Newsvendor Networks: Inventory Management and Capacity Investment with Discretionary Activities

Jan A. Van Mieghem • Nils Rudi

Kellogg School of Management, Northwestern University, Evanston, Illinois 60208-2009
Simon Graduate School of Business Administration, University of Rochester, Rochester, New York 14627
vanmieghem@kellogg.northwestern.edu • rudi@simon.rochester.edu

---

We introduce a class of models, called *newsvendor networks*, that allow for multiple products and multiple processing and storage points and investigate how their single-period properties extend to dynamic settings. Such models provide a parsimonious framework to study various problems of stochastic capacity investment and inventory management, including assembly, commonality, distribution, flexibility, substitution and transshipment. Newsvendor networks are stochastic models with recourse that are characterized by linear revenue and cost structures and a linear input-output transformation. While capacity and inventory decisions are locked in before uncertainty is resolved, some managerial discretion remains via ex-post input-output activity decisions. Ex-post decisions involve both the choice of activities and their levels and can result in subtle benefits. This discretion in choice is captured through alternate or "nonbasic" activities that can redeploy inputs and resources to best respond to resolved uncertain events. Nonbasic activities are never used in a deterministic environment; their value stems from discretionary flexibility to meet stochastic demand deviations from the operating point.

The optimal capacity and inventory decisions balance overages with underages. Continuing the classic newsvendor analogy, the optimal balancing conditions can be interpreted as specifying multiple "critical fractiles" of the multivariate demand distribution; they also suggest appropriate measures for and trade-offs between product service levels. This paper shows that the properties of optimal newsvendor network solutions extend to a dynamic setting under plausible conditions. Indeed, we establish dynamic optimality of inventory and capacity policies for the lost sales case. Depending on the nonbasic activities, this also extends to the backordering case. Analytic- and simulation-based solution techniques and graphical interpretations are presented and illustrated by a comprehensive example that features discretionary input commonality and a flexible processing resource.

(*Inventory; Capacity; Assembly; Commonality; Distribution; Flexibility; Substitution; Transshipment; Multiple Products*)

---

1. Introduction

The classic "newsboy" model provides a very simple, yet effective framework for studying a variety of stochastic economic decision problems. Using its gender-neutral name, the newsvendor must decide how much of a particular divisible asset—for example, newspapers—to buy at a constant unit cost knowing only a probabilistic forecast of the demand for that asset, which is to be sold at a given unit price. After demand is observed and sales are made, either excess demand results in lost sales or excess assets are salvaged at a loss. In deciding the optimal quantity the

newsvendor must trade off the cost of overage and underage, resulting in the well-known ''critical fractile'' solution. This simple, yet powerful model forms the basis for inventory, cash, capacity, and a variety of other managerial problems. A vast literature has modified or extended the classic newsvendor model. The ''multidimensional newsvendor model'' introduced by Harrison and Van Mieghem (1999) and Van Mieghem (1998a) generalizes the classic newsvendor model by allowing for multiple products and multiple processing points. Multidimensional newsvendor models focus on processing capacity constraints and consider essentially a single period by excluding inventory holding and carryover.

This paper continues the agenda of Harrison and Van Mieghem by incorporating multiple storage points into the multidimensional newsvendor model. We call the resulting broader class of models *newsvendor networks.* The inclusion of inventory holding points allows a direct extension of the single-period model into a truly dynamic setting where any leftover stock at the end of one period carries over as input to the next period. While *processing-storage networks* would be a perfectly appropriate name for the class of models that we consider, we chose ''newsvendor networks'' to stress the link to the classic newsvendor model. Indeed, newsvendor networks inherit many of its classic predecessor's advantages—such as parsimony, tractability, and effectiveness in yielding insights into stochastic planning—but also its main disadvantage: The model may be too stylized to capture details necessary for practical decision support systems.

To illustrate the features of a newsvendor network consider the example depicted in Figure 1, which we will carry throughout our discussion. Adopting the process nomenclature of Anupindi et al. (1999), the entities that flow through the network are called ''flow units.'' Before demand is known, a set of ''ex-ante'' activities are performed onto the inputs and their results are stored in ''stocks'' or inventories. These activities can be simple purchasing or preprocessing activities. After demand is realized, ''ex-post'' activities process stocked inputs into demanded outputs using resources. In addition to being constrained

Figure 1   **A Newsvendor Network Example Featuring a Discretionary Common Input 1 and a Flexible Processing Resource 2**
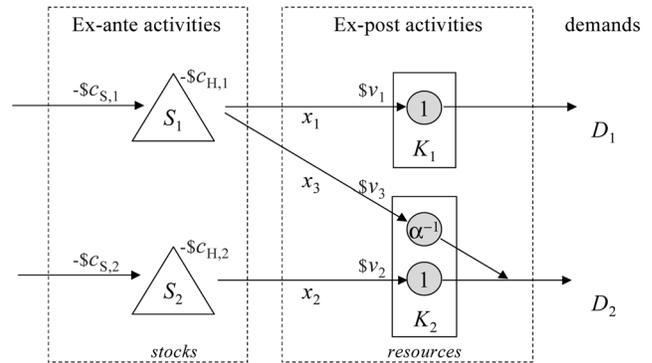
by demand, the sales or output rate is also constrained both by the input stock levels and by the resource capacities, denoted by vectors S and K, respectively. Both types of activities generate cash flows: The ex-ante activities incur marginal cost vector c_S, while the ex-post activities generate net marginal value vector v, which includes the price minus marginal processing and demand shortage penalty costs. Finally, units carried over to a subsequent period incur a holding cost c_H.

The example captures some key characteristics of newsvendor networks. First, there are multiple inputs—and thus inventories—that are transformed into multiple outputs (products) by utilizing a network of resources that are linearly capacity constrained. The ex-post activity levels x in this *linear production technology* with *linear financial structure* are thus constrained by both input stocks S and capacities K. For example, Activities 3 and 2 deplete stocks 1 and 2, respectively, and consume Resource 2's capacity at rate α^{-1} and 1, respectively. (As we shall illustrate later, newsvendor networks can also easily handle an activity that simultaneously depletes several complementary stocks or a less traditional activity that simultaneously requires multiple complementary resources.) The inventory constraints are: $x_1 + x_3 \leq S_1$ and $x_2 \leq S_2$, while the capacity constraints are: $x_1 \leq K_1$ and $x_2 + \alpha^{-1}x_3 \leq K_2$. Newsvendor networks are thus about three decisions: capacity investment decisions K, input inventory procurement decisions S, and activity decisions $x(K, S, D)$. (The multidimensional newsvendor model is only about K and x.)

A second characteristic is that while both the capacity investment and inventory procurement decisions are locked in before demand uncertainty is resolved as usual, newsvendor networks allow for ex-post ''discretionary'' or ''alternate'' activities during the input-output transformation. Thus, multiple options may exist to produce a given output and the ex-post decisions thus involve both choice and quantity: *how* to convert inputs into outputs, as well as *how many.* The example features input substitution, which we will also refer to as *discretionary commonality*: While the ''normal'' or ''basic'' activity is 2, the process manager has the option to draw from the more costly Input 1 to process Product 2. In other words, the manager has ex-post discretion in choosing whether Input 1 should be common to both products or not (by using the ''alternate'' (nonbasic) Activity 3 or not). This discretion requires Resource 2 to be flexible in the sense that it can transform either type of input. Typically, the resource would be ''better'' at its basic activity than its alternate activity because by design the latter is to be used only sparingly. Hence, the capacity consumption rate and the processing cost of Activity 3 may exceed those of Activity 2: $\alpha^{-1} \geq 1$ and $v_3 \leq v_2$. Thus, $0 \leq \alpha \leq 1$ and α can be interpreted as a measure of the product flexibility of Resource 2 with $\alpha = 1$ representing perfect processing flexibility, while $v_3 = v_2$ represents perfect financial flexibility.

This redundancy in activities highlights the flexibility inherent in the formulation that admits a variety of interesting applications that have more activities than end-products and/or inputs. Obviously, newsvendor networks can capture assembly operations where one activity simultaneously consumes multiple input types in fixed proportions, distribution activities that fill multiple demands from a single input, or commonality settings where multiple products draw from a common input. While distribution and commonality result in well-known inventory pooling, none of those activities are discretionary or nonbasic. Indeed, nonbasic activities are never used in a deterministic environment; their value stems from the discretionary flexibility to meet stochastic demand deviations from the operating point. Discretionary activities thus model input- or resource-substitution and provide subtle benefits in additional to traditional ''risk pooling.'' The analysis here will distinguish resource pooling (i.e., when alternative means for processing a given set of inputs are available) and inventory pooling (i.e., when a set of outputs can be provided from a smaller set of inputs). The discretion imbedded in resource pooling generates an ex-post revenue maximization option that can remain valuable even when risk pooling is not, as discussed in Van Mieghem (2002b).

Whereas these first two characteristics of multidimensionality and discretionary activities illustrate modeling objectives, the second objective of this paper addresses the analysis of newsvendor networks in a dynamic or multiperiod setting. It is in this setting that the difference between inventories, which store flow units and link successive periods, and capacities, which limit processing activity levels, matters and is clarified. In addition, we will establish the dynamic optimality of the myopic policy for the case where excess demands result in lost sales. This result will be extended to the case of backlogging for a restricted set of newsvendor networks that are uncapacitated and only have ''strong'' nonbasic activities, which will be defined later. Thus, while the lost sales case is a generalization of the optimality of the classic single-item base-stock policy, this paper highlights some of the technical difficulties that arise when unmet demand is backlogged in a network with discretionary, capacitated activities.

The objectives of newsvendor networks and this paper can now be summarized as follows. Newsvendor networks are used to synthesize newsvendor literature, to develop richer understanding of the value and purpose of basic and nonbasic or discretionary activities, and to investigate how insights apply to multiple-period settings. From a modeling standpoint, newsvendor networks are a direct extension to the multidimensional newsvendor model by allowing for multiple storage points, and their solution can be obtained by analogy, as shall be shown in §2. The key difference between this paper and those on the multidimensional newsvendor model is not modeling, but how the model is *used* and what we *learn* from that application. By formalizing the natural extension

of multidimensional newsvendor models to include storage points and simultaneously consider ex-ante inventory and capacity decisions, newsvendor networks can be used to synthesize a variety of newsvendor-type models, and they can be used to develop a much richer understanding of the value and the purpose of discretionary activities. (Multidimensional newsvendor models were not used to achieve either of these effects.) In passing, we will also suggest appropriate measurements of and trade-offs between product service levels, which are not obvious in a multiproduct setting.

This paper shows that the structure of the optimal policy in a single-period newsvendor network extends to a dynamic setting under plausible conditions. Thus we provide a rigorous proof of dynamic optimality under certain conditions and give some initial indications of factors that may prevent us from establishing such dynamic optimality. Obviously, under the conditions that we give, the result is not surprising to anyone familiar with dynamic control. (However, hardly any dynamic optimality result of a simple stationary policy is surprising.) The contribution here must be found in providing a stepping stone for future work and in "checking off" a necessary task in building newsvendor networks theory. Indeed, the dynamic optimality result directly generalizes a set of previously published, single-period studies and insights to a dynamic setting. Specifically, our results suggest that earlier papers that focused only on capacity *or* inventory in a single-period newsvendor model will retain their insights when dynamic inventory management is added to the modeling setup.

To put the two objectives in perspective, it is helpful to clearly delineate *the boundaries of this paper.* First, while newsvendor networks attempt to contribute to the literature on capacity and the literature on inventory, there are many other existing models that incorporate both capacity and inventory and where both of these factors affect system performance. By design, newsvendor networks are clearly not the most detailed models of reality, yet they do bridge the two literatures and several functional or problem-specific areas. In addition, they allow for a rather general set-

ting of multiple products flowing through both multiple storage inventories and processing capacities in a multiperiod context. Second, continuing the abstraction of reality: The models as presented here involve centralized optimization in a risk-neutral setting. Newsvendor networks do extend to multiperson, decentralized game-theoretic settings (e.g., Anupindi et al. 2001, Rudi et al. 2001, Van Mieghem 1999) as well as risk-sensitive settings, but no attempt will be made to prove that assertion in this article. (Van Mieghem 2002a discusses how to incorporate risk considerations.) Yet centralized control does imply that certain customer-driven activities, such as customers randomly substituting among the set of available products (e.g., Lippman and McCardle 1997, Netessine and Rudi 2002, Parlar 1988), are not captured in our current framework. Also, by design we fix the capacity of the processing resources at the beginning of the time period to highlight the difference between capacity, which is constant throughout time periods, and inventory, whose fluctuations connect time periods. Multiresource dynamic capacity adjustment, as in Eberly and Van Mieghem (1997), could be incorporated, but the added complication would bring us beyond the scope of this first newsvendor networks paper. Finally, the model presented here is "single-stage" in the sense that inventory storage points are not connected. Section 5 will discuss the complications that arise when one extends the newsvendor setting and allows for output inventories, in addition to input inventories. In the "true newsvendor setting" the basic trade-off is between holding inventory and losing sales or having backorders. When we consider both input and output inventories, we must decide *where* to keep inventory in addition to *how much.* This relates to echelon inventories, the marvelous concept of a simple, plausible local control scheme, using a base-stock policy at each stage that is equivalent to fully centralized control in uncapacitated supply chains without discretionary activities. In the multiproduct setting with discretionary activities, such equivalence does not always exist. Indeed, this is due to the distinct functions of input vs. output inventories: Input inventories are low-cost holding points against stockouts and

partially pool output demands, whereas output inventories offer a hedge against production capacity constraints. As such, input inventories and capacities are typically economic complements, while output inventories and capacities tend to be economic substitutes.

Structure of the Paper. After a literature review, §2 starts with the formulation of a newsvendor network, illustrates some network examples and summarizes analytic properties of the optimal inventory and capacity decision. Section 3 analyzes the illustrative example of the newsvendor network in Figure 1. Section 4 extends the formulation to a multiperiod setting and presents dynamic optimality results of the myopic inventory base-stock policy and the capacity decision. Section 5 discusses complications that may arise when allowing for backlogging or output inventories.

Related Literature. There is a vast literature using the classic (single-dimensional) newsvendor model or some variation of it. Khouja (1999) gives an extensive literature overview of the variations and extensions to the classic model.

Several models of multi-item management of either inventories or capacities fit into the framework of newsvendor networks. In addition to the multidimensional newsvendor model of Harrison and Van Mieghem (1999), newsvendor networks also capture existing models of centralization (Eppen 1979), commonality (Gerchak and Henig 1989, Tayur 1995, Van Mieghem 2002b), flexibility (Van Mieghem 1998a, Rudi and Zheng 1997), substitution (Bassok et al. 1999, Netessine et al. 2002), and transshipment (Krishnan and Rao 1965, Robinson 1990), as will be discussed in §2. Multiechelon inventory models are related to our framework in that it deals with inventories in networks, although the typical focus is to address the impact of holding costs, setup costs, and leadtimes (see Chen and Zheng 1994 and references therein). Federgruen and Zipkin (1986a–b) are the classic references for capacitated inventory management. Other papers that have combined inventory and capacity management include Angelus and Porteus (2002), for a single item, and work by Kapuscinski and Tayur (1998). The latter builds on pioneering

work by Glasserman and Tayur (1995) on the use of infinitesimal perturbation analysis (IPA) for calculating optimal base-stock policies in complex multiitem, capacitated inventory systems. While our framework is simpler, in that we focus on analytic results for single-stage systems, we also discuss IPA as the method to be used in practice. The formulation of our newsvendor network is strongly influenced by the recent and ongoing work of Harrison (2001a–b) that presents stochastic processing networks in relation to the classical field that T.C. Koopmans called activity analysis. That general linear model of the firm with activities performed on flow units by resources is also our building block, and linear programming is an important analysis tool. The name discretionary or alternate activity was coined by Harrison and our coverage of such activities mirrors his. The name is an abstraction of existing ideas relating substitution but also allows more general and unexplored types of flexibility or ''real options'' as will be illustrated in the next section. There are, however, some important distinctions between our newsvendor network and Harrison's processing network. Newsvendor networks are a much simpler subset of stochastic networks. Our setting is discrete-time and assumes deterministic ex-post processing. More importantly, we adopt a setting more in line with supply chains, where output demand is exogenous but inputs, capacities, and activities are controlled. The simplification of deterministic processing allows us to get more structural results. In that sense, this paper is complementary to Harrison's work.

It is appropriate to finish this introduction with some of Zipkin's caveats (2000). As in all stylized models, one walks a fine line between abstraction, which favors tractability, and realism, which leads to notoriously hard problems in multidimensional systems. It is important to recognize the implicit assumptions in newsvendor networks: As in many inventory models, the supply system is exogenous and has constant, small leadtimes. This is clearly a simplifying assumption. A more realistic model endogenizes the supply system and the leadtimes into a multistage processing network, which is much harder to analyze. (In many situations, however, it is reason-

able to assume that ''our process'' represents only a small part of the overall demands on the supply system.) Regardless, newsvendor networks are a useful and tractable starting point to analyze and understand the fundamentals of the harder problems.

2. Newsvendor Network Formulation

A newsvendor network is a stochastic, linear decision model with recourse. (As usual, newsvendor networks are defined as a single-period model; §4 presents its dynamic extension.) Consistent with the multidimensional newsvendor model, it is defined by a linear production technology, which describes how inputs (supply) are transformed into outputs to fill end-product demand, a linear financial structure, and a probability distribution of end-product demand. Its primitive attributes are l different resources that consume m distinct inputs (or stocks) to produce n distinct outputs by means of p different processing activities, in addition to the m ordering activities. Inputs and outputs are also called flow units, which may include traditional ''materials.'' The timing is exactly as it is in the classic newsvendor: Only the probability distribution P of end-product demand is known when the ordering and capacity investment activities are made. Then demand uncertainty is resolved, after which the processing activities are chosen with full knowledge of the actual demand. Hence, we shall also refer to the sourcing and processing activities as the ex-ante and ex-post activities, respectively.

Let R_{ij} denote the amount of flow unit i consumed per unit of activity j, with a negative value interpreted to mean that activity j is a net producer of flow unit i. In a supply chain setting, one traditionally partitions flow units in two classes: The m input stocks are being consumed, while n distinct outputs are being generated. Accordingly, partition the input-output matrix R as $R = [R_{Sj} - R_D]$, where both submatrices R_S and R_D are nonnegative. Notice that this setup allows for an activity to simultaneously consume multiple (complementary) inputs or produce multiple outputs. Similarly, an output may be produced via different activities, possibly drawing from different (substitutable) inputs.

Let x be a nonnegative p-vector of processing activity rates. Next, let A_{kj} be the amount of resource k capacity consumed per unit of activity j, and let K_k be the available capacity during the period of resource k. The $l \times p$ capacity consumption matrix A is nonnegative. Let S be the available input stock vector (after ordering), and let D be the demand vector for output materials.

The last set of the data for the formulation are financials. Let p-vectors r and c be the revenue and processing cost rates associated with the various processing activities; $r - c$ can thus be thought of as gross margins. For the processing resources, we assume linear[1] capacity costs:

- c_K: per unit capacity investment cost.

For the inventories, the usual data as defined by Porteus (1990), as cost at the beginning of period apply:

- c_S: per unit order cost for input materials,
- c_H: per unit effective holding cost assessed against any leftover input stock at the end of the period. (If each unit of leftover stock has a salvage value of c_L, then the effective holding cost is the actual holding cost c_{Ha} less c_L.)

In newsvendor networks, insufficient inputs or capacity may lead to output shortages, which may inflict a penalty (e.g., loss of customer goodwill) that is captured, as usual, by

- c_P: per unit output shortage penalty cost assessed on any unfilled demand $(D - R_D x)^+$. (The shortage cost may differ when shortages result in lost sales versus backlogs.)

We can now summarize the formulation of a newsvendor network:

DEFINITION 1. A newsvendor network is defined by three data sets:

(1) Demand data: the probabilistic demand forecast represented by measure P over demand space.

(2) Financial data: gross margins $r - c$, capacity in-

---

[1]Economies of scale, modeled via concave affine capacity costs resulting from the inclusion of fixed cost components, can relatively easily be accomodated as described in an unabridged version of this paper available from the authors.

vestment cost c_K, inventory costs c_S and c_H, and demand (output) shortage cost c_P.

(3) Network data: input-output matrices R_S and R_D and capacity consumption matrix A.

Notation. We will use matrix notation, such that vectors are assumed to be arranged in columns and primes denote transposes. Expectation, with respect to the demand distribution P, is denoted by $\mathbb{E}$. (For simplicity we will assume that P has a continuous distribution function, although results directly extend to mixed and discrete distributions by using subgradients.) The terms decreasing and increasing are used in the weak sense throughout this paper.

Objective. The system's manager objective is to maximize the expected firm value by choosing capacity K and inventory S before demand is known, and activity x afterwards. The activity vector x maximizes operating profit by transforming $R_S x$ of input stock into output $R_D x$. Let $\Pi(K, S)$ denote the expected maximal operating profit, which is net value from processing minus the shortage penalty cost and holding cost:

$$\Pi(K, S) = \mathbb{E} \max_{x \in X(K,S,D)} [(r - c)'x - c_P'(D - R_D x)^+$$
$$- c_H'(S - R_S x)^+], \qquad (1)$$

where the set of feasible activities are constrained by supply S, demand D, and capacity K:

$$X(K, S, D) = \{x \geq 0 : R_S x \leq S, R_D x \leq D, Ax \leq K\}. \qquad (2)$$

The expected firm value to be maximized then is

$$V(K, S) = \Pi(K, S) - c_S'S - c_K'K. \qquad (3)$$

2.1. Newsvendor Network Properties and Solution

Given the set of feasible activities (2), the expected maximal operating profit (1) equals

$$\Pi(K, S) = \mathbb{E} \max_{x \in X(K,S,D)} [(r - c)'x - c_P'(D - R_D x)$$
$$- c_H'(S - R_S x)]$$
$$= \mathbb{E}\pi(K, S, D) - c_P'\mathbb{E}D - c_H'S,$$

where the optimal effective operating profit is

$$\pi(K, S, D) = \max_{x \in X(K,S,D)} v'x, \qquad (4)$$

and v is the *net value vector* associated with the various processing activities:

$$v = r - c + R_D'c_P + R_S'c_H. \qquad (5)$$

The net value vector thus includes not only the revenue r and marginal processing cost c of each activity, but also the mitigating impact each activity has on total demand shortage penalty and inventory holding costs. The value vector is assumed to be positive to make each ex-post processing activity economically viable. (If v_j were negative, the revenue generated by activity j would not outweigh its cost. If it were zero, we are indifferent and may as well not do it.)

The properties and solution of newsvendor networks are analogous to those of the multidimensional newsvendor model. They build on linear programming theory and are presented as Proposition 1 (see Harrison and Van Mieghem 1999 for a proof[2]):

PROPOSITION 1. *The optimal effective operating profit $\pi(K, S, D)$, the expected operating profit $\Pi(K, S)$ and the firm value $V(K, S)$ are jointly concave. There exists an Ω-partition of the demand space such that the gradients of Π and V simplify to:*

$$\nabla_K V(K, S) = \nabla_K \Pi(K, S) - c_K,$$

where

$$\nabla_K \Pi(K, S) = \mathbb{E}\nabla_K \pi(K, S, D) = \sum_j \lambda_{K,j} P_j,$$

$$\nabla_S V(K, S) = \nabla_S \Pi(K, S) - c_S,$$

where

$$\nabla_S \Pi(K, S) = \mathbb{E}\nabla_S \pi(K, S, D) = \sum_j \lambda_{S,j} P_j - c_H,$$

where P_j is shorthand for $P(\Omega_j(K, S, D))$ and $\lambda_{k,j}$ and $\lambda_{S,j}$ are the constant Lagrange multipliers of the capacity and inventory constraints, respectively, in Ω_j.

Clearly, the optimal inventory and capacity vectors K^* and S^* solve $\nabla_K V(K^*, S^*) = 0$ and $\nabla_S V(K^*, S^*) = 0$, and the optimal inventory levels S^* are, in general,

[2]An unabridged version of this paper presents a shorter proof based on the monotone convergence theorem.

capacity dependent. Similar to traditional newsvendor results, these sufficient first-order conditions specify the optimal balance between overages and underages. Indeed, any of the dual variables is non-zero only if the optimal activity $x^*(K, S, D)$ is either capacity or supply constrained. Each component of the vector sum thus denotes the expected "underage cost" of a particular resource or input, while the component of the marginal cost c_K or $c_H + c_S$ denotes its "overage cost." The underage cost measures the risk of output shortages due to capacity or input shortages. The overage costs measures the risk of leftover (unused) capacity or inputs. (Notice, that output decisions are made ex-post so that there is no risk of leftover outputs.) The optimal coupled balance between underages and overages of all resources and inputs thus uniquely specifies the optimal probabilities P_j^*, which can be interpreted as "generalized critical fractiles" of the multivariate demand distribution. The determination of these critical fractiles P_j^*, however, is much more difficult than for the single-dimensional newsvendor model that can always be solved in closed form.

For small problems or if the linear program exhibits special structure such that a greedy allocation policy is optimal (see Federgruen and Groenevelt 1986 for necessary and sufficient conditions), the linear program and the regions Ω_j can be solved in closed form for any value of the parameters K and S and a realization of D. Otherwise, the newsvendor network solution can effectively be solved numerically with a steepest-ascent method using the gradient of V as follows. Draw a large set of sample demand vectors, and keep these fixed. Assume we have an initial estimate $(K^{(0)}, S^{(0)})$, and set $i = 0$. Now iterate as follows: Given capacity $K^{(i)}$ and supply $S^{(i)}$, solve the linear program and associated dual variables $\lambda(K^{(i)}, S^{(i)}, D^{(j)})$ numerically for each sample demand vector $D^{(j)}$. Take the average of the $\lambda(K^{(i)}, S^{(i)}, D^{(j)})$ over all j as an unbiased estimate of $\mathbb{E}\lambda(K^{(i)}, S^{(i)})$, and use it to compute an unbiased (given that differentiation and integration interchange) estimate of $\nabla V(K^{(i)}, S^{(i)})$. If $|\nabla V(K^{(i)}, S^{(i)})|$ is smaller than some tolerance level, stop: $(K^{(i)}, S^{(i)})$ is close to the optimal vector (K^*, S^*). Otherwise, adjust capacity and inventory in the di-

Table 1 Comparing Newsvendor Networks and the Multidimensional Newsvendor Model of Harrison and Van Mieghem (1999)

	Multidimensional Newsvendor Model	Newsvendor Network Model
Ex-post (Recourse) Problem:	$\pi(K, D) = \max\limits_{x \geq 0} v'x$ s.t. $Ax \leq K,\quad x \leq D.$	$\pi(K, S, D) = \max\limits_{x > 0} v'x$ s.t. $Ax \leq K,\quad R_D x \leq D,$ $R_S x \leq S.$
Ex-ante Problem:	$\max\limits_{K \geq 0} \mathbb{E}[\pi(K, D)] - c_K'K$	$\max\limits_{K,S \geq 0} \mathbb{E}[\pi(K, S, D)]$ $- [c_K'K + (c_S + c_H)'S]$
"Critical Fractile" Solution:	$\sum\limits_j \lambda_{K,j} P_j = c_K$	$\sum\limits_j \lambda_{K,j} P_j = c_K,$ $\sum\limits_j \lambda_{S,j} P_j = c_S + c_H.$

rection of the gradient: $K^{(i+1)} = K^{(i)} + \xi \nabla_K V(K^{(i)}, S^{(i)})$, and similarly for $S^{(i)}$, where ξ is some step-size (or perform a line-search), and iterate. This is nothing more than optimization through simulation, also called infinitesimal perturbation analysis (IPA). Obviously, if the dual variables $\lambda(K, S)$ are known analytically, the iteration is vastly accelerated because the linear programs do not need to be resolved. This numerical optimization through simulation is easily implemented in Excel for virtually any demand distribution (or forecast) which enhances the practical and didactical value of newsvendor networks.

The preceding presentation has closely followed those of the multidimensional newsvendor model provided in Harrison and Van Mieghem (1999) and Van Mieghem (1998a). As illustrated in Table 1, newsvendor networks extend the multidimensional newsvendor model in a natural way by including inventories and enriching the processing formulation via activities and more general demand constraints. Further, newsvendor networks feature a richer activity set that allows for nonbasic activities and inventories that connect periods in a dynamic setting. (In multidimensional newsvendor models, which are essentially single-period, inventories and capacities are indistinguishable. Their fundamental difference only arises in a dynamic setting where inventories change on a much smaller time-scale than capacities, as is elaborated on in Van Mieghem 2002a.) The remainder

Figure 2   Four Illustrative Newsvendor Networks
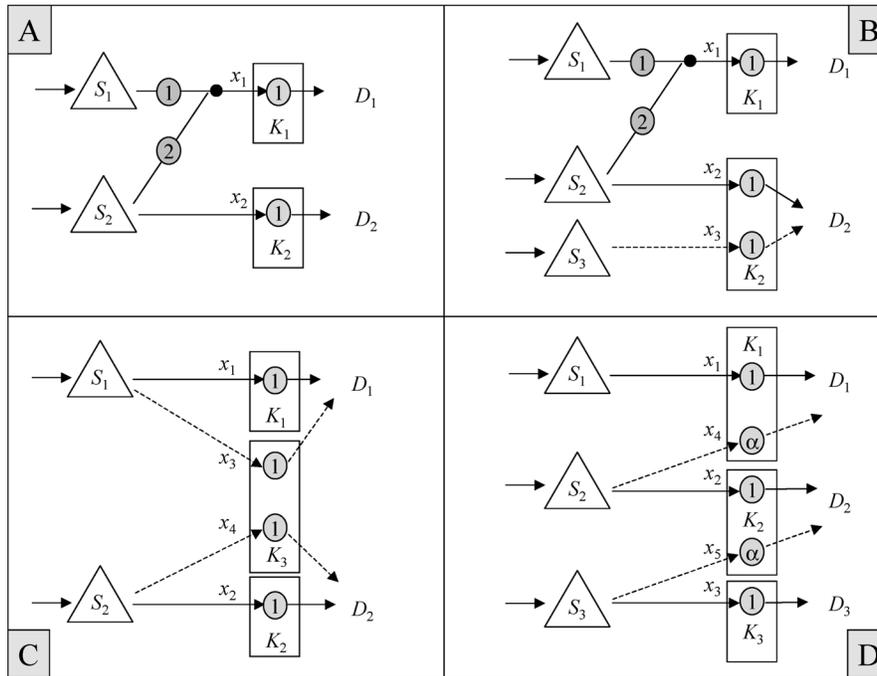
Note. A features an assembly activity and input commonality; B adds nonbasic (in dotted line) input substitution; C has a flexible resource; and D enjoys resource pooling.

of this section discusses the richer processing formulation and additional properties and sets the groundwork for the extension to the multiperiod setting of §4.

2.2.   Modeling Processing: Basic vs. Nonbasic Activities

The processing formulation in newsvendor networks allows for ex-post redeployment of inputs and resources to best respond to resolved uncertain events. Redeployment implies a deviation from the normal (or planned) regime or operating point. Loosely speaking, utilized activities in the normal regime are called basic activities, while nonbasic activities perform redeployment.

Formally, it is useful to define *basic activities* as follows. For a given demand or ''operating point'' D^*, consider the relaxed linear program $\max_{x \in X^*} v'x$, where $X^* = X(\infty, \infty, D^*) = \{x \geq 0 : R_D x \leq D^*\}$. If this relaxed problem has a unique optimal solution x^*, then we call the activities j, such that $x_j^* > 0$, ''basic

activities'' and the other activities ''nonbasic'' or ''discretionary.'' Thus, basic activities are those that satisfy the demand D^* in the most profitable manner *if* the system is not supply or capacity constrained. The corresponding process that only uses basic activities and the minimal cost inputs and capacities S^* and K^* such that $X^* = X(K^*, S^*, D^*)$ is called the *basic process.* It is the process that would be optimal if there were no uncertainty. The value of nonbasic activities then reflects the discretionary flexibility to meet stochastic demand deviations from the operating point. Nonbasic activities thus provide a redundancy that is only valuable in the presence of uncertainty.

Figure 2 illustrates four newsvendor networks, three of which feature nonbasic activities, which are drawn with a dotted line. Network A features simple assembly and commonality: Activity 1 depletes two inputs simultaneously in fixed proportions: $R_{S,11} = 1$ and $R_{S,21} = 2$, while Input 2 is a joint input for Products 1 and 2. Notice that simple assembly and com-

monality do not involve discretionary activities. Both activities are basic and the common Input 2 exhibits traditional inventory pooling. Network B adds a nonbasic input-substitution Activity 3 to Network A (assuming $v_2 > v_3$). Network C illustrates a flexible Resource 3 and two nonbasic activities (assuming $v_1 > v_3$ and $v_2 > v_4$). Network D has two processing resources with limited flexibility. Its two nonbasic activities (assuming $v_1 > v_4$ and $v_2 > v_5$) generate resource pooling: Input 2 can be processed on Resource 1 or on Resource 2 and generates dynamic pooling among the inputs.

Given that discretionary activities enlarge the choice set of activities, it is obvious that:

PROPOSITION 2. *The presence or addition of discretionary activities increases optimal expected network value.*

Indeed, having more processing options can never be worse because one can choose not to use them. Related interesting questions are: Under which conditions does the presence of discretionary activities lead to a difference between inventories and capacities? Similarly, when does the presence of common inputs or resources lead to lower inventories or capacities and higher inventories or capacities of noncommon inventories or resources to exploit ex-post flexibility? Typically, the answers to these questions will depend on the three data sets (demand, financial, and network data) of the newsvendor network. Section 3 will illustrate some of these answers.

The rich processing formulation allows newsvendor networks to unify various elements in the existing literature on newsvendor-type decisions. Indeed, by appropriately structuring the capacity consumption matrix A and the input-output matrices R_S and R_D various classical problems are recovered. For example:

(1) A network with independent products implies $R_S = R_D = I$, where I is the identity matrix. If, in addition, there are no capacity constraints ($A = 0$), then all decisions are decoupled per product: $x = \min(S_i, D_i)$. Thus, $\pi(S, D) = \Sigma_i v_i \min(S_i, D_i)$ and $\lambda_{S,i} = v_i \mathbf{1}\{D_i > S_i\}$, where $\mathbf{1}\{\cdot\}$ is the indicator operator whose expectation yields the probability so that: $\mathbb{E}\lambda_{S,i} = v_i \mathbb{E}\mathbf{1}\{D_i > S_i\} = v_i \mathbb{P}\{D_i > S_i\}$. The solution thus reduces to m independent critical fractiles: $v_i\mathbb{P}\{D_i > S_i\} = c_{H,i} - c_{S,i}$.

(2) When activities correspond one-to-one to output production quantities, then $R_D = I$ and R_S represent the familiar ''bill-of-materials.''

(3) Assembly activities (i.e., when an output requires multiple inputs) fill the corresponding column of R_S with multiple positive entries.

(4) Component commonality of an input (i.e., when multiple outputs require a common input) fills the corresponding row of R_S with multiple positive entries; c.f. Gerchak and Henig (1989), Tayur (1995), and Van Mieghem (2002b), who also establish an equivalence with resource flexibility.

(5) Input substitution and/or transshipment: Row j of R_D has multiple positive entries when output j normally uses input i with basic activity k but can substitute i with an equal amount of i' using nonbasic activity k' ($R_{D,jk} = R_{D,jk'}$ and $R_{S,ik} = R_{S,i'k'}$). With ''one-level downward substitution,'' $R_D = (I, I)$ and $R_S = (I\ B)$, where B has zeros except for the one-off diagonal. For examples, see Bassok et al. (1999) and Netessine et al. (2002) for substitution and Krishnan and Rao (1965) and Robinson (1990) for transshipment.

(6) Resource flexibility: Multiple positive entries in row i of A mean that resource i can perform multiple activities. For examples, see Rudi and Zheng (1997) and Van Mieghem (1998a) who study a setting where two products can be produced either on product-dedicated resources or on one flexible resource, which is captured by

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \text{ and}$$

$$R_D = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}.$$

(7) Simultaneous resource requirements: Multiple positive entries in column j of A mean that activity j requires the simultaneous use of multiple resources.

Finally, aside from network optimization, newsvendor networks can also be used for network design. For example, the structure of newsvendor network D exhibits ''chaining'' in the sense introduced and stud-

ied by Jordan and Graves (1995). Their objective was to study the higher-level decision of network design, while this paper focuses on optimal control of a given network design. In particular, the extremely important decision that Jordan and Graves (1995) address so beautifully can be phrased using our terminology as follows. If one wants to add ''process flexibility'' to a basic network by adding nonbasic activities (i.e., by adding links as the dotted lines in our Figure 2), which nonbasic activities have the highest impact on expected profits or shortages? Thus, the input-output matrix R, which defines the activity set or routing structure, becomes a decision variable. Recent work by Graves and Tomlin (2001) studies this network design question in a multistage setting.

2.3. The Effect of Demand Uncertainty
In many newsvendor networks, one is also interested in the effect of demand uncertainty on optimal inventory and capacity decisions, as well as on the optimal value. An increase in the mean vector of the demand distribution is easy because it affects only the basic activities by a similar increase and yields a corresponding increase in the optimal value and the inputs and capacities necessary to support the increase of those basic activities. To analyze the comparative statics of other demand parameters, such as standard deviations or pairwise correlations, the traditional approach is to use Jacobians and the implicit function theorem. Unfortunately, such analysis quickly becomes involved (e.g., Netessine et al. 2002, Van Mieghem 1998a) and often has to be numerically evaluated for specific parameter values so that one may as well directly vary demand parameters in the optimization through simulation. Sometimes, however, special analytic structure allows one to draw rather general conclusions for the optimal value[3] by drawing on Müller's (2001) recent work on stochastic orders. (Corbett and Rajaram 2001 give other useful stochastic order results.)

PROPOSITION 3. *Assume D is normally distributed with mean vector μ and covariance matrix Σ. The optimal value V is increasing in μ and decreasing in any variance term*

[3]Comparative statics on the optimal inventory and capacity levels, however, are much more difficult to establish.

Σ_{ii}. *In addition, if the operating profit $\pi(K, S, D)$ is submodular in D, then the optimal value V is decreasing in any covariance term Σ_{ij} (and thus pairwise demand correlation), whether K and S are held constant or are adjusted optimally.*
(All proofs are relegated to the Appendix.)

Recall that a function f is submodular in D if the marginal returns of D_i are decreasing in D_j. This is a condition that holds in uncapacitated transportation problems whose optimal profit $\pi(S, D)$ is submodular in $(-S, D)$ (Topkis 1998, Theorem 3.4.1): As other products consume higher levels of available inputs and capacity, the return of an additional unit of demand for product i typically does not increase; similarly, returns are supermodular in supply S. Hence, the demands of any two products are substitutes, the supplies of any two inputs are substitutes, and the demand of any product and the supply of any input are complements. The transportation problem is a special case of our input-output activity problem (4), where activities correspond to an arc between one supply node and one demand node. Unfortunately, the proof by Topkis does not address capacity constraints on arcs, let alone joint capacity constraints of the type considered here. Our formulation also allows one activity to simultaneously deplete multiple inputs or generate multiple outputs.

Although our formulation thus is more general than a transportation problem, economic intuition suggests that the profit be submodular in D for most systems. To verify submodularity in our setting, one must verify whether the (sub)gradient $v'\partial x/\partial D_i$, which is constant in each domain Ω_j, is decreasing over the sequence of the domains that are traversed when one increases D_j for any given D_i. For stylized problems this is easy to verify. For example, this holds for the flexible system studied in Van Mieghem (1998a) and for our example, as will be shown in the next section.

2.4. Economical Service Levels
By definition, the firm's *processing capacity region* is the demand region where all demand can be met. It is the feasible region of the linear program (4), denoted by Ω_0 for now. Baker et al. (1986) define the *aggregate*

service level as the probability of meeting all demand, $P(\Omega_0)$. They further point out that while ''the concept of service level is fundamental to single-product inventory analysis, it is not obvious how to generalize that concept to the multiproduct situation.'' Indeed, for cases where aggregate demand exceeds total capacity, the service level of a particular product depends on how available inputs and capacity are rationed and allocated to outputs. Baker et al. suggest that ''in general the optimal rationing policy is a function of the service level measure.'' Not only is the optimal rationing policy endogenous in any newsvendor-based framework, but so is the *type* as well as the *value* of service. Indeed, by capturing financial data, the newsvendor network solution has a clear objective and automatically specifies the economically optimal service levels and rationing policy. This overcomes a major weakness of using the probability of meeting all demand as an exogenous service measure in multi-item systems. (To illustrate this, consider n independent newsvendor problems, each with a probability θ of meeting its demand. The probability of meeting all demand is then θ^n, which not only depends on the size of the problem n but approaches zero as n gets sufficiently large.) The different setting of Baker et al. also leads to a more subtle difference. In their setting, where they minimize overall inventory subject to an exogenous service measure, they conclude that the pooling effect of commonality leads to lower overall inventory. Such general conclusions cannot be made for newsvendor networks. It is well known that, depending on financial data, pooled inventory in newsvendor-type systems can be smaller, equal to, or larger than the aggregate inventory when managing each product's inputs individually. As an example, consider the classic paper by Eppen (1979) that considers the effect of centralization (i.e., merging multiple independent newsvendors into a single newsvendor) under a multivariate normal distribution. While centralization always increases the optimal expected profit, it leads to (i) increased total inventory when the ''newsvendor fractile'' is less than half, (ii) no change in inventory when the fractile is equal to one half, and (iii) decreased total inventory when the fractile exceeds one half. In summary, for

normally distributed demands, the pooling considered by Eppen makes the total inventory gravitate towards the mean.

2.5. Initial Input Stock z

For the extension to the dynamic setting, it is useful to consider the case where we start with an initial level z of input inventory before ordering. Let y denote the input inventory level after ordering and denote the value function excluding capacity investment costs by

$$g(K, y) = \Pi(K, y) - c'_S y. \tag{6}$$

Restricting attention to the inventory decisions, the objective function now becomes

$$G(K, y; z) = \max_{y \geq z} \{\Pi(K, y) - c'_S(y - z)\}$$
$$= \max_{y \geq z} \{g(K,y) + c'_S z\}.$$

If we do order up to y, the value is $G(K, y; x)$; if we don't order, the value is $G(K, z; z)$. Hence, we order iff

$$G(K, y; z) - G(K, z; z) = g(K, y) - g(K, z) > 0.$$

Given that S^* is an unconstrained optimizer of the concave function $g(K, \cdot)$, it is optimal to order up to $y = S^*$ for any vector $z \leq S^*$. (If any $z_i > S^*$, the optimal policy is more complicated, but we are not concerned about such transient initial conditions.) This means that each input component is managed via a base-stock policy iff $z < S^*$. It also means that the optimal value function is affine in its starting states, for $z < S^*$:

$$G^*(K, z) = g(K, S^*) + c'_S z,$$

and its gradient is c_S. These results are similar to the conventional, one-dimensional newsvendor model.

2.6. Incorporating Random Yield

In many applications, there might not be a 100% yield of the resources. (See Yano and Lee 1995 for a review of research on random yield and Hsu and Bassok 1999 for an example of a newsvendor network with random yield.) Input stocks might have defects, while capacities might not be fully available due to maintenance or employee absence. Let $U = (U_S, U_K)$ be a diagonal matrix where $U_{S,ii}$ and $U_{K,ii}$ are the random yield of input stocks i and the random fraction of time

that resource j capacity is available. For the case where U is realized at the same time as D, the constraint set (2) can easily be adjusted to reflect the random yield U as follows:

$$X(K, S, D, U)$$

$$= \{x \geq 0 : R_S x \leq U_S S, R_D x \leq D, Ax \leq U_K K\},$$

and (1) is adjusted by taking the expectation over both D and U. It follows that Proposition 1 still holds.

3. An Example with Discretionary Commonality

This section illustrates the typical properties of a newsvendor network and the insights that it generates by considering the example presented in the introduction. The matrices that define the feasible activity set X in (2) are:

$$R_S = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \qquad R_D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix},$$

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & \alpha^{-1} \end{pmatrix}.$$

3.1. Optimal Activities, Inventories, and Capacities

Assume now that $v_1 > v_2 \geq v_3 > 0$ and $0 \leq \alpha \leq 1$. For this (or any other) ranking, the linear program (4) is easily solved for the optimal activity: Prioritize Activities 1 and 2, and let Activity 3 take care of the residual demand of Product 2, provided there is sufficient capacity and input. Hence, the parametric solution of the linear program is:

$$x_1(K, S, D) = \min\{D_1, K_1, S_1\},$$

$$x_2(K, S, D) = \min\{D_2, K_2, S_2\},$$

$$x_3(K, S, D) = \min\{D_2 - x_2, \alpha(K_2 - x_2), S_1 - x_1\}$$

$$= \min\{[D_2 - \min\{K_2, S_2\}]^+,$$

$$\alpha[K_2 - \min\{D_2, S_2\}]^+,$$

$$[S_1 - \min\{D_1, K_1\}]^+\}.$$

Clearly, without supply or capacity constraints and for any chosen operating point D^* we have that $x_1^* =$

Figure 3   **Ex-post Activity Vector x and Total Output $q = (x_1, x_2 + x_3)$ Depends on Supply S, Capacity K, and Demand D**
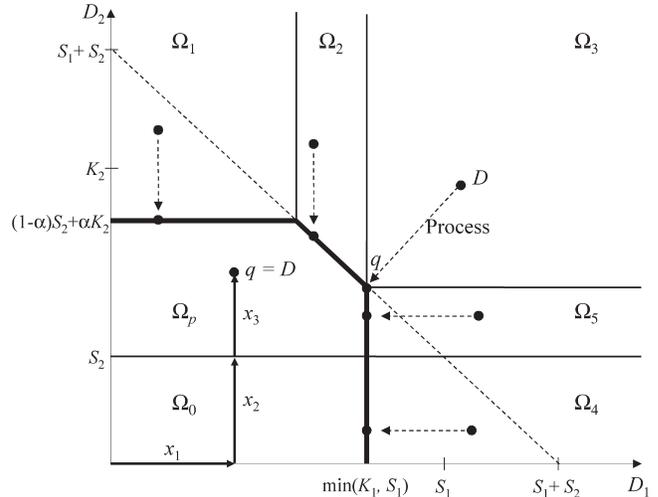
Note. Graph assumes $K_1 < S_1$, otherwise $\Omega_5 = \varnothing$ and the capacity K_1 constraint is nonbinding.

D_1^*, $x_2^* = D_2^*$, and $x_3^* = 0$. Therefore, Activities 1 and 2 are basic, while Activity 3 is nonbasic. These activity decisions and the aggregate output vector $q = (x_1, x_2 + x_3)$ can be represented graphically in the demand space, shown in Figure 3. The demand space is partitioned into seven domains: $\mathbb{R}_+^2 = \cup_i \Omega_i$, and Table 2 gives the optimal activity and dual variables in each domain. As defined earlier, the firm's processing region here is the thick-lined $\Omega_0 \cup \Omega_p$ where all demand can be met through the basic activities (x_1, x_2) and the nonbasic activity x_3. The newsvendor network framework shows that the aggregate service level as defined by Baker et al. (1986) should be measured by the probability of the rectangle with cut-off at the upper right corner: $SL_{agg} = P(\Omega_0 \cup \Omega_p)$, which is a function of K and S. It also shows how trade-offs between individual service levels should be measured: The individual service levels, as measured by $SL_1(K, S) = 1 - P(\Omega_3 \cup \Omega_4 \cup \Omega_5)$ and $SL_2(K, S) = 1 - P(\Omega_1 \cup \Omega_2 \cup \Omega_3)$ satisfy the trade-off, for arbitrary K and S:

$$SL_1 + SL_2 = SL_{agg} + 1 - P(\Omega_3).$$

Notice that for the basic process the domains Ω_p, Ω_2, and Ω_5 would be empty. Hence, intuitively, the sub-

stitution-flexibility option embedded in the presence of the nonbasic activity x_3 will increase in value as these domains ''cover more demand,'' that is, as $P(\Omega_p)$, $P(\Omega_2)$, and $P(\Omega_5)$ increase. Clearly, a necessary condition is that there is some extra capacity and some flexibility: $K_2 - S_2 > 0$ and $\alpha > 0$.

For any capacity vector K, it is suboptimal to procure more Input 2 than can ever be used, so that $S_2^* \leq K_2$. What is more interesting, and perhaps surprising at first, given the substitution option is that an identical result holds for the discretionary common Input 1.

RESULT 1. If $c_{S,1} + c_{H,1} > c_{S,2} + c_{H,2}$, then, for any K, optimal sourcing sets $S_1^* \leq K_1$ and $S_2^* \leq K_2$.

This result becomes intuitive by considering the counter case that $\exists \epsilon > 0$ such that $S_1 > K_1 + \epsilon$. Hence, at least ϵ of Input 1 is held only for Output 2 production. Clearly, under the assumptions, one can do better by reducing S_1 by ϵ and increasing S_2 by ϵ: We save in procurement by $(c_{S,1} - c_{S,2})\epsilon \geq 0$; we may gain an operating margin $(v_2 - v_3)\epsilon \geq 0$ while consuming

equal or less of K_2 capacity. (A formal proof can be found in the Appendix.) Given Result 1, the optimal sourcing conditions $E\lambda_S = c_S$ simplify to:

$$v_3 P(\Omega_2(K, S^*)) + v_1 P(\Omega_{3+4}(K, S^*)) = c_{S,1} + c_{H,1} + \mu_1^S,$$

$$(v_2 - v_3)P(\Omega_p(K, S^*)) + (v_2 - \alpha v_3)P(\Omega_1(K, S^*))$$
$$+ v_2 P(\Omega_{2+3}(K, S^*)) = c_{S,2} + c_{H,2},$$

where $\mu_1^S = 0$ if $S_1^* < K_1$ and is positive if $S_1^* = K_1$. This defines the unique optimal sourcing vector $S^*(K)$ for any K. (Note that without capacity constraints, $P_1 = 0$ and $\mu_1^S = 0$.) Now we can strengthen the bounds on the optimal S^* and K^*. Clearly, it is suboptimal to invest in more capacity than can ever be used given the input stocks S. Hence, $S_1^* = K_1^*$ and $S_2^* \leq K_2^* \leq S_2^* + \alpha^{-1}S_1^*$, and the optimality conditions become:

RESULT 2. If $c_{S,1} + c_{H,1} > c_{S,2} + c_{H,2}$, then optimal sourcing and investment sets: $S_1^* = K_1^*$ and $S_2^* \leq K_2^* \leq S_2^* + \alpha^{-1}S_1^*$ and the value function and its gradient (and thus optimality equations) simplify:

$$V(K_1, K_2, S_2) = v_1 \mathbb{E} \min\{D_1, K_1\} + v_2 \mathbb{E} \min\{D_2, S_2\} + v_3 \mathbb{E} \min\{[D_2 - S_2]^+, \alpha[K_2 - \min\{D_2, S_2\}], [K_1 - D_1]^+\}$$

$$- c_P' \mathbb{E}D - (c_{S,1} + c_{H,1} + c_{K,1})K_1 - c_{K,2}K_2 - (c_{S,2} + c_{H,2})S_2,$$

$$\nabla V(K_1, K_2, S_2) = \begin{pmatrix} v_3 P(\Omega_2(K, S)) + v_1 P(\Omega_{3+4}(K, S)) - (c_{S,1} + c_{H,1} + c_{K,1}) \\ \alpha v_3 P(\Omega_1(K, S)) - c_{K,2} \\ (v_2 - v_3)P(\Omega_p(K, S)) + (v_2 - \alpha v_3)P(\Omega_1(K, S)) + v_2 P(\Omega_{2+3}(K, S)) - c_{S,2} - c_{H,2} \end{pmatrix}.$$

3.2. Discussion and Insights on Discretionary Commonality

Recall the earlier questions of interest: Under which conditions do the presence of discretionary activities lead to a difference between inventories and capacities? Similarly, when does the presence of common inputs or resources lead to lower inventories or capacities and higher inventories or capacities of noncommon inventories or resources to exploit ex-post flexibility? This example illustrates the answers and the type of insights newsvendor networks may generate.

Impact of Discretionary Commonality (Substitutive Flexibility Option). An important question is how the

presence of the substitutive flexibility option changes decisions compared to the basic system (i.e., a system without the substitution Activity 3). Let (K^b, S^b) denote the optimal solution for the basic system, which satisfies simple critical fractiles: $v_i P(D_i > K_i^b = S_i^b) = c_{K,i} + c_{S,i} + c_{H,i}$. Comparing these with the generalized critical fractile solutions directly shows that:

RESULT 3. Compared to a basic system, discretionary commonality (substitutive flexibility) warrants a *higher* inventory of the expensive discretionary common Input 1 and higher Capacity 1; a *lower* inventory of the cheaper unique Input 2; and a higher flexible Capacity 2 to be able to exercise on the substitution option:

Table 2 The Optimal Activity Vector and Marginal Values of Inputs and Capacities in Each Demand Domain for the Example

Domain	Activity Vector x		Marginal Value λ_S		Marginal Value λ_K	
	$S_1 > K_1$	$S_1 \leq K_1$	$S_1 > K_1$	$S_1 \leq K_1$	$S_1 > K_1$	$S_1 \leq K_1$
Ω_0	$(D_1, D_2, 0)$	$(D_1, D_2, 0)$	$(0, 0)$	$(0, 0)$	$(0, 0)$	$(0, 0)$
Ω_p	$(D_1, S_2, D_2 - S_2)$	$(D_1, S_2, D_2 - S_2)$	$(0, v_2 - v_3)$	$(0, v_2 - v_3)$	$(0, 0)$	$(0, 0)$
Ω_1	$(D_1, S_2, \alpha(K_2 - S_2))$	$(D_1, S_2, \alpha(K_2 - S_2))$	$(0, v_2 - \alpha v_3)$	$(0, v_2 - \alpha v_3)$	$(0, \alpha v_3)$	$(0, \alpha v_3)$
Ω_2	$(D_1, S_2, S_1 - D_1)$	$(D_1, S_2, S_1 - D_1)$	(v_3, v_2)	(v_3, v_2)	$(0, 0)$	$(0, 0)$
Ω_3	$(K_1, S_2, S_1, - K_1)$	$(S_1, S_2, 0)$	(v_3, v_2)	(v_1, v_2)	$(v_1 - v_3, 0)$	$(v_1, 0)$
Ω_4	$(K_1, D_2, 0)$	$(S_1, D_2, 0)$	$(0, 0)$	$(v_1, 0)$	$(v_1, 0)$	$(v_1, 0)$
Ω_5	$(K_1, S_2, D_2 - S_2)$	—	$(0, v_2 - v_3)$	—	$(v_1, 0)$	—

$$K_1^b = S_1^b \leq K_1^* = S_1^* \quad \text{and}$$

$$S_2^* \leq S_2^b = K_2^b \leq K_2^* \leq S_2^* + \alpha^{-1} S_1^*.$$

Thus the procurement savings from substitution derive from holding less of the less costly input rather than more of the more costly input. (Clearly, this all assumes—quite optimistically—that the flexible capacity has the same investment cost as a dedicated resource: $c_{K,2}^b = c_{K,2}$. In reality, one would expect $c_{K,2}^b < c_{K,2}$, and if the discount is sufficiently high, the substitutive option may become worthless.) The example directly shows how the presence of the substitutive discretionary activity leads to an essential difference between inventory and capacity levels: $S_2^* \leq K_2^*$, an effect that can only be captured if both inventories and capacities are modeled. It is the discretionary commonality and the flexible resource that lead to higher K_1^*, S_1^*, and K_2^*, but lower S_2^*, compared to the basic system. And both commonality and flexibility are necessary to produce the effects and higher profitability: This real option is worthless without the simultaneous presence of the discretionary Activity 3, processing flexibility ($\alpha > 0$), and financial flexibility ($v_3 > 0$).

The effect of discretionary commonality in newsvendor networks is different from the effect of commonality in inputs in the insightful article by Baker et al. (1986). They showed that "commonality permits a given service level to be attained with a smaller amount of safety stock than would be attainable without commonality" and, more interestingly, that inventory of the common component decreases while those of unique components increase, again with a constraint on service level. In contrast, our example shows that inventory of Input 1 (which is a discretionary common component) actually increases, while inventory of the Input 2 (which is unique to Product 2) decreases, while increasing overall expected profit. Our effect is thus very different from that in Baker et al. due to the inherent difference in model setup.[4] Similar to the centralization benefit in Eppen (1979) mentioned earlier, total inventories (and thus safety stock) $S_1^* + S_2^*$ can be smaller, equal to, or larger than $S_1^b + S_2^b$, depending on the financial data. This is a comparison of economically optimal stocking levels whose corresponding economic optimal service levels may differ between the basic network and the network with discretionary commonality. Also, discretionary commonality leads to weaker pooling effects than "simple" or "ordinary commonality." In our example, the "base case" is that both products are processed each from their unique inputs. Only if Demand 2 is "much higher than expected" while Demand 1 is lower (loose language for $D \in \Omega_p$) will Input 1 be a common component for both products. The graphical representation is also useful to estimate some nonobvious comparative statics related to demand uncertainty. For example:

Impact of Demand Correlation and Variability. Assume for concreteness that D is multivariate normal with mean vector $\mathbb{E}D$ and covariance matrix

---

[4]Newsvendor networks automatically lead to economically optimal service levels, for example, inventory levels that maximize expected profits. Therefore, we do not need service-level constraints in our analysis, as Baker et al. (1986) did because they did not capture financials.

$$\Sigma = \gamma \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

where $\gamma \geq 0$ is a measure for the amount of variability and $-1 \leq \rho \leq 1$ is the correlation coefficient. Now set $\gamma = 1$, and fix a value for ρ. It is easily verified[5] that $\pi(K, S, D)$ is submodular in D, so that Proposition 3 yields that the value of the system decreases in correlation ρ. Establishing general comparative statics of the optimal inventory and capacity levels, however, is more difficult. The following discussion is meant to illustrate how one can build intuition and appreciation for the complexity of the impact of various demand parameters from a graphical representation. For stylized problems as in Van Mieghem (1998b), the graphical approach is very effective in the classroom: It typically gets to the key effects quickly. Nevertheless, this reasoning is rather intuitive and case specific.

Assume that the stock and capacity levels shown in Figure 4 are optimal for the normal distribution with those parameter values and thick isoplot. Now consider a small increase in correlation to the dotted isoplot, *ceteris paribus.* A plausible reasoning may go as follows, which illustrates what may happen if correlation increases for the situation as shown in Figure 4. Given that the optimality conditions must remain to hold, one can estimate the change in the stock and capacity levels to counteract the change in the probability of the domains. To counteract the increase of P_2, increase $S_1^* + S_2^*$; to counteract the small decrease in P_{3+4}, decrease $K_1^* = S_1^*$. Hence, S_2^* must increase, which indeed counteracts the decrease in P_4. Finally, to counteract the decrease in P_1, the weighted average of S_2^* and K_2^* must decrease. Given that S_2^* increases, K_2^* must decrease. Thus, as correlation increases for the situation given in Figure 4, we expect Input 1 stock to decrease and Input 2 stock to increase while

Figure 4 **Superimposed Isoplots of the Demand Density to Estimate the Impact of Increasing Correlation and Decreasing Uncertainty on Optimal Stock and Capacity Levels**
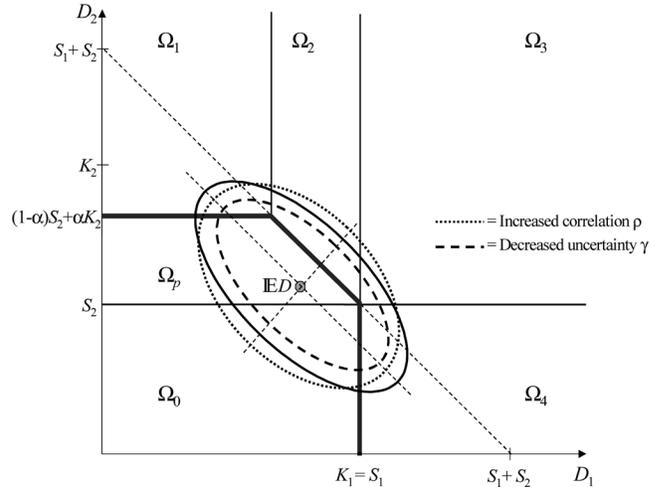
total stock levels increase. At the same time, we expect *both* capacity levels to decrease, which reflects the decreasing option value imbedded in the substitutive flexibility. Given that S_2^* increases while K_2^* decreases, there may exist a threshold value $\bar{\rho} \leq 1$ at which $S_2^* = K_2^*$ (and thus $P_2 = P_p = 0$ and x_3 are always zero). The above trends are indeed observed in numerical studies. Thus, $\bar{\rho}$ is the maximal correlation for the substitution option to be valuable. Beyond $\bar{\rho}$ the dedicated solution is optimal. (Clearly, $\bar{\rho}$ is a function of the value and cost parameters that can be analytically studied. Similar to the flexible and commonality systems of Van Mieghem 1998a and 2002b, there may be instances such that the substitution option remains valuable—and thus $S_2^* > K_2^*$—even for perfectly positive correlation. This highlights the revenue maximization benefit imbedded in discretionary activities, in addition to traditional risk pooling.)

Similarly, consider any other amount of variability $\gamma \neq 1$. Rescaling the demand space directly yields that the optimal values of $(1 - \alpha)S_2 + \alpha K_2 - \mathbb{E}D_2$, $K_1 - \mathbb{E}D_1$ and $S_1 + S_2 - \mathbb{E}(D_1 + D_2)$ are also scaled by γ, generalizing the critical fractile scaling of the one-dimensional newsboy. Hence, as variability γ decreases to zero, $K_1^* = S_1^* \to \mathbb{E}D_1$, $(1 - \alpha)S_2^* + \alpha K_2^* \to \mathbb{E}D_2$

---

[5]For example, to show that the (sub)gradient $\partial\pi/\partial D_1 = v'\partial x/\partial D_1$ is decreasing in D_2, one must consider three scenarios: (1) if $D_1 < S_1 + \alpha(S_2 + K_2)$, then, as D_2 increases from 0, $v'\partial x/\partial D_1$ remains constant at v_1 throughout Ω_0, Ω_p, and Ω_1; (2) if $S_1 + \alpha(_2 + K_2) < D_1 < \min(K_1, S_1)$, then, as D_2 increases from 0, $v'\partial x/\partial D_1$ remains constant at v_1 throughout Ω_0 and Ω_p, and decreases to $v_1 - v_3$ in Ω_2; (3) finally, if $D_1 > \min(K_1, S_1)$, $v'\partial x/\partial D_1$ remains constant at 0 throughout Ω_4, Ω_5, and Ω_3.

and $S_1^* + S_2^* \to \mathbb{E}(D_1 + D_2)$, so that S_2^* and K_2^* both approach $\mathbb{E}D_2$. Thus, P_p and P_2 both decrease to zero as does the value of the nonbasic activity.

Impact of Processing Flexibility α. Recall that $0 \leq \alpha \leq 1$ captures the degree of processing flexibility of Resource 2. From the second optimality equation, it follows that an increase in flexibility decreases P_1 and the effective investment cost $c_{K,2}/\alpha$ for Resource 2 so that K_2^* will tend to increase. At the same time, while the effective procurement cost of Input 2, $c_{S,2} + \alpha v_3 P_1 = c_{S,2} + c_{K,2}$, remains constant, P_{p+2+3} must increase to counterbalance the decrease in P_1. Hence, we expect the individual Input 2 stock level to decrease, increasing P_p, P_2, and P_3. To counterbalance the increase in P_2 and P_3, the discretionary common input S_1^* and $K_1^* = S_1^*$ must increase. Reversing the argument, as α decreases, K_2^* decreases while S_2^* increases so that there exists a threshold value $\underline{\alpha}$ under which $S_2^* = K_2^*$ and $P_2 = 0$, and x_3 is always zero. Thus, similar to $\bar{\rho}$, $\underline{\alpha}$ is the minimum amount of processing flexibility that is necessary to make the substitution option worthwhile. Clearly, total value is increasing in α as seen from the expression of V.

4. Dynamic Optimality of the Base-Stock Policy

This section extends the newsvendor network to a dynamic (multiperiod) setting. Similar to the single-item inventory model (Porteus 1990, p.628), there now are a sequence of discrete periods in which demands, denoted by $\{D_t : t > 0\}$, occur. There is a single-capacity investment decision K at the beginning of Period 1 and K remains in effect ever after. Other timing follows the standard inventory setup: At the beginning of each period, stock levels are reviewed and an order is made for the current period. Any order is received in time to satisfy any demand in that period. Then demand for that period is observed, after which production decisions for that period are made. Demands in different periods are independent and identically distributed according to P. Revenues and costs are discounted using the discount factor δ, where $0 < \delta \leq 1$. The marginal costs c_S, c_K, c_P and the actual holding cost c_{Ha} remain as before. (Even though these costs may be incurred at the end of a period, they are expressed in beginning-of-period monetary units.) At the end of the last period in the time horizon, each unit of leftover input stock has a value of c_S. Otherwise, leftover stock at the end of one period is the initial inventory for the following period. We can now summarize the formulation of a (dynamic) newsvendor network:

DEFINITION 2. A dynamic newsvendor network is defined by three data sets of the newsvendor network, augmented by:

(1) Demand data: demand $\{D_t : t > 0\}$ is i.i.d. with measure P; treatment of demand shortages (lost, backlogged, or a combination).

(2) Financial data: discount factor δ.

The key result is that the optimal dynamic policy is myopic when shortages result in lost sales; that is, it then equals the optimal stationary base-stock policy for the single-period model with holding and investment cost parameters adjusted for discounting. When shortages are backlogged, some restrictions must be imposed for the optimal policy to remain a myopic stationary base-stock policy.

4.1. Dynamic Optimality with Lost Sales
This section considers the case when unmet demand results in lost sales.

PROPOSITION 4. *A stationary base-stock inventory policy with level S* is optimal for any finite horizon problem.*

Similar traditional extensions to the single-item case also hold: If $\delta < 1$, the policy is also optimal for the infinite horizon problem. The policy is also optimal if the discounted cost is replaced by the average-cost criterion. Karlin (1960) and Veinott's (1965) results for nonstationary independent demands also hold: Compute S_t^* with the parameters for period t. If $S_t^* \leq S_{t+1}^*$ (which is the case if demands increase stochastically over time and financial parameters are stationary), then everything works out and using myopic base-stock in each period is an optimal strategy.

Finally, returning to the original i.i.d. setting, the optimal capacity level is also derived from the one-period problem, provided we use the effective single-period capacity cost $c_K(1 - \delta)/(1 - \delta^T)$. Indeed, the optimal net present value of investing in capacity K

and following the inventory base-stock S^* policy during T periods, starting with zero initial inventory, is

$$V_1(K, 0) = \frac{1 - \delta^T}{1 - \delta}(\Pi(S^*, K) - c_S'S^*) - c_K'K.$$

The optimal capacity level that maximizes this concave function is as before, provided we adjust the capacity cost. (The special choice of a salvage value of c_S eliminates any end-of-horizon effects and makes the optimal base-stock policy stationary. Its cost in each period is i.i.d., and its optimal value is the discounted sum of T single-period values.)

4.2. Dynamic Optimality with Backlogging

Complications arise when, instead of being lost, unmet demand is backlogged to be filled in the future. First, we must keep track of the backorders for end units. As usual, this is most easily accomplished by establishing a cumulative backorder counter as an inventory level z_D for outputs that must be added to the state descriptor of the system. In true newsvendor networks, $z_D \leq 0$ with negative values signifying the backlog $b = -z_D$.

Second, in traditional inventory systems one can map backlogs of outputs into certain demand for inputs upstream. Adding to that the stochastic component of next period's demand leads to the marvelous concept of echelon inventory, which in essence allows a reduction of the state space. With nonbasic activities, however, this is not possible in general because the flexibility of the nonbasic activities makes it impossible to ex-ante designate inputs to fill known backlogs. Indeed, when an input is consumed by both a basic and a nonbasic activity, it may be ex-post optimal to prioritize the nonbasic activity over the basic one. Consider our example: Assume there is a backlog b_1 for Output 1 and no capacity constraints. Traditionally, one would map the backlog into an increased demand for Input 1 and order $S_1 + b_1$, with the intent to use the deterministic component b_1 to fill the backlog next period. However, it may be better ex-post to allocate some portion of b_1, which was intended to fill the Output 1 backlog, to Output 2 if that generates higher profits, for example, if $v_3 > v_1$. (Notice that Activity 3 remains nonbasic if $v_2 > v_3 > v_1$.) Or, sticking with the earlier parameter values $v_1 > v_2 \geq v_3$, a

high backlog b_1 will make the expected usage of the discretionary Activity 3 very small and may lead to an increase of the optimal order-up-to level S_2 for Input 2.

These complications illustrate why the myopic policy may no longer be dynamically optimal for newsvendor networks under backlogging. In restricted classes of newsvendor networks, however, a myopic solution remains dynamically optimal with backlogging. As in standard inventory systems, it is natural to consider whether base-stock inventory control is optimal. An alternative way to think about an output backorder b is that the effective demand is the sum of b and the stochastic single-period demand D. Thus, the expected operating profit becomes

$$\Pi(S, K; b) = \mathbb{E}\pi(S, K, D + b),$$

and the optimal order-up-to input levels are a function of the backlog b. If this function is *linear* in the backlog, then one can separate the deterministic backlog component b in the effective demand from the stochastic component D. If, in addition, we can translate backorders for outputs into backorders for inputs, the backorder can then be accounted for as usual with a negative input inventory and a base-stock policy for those input inventories remains optimal. Provided we add some restrictions, the myopic policy then remains dynamically optimal with backlogging. It is known that this holds in networks without discretionary activities and capacity constraints; in terms of our primitives:

PROPOSITION 5. *In an uncapacitated newsvendor network without discretionary activities, a stationary policy is optimal under backlogging: Let z be the input stock on-hand just before ordering, and let b be the output backlog; it is optimal to order $S^* - z + R_S R_D^{-1} b$.*

Thus, a base stock S^* policy remains optimal on the quantity $x - R_S R_D^{-1} b$, which is like an "echelon-like" inventory because it includes both on-hand input inventory and a backlog from outputs.

Notice that the presence of a joint capacity constraint may prevent the optimality of such a simple policy. Consider, for example, a simple newsvendor network with two activities (and no discretionary activities) and one simple capacity constraint $x_1 + x_2 \leq$

K. The demand for inputs then becomes $R_S x^*(D + b, K)$, where $x^*(D + b, K) = \arg \max_x \{v'x : 0 \leq x \leq R_D^{-1}(D + b)$ and $x_1 + x_2 \leq K\}$. Now assume a large backlog b so that the capacity constraint is binding for almost any D. It then would be optimal to stock only one input (the one with the highest component in $R_S v - c_S$). The point is that a joint capacity constraint may introduce nonlinearity so that a simple echelon policy would no longer be optimal.

Can we say anything at all for newsvendor networks with discretionary activities? Yes, but only for networks with a restricted type of discretionary activities, which we will simply call *strong nonbasic* activities and define as follows. A network has strong nonbasic activities if:

(1) All basic activities remain basic for any operating point $D^* > 0$.

(2) Basic activities do not share inputs.

(3) Basic activities dominate nonbasic activities ex-ante and ex-post. That is, for any known $D^* > 0$, it is optimal ex-ante to procure only those inputs that are needed by the basic activities and, ex-post, it is optimal to have basic activities first consume that stock before nonbasic activities.

In terms of our model primitives, this requires that:

(1) The number of strong basic activities must equal the number of outputs m. If we label those activities from 1 to m, then R_D can be decomposed as $R_D = [B, N]$, where the $m \times m$ matrix B forms a basis and is invertible. Equivalently, R_D has a left inverse: $R_D^+ = [B^{-1}, 0]'$. This basis remains unchanged for any $D^* > 0$ in the linear program $\max_{x \in X^*} v'x$, where $X^* = \{x \geq 0 : R_D x \leq D^*\}$. Hence, with ample supply, the optimal activity vector x^* for any output demand D^* is $x^* = [x_B^*, x_N^*]$ with basic activities $x_B^* = B^{-1}D^*$ and nonbasic activities $x_N^* = 0$, and would consume the input quantity $R_S x^*$.

(2) Let i_j denote the number of inputs that are depleted by strong basic activity j, and let $i = \Sigma_j i_j$. If we label the inputs that are depleted by the first strong-basic activities by $1, 2, \ldots, i_1$, followed by those depleted by the second and so on, then R_S can be decomposed as

$$R_S = \begin{pmatrix} R_b & R_{n1} \\ 0 & R_{n2} \end{pmatrix}$$

where the $i \times m$ matrix R_B is block-diagonal and block j is $i_j \times 1$.

(3) For any $D^* > 0$, strong basic activities remain the only basic variables in the following two linear programs—(i) ex-ante: $\max_{(x, y \in Y^*)} v'x - c_S'y$, where $Y^* = \{x, y \geq 0 : R_S x \leq y, R_D x \leq D^*\}$ and (ii) ex-post: $\max_{x \in Z^*} v'x$, where $Z^* = \{x \geq 0 : R_S x \leq R_S R_D^+ D^*\}$.

Newsvendor networks with strong nonbasic activities are a nontrivial and a useful subset of networks that can model problems with substitution, flexibility, and transshipment. For example, it is easy to verify that in our example newsvendor network of Figure 1, Activities 1 and 2 are strong basic, while Activity 3 is strong nonbasic, if the natural conditions $v_1 - c_{S,1} \geq v_2 - c_{S,2}$ and $v_1 \geq v_2 \geq v_3$ hold. Matrix B becomes the identity matrix, while $N = (0, 1)'$. Compared to basic activities, strong basic activities thus have two additional properties: First, they remain the only activities to be used for a deterministic problem of procuring stock to fill a known demand. Second, they always dominate nonbasic activities. These two properties are exactly what we need to separate deterministic backlog from stochastic demand: They allow us to ex-ante map output backlog b into unique input requirements, while a weakly greedy ex-post allocation is optimal—First satisfy the backlog, then fill the stochastic demand with basic activities, and finally fill the remaining stochastic demand with nonbasic activities. (The first two allocations via basic activities are separable for each output, while the third allocation using nonbasic activities is not separable and need not be greedy.)

PROPOSITION 6. *In an uncapacitated newsvendor network with strong nonbasic activities, a stationary policy is optimal under backlogging: Let z be the input stock on-hand just before ordering, and let b be the output backlog; it is optimal to order* $S^* - z + R_S R_D^+ b$.

Thus, a base stock S^* policy remains optimal provided we consider "echelon-like" inventory $z - R_S R_D^+ b$. Given that $R_D^+ = [B^{-1}, 0]'$, it follows that only the i inputs that are drawn by strong basic activities have backlog-adjusted, echelon-like inventories.

5. Concluding Remarks

This paper has introduced a rather broad class of capacitated processing networks with single-stage inventories. A key feature of such newsvendor networks is that they allow for ex-post managerial discretion through nonbasic activities. Such activities can capture various subtle inventory- and resource-pooling effects. We presented single-period optimality conditions and showed that they retain their optimality in a dynamic setting, so that a stationary base-stock policy is optimal. As such, this paper directly extends the results in a variety of prior papers that focus either on inventory or on capacity in a multiperiod setting with both inventories and capacities. This paper has provided a first thrust of analysis of how multiple end products become linked due to the presence of static or discretionary sharing of inventories or capacities. Future work should continue exploring the implications of these intricate links.

Our intent was to present a network model involving both capacity, inventory, and discretionary activity decisions that retains many of the features of the traditional newsvendor model, yet is quite a bit more general. The restrictions we impose on newsvendor networks were made to maintain parsimony and tractability. Our discussion also gave a first hint of the limits of this tractability. As in most inventory settings, lost sales are more tractable in newsvendor networks than backlogging. Our discussion suggests that the culprits are discretionary activities or joint ex-post capacity constraints, both of which make the order-up-to levels of inputs dependent on the backlog in a nonlinear manner so that simple echelon stocks are no longer optimal. (They are, however, for a restricted class of newsvendor models as shown.) Backlogging inherits its difficulty because it is very closely related to a multistage inventory model that allows for holding output inventories, in addition to input inventories. Such networks are notoriously hard to analyze. This echoes what is well known for multiechelon inventory systems with a distribution structure (e.g., one warehouse serving multiple retailers must decide how much to ship to each retailer and how much to hold back at the warehouse for later allocation). This distribution problem is a dynamic newsvendor network problem with discretionary activities and backlogging for which an optimal policy is still unknown after Clark and Scarf (1960) pointed it out several decades ago.

By enriching newsvendor networks with output inventories, leadtimes, or setup costs, tractability will suffer. For example, the well-known concept of "echelon inventories" does not readily extend to networks with discretionary activities that are not strongly greedy. The problem is that one can no longer ex-ante map output backlogs into needed input stock. Also, as in typical inventory models, lost sales and positive leadtimes would be a deadly combination. Setup costs and distributive networks also are hard.

In fact, discretionary activities in newsvendor networks, which focus on inventory and production to meet exogenous output demand, create the same difficulties as dynamic routing in queuing networks, which focus on production and the input buffers from arrivals. Concepts developed for queuing networks may very well be useful for newsvendor networks. For example, the accounting problem of backlog may be approached through an "equivalent workload formulation" (Harrison and Van Mieghem 1997), which is the minimal state descriptor needed to account for inventory in discretionary networks in an appropriately scaled asymptotic regime. Future work that investigates a series of scaled newsvendor networks may provide fruitful insights and simplifications similar to heavy-traffic queuing networks. Nevertheless, when the frontier of analytic tractability is reached, one has no choice but to adopt approximating network control problems (i.e., simplify the network flows using fluid or Brownian approximations and find optimal controls for that simplified network in an appropriately scaled asymptotic regime) or to restrict the policy space ex-ante (i.e., restrict the analysis to base-stock policies, for example).

Finally, as mentioned in the introduction, we introduced newsvendor networks as being controlled by a single decision maker. Clearly, following recent trends in supply chain theory, multiple decision makers can directly be incorporated into a game-theoretic formulation. Newsvendor networks then become a useful tool to study subcontracting (as in Van Mieghem

1999) and other contingent relationships, including pricing, in the supply chain.

Acknowledgments

Appendix. Proofs

PROOF OF PROPOSITION 3. This proof follows directly from Müller (2001): Let D' also be normally distributed with mean μ' and a covariance matrix Σ'. If $\mu \leq \mu'$ and $\Sigma = \Sigma'$, then $D \leq D'$ in the increasing concave order so that $\mathbb{E}f(D) \leq \mathbb{E}f(D')$ for any increasing concave function, which includes $\pi(K, S, \cdot)$. If $\mu = \mu'$ and Σ' differs from Σ in only one variance term ($\Sigma'_{ii} \geq \Sigma_{ii}$), then $M = \Sigma' - \Sigma$ is a zero matrix except for $M_{ii} > 0$, so that M is positive and semi-definite. Hence, $D \leq D'$ in the convex order, so that $\mathbb{E}f(D) \leq \mathbb{E}f(D')$ for any convex function, which includes $-\pi(K, S, \cdot)$. Finally, if $\mu = \mu'$ and Σ' differs from Σ in only one covariance term ($\Sigma'_{ij} \geq \Sigma_{ij}$), then $D' \geq D$ in the supermodular order, so that $\mathbb{E}f(D^1) \geq \mathbb{E}f(D^2)$ for any supermodular function. If K and S remain fixed, the proposition follows directly by considering $f = -\pi$. If they are optimally adjusted, $\partial\Pi/\partial K_k = \partial\Pi/\partial S_l = 0$ for all k and l, so that $d\Pi/d\Sigma_{ij} = \partial\Pi/\partial\Sigma_{ij}$ and the ordering remains. $\square$

PROOF OF RESULT 1. It is obvious that $S_2 > K_2$ is suboptimal. Assume on the contrary that $S_1 > K_1$. Consider the directional derivative $d\Pi$ in direction $dS = (-1, 1)d\epsilon$, where $d\epsilon > 0$. With $S_2 \leq K_2$ and $\alpha \leq 1$, the marginal values λ_S of Table 1 apply:

$$2^{1/2}d\Pi = (E\lambda_S - c_S)\cdot(-1, 1)d\epsilon$$

$$2^{1/2}d\Pi/d\epsilon = (v_2 - v_3)P_{p+2+3+5} + (v_2 - \alpha v_3)P_1$$

$$+ (c_{S,1} + c_{H,1}) - (c_{S,2} + c_{H,2}) > 0,$$

so that $S_1 > K_1$ cannot be optimal. $\square$

PROOF OF PROPOSITION 4. By induction, let $V_t(K, z)$ denote the expected present value starting and evaluated at the beginning of period t with initial input inventory z and capacity K. Define the set of structured functions V^* as the set of continuous functions f that are concave and affine in the starting state; for example, $f(K, z) = f(K S^*) - c'_S(S^* - z)$ for $z \leq S^*$, where S^* is the base-stock level defined earlier and the minimizer of $g(K, \cdot)$, defined in (6). Let Δ^* be the set of decision rules that bring the input inventory level after ordering up to S^* if $z \leq S^*$, and order nothing otherwise. Recall that with lost sales, $z \geq 0$. As before, we assume the starting state $z \leq S^*$. (If some components $z_i > S^*_i$, there is a transient policy that is more complicated but eventually will bring $z \leq S^*$.)

For the last period $T + 1$, we clearly have that the optimal value function $V_{T+1}(K, z) = c'_S z$, which is structured. Now assume that $V_{t+1}(K, z) \in V^*$, we show that V_t, which solves the Belmann equations, is also structured:

$$V_t(K, z) = \max_{y \geq z}\{\mathbb{E} \max_{x \in X(K,y,D)} [(r - c)'x - c'_P(D - R_D x)$$

$$- c'_{Ha}(y - R_S x) - c'_S(y - z)$$

$$+ \delta V_{t+1}(K, y - R_S x)]\}$$

$$= \max_{y \geq x} G_t(K, y).$$

Concavity preservation under maximization says that the function behind the expectation operator, and hence its expectation $G_t(K, y)$, is a concave function. Another application of the concavity preservation theorem directly yields that $V_t(K, z)$ is concave. We now show that V_t is affine with slope c_S for $z \leq S^*$ and that it is optimized by a base-stock policy.

Consider any $y \leq S^*$. Then, for any $x \in X(K, y, D)$, we have that $y - R_S x \leq y \leq S^*$, thus

$$G_t(K, y)$$

$$= \mathbb{E} \max_{x \in X(K,y,D)} [(r - c)'x - c'_P(D - R_D x)$$

$$- c'_{Ha}(y - R_S x) - c'_S(y - z)$$

$$+ \delta V_{t+1}(K, S^*) - \delta c'_S(S^* - y + R_S x)]$$

$$= \mathbb{E} \max_{x \in X(K,y,D)} [(r - c)'x - c'_P(D - R_D x)$$

$$- (c'_{Ha} - \delta c'_S)(y - R_S x) - c'_S(y - z)$$

$$+ \delta V_{t+1}(K, S^*) - \delta c'_S S^*]$$

$$= \Pi(K, y) - c'_S(y - z) + \delta V_{t+1}(K, S^*) - \delta c'_S S^*$$

$$= g(y, K) + c'_S z + \delta V_{t+1}(K, S^*) - \delta c'_S S^*.$$

Clearly, $y^* = S^*$ is a maximizer of $G_t(K, y)$ for $y \leq S^*$. Given that $\nabla G_t(K, S^*) = 0$ and G_t is concave, S^* is also a global maximizer of $G_t(K, y)$. Thus, again the optimal policy is a base-stock S^* policy: $y = S^*$ if $z \leq S^*$. (If some components $z_i > S^*_i$, there is a transient policy that is more complicated but eventually will bring $z \leq S^*$.) The optimal value function is not only concave again, but also structured:

$$V_t(K, z) = \Pi(K, S^*) - c'_S(S^* - z) + \delta V_{t+1}(K, S^*) - \delta c'_S S^*$$

$$\text{if } z \leq S^*. \quad \square$$

PROOF OF PROPOSITION 5. If there are no discretionary activities, all activities are basic in the relaxed linear program $\max_{x \in X^*} v'x$, where $X^* = X(\infty, \infty, D^*) = \{x \geq 0 : R_D x \leq D^*\}$ for any operating point D^*. Hence, linear algebra shows that the number of activities must equal the number of outputs m and the matrix R_D forms a basis and is invertible: The optimal solution is $R_D x^* = D^*$ or $x^* = R_D^{-1}D^*$. Thus, without capacity constraints, an effective output demand $D^* = D + b$ yields a demand for inputs $R_S x^* = R_S R_D^{-1}(D + b)$. Essentially, the multi-item inventory system decouples and each input can be analyzed independently; the base-stock S^* reorder policy remains optimal for the stochastic part, to which we add the deterministic order $R_S R_D^{-1}b$. $\square$

PROOF OF PROPOSITION 6. First, in an uncapacitated network with strong nonbasic activities, the optimal activity vector $x(S, D)$ satisfies for any backlog vector $b \geq 0$ and $S, D \geq 0$:

$$x(S + R_S R_D^+ b, D + b) = x(S, D) + x(R_S R_D^+ b, b).$$

To see this, consider the ex-post problem of allocating demand $D + b$ to available supply. If supply were not constrained, demand $D + b$ would be allocated optimally via the strong basic activities to the supply $R_S R_D^+ b$. Given that basic activities dominate nonbasic activities and that no basic activities share inputs, the feasible allocation of the backlog b to the available supply $R_S R_D^+ b$ is optimal for any S and D. Hence, it only remains to allocate the remaining demand D to the remaining supply S, which is by definition achieved through activity $x(S, D)$. Thus, with strong basic activities, one can again ex-ante translate output backorders into a deterministic input requirement that ex-post also will be used to fill the backlog. In essence, using the designed supply $R_S R_D^+ b$, the backlog problem can be separated from the stochastic problem: The expected operating profit given a backlog b, $\Pi(S + R_S R_D^+ b; b) = \mathbb{E}\pi(S + R_S R_D^+ b, D + b)$, is linear in the backlog: $\Pi(S + R_S R_D^+ b; b) = \Pi(S) + \Pi(R_S R_D^+ b)$. The backlog profit is deterministic and the stochastic problem retains its dynamic optimality. $\square$

References

Angelus, A., E.L. Porteus. 2002. Simultaneous capacity and production management of short-life-cycle, produce-to-stock goods under stochastic demand. *Management Sci.* **48**(3) 399–413.

Anupindi, R., Y. Bassok, E. Zemel. 2001. A general framework for the study of decentralized distribution systems. *Manufacturing and Service Oper. Management* **3**(4) 349–368.

——, S. Chopra, S. Deshmukh, J.A. Van Mieghem, E. Zemel. 1999. *Managing Business Process Flows.* Prentice Hall, Upper Saddle River, NJ.

Baker, K. R., M. J. Magazine, H. L. W. Nuttle. 1986. The effect of commonality on safety stock in a simple inventory model. *Management Sci.* **32**(8) 982–988.

Bassok, Y., R. Anupindi, R. Akella. 1999. Single-period multiproduct inventory models with substitution. *Oper. Res.* **47**(4) 632–642.

Chen, F., Y.S. Zheng. 1994. Lower bounds for multi-echelon stochastic inventory systems. *Management Sci.* **40**(11) 1426–1443.

Clark, A. J., H. Scarf. 1960. Optimal policies for a multi-echelon inventory problem. *Management Sci.* **6**(4) 475–490.

Corbett, C. J., K. Rajaram. 2001. Stochastic dominance, multivariate dependence and aggregation of uncertainty. Working paper, UCLA.

Eberly, J. C., J. A. Van Mieghem. 1997. Multi-factor dynamic investment under uncertainty. *J. Econom. Theory* **75**(2) 345–387.

Eppen, G. D. 1979. Effects of centralization on expected costs on multi-location newsboy problem. *Management Sci.* **25** 498–501.

Federgruen, A., H. Groenevelt. 1986. The greedy procedure for resource allocation problems: Necessary and sufficient conditions for optimality. *Oper. Res.* **34**(6) 909–918.

——, P. Zipkin. 1986a–b. An inventory model with limited production capacity and uncertain demands, I: The average-cost criterion, II: The discounted-cost criterion. *Math. Oper. Res.* **11**(2) 193–215.

Gerchak, Y., M. Henig. 1989. Component commonality in assemble-to-order system: Models and properties. *Naval Res. Logist.* **36** 61–68.

Glasserman, P., S. Tayur. 1995. Sensitivity analysis for base-stock levels in multiechelon production-inventory systems. *Management Sci.* **41**(2) 263–280.

Graves, S. C., B. T. Tomlin. 2001. Process flexibility in supply chains. Working paper, MIT, Cambridge, MA.

Harrison, J. M. 2001a. A broader view of brownian networks. Forthsoming, *Ann. Appl. Probability.*

——. 2001b. Stochastic networks and activity analysis. Y. Suhov, ed. *Analytic Methods in Applied Probability.* Forthcoming, American Mathematical Society, Providence, RI.

——, J.A. Van Mieghem. 1997. Dynamic control of Brownian networks: State space collapse and equivalent workload formulations. *Ann. Appl. Probability* **7**(3) 747–771.

——, ——. 1999. Multi-resource investment strategies: Operational hedging under demand uncertainty. *Eur. J. Oper. Res.* **113** 17–29.

Hsu, A., Y. Bassok. 1999. Random yield and random demand in a production system with downward substitution. *Oper. Res.* **47**(2) 277–290.

Jordan, W. C., S. C. Graves. 1995. Principles on the benefits of manufacturing process flexibility. *Management Sci.* **41**(4) 577–594.

Kapuscinski, R., S. Tayur. 1998. Optimal policies and simulation-based optimization for multi-stage capacitated production inventory systems. S. Tayur, ed. *Quantitative Methods for Supply Chain Management.* Kluwer Academic Publishers, Boston, MA.

Karlin, S. 1960. Dynamic inventory policy with varying stochastic demands. *Management Sci.* **6** 231–258.

Khouja, M. 1999. The single-period (news-vendor) problem: Literature review and suggestions for future research. *Omega Internat. J. Management Sci.* **27** 537–553.

Krishnan, K., V. Rao. 1965. Inventory control in N warehouses. *J. Indust. Engrg.* **XVI** 212–215.

Lippman, S. A., K. F. McCardle. 1997. The competitive newsboy. *Oper. Res.* **45**(1) 54–65.

Müller, A. 2001. Stochastic ordering of multivariate normal distributions. *Ann. Inst. Statist. Math.* **53** 567–575.

Netessine, S., N. Rudi. 2002. Centralized and competitive inventory models with demand substitution. Forthcoming, *Oper. Res.*

——, R. Dobson, R. Shumsky. 2002. Flexible service capacity: Optimal investment and the impact of demand correlation. *Oper. Res.* **50**(2) 375–388.

Parlar, M. 1988. Game theoretic analysis of the substitutable product inventory problem with random demands. *Naval Res. Logist.* **35** 397–409.

Porteus, E. L. 1990. Stochastic inventory theory. D.P. Heyman and M.J. Sobel, eds. *Handbooks in OR & MS*, vol. 2. Elsevier Science Publishers, North-Holland, Amsterdam, The Netherlands.

Robinson, L.W. 1990. Optimal and approximate policies in multi-

period, multilocation inventory models with transshipments. *Oper. Res.* **38** 278–295.

Rudi, N., Y. S. Zheng. 1997. Multi-item newsboy model with partial variety postponement. Presented at the *INFORMS Dallas Meeting* (October).

——, S. Kapur, D.F. Pyke. 2001. A two-location inventory model with transshipment and local decision making. *Management Sci.* **47**(12) 1668–1680.

Tayur, S. 1995. Computing optimal stock levels for common components in an assembly system. Working paper, Carnegie Mellon University, Pittsburgh, PA.

Topkis, D. M. 1998. *Supermodularity and Complementarity.* Princeton University Press, Princeton, NJ.

Van Mieghem, J. A. 1998a. Investment strategies for flexible resources. *Management Sci.* **44**(8) 1071–1078.

——. 1998b. Seagate Technologies: Operational hedging. Case study, Kellogg School of Management, Northwestern University, Evanston, IL.

——. 1999. Coordinating investment, production, and subcontracting. *Management Sci.* **45**(7) 954–971.

——. 2002a. Capacity portfolio investment and hedging: Review and new directions. Forthcoming, *Manufacturing and Service Oper. Management.*

——. 2002b. Component commonality strategies: Value drivers and equivalence with flexible capacity strategies. Working paper, Kellogg School of Management, Northwestern University, Evanston, IL.

Veinott, A., Jr. 1965. Optimal policy for a multi-product, dynamic non-stationary inventory problem. *Management Sci.* **12** 206–222.

Yano, C.A., H.L. Lee. 1995. Lot sizing with random yields: A review. *Oper. Res.* **43**(2) 311–334.

Zipkin, P.H. 2000. *Foundations of Inventory Management.* McGraw-Hill, Boston, MA.