Operations, Information and Decisions Papers

Wharton Faculty Research

3-2002

Flexible Service Capacity: Optimal Investment and the Impact of Demand Correlation

Serguei Netessine
University of Pennsylvania

Gregory Dobson

Robert A. Shumsky

Flexible Service Capacity: Optimal Investment and the Impact of Demand Correlation

Abstract

We consider a firm that provides multiple services using both specialized and flexible capacity. The problem is formulated as a two-stage, single-period stochastic program. The firm invests in capacity before the actual demand is known and optimally assigns capacity to customers when demand is realized. Sample applications include a car rental company's use of mid-sized cars to satisfy unexpectedly high demand for compact cars and an airline's use of business-class seats to satisfy economy-class demand. We obtain an analytical solution for a particular case, when services may be upgraded by one class. The simple form of the solution allows us to compare the optimal capacities explicitly with a solution that does not anticipate flexibility. Given that demand follows a multivariate normal distribution, we analytically characterize the effects of increasing demand correlation on the optimal solution. For the case with two customer classes, the effects of demand correlation are intuitive: Increasing correlation induces a shift from flexible to dedicated capacity. When there are three or more classes, there are also adjustments to the resources not directly affected by the correlation change. As correlation rises, these changes follow an alternating pattern (for example, if the optimal capacity of one resource rises, then the optimal capacity of the adjacent resource falls). These results make precise conjectures based on numerical experiments that have existed in the literature for some time.

Keywords

Disciplines

Other Business | Other Economics

Flexible Service Capacity:

Optimal Investment and the Impact of Demand Correlation

Serguei Netessine

The Wharton School

of Business Administration

University of Pennsylvania

Philadelphia, PA 19104

netessin@wharton.upenn.edu

Gregory Dobson

William E. Simon Graduate School of

Business Administration

University of Rochester,

Rochester, NY 14627

dobson@simon.rochester.edu

Robert A. Shumsky[*]

William E. Simon Graduate School of

Business Administration

University of Rochester,

Rochester, NY 14627

shumsky@simon.rochester.edu

.

September, 2000

[*] Corresponding author. Phone (716) 275-7956, FAX (716) 273-1140

Flexible Service Capacity:
Optimal Investment and the Impact of Demand Correlation

Abstract: We consider a firm that provides multiple services using both specialized and flexible capacity. The problem is formulated as a two-stage single-period stochastic program. The firm invests in capacity before the actual demand is known and optimally assigns capacity to customers when demand is realized. Sample applications include a car-rental company's use of mid-sized cars to satisfy unexpectedly high demand for compact cars and an airline's use of business-class seats to satisfy economy-class demand. We obtain an analytical solution for a particular case, when services may be upgraded by one class. The simple form of the solution allows us to compare the optimal capacities explicitly with a solution that does not anticipate flexibility. Given that demand follows a multivariate Normal distribution, we analytically characterize the effects of increasing demand correlation on the optimal solution. For the case with two customer classes, the effects of demand correlation are intuitive: increasing correlation induces a shift from flexible to dedicated capacity. When there are three or more classes, there are also adjustments to the resources not directly affected by the correlation change. As correlation rises, these changes follow an alternating pattern (for example, if the optimal capacity of one resource rises, then the optimal capacity of the adjacent resource falls). These results make precise conjectures based on numerical experiments that have existed in the literature for some time.

1. Introduction

Consider a telecommunications company that must construct daily service assignments for its hardware technicians and must also, in the long run, determine the number of technicians to hire and train. There are three major functions that the technicians perform. They are, in order of increasing complexity, equipment installation, testing, and repair. Technicians from the repair department are able to handle the other two jobs, those in the testing department can also accomplish installation, and members of the installation department are not yet cross-trained in any other job. When constructing the daily schedule, if the company runs out of technicians in one department, it may try to borrow a specialist with expertise at the next highest level. Due to union rules and employee preferences, borrowing from a department with skills *two* levels higher may not be feasible. In the long run, this limited flexibility may be taken into account when specifying the number of technicians needed to staff each department.

This paper considers such environments, in which customers may be upgraded to a higher level of service at no cost to the customer. We consider the short-term assignment problem, when actual demand is known, as well as the long-term capacity decision in the face of uncertain demand. This model applies to problems in which customers may not be aware of the upgrade (as in the hardware technician example, above), as well as service environments in which the customer experiences the upgrade directly. Examples include commercial aviation (economy to business-class upgrades), time-shared executive jets (the use of larger, faster jets to substitute for smaller, slower aircraft, as described by Keskinocak, 1999) and the car rental industry (customers who request a compact car may be offered a mid-sized car).

We formulate the problem as a one-period model (in Section 2, after describing the formulation, we will discuss the limitations of such a model for real-world service applications). The firm must commit to resource capacities at the beginning of the period, when demand is

1

uncertain but a distribution is known. During the period the firm must assign available resources to realized demand. We find, below, that significant benefits can be gained if the possibility of service substitution is accounted for at the time of capacity planning, rather than only at the time of service delivery. We also find that the correlation structure of demand has a significant impact on the value of resource flexibility.

There are two major streams of work in the Operations Management literature that have focused on the benefits of flexible capacity in the face of demand uncertainty. We will characterize these as the *capacity investment* literature and the *inventory* literature. The capacity investment literature is quite large, with strong connections to production economics and finance (options pricing). Examples of work close to ours are Fine and Freund (1990), who also formulate a single-period capacity acquisition problem as a two-stage stochastic program. While their formulation allows for an unlimited number of products, their analysis focuses on two products, two dedicated resources, one flexible resource, and a discrete demand distribution. We assume that the demand follows a continuous distribution and that the firm has multiple flexible resources that accept customer upgrades. Van Mieghem (1998) also focuses on a two-product firm and, with the assumption of continuous demand and capacity, he shows that the optimal capacities can be found by solving a multidimensional news-vendor problem. Harrison and Van Mieghem (1999) examine the multi-period capacity problem with uncertain demand in each period and a cost to adjust capacity between periods. Our environment is substantially simpler, with the advantage that the solution reduces to a relatively simple set of simultaneous equations. Finally, Jordan and Graves (1995) examine the benefits of process flexibility and demonstrate that limited flexibility, carefully configured, can be nearly as valuable as complete flexibility. Here we also consider a form of limited flexiblity, although we focus our efforts on finding the optimal resource capacities, while Jordan and Graves assumed that capacity is fixed.

The related inventory literature is also quite voluminous, and we will only mention a few closely-related works here. Good summaries can be found in Khouja (1999) and Bassok, Anapundi and Akella (1999, hereafter referred to as BAA). The earliest work on inventory substitution is by McGillivray and Silver (1978) who compare cases with no substitution and complete substitution, and present a heuristic for finding optimal order-up-to levels with partial substitution. Pasternak and Drezner (1991) examine the value of substitution in a two-product inventory system. They derive optimality conditions similar to those presented below in Section 3, although our approach is more concise and is extended to the general case with 'n' services and 'n' resources. Mathematically, the model presented in this paper is a special case of the model developed in BAA, 1999. Our model was developed with the service application in mind, and in this paper we will continue to interpret its parameters in terms of the application described above. However, with minor modifications, the following results apply to the inventory problem as well. Gans and Zhou (1999), in a different setting, also describe the relationship between an inventory problem and a problem in staffing and capacity planning. Finally, Rudi and Netessine (1999) consider a related problem where customers (rather than firm) decide how to substitute a product that is out of stock.

The most important difference between our model and that of Van Mieghem (1998) and BAA is in the description of capacity flexibility. Van Mieghem's formulation can be adapted to an arbitrary substitution structure, while BAA allow "full downward substitution", i.e. a customer can be served by any capacity of a higher, more expensive class. Here we assume that substitution by only one level is profitable. In many service environments this is a reasonable approximation of reality, for substitution by more than one class may dedicate a prohibitively expensive resource, such as a luxury car, to service generating little revenue, as from a customer only willing to pay for a subcompact. Our assumption enables us to find simple expressions for

the first-order optimality conditions, to make explicit comparisons among solutions with and without considering flexibility, and to derive efficient methods for calculating the optimal solutions for problems with large numbers of products and services. With a more general structure it is difficult to find solutions to problems with more than a few products. For example, the solution method proposed by BAA requires the computation of nested integrals whose number "increases exponentially with the number of products." (pg. 637).

Our single-upgrade structure, as well as the assumption that we make later in the paper, that demand follows a multivariate Normal distribution, also allows us to characterize the effect of changes in demand correlation on the optimal solution. To our knowledge, there are no previous analytical results describing these effects on flexible capacity decisions. Eppen (1979) discusses the impact of correlation on the centralization decision for a single-product inventory system. Eynan (1996) studies the effect of demand correlation on the benefits of component commonality. Using a bivariate uniform distribution with correlation, he derives algebraic expressions for the optimal inventory levels of specialized and shared components. These expressions are quite complex, and Eynan uses numerical experiments to examine the impact of correlation. Fine and Freund (1990) demonstrate numerically that the value of flexible capacity can disappear in the presence of perfect correlation. Van Mieghem (1998) shows analytically that, given perfect correlation, it may yet be optimal to invest in flexible capacity if there are differences in the profit margins of the products. Van Mieghem (1997) describes numerical evidence that the optimal level of dedicated (flexible) capacity increases (decreases) as correlation increases.

Our analytical results demonstrate that when there are just two types of service, an increase in correlation will indeed lead to a shift from flexible to dedicated capacity. This shift may also occur as *both* types of capacity rise or fall, a phenomenon not described in the literature. When

4

there are three or more classes, there are also adjustments in the resources not directly affected by the correlation change. The adjustments follow an alternating pattern, with every other resource rising, and the remaining resources falling.

We present the general model in the following section, while in Section 3 we restrict the model to the single-upgrade case and derive simple optimality conditions. Section 4 describes the impact of the demand correlation structure on the optimal capacity decision. Section 5 suggests algorithms for solving the problem numerically and provides numerical examples based on parameters gleaned from the car rental industry. Finally, in Section 6 we summarize the results and describe directions for the future research.

2. Model formulation

Consider a company that provides services indexed by $i = 1,...,n$, for which each service has an associated revenue per unit (e. g., billing rate) p_i and a penalty cost per unit for not fulfilling the demand for this service C_i. To provide these services, the company invests in resources $j = 1,...,n$. Service i can be delivered by resource i, but it can also be delivered by any resource $j \leq i$, as in BAA. The amount of capacity of resource j purchased by the firm is x_j, with a cost paid per unit of capacity F_j. This cost is incurred whether the resource is used or not. In addition to this fixed or purchase cost, there is a variable or usage cost V_j, which is incurred if the unit of capacity is used.

After the capacity investment decision is made, the demand is observed. Demands for services $\{D_i\}$ have realizations $\{d_i\}$ and a joint probability density function $f_{\underline{D}}(\bullet)$ (we denote vectors by underlined letters ($\underline{D}, \underline{x}$) and assume that all demand levels, capacities, and allocations are continuous variables). Our objective is to find the optimal capacity for each resource in order to maximize expected profit under the assumption that once demand $\{d_i\}$ is observed, we allocate optimally the existing resources to fulfill demand. Let y_{ij}, $i = 1,...,n$, $j = 1,...,n$ represent

5

the amount of service i delivered by resource j once demand is observed. The problem

formulation is

$$
\begin{aligned}
\text{Stage 1:} \quad &\max_{\underline{x} \geq 0} \Pi(\underline{x}) = \max_{\underline{x} \geq 0}\left[\underset{\underline{D}}{E}\big[\Theta(\underline{D},\underline{x})\big] - \sum_j F_j x_j\right] \\
\text{Stage 2:} \quad &\Theta(\underline{d},\underline{x}) = \max_{\underline{y}} \sum_{i,j} \alpha_{ij} y_{ij} - \sum_i C_i d_i \\
&s.t.: \sum_j y_{ij} \leq d_i \qquad \forall i; \\
&\qquad\quad \sum_i y_{ij} \leq x_j \qquad \forall j; \\
&\qquad\quad y_{ij} \geq 0 \qquad\quad \forall i,j; \\
&\qquad\quad y_{ij} = 0 \qquad\quad if \ \ i < j.
\end{aligned}
\tag{1}
$$

where $\alpha_{ij} = p_i - V_j + C_i$ is the contribution margin when resource j handles a demand i.

Although the firm will often have a choice of which resource to use to provide a service we

assume that the customers either do not perceive a difference in the service based on the resource

or if they do perceive a difference they do not attempt to game the system. An example of such

gaming would be when a customer asks for a compact car when he actually wants a midsized

car, knowing that it is likely that he will be upgraded while paying the price of a compact car.

As mentioned in Section 1, similar models have appeared in the literature (Fine and Freund

(1990) for capacity investment decisions, Hsu and Bassok (1999) for production management

with random yields, and BAA (1999) for inventory order-up-to policies). Our model is closest to

BAA in mathematical structure but is different in interpretation. Our variables are capacity

decisions; theirs are order-up-to levels. Our costs, $F_j, V_j,$ and C_j represent capacity costs and

contribution margins from use of a unit of the capacity; their costs are inventory purchase,

holding, and lost sales costs.

Our general focus is on the long-term capacity choice, and we seek insights into the value of

resource flexibility. For analytical tractibility we choose a simple, single-period model to

accomplish short-term scheduling. Implicitly, our model assumes that the capacity (Stage 1)

6

decision is fixed for some period of time, and that the Stage 2 problem is solved repeatedly (e.g., day after day) within that period. Note that the fixed costs, F_j, must be appropriately scaled by prorating the total fixed costs over the period. In addition, the distribution of demand $f_{\underline{D}}(\bullet)$ is derived from the distributions of daily demands. For example, $f_{\underline{D}}(\bullet)$ may incorporate both 'peak' and 'lull' demand days, with the probability of each demand scenario determined by that scenario's relative frequency within the long-term period. Since in most service applications the demands from various customer classes may move together, or in opposite directions, it is particularly important to understand the effects of demand correlation on the overall capacity decision. This discussion motivates our efforts in Section 4, below.

This single-period model has a few potential pitfalls, and we discuss these shortcomings here. First, the single period model does not capture situations in which customers order a rental car, for example, for several days. Such multi-day rentals create a correlation in demand *among* days that is not captured by our model. Our model is an approximation, and we suspect that if the percentage of multi-day rentals is small then the effect of the multi-day orders on the optimal capacity is small. Second, we assume that all demand in a period is observed before resources are allocated to it. In reality orders would come in more continuously and resources would need to be allocated to them continuously. Thus, our model gives an optimistic estimate of the value of a given set of resources. There is a great deal of research in the revenue management literature that discusses the problem of allocating a given capacity as demand is observed. The dynamic capacity investment and allocation problem is beyond the scope of this paper, but is a promising area for research.

Third, given multiple periods and continuous arrivals, one might consider a queuing model to represent this situation. Again we feel this approach, although possibly more accurate, would limit our ability to derive analytic results. Green (1985) and Shumsky (1999) must resort to

numerical procedures and approximations for steady-state performance analysis of queueing systems with both general and flexible servers, while Harrison and Lopez (1999) use a heavy traffic analysis for the dynamic scheduling problem.

Fourth, the assumption that all customers are willing to upgrade to a higher service level may not be true in reality (if, for instance, a person has a strong preference for a smaller car due to the fuel consumption). However, our model can be altered to incorporate a parameter designating the proportion of customers willing to upgrade, although in the following text, to keep the number of parameters manageable, we shall assume that all customers will accept upgrades. Finally, the continuity assumption makes it difficult to apply this model to problems of small size where every unit of capacity can have a significant impact on profits. In practice, however, such problems may be solved by complete enumeration.

We end this section by noting that our formulation is solvable at least in theory because the problem is concave.

Proposition 1. The objective function $\Pi(\underline{x})$ is jointly concave in capacity variables $\underline{x}$.

Proof: Observe that function $\Theta(\underline{d},\underline{x})$ is jointly concave in $\underline{d}$ and $\underline{x}$ since $\underline{d}$ and $\underline{x}$ determine the right-hand-side of a linear program. Further, the expectation of a concave function, e.g., $E_D(\Theta(\underline{D},\underline{x}))$, is concave in $\underline{x}$ and hence $\Pi(\underline{x})$ is concave as it is a sum of concave and linear functions. ∎

On the other hand, if arbitrary upgrades are allowed the problem is still quite complex. Thus in the next section we limit ourselves to single level upgrades and a further restriction of the costs that allow the stage 2 transportation problem to be solved explicitly.

3. Single Level Upgrades

We now make the following additional assumptions about the structure of the problem. We restrict the firm to using only resources i and i-1 for service i by restricting the parameters of the

problem as follows:

(I) $V_n \leq V_{n-1} \leq \cdots \leq V_2 \leq V_1$

(II) $(p_n + C_n) \leq (p_{n-1} + C_{n-1}) \leq \cdots \leq (p_2 + C_2) \leq (p_1 + C_1)$ (2)

(III) $\alpha_{ij} \geq 0$ for all $j \leq i \leq j+1$ and $\alpha_{ij} < 0$ otherwise

Although these assumptions may seem restrictive we claim that many practical problems would fall in this category. For example in rental cars, it is fairly common to see customers upgraded for lack of capacity of the desired car type but rarely do companies upgrade more than one level, the implication of constraint III. Rental cars have a natural ordering where Assumptions I and II apply, namely that as you move up in quality, the sizes of revenues, lost-sales costs, and variable costs all increase. These assumptions allow us to solve the inner "transportation problem" so as to obtain analytical expressions for the objective function without constraints. Furthermore, under these assumptions there exists a very efficient procedure for computing solutions.

Proposition 2. Suppose that there are n service classes and n resource classes with one level upgrades and where conditions (2) apply. The stage 2 problem can be solved to optimality by a greedy algorithm as follows:

$$y_{ii} = \min(x_i, d_i) \qquad\qquad \forall i = 1, \ldots, n$$
$$y_{i+1,i} = \min((d_{i+1} - x_{i+1})^+, (x_i - d_i)^+) \qquad \forall i = 1, \ldots, n-1.$$

Note that the allocation method described by the algorithm is very natural: first allocate each customer to her requested class and then upgrade the rest if there remains excess capacity. Hoffman (1966) identified conditions under which the transportation problem is solvable by the greedy algorithm, and conditions (2) meet his criteria. BAA provide additional references and demonstrate that the result applies to their generalization of our formulation (recall that their formulation allows multi-level upgrades).

In Figure 1 we show the transportation problem, including the demand realizations for n

9

services on the left, and the capacities for the *n* resources on the right. The y_{ij} indicate the allocation of services to resources and are the flows on these arcs.
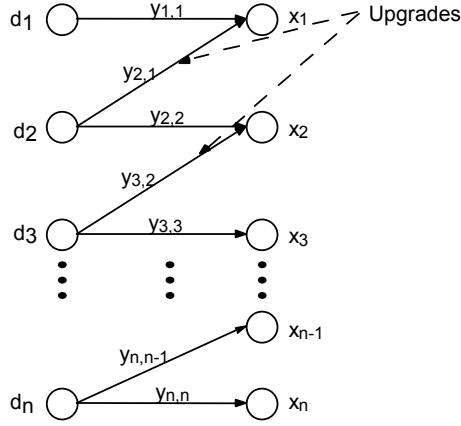

Figure 1: Transportation problem for n types of service and n resources.

From Proposition 2 we can see that the Stage 1 objective function is:

$$\Gamma(\underline{x}) = \sum_{i=1}^{n} \left\{ \alpha_{ii} E[\min(D_i, x_i)] + \alpha_{i+1,i} E[\min((D_{i+1} - x_{i+1})^+, (x_i - D_i)^+)] - F_i x_i - C_i E[D_i] \right\} \qquad (3)$$

where we define $\alpha_{n+1,n} \equiv 0$. Note that $\Gamma(\underline{x}) \equiv \Pi(\underline{x})$ if the demand distribution has positive support. For the remainder of the paper we will focus on $\Gamma(\underline{x})$ rather than $\Pi(\underline{x})$, and in Section 4 we will use $\Gamma(\underline{x})$ as an approximation for $\Pi(\underline{x})$ when the demand distribution is Normal and the coefficients of variation for all demand classes are small ($\sigma_i << \mu_i$ for all *i*).

Now note that the second expectation within each term of (3) represents the benefits of the flexibility in an explicit form. Since the objective function is concave, the first order conditions are necessary and sufficient. They are

$$\frac{\partial \Gamma(\underline{x})}{\partial x_1} = \alpha_{11} P(D_1 \geq x_1) + \alpha_{21} P(D_1 \leq x_1, D_1 + D_2 \geq x_1 + x_2) - F_1 = 0 \qquad (4)$$

$$\frac{\partial \Gamma(\underline{x})}{\partial x_i} = \alpha_{ii} P(D_i \geq x_i) + \alpha_{i+1,i} P(D_i \leq x_i, D_i + D_{i+1} \geq x_i + x_{i+1})$$
$$- \alpha_{i,i-1} P(D_i \geq x_i, D_{i-1} + D_i \leq x_{i-1} + x_i) - F_i = 0 \qquad \text{for } i = 2,...,n\text{-}1 \qquad (5)$$

10

$$\frac{\partial \Gamma(\underline{x})}{\partial x_n} = \alpha_{nn} P(D_n \geq x_n) - \alpha_{n,n-1} P(D_n \geq x_n, D_{n-1} + D_n \leq x_{n-1} + x_n) - F_n = 0 \tag{6}$$

BAA derive conditions for their more general case by a systematic, logical argument. For our case the greedy algorithm results in an analytic expression for the objective, (3), which in turn allows us to derive (4)-(6) by differentiating (3). From (4)-(6) it is easy to compute the second partial derivatives. Specifically, $\partial^2 \Gamma(\underline{x})/\partial x_i \partial x_j = 0$ for $j > i+1$ and $j < i$, and for all $i=1...n-1$,

$$\frac{\partial^2 \Gamma(\underline{x})}{\partial x_i \partial x_{i+1}} = -\alpha_{i+1,i} \int_{-\infty}^{x_i} f_{D_i, D_i + D_{i+1}}(t, x_i + x_{i+1}) dt \tag{7}$$

These derivatives are non-positive, thus proving that the profit function, $\Gamma(x)$, is sub-modular in x (see Topkis, 1978). The submodularity of the objective function shows that an increase in one of the variables (capacity) decreases the returns due to increasing the other variable. Hence, once flexibility is introduced, the two resources become substitutes.

Recall that, without accounting for flexibility, a solution can be obtained by using the classical news-vendor formula:

$$P(D_i \leq x_i^{NV}) = \frac{\alpha_{ii} - F_i}{\alpha_{ii}} \quad \text{for } i = 1,...,n \tag{8}$$

We can rearrange the first order conditions for our problem to obtain

$$P(D_1 \leq x_1^*) = \frac{\alpha_{11} - F_1}{\alpha_{11}} + \frac{\alpha_{21}}{\alpha_{11}} P(D_1 \leq x_1^*, D_1 + D_2 \geq x_1^* + x_2^*) \tag{9}$$

$$P(D_i \leq x_i^*) = \frac{\alpha_{ii} - F_i}{\alpha_{ii}} + \frac{\alpha_{i+1,i}}{\alpha_{ii}} P(D_i \leq x_i^*, D_i + D_{i+1} \geq x_i^* + x_{i+1}^*)$$
$$- \frac{\alpha_{i,i-1}}{\alpha_{ii}} P(D_i \geq x_i^*, D_{i-1} + D_i \leq x_{i-1}^* + x_i^*) \qquad \text{for } i = 2,...,n\text{-}1 \tag{10}$$

$$P(D_n \leq x_n^*) = \frac{\alpha_{nn} - F_n}{\alpha_{nn}} - \frac{\alpha_{n,n-1}}{\alpha_{nn}} P(D_n \geq x_n^*, D_{n-1} + D_n \leq x_{n-1}^* + x_n^*) \tag{11}$$

where x_i^* shall denote the optimal capacity decision for the case with flexibility. Observe that the first terms of the expressions on the right of (9) and (11) are the news-vendor ratios. One

would expect that the optimal capacity levels should be different from the news-vendor

quantities when accounting for flexibility. This is indeed the case, as shown by (9) and (11).

Moreover, the relationship between the optimal capacity levels and the news-vendor quantities is

quite specific:

Proposition 3. $x_1^* \geq x_1^{NV}$, $x_n^* \leq x_n^{NV}$.

Proof: This follows immediately from equations (9) and (11) above and the sign on the second

probability terms. ∎

Note that we cannot conclude that a similar inequality holds for x_i^* for $i = 2,\ldots,$ n-1. The first

terms on the right of (10) are the news-vendor ratios. However, it is no longer obvious if each

optimal capacity x_i^* is more or less than the news-vendor quantity for $i = 2, \ldots,$ n-1, since the

right-hand side of (10) involves both a positive and a negative term and it is not clear which one

dominates. We provide an example later where the optimal quantity of a "middle" resource rises

above and below the news-vendor quantity as a parameter changes.

Finally, Equations (9)-(11) allow us to derive simple bounds on the optimal capacity choice.

Given $\alpha_{10} = \alpha_{n+1,n} \equiv 0$, then we have

Proposition 4. $\quad \dfrac{\alpha_{ii} - F_i - \alpha_{i,i-1}}{\alpha_{ii} - \alpha_{i,i-1}} \leq P(D_i \leq x_i^*) \leq \dfrac{\alpha_{ii} - F_i}{\alpha_{ii} - \alpha_{i+1,i}} \quad$ for $i = 1,\ldots,n.$

Proof: To prove the upper bound note that

$$P(D_i \leq x_i^*) = \frac{\alpha_{ii} - F_i}{\alpha_{ii}} + \frac{\alpha_{i+1,i}}{\alpha_{ii}} P(D_i \leq x_i^*, D_i + D_{i+1} \geq x_i^* + x_{i+1}^*) - \frac{\alpha_{i,i-1}}{\alpha_{ii}} P(D_i \geq x_i^*, D_{i-1} + D_i \leq x_{i-1}^* + x_i^*)$$

$$\leq \frac{\alpha_{ii} - F_i}{\alpha_{ii}} + \frac{\alpha_{i+1,i}}{\alpha_{ii}} P(D_i \leq x_i^*, D_i + D_{i+1} \geq x_i^* + x_{i+1}^*) \leq \frac{\alpha_{ii} - F_i}{\alpha_{ii}} + \frac{\alpha_{i+1,i}}{\alpha_{ii}} P(D_i \leq x_i^*).$$

Therefore, $P(D_i \leq x_i^*) \leq (\alpha_{ii} - F_i)/(\alpha_{ii} - \alpha_{i+1,i}).$ The proof for the lower bound is similar. ∎

Note that when $i=1$, the lower bound is the news-vendor ratio, while when $i=n$ the upper bound

is the news-vendor ratio. The numerical experiments in Section 5 will demonstrate that these

bounds are quite loose for many problem parameters (see, for example, Figure 4).

4. Correlation and Capacity

In a variety of business environments it is important to understand the impact of correlation on the optimal capacity choice. For example, a car rental company may experience a heavy load throughout the week due to business travelers who tend to rent mid-sized and luxury cars. During the weekend the demand shifts to leisure travelers who, on average, prefer either smaller, cheaper vehicles or minivans (Carroll and Grimes, 1995). When daily demands are aggregated into a general distribution, demand among certain cars classes will be negatively correlated. This pattern will be altered during holidays and special events when demand for all car classes is high. What effect does the correlation structure of demand have on the optimal capacity decision?

Van Mieghem (1997) examines the relative benefits of dedicated and flexible capacity for the production of two goods. Numerical experiments using a demand distribution with positive support over an elliptical region demonstrate that "the optimal levels of dedicated capacity increase in a concave manner as correlation increases, while the optimal level of flexible capacity decreases in a convex manner." (Van Mieghem, 1997, pg. 12). Is this experimental result true, in general? And what can we say about the $n \times n$ case? In this section we present our main results about how the optimal capacities vary when one correlation, e.g., $\rho_{i,i+1}$, increases. For the 2×2 case, Van Mieghem's experiments reveal just one of three possible results: as correlation rises, both types of capacity may rise, both may fall, or dedicated capacity may rise while flexible capacity falls. However, given any of the three cases, there will be a shift from flexible to dedicated capacity.

Initially, we will make no assumptions about the distribution of demand. In fact, in Section 4.1 we show that our objective function, as defined in (3), is concave for any demand distribution, including those with negative support. In Section 4.2 we restrict our attention to the

13

multivariate normal distribution, and in Section 4.3 we further limit the range of possible parameters. Each assumption allows us to characterize more strongly the impact of a rise in correlation on the optimal capacity decision.

4.1 Impact of Correlation for an Arbitrary Demand Distribution

In Section 4.2 we will assume that the products follow the multivariate Normal demand distribution. However, in order to make this statement precise we must first define what we mean when we say that demand can take on negative values, which would be meaningless as demands, and render model (1) infeasible.

If $\mu \gg \sigma$ then truncating the distribution and giving it a probability mass at 0 (or distributing that mass over the rest of the distribution) is a common approach used by practitioners. However, this method would complicate our analysis and would exclude the application of the substantial mathematical machinery that has been developed for the Normal distribution.

Thus, our approach is to approximate the true optimum, $\Pi(\underline{x})$ evaluated using a truncated distribution, with the expression $\Gamma(\underline{x})$ given by (3) but with an arbitrary distribution for demand. From now on it is understood that $\Gamma(\underline{x})$ denotes an approximation to the optimal $\Pi(\underline{x})$ in cases where demand has negative support. We have confirmed, through both analytical results and numerical experiments, that this approximation is a good one when demand is normally distributed and when the coefficients of demand variation are reasonably small (i.e. $\sigma_i/\mu_i < 0.5$ for all i). We will confirm that the expression $\Gamma(\underline{x})$ expressed as (3) is concave for any distribution, but first we state and prove two auxiliary results.

Lemma 1. Let $H(\Gamma)$ be the Hessian of $\Gamma(\underline{x})$. For any demand distribution, $H(\Gamma) \leq 0$ and all elements on the diagonal of $H(\Gamma)$ are strictly negative.

Proof: see Appendix A.

Lemma 2. For any demand distribution $H(\Gamma)$ is strictly diagonally dominant.

Proof: see Appendix A.

Proposition 5. The objective function $\Gamma(\underline{x})$ as expressed in (3) is concave for any demand distribution.

Proof: A sufficient condition for strict concavity of a function is that its Hessian is negative definite. According to Horn and Johnson (1996, Corollary 7.2.3) the sufficient condition of negative definiteness is that all diagonal elements are strictly negative (Lemma 1) and strict diagonal dominance (Lemma 2). Hence, $\Gamma(\underline{x})$ is strictly concave. ∎

We are interested in how the optimal capacity levels change as a single coefficient of correlation changes. It is understandable that a complete characterization will depend on the demand distribution. There are, however, some results that we would expect to hold for any demand distribution. For example, it is conceivable that a change in the coefficient of correlation $\rho_{i-1,i}$ will have the largest effect on the capacities x_{i-1} and x_i and that the effect will diminish as we move away (in terms of the ordering $1,\ldots,n$) from i and $i-1$. We might also expect that the direction of capacity change (i.e., decrease or increase) as correlation rises will be reversed for neighboring capacities. This result is suggested by the submodularity of the objective function. In other words, an increase in the capacity x_k should lead to a decrease in the capacity x_{k-1}. This change will in turn lead to an increase in the capacity x_{k-2} and so on. The next proposition verifies that both of these hypotheses are correct.

Let $\underline{x}^*$ be an optimal capacity vector, given correlation vector $\underline{\rho}$. In the following proposition we will examine the effect of a change in $\rho_{i-1,i}$, while the other elements of $\underline{\rho}$ are held constant. We will see that $\underline{x}^*$ can be seen as a function of the correlation coefficient, e.g.,

$$\underline{x}^* = f(\rho_{i-1,i}).$$

Proposition 6. Let sgn($\bullet$) indicate the sign of its argument. Then for any demand distribution and any $i=2,...,n$:

(i) $\qquad \left| \dfrac{\partial x_1^*}{\partial \rho_{i-1,i}} \right| < \left| \dfrac{\partial x_2^*}{\partial \rho_{i-1,i}} \right| < \cdots < \left| \dfrac{\partial x_{i-1}^*}{\partial \rho_{i-1,i}} \right|;$

(ii) $\qquad \left| \dfrac{\partial x_n^*}{\partial \rho_{i-1,i}} \right| < \left| \dfrac{\partial x_{n-1}^*}{\partial \rho_{i-1,i}} \right| < \cdots < \left| \dfrac{\partial x_i^*}{\partial \rho_{i-1,i}} \right|;$

(iii) $\qquad \mathrm{sgn}\left[\dfrac{\partial x_1^*}{\partial \rho_{i-1,i}} \right] = -\mathrm{sgn}\left[\dfrac{\partial x_2^*}{\partial \rho_{i-1,i}} \right] = \cdots = -\mathrm{sgn}\left[\dfrac{\partial x_{i-2}^*}{\partial \rho_{i-1,i}} \right] = \mathrm{sgn}\left[\dfrac{\partial x_{i-1}^*}{\partial \rho_{i-1,i}} \right];$

(iv) $\qquad \mathrm{sgn}\left[\dfrac{\partial x_n^*}{\partial \rho_{i-1,i}} \right] = -\mathrm{sgn}\left[\dfrac{\partial x_{n-1}^*}{\partial \rho_{i-1,i}} \right] = \cdots = -\mathrm{sgn}\left[\dfrac{\partial x_{i+1}^*}{\partial \rho_{i-1,i}} \right] = \mathrm{sgn}\left[\dfrac{\partial x_i^*}{\partial \rho_{i-1,i}} \right].$

In Figure 2 we illustrate Proposition 6 using an example with six resources. Arrows show the direction of change for each optimal capacity, and the size of each arrow indicates the relative magnitude of change. Note that the directions of change for capacities 3 and 4 are picked arbitrarily at this point, since Proposition 6 does not say anything about changes in x_{i-1} and x_i. We address this issue in the next section.
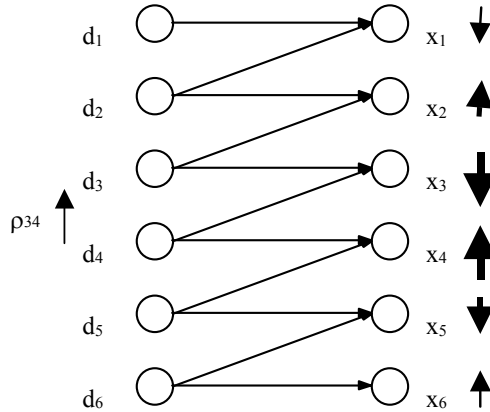

Figure 2. Direction and magnitude of the change in optimal capacity.

Proof: To simplify the notation, let $\rho_{i-1,i} \equiv \rho$. We will prove statements (i) and (iii), while the proofs of statements (ii) and (iv) are analogous. Consider the first order conditions, (4) through

16

(6) above. Let us denote the expressions on the left with the function $\Omega(\underline{x},\rho): R^{n+1} \to R^n$, where $\underline{x}$ is the vector of capacities and ρ is a single correlation coefficient. This function is continuous and is equal to zero at the optimal capacity vector $\underline{x}^0$ for a given ρ^0: $\Omega(\underline{x}^0,\rho^0) = \underline{0}$. Note that $\underline{x}^0$ is defined implicitly by ρ^0 through function Ω.

We will use the implicit function theorem (IFT) to characterize this relationship. First we make sure that all the relevant conditions are satisfied. The function Ω has continuous partial derivatives $\partial\Omega_i / \partial x_j$ and $\partial\Omega_i / \rho$ in a neighborhood of $(\underline{x}^0,\rho^0)$, and the Jacobian matrix of partial derivatives of Ω with respect to $\underline{x}$, $J_\Omega(\underline{x})$, is invertible when evaluated at $(\underline{x}^0,\rho^0)$. The IFT states that under these conditions, there is an open set $V \subset R^1$ containing ρ^0 and a unique function $f : V \to R^n$ such that $f(\rho^0) = \underline{x}^0$, $\Omega(f(\rho),\rho) = 0$ for all $\rho \in V$. In other words,

$$\underline{x}^* = f(\rho).$$

We are interested in the derivative of the function f. Let $J_f(\rho)$ and $J_\Omega(\rho)$ denote the vectors of partial derivatives with respect to ρ of functions f and Ω, respectively. The IFT states that

$$J_f(\rho) = -[J_\Omega(\underline{x})]^{-1} J_\Omega(\rho). \tag{12}$$

We left-multiply this expression by the matrix $[J_\Omega(\underline{x})]$ and note that $[J_\Omega(\underline{x})]$ is a Hessian matrix of the objective function $\Gamma(\underline{x})$. Therefore:

$$H(\Gamma(\underline{x}))J_f(\rho) = -J_\Omega(\rho). \tag{13}$$

We will denote the (ij)-th entry of the Hessian by $H_{ij} = \partial^2\Gamma / \partial x_i \partial x_j$. Since H is symmetric $H_{ij} = H_{ji}$. Equation (13) in a more explicit form is

$$
\begin{bmatrix}
H_{11} & H_{12} & 0 & 0 & \ldots & 0 & 0 & 0 & 0 \\
H_{12} & H_{22} & H_{23} & 0 & \ldots & 0 & 0 & 0 & 0 \\
0 & H_{23} & H_{33} & H_{34} & \ldots & 0 & 0 & 0 & 0 \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
0 & 0 & 0 & 0 & \ldots & H_{n-3,n-2} & H_{n-2,n-2} & H_{n-2,n-1} & 0 \\
0 & 0 & 0 & 0 & \ldots & 0 & H_{n-2,n-1} & H_{n-1,n-1} & H_{n-1,n} \\
0 & 0 & 0 & 0 & \ldots & 0 & 0 & H_{n-1,n} & H_{nn}
\end{bmatrix}
\begin{bmatrix}
\partial x_1^* / \partial \rho \\
\partial x_2^* / \partial \rho \\
\partial x_3^* / \partial \rho \\
\ldots \\
\partial x_{n-2}^* / \partial \rho \\
\partial x_{n-1}^* / \partial \rho \\
\partial x_n^* / \partial \rho
\end{bmatrix}
= (-1)
\begin{bmatrix}
0 \\
0 \\
\ldots \\
\partial \Omega_{i-1} / \partial \rho \\
\partial \Omega_i / \partial \rho \\
\ldots \\
0
\end{bmatrix}.
$$

The remainder of the proof is by mathematical induction. To simplify the proof we will modify the result (i) by making it stronger. Instead of proving that $\left|\partial x_{k-1}^* / \partial \rho\right| < \left|\partial x_k^* / \partial \rho\right|$ we will demonstrate that

$$
\left|\frac{\partial x_{k-1}^*}{\partial \rho}\right| \le \frac{H_{k-1,k}}{H_{k-1,k-1}} \left|\frac{\partial x_k^*}{\partial \rho}\right| \text{ for } k=2,\ldots,i\text{-}1. \tag{14}
$$

Since, according to Lemma 2, $\left|H_{k-1,k}\right| < \left|H_{k-1,k-1}\right|$, the result (i) will immediately follow. First, we verify that conjectures (14) and (iii) hold for $i=2$. Consider the equality obtained by multiplying the first row of the Hessian:

$$
H_{11} \frac{\partial x_1^*}{\partial \rho} + H_{12} \frac{\partial x_2^*}{\partial \rho} = 0 \Leftrightarrow \frac{\partial x_1^*}{\partial \rho} = -\frac{H_{12}}{H_{11}} \frac{\partial x_2^*}{\partial \rho}. \tag{15}
$$

It immediately follows that (14) holds and hence $\left|\partial x_2^* / \partial \rho\right| > \left|\partial x_1^* / \partial \rho\right|$. Further, recall from Lemma 1 that $H_{11} < 0$ and $H_{12} < 0$. Hence, $\text{sgn}\left[\partial x_1^* / \partial \rho\right] = -\text{sgn}\left[\partial x_2^* / \partial \rho\right]$.

Next, we assume that (14) and (iii) hold for some k, $k=3,\ldots,i\text{-}2$ and use this assumption to prove that the same is true for $k+1$. Consider the equality obtained by multiplying the k-th row of the Hessian:

$$
H_{k-1,k} \frac{\partial x_{k-1}^*}{\partial \rho} + H_{kk} \frac{\partial x_k^*}{\partial \rho} + H_{k,k+1} \frac{\partial x_{k+1}^*}{\partial \rho} = 0. \tag{16}
$$

First, re-write (16):

$$
\frac{\partial x_{k+1}^*}{\partial \rho} = \frac{-H_{k-1,k} \dfrac{\partial x_{k-1}^*}{\partial \rho} - H_{kk} \dfrac{\partial x_k^*}{\partial \rho}}{H_{k,k+1}}. \tag{17}
$$

18

By diagonal dominance, $|H_{k-1,k}| < |H_{kk}|$, and by the induction assumption, $|\partial x_{k-1}^* / \rho| < |\partial x_k^* / \partial \rho|$. It

then follows that $|H_{k-1,k} \partial x_{k-1}^* / \rho| < |H_{kk} \partial x_k^* / \partial \rho|$ and hence $\text{sgn}[\partial x_{k+1}^* / \partial \rho] = -\text{sgn}[\partial x_k^* / \partial \rho]$

meaning that (iii) holds for all k. Next, re-arrange (16) again using induction assumptions (iii)

and (14):

$$\left| \frac{\partial x_{k+1}^*}{\partial \rho} \right| = \left| \frac{H_{kk}}{H_{k,k+1}} \frac{\partial x_k^*}{\partial \rho} + \frac{H_{k-1,k}}{H_{k,k+1}} \frac{\partial x_{k-1}^*}{\partial \rho} \right| = \frac{H_{kk}}{H_{k,k+1}} \left| \frac{\partial x_k^*}{\partial \rho} \right| - \frac{H_{k-1,k}}{H_{k,k+1}} \left| \frac{\partial x_{k-1}^*}{\partial \rho} \right|.$$

Finally, using induction assumption (14),

$$\left| \frac{\partial x_{k+1}^*}{\partial \rho} \right| \geq \left(\frac{H_{kk}}{H_{k,k+1}} - \frac{H_{k-1,k}^2}{H_{k-1,k-1} H_{k,k+1}} \right) \left| \frac{\partial x_k^*}{\partial \rho} \right| \geq \frac{H_{kk}}{H_{k,k+1}} \left| \frac{\partial x_k^*}{\partial \rho} \right|.$$

Hence, (14) holds and therefore (i) holds as well. ∎

4.2 Impact of Correlation for the Normal Demand Distribution

Proposition 6 describes how the elements of the set of optimal capacities $(x_1^*, x_2^*, ..., x_{i-1}^*)$ change

relative to each other as $\rho_{i-1,i}$ changes, and it does the same for the set $(x_i^*, x_{i+1}^*, ..., x_n^*)$.

However, Proposition 6 does not describe the relationship between x_{i-1}^* and x_i^*. To characterize

this relationship, we will assume that demand is normally distributed. This assumption allows us

to use a standard result from probability theory, Slepian's inequality (see Tong, 1980, pg. 10).

Slepian's inequality states that the probability $P(D_1 \leq a_1, D_2 \leq a_2)$ is increasing (and hence

probability $P(D_1 \leq a_1, D_2 \geq a_2)$ is decreasing) in the coefficient of correlation for any fixed a_1

and a_2. Note that this result, and thus the results for the remainder of this section, apply to any

elliptically contoured distribution, a set of distributions that include the multivariate normal and t

distributions (Tong, 1980, Theorem 4.3.6).

We will first characterize the impact of correlation on the function $\Omega(\underline{x}, \rho_{i-1,i})$, the first

order conditions defined previously in (4)-(6). Later, we will use this characterization together

with the implicit function theorem to obtain a final characterization of the change in the optimal

capacity levels with the change in correlation.

Lemma 3. Let Ω_i represent the *i*th element of $\Omega(\underline{x}, \rho_{i-1,i})$. If demand is Normally distributed,

then $\partial\Omega_{i-1}/\partial\rho_{i-1,i} - \partial\Omega_i/\partial\rho_{i-1,i} < 0$.

Proof: Again, let $\rho_{i-1,i} \equiv \rho$. From the first order conditions (4)-(6), $\partial\Omega_{i-1}/\partial\rho = \alpha_{i-1,i}\partial P_1/\partial\rho$,

where $P_1 = P(D_{i-1} \le x_{i-1}, D_{i-1} + D_i \ge x_{i-1} + x_i)$. In addition, $\partial\Omega_i/\partial\rho = -\alpha_{i-1,i}\partial P_2/\partial\rho$,

where $P_2 = P(D_i \ge x_i, D_{i-1} + D_i \le x_{i-1} + x_i)$.

Therefore,

$$\frac{\partial\Omega_{i-1}}{\partial\rho} - \frac{\partial\Omega_i}{\partial\rho} = \alpha_{i-1,i}\frac{\partial P_1}{\partial\rho} + \alpha_{i-1,i}\frac{\partial P_2}{\partial\rho} = \alpha_{i-1,i}\frac{\partial P(D_{i-1} \le x_{i-1}, D_i \ge x_i)}{\partial\rho} < 0$$

where the last inequality follows directly from the Slepian's inequality (Tong, 1980). ∎

Lemma 3 leads to an intuitive result: as $\rho_{i-1,i}$ rises there will be a shift from more to less

expensive capacity. When there are two classes of customers and services, the relatively

expensive capacity, x_1, is flexible while the relatively inexpensive capacity, x_2, is specialized.

Therefore, in the *2×2* case, as ρ_{12} rises there will be a shift from flexible to specialized capacity.

Proposition 7. Given that demand is normally distributed, $\dfrac{\partial x_{i-1}^*}{\partial\rho_{i-1,i}} < \dfrac{\partial x_i^*}{\partial\rho_{i-1,i}}$.

Figure 3 displays this result by showing the direction of change in capacity, given a

change in $\rho_{i-1,i}$. The proposition states that the shift will lie somewhere in the halfplane, in

regions I, II, or III. Note that the total amount of capacity may either rise or fall.
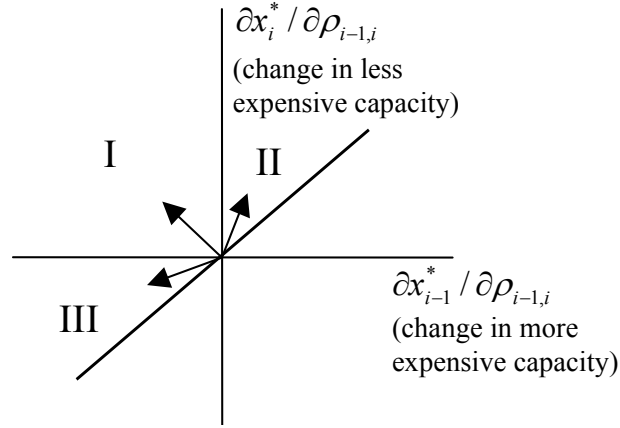
$\partial x_i^* / \partial \rho_{i-1,i}$
(change in less expensive capacity)

I

II

III

$\partial x_{i-1}^* / \partial \rho_{i-1,i}$
(change in more expensive capacity)

Figure 3: Change in capacity, given an increase in ρ.

A shift to region I indicates a decline in the relatively expensive capacity while inexpensive capacity rises. A shift to region II or region III indicates a rise or decline, respectively, in both types of capacity. However, in region II we add more inexpensive capacity than expensive capacity, and in region III we subtract more expensive capacity than inexpensive capacity.

While the Proposition does not rule out a shift to regions II or III, we can do so under particular parametric assumptions (see the next section). In addition, our numerical experiments demonstrate that while a shift to region II or III is possible, a shift to region I is the norm.

Proof of Proposition 7. As in Proposition 6, let $\rho_{i-1,i} \equiv \rho$, let $\underline{x}^* = f(\rho)$ be the optimal capacity vector that maximizes $\Gamma(\underline{x})$, and let $\underline{x}^0 = f(\rho^0)$ be the optimal capacity vector for a given ρ^0. We will examine the impact on x_{i-1}^* and x_i^* as ρ moves away from ρ^0. To this end, define the function

$$g(\delta,\varepsilon) = \Gamma\left(x_1^*,\ldots,x_{i-2}^*,x_{i-1}^0 - \delta + \varepsilon, x_i^0 + \delta + \varepsilon, x_{i+1}^*,\ldots,x_n^*\right) \equiv \Gamma\left(\underline{x}^*,\delta,\varepsilon\right)$$

where δ represents the optimal decrease below x_{i-1}^0 and the optimal increase above x_i^0 with a change from ρ^0 to an arbitrary ρ, and ε represents the optimal orthogonal shift that takes both capacities in the same direction.

21

As in Proposition 6, we will use the implicit function theorem (IFT) and Slepian's inequality to show that δ rises as ρ rises. To apply the IFT, we note that the first-order condition $\partial g / \partial \delta \equiv K(\delta, \rho) = 0$ implicitly defines δ as a function of ρ. Therefore,

$$\frac{\partial \delta}{\partial \rho} = -\left(\frac{\partial K}{\partial \delta}\right)^{-1} \frac{\partial K}{\partial \rho} = -\left(\frac{\partial^2 g}{\partial \delta^2}\right)^{-1} \frac{\partial^2 g}{\partial \rho \partial \delta} \tag{18}$$

The first of the two terms is a second-order directional derivative of a concave function, so that $\partial^2 g / \partial \delta^2 < 0$. To find the sign of the second term, $\partial^2 g / \partial \rho \partial \delta$, we first note that the derivative of the function g with respect to δ is

$$\frac{\partial g}{\partial \delta} = \nabla_x \Gamma \circ \left(\frac{\partial x_1}{\partial \delta}, ..., \frac{\partial x_n}{\partial \delta}\right)^T = \nabla_x \Gamma \circ (0,...,0,-1,1,0,...0)^T = -\frac{\partial \Gamma}{\partial x_{i-1}} + \frac{\partial \Gamma}{\partial x_i}. \tag{19}$$

From first order conditions (4) through (6) and Lemma 3,

$$\frac{\partial^2 g}{\partial \rho \partial \delta} = \frac{\partial}{\partial \rho}\left(-\frac{\partial \Gamma}{\partial x_{i-1}} + \frac{\partial \Gamma}{\partial x_i}\right) = -\left(\frac{\partial \Omega_{i-1}}{\partial \rho} - \frac{\partial \Omega_i}{\partial \rho}\right) > 0. \tag{20}$$

Therefore, $\partial \delta / \partial \rho > 0$. We complete the proof by linking a change in δ with changes in x_i^* and x_{i-1}^*: $x_i^* = x_i^0 + \varepsilon + \delta$ and $x_{i-1}^* = x_{i-1}^0 + \varepsilon - \delta$. Therefore,

$$\frac{\partial x_i^*}{\partial \rho} = \frac{\partial \varepsilon}{\partial \rho} + \frac{\partial \delta}{\partial \rho} \text{ and } \frac{\partial x_{i-1}^*}{\partial \rho} = \frac{\partial \varepsilon}{\partial \rho} - \frac{\partial \delta}{\partial \rho} \Rightarrow \frac{\partial x_i^*}{\partial \rho} - \frac{\partial x_{i-1}^*}{\partial \rho} = 2\frac{\partial \delta}{\partial \rho} > 0. \blacksquare$$

4.3 Impact of Correlation for Specified Parameters

Thus far we have not made any assumptions about parameters of the demand distribution. It is, however, possible to somewhat strengthen the results of Proposition 7 if appropriate restrictions are imposed on the parameters of the distribution. In fact, given certain parameter restrictions, it is possible to eliminate region II or region III of Figure 3 from the feasible set of capacity changes. In what follows, we will show that if $\mu_{i-1} + \mu_i < x_{i-1}^* + x_i^*$ (the firm maintains a high service level), then region III may be ruled out. In that case, a rise in correlation cannot reduce

the optimal levels of *both* types of capacity. On the other hand, if $\mu_{i-1} + \mu_i > x_{i-1}^* + x_i^*$ (the firm maintains a low service level) then region II may be ruled out, and both types of capacity cannot increase. In each case we will require moderate restrictions on the correlation coefficient to obtain our results.

Proposition 8. Suppose demand is Normally distributed. Then,

(i)   if $\mu_{i-1} + \mu_i < x_{i-1}^* + x_i^*$ and either $\rho_{i-1,i} < -\sigma_i / \sigma_{i-1}$ or $\rho_{i-1,i} > -\sigma_{i-1} / \sigma_i$, then $\partial x_i^* / \partial \rho_{i-1,i} > 0$,

(ii) if $\mu_{i-1} + \mu_i > x_{i-1}^* + x_i^*$ and either $\rho_{i-1,i} < -\sigma_{i-1} / \sigma_i$ or $\rho_{i-1,i} > -\sigma_i / \sigma_{i-1}$, then $\partial x_{i-1}^* / \partial \rho_{i-1,i} < 0$.

Proof: See Appendix B.

Note that the assumptions on the correlation coefficients in Proposition 8 are not very restrictive. The conditions $\rho_{i-1,i} > -\sigma_{i-1}/\sigma_i$ in (i) and $\rho_{i-1,i} > -\sigma_i/\sigma_{i-1}$ in (ii) are true for a variety of situations. They both hold if $\rho_{i-1,i} \geq 0$ or if $\sigma_{i+1} = \sigma_i$ or if σ_{i+1} is close to σ_i and $\rho_{i-1,i}$ is not too negative. However, we will later verify through numerical experiments that it is not possible to further restrict the direction of the capacity shift. We will demonstrate that shifts to any of the regions are possible, given appropriate parameters.

5. Algorithms and Numerical Experiments

Here we describe an algorithm for finding the optimal capacities x_i^*. Let us denote the inverse of the cumulative distribution function by $\Psi_i^{-1}(\bullet)$ and define the function $\eta_i(\underline{x})$ as:

$$\eta_i(\underline{x}) = \Psi_i^{-1}\left[\frac{\alpha_{ii} - F_i}{\alpha_{ii}} + \frac{\alpha_{i+1,i}}{\alpha_{ii}} P(D_i \leq x_i, D_i + D_{i+1} \geq x_i + x_{i+1}) - \frac{\alpha_{i,i-1}}{\alpha_{ii}} P(D_i \geq x_i, D_{i-1} + D_i \leq x_{i-1} + x_i)\right]$$

After developing their more general model, BAA propose a successive approximation algorithm to find the optimal capacities. Their algorithm may be applied to our problem, but convergence can be slow, particularly when demand classes are strongly negatively correlated. To increase the rate of convergence we employ a standard technique called Steffenson extrapolation (Ortega

23

and Rheinboldt 1970, pg. 198). This method estimates the derivative of convergence using the three successive iterations. In what follows, $\underline{\eta}(\underline{x})$ will denote a vector $(\eta_1(\underline{x}),...,\eta_n(\underline{x}))$.

Algorithm
Step 1: $\quad\quad\quad x_i = x_i^{NV}$ *for all i,*

Step 2: for all i: $\quad x_i \leftarrow \eta_i(\underline{\eta}(\underline{x})) - \dfrac{\left|\eta_i(\underline{\eta}(\underline{x})) - \eta_i(\underline{x})\right|}{\eta_i(\underline{\eta}(\underline{x})) - 2*\eta_i(\underline{x}) + x_i}$, *repeat until convergence.*

While conducting the following numerical experiment, we found that this algorithm converges within 5-7 iterations for the majority of problem parameters. Note that there are other methods for determining the optimal capacities (see, for example, the discrete programming approach used by Fine and Freund, 1990), and we did not measure the performance of our algorithm against these alternatives.

We now consider a particular example, an automobile rental company with multiple classes of demand and capacity. First, assume that there are only two classes of demand ($i=1,2$) and two types of capacity ($j=1$ can serve either demand class while $j=2$ can serve only demand class 2). Table 1 presents the parameters for the first set of numerical experiments; note that demand class 2 has a larger mean and standard deviation than class 1.

Demand and capacity classes	Revenue (p_i)	Variable cost (V_j)	Penalty cost (C_i)	Marginal contribution (α_{ij})	Fixed cost (F_j)	μ_i	σ_i
$i=1, j=1$	\$42	\$18	\$12	\$36	\$20	120	50
$i=2, j=2$	\$35	\$10	\$7	\$32	\$18	200	80
$i=2, j=1$	\$35	\$18	\$7	\$24			

Table 1. Parameters for car rental example – two car classes.

We assume that demand follows a bivariate Normal distribution. To characterize the benefits of flexibility, we compare our optimal profits with profits generated by a firm that allows service substitution in the second stage (when demand is realized) but does not consider this flexibility in the first stage (when the capacity decision is made). In other words, we compare our results to the alternative in which the capacity is chosen at the news-vendor quantity, and the transportation problem is solved to optimally allocate this capacity.

Figure 4 shows how the optimal capacity decision varies with the coefficient of correlation, ρ. As has been shown in Proposition 3, the optimal capacity x_2^* is always below the news-vendor quantity and x_1^* is always above the news-vendor quantity. Also, in accordance with Proposition 7, x_2^* is increasing and x_1^* is decreasing in the coefficient of correlation.

The effect of varying the coefficient of correlation on the relative profitability of our procedure is shown in Figure 5. There is a significant increase in profitability when we account for flexibility in the capacity decision stage, as compared to the objective function value given the news-vendor solution. The profit increase is 20% when the demand classes are not correlated, and relative profits increase as correlation declines.
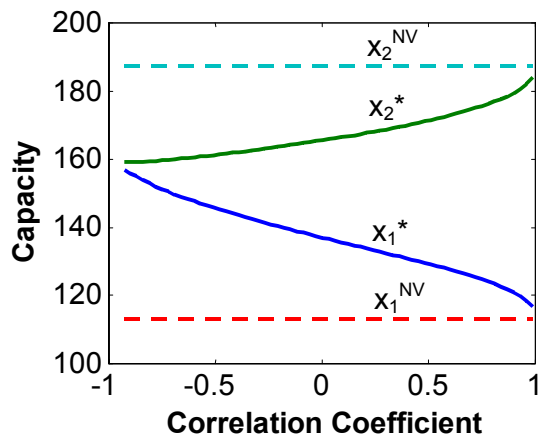

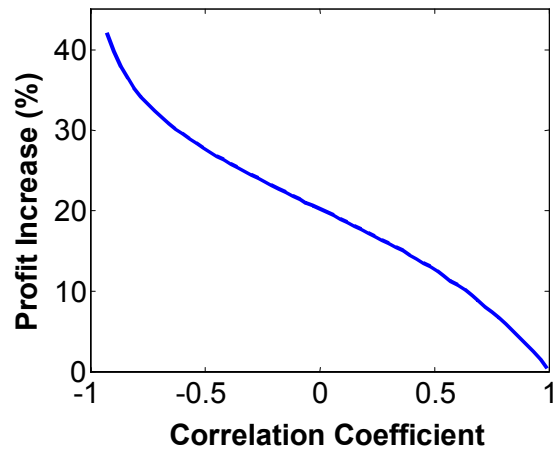Figure 4: Resource Capacities
as a function of correlation

Figure 5: Increase in profit
as a function of the correlation

To demonstrate that, as correlation rises, a decrease in capacity x_{i-1} and increase in capacity x_i is not the only possibility, we offer two examples. First, we construct a situation in which x_{i-1} and x_i increase (region II in Figure 3). As we demonstrated analytically, this may occur when $x_1^* + x_2^* > \mu_1 + \mu_2$, a condition that can be satisfied when fixed costs are low. We will therefore use $F_1 = 8$ and $F_2 = 7$ (all the other parameters shown in Table 1 are held constant).

The results are presented in Figure 6.

Now, we construct an example in which x_{i-1} and x_i decrease (region III in Figure 3). Since

this can be achieved only when $x_1^* + x_2^* < \mu_1 + \mu_2$, we shall increase fixed costs to $F_1 = 30$ and F_2

= 25, and leave all other parameters unchanged. The result is presented in Figure 7. Observe

that in both cases, non-intuitive responses by the optimal capacities occur only when correlation
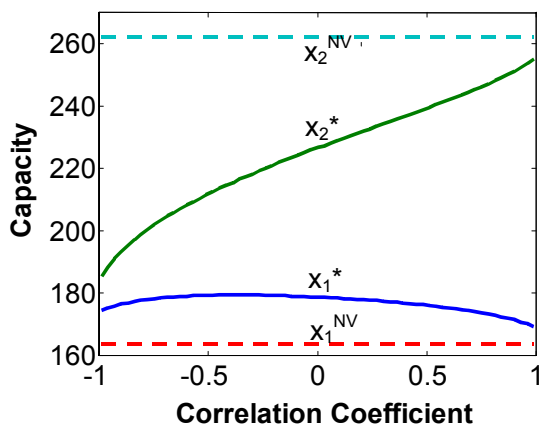
is close to -1.

Figure 6: Both optimal capacities
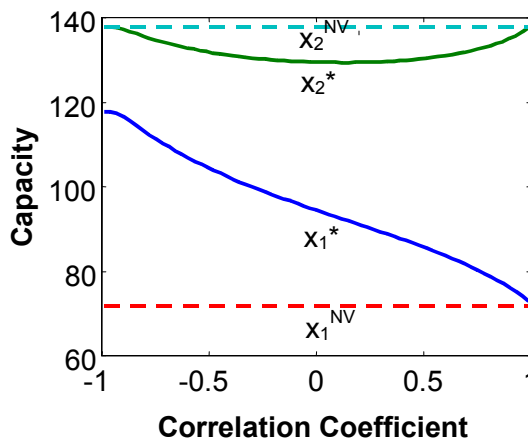may rise with correlation

Figure 7: Both optimal capacities
may fall with correlation

We now demonstrate that, as noted in section 3, the relationship between the optimal

capacity levels and the news-vendor quantities is not obvious when there are more than two

demand classes. In Table 2 we introduce a third car class.

Demand and capacity classes	Revenue (p_i)	Variable cost (V_j)	Penalty cost (C_i)	Marginal contribution (α_{ij})	Fixed cost (F_j)	μ_i	σ_i
$i=1, j=1$	$70	$40	$7	$37	$20	120	50
$i=2, j=2$	$50	$30	$5	$25	$15	165	80
$i=3, j=3$	$35	$20	$3	$18	$12	220	100
$i=2, j=1$	$50	$40	$5	$15			
$i=3, j=2$	$35	$30	$3	$8			

Table 2. Parameters for the automobile rental example – three car classes.

We again assume that demand classes follow a multivariate Normal distribution with correlation

26

coefficients $\rho_{23}=0$ and $\rho_{12}\in(-1,1)$. The resulting optimal capacities are presented in Figure 8. As has been noted earlier, the optimal capacity for class 2 can be more than, or less than, the corresponding news-vendor quantity. As predicted by Proposition 6, x_3^* rises as x_2^* falls, and x_3^* falls as x_2^* rises.
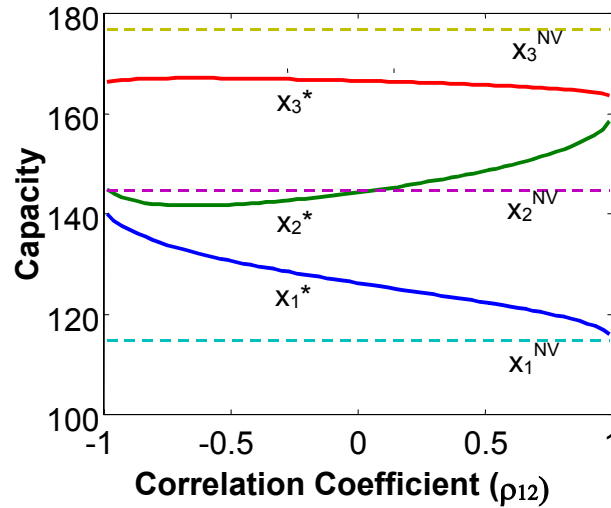

Figure 8: Optimal capacities for three resources
as a function of correlation coefficient ρ_{12}

Finally, we examine the impact of demand variability on the optimal capacities. Additional numerical experiments have demonstrated that as the standard deviation of demand for type i increases the optimal capacity for type i-1 increases. This observation is consistent with a result of Van Mieghem (1998), that as the level of uncertainty in the demand distribution increases, the optimal level of the flexible resource rises.

6. Conclusions and Future Research

By restricting service upgrades to at most one class, we have found a relatively simple method to determine the optimal capacity of flexible resources. While we have described the problem in terms of service delivery (automobile rental, flexible staffing), the method is also applicable to

inventory problems in which inventory may be used to satisfy multiple demand streams (see BAA). We compared our solution with the traditional news-vendor solution, and we found that a strategy in which higher (lower) capacity of more flexible (specialized) resources is an optimal strategy when there are two service classes. However, we found that this does not necessarily hold in the general case. We have also determined that a change in correlation between adjacent demand streams has an 'alternating' impact on the optimal capacity, for capacities of adjacent resources move in opposite directions. We then developed an efficient algorithm to find the optimal capacities and demonstrated these results with numerical experiments.

It is plausible to assume that our results can be generalized to probability distributions other than the Normal distribution and its elliptical cousins, and that similar results may be found for demand-capacity structures with upgrades to multiple levels. The results may also be employed within a dynamic program to solve this problem in a multi-period setting. This extension is particularly important for applications such as automobile rentals and technician staffing, in which resources may be consumed for more than one period.

Similar techniques may also be successfully applied to additional revenue management problems. For example, Karaesmen and Van Ryzin (1998) present a problem formulation for determining optimal overbooking levels that is similar to ours. However, their formulation allows a more general structure for substitution. It may be possible to derive analytical results, as well as additional intuition, by restricting substitution to one-level upgrades.

Appendix A

Proof of Lemma 1: We need to show that the second order partial derivatives of the objective function are all non-positive.

$$\frac{\partial^2 \Gamma}{\partial x_i \partial x_j} = 0 \quad \text{for all } i \text{ and } j < i\text{-}1, j > i+1 \tag{A.1}$$

$$\frac{\partial^2 \Gamma}{\partial x_i \partial x_{i-1}} = -\alpha_{i,i-1} \int_{x_i}^{\infty} f_{D_i,D_{i-1}+D_i}(t,x_{i-1}+x_i)dt < 0 \quad \text{for all } i = 2...n \tag{A.2}$$

$$\frac{\partial^2 \Gamma}{\partial x_i^2} = -\alpha_{ii} f_{D_i}(x_i) - \alpha_{i+1,i} \int_{-\infty}^{x_i} f_{D_i,D_i+D_{i+1}}(t,x_i+x_{i+1})dt + \alpha_{i+1,i} \int_{x_i+x_{i+1}}^{\infty} f_{D_i,D_i+D_{i+1}}(x_i,t)dt$$
$$- \alpha_{i,i-1} \int_{x_i}^{\infty} f_{D_i,D_i+D_{i-1}}(t,x_i+x_{i-1})dt + \alpha_{i,i-1} \int_{-\infty}^{x_i+x_{i-1}} f_{D_i,D_i+D_{i-1}}(x_i,t)dt \tag{A.3}$$

$$< -\alpha_{ii} f_{D_i}(x_i) + \alpha_{i+1,i} \int_{x_i+x_{i+1}}^{\infty} f_{D_i,D_i+D_{i+1}}(x_i,t)dt + \alpha_{i,i-1} \int_{-\infty}^{x_i+x_{i-1}} f_{D_i,D_i+D_{i-1}}(x_i,t)dt \tag{A.4}$$

$$< -\alpha_{ii} f_{D_i}(x_i) + \alpha_{i+1,i} \int_{-\infty}^{\infty} f_{D_i,D_i+D_{i+1}}(x_i,t)dt + \alpha_{i,i-1} \int_{-\infty}^{\infty} f_{D_i,D_i+D_{i-1}}(x_i,t)dt \tag{A.5}$$

$$= -\alpha_{ii} f_{D_i}(x_i) + \alpha_{i+1,i} f_{D_i}(x_i) + \alpha_{i,i-1} f_{D_i}(x_i) \tag{A.6}$$

$$= f_{D_i}(x_i)(-\alpha_{ii} + \alpha_{i+1,i} + \alpha_{i,i-1}) = f_{D_i}(x_i)(p_{i+1} - V_{i-1} + C_{i+1}) = f_{D_i}(x_i)(\alpha_{i+1,i-1}) < 0 \tag{A.7}$$

where the last inequality is a result of assumption (iii) of (2). ∎

Proof of Lemma 2: In our case diagonal dominance means:

$$\left| \frac{\partial^2 \Gamma}{\partial x_i^2} \right| > \left| \frac{\partial^2 \Gamma}{\partial x_i \partial x_{i-1}} \right| + \left| \frac{\partial^2 \Gamma}{\partial x_i \partial x_{i+1}} \right|. \tag{A.8}$$

After writing out terms and employing manipulations similar to equations (A.4) through (A.7), we find that

$$\left| \frac{\partial^2 \Gamma}{\partial x_i^2} \right| - \left(\left| \frac{\partial^2 \Gamma}{\partial x_i \partial x_{i-1}} \right| + \left| \frac{\partial^2 \Gamma}{\partial x_i \partial x_{i+1}} \right| \right) > -f_{D_i}(x_i)(\alpha_{i+1,i-1}) > 0. \ \blacksquare \tag{A.9}$$

Appendix B

To simplify the notation, let $\rho_{i-1,i} \equiv \rho$. Before proving Proposition 8, we will first prove the following lemma:

<u>Lemma 4</u>. Suppose demand is Normally distributed. Then:

(i)   if $\mu_{i-1} + \mu_i < x^*_{i-1} + x^*_i$ and either $\rho < -\sigma_i / \sigma_{i-1}$ or $\rho > -\sigma_{i-1} / \sigma_i$, then $\partial P_2 / \partial \rho < 0$,

(ii)   if $\mu_{i-1} + \mu_i > x^*_{i-1} + x^*_i$ and either $\rho < -\sigma_{i-1} / \sigma_i$ or $\rho > -\sigma_i / \sigma_{i-1}$, then $\partial P_1 / \partial \rho < 0$.

Proof: We will prove (i) only, and the proof of (ii) is analogous. We will now show that, given $\mu_{i-1} + \mu_i < x^*_{i-1} + x^*_i$ and $\rho > -\sigma_{i-1} / \sigma_i$, $\partial P_2 / \partial \rho < 0$. We will assume that $i=2$, and therefore $\rho \equiv \rho_{i,i-1} \equiv \rho_{12}$ (the proof is identical for all $i = 3 \ldots n$). Our challenge is to apply Slepian's inequality to the probability term of interest, $P_2 = P(D_2 \geq x_2, D_1 + D_2 \leq x_1 + x_2)$.

The joint density function of D_1 and D_2 is the bivariate Normal distribution with mean (μ_1, μ_2), correlation coefficient ρ. Define random variables

$$(Z_1, Z_2) = \left(\frac{-D_2 + \mu_2}{\sigma_2}, \frac{D_1 + D_2 - \mu_1 - \mu_2}{\sqrt{\sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2}} \right). \tag{B.1}$$

The pair (Z_1, Z_2) is distributed according to the standard bivariate Normal distribution with correlation coefficient $r(\rho) \equiv (-\sigma_2 - \rho\sigma_1) / \sqrt{\sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2}$. Let $f_{Z_1,Z_2}(z_1, z_2 \mid r(\rho))$ be its density function and define

$$a_1 = \frac{-x_2 + \mu_2}{\sigma_2} \quad \text{and} \quad a_2(\rho) = \frac{x_1 + x_2 - \mu_1 - \mu_2}{\sqrt{\sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2}}. \tag{B.2}$$

We examine the total derivative:

$$\frac{\partial P(D_2 \geq x_2, D_1 + D_2 \leq x_1 + x_2)}{\partial \rho} = \frac{\partial P(Z_1 \leq a_1, Z_2 \leq a_2(\rho))}{\partial \rho}$$

$$= \frac{\partial}{\partial \rho}\left[\int_{-\infty}^{a_2(\rho)} \int_{-\infty}^{a_1} f_{Z_1,Z_2}(z_1, z_2 \mid r(\rho)) dz_1 dz_2\right]$$

$$= \frac{\partial a_2(\rho)}{\partial \rho} \cdot \int_{-\infty}^{a_1} f_{Z_1,Z_2}(z_1, a_2(\rho) \mid r(\rho)) dz_1 + \int_{-\infty}^{a_2(\rho)} \int_{-\infty}^{a_1} \left[\frac{\partial f_{Z_1,Z_2}(z_1, z_2 \mid r(\rho))}{\partial \rho}\right] dz_1 dz_2$$

(B.3)

By the Chain Rule, the second term of (B.3) is

$$\int_{-\infty}^{a_2(\rho)} \int_{-\infty}^{a_1} \frac{\partial r(\rho)}{\partial \rho} \cdot \frac{\partial f_{Z_1,Z_2}(z_1, z_2 \mid r(\rho))}{\partial r(\rho)} dz_1 dz_2 = \frac{\partial r(\rho)}{\partial \rho} \cdot \frac{\partial P(Z_1 \leq a_1, Z_2 \leq a_2(\rho))}{\partial r(\rho)}$$

(B.4)

where $a_2(\rho)$ is taken as a constant in the final derivative of (B.4). By taking derivatives and

applying some algebra we find that:

$$\frac{\partial a_2(\rho)}{\partial \rho} < 0 \quad \text{iff } x_1 + x_2 > \mu_1 + \mu_2 \text{ and } \frac{\partial r(\rho)}{\partial \rho} < 0 \quad \text{iff } \rho > -\frac{\sigma_1}{\sigma_2}.$$

(B.5)

In addition, from Slepian's inequality,

$$\frac{\partial P(Z_1 \leq a_1, Z_2 \leq a_2(\rho))}{\partial r(\rho)} > 0.$$

(B.6)

Given the assumptions of the lemma, the sufficient conditions (B.5) are true, so that expression

(B.3) is strictly negative. Therefore, the expression $P_2 = P(D_2 \geq x_2, D_1 + D_2 \leq x_1 + x_2)$ is decreasing

in the coefficient of correlation ρ. Further, one can verify that $\partial P_1 / \partial \rho > 0$ for $\rho < -\sigma_i / \sigma_{i-1}$.

Since $\partial P_1 / \partial \rho + \partial P_2 / \partial \rho < 0$ (see the proof of Lemma 3), it follows that $\partial P_2 / \partial \rho < 0$ holds also for

$\rho < -\sigma_i / \sigma_{i-1}$. This completes the proof of Lemma 4. ∎

Proof of Proposition 8: We will prove statement (i) only and statement (ii) can be proven in a

similar manner. Using equation (13) in the proof of Proposition 6, Lemma 3, and Lemma 4 we

write:

$$H_{i-1,i}\frac{\partial x_{i-1}^*}{\partial\rho} + H_{i,i}\frac{\partial x_i^*}{\partial\rho} + H_{i,i+1}\frac{\partial x_{i+1}^*}{\partial\rho} = -\frac{\partial\Omega_i}{\partial\rho} = \alpha_{i-1,i}\frac{\partial P_2}{\partial\rho} < 0 \qquad\text{(B.7)}$$

The rest of the proof is by contradiction. Assume $\partial x_i^*/\partial\rho < 0$. Then, by Propositions 7,

$\partial x_{i-1}^*/\partial\rho < 0$. Finally, using Proposition 6 and by invoking diagonal dominance (Lemma 2), it is

straightforward to verify that (B.7) can not hold. This is a contradiction and hence $\partial x_i^*/\partial\rho > 0$. ■

Acknowledgements

References

Bassok, Y., R. Anupindi and R. Akella. 1999. Single-period multiproduct inventory models
with substitution. *Operations Research,* Vol. 47, 632-642.

Carroll, W. J. and Grimes, R. C. 1995. Evolutionary change in product management:
Experiences in the car rental industry. *Interfaces,* Vol. 25, 84-104.

Eppen, G. D. 1979. Effects of centralization on expected costs in a multi-location newsboy
problem. *Management Science,* Vol. 25, 498-501.

Eynan, A. 1996. The impact of demands' correlation on the effectiveness of component
commonality. *Int. J. Prod. Res.,* Vol. 34, No. 6, 1581-1602.

Fine, C. H. and R. M. Freund. 1990. Optimal investment in product-flexible manufacturing
capacity. *Management Science,* Vol. 36, 449-466.

Green, L. 1985. A queuing system with general-use and limited-use servers. *Operations
Research,* Vol. 33, 168-182.

Gans, N. and Y. Zhou. 1999. Managing Learning and Turnover in Employee Staffing. Working
Paper, Operations and Information Management Department, The Wharton School,

University of Pennsylvania, Philadelphia, PA, 19104.

Harrison, J. M. and M. J. Lopez. 1999. Heavy traffic resource pooling in parallel-server systems. Unpublished working paper, Stanford University Graduate School of Business. Available at http://faculty-gsb.stanford.edu/harrison/Documents/parallel.pdf.

Harrison, J. M. and J.A.Van Mieghem. 1999. Multi-resource investment strategies: operational hedging under demand uncertainty. *European Journal of Operational Research*, Vol. 113, 17-29.

Hoffman, A. 1963. On simple linear programming problems. Convexity: proceedings of symposia in pure mathematics. Vol. 7.

Horn A. R. and C. A. Johnson. 1985. *Matrix analysis*. Cambridge University Press.

Hsu, A., and Y. Bassok. 1999. Random yield and random demand in a production system with downward substitution. *Operations Research*, Vol. 47, 277-290.

Jordan, W. C. and S. C. Graves. 1995. On the principles of the benefits of manufacturing process flexibility. *Management Science*, Vol. 41, 577-594.

Karaesmen, I., G. and van Ryzin. 1998. Overbooking with substitutable inventory classes. Working paper, Columbia University. Available at http://www.columbia.edu/~gjv1/subob1.PDF.

Keskinocak P. 1999. Corporate high flyers. *ORMS Today,* Vol. 26, No. 6, 22-29.

Khouja, M. 1999. The single-period (news-vendor) problem: literature review and suggestions for further research. *Omega*, Vol. 27, 537-553.

McGill, J. I. and G. J. van Ryzin. 1999. Revenue management: research overview and prospects. *Transportation Science*, Vol. 33, 233-256.

McGillivray, A. R. and E. A. Silver. 1978. Some concepts for inventory control under substitutable demand. *INFOR*, Vol. 16, No. 1, 47-63.

Ortega, J. M. and W. C. Rheinboldt. 1970. *Iterative solution of nonlinear equations in several variables*. Academic Press.

Pasternak, B. A. and Z. Drezner. 1991. Optimal inventory policies for substitutable commodities with stochastic demand. *Naval Research Logistics*, Vol. 38, 221-240.

Patel, J. K. and C. B. Read. 1996. *Handbook of the Normal distribution*. Marcel Dekker.

Rudi, N. and S. Netessine. 1999. Centralized and competitive inventory models with demand substitution. Simon School working paper OP 99-04, University of Rochester. Available at http://www.netessine.com

Shumsky, R. A. 1999. Approximation and analysis of a queueing system with flexible and specialized servers. Simon School Working Paper OP 99-02, University of Rochester. Available at http://omg.ssb.rochester.edu/omgHOME/shumsky/flex_serv.pdf

Tong, Y. L. 1980. *Probability inequalities in multivariate distributions*. Academic Press.

Topkis, D. M. 1978. Minimizing a submodular function on a lattice. *Operations Research*, Vol. 26, 305-321.

Van Mieghem, J. A. 1997. Investment strategies for flexible resources. Technical report, Center for Mathematical Studies in Economics and Management Science, Northwestern University. Available at http://www.kellogg.nwu.edu/faculty/VanMieghem/htm/flex0tec.pdf.

Van Mieghem, J. A. 1998. Investment strategies for flexible resources. *Management Science*, Vol. 44, 1071-1078.