Where To Differentiate Your Product When Stocking Levels Are Coupled

Erik Lystad
The College of Management
Georgia Institute of Technology
800 West Peachtree Street
Atlanta, GA 30332-0520
Tel: (404) 206-2020
Fax: (404) 894-6030
Email: Erik.Lystad@prism.gatech.edu

Mark Ferguson
The College of Management
Georgia Institute of Technology
800 West Peachtree Street
Atlanta, GA 30332-0520
Tel: (404) 894-4330
Fax: (404) 894-6030
Email: Mark.Ferguson@mgt.gatech.edu

February 12, 2006

Where To Differentiate Your Product
When Stocking Levels Are Coupled

A critical assumption of Lee and Tang's (1997) analysis of where in the production process a company should delay differentiation of its product is the independent treatment of installations in the production network. We show this "decoupling" approach gives rise to inaccuracies in assessing the value of delayed differentiation, frequently overestimating but also potentially underestimating the savings in inventory costs by failing to appropriately exploit the risk pooling effect. By doing so, we reveal a previously hidden factor in determining the optimal delayed differentiation strategy: the pattern of holding costs assessed for the various stages of work-in-process, which we refer to as the holding cost profile, plays an important role in the determination of the optimal strategy. Prior work has established the importance of the absolute holding cost at each stage in this decision but the relative holding costs are also important; as sharp increases in the local holding costs indicate potential cost reduction opportunities. Finally, we provide insight on the conditions when the decoupling assumption may lead to significant errors and cause a firm to make a costly mistake when determining where in the process to differentiate its product.

1. Introduction

In many industries and product lines, the growing diversity of customer demands is causing firms to dramatically increase product variety. For instance, in 2003 alone, 26,893 new food and household products were introduced, including 115 deodorants, 187 breakfast cereals, and 303 women's fragrances (Bianco, 2004). Such an increase poses significant challenges to firms as product proliferation is typically accompanied by increasingly inaccurate forecasts, higher inventories, and more frequent stockouts (Lee, 1996). The delay of differentiation between multiple product lines, through use of common components or modularity, has been examined as one solution to this problem. Delayed differentiation, first introduced by Alderson (1950), refers to the redesign of production processes to delay the stage where a universal set of product components is modified to their final distinct configurations. A delay in differentiation typically allows for greater service levels at decreased inventory costs, as firms exploit better information and risk-pooling effects. Thus, benefits of delayed differentiation tend to increase the further down the supply chain that differentiation takes place (Lee, 1996).

Unfortunately, implementing delayed differentiation is not free. There are often significant costs involved in redesigning (and in many cases, over-designing) the product (Fisher et. al., 1999). In addition, configuring the product further down the supply chain (for instance, at the warehouse or retailer stage) is rarely as cost efficient as at the primary manufacturing facility. Both of these costs (design and assembly) tend to increase the further down the supply chain that differentiation takes place. Since both the benefits and the costs of delayed differentiation increase the further down the supply chain, the question of where in the supply chain to optimally differentiate arises. In this paper, we explicitly model this tradeoff to determine the optimal point of differentiation. While this problem has been previously modeled in the literature (Lee and

Tang, 1997), we relax one of their major assumptions and find some instances where our solutions are significantly different.

As in Lee and Tang (1997), we assume the production process may be modeled by a series of discrete processing stages or installations. We refer to the last stage that the generic product exists as the Last Common Operation (LCO), and seek to determine which stage should be selected as the LCO to minimize total cost. For example, a comparison of the total cost of the two networks shown in Figure 1 will assist a firm in deciding if it is better to differentiate its product either one or two stages from the last stage.

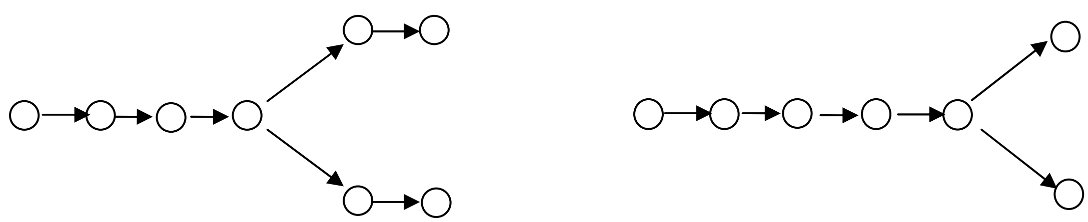


Figure 1: Delayed Differentiation Product Networks

When Lee and Tang (1997) choose the amount of product to stock at each stage, they assume the stages may be treated independently; i.e. there is no connection between the service level of one stage with the stocking level of the stage preceding it. This assumption is referred to as the decoupling heuristic (DH) because it allows each stage to be analyzed in isolation. While the DH is appropriate when every stage is required to maintain a high service level, the literature on multi-echelon inventory systems shows it is rarely optimal to maintain high service levels in the stages of the system that do not directly serve the final customer (see for example Chapter 8 in Zipkin, 2000). We relax the DH and provide guidance on the conditions where its use can lead to far from optimal decisions.

Because the analysis of arboreal multi-echelon supply chains without the DH becomes prohibitively complex, we approximate the echelon stocking levels for each installation through an extension of the Newsvendor Bounds Heuristic (NBH) of Lystad and Ferguson (2005) to networks of more than two echelons. The heuristic is tested using these stocking levels as parameters in a series of supply chain simulation experiments and compare the resulting supply chain costs to the costs obtained using the best stocking levels found from a full enumeration search. After verifying the heuristic performs well under common conditions, we compare the results against the results generated using the DH.

We show that the NBH performs much better than the DH and the optimal point of differentiation shifts towards the end of the supply chain as backorder costs increase and towards the beginning of the supply chain as echelon holding costs increase. These results are in agreement with the results of Lee and Tang (1997), however, we also find the DH may *over or underestimate* the value of delaying differentiation. Specifically, the DH often overestimates the value of delaying differentiation, except for cases when the echelon holding costs are high in the initial stages of production and the backordering costs are also high. In these cases, the DH results in the firm failing to carry sufficient inventory in the intermediate stages, and correspondingly underestimates the risk pooling benefits of delaying differentiation. Both the over and underestimation are often significant and may lead a firm to select suboptimal supply chain structures.

Finally, by not assuming the stages are decoupled, we discover a non-intuitive and previously hidden insight that the *shape* of the holding cost profile (how much the holding cost increases from one stage to the next) significantly affects the choice of where in the process product differentiation occurs. We find the presence of sharp rises in local holding cost between stages is associated with increased cost savings due to the reduction in effective backordering costs at the downstream stages. Thus, capturing holding cost 'spikes' under a common component form is more valuable than previously believed, and may serve as justification for a more extensive use of delayed differentiation strategies.

The remainder of this paper is organized as follows. We provide a summary of the relevant literature in §2. The model is defined in §3 and a simulation study is presented and discussed in §4. We state and explain our experimental results in §5 and conclude in §6. Appendix 1 contains additional numerical results and Appendix 2 provides proofs to our propositions.

2. Literature Review

The realities of increasing product variety has led to many proposals to address the corresponding complexities, such as part commonalities, process sequencing, delayed differentiation, and lead-time reduction (see Chapters 15, 16, and 18 in Tayur et al.(1999) for a representative sample). Delayed differentiation exploits the variance reduction through the risk pooling effect, reducing the required safety stock to meet a given service level: Lee et al.(1993), Lee (1996), and Lee and Tang (1997). Garg and Tang (1997) add a second level of differentiation to the firm's decision resulting in even more benefits of late differentiation.

Historically, the analysis of delayed differentiation of multiple product lines typically assumes installations along the supply chain may be treated independently from one another (e.g. Lee and Tang 1997; Ma, Wang, and Wu 2002), resulting in single stage inventory policies. Aviv and Federgruen (2001) avoid the use of this decoupling assumption when considering capacitated, two-echelon production systems under seasonally fluctuating demand distributions. Instead, they utilize a balance relaxation, where negative inventory allocations may be applied to the differentiated installations. In this paper, we avoid the use of decoupling and balance relaxations.

Decoupling assumptions are also made in the literature on component commonalities (e.g. Hsu and Wang 2003).   The use of common components is a special case of delayed differentiation, postponing the differentiation of an assembly system until the first non-common component is included. By incorporating common components, firms may increase their flexibility to meet diverse customer demands at a higher service level but often face the tradeoff of increased redesign and component costs. Baker et al. (1986), Gerchak, Magazine, and Gamble (1988) and Eynan and Rosenblatt (1996) show under general demand distributions and correlations, the optimal levels of common component inventories are non-intuitive. Furthermore, Eynan and Rosenblatt (1996) show it may be optimal to not utilize commonality when the costs of a common assembly are substantially more expensive than the specific assemblies it replaces. Although these works suggest the optimal inventory policy of stages of a product line poses an interesting, non-intuitive problem, they also limit their analysis to 2-echelon systems and focus on service levels rather than minimizing expected costs. In this paper, we use a heuristic to determine inventory base-stock levels for an *N*-echelon network, where total holding and penalty costs are minimized.

Our heuristic is an extension of the NBH provided by Lystad and Ferguson (2005) who suggest the expected periodic cost of a two-echelon distribution system may be bounded from above and below by two constructed serial systems. The heuristic calculates the inventory base-stock levels at each stage through the use of single-stage newsvendor solutions. To estimate the inventory cost effects on the value of delaying differentiation, we extend this technique to analyze and compare three-echelon network topologies.

3. Model

Consider a multi-echelon supply chain with N stages, a single initial common product form and M final products as shown in Figure 2 (the following constructs assume that all differentiation occurs at the single node k, although our approach may be extended to any arborescent topology).

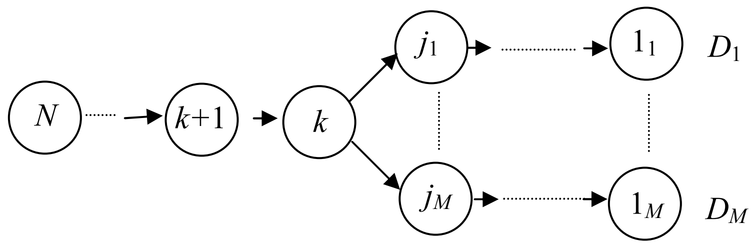


Figure 2: Model of Supply Chain Network

We assume a periodic review policy where the system updates as follows: Allow L_n to represent the known, deterministic lead-time of shipments from the $n+1^{th}$ to the n^{th} installation. At the beginning of period t, a shipment arrives to installation n that was shipped L_n time units ago from installation $n+1$. Demand occurs at the final stage of the echelon tree, nodes 1_1 through 1_M. Unmet demand is fully backordered. A centralized decision maker observes the current state of inventory in the entire chain and places orders for each stage. The upstream stage consequently ships downstream the minimum of the net order quantity and the on-hand inventory at stage $n+1$.

Our decision of interest is identical to Lee and Tang (1997)'s; i.e. at which stage should product differentiation occur. Our objective is to minimize the expected periodic total supply chain cost by properly selecting stage k (between N and 1) as the point of differentiation. Before presenting our model, we introduce the following terminology:

n = the stage index, $n \in (1,..N)$

i = the product (equivalently sub-chain) index, $i \in (1,..M)$

k = the stage of last common operations (LCO), our decision variable

L_n = the constant transportation lead-time from stage n+1 to stage n when $n \geq k$.

$L_{i,n}$ = the constant transportation lead-time for product i from stage n+1 to stage n when $n < k$.

S_k = the annuitized cost per period for an investment in the ability to perform a common operation at stage k

P_n = the common processing costs per unit at stage n when $n \geq k$.

$P_{i,\mathrm{n}}$ = the processing cost per unit of product i at stage n when $n < k$.

$F_{i,n}(x)$ = the $\sum_{j=1}^{n} L_{i,j}$ fold convolution of the distribution function of periodic demand for product i.

$D_i(t)$ = the cumulative demand for product i in the time interval (0,t]

$D_{i,n} = D_i(t+L_{i,n}) - D_i(t)$, the lead-time demand for product i in installation n

μ_i = the mean single period demand of product i

H_n = the common installation holding costs per unit per period at stage n when $n \geq k$.

$H_{i,n}$ = the installation holding cost per period and per unit of product i at stage n when $n < k$.

$h_n = H_n - H_{n+1}$, echelon inventory holding cost per unit, per period at stage n when $n \geq k$.

$h_{i,n} = H_{i,n} - H_{i,n+1}$, $H_{i,k} = H_k$, echelon inventory holding cost per unit, per period in chain i at stage n when $n < k$.

I_n^k = the echelon inventory base-stock level for installation n, given LCO = k

$I_{i,n}^k$ = the echelon inventory base-stock level of product i for installation n, given LCO = k

T_n = the common transportation cost per unit at stage n.

$T_{i,n}$ = the transportation cost per unit of product i at stage n.

R^k_n = the net echelon inventory at stage n, given the last common operation is stage k

$R_{i\ n}^{\ k}$ = the net echelon inventory of product i at stage n, given the last common operation is stage k

b_i = the per period unit cost of a backorder at retail installation i

$b = \dfrac{\sum_i \mu_i b_i}{\sum_i \mu_i}$ the demand weighted average unit backorder cost

The expected periodic cost is comprised of annuitized investment costs that allow a product to remain in a generic form until the LCO (stage k), processing costs, inventory holding and stockout costs, and transportation costs between the installations. Let the sum of these costs be $Z(k)$, so the firm's problem is to find the stage $\bar{k}$ where

$$\begin{aligned}
\bar{k} = \underset{1 \le k \le N-1}{\arg\min}\, Z(k) &= \sum_{n=k}^{N} S_n + \left[\sum_{n=k}^{N} \left(\sum_{i=1}^{M} \mu_i \right) * P_n + \sum_{n=1}^{k-1} \left(\sum_{i=1}^{M} P_{i,n} * \mu_i \right) \right] \\
&+ \mathrm{E}\left[\sum_{n=k+1}^{N} \left(H_n * \left[R_n^k - \sum_{i=1}^{M} D_i \right]^+ \right) + \sum_{n=1}^{k} \left(\sum_{i=1}^{M} H_{i,n} * \left[R_{i,n}^k - D_i \right]^+ \right) \right] \\
&+ \left[\sum_{n=k}^{N} \left(\sum_{i=1}^{M} \mu_i \right) * T_n + \sum_{n=1}^{k-1} \left(\sum_{i=1}^{M} T_{i,n} * \mu_i \right) \right] + \sum_{i=1}^{M} b_i \mathrm{E}\left[D_i - R_{i,1}^k \right]^+
\end{aligned} \tag{1}$$

The first term in (1) is the total investment costs, the terms in the first bracket are the total processing costs, the second bracket includes the total holding costs, the third bracket includes the transportation costs, and the fifth and sixth terms include the backorder costs. As in Lee and Tang (1997), the optimal LCO is determined by comparing the objective function $Z(k)$ for stages $k = 1$ through N-1. Unlike their model however, which uses a decoupling argument to determine an inventory stocking level (or safety stock factor) at each stage in the chain, we determine the echelon stocking level for each stage. An optimal policy for this type of system has not yet been solved (although Federgruen and Zipkin (1984) provide bounds and Zipkin (2000) provides approximations). Even these approximations are complex and difficult to compute, thus we solve the problem using the NBH from Lystad and Ferguson (2005). We provide a brief description of this heuristic below.

To analyze a given multiechelon topology, we construct two serial supply chain systems whose costs bound the optimal costs of a branched chain from above and below. Our illustrative network is depicted in Figure 1, and faces demand processes $D_1, D_2, \ldots, D_M$ at the terminal ends of the chain segments. Following the approach of Lystad and Ferguson (2005), we construct an upper bound by decomposing the system depicted in Figure 2 into the set of serial chains depicted in Figure 3.

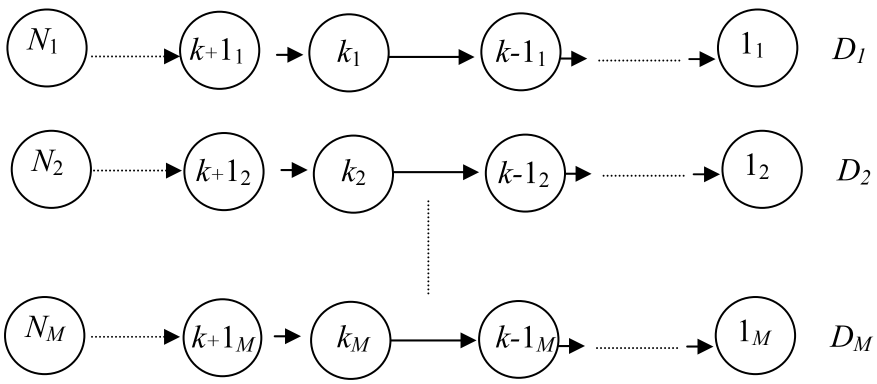
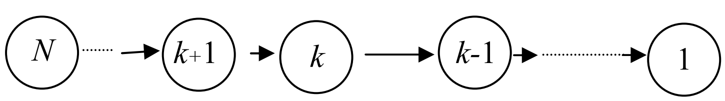


Figure 3: Decomposed Serial Chains

Likewise, we construct a lower bound by collapsing the system depicted in Figure 1 into a single serial chain as depicted in Figure 4.



D_1+D_2+…+D_M

Figure 4: Collapsed Supply Chain

To describe our heuristic, we need the following additional notation:

$s_{i,n}$ = a base-stock level for stage n in serial chain i

$s_{i,n}^{x*}$ = the "best found" base-stock level for stage n in serial chain i, where $x \in (d,c,e,a)$

$C_{i,n}^{x}\left(s_{i,n}\right)$ = the expected per period cost of the first n stages of serial chain i under base-stock policy $s_{i,n}$, where $x \in (d,c,e,a)$

u, l = superscripts denoting upper and lower bounds, respectively

d, c = superscripts denoting decomposed and collapsed systems, respectively

e = superscript denoting the arborescent topology

a = superscript denoting the heuristic policy for the arborescent topology

The decomposition and collapsed system results combine to give

$$C_N^c\left(s_N^{c*}\right) \le C_N^e\left(s_N^{e*}\right) \le \sum_i C_{i,N}^d\left(s_{i,N}^*\right) \tag{2}$$

and suggest that $s_n^{c*} \le s_n^{e*} \le \sum_{i=1}^{M} s_{i,n}^{d*}$ (for $n = k...N$). (3)

We use the stocking level results of the serial systems to approximate the optimal base-stock levels for the arborescent network. Our approach is similar to the Shang and Song (2003) heuristic for each of the M+1 constructed chains. Using an illustrative two-retailer system, for the collapsed serial chain system, the stocking level at stage n is

$$s_n^c = \frac{F_{1+2,n}^{-1}\left(\dfrac{b + \sum_{j=n+1}^{N} h_j}{b + \sum_{j=1}^{N} h_j}\right) + F_{1+2,n}^{-1}\left(\dfrac{b + \sum_{j=n+1}^{N} h_j}{b + \sum_{j=n}^{N} h_j}\right)}{2}. \tag{4}$$

For the decomposed serial chain system, the stocking levels at stage n are, for our illustrative system,

$$s_{1,n}^d = \frac{F_{1,n}^{-1}\left(\dfrac{b_1 + \sum_{j=n+1}^{N} h_{1,j}}{b_1 + \sum_{j=1}^{N} h_{1,j}}\right) + F_{1,n}^{-1}\left(\dfrac{b_1 + \sum_{j=n+1}^{N} h_{1,j}}{b_1 + \sum_{j=n}^{N} h_{1,j}}\right)}{2} \quad \text{and} \quad s_{2,n}^d = \frac{F_{2,n}^{-1}\left(\dfrac{b_2 + \sum_{j=n+1}^{N} h_{2,j}}{b_2 + \sum_{j=1}^{N} h_{2,j}}\right) + F_{2,n}^{-1}\left(\dfrac{b_2 + \sum_{j=n+1}^{N} h_{2,j}}{b_2 + \sum_{j=n}^{N} h_{2,j}}\right)}{2}. \tag{5,6}$$

The sum of the base-stock levels, $s_n^d = s_{1,n}^d + s_{2,n}^d$, represents an approximation for the echelon inventory of the arborescent chain. If the backorder costs or holding costs differ between product chains, Equation 4 is adjusted to use average backorder and holding costs weighted by the mean demands.

Our heuristic for setting echelon base stock levels at stage n ($k \le n \le N$) is to average the approximations resulting from the collapsed and decomposed chains,

$$s_n^a = \frac{s_n^c + s_n^d}{2}. \tag{7}$$

4. Simulation Experiments

The NBH is extensively tested for a two-echelon distribution network in Lystad and Ferguson (2005). To demonstrate its usefulness in a delayed differentiation analysis, however, we need to test it on systems of at least three echelons. Because closed form solutions are as of yet unavailable, we do so using simulation. We consider two candidate supply chains of three echelons with LCOs of 2 and 3, respectively, as shown below in Figures 5 and 6. Comparing these topologies captures the critical elements of a delayed differentiation process.

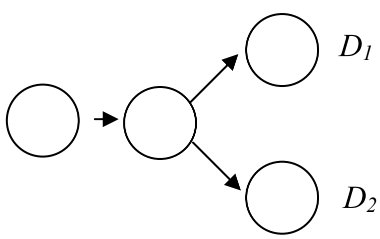


Figure 5: k=2

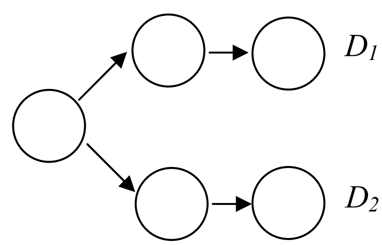


Figure 6: k=3

The simulations are conducted as follows; for a single, steady state replication, random demands are generated for 100,000 periods. In each period, demand is satisfied or backordered, orders are placed and filled, and linear holding and backordering costs are assessed. These costs are aggregated into 1000 batch means of 100 periods each. The first batch mean is removed to eliminate initialization effects. The removal of the first 100 periods of data is overly conservative because, under base-stock policies, the state spaces of the investigated supply chains are independent after $L_3 + L_2 + L_1 + 1$ periods. The average costs and standard deviations for the remaining 99 batch means (periods 1001 through 100,000) are reported. We utilize common random numbers across systems during demand generation, for computational simplicity and to potentially exploit variance reduction. Demand for both products is assumed Poisson with a mean rate of 10 units per period.

We compare the NBH results to those generated by the DH used in Lee and Tang's (1997) procedure and to the set of echelon base-stock levels that results in the lowest average system cost found via the simulation. The DH stocking levels are determined by calculating the optimal local fill rate at the retail stages and applying this rate to each upstream installation. The lowest average system costs are found by conducting a full enumeration across the expected minimizing local base-stock level ± ½μ. If lowest cost set of stock levels are potentially constrained by these limits, the study is widened so that the set contains no elements at the limits. We conduct simple difference of means tests between the NBH and DH to establish statistically significant results. We achieve significant results in 46 of 48 experiments, while the remaining 2 experiments fail to achieve a significant difference at the 5% level using a two-tailed t-test.

5. Experimental Results

5.1 Symmetric Costs

The first series of experiments assumes the echelon holding costs and backordering costs for each installation at each stage are identical. We further assume that the processing and transportation costs are constant regardless of k (e.g. $P_n = P_{1,n} = P_{2,n}$), and that $\sum_{n=1}^{k} S_k = 0$ for all k.

Under these assumptions, differences in the solution to Equation (1) arise solely from inventory related effects.

We begin establishing the accuracy of the two heuristics by varying the backorder costs across a 4,000% range while keeping the echelon holding costs constant. For these examples, the echelon holding costs of each of three levels of installations is set to 1 (i.e. the local installation holding costs are 1, 2, and 3, respectively). Simulations were conducted as described in §4 and the results are depicted in Appendix 1 in Tables A.1 and A.2.

Observation 1: *The NBH policies result in lower costs than the DH policies.*

For the non-delayed case, the NBH generates inventory policies that exceed the cost of the best found policy by an average of 0.2%, compared to 1.7% for the DH. For the delayed case, the errors of the NBH policies and DH policies are 0.7% and 2.0%, respectively.

Observation 2: *The NBH performs progressively better at higher backordering costs. The DH also performs better as the backordering costs increase, but reaches a point where it begins to perform worse.*

As the backordering costs increase, the consistently high service levels at each installation required by the DH are justified. Thus it is unsurprising that here, the DH policies perform well. However, these policies are still outperformed by the NBH. In the non-delayed case, the NBH finds the best policy in over half of the experiments with the highest backordering costs.

Observation 3: *The largest discrepancies between the NBH and the DH occur in the low backordering cost range.*

We focus on this area in the remainder of this paper, believing the greatest contribution may be achieved in this range. Intuitively, it is in this region that the optimal upstream fill rates are significantly smaller than at the retail stage. Thus, in the low backordering cost range, the DH is particularly inappropriate. We specifically consider backordering to holding cost ratios of approximately 7:1 at the retailer, noting that this ratio describes a wide variety of products. For instance, by taking the backordering cost as the lost revenue of a sale of a product with a 50% profit margin, this ratio applies as long as the holding costs exceed 7% of the value of the product.

5.2 Asymmetric Costs

The second series of experiments addresses asymmetric costs where the backordering cost and/or the holding costs at the terminal stage are allowed to differ. We continue to assume the processing and transportation costs are constant regardless of the topology given by the selection of *k*, and that $\sum_{n=1}^{k} S_k = 0$ for all *k*, again resulting in an investigation into solely the role of inventory effects on Equation (1).

We test two levels of holding cost at each echelon, $h_n = \{1,2\}$ and three backordering costs, b=$\{5,10,20\}$. The NBH and DH stocking levels are generated as before. The findings of these experiments are summarized in Appendix 1 in Tables A.3 and A.4.

Observation 4: *The NBH is robust to asymmetric costs in the production network beyond the last common operation.*

For the non-delayed and delayed production networks, the NBH produces an average error of 0.6% and 1.2%, respectively. In contrast, use of the DH leads to errors of 1.7% and 1.8%, respectively. We note that when both holding and backordering costs differ between the chains, the NBH performs worse than the DH in the delayed network. The allocation policy utilized in the NBH is suboptimal for asymmetric retailers but the errors induced by this allocation policy are small relative to the DH results. These results should also be viewed in the context of the range of diversity in holding costs (100%) and backordering costs (400%) between installations. Products originating from a common form are unlikely to experience this degree of cost parameter asymmetry. We leave an investigation for a slight correction to the heuristic for future work. For both the DH and NBH, the errors are greater in the asymmetric cost experiments than in the symmetric cost experiments.

5.3 The Value of Delaying Differentiation

Having established the NBH performs well in three-echelon topologies, we expand the experiment to compare the NBH results to the DH results as the echelon holding costs and backordering costs vary. These results are summarized in Table 1.

Composite Results							
Echelon Holding Costs			Backorder Cost	Non-delayed Chain Costs		Delayed Chain Costs	
h_3	h_2	h_1	b	NBH	DH	NBH	DH
1	1	1	2.5	8110	8603	7970	8385
1	1	1	5	8928	9226	8733	8940
1	1	1	7.5	9426	9637	9183	9302
1	1	1	10	9814	9980	9544	9603
1	1	1	20	10676	10819	10331	10380
1	1	2	5	9341	9460	9132	9175
1	1	2	10	10411	10484	10136	10088
1	1	2	20	11592	11644	11155*	11146*
1	2	1	5	11473	12382	11180	11859
1	2	1	10	12597	13232	12236	12578
1	2	1	20	13773	14176	13286	13458
1	2	2	5	11798	12208	11507	11741
1	2	2	10	13159	13420	12667	12861
1	2	2	20	14561	14706	14076*	14027*
2	1	1	5	13428	14473	13201	14156
2	1	1	10	14621	15324	14308	14947
2	1	1	20	15852	16268	15399	15827
2	1	1	50	17461	17509	16880	17090
2	1	2	5	13840	14300	13613	13973
2	1	2	10	15175	15550	14842	15158
2	1	2	20	16643	16835	16210	16324
2	2	1	5	15934	17326	15663	16595
2	2	1	10	17317	18331	16919	17926
2	2	1	20	18845	19439	18322	18904
2	2	2	5	16221	17206	15940	16770
2	2	2	10	17856	18452	17467	17880
2	2	2	20	19574	19959	19087	19206

* Denotes an insignificant difference of means on a two-tailed t-test at the 5% level

Table 1: Comparative Costs

Observation 5: *The experiments show that delaying differentiation is consistently valuable.*

This observation is expected since we assume the redesign cost to be zero.

Observation 6: *The NBH outperforms the results of the DH.*

In 45 of the 48 cases the NBH outperforms the DH at the 5% significance level. In 2 of the remaining 3 cases, the difference of means is statistically insignificant at the 5% level after 200,000 periods. In only one case did the DH significantly outperform the NBH. Thus, when determining stocking levels for a multi-level production system, our results indicate NBH stocking levels result in lower inventory costs for a given service level than DH stocking levels.

In designing the manufacturing process, a firm seeks to answer the question of whether the savings in inventory costs from delaying differentiation are worth the additional costs from processing, transportation, or redesign. To examine the effects of the DH on this decision, we revisit the data presented in Table A.3 and compare the differences in the expected inventory costs between the non-delayed and delayed chains. This data is presented in Table 2

.

Expected Value of Delayed Differentiation						
Echelon Holding Costs			Backorder Cost	Delay Value		
h_3	h_2	h_1	b	NBH	DH	DH Overestimation
1	1	1	2.5	140.65	218.3	35.6%
1	1	1	5	194.56	286.28	32.0%
1	1	1	7.5	242.65	334.8	27.5%
1	1	1	10	270.13	376.8	28.3%
1	1	1	20	344.6	438.1	21.3%
1	1	2	5	209.02	284.44	26.5%
1	1	2	10	274.8	395.6	30.5%
1	1	2	20	437.8	498.1	12.1%
1	2	1	5	293.7	523.1	43.9%
1	2	1	10	360.7	654.7	44.9%
1	2	1	20	487	718.6	32.2%
1	2	2	5	291.4	467.9	37.7%
1	2	2	10	491.1	559.1	12.2%

1	2	2	20	485.2	678.5	28.5%
2	1	1	5	226.6	317.5	28.6%
2	1	1	10	312.6	376.9	17.1%
2	1	1	20	453.2	440.9	-2.8%
2	1	1	50	581	418.8	-38.7%
2	1	2	5	227	326.9	30.6%
2	1	2	10	333.8	391.6	14.8%
2	1	2	20	433.5	510.9	15.1%
2	2	1	5	271	730.3	62.9%
2	2	1	10	398.5	405.2	1.7%
2	2	1	20	523.3	534.7	2.1%
2	2	2	5	281.3	436.6	35.6%
2	2	2	10	389.1	572.6	32.0%
2	2	2	20	486.5	753.6	35.4%

Table 2: Value of Delaying Differentiation

Observation 7: *In the majority of cases studied, the DH overestimates the potential cost savings of delaying differentiation.*

In 93% of the cases investigated, the cost savings generated by delaying differentiation under the DH exceeds that generated by delaying differentiation under the NBH, and this discrepancy ranged as high as 62.9%. The overestimation of value is greatest when a significant increase in holding cost occurs at the potential LCO stage. Here, the DH overestimates the cost savings by decreasing the inventory held at the LCO by amounts greater than the NBH. Because the DH relies on the decoupling assumption, it carries excessive inventory at upstream stages. This excessive inventory is reduced by greater amounts upon realization of the pooling effect than the more appropriately set inventory levels under the NBH. Thus, the DH often leads to an overestimation of the benefits of delayed differentiation.

Observation 8: *Use of the DH may also significantly underestimate the potential cost savings of delaying differentiation.*

The underestimation of value is greatest when the local holding costs are high at the beginning of the process and backordering costs are also high. This may occur when there is considerable value in the raw materials compared to the value added during the production process, or when dealing with materials that require expensive or dangerous handling such as molten metal. Here, the DH fails to capitalize fully on the benefits of holding product in intermediate stages. Because a majority of the holding costs are applied regardless of the position of the inventory, the effective cost of shifting base-stock levels downstream towards the intermediate installations decreases. This increases inventory at the potential LCO stage under the NBH, affording a greater savings upon merging the chains and exploiting the risk pooling effect. By ignoring the effects of the delayed differentiation decision on installations upstream of the LCO echelon, the DH also fails to appreciate potential cost savings and may underestimate the value of delaying differentiation.

5.4 Holding Cost Profile Insights

The use of the NBH allows the development of a non-intuitive and previously obscured result in the behavior of the value of delayed differentiation as a function of the local holding costs. To demonstrate, consider the data in Table 2 where $H_1 = 4$ and $b = 20$. Three local holding cost profiles meet this criteria; P1: $\{H_3 = 2, H_2 = 3, H_1 = 4\}$, P2: $\{H_3 = 1, H_2 = 3, H_1 = 4\}$, and P_3: $\{H_3 = 1, H_2 = 2, H_1 = 4\}$. These data are plotted below in Figure 7, where the plotted areas represent the local holding cost of a unit of product as it progresses through the production process, and the values in parenthesis are the associated values of delaying differentiation. Under the DH, the value of delaying differentiation under holding cost profiles P1 and P2 are equal, and larger than under P3, due to the inventory savings at the second echelon. Using the NBH, however, clearly shows the value under profile P2 > P1 > P3.

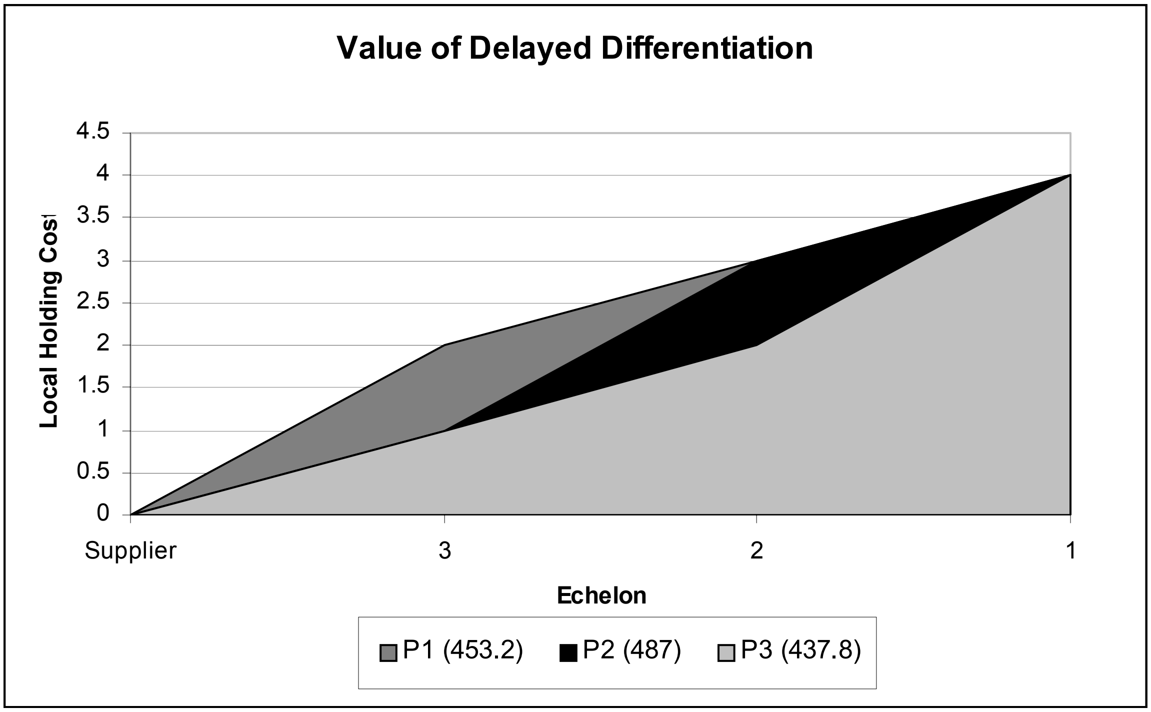


Figure 7: Values of Delaying Differentiation for Three Holding Cost Profiles

The example above indicates the value of delaying differentiation is related to the shape of the holding cost profile in addition to the absolute level of local holding cost. When comparing P1 and P2, we see that under P2, the holding costs increase at the potential point of delayed differentiation. We refer to this as a holding cost 'spike'; that is, a region in the production process where the slope of the holding cost profile is large. We find that it is more valuable than previously expected to capture this spike, a finding that is consistent with all of our numerical examples. This value may even exceed that obtained from the absolute holding cost savings. Note that in the above example, the increase in savings due to reduction of inventory when holding costs are 2 rather than 3 (e.g. P1 and P3) is 15.4 cost units. However, the additional benefit obtained from including a holding cost spike in delayed differentiation (e.g. P1 and P2) is an additional 33.8 cost units beyond that predicted solely from the absolute holding cost related savings. We formalize this finding in Propositions 1 and 2, whose proofs are presented in Appendix 2. The propositions require the following two assumptions:

Assumption 1: The investment, processing, and transportation costs are independent of the selection of k. That is, $\sum_{n=1}^{k} S_k = 0$, $P_n = P_{1,n} = P_{2,n}$, and $T_n = T_{1,n} = T_{2,n}$. Assumption 1 limits our consideration to the effects of inventory costs only. Even if the conditions of the assumption never occur in practice, the analysis allows us to isolate the role of these costs from those of the investment, processing, and transportation costs, which are additive in the objective function and may be treated independently from the inventory considerations.

Assumption 2: Consider two network topologies, each with the same number of echelons, where the first consists of serial chains and the second consists of a distribution center feeding the same chains, as depicted in Figures 7 and 8, respectively. We assume the addition of an upstream echelon that converts the networks into Figures 9 and 10 affects the costs of operating each system identically. That is, the difference in expected periodic costs between the topologies represented by Figures 7 and 8 is equal to the difference in expected periodic costs between the topologies represented by Figures 9 and 10.

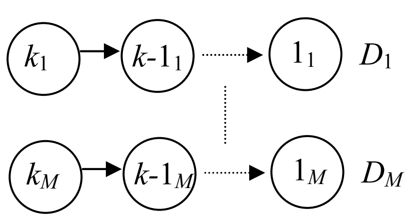


Figure 7

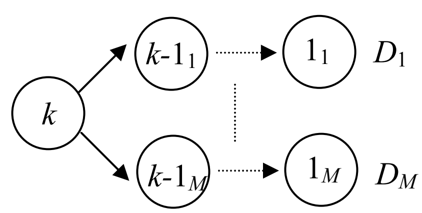


Figure 8

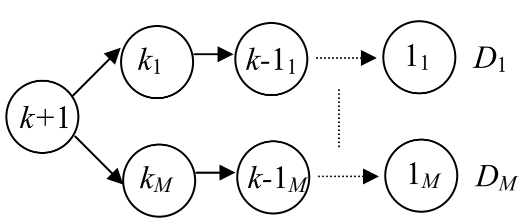


Figure 9

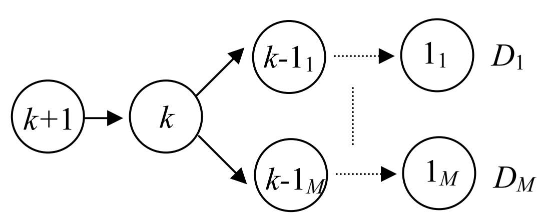


Figure 10

Assumption 2 is weaker than a decoupling assumption, as we need not assume the upstream echelon behaves as an infinite supplier for both topologies but rather it affects the downstream system in identical ways. Under a base-stock policy, the demand process observed at echelon $k+1$ is identical under both topologies and the additional echelon holding costs and implied backordering costs are constant across topologies. We can now present our two propositions.

Proposition 1: *Let $h'_{k-1} = h_{k-1} + \Delta$ and $h'_k = h_k - \Delta$. Then, under Assumptions 1 and 2, and holding all other parameters constant, the value of delaying differentiation is non-increasing in Δ. This effect becomes more pronounced as the number of final product forms increase.*

Proposition 2: *Let $h'_k = h_k + \Delta$ and $h'_{k+1} = h_{k+1} - \Delta$. Then, under Assumptions 1 and 2, and holding all other parameters constant, the value of delaying differentiation is non-decreasing in Δ. This effect becomes more pronounced as the number of final product forms increase.*

Proposition 1 states when larger holding costs are applied early in the process, the value of delayed differentiation is greater than when the holding costs are applied later in the process. This result is largely intuitive and arises from the decrease in inventory at the second stage associated with the pooling effect. Proposition 2 considers changes in the local holding costs that apply before the stage of differentiation and states the value of delayed differentiation increases as holding costs shift downstream towards the point of differentiation.

This non-intuitive result arises from the decrease in the effective backordering cost for carrying insufficient inventory in the LCO echelon. The lower upstream holding cost allows for greater inventory to be held at the installation just upstream of the LCO, insulating the LCO from stockouts in a manner that has previously been unobserved. By failing to capture the role of the entire supply chain when determining the value of delayed differentiation, the DH misses these important cost savings. In this instance, the holding cost profile serves as a qualitative indicator for the presence of potential savings via delayed differentiation, and the presence of 'spikes' in the holding cost from one stage to the next gives rise to the most favorable conditions for the use of delayed differentiation strategies.

An example of Proposition 2 may be found in the products of Experimental Craftworks (http://www.experimentalcraftworks.com), a handmade jewelry boutique. Experimental Craftworks designs and produces faceted gemstone and woven seed glass chandelier earrings. The gemstone products are simple in design and require little labor (approximately one half-hour) in assembly, but have relatively expensive raw materials, costing on average $35 per pair. The woven seed glass earrings, however, have inexpensive raw materials, costing only $5 per pair, but require substantial assembly time and expertise. Thus, although the raw materials costs differ, at the earring drop (completed subassembly) state, their values are approximately equal. Both products are also candidates for delayed differentiation strategies, as customers are sensitive to the type of earring backing, such as French hook, lever back, clip-on, or post. These backings are finalized in a separate production stage. Proposition 2 states the woven seed crystal earrings are more suitable for delayed differentiation. In essence, the lower raw material value of the seed crystals allows larger quantities to be held in raw inventory, reducing the effective backordering rate of stockouts.

5.5 Numerical Example Demonstrating How the DH May Lead To More Costly Supply Chain Configurations

To see the possible implications of the error induced by the DH, consider the case of a manufacturer who produces a product in three separate steps. Initially the product utilizes a generic component that is differentiated upon completion of the first stage. For ease of exposition, suppose the processing, shipping, and holding costs of the intermediate stage product are identical to that of a possible generic component at the second stage (i.e. the case of delayed differentiation). In other words, let $H_2 = H_{1,2} = H_{2,2}$, $T_2 = T_{1,2} = T_{2,2}$ and $P_2 = P_{1,2} = P_{2,2}$.

Suppose the manufacturer faces the following inventory and backorder costs: $H_3 = 1$, $H_2 = 3$, $H_{1,1} = H_{2,1} = 4$, and $b_1 = b_2 = 10$, corresponding to the parameters presented in the 10th row of example problems in Table 1. Suppose further that the firm faces an average demand of $\mu_1 = \mu_2 = 10$ units per period for each product, and the per-period annuitized cost of redesigning the process to allow for use of the generic component at the second stage is $S_2 = 6$. This is a set of parameters where we expect the DH to overestimate the value of delaying differentiation. From Table 2, we see that under the NBH, the firm is only willing to pay up to 3.61 cost units per period to enable delayed differentiation, and will thus opt to not implement the strategy. However, under the DH, the firm will pay up to 6.55 cost units, and will delay the differentiation of the products. In this case, solving the problem using the DH compares an expected cost of

$$Z(3)_{\text{Decoupling}} = 132.3+20*[P_1+T_1]+10*[T_{1,2}+T_{2,2}+T_{1,3}+T_{2,3}]+10*[P_{1,2}+P_{2,2}+P_{1,3}+P_{2,3}]$$

to

$$Z(2)_{\text{Decoupling}} = 125.8+6+20*[P_1+T_1+P_2+T_2]+10*[T_{1,3}+T_{2,3}] + 10*[P_{1,3}+P_{2,3}] .$$

By assumption, the processing and transportation costs are equivalent in the two cases and thus are ignored, leaving $Z(3)_{\text{Decoupling}} = 152.3$ and $Z(2)_{\text{Decoupling}} = 151.8$. Because $Z(2)_{\text{Decoupling}} < Z(3)_{\text{Decoupling}}$, the firm delays differentiation, believing the savings in inventory costs exceed the cost of redesigning the product for delayed differentiation. By a similar analysis, if the firm utilizes the NBH, it opts to *not* delay differentiation because $Z(3)_{\text{Bounds}} < Z(2)_{\text{Bounds}}$.

To see the importance of this difference, consider the difference between the total supply chain costs of the two strategies; that is, $Z(3)_{\text{Bounds}} - Z(2)_{\text{Decoupling}}$. Again, the processing and transportation costs are equal and thus cancel. Therefore the difference in expected costs between the two options is 5.81 per period. This represents 4.6% of the total expected inventory related costs per period.

6. Conclusions

When firms face increasing operational costs driven by product proliferation, they often turn to delayed differentiation as a potential cure. To properly assess the benefits of delayed differentiation, firms need to balance the savings from inventory risk-pooling with the costs of process and design modifications. In this paper, we make three major contributions: I) we provide guidance for when the decoupling assumption used by Lee and Tang (1997) may mislead a firm attempting to determine the optimal point to differentiate its products, II) we verify the Newsvendor Bounds Heuristic of Lystad and Ferguson (2005) is robust in three echelon topologies, and most importantly, III) we show the shape of the holding cost profile impacts the optimal point in the production process to delay differentiation.

We show in most cases, the benefits of delaying differentiation are smaller than those predicted by Lee and Tang (1997) due to their decoupling assumption, especially when the echelon holding costs at the last common operation are relatively large. This situation occurs when the majority of the value added processing occurs at the potential point of differentiation, because the inventory becomes relatively expensive at this point. Because the high value add stages also, sometimes incorrectly, appear to benefit the most from delaying differentiation, the decoupling assumption may lead to significant errors in supply chain design.

The decoupling assumption also underestimates the risk pooling savings when the echelon holding costs at stages upstream from the differentiating stage are high relative to the holding costs at other echelons and the backordering costs are high. This is due to a failure to exploit the holding cost structure in the intermediate installations, resulting in lower inventory levels, and consequently, smaller cost savings from risk pooling. This effect is exacerbated when inventories are large due to significant backordering costs.

Finally, we discover the non-intuitive and previously hidden insight that the shape of the holding cost profile significantly affects the optimal point in the process to delay differentiation. We show that the presence of sharp rises in local holding cost is associated with increased cost savings due to the reduction in effective backordering costs at the downstream stages. In other words, capturing holding cost spikes through the use of a common component is more valuable than previously believed, and may serve as an additional justification for using delayed differentiation strategies.

References

Alderson, W. 1950. "Marketing Efficiency and the Principle of Postponement", *Cost and Profit Outlook*, 3.

Aviv, Y., and Federgruen, A. "Capacitated Multi-Item Inventory Systems with Random and Seasonally Fluctuating Demands: Implications for Postponement Strategies", *Management Science*, Vol 47, No. 4, 512-531.

Baker, K., Magazine, M., and Nuttle, H. 1986. "The Effect of Commonality on Safety Stock in a Simple Inventory Model", *Management Science*, Vol. 32, No. 8, 982-988.

Bianco, A., 2004. "The Vanishing Mass Market", Business Week, July 12th.

Eynan, A., and Rosenblatt, M. 1996. "Component Commonality Effects on Inventory Costs", *IIE Transactions* 28, 93-104

Federgruen, A., and Zipkin, P. 1984. "Computational Issues in an Infinite-Horizon, Multi-echelon Inventory Problem with Stochastic Demand", *Operations Research*, 32, 4, 818-836.

Fisher, M., Ramdas, K., and Ulrich, K. 1999. "Component Sharing in the Management of Product Variety: A Study of Automotive Braking Systems", *Management Science*, 45, 5, 297-315.

Garg, A., and C. Tang. 1997. "On Postponement Structures for Product Families with Multiple Points of Differentiation", *IIE Transactions* 29, 8, 641-650.

Gerchak, Y., Magazine, M., and Gamble, B. 1988. "Component Commonality with Service Level Requirements", *Management Science*, Vol. 34, No. 6, 753-760.

Hsu, H., and Wang, W. 2004. "Dynamic Programming for Delayed Product Differentiation", *European Journal of Operations Research*,. Vol. 156, No. 1, 183-193.

Lee, H. 1996. "Effective Inventory and Service Management Through Product and Process Redesign", *Operations Research*, Vol 44, No. 1, 151-159.

Lee, H., and Tang, C. 1997. "Modelling the Costs and benefits of Delayed Product Differentiation", *Management Science*, Vol. 43, No. 1, 40-53.

Lystad, E,. and Ferguson, M. 2005. "Simple Newsvendor Heuristics for Multiechelon Distribution Systems," Working Paper, Georgia Institute of Technology.

Ma, S., Wang, W., and Liu, L. 2002. "Commonality and Postponement in multistage assembly systems", *European Journal of Operations Research*, 142, 523-538.

Tayur, S., Ganeshan, R., and Magazine, M. 1999. Quantitative Models for Supply Chain Management, Kluwer Academic Publishers

Zipkin, P. 2000. Foundations of Inventory Management, Irwin

Appendix 1

Non-Delayed (k=3) Cost Results					
Best Found Base Stock Policy					
b	s_3	s_2	s_1	Cost	% Error
2.5	15	9	13	8109.7	N/A
5	18	9	14	8935.5	N/A
7.5	17	11	14	9452.7	N/A
10	18	10	15	9822.9	N/A
20	19	11	16	10710.9	N/A
30	19	12	16	11210.8	N/A
50	20	12	17	11824.1	N/A
100	19	13	18	12599.2	N/A
Newsvendor Bounds Heuristic Policy					
b	s_3	s_2	s_1	Cost	% Error
2.5	18	9	13	8169.69	0.7
5	18	10	14	8985.39	0.6
7.5	18	11	14	9475.62	0.2
10	18	11	15	9864.56	0.4
20	19	11	16	10710.9	0.0
30	19	12	16	11210.8	0.0
50	20	12	17	11824.1	0.0
100	19	13	18	12599.2	0.0
Decoupling Heuristic Policy					
b	s_3	s_2	s_1	Cost	% Error
2.5	19	11	13	8588.5	5.9
5	19	11	14	9204.5	3.0
7.5	18	12	14	9606.4	1.6
10	17	12	15	9944.1	1.2
20	18	12	16	10758.6	0.4
30	17	13	16	11234.9	0.2
50	17	13	17	11850.4	0.2
100	17	13	18	12681.5	0.7

Table A.1: Stocking Level and Costs for Non-Delayed Chain (k=3)

Delayed (k=2) Cost Results					
Best Found Base Stock Policy					
b	s_3	s_2	s_1	Cost	% Error
2.5	17	16	13	7882.2	N/A
5	18	17	14	8632.6	N/A
7.5	19	19	14	9100.5	N/A
10	18	19	15	9444.2	N/A
20	19	20	16	10243.7	N/A
30	20	21	16	10694.3	N/A
50	21	21	17	11256.3	N/A
100	22	21	18	11952.2	N/A
Newsvendor Bounds Heuristic Policy					
b	s_3	s_2	s_1	Cost	% Error
2.5	18	18	13	7954.43	0.9
5	19	19	14	8715.86	1.0
7.5	20	20	14	9160.29	0.7
10	21	19	15	9519.34	0.8
20	21	20	16	10287.7	0.4
30	21	22	16	10748.9	0.5
50	22	22	17	11329.2	0.6
100	23	22	18	12017.9	0.5
Decoupling Heuristic Policy					
b	s_3	s_2	s_1	Cost	% Error
2.5	21	20	13	8376.9	6.3
5	22	19	14	8926.1	3.4
7.5	22	20	14	9282.6	2.0
10	22	19	15	9579.1	1.4
20	23	19	16	10338.5	0.9
30	23	20	16	10754.2	0.6
50	23	20	17	11289.7	0.3
100	24	19	18	12071.7	1.0

Table A.2: Stocking Level and Costs for Delayed Chain (k=2)

Asymmetric Costs, Nondelayed (k=3) Results												
Echelon Holding Costs				Backordering Costs		Best Found Base Stock Policy						
h_3	h_2	$h_{1,1}$	$h_{1,2}$	b_1	b_2	s_3	$s_{2,1}$	$s_{2,2}$	$s_{1.1}$	$s_{1.2}$	Cost	%Error

1	1	1	2	10	10	18	11	12	14	13	10131	N/A
1	2	1	2	10	10	19	9	9	15	14	12891	N/A
2	1	1	2	10	10	15	10	12	15	13	14969	N/A
2	2	1	2	10	10	17	9	9	15	14	17625	N/A
1	1	1	1	5	10	18	10	10	13	15	9388	N/A
1	1	1	1	5	20	18	10	11	13	16	9839	N/A
1	1	1	1	10	20	18	10	11	13	16	10386	N/A
1	1	1	2	5	5	16	10	11	14	12	9134	N/A
1	1	1	2	5	10	18	10	12	13	13	9696	N/A
1	1	1	2	5	20	18	10	12	13	15	10274	N/A
1	1	1	2	10	5	18	11	10	14	12	9572	N/A
1	1	1	2	10	20	18	11	12	14	15	10709	N/A
1	1	1	2	20	5	19	12	10	15	12	10008	N/A
1	1	1	2	20	10	19	12	11	15	13	10566	N/A
1	1	1	2	20	20	18	12	12	15	15	11146	N/A
Echelon Holding Costs				Backordering Costs		Newsvendor Bounds Heuristic Policy						
h_3	h_2	$h_{1,1}$	$h_{1,2}$	b_1	b_2	s_3	$s_{2,1}$	$s_{2,2}$	$s_{1.1}$	$s_{1.2}$	Cost	%Error
1	1	1	2	10	10	18	11	12	15	13	10150	0.2
1	2	1	2	10	10	21	9	10	15	14	12988	0.8
2	1	1	2	10	10	15	11	12	15	14	15016	0.3
2	2	1	2	10	10	17	10	10	15	14	17703	0.4
1	1	1	1	5	10	18	10	11	14	15	9499	1.2
1	1	1	1	5	20	20	10	11	14	16	9930	0.9
1	1	1	1	10	20	19	11	11	15	16	10423	0.4
1	1	1	2	5	5	17	10	12	14	12	9174	0.4
1	1	1	2	5	10	19	10	12	14	13	9757	0.6
1	1	1	2	5	20	19	10	12	14	15	10332	0.6
1	1	1	2	10	5	18	11	12	15	12	9650	0.8
1	1	1	2	10	20	19	11	12	15	15	10764	0.5
1	1	1	2	20	5	19	11	12	16	12	10103	1.0
1	1	1	2	20	10	20	11	12	16	13	10618	0.5
1	1	1	2	20	20	19	11	12	16	15	11157	0.1
Echelon Holding Costs				Backordering Costs		Decoupling Heuristic Policy						
h_3	h_2	$h_{1,1}$	$h_{1,2}$	b_1	b_2	s_3	$s_{2,1}$	$s_{2,2}$	$s_{1.1}$	$s_{1.2}$	Cost	%Error
1	1	1	2	10	10	18	12	12	15	13	10226	0.9
1	2	1	2	10	10	18	12	11	15	14	13256	2.8
2	1	1	2	10	10	18	12	11	15	14	15345	2.5
2	2	1	2	10	10	18	12	11	15	14	18367	4.2
1	1	1	1	5	10	18	11	12	14	15	9575	2.0
1	1	1	1	5	20	19	11	12	14	16	10017	1.8
1	1	1	1	10	20	18	12	12	15	16	10475	0.9
1	1	1	2	5	5	19	11	11	14	12	9288	1.7
1	1	1	2	5	10	19	11	12	14	13	9840	1.5

1	1	1	2	5	20	19	11	11	14	15	10406	1.3
1	1	1	2	10	5	19	12	11	15	12	9730	1.7
1	1	1	2	10	20	18	12	11	15	15	10829	1.1
1	1	1	2	20	5	19	12	11	16	12	10080	0.7
1	1	1	2	20	10	18	12	12	16	13	10593	0.2
1	1	1	2	20	20	18	12	11	16	15	11195	0.4

Table A.3: Asymmetric Results for Non-Delayed Differentiation Network

Asymmetric Costs, Delayed (k=2) Results											
Echelon Holding Costs				Backordering Costs		Best Found Base Stock Policy					
h_3	h_2	$h_{1,1}$	$h_{1,2}$	b_1	b_2	s_3	s_2	$s_{1,1}$	$s_{1,2}$	Cost	% Error
1	1	1	2	10	10	19	21	14	13	9740.4	N/A
1	2	1	2	10	10	20	17	15	14	12373.4	N/A
2	1	1	2	10	10	17	19	15	14	14481.6	N/A
2	2	1	2	10	10	18	18	14	14	17033.9	N/A
1	1	1	1	5	10	18	19	13	15	9040.7	N/A
1	1	1	1	5	20	19	18	13	17	9456.5	N/A
1	1	1	1	10	20	19	20	14	16	9845.9	N/A
1	1	1	2	5	5	18	20	13	12	8826.0	N/A
1	1	1	2	5	10	19	19	13	14	9343.1	N/A
1	1	1	2	5	20	19	21	12	15	9893.7	N/A
1	1	1	2	10	5	19	20	14	12	9229.5	N/A
1	1	1	2	10	20	20	20	14	15	10266.8	N/A
1	1	1	2	20	5	19	20	16	12	9627.4	N/A
1	1	1	2	20	10	20	21	15	13	10137.6	N/A
1	1	1	2	20	20	20	21	15	15	10667.8	N/A
Echelon Holding Costs				Backordering Costs		Newsvendor Bounds Heuristic Policy					
h_3	h_2	$h_{1,1}$	$h_{1,2}$	b_1	b_2	s_3	s_2	$s_{1,1}$	$s_{1,2}$	Cost	% Error
1	1	1	2	10	10	20	21	15	13	9826.5	0.9
1	2	1	2	10	10	22	18	15	14	12520.6	1.2
2	1	1	2	10	10	18	20	15	14	14581.9	0.7
2	2	1	2	10	10	19	18	15	14	17150.7	0.7
1	1	1	1	5	10	20	19	14	15	9126.9	1.0
1	1	1	1	5	20	21	20	14	16	9623.2	1.8
1	1	1	1	10	20	21	20	15	16	9943.3	1.0
1	1	1	2	5	5	19	20	14	12	8905.7	0.9
1	1	1	2	5	10	20	21	14	13	9510.1	1.8
1	1	1	2	5	20	20	21	14	15	10040.3	1.5
1	1	1	2	10	5	20	21	15	12	9354.3	1.4
1	1	1	2	10	20	22	20	15	15	10392.8	1.2
1	1	1	2	20	5	20	22	16	12	9775.0	1.5

1	1	1	2	20	10	22	21	16	13	10265.6	1.3
1	1	1	2	20	20	21	21	16	15	10739.9	0.7
Echelon Holding Costs				Backordering Costs		Decoupling Heuristic Policy					
h_3	h_2	$h_{1,1}$	$h_{1,2}$	b_1	b_2	s_3	s_2	$s_{1,1}$	$s_{1,2}$	Cost	% Error
1	1	1	2	10	10	22	20	15	13	9869.2	1.3
1	2	1	2	10	10	22	19	15	14	12650.0	2.2
2	1	1	2	10	10	22	19	15	14	14946.1	3.2
2	2	1	2	10	10	22	19	15	14	17748.7	4.2
1	1	1	1	5	10	22	19	14	15	9254.2	2.4
1	1	1	1	5	20	22	20	14	16	9695.7	2.5
1	1	1	1	10	20	23	19	15	16	9995.7	1.5
1	1	1	2	5	5	21	20	14	12	9017.9	2.2
1	1	1	2	5	10	21	20	14	13	9498.5	1.7
1	1	1	2	5	20	22	19	14	15	10018.5	1.3
1	1	1	2	10	5	19	20	15	12	9247.8	0.2
1	1	1	2	10	20	22	19	15	15	10343.3	0.7
1	1	1	2	20	5	22	20	16	12	9739.8	1.2
1	1	1	2	20	10	22	20	16	13	10204.3	0.7
1	1	1	2	20	20	22	19	16	15	10695.5	0.3

Table A.4: Asymmetric Results for Delayed Differentiation Network

Appendix 2: Proofs of Propositions 1 and 2

Propositions 1 and 2 investigate the difference in inventory costs associated with varying production network topologies. Under Assumption 2, we may analyze the network topologies below in Figures A1 and A2.

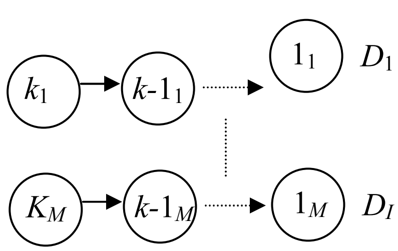


Figure A1

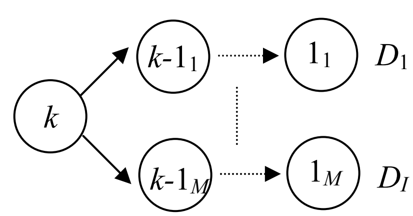


Figure A2

We define the critical fractiles

$$\Theta_n^{l,c} = \frac{b + \sum_{j=n+1}^{N} h_j}{b + \sum_{j=1}^{N} h_j} \qquad \Theta_n^{u,c} = \frac{b + \sum_{j=n+1}^{N} h_j}{b + \sum_{j=n}^{N} h_j}$$

and let

$z_n^{l,c} = \Phi^{-1}\left(\Theta_n^{l,c}\right)$ and $z_n^{u,c} = \Phi^{-1}\left(\Theta_n^{u,c}\right)$.

Let $\phi(\cdot)$ and $\Phi(\cdot)$ represent the standard normal pdf and cdf, respectively. Following the approach in Zipkin (2000) (see also Shang and Song, 2003),

$$s_{i,n}^{l,d} = \mu_i \tilde{L}_n + z_n^{l,d} \sqrt{\sigma_i^2 \tilde{L}_n} \tag{A1}$$

$$s_n^{l,d} = \sum_{i=1}^{M} s_{i,n}^{l,d} \tag{A2}$$

$$s_{i,n}^{u,d} = \mu_i \tilde{L}_n + z_n^{u,d} \sqrt{\sigma_i^2 \tilde{L}_n} \tag{A3}$$

$$s_n^{u,d} = \sum_{i=1}^{M} s_{i,n}^{u,d} \tag{A4}$$

$$s_n^{l,c} = \sum_{i=1}^{M} \left(\mu_i \tilde{L}_n\right) + z_n^{l,c} \sqrt{\sum_{i=1}^{M} \left(\sigma_i^2 \tilde{L}_n\right)} \tag{A5}$$

$$s_n^{u,c} = \sum_{i=1}^{M} \left(\mu_i \tilde{L}_n\right) + z_n^{u,c} \sqrt{\sum_{i=1}^{M} \left(\sigma_i^2 \tilde{L}_n\right)} \tag{A6}$$

$$C_{i,n}^{l,d}\left(s_{i,n}^{u,d}\right) = \left(b_i + \sum_{j=n}^{k+1} h_{i,j}\right) \phi\left(z_n^{u,d}\right) \sqrt{\sigma_i^2 \tilde{L}_n} + \sum_{j=1}^{n-1} h_{i,j+1} \mu_i \tilde{L}_n \tag{A7}$$

$$C_n^{l,d} = \sum_{i=1}^{M} C_{i,n}^{l,d} \tag{A8}$$

$$C_{i,n}^{u,d}\left(s_{i,n}^{l,d}\right) = \left(b_i + \sum_{j=1}^{k+1} h_{i,j}\right) \phi\left(z_n^{l,d}\right) \sqrt{\sigma_i^2 \tilde{L}_n} + \sum_{j=1}^{n-1} h_{i,j+1} \mu_i \tilde{L}_n \tag{A9}$$

$$C_n^{u,d} = \sum_{i=1}^{M} C_{i,n}^{u,d} \tag{A10}$$

$$C_n^{l,c}\left(s_n^{u,c}\right) = \left(b + \sum_{j=n}^{k+1} h_j\right) \phi\left(z_n^{u,c}\right) \sqrt{\sigma^2 \tilde{L}_n} + \sum_{j=1}^{n-1} h_{j+1} \mu \tilde{L}_n \tag{A11}$$

$$C_n^{u,c}\left(s_n^{l,c}\right) = \left(b + \sum_{j=1}^{k+1} h_j\right) \phi\left(z_n^{l,c}\right) \sqrt{\sigma^2 \tilde{L}_n} + \sum_{j=1}^{n-1} h_{j+1} \mu \tilde{L}_n \tag{A12}$$

Let the value of delaying differentiation,

$V_d = Z(k+1)-Z(k)$ (A13)

Proof of Proposition 1

By decomposing the network in Figure A2, a lower bound on V_d is obtained. Under the decomposition, the two networks are identical, thus

$$V_d \geq 0 \tag{A14}$$

By allowing instantaneous and costless transshipments between echelons in Figure A2, an upper bound on V_d is obtained. This transforms Figure A2 into a serial network. By reducing the serial chains to single stage problems (see Shang and Song (2003)), we have an upper bound for V_d which is in turn bounded by

$$C_k^{l,d}\left(s_k^{u,d}\right)-C_k^{l,c}\left(s_k^{u,c}\right) \tag{A15}$$

and

$$C_k^{u,d}\left(s_k^{l,d}\right)-C_k^{u,c}\left(s_k^{l,c}\right). \tag{A16}$$

Equation A15 is equal to

$$M\left(\left(b+\sum_{j=k}^{k+1}h_j\right)\phi\left(z_k^{u,c}\right)\sqrt{\sigma_i^2\tilde{L}_k}+\sum_{j=1}^{k-1}h_{j+1}\mu_i\tilde{L}_j\right)-\left(b+\sum_{j=k}^{k+1}h_j\right)\phi\left(z_k^{u,c}\right)\sqrt{M\sigma_i^2\tilde{L}_k}+\sum_{j=1}^{k-1}h_{j+1}M\mu_i\tilde{L}_j$$

$$=M\left(b+\sum_{j=k}^{k+1}h_j\right)\phi\left(z_k^{u,c}\right)\sqrt{\sigma_i^2\tilde{L}_k}-\left(b+\sum_{j=k}^{k+1}h_j\right)\phi\left(z_k^{u,c}\right)\sqrt{M\sigma_i^2\tilde{L}_k}$$

$$=\left(b+\sum_{j=k}^{k+1}h_j\right)\phi\left(z_k^{u,c}\right)\sqrt{\left(M^2-M\right)\sigma_D^2\tilde{L}_k} \tag{A17}$$

Let $h_{k-1}{}' = h_k+\Delta$ and $h_k{}' = h_k-\Delta$. Then Equation A17 becomes

$$\left(b+h_{k+1}+h_k{}'\right)\phi\left(\Phi^{-1}\left(\frac{b+h_k}{b+h_{k+1}+h_k{}'}\right)\right)\sqrt{(M^2-M)\sigma_i^2\tilde{L}_k}$$

$$=\left(b+h_{k+1}+h_k-\Delta\right)\phi\left(\Phi^{-1}\left(\frac{b+h_k}{b+h_{k+1}+h_k-\Delta}\right)\right)\sqrt{(M^2-M)\sigma_d^2\tilde{L}_k}$$

$$\leq \left(b+h_{k+1}+h_k\right)\phi\left(\Phi^{-1}\left(\frac{b+h_{k+1}}{b+h_{k+1}+h_k}\right)\right)\sqrt{(M^2-M)\sigma_d{}^2\tilde{L}_k} \tag{A18}$$

Hence Equation A17 is decreasing in Δ.

Likewise, Equation A16 is equal to

$$\left(b+\sum_{j\neq k,k-1} h_j+h_k'+h_{k-1}'\right)\phi\left(\Phi^{-1}\left(\frac{b+h_{k+1}}{b+\sum_{j\neq k,k-1} h_j+h_k'+h_{k-1}'}\right)\right)\sqrt{(M^2-M)\sigma_i{}^2\tilde{L}_k}$$

$$=\left(b+\sum_{j\neq k,k-1} h_j+h_k+h_{k-1}\right)\phi\left(\Phi^{-1}\left(\frac{b+h_{k+1}}{b+\sum_{j\neq k,k-1} h_j+h_k+h_{k-1}}\right)\right)\sqrt{(M^2-M)\sigma_i{}^2\tilde{L}_k} \tag{A19}$$

Hence Equation A16 is independent of Δ.

Thus the lower bound of V_d is independent of Δ, while the upper bound is non-increasing (and potentially decreasing) in Δ. Hence shifting holding costs downstream from the point of differentiation may decrease the value of delaying differentiation. As *M* increases, the factor $\sqrt{(M^2-M)\sigma_i{}^2\tilde{L}_k}$ also increases, hence the decrease in the value of delayed differentiation attributable to the holding cost shift is increasing in *M*.

Proof of Proposition 2

By a similar argument as in Proposition 1, we have

$$V_d \geq 0 \tag{A20}$$

$$C_{k+1}^{l,d}\left(s_{k+1}^{u,d}\right)-C_{k+1}^{l,c}\left(s_{k+1}^{u,c}\right)$$

$$=M\left(b+h_{k+1}+h_k\right)\phi\left(z_k^u\right)\sqrt{\sigma_D{}^2\tilde{L}_k}+M\sum_{j=1}^{k-1}h_{j+1}\mu_i\tilde{L}_{\boldsymbol{j}}-\left(b+h_{k+1}+h_k\right)\phi\left(z_k^u\right)\sqrt{M\sigma_i^2\tilde{L}_k}+\sum_{j=1}^{k-1}h_{j+1}M\mu_i\tilde{L}_k$$
$$=M\left(b+h_{k+1}+h_k\right)\phi\left(z_k^u\right)\sqrt{\sigma_i^2\tilde{L}_k}-\left(b+h_{k+1}+h_k\right)\phi\left(z_k^u\right)\sqrt{M\sigma_i^2\tilde{L}_k}$$
$$=\left(b+h_{k+1}+h_k\right)\phi\left(z_k^u\right)\sqrt{\left(M^2-M\right)\sigma_i^2\tilde{L}_k} \tag{A21}$$

and

$$C_k^{u,d}\left(s_k^{l,d}\right)-C_k^{u,c}\left(s_k^{l,c}\right)$$
$$=M\left(b\ +\sum_{j=1}^{k+1}h_i\right)\phi\left(z_k^{l,d}\right)\sqrt{\sigma_i{}^2\tilde{L}_k}+M\sum_{j=1}^{i-1}h_{j+1}\mu\tilde{L}_j-\left(b+\sum_{j=1}^{k+1}h_i\right)\phi\left(z_k^{l,d}\right)\sqrt{M\sigma_i{}^2\tilde{L}_k}+\sum_{j=1}^{i-1}h_{j+1}M\,\mu_i\tilde{L}_j$$
$$=M\left(b\ +\sum_{j=1}^{k+1}h_i\right)\phi\left(z_k^{l,d}\right)\sqrt{\sigma_i{}^2\tilde{L}_k}-\left(b+\sum_{j=1}^{k+1}h_i\right)\phi\left(z_k^{l,d}\right)\sqrt{M\sigma_i{}^2\tilde{L}_k}$$
$$=\left(b\ +\sum_{j=1}^{k+1}h_i\right)\phi\left(z_k^{l,d}\right)\sqrt{(M^2-M)\sigma_i{}^2\tilde{L}_k} \tag{A22}$$

Let $h_2^{'}=h_2+\Delta$ and $h_3^{'}=h_3-\Delta$. Then (A21) becomes

$$\left(b+h_{k+1}{}^{'}+h_k{}^{'}\right)\phi\left(\Phi^{-1}\left(\frac{b+h^{'}_{k+1}}{b+h^{'}_{k+1}+h_k^{'}}\right)\right)\sqrt{\left(M^2-M\right)\sigma_i{}^2\tilde{L}_k}$$
$$=\left(b+h_{k+1}+h_k\right)\phi\left(\Phi^{-1}\left(\frac{b+h_{k+1}-\Delta}{b+h_{k+1}+h_k}\right)\right)\sqrt{\left(M^2-M\right)\sigma_i{}^2\tilde{L}_k}$$
$$\geq\left(b+h_{k+1}+h_k\right)\phi\left(\Phi^{-1}\left(\frac{b+h_{k+1}}{b+h_{k+1}+h_k}\right)\right)\sqrt{\left(M^2-M\right)\sigma_i{}^2\tilde{L}_k} \tag{A23}$$

Hence (A21) is increasing in Δ.

Likewise, (A22) becomes

$$\left(b+\sum_{j\neq k+1,k}h_j+h^{'}_{k+1}+h_k{}^{'}\right)\phi\left(\Phi^{-1}\left(\frac{b+h^{'}_{k+1}}{b+\sum_{j\neq k+1,k}h_j+h^{'}_{k+1}+h_k^{'}}\right)\right)\sqrt{(M^2-M)\sigma_i{}^2\tilde{L}_k}$$
$$=\left(b+\sum_{j=1}^{k=1}h_j\right)\phi\left(\Phi^{-1}\left(\frac{b+h_{k+1}-\Delta}{b+\sum_{j=1}^{k=1}h_j}\right)\right)\sqrt{(M^2-M)\sigma_i{}^2\tilde{L}_k}$$
$$\geq\left(b+\sum_{j=1}^{k=1}h_j\right)\phi\left(\Phi^{-1}\left(\frac{b+h_{k+1}}{b+\sum_{j=1}^{k=1}h_j}\right)\right)\sqrt{(M^2-M)\sigma_i{}^2\tilde{L}_k} \tag{A24}$$

Hence (A22) is increasing in Δ.

Thus the upper bound of V_d is increasing in Δ, the shift of holding costs from the third echelon to the second, through the mechanism of effective backordering cost rate at the second echelon. The lower bound remains independent of Δ, hence the value of delaying differentiation is non-decreasing, and likely increasing in Δ. As M increases, the factor $\sqrt{(M^2 - M)\sigma_i^2 \tilde{L}_k}$ also increases, hence the increase in the value of delayed differentiation attributable to the holding cost shift is increasing in M.