let Z denote a standard normal random variable with distribution function Φ, and $z^i = \Phi^{-1}(w^i)$, $i = 1, 2$. Then (32) is equivalent to*

$$\Delta h\sqrt{m}(\sqrt{L_A} + \sqrt{L_B})\frac{\mu}{\sigma} \ge E\left(h^1(z^1 - Z)^+ + p^1(Z - z^1)^+ - \sqrt{m}\left(h^2(z^2 - Z)^+ + p^2(Z - z^2)^+\right)\right). \tag{33}$$

In practice, we often have $h^1 < h^2$ and $p^1 < p^2$ due to value adding along the supply chain. Furthermore, h^2 and p^2 are multiplied by $\sqrt{m}$ in (33). So the right hand side of (33) is negative in most cases, making (33) hold trivially, i.e., it is optimal to process feature A first. In addition, note that the left hand side of (33) increases in m while the right hand side decreases in m. Therefore, Corollary 3 also indicates that, under normal demand, with a larger number of feature choices (m), it is more advantageous to process feature A first. Another interesting observation is that with normally distributed demand, as long as $L_A > L_B$, the sufficient condition to guarantee feature A's priority is independent of the absolute value of $L_A - L_B$, but instead depend on the summation of the square roots of the lead times.

Next we present a result for the case when features A and B have the same lead time (could be larger than one), but have different number of choices.

Proposition 9. *Assume that D_{ij} are independent and identically distributed. If $L_A = L_B = L$, and $n = km$ with $k \in \mathbb{Z}^+$, then it is optimal to process feature A first.*

Proposition 9 states an intuitive result that given everything else being equal, the feature with fewer choices should be given priority. This is because by postponing the larger differentiation (processing the feature with more choices), the benefit of pooling is maximally utilized.

When end product demands are not independent and identically distributed, the analysis of the lead time effect becomes less tractable. The following numerical examples demonstrate that, with same lead times and cost parameters, the optimal sequence could vary with demand structures. Further analysis is left for future research.

Example 8. Suppose $M = \{1,2\}, N = \{1,2,3\}$, and $\{D_{ij}, i = 1,2,\ j = 1,2,3\}$ are multivariate normally distributed with mean vector $\mu = \{4,10,7,8,6,8\}$ and covariance matrix

$$\Sigma = \begin{pmatrix} 8 & -5 & -4 & 9 & -1 & 0 \\ -5 & 11 & -4 & -1 & 10 & -1 \\ -4 & -4 & 9 & -3 & -1 & 9 \\ 9 & -1 & -3 & 8 & -4 & -3 \\ -1 & 10 & -1 & -4 & 8 & -4 \\ 0 & -1 & 9 & -3 & -4 & 8 \end{pmatrix}.$$

Assume that $h_0 = 0.5,\ h_1 = 1,\ h_2 = 2,\ p_1 = 5,\ p_2 = 10,\ L_A = 1,\ L_B = 2$. It can be verified that $\Delta C^* = 147.67 - 149.09 < 0$, i.e., feature A should be processed first.

If we reset the demand parameters to be $\mu = \{5,10,6,8,7,9\}$ and covariance matrix

$$\Sigma = \begin{pmatrix} 6 & -4 & -3 & 4 & -1 & 0 \\ -4 & 8 & -1 & -1 & 5 & -1 \\ -3 & -1 & 6 & -3 & -1 & 4 \\ 4 & -1 & -3 & 8 & -2 & -3 \\ -1 & 5 & -1 & -2 & 7 & -3 \\ 0 & -1 & 4 & -3 & -3 & 8 \end{pmatrix}.$$

It can be verified that $\Delta C^* = 154.66 - 145.22 > 0$, i.e., feature B should be processed first.

5. Conclusions

Summary of Results and Contributions

In this paper we revisited the operations reversal problem inspired by industrial practice and studied by several authors previously. The focus is on the effect of demand uncertainty arising from multiple products. While the previous studies use the objective of the total variability of production volumes, we examine the same problem with the objective of the long-run average inventory cost. We find that reducing total variability, measured either by total variance or total standard deviation, does not necessarily imply inventory cost reduction under our measure, although the two objectives are related in certain cases. Unlike the conditions in the previous works that focus on product feature level, we look into the choice level of each feature. By leveraging stochastic orders criterions at the individual choice level, we provide a set of sufficient conditions under which operations reversal yields cost reduction. Specifically, we utilize the convex order and the dilation order to obtain sufficient conditions that cover scenarios of intermediate product demands in the original system and in the reversed system with both equal and unequal means. Several analytical and numerical examples illustrate the key determinants of cost improvement under operations reversal. These include aggregate variance at the choice level of each feature, the number of choices for each feature, and demand correlations. As an extension, we also examine the effects of varying cost parameters under different production sequences and the effect of general lead times for product features. Again, we adopt the cost measure and provide conditions on when operations reversal is optimal.

A Roadmap for Implementation

In practice, the companies usually keep historical records of the demands for end products. In the sweater context, the accurate sales data from past seasons is available with the aid of point of sale (POS) information. By leveraging on these data, there are in general two ways to verify the conditions we have obtained in this paper. The first way is to obtain the demand distributions of both end and intermediate products through parametric distribution fitting (see for example, Sheskin (2003)), and then to work on the known distributions. The second way is to directly verify the relevant stochastic orders using the empirical data (see Belzunce et al. (2005) and Mosler and Scarsini (2012)). We refer to the aforementioned literatures for details of data manipulations.

Acknowledgments

The authors would like to thank Hau Lee and Paul Zipkin for their helpful comments on an earlier draft of this paper. They are also grateful to the editor and the reviewers for their constructive comments. This research was supported in part by the Hong Kong Research Grants Council under Grant 640808 and the Chinese Natural Science Foundation under Grant No. 71390331.

References

Bard, Jonathan F, Thomas A Feo. 1989. Note-operations sequencing in discrete parts manufacturing. *Management science* **35**(2) 249–255.

Belzunce, F, Franco Pellerey, JM Ruiz, Moshe Shaked. 1997. The dilation order, the dispersion order, and orderings of residual lives. *Statistics & probability letters* **33**(3) 263–275.

Belzunce, F, JF Pinar, JM Ruiz. 2005. On testing the dilation order and hnbue alternatives. *Annals of the Institute of Statistical Mathematics* **57**(4) 803–815.

Clark, Andrew J., Herbert Scarf. 1960. Optimal Policies for a Multi-Echelon Inventory Problem. *Management Science* **6**(4) 475–490.

Corbett, Charles J, Kumar Rajaram. 2006. A generalization of the inventory pooling effect to nonnormal dependent demand. *Manufacturing & Service Operations Management* **8**(4) 351–358.

Fagiuoli, Enrico, Franco Pellerey, Moshe Shaked. 1999. A characterization of the dilation order and its applications. *Statistical papers* **40**(4) 393–406.

Federgruen, Awi, Min Wang. 2013. Monotonicity properties of a class of stochastic inventory systems. *Annals of Operations Research* **208**(1) 155–186.

Gerchak, Yigal, Qi-Ming He. 2003. On the relation between the benefits of risk pooling and the variability of demand. *IIE transactions* **35**(11) 1027–1031.

Gerchak, Yigal, David Mossman. 1992. On the Effect of Demand Randomness on Inventories and Costs. *Operations Research* **40**(4) 804–807.

Gupta, Diwakar, William L. Cooper. 2005. Stochastic Comparisons in Production Yield Management. *Operations Research* **53**(2) 377–384.

Harvard Business School. 1986. Benetton(A) and (B). *Harvard Teaching Case* (9-685-014).

Hickey, Raymond J. 1986. Concepts of dispersion in distributions: a comparative note. *Journal of Applied Probability* 914–921.

Huang, Lu, Jing-Sheng Song, Jordan D Tong. 2015. Supply chain planning for random demand surges: Reactive capacity and safety stock. *Available at SSRN 2622510, to appear in Manufacturing & Service Operations Management* .

Iyer, Ananth V, Apurva Jain. 2003. The Logistics Impact of a Mixture of Order-streams in a Manufacturer-retailer System. *Management Science* **49**(7) 890–906.

Jain, Nikhil, Anand Paul. 2001. A generalized model of operations reversal for fashion goods. *Management Science* **47**(4) 595–600.

Jemaï, Zied, Fikri Karaesmen. 2005. The Influence of Demand Variability on the Performance of a Make-to-stock Queue. *European Journal of Operational Research* **164**(1) 195–205.

Kapuscinski, Roman, Sridhar Tayur. 1999. Variance vs. Standard Deviation: Variability Reduction Through Operations Reversal. *Management Science* **45**(5) 765–767.

Lee, Hau L., Kut C. So, Christopher S. Tang. 2000. The Value of Information Sharing in a Two-Level Supply Chain. *Management Science* **46**(5) 626–643.

Lee, Hau L., Christopher S. Tang. 1998. Variability Reduction Through Operations Reversal. *Management Science* **44**(2) 162–172.

Lu, Yingdong, Jing-Sheng Song, David D. Yao. 2003. Order Fill Rate, Lead Time Variability, and Advance Demand Information in an Assemble-to-order System. *Operations Research* **51**(2) 292–308.

Mosler, Karl, Marco Scarsini. 2012. *Stochastic orders and applications: a classified bibliography*, vol. 401. Springer Science & Business Media.

Müller, Alfred, Dietrich Stoyan. 2002. *Comparison Methods for Stochastic Models and Risks*. Wiley Series in Probability and Statistics, John Wiley & Sons, Ltd, Baffins Lane, Chichester, West Sussex, PO19 IUD, England.

Ridder, Ad, Erwin Van Der Laan, Marc Salomon. 1998. How Larger Demand Variability May Lead to Lower Costs in the Newsvendor Problem. *Operations Research* **46**(6) 934–936.

Schraner, Erika, Warren H Hausman. 1997. Optimal production operations sequencing. *IIE transactions* **29**(8) 651–660.

Shaked, Moshe, J George Shanthikumar. 1994. *Stochastic orders and their applications*. Academic Press.

Sheskin, David J. 2003. *Handbook of parametric and nonparametric statistical procedures*. CRC Press.

Shi, Jim Junmin, Xiaohang Yue, Yao Zhao. 2014. Operations sequencing for a multi-stage production inventory system. *Naval Research Logistics (NRL)* **61**(2) 144–154.

Song, Jing-Sheng. 1994a. The Effect of Leadtime Uncertainty in a Simple Stochastic Inventory Model. *Management Science* **40**(5) 603–613.

Song, Jing-Sheng. 1994b. Understanding the Lead Time Effects in Stochastic Inventory Systems with Discounted Costs. *Operations Research Letters* **15**(2) 85–93.

Song, Jing-Sheng, David D. Yao. 2002. Performance Analysis and Optimization in Assemble-to-order Systems with Random Lead Times. *Operations Research* **50**(5) 889–903.

Song, Jing-Sheng, Hanqin Zhang, Yumei Hou, Mingzheng Wang. 2010. The Effect of Leadtime and Demand Uncertainty in (r,q) Inventory Systems. *Operations Research* **58**(1) 68–80.

Van Houtum, GJ, Karl Inderfurth, Willem HM Zijm. 1996. Materials coordination in stochastic multi-echelon systems. *European Journal of Operational Research* **95**(1) 1–23.

Zhang, Sheng Hao, Ki Ling Cheung. 2008. Replenishment Sequencing in a Supply Chain with Balanced Ordering. *Manufacturing & Service Operations Management* **10**(2) 198–203.

Zipkin, Paul Herbert. 2000. *Foundations of inventory management*. McGraw-Hill New York.

Appendix A (Proofs of the Propositions and Corollaries)

Proof of Proposition 1. Firstly, notice that $\sum_{i=1}^{m} D_i = \sum_{i\in M, j\in N} D_{ij} = \sum_{j=1}^{n} \tilde{D}_j$, so $E\left(\sum_{i=1}^{m} D_i\right) = E\left(\sum_{j=1}^{n} \tilde{D}_j\right)$. Secondly, the optimal solution to the minimization problem $\min_y E\left(h(y-Z)^+ + p(Z-y)^+\right)$, where Z is a random variable with distribution function $F_Z(x)$ and inverse distribution function F_Z^{-1}, is $y^* = F_Z^{-1}(\frac{p}{h+p})$. Now we can derive

$$\begin{aligned}
E(R(s_i^*, D_i)) &= E(h(s_i^* - D_i)^+ + p(D_i - s_i^*)^+) \\
&= h\int_0^{s_i^*} (s_i^* - x) f_i(x)dx + p\int_{s_i^*}^{\infty} (x - s_i^*) f_i(x)dx \\
&= h\left(s_i^* F_i(s_i^*) - \int_0^{s_i^*} x f_i(x)dx\right) + p\left(\int_{s_i^*}^{\infty} x f_i(x)dx - s_i^*(1 - F_i(s_i^*))\right) \\
&= (h+p)s_i^* F_i(s_i^*) - p s_i^* - h\int_0^{s_i^*} x f_i(x)dx + p\int_{s_i^*}^{\infty} x f_i(x)dx \\
&= (h+p)s_i^* \frac{p}{h+p} - p s_i^* + pE(D_i) - (h+p)\int_0^{s_i^*} x f_i(x)dx \\
&= pE(D_i) - (h+p)\int_0^{s_i^*} x f_i(x)dx.
\end{aligned}$$

Similarly, we can derive $E(R(\tilde{s}_j^*, \tilde{D}_j)) = pE(\tilde{D}_j) - (h+p)\int_0^{\tilde{s}_j^*} x\tilde{f}_j(x)dx$. Therefore,

$$\begin{aligned}
&C_{AB}(\mathbf{s}^*; \mathbf{S}^*) - C_{BA}(\tilde{\mathbf{s}}^*; \mathbf{S}^*) \\
&= \sum_{i\in M} E(R(s_i^*, D_i)) - \sum_{j\in N} E(R(\tilde{s}_j^*, \tilde{D}_j)) \\
&= \sum_{i\in M}\left(pE(D_i) - (h+p)\int_0^{s_i^*} x f_i(x)dx\right) - \sum_{j\in N}\left(pE(\tilde{D}_j) - (h+p)\int_0^{\tilde{s}_j^*} x\tilde{f}_j(x)dx\right) \\
&= p\left(\sum_{i\in M} E(D_i) - \sum_{j\in N} E(\tilde{D}_j)\right) + (h+p)\left(\sum_{j\in N}\int_0^{\tilde{s}_j^*} x\tilde{f}_j(x)dx - \sum_{i\in M}\int_0^{s_i^*} x f_i(x)dx\right) \\
&= (h+p)\left(\sum_{j\in N}\int_0^{\tilde{s}_j^*} x\tilde{f}_j(x)dx - \sum_{i\in M}\int_0^{s_i^*} x f_i(x)dx\right).
\end{aligned}$$

Hence $\sum_{i=1}^{m}\int_0^{s_i^*} x f_i(x)dx \geq \sum_{j=1}^{n}\int_0^{\tilde{s}_j^*} x\tilde{f}_j(x)dx$ is equivalent with $C_{AB}(\mathbf{s}^*; \mathbf{S}^*) - C_{BA}(\tilde{\mathbf{s}}^*; \mathbf{S}^*) \leq 0$. We can thus conclude that if and only if $\sum_{i=1}^{m}\int_0^{s_i^*} x f_i(x)dx \geq \sum_{j=1}^{n}\int_0^{\tilde{s}_j^*} x\tilde{f}_j(x)dx$, it is optimal to process feature A first. $\square$

Proof of Corollary 1 Let Z denote a standard normal random variable with distribution function Φ, and $z^* = \Phi^{-1}(w)$. Then $s_i^* = F_i^{-1}(w) = E(D_i) + SD(D_i)z^*$, and $\tilde{s}_j^* = \tilde{F}_j^{-1}(w) = E(\tilde{D}_j) + SD(\tilde{D}_j)z^*$. By substituting into (12), we have

$$\sum_{i=1}^{m}\Big(E\left(E(D_i)+SD(D_i)z^*-D_i\right)^+ - w\left(E(D_i)+SD(D_i)z^*\right)\Big)$$
$$\leq \sum_{j=1}^{n}\left(E\left(E(\tilde{D}_j)+SD(\tilde{D}_j)z^*-\tilde{D}_j\right)^+ - w\left(E(\tilde{D}_j)+SD(\tilde{D}_j)z^*\right)\right),$$

which is equivalent to

$$\sum_{i=1}^{m}\left(SD(D_i)E\left(z^*-\frac{D_i-E(D_i)}{SD(D_i)}\right)^+\right) - \sum_{i=1}^{m}\left(w\left(E(D_i)+SD(D_i)z^*\right)\right)$$
$$\leq \sum_{j=1}^{n}\left(SD(\tilde{D}_j)E\left(z^*-\frac{\tilde{D}_j-E(\tilde{D}_j)}{SD(\tilde{D}_j)}\right)^+\right) - \sum_{j=1}^{n}\left(w\left(E(\tilde{D}_j)+SD(\tilde{D}_j)z^*\right)\right),$$

which can be further simplified as

$$\left(E(z^*-Z)^+ - wz^*\right)\left(\sum_{i=1}^{m}SD(D_i) - \sum_{j=1}^{n}SD(\tilde{D}_j)\right) \leq w\left(\sum_{i=1}^{m}E(D_i) - \sum_{j=1}^{n}E(\tilde{D}_j)\right),$$

The right hand side is equal to zero, and note that

$$\begin{aligned} E(z^*-Z)^+ - wz^* &= \int_{-\infty}^{z^*}(z^*-x)\phi(x)dx - z^*w \\ &= \int_{-\infty}^{z^*} z^*\phi(x)dx - \int_{-\infty}^{z^*} x\phi(x)dx - z^*w \\ &= z^*\Phi(z^*) - \int_{-\infty}^{z^*} x\phi(x)dx - z^*w \\ &= -\int_{-\infty}^{z^*} x\phi(x)dx \geq 0 \end{aligned}$$

Therefore the necessary and sufficient condition now becomes

$$\sum_{i=1}^{m}SD(D_i) - \sum_{j=1}^{n}SD(\tilde{D}_j) \leq 0. \quad \square$$

Proof of Proposition 2 Since $D_i \leq_{cx} \tilde{D}_i$ for $i \in M$, by the definition of convex order, $E\left(R(x,D_i)\right) \leq E\left(R(x,\tilde{D}_i)\right)$ for any $i \in M$ and any value of x. Hence, $\sum_{i\in M} E\left(R\left(\tilde{s}_i, D_i\right)\right) \leq \sum_{i\in M} E\left(R\left(\tilde{s}_i, \tilde{D}_i\right)\right)$. For any set of $\{\tilde{\mathbf{s}}, \tilde{\mathbf{S}}\}$,

$$C_{AB}(\tilde{\mathbf{s}};\tilde{\mathbf{S}}) - C_{BA}(\tilde{\mathbf{s}};\tilde{\mathbf{S}}) = \sum_{i\in M} E\left(R\left(\tilde{s}_i, D_i\right)\right) - \sum_{i\in M} E\left(R\left(\tilde{s}_i, \tilde{D}_i\right)\right) \leq 0.$$

That is, $C_{AB}(\tilde{\mathbf{s}};\tilde{\mathbf{S}}) \leq C_{BA}(\tilde{\mathbf{s}};\tilde{\mathbf{S}})$ for any set of $\{\tilde{\mathbf{s}},\tilde{\mathbf{S}}\}$. Therefore, $C_{AB}(\mathbf{s}^*;\mathbf{S}^*) \leq C_{AB}(\tilde{\mathbf{s}}^*;\mathbf{S}^*) \leq C_{BA}(\tilde{\mathbf{s}}^*;\mathbf{S}^*)$. It is optimal to process feature A first. $\square$

Proof of Proposition 3 Since $D_i \leq_{dil} \tilde{D}_i$ for $i \in M$, by the definition of dilation order, $D_i - E(D_i) \leq_{cx} \tilde{D}_i - E(\tilde{D}_i)$, which further implies $D_i \leq_{cx} \tilde{D}_i + E(D_i) - E(\tilde{D}_i)$ (the composition of a convex function and an increasing convex function is still convex). Since $R(x,y)$ is convex in y for given x, by the definition of convex order, we have

$$E\left(R\left(x,D_i\right)\right) \leq E\left(R\left(x,\tilde{D}_i + E(D_i) - E(\tilde{D}_i)\right)\right)$$

for any $i \in M$ and any value of x.

Note that $R(x,y) = R(x+z, y+z)$ for any z, hence we have

$$R\left(x,\tilde{D}_i\right) = R\left(x + E(D_i) - E(\tilde{D}_i), \tilde{D}_i + E(D_i) - E(\tilde{D}_i)\right).$$

Therefore,

$$\begin{aligned}
& C_{AB}(\mathbf{s}^*;\mathbf{S}^*) - C_{BA}(\tilde{\mathbf{s}}^*;\tilde{\mathbf{S}}^*) \\
&= \sum_{i\in M} E\left(R\left(s_i^*,D_i\right)\right) - \sum_{j\in M} E\left(R\left(\tilde{s}_j^*,\tilde{D}_j\right)\right) \\
&\leq \sum_{i\in M} E\left(R\left(\tilde{s}_i^* + E(D_i) - E(\tilde{D}_i), D_i\right)\right) - \sum_{i\in M} E\left(R\left(\tilde{s}_i^*,\tilde{D}_i\right)\right) \\
&\leq \sum_{i\in M} E\left(R\left(\tilde{s}_i^* + E(D_i) - E(\tilde{D}_i), \tilde{D}_i + E(D_i) - E(\tilde{D}_i)\right)\right) - \sum_{i\in M} E\left(R\left(\tilde{s}_i^*,\tilde{D}_i\right)\right) \\
&= \sum_{i\in M} E\left(R\left(\tilde{s}_i^* + E(D_i) - E(\tilde{D}_i), \tilde{D}_i + E(D_i) - E(\tilde{D}_i)\right)\right) \\
&\quad - \sum_{i\in M} E\left(R\left(\tilde{s}_i^* + E(D_i) - E(\tilde{D}_i), \tilde{D}_i + E(D_i) - E(\tilde{D}_i)\right)\right) = 0,
\end{aligned}$$

where the first inequality is due to the optimality of s_i^*.

It is optimal to process feature A first. $\square$

Proof of Proposition 4 Since $D_i \leq_{cx} \tilde{D}_i, \quad i=1,...,m-1,$ and $D_m \leq_{cx} \tilde{D}_m + \tilde{D}_{m+1}$, then for any given set of $\{\tilde{\mathbf{s}}, \tilde{\mathbf{S}}\}$,

$$\begin{aligned}
&\sum_{i=1}^{m-1} E\left(R\left(\tilde{s}_i, D_i\right)\right) + E\left(R\left(\tilde{s}_m + \tilde{s}_{m+1}, D_m\right)\right) \\
&\leq \sum_{i=1}^{m-1} E\left(R\left(\tilde{s}_i, \tilde{D}_i\right)\right) + E\left(R\left(\tilde{s}_m + \tilde{s}_{m+1}, \tilde{D}_m + \tilde{D}_{m+1}\right)\right) \\
&\leq \sum_{i=1}^{m-1} E\left(R\left(\tilde{s}_i, \tilde{D}_i\right)\right) + E\left(R\left(\tilde{s}_m, \tilde{D}_m\right)\right) + E\left(R\left(\tilde{s}_{m+1}, \tilde{D}_{m+1}\right)\right) \\
&= \sum_{j=1}^{m+1} E\left(R\left(\tilde{s}_j, \tilde{D}_j\right)\right),
\end{aligned}$$

where the second inequality above is due to the pooling effect, which can be proved easily as follows.

To show that $E(R(s_1+s_2, D_1+D_2)) \leq E(R(s_1, D_1)) + E(R(s_2, D_2))$ for any s_1, s_2 and random variables D_1, D_2. By definition,

$$\begin{aligned}
E(R(s_1+s_2, D_1+D_2)) &= E\left(h(s_1+s_2-D_1-D_2)^+ + p(D_1+D_2-s_1-s_2)^+\right) \\
&= E\left(h(s_1-D_1+s_2-D_2)^+ + p(D_1-s_1+D_2-s_2)^+\right) \\
&= E\left(2h\left(0.5(s_1-D_1)+0.5(s_2-D_2)\right)^+ + 2p(0.5(D_1-s_1)+0.5(D_2-s_2))^+\right) \\
&\leq E\left(h(s_1-D_1)^+ + h(s_2-D_2)^+ + p(D_1-s_1)^+ + p(D_2-s_2)^+\right) \\
&= E(R(s_1, D_1)) + E(R(s_2, D_2)),
\end{aligned}$$

where the inequality is due to convex property of the function $f(x) = x^+$.

Hence, for any given set of $\{\tilde{\mathbf{s}}, \tilde{\mathbf{S}}\}$,

$$\begin{aligned}
&C_{AB}(\tilde{s}_1, ..., \tilde{s}_{m-1}, \tilde{s}_m + \tilde{s}_{m+1}; \tilde{\mathbf{S}}) - C_{BA}(\tilde{\mathbf{s}}; \tilde{\mathbf{S}}) = \\
&\sum_{i=1}^{m-1} E\left(R\left(\tilde{s}_i, D_i\right)\right) + E\left(R\left(\tilde{s}_m + \tilde{s}_{m+1}, D_m\right)\right) - \sum_{j=1}^{m+1} E\left(R\left(\tilde{s}_j, \tilde{D}_j\right)\right) \leq 0.
\end{aligned}$$

That is, $C_{AB}(\tilde{s}_1, ..., \tilde{s}_{m-1}, \tilde{s}_m + \tilde{s}_{m+1}; \tilde{\mathbf{S}})$ is less than or equal to $C_{BA}(\tilde{\mathbf{s}}; \tilde{\mathbf{S}})$ for any given set of $\{\tilde{\mathbf{s}}, \tilde{\mathbf{S}}\}$. Note that $\mathbf{S}^* = \tilde{\mathbf{S}}^*$, therefore $C_{AB}(\mathbf{s}^*; \mathbf{S}^*) \leq C_{AB}(\tilde{s}_1^*, ..., \tilde{s}_{m-1}^*, \tilde{s}_m^* + \tilde{s}_{m+1}^*; \tilde{\mathbf{S}}^*) \leq C_{BA}(\tilde{\mathbf{s}}^*; \tilde{\mathbf{S}}^*)$. It is optimal to process feature A first. $\square$

Proof of Proposition 5 Since $D_i \le_{cx} \sum_{j\in K_i} \tilde{D}_j$, by the definition of $R(x,y)$, for any given set of $\{\tilde{\mathbf{s}}, \tilde{\mathbf{S}}\}$,

$$\begin{aligned}\sum_{i=1}^{m} E\left(R\left(\sum_{j\in K_i} \tilde{s}_j, D_i\right)\right) &\le \sum_{i=1}^{m} E\left(R\left(\sum_{j\in K_i} \tilde{s}_j, \sum_{j\in K_i} \tilde{D}_j\right)\right)\\ &\le \sum_{i=1}^{m}\sum_{j\in K_i} E\left(R\left(\tilde{s}_j, \tilde{D}_j\right)\right)\\ &\le \sum_{j=1}^{n} E\left(R\left(\tilde{s}_j, \tilde{D}_j\right)\right),\end{aligned}$$

where the second inequality is due to the pooling effect which we have shown in the proof of Proposition 4.

Now we can derive, for any given set of $\{\tilde{\mathbf{s}}, \tilde{\mathbf{S}}\}$,

$$\begin{aligned}&C_{AB}\left(\sum_{j\in K_i} \tilde{s}_j, i\in M; \tilde{\mathbf{S}}\right) - C_{BA}\left(\tilde{\mathbf{s}}; \tilde{\mathbf{S}}\right)\\ =&\sum_{i=1}^{m} E\left(R\left(\sum_{j\in K_i} \tilde{s}_j, D_i\right)\right) - \sum_{j=1}^{n} E\left(R\left(\tilde{s}_j, \tilde{D}_j\right)\right) \le 0.\end{aligned}$$

Hence, $C_{AB}\left(\sum_{j\in K_i} \tilde{s}_j, i\in M; \tilde{\mathbf{S}}\right) \le C_{BA}\left(\tilde{\mathbf{s}}, \tilde{\mathbf{S}}\right)$ for any given set of $\{\tilde{\mathbf{s}}, \tilde{\mathbf{S}}\}$. Therefore,

$$C_{AB}\left(\mathbf{s}^*; \mathbf{S}^*\right) \le C_{AB}\left(\sum_{j\in K_i} \tilde{s}_j^*, i\in M; \tilde{\mathbf{S}}^*\right) \le C_{BA}\left(\tilde{\mathbf{s}}^*; \tilde{\mathbf{S}}^*\right).$$

It is optimal to process feature A first. □

Proof of Proposition 6 It is easy to check that $\Delta C^* \le 0$ is equivalent to

$$\begin{aligned}&\gamma_{AB} p_{AB}^1 E\left(F_{D_0}^{-1}\left(\frac{1}{1+\gamma_{AB}}\right) - D_0\right)^+ + p_{AB}^1 E\left(D_0 - F_{D_0}^{-1}\left(\frac{1}{1+\gamma_{AB}}\right)\right)^+\\ \le\ &\gamma_{BA} p_{BA}^1 E\left(F_{D_0}^{-1}\left(\frac{1}{1+\gamma_{BA}}\right) - D_0\right)^+ + p_{BA}^1 E\left(D_0 - F_{D_0}^{-1}\left(\frac{1}{1+\gamma_{BA}}\right)\right)^+.\end{aligned}$$

□

Proof of Corollary 2 The if and only if condition (29) given by Proposition 6 becomes

$$\begin{aligned}&\eta\left(\gamma_{AB} E\left(\mu_0 + \sigma_0 z_{AB}^* - D_0\right)^+ + E\left(D_0 - \mu_0 - \sigma_0 z_{AB}^*\right)^+\right)\\ \le\ &\gamma_{BA} E\left(\mu_0 + \sigma_0 z_{BA}^* - D_0\right)^+ + E\left(D_0 - \mu_0 - \sigma_0 z_{BA}^*\right)^+,\end{aligned}$$

which can be shown to be equivalent to

$$\eta\left(\gamma_{AB}E\left((z^*_{AB}-Z)^+\right)+E\left((Z-z^*_{AB})^+\right)\right)\le\gamma_{BA}E\left((z^*_{BA}-Z)^+\right)+E\left((Z-z^*_{BA})^+\right).$$

□

Proof of Proposition 7 By definition, $E(R^1_{AB}(s,D_0))\le E(R^1_{BA}(s,D_0))$ for any s. Therefore we have

$$\begin{aligned}\Delta C^*=C^*_{AB}-C^*_{BA}&=\sum_{i\in M}E(R^1_{AB}(s^*_i,D_0))-\sum_{j\in N}E(R^1_{BA}(\tilde{s}^*_j,D_0))\\&=\sum_{i\in M}E(R^1_{AB}(s^*_i,D_0))-\sum_{i\in M}E(R^1_{AB}(\tilde{s}^*_i,D_0))\\&\quad+\sum_{j\in M}E(R^1_{AB}(\tilde{s}^*_j,D_0))-\sum_{j\in N}E(R^1_{BA}(\tilde{s}^*_j,D_0))\le 0.\end{aligned}$$

So it is optimal to process feature A first. □

Proof of Proposition 8 When $m=n$, by (31), $\Delta C^*\le 0$ is equivalent to

$$\begin{aligned}&mE(D)(L_B-L_A)(h^1-h^0)+E\left(R^1\left(F^{-1}_{D(mL_A)}(w^1),D(mL_A)\right)\right)+mE\left(R^2\left(F^{-1}_{D(L_B)}(w^2),D(L_B)\right)\right)\\&-E\left(R^1\left(F^{-1}_{D(mL_B)}(w^1),D(mL_B)\right)\right)-mE\left(R^2\left(F^{-1}_{D(L_A)}(w^2),D(L_A)\right)\right)\le 0.\end{aligned}$$

With elementary transformations and noting that $L_A>L_B$, the desired result can be obtained immediately. □

Proof of Corollary 3 Since D_{ij} are independent and identically distributed and D is normally distributed with parameter (μ,σ), $D(mL)$ is normally distributed with parameter $(mL\mu,\sqrt{mL}\sigma)$ for any positive integer m. Then we have $F^{-1}_{D(mL)}(w^1)=mL\mu+\sqrt{mL}\sigma z^1$ and $F^{-1}_{D(L)}(w^2)=L\mu+\sqrt{L}\sigma z^2$. Therefore it is easy to verify that

$$\begin{aligned}\Delta R(L)&=R^1\left(F^{-1}_{D(mL)}(w^1),D(mL)\right)-mR^2\left(F^{-1}_{D(L)}(w^2),D(L)\right)\\&=\sqrt{mL}\sigma\left(h^1(z^1-Z)^++p^1(Z-z^1)^+\right)-m\sqrt{L}\sigma\left(h^2(z^2-Z)^++p^2(Z-z^2)^+\right)\\&=\sqrt{mL}\sigma\left(h^1(z^1-Z)^++p^1(Z-z^1)^+-\sqrt{m}\left(h^2(z^2-Z)^++p^2(Z-z^2)^+\right)\right),\end{aligned}$$

which implies,

$$\begin{aligned}&\Delta R(L_A)-\Delta R(L_B)\\&=(\sqrt{L_A}-\sqrt{L_B})\sqrt{m}\sigma\left(h^1(z^1-Z)^++p^1(Z-z^1)^+-\sqrt{m}\left(h^2(z^2-Z)^++p^2(Z-z^2)^+\right)\right).\end{aligned}$$

Therefore (32) is equivalent to

$$\Delta h \Delta L \geq \frac{\sigma}{\sqrt{m}\mu}(\sqrt{L_A} - \sqrt{L_B}) E\left(h^1(z^1 - Z)^+ + p^1(Z - z^1)^+ - \sqrt{m}\left(h^2(z^2 - Z)^+ + p^2(Z - z^2)^+\right)\right),$$

or

$$\frac{\Delta h(\sqrt{L_A} + \sqrt{L_B})\sqrt{m}\mu}{\sigma} \geq E\left(h^1(z^1 - Z)^+ + p^1(Z - z^1)^+ - \sqrt{m}\left(h^2(z^2 - Z)^+ + p^2(Z - z^2)^+\right)\right).$$

□

Proof of Proposition 9 By equation (31), we have

$$\Delta C^* = mE\left(R^1\left(F^{-1}_{D(nL)}(w^1), D(nL)\right)\right) - nE\left(R^1\left(F^{-1}_{D(mL)}(w^1), D(mL)\right)\right). \tag{34}$$

To show $\Delta C^* \leq 0$, which is trivial for $k = 1$. We show the inequality for $k \geq 2$ by induction on k.

When $k = 2$,

$$\begin{aligned}
\Delta C^* &= mE\left(R^1\left(F^{-1}_{D(2mL)}(w^1), D(2mL)\right)\right) - 2mE\left(R^1\left(F^{-1}_{D(mL)}(w^1), D(mL)\right)\right) \\
&\leq mE\left(R^1\left(2F^{-1}_{D(mL)}(w^1), D(2mL)\right)\right) - 2mE\left(R^1\left(F^{-1}_{D(mL)}(w^1), D(mL)\right)\right) \\
&\leq 2mE\left(R^1\left(F^{-1}_{D(mL)}(w^1), D(mL)\right)\right) - 2mE\left(R^1\left(F^{-1}_{D(mL)}(w^1), D(mL)\right)\right) = 0.
\end{aligned}$$

Suppose it holds for $k \leq N$, to show it also holds for $k = N + 1$.

$$\begin{aligned}
\Delta C^* &= mE\left(R^1\left(F^{-1}_{D((N+1)mL)}(w^1), D((N+1)mL)\right)\right) - (N+1)mE\left(R^1\left(\tilde{s}^*(mL), D(mL)\right)\right) \\
&\leq mE\left(R^1\left(F^{-1}_{D(mL)}(w^1) + F^{-1}_{D(NmL)}(w^1), D((N+1)mL)\right)\right) \\
&\quad -(N+1)mE\left(R^1\left(F^{-1}_{D(mL)}(w^1), D(mL)\right)\right) \\
&\leq mE\left(R^1\left(F^{-1}_{D(mL)}(w^1), D(mL)\right)\right) + mE\left(R^1\left(F^{-1}_{D(NmL)}(w^1), D(NmL)\right)\right) \\
&\quad -(N+1)mE\left(R^1\left(F^{-1}_{D(mL)}(w^1), D(mL)\right)\right) \\
&\leq mE\left(R^1\left(F^{-1}_{D(mL)}(w^1), D(mL)\right)\right) + NmE\left(R^1\left(F^{-1}_{D(mL)}(w^1), D(mL)\right)\right) \\
&\quad -(N+1)mE\left(R^1\left(F^{-1}_{D(mL)}(w^1), D(mL)\right)\right) = 0,
\end{aligned}$$

where the last inequality is due to the induction assumption. Therefore it is optimal to process feature A first. □