Capacity and Backlog Management in Service-Oriented Supply Chains

EDWARD G. ANDERSON JR.

McCombs School of Business
The University of Texas at Austin
Management Department, CBA 4.202
Austin, Texas 78712
Phone: (512) 471-6394
Fax: (512) 471-3937

DOUGLAS J. MORRICE

McCombs School of Business
The University of Texas at Austin
MSIS Department, CBA 5.202
Austin, Texas 78712-1175
Phone: (512) 471-7857
Fax: (512) 471-0587

Revised October 9, 2001

Abstract

We investigate the dynamic behavior of service-oriented supply chains by developing a two-stage serial capacity management model. Reflecting the reality of many service (and custom manufacturing) supply chains, each stage holds no finished goods inventory, but rather only backlogs that can be managed solely by adjusting capacity. Using control theory, we develop an optimal policy that trades off backlog costs against capacity adjustment costs when information is shared. We then establish the potential for an increase in demand variability along an optimally managed supply chain. Contrary to conventional wisdom, we also show that, while lead-time reduction does generally *reduce* backlog variance, it also *increases* capacity variance, resulting in a trade-off between service quality and personnel costs at each stage. Furthermore, such lead-time reductions *increase* backlog variances at subsequent stages resulting in a service quality trade-off between stages. Finally, we show that sharing backlog information will not materially improve overall supply chain performance if the target lead- and capacity adjustment times of the stage closest to end-customer demand are much smaller than subsequent stages'.

1 Introduction

In this paper, we focus on the issue of capacity management in supply chains that hold no finished goods inventory such as those typically associated with many service industries. Without finished goods inventories, customer order backlogs can only be managed by adjusting the level of resources required to process them such as labor and capital equipment. For example, Figure 1 contains a simplified overview of a mortgage services supply chain from Anderson and Morrice (2000). Once generated, a customer order (mortgage application) queues up at each stage (credit check, appraisal and surveying, and title check) for service because no service is—or can be—prepared in advance of the order's arrival at each stage. Many other services including insurance, consulting, professional services, and health care possess a similar process structure.

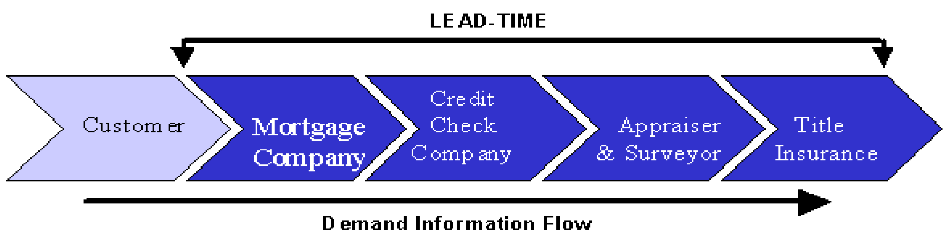

Figure 1: Mortgage Services Supply Chain

Because these supply chains are so typical of the service industry, we will call them "service-oriented" supply chains. However, such supply chains are not limited to the service sector. For example, custom manufactured products produced by capital equipment manufacturers such as Cincinnati Milacron or Applied Materials are similar to services because production may only begin after an order including specifications for both final product and supplier components has been placed. Furthermore, with recent advances in technology and increased competition, many other industries are choosing to design supply chains that hold no finished goods inventory. For example, companies such as Dell Computer Corporation use direct sales and mass customization to achieve supply chains that contain virtually no finished goods inventory (Dell and Magretta 1998). On the other hand, the supply chains of some service industries such as fast foods are essentially inventory supply chains more similar to that of automobile production than to those in the mortgage, consulting, or insurance industries.

In this paper, we develop a capacity management model for a two-stage serial service-oriented supply chain, based on the System Dynamics Modeling approach, using differential equations to mimic the decision rules and other dynamic behavior found in real service-oriented supply chains. (Forrester 1958; Sterman 2000). At each stage in the supply chain, our model relates capacity, processing, backlog, and lead-times to capture the aggregate dynamic interactions between the different stages. The model builds on Anderson and Morrice (2000), which examine service-oriented supply chains through empirical and simulation analysis. Similar approaches are used in Oliva (1996) to study service quality erosion in a single firm, Anderson (2001) to analyze staff planning problems in the face of non-stationary demand, and Cohen and Whang (1997) to explore pricing of after-sales service with a game-theoretic model.

Using control theory and signal analysis techniques, we formulate and solve an optimal control problem that minimizes backlog and capacity adjustment costs. The resulting control rule is intuitive and relatively easy to implement. Backlog management is critical to service firms because waiting time length and variation can have a highly negative impact on service quality (Fitzsimmons and Fitzsimmons 1998) that can in turn depress long-term profits (Heskett et al. 1997). However, adjusting capacity in services—primarily by hiring, training, and firing personnel—is also quite expensive (Anderson 2001). Based on the resulting optimal control rule, we employ a one-stage model with a non-stationary input in the form of a step change to establish the potential for a "bullwhip effect" (Forrester 1958, Lee et al. 1997), i.e., an increase in demand variation along the supply chain (order rates in inventory supply chains, processing rates in service-oriented supply chains). We then study the impact of various management strategies such as target lead-time reduction, capacity adjustment time reduction, and sharing first stage information (demand, backlog, and capacity) on backlog variance (i.e., the variance of the number of units of backlog) and capacity variance (i.e., variance of the number of units of capacity) in a serial two-stage model under stochastic (Gaussian white noise) end-customer demand.

The dimensions of service quality include reliability, responsiveness, assurance, empathy, and tangibles. Reliability, the most important dimension, is the customer expectation that the service is completed, *every time, on time*, in the same manner, and without errors (Parsuraman et al. 1988). We focus on the reliability and measure it using backlog variance. Conventional wisdom, largely based on the inventory management literature, strongly supports lead-time reduction in order to mitigate the bullwhip effect. We show that lead-time reduction can exacerbate the bullwhip effect if it is not carefully coordinated with capacity adjustment. In particular, lead-time reduction generally *reduces* backlog variance but *increases* the capacity variance within a given stage resulting a trade-off between service quality costs and net hiring costs. Furthermore, the same change generally *increases* backlog variances at higher stages resulting in a service quality trade-off at different stages of the supply chain. Finally, lead time *reduction* at the first stage also often *increases* capacity adjustment costs at higher stages, thus inducing overall performance deterioration. Conventional wisdom also suggests that sharing first stage information should help minimize capacity and backlog variance at the second stage of the supply chain. In general, we show that sharing this information does improve second stage performance as measured by the objective function of the optimal control rule. We quantify the value of the information and characterize under what circumstances information is most beneficial.

There is a growing body of research on managing the dynamics (e.g. the bullwhip effect) of finished goods inventory in supply chains (Lee et al. 1997; Lee et al. 2000; Chen et al. 2000; Chen et al. 1997; Baganha and Cohen 1998). However, little research exists on managing supply chain dynamics in the absence of finished goods inventory. Some related work exists in the queuing theory literature on tandem queues. Rosberg et al. (1982) develop a model for optimally controlling first stage service capacity in a tandem queue by selecting the first stage service rate as a function of the number of customers at both stages in order to minimize waiting costs at the two stations. Chen et al. (1994) extends this model to include different job types and their arrival processes as controlling factors. However, these papers assume that the total capacity shared between the two stages is fixed and may be transferred without cost between stages. In this paper, we develop optimal control policies for both stages using a broader objective function that trades off backlog costs against non-zero capacity adjustment costs. Related work also exists in the job shop literature, particularly Graves (1986), which examines the problem of setting desired lead-times in a job shop using a control-theoretic approach. Tang (1990) and Gong and Matsuo (1997) extend Grave's work to account respectively for production variability and an explicit objective function. These papers differ from the current paper because we (1) focus on issues more relevant to services supply chains rather than job shops, particularly the effect of one stage's policy upon other stages' performance, (2) assume a capacity constraint and a cost associated with changing capacity, and (3) develop an optimal rather than heuristic control policy. (Gong and Matsuo is an exception to (3) because they develop an optimal policy that considers deviations from a target production rate, which is important in a production setting. We instead consider hiring, training, firing, and other capacity adjustment costs, which are more significant in services.)

Section 2 contains an optimal control problem formulation and solution for a two-stage service-oriented supply chain. Section 3 quantifies the bullwhip effect for a single stage model with non-stationary (a step change) demand and develops analytical expressions for backlog, capacity, and processing variances with stochastic (white noise) demand. Section 4 extends the stochastic demand results to a two-stage model and discusses the impact on capacity and processing variances of sharing backlog information and reducing lead- and capacity adjustment times. Section 5 discusses managerial insights and future research directions.

2 Model Logic

Figure 2 contains a generic 2-stage serial supply chain network. At each stage, demand feeds backlog, and capacity processes backlog, which creates demand for the next stage. We consider two types of end-customer demand: a deterministic non-stationary step function, and an independent and identically distributed (iid) random sequence. Many marketing and economic studies suggest that the nonstationarity of mean demand is important (Lilien et al. 1992). We use one of the most basic of nonstationary demand forms, the step function, to establish the bullwhip effect. Additionally, we characterize the dynamic behavior resulting from this demand signal because of its similarity to the demand stream in the famous MIT Beer Game (Sterman 1989) and the mortgage services game in Anderson and Morrice (2000). The iid random sequence permits analysis of the dynamic behavior of the two-stage supply chain with stochastic demand. Although more complex stochastic demand models could be considered (e.g. Graves 1999), the iid model is sufficient to provide novel and interesting insights concerning supply chain dynamics in services (and has been used numerous times for similar purposes in the inventory supply chain literature, see for recent examples Baganha and Cohen 1998 and Lee and Whang 1999.)

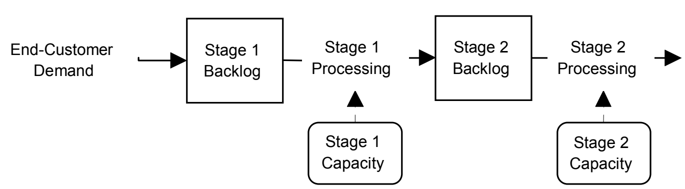

Figure 2: Service-Oriented Supply Chain

The optimal control problem for capacity and backlog management is formulated and then solved for each stage. Stage 1 selects an optimal control policy independent of Stage 2 objectives. However, because stage 1 processing drives stage 2 demand, optimal policy at stage 2 is conditioned on the stage 1's results.

2.1 The General Closed-Loop Problem

Let the index $t \geq 0$ represent time and $i\,(i = 0,1,2)$ be the index denoting the stage in the supply chain. The stage $i = 0$ is the end customer. Because the optimal solution to a step change in demand is an almost-trivial derivation of that for iid demand (see Section 3.1), we model the latter case first. In particular, we model demand as the continuous equivalent of an iid normal random demand process, often termed Gaussian white noise.

Let: $r(t)$ = end-customer demand, a Gaussian, white noise process at time *t.*
σ_r = the standard deviation of *r(t).*
$\overline{r}$ = the expected value of *r(t), a constant.*

Hence,

$$r(t) \sim Norm\left(\overline{r}, \sigma_r^2\right) \qquad \text{for } t \geq 0 \tag{2.1}$$

$$\text{cov}\left[r(t), r(t+\varepsilon)\right] = 0 \text{ for } \varepsilon \neq 0 \quad \text{for } t \geq 0 \tag{2.2}$$

Let $C_i(t)$ = the capacity at stage *i* at time *t*. The productivity of $C_i(t)$ is assumed to be unity.
$P_i(t)$ = the processing rate at stage *i* at time *t.* $P_i(t) \leq C_i(t)$ for all $t \geq 0$.
$W_i(t)$ = the net rate of change in capacity (or *net hiring*) at stage *i* at time *t.*
$B_i(t)$ = the backlog at stage *i* at time *t*
$\tilde{B}_i$ = the target backlog at stage *i.*
φ_i^2 = the fractional cost of excess or insufficient backlog at stage *i*.
ψ_i^2 = the fractional cost of changing capacity at stage *i.*
ρ_i = the discount rate at stage *i.*

To obtain a reasonable employment policy at each stage, we look forward from the present time $t = 0$ seeking to balance backlogs against capacity adjustment costs. This defines the following dynamic program at each stage.

Closed-Loop Optimal Control Problem *$S_i[0, B_1(0), B_2(0), C_1(0), C_2(0)]$*:

Assume $W_i(t)$ is continuous for i=1,2 and t≥0. Given the initial backlogs $B_i(0)$ and capacities $C_i(0)$ for i=1,2 at time t=0:

$$Z_i = \min_{W_i(t)} E\int_0^{\infty} \exp(-\rho_i t)\left\{\phi_i^2\left[\frac{B_i(t)-\tilde{B}_i}{\tilde{B}_i}\right]^2 + \psi_i^2\left[\frac{W_i(t)}{E\left[C_i(t)\right]}\right]^2\right\}dt \tag{2.3}$$

subject to

$$\frac{dB_i(t)}{dt} = P_{i-1}(t) - P_i(t) \quad \text{for } t\geq 0 \tag{2.4}$$

$$P_i(t) = \begin{cases} C_i(t) & \text{if } B_i(t) > 0 \\ \min\left(P_{i-1}(t), C_i(t)\right) & \text{otherwise} \end{cases} \quad \text{for } t\geq 0 \tag{2.5}$$

$$\frac{dC_i(t)}{dt} = W_i(t) \quad \text{for } t\geq 0 \tag{2.6}$$

$$B_1(t), C_1(t) \geq 0 \quad \text{for } t\geq 0 \tag{2.7}$$

where $P_0(t) = r(t)$. By "closed-loop" we indicate that the determination of the value of all control variables $W_i(t)$ will incorporate the values of $B_i(t)$, $C_i(t)$, and all *r(u)* for *0≤u≤t*.

Fundamentally, the objective function in (2.3) trades off the net present value of smoothing the firm's backlog against smoothing its capacity from the present infinitely far into the future. In the operations literature, smoothing is often seen as an inherently desirable goal (e.g. Graves 1986). At a more intuitive level, (2.3) quadratically weights the difference between actual backlog and some positive, non-zero target because excess backlog—by Little's Law—necessitates longer delivery lead times to customers if the average processing rate is stationary while insufficient backlog starves some employees of work. We have talked to many service companies and make-to-order manufacturers such as Dell and Applied Materials and most prefer to maintain a non-negative target backlog at all times to keep all employees utilized. For similar reasons, Holt et al. (1960 p. 58) quadratically control around a target nonnegative inventory in a similar manner.

Changing employee capacity is also problematic. Hence, (2.3) also quadratically weights net hiring as a fraction of the average capacity. Quadratic weighting is common in many capacity-planning problems (e.g. Holt et al. 1960) for a number of reasons. Hiring is expensive because of search and training costs. Firing necessitates severance pay and often leads to loss of morale. Furthermore, the problems inherent in hiring and firing tend to escalate in a concave manner. For example, absorbing, training, and acculturating five percent new hires per year is much simpler than fifty percent (Brooks 1982). Similarly, a small workforce reduction can often be accomplished through attrition, avoiding the morale losses induced by firing (Kofman et al. 1994).

Constraint (2.4) represents the change in stage i backlog, which ordinarily increases with demand and decreases as the capacity at stage i works the backlog off. However, processing cannot exceed local demand without some backlog, which is represented by constraint (2.5). Constraint (2.6) represents the rate of change of stage i capacity. The rationality constraints in (2.7) prohibit negative backlogs and employees.

2.2 Linear Relaxation of the Closed Loop Problem

As stated earlier, many service and make-to-order firms manage backlog so that it very rarely, if ever, reaches zero in order to avoid idling workers. If we follow this logic in problem S_i by assuming that backlog is always nonnegative, the processing rate $P_i(t)$ and the capacity $C_i(t)$ at each stage must always be identically equal (and nonnegative) for $t \geq 0$. Under these assumptions, we can rewrite S_i for $i=1,2$ as:

$$Z_i = \min_{W_i(t)} E\int_0^{\infty} \exp(-\rho_i t)\left\{\phi_i^2\left[\frac{B_i(t)-\tilde{B}_i}{\tilde{B}_i}\right]^2 + \psi_i^2\left[\frac{W_i(t)}{E\left[C_i(t)\right]}\right]^2\right\}dt \qquad (2.8)$$

subject to

$$\frac{dB_i(t)}{dt} = C_{i-1}(t) - C_i(t) \qquad \text{for } t\geq 0 \quad (2.9)$$

$$\frac{dC_i(t)}{dt} = W_i(t) \qquad \text{for } t\geq 0 \quad (2.10)$$

where $C_0(t) = r(t)$ and $W_i(t)$ is continuous *for* t≥0.

Furthermore, it is a well-known result that the optimal solution to $W_i(t)$ in control problems of this form are linear in the state variables (Sethi and Thompson 2000, p. 85). Hence, the expected value of the capacity at each stage $C_i(t)$ will be identically equal to the expected demand $\bar{r}$ for *t≥0*. If we then define

$$\lambda_i \triangleq \frac{\tilde{B}_i}{\bar{r}}, \tag{2.11}$$

so that λ_i represents a "target" number of periods of backlog or *target lead-time*, then we can rewrite (2.8) as

$$Z_i = \min_{W_i(t)} E \int_0^{\infty} \exp(-\rho_i t) \left\{ \phi_i^2 \left[\frac{B_i(t) - \lambda_i \bar{r}}{\lambda_i \bar{r}} \right]^2 + \psi_i^2 \left[\frac{W_i(t)}{\bar{r}} \right]^2 \right\} dt\ . \tag{2.12}$$

The objective function in (2.12) balances the *proportionate* excess (or shortfall) of backlog against the *proportionate* hiring or firing rate at each instant. Hence, 2 days of excess backlog will create a greater relative penalty for a target lead-time of 1 week than for 12 weeks. Similarly, hiring and training 1 worker is more problematic for a firm with 10 workers than with a 1000.

Without loss of generality we can set $\psi_i = 1$ in (2.12) and redefine ϕ_i as a relative cost of backlog versus hiring and firing. Additionally, $\bar{r}$ can be dropped from (2.12) because it is a constant factor that appears in the denominator of both terms of the integrand. Hence, the entire problem S_i can be rewritten in final form as:

Final Form of Closed-Loop Optimal Control Problem $S_i[0, B_1(0), B_2(0), C_1(0), C_2(0)]$:

$$Z_i = \min_{W_i(t)} E \int_0^{\infty} \exp(-\rho_i t) \left\{ \frac{\phi_i^2}{\lambda_i^2} \left[B_i(t) - \lambda_i \bar{r} \right]^2 + \left[W_i(t) \right]^2 \right\} dt \tag{2.13}$$

subject to (2.9) and (2.10), and assuming $W_i(t)$ is continuous for all *t*.

2.3 Certainty-Equivalent Control Policies

The separability theorem of linear control theory ensures that, for a linear control problem with a linear-quadratic penalty function and a wide-sense stationary input, one can separate the problems of forecasting and control. This is done by solving for the optimal forecast (in a minimum-mean square error sense) directly without concern for the control policy, and by treating the resulting forecast as a *deterministic* input to a certainty-equivalent control problem. Under these conditions, the resulting policy will be optimal (Stengel 1994, p.448). The same approach to a single-echelon staffing planning problem was used in Anderson (2001).

Let $\hat{C}_0(t,u)$ = the optimal forecast of end-customer demand *r(t)* at time *t+u* determined at time *t,*

$\hat{C}_1(t,u)$ = the optimal forecast of stage 1 capacity $C_1(t,u)$ (i.e. the local demand seen at stage 2) at time *t+u* determined at time *t.*

$X(t)$ = the state vector at time *t*, i.e., $\left[C_1(t), B_1(t), C_2(t), B_2(t)\right]'$.

X(t,u) = the future state vector at time *t+u,* i.e., $\left[C_1(t,u), B_1(t,u), C_2(t,u), B_2(t,u)\right]'$ and

V(t,u) = the future control vector at time *t+u,* i.e., $\left[W_1(t,u), W_2(t,u)\right]'$ where

$C_i(t,u)$ = the stage i capacity at time *t+u* given a policy determined at time *t,*

$B_i(t,u)$= the stage i backlog at time *t+u* given a policy determined at time *t.*

$W_i(t,u)$= the net stage *i* hiring at time *t+u* given a policy determined at time *t.*

If the optimal forecast of the demand seen at stage *i*, that is $\hat{C}_{i-1}(t,u)$, is substituted for future demand $C_{i-1}(t,u)$ in the closed-loop problem $S_i[t,X(0)]$, the following certainty equivalent control problem $R_i[t,X(t,0)]$ results:

Certainty Equivalent Control Problem $R_i[t,X(t,0)]$:

Given the initial state vector X(t,0):

$$\min_{W_i(t,u)} E\int_0^\infty \exp\left[-\rho_i(t+u)\right]\left\{\frac{\phi_i^2}{\lambda_i^2}\left[B_i(t,u)-\lambda_i\bar{r}\right]^2+\left[W_i(t,u)\right]^2\right\}du \qquad (2.14)$$

subject to

$$\frac{dB_i(t,u)}{dt}=\hat{C}_{i-1}(t,u)-C_i(t,u) \qquad \text{for } t,u\geq 0 \quad (2.15)$$

$$\frac{dC_i(t,u)}{dt} = W_i(t,u) \qquad \text{for } t,u\geq 0. \quad (2.16)$$

Let $X_i^*(t)$ = the optimal state vector at time t with respect to the closed-loop problem $S_i(0,X(0))$ for any $t\geq 0$,

$X_i^*(t,u)$ = the optimal future state vector at time $t+u$ with respect to the certainty-equivalent problem $R_i(t,X(t,0))$ for any $t\geq 0$,

$V_i^*(t)$ = the optimal control vector at time t with respect to the closed-loop problem $S_i(0,X(0))$ for any $t\geq 0$,

$V_i^*(t,u)$ = the optimal control vector at time $t+u$ with respect to the certainty-equivalent problem $R_i(t,X(t,0))$ for any $t\geq 0$.

Proposition 2.1 (equivalence between closed-loop and certainty-equivalent solutions): Assume that all elements of the control vector $V_i(t)$ are finite *for* $t\geq 0$. Then, for $t\geq 0$, the solution to the closed-loop problem $S_i(0,X(0))$ will be equivalent to the solution for the certainty-equivalent problem $R_i(t,X(t,0))$ evaluated when $u=0$. That is,

$$X_i^*(t)=X_i^*(t,0) \qquad \text{for } t\geq 0 \qquad (2.17)$$

$$V_i^*(t)=V_i^*(t,0) \qquad \text{for } t\geq 0 \qquad (2.18)$$

Furthermore,

$$Z_i = E\int_0^{\infty} \exp(-\rho_i t)\left\{\frac{\phi_i^2}{\lambda_i^2}\left[B_i(t)-\lambda_i \overline{r}\right]^2 + \left[W_i(t)\right]^2\right\}dt\Bigg|_{X(t)=X_i^*(t,0),V(t)=V_i^*(t,0)} \qquad \text{for } t\geq 0. \; (2.19)$$

Thus, if we can find the control law that is optimal for the certainty equivalent problem $R_i(t,X(t,0))$ with control vector $V_i(t,0)$, then we have found the optimal control law for the closed-loop problem $S(t,X(0))$ with control vector $V_i(t)$. The proof for Proposition 2.1 (and all other non-trivial proofs) is in the appendix.

2.4 Stage 1 Control Law

Examining problem S_1, it is clear that the states $B_2(t)$ and $C_2(t)$ are irrelevant to its solution. Hence, we need only solve the forecast problem for future end-customer demand $\hat{C}_0(t,u)$, that is the forecast for future values of $r(t)$.

Proposition 2.2 (MMSE forecast of Stage 1 demand): If end-customer demand *r(t)* is of the form stated in (2.1)-(2.2), then its minimum mean square error forecast $\hat{C}_0(t,u)$ will be $\bar{r}$.

Proposition 2.3 (solution to the Closed-Loop Problem *$S_1[t,X(0)]$*): Under the conditions given in the closed-loop dynamic problem $S_1[t,X(0)]$), the optimal net hiring is:

$$W_1(t) = (\delta_{11} + \delta_{12})\left[\bar{r} - C_1(t)\right] + \delta_{11}\delta_{12}\left[B_1(t) - \lambda_1\bar{r}\right] \qquad \text{for } t \geq 0 \quad (2.20)$$

where $$\delta_{11} \triangleq -\frac{1}{2}\left(\rho_1 - \sqrt{\rho_1^2 - 4\frac{\phi_1}{\lambda_1}i}\right), \delta_{12} \triangleq -\frac{1}{2}\left(\rho_1 - \sqrt{\rho_1^2 + 4\frac{\phi_1}{\lambda_1}i}\right).$$

Corollary 2.1 (solution to *$S_1(0,X(0))$* for small penalty discounts): If the conditions for Proposition 2.3 hold and ρ_1 is small, then the optimal net hiring is approximately:

$$W_1(t) = \frac{1}{\tau_1\sqrt{\lambda_1}}\left[\bar{r} - C_1(t)\right] + \frac{1}{2\tau_1^2\lambda_1}\left[B_1(t) - \lambda_1\bar{r}\right] \qquad \text{for } t \geq 0 \quad (2.21)$$

where $$\tau_1 \triangleq (2\phi_1)^{-\frac{1}{2}}.$$

Proof: Substituting $\rho_1 = 0$ into (2.20) and simplifying yields the required result.

Two comments are in order for Corollary 2.1. First, assuming that the discount rate is small is not heroic. For example, real interest rates have been in recent decades typically less than 5% (Barro 1990). Second, from the first term in (2.21) we will refer to τ_1 as the base capacity adjustment time (or simply, *capacity adjustment time*) because it can be thought of as the time to correct the gap between the desired and actual capacity when there is a one-period target lead-time and no deviation from target backlog.

While useful in itself, the simplified policy in (2.21) can improve our insight into service supply chain management. For example, a firm following that policy will react more vigorously to deviations from the target backlog when it is easy to adjust capacity (the capacity adjustment time τ_1 is small) or the target lead-time λ_1 is short. These factors will be responsible for many of the interesting performance measure trade-offs that appear later in the paper.

2.5 Stage 2 Control Policy

We repeat the same process to determine the optimal control strategy at stage 2. However, because stage 2 has no control over the solution to the problem *$S_i[0,X(0)]$,* it must take the values of the stage 1 backlog $B_1(t)$ and capacity $C_1(t)$ and their associated certainty-equivalent variables as givens. This simplifies the need to estimate the local demand seen at stage 2, i.e., the stage 1 capacity $C_1(t,u)$.

Lemma 2.1 Solution to *$C_1(t,u)$* in *$R_1[0,X(t,0)]$*:

Under the conditions for Proposition 2.3, the solution to the certainty-equivalent value of stage 1 capacity at time *t+u* is (where δ_{11}, δ_{12} are previously defined in Prop. 2.4.2.)

$$C_1(t,u)=\overline{r}+\frac{\left(\delta_{12}e^{-\delta_{12}u}-\delta_{11}e^{-\delta_{11}u}\right)\left[\overline{r}-C_1(0)\right]+\delta_{11}\delta_{12}\left(e^{-\delta_{12}u}-e^{-\delta_{11}u}\right)\left[B_1(0)-\lambda_1\overline{r}\right]}{\delta_{11}-\delta_{12}} \quad \text{for } t,u\geq 0. \quad (2.22)$$

Let $\hat{C}_1(t,u)$ be the expected stage 1 capacity at time *t+u* forecast at time *t.*

Proposition 2.4 (MMSE forecast of Stage 2 demand): The minimum mean square error forecast for the stage 1 capacity under optimal control with respect to Problem *$S_1[t,X(0)]$* is

$$\hat{C}_1(t,u)=\overline{r}+\frac{\left(\delta_{12}e^{-\delta_{12}u}-\delta_{11}e^{-\delta_{11}u}\right)\left[\overline{r}-C_1(0)\right]+\delta_{11}\delta_{12}\left(e^{-\delta_{12}u}-e^{-\delta_{11}u}\right)\left[B_1(0)-\lambda_1\overline{r}\right]}{\delta_{11}-\delta_{12}} \quad \text{for } t,u\geq 0. \quad (2.23)$$

Proposition 2.5 (solution to the Closed-Loop Problem *$S_2[t,X(0)]$*): Under the conditions given in the closed-loop dynamic problem *$S_2[t,X(0)]$*) and assuming that stage 2 local demand evolves according to Lemma 2.1, then the optimal net hiring for the closed-loop dynamic problem *$S_2[t,X(t])$* is:

$$\begin{aligned}W_2(t)=&Q_{C1}\left[C_1(t)-\overline{r}\right]+Q_{C2}\left[C_2(t)-C_1(t)\right]\\&+Q_{B1}\left[B_1(t)-\lambda_1\overline{r}\right]+Q_{B2}\left[B_2(t)-\lambda_2\overline{r}\right]\end{aligned} \qquad \text{for } t\geq 0 \qquad (2.24)$$

where

$$Q_{B1}=\delta_{21}\delta_{22}+\frac{\phi_2^2}{\lambda_2^2\left(\delta_{11}-\delta_{12}\right)}\left[\frac{\delta_{12}\left(\delta_{11}-\delta_{21}\right)\left(\delta_{11}-\delta_{22}\right)}{\left(\phi_2^2/\lambda_2^2\right)+\lambda_2^2\delta_{11}^2\left(\rho_2+\delta_{11}\right)^2}-\frac{\delta_{11}\left(\delta_{12}-\delta_{21}\right)\left(\delta_{12}-\delta_{22}\right)}{\left(\phi_2^2/\lambda_2^2\right)+\lambda_2^2\delta_{12}^2\left(\rho_2+\delta_{12}\right)^2}\right],$$

$$Q_{C1} = -\delta_{21} - \delta_{22} + \frac{\phi_2^2}{\lambda_2^2\left(\delta_{12}-\delta_{11}\right)}\left[\frac{\left(\delta_{11}-\delta_{21}\right)\left(\delta_{11}-\delta_{22}\right)}{\left(\phi_2^2/\lambda_2^2\right)+\lambda_2^2\delta_{11}^2\left(\rho_2+\delta_{11}\right)^2} - \frac{\left(\delta_{12}-\delta_{21}\right)\left(\delta_{12}-\delta_{22}\right)}{\left(\phi_2^2/\lambda_2^2\right)+\lambda_2^2\delta_{12}^2\left(\rho_2+\delta_{12}\right)^2}\right], \quad (2.25)$$

$$Q_{C2} = -\delta_{21} - \delta_{22},$$

$$Q_{B2} = \delta_{21}\delta_{22}.$$

where δ_{11}, δ_{12} are defined in (2.20) and

$$\delta_{22} \triangleq -\frac{1}{2}\left(\rho_2 - \sqrt{\rho_2^2 - 4\frac{\phi_2}{\lambda_2}i}\right), \delta_{22} \triangleq -\frac{1}{2}\left(\rho_2 - \sqrt{\rho_2^2 + 4\frac{\phi_2}{\lambda_2}i}\right).$$

As with the first stage results, we shall find the following corollary for small penalty discounts to be useful for measuring peformance.

Corollary 2.2 (solution to *$S_2(0,X(0))$* for small penalty discounts): If the conditions for Proposition 2.5 hold, and ρ_1 and ρ_2 are small, then the optimal control law coefficients for the closed-loop problem *$S_2(0,x(0))$* are approximately:

$$\begin{aligned} W_2(t) &= Q_{C1}\left[C_1(t)-\bar{r}\right] + Q_{C2}\left[C_2(t)-C_1(t)\right] \\ &\quad + Q_{B1}\left[B_1(t)-\lambda_1\bar{r}\right] + Q_{B2}\left[B_2(t)-\lambda_2\bar{r}\right] \end{aligned} \qquad \text{for } t \geq 0 \quad (2.26)$$

where $$Q_{B1} = \frac{1}{2\left(\lambda_1\tau_1^2+\lambda_2\tau_2^2\right)}, \; Q_{C1} = -\frac{\lambda_1\tau_1^2 + \tau_1\tau_2\sqrt{\lambda_1\lambda_2} + \lambda_2\tau_2^2}{\lambda_1^{\frac{3}{2}}\tau_1^3 + \lambda_1\tau_1^2\tau_2\sqrt{\lambda_2} + \lambda_2\tau_2^2\tau_1\sqrt{\lambda_1} + \lambda_2^{\frac{3}{2}}\tau_2^3},$$

$$Q_{C2} = -\frac{1}{\tau_2\sqrt{\lambda_2}}, \; Q_{B2} = \frac{1}{2\tau_2^2\lambda_2} \qquad (2.27)$$

where $$\tau_1 \triangleq \left(2\phi_1\right)^{-\frac{1}{2}}, \tau_2 \triangleq \left(2\phi_2\right)^{-\frac{1}{2}}$$

Proof: Substituting $\rho_1 = 0$ and $\rho_2 = 0$ into (2.25) and simplifying yields the required result.

Observe from Corollary 2.2 that if the stage 1 capacity and backlog are at their mean levels, the stage 2 control law is identical in form to stage 1's. However, if the stage 1 backlog exceeds its mean level, the net hiring rate at stage 2 will increase in anticipation of stage 1 increasing its capacity, which will in turn increase stage 2's local demand. This effect is exacerbated if stage 1's fractional capacity adjustment rate $\left(\lambda_1\tau_1^2\right)^{-1}$ is fast. Hence, all other things being equal, the smaller the backlog or the easier the hiring at stage 1, the more quickly stage 2 itself needs to hire. On the other hand, these effects will be counteracted to some extent if stage 1's capacity exceeds its mean level. In this case, stage 2 net hiring will decrease in anticipation of stage 1 decreasing its capacity.

3. Results for the Single Stage Model

In order to improve our intuition for the characteristics of capacity, backlog, and information management in a service-oriented supply chain, we shall develop several measures of performance assuming optimal control by both firms under conditions of small penalty discounts. In fact, the results in the rest of the paper will be based on these assumptions.

3.1 Conditions for the Bullwhip to exist in response to a deterministic step input.

The most basic question with respect to service-oriented supply chains is under what conditions, if any, does the bullwhip effect occur. Bullwhip is defined as a change in the processing rate at least temporarily greater than the change in the demand rate. We are also interested in whether the percentage change in the backlog may exceed that in the local demand. For simplicity in exposition we will refer to this bullwhip-like effect in the backlog as the "backlog bullwhip" when necessary to distinguish it from the standard bullwhip effect seen in order processing.

In order to improve our intuition for these issues, we digress temporarily from examining the case of white noise end-customer demand. Instead, we first characterize the bullwhip effect when end-customer demand experiences a one-time step change at time *t=0*. That is,

$$r(t) = \{ \begin{array}{ll} \mu_i & \text{for } t < 0 \\ \mu_f & \text{for } t \geq 0 \end{array} \qquad (3.1)$$

Because the demand evolution in (3.1) for $t \geq 0$ is identical to that of the optimal MMSE forecast in Proposition 2.2, we can use the solution to $S_1[0,X(0)]$ from Corollary 2.1 by substituting μ_f for $\overline{r}$ and assuming that $C_1(0) = \mu_i$ and $B_1(0) = \lambda_1 \mu_i$.

Note that in the following theorem and also for the remainder of the paper we will often make use of the following identities to simplify exposition.

$$\theta_1 \triangleq \frac{1}{\tau_1\sqrt{2\lambda_1}} \quad \text{and} \quad \theta_2 \triangleq \frac{1}{\tau_2\sqrt{2\lambda_2}} \tag{3.2}$$

Since θ_1 is inversely related to λ_1 and τ_1, the value of θ_1 can be thought of as the optimal fractional rate at which the gap between the desired and actual capacities at stage 1 closes. The parameter θ_2 permits a similar interpretation except that the total control law from Corollary 2.2 will also be modified by the values of stage 1's backlog and capacity.

Theorem 3.1: If the end-customer demand follows equation (3.1), then the under optimal control:

a. The time behavior response of stage 1 capacity (and consequently its processing rate) equals

$$C_1(t) = \mu_i + \left(\mu_f - \mu_i\right)\left\{1 - \exp\left(-\frac{t\theta_1}{\sqrt{2}}\right)\left[\cos\left(\frac{t\theta_1}{\sqrt{2}}\right) + \sin\left(\frac{t\theta_1}{\sqrt{2}}\right)\right]\right\} \quad \text{for } t \geq 0. \tag{3.3}$$

b. The processing rate exhibits a bullwhip effect (i.e., $C_1(t) > \mu_f$ for some $t \geq 0$ if $\mu_f > \mu_i$, or $C_1(t) < \mu_f$ for some $t \geq 0$ if $\mu_f < \mu_i$) with a maximum overshoot of $\exp(-\pi)\left(\mu_f - \mu_f\right)$—approximately 4.3 percent—occurring at $t = \frac{\pi\sqrt{2}}{\theta_1}$.

c. The time behavior response of the stage 1 backlog equals:

$$B_1(t) = \theta_1^{-1}\sqrt{2}\lambda_1\left[\mu_i + \left(\mu_f - \mu_i\right)\right]\left[1 - \exp\left(-\frac{t\theta_1}{\sqrt{2}}\right)\cos\left(\frac{t\theta_1}{\sqrt{2}}\right)\right] \quad \text{for } t \geq 0. \tag{3.4}$$

d. The backlog exhibits a bullwhip effect (i.e., for some $t \geq 0$, $B_1(t) > \lambda_1\mu_f$ if $\mu_f > \mu_i$, or $B_1(t) < \lambda_1\mu_f$ if $\mu_f < \mu_i$) with maximum overshoot of $\lambda_1 \exp\left(-\frac{3\pi}{4}\right)\left(\mu_f - \mu_f\right)$—approximately 6.7 percent—at $t = \frac{3\pi}{2\theta_1\sqrt{2}}$.

From the theorem above, regardless of the value of any parameters, a bullwhip in the capacity and the backlog will necessarily occur. Furthermore, for both performance measures, the fractional capacity adjustment rate θ_1—and by definition the target lead-time λ_1 and capacity adjustment time τ_1—*does not affect the magnitude of the overshoot*. However, the times of peak overshoot for both capacity and backlog decrease in θ_1 or, conversely, increase in λ_1 and τ_1 because a longer target lead-time or base capacity adjustment time results in delaying the impact of the step change.

3.2 Analysis with Random Demand

Typically demand data are uncertain. Therefore, we extend our analysis to determine the effects of random demand in the optimally controlled service-oriented supply chain. Under the linear model given by the control laws developed in the previous section, the mean capacity will equal the average demand, and the mean backlog will simply be the product of the target lead-time and the average demand. Accordingly, we will focus on the effects of various parameters on the *variances* of capacity, net hiring, and backlog at each stage. The variance in capacity is important because it implies greater variation in demand to higher echelons in the supply chain. The variance in net hiring is of interest because it is proportionate to the square of the mean absolute value of the net hiring rate and, hence, the total hiring, training, and firing costs associated with managing variable demand. The variance in the backlog is also a key factor in service quality. As the variation of the backlog increases, the variation of the waiting time may also increase[1].

The bulk of the remaining analysis in this section will concentrate on illuminating the interactions of the capacity, net hiring, and backlog variances and their relationships with the target lead-time and capacity adjustment time at each stage. Theorem 3.2 contains the main results. Because the theorem contains exact expressions, we broaden our discussion beyond the bullwhip effect to provide a complete characterization of service-oriented supply chain dynamics.

Theorem 3.2: Under the assumptions in $S_1[t,X(0)]$ and $S_2[t,X(0)]$, the following results hold:

---

[1] Because capacity is changing, increased backlog variance does not necessarily imply increased waiting time variance. However, a heuristic argument can be made that there is, in general, a positive relationship between backlog variance and waiting time variance. If end-customer demand has a Gaussian distribution, so too will *B(t)* and *C(t)* and in a linear time-invariant system. Consequently, the ratio *B(t)/C(t)*, which represents instantaneous waiting time, has a Cauchy distribution with scale parameter $\sigma_B \sigma_C^{-1}\sqrt{1-r^2}$ (Papoulis 1991, p. 138) where *r* is the correlation of *B(t)*

a. The variance of the backlog at the first stage equals:

$$V_{B1} = \sigma^2 \frac{3\tau_1\sqrt{\lambda_1}}{2}. \tag{3.5}$$

b. The variance of the capacity at the first stage equals:

$$V_{C1} = \frac{\sigma^2}{4\tau_1\sqrt{\lambda_1}}. \tag{3.6}$$

c. The variance of net hiring at the first stage equals:

$$V_{W1} = \frac{\sigma^2}{8\tau_1^3\lambda_1^{\frac{3}{2}}} \tag{3.7}$$

Proposition 3.1: If the conditions assumed in Theorem 3.2 hold, then

a. V_{B1} *increases* in the target lead time λ_1 and the base capacity adjustment time τ_1, and

b. V_{C1} and V_{W1} *decrease* in the target lead-time λ_1 and the base capacity adjustment time τ_1.

The proof of Proposition 3.1 follows by inspection of Theorem 3.2. Proposition 3.1 demonstrates that the capacity and backlog variances can respond differently to the same stimulus. More specifically, the variances of capacity and backlog move in different directions as the target lead-time and capacity adjustment time are adjusted. Thus, the modern managerial tendency to reduce target lead-time and base capacity adjustment time will reduce backlog variance and by extension may improve service quality. But there is a tradeoff. Reducing these two delays exacerbates the net-hiring variance, thus increasing hiring and firing costs. It also increases the capacity variance, thus passing on increased demand variability to higher echelons. If adjusting capacity is expensive, a tradeoff between customer service and personnel costs may result. (For a more extensive look at these issues under high growth conditions, see Anderson Forthcoming.)

4 Results for the Two-Stage Model

Theorem 4.1: Under the assumptions in $S_1[t,X(0)]$ and $S_2[t,X(0)]$, the following results hold:

---

and *C(t)* and σ_B and σ_C are their respective standard deviations. Hence the scale (or spread) of the instantaneous waiting time distribution increases in σ_B.

a. The variance of the backlog at stage 2 equals:

$$V_{B2} = \sigma^2 \theta_1^7 \frac{3\theta_1^2 + 4\theta_1\theta_2 + 3\theta_2^2}{2\sqrt{2}\theta_2 \left(\theta_1 + \theta_2\right)^3 \left(\theta_1^2 + \theta_2^2\right)^3}. \tag{4.1}$$

b. The variance of capacity at stage 2 equals:

$$V_{C2} = \sigma^2 \theta_1 \theta_2 \frac{\theta_1^8 + 3\theta_1^7\theta_2 + 6\theta_1^6\theta_2^2 + 10\theta_1^5\theta_2^3 + 14\theta_1^4\theta_2^4 + 10\theta_1^3\theta_2^5 + 6\theta_1^2\theta_2^6 + 3\theta_1\theta_2^7 + \theta_2^8}{2\sqrt{2}\left(\theta_1 + \theta_2\right)^3 \left(\theta_1^2 + \theta_2^2\right)^3}. \tag{4.2}$$

c. The variance of net hiring at stage 2 equals:

$$V_{W2} = \sigma^2 \theta_1^3 \theta_2^3 \frac{\theta_1^6 + 3\theta_1^5\theta_2 + 6\theta_1^4\theta_2^2 + 10\theta_1^3\theta_2^3 + 6\theta_1^2\theta_2^4 + 3\theta_1\theta_2^5 + \theta_2^6}{2\sqrt{2}\left(\theta_1 + \theta_2\right)^3 \left(\theta_1^2 + \theta_2^2\right)^3}. \tag{4.3}$$

d. The variance of the total backlog ($\operatorname{var}\left(B_1(t) + B_2(t)\right)$ equals:

$$V_{B1+B2} = \sigma^2 \frac{\begin{bmatrix} 3\theta_1^{10} + 9\theta_1^9\theta_2 + 18\theta_1^8\theta_2^2 + 30\theta_1^7\theta_2^3 + 42\theta_1^6\theta_2^4 + 46\theta_1^5\theta_2^5 \\ +42\theta_1^4\theta_2^6 + 30\theta_1^3\theta_2^7 + 18\theta_1^2\theta_2^8 + 9\theta_1\theta_2^9 + 3\theta_2^{10} \end{bmatrix}}{2\sqrt{2}\theta_1\theta_2 \left(\theta_1 + \theta_2\right)^3 \left(\theta_1^2 + \theta_2^2\right)^3}. \tag{4.4}$$

Theorem 4.1 characterizes the bullwhip effect in the service-oriented supply chain and can describe how target lead-times and capacity adjustment times should be coordinated throughout such a chain when there is complete information sharing from the first stage to the second. While the results of Theorem 4.1 are complex, they do provide easily interpreted and interesting sensitivity results.

4.1 Second-Stage Sensitivity to First-Stage Parameters

Proposition 4.1: If the conditions assumed in Theorem 3.2 hold, then the first-stage target lead-time λ_1 and capacity adjustment time τ_1 have the following effects:

a. The second-stage backlog and net-hiring variances V_{B2} and V_{W2} *decrease* in λ_1 and τ_1.

b. Stage 2 capacity variance V_{C2} *increases* in λ_1 and τ_1 for $\lambda_1\tau_1^2 < a\lambda_2\tau_2^2$ (where $a \approx 0.1106$) to a maximum value $V_{C2}^{\max}$. Furthermore, $V_{C2}^{\max}$ exceeds V_{C2} by no more than 0.3 percent for $\lambda_1\tau_1^2 < a\lambda_2\tau_2^2$. However, V_{C2} *decreases* in λ_1 and τ_1 asymptotically to zero for $\lambda_1\tau_1^2 > a\lambda_2\tau_2^2$.

c. The total backlog variance V_{B1+B2} *increases* in λ_1 and τ_1.

For sake of brevity we shall discuss only the impact of the stage 1 target lead-time λ_1 because τ_1 appears in an interaction term with λ_1 in the denominator of the definition of θ_1. However, all of the following insights apply equally for both parameters.

Proposition 3.1 indicates that reducing the stage 1 target lead-time will reduce the variance of the stage 1 backlog. This accords with conventional wisdom derived from inventory supply chains. However, the primary result of Proposition 4.1 is that reducing the stage 1 target lead-time will actually *increase* the backlog and net hiring variances at stage 2. For all practical purposes, the same action will generally *increase* the capacity variance at stage 2 as well, because a reduction of at most 0.3 percent for $\lambda_1 < a\lambda_2\tau_2^2\tau_1^{-2}$ is, from a managerial perspective, negligible. Thus, reducing target lead-time at stage 1 myopically appears to improve local service quality but hurts higher echelon service quality and personnel costs. Figure 3 shows these effects on backlog variances for one set of parameters.

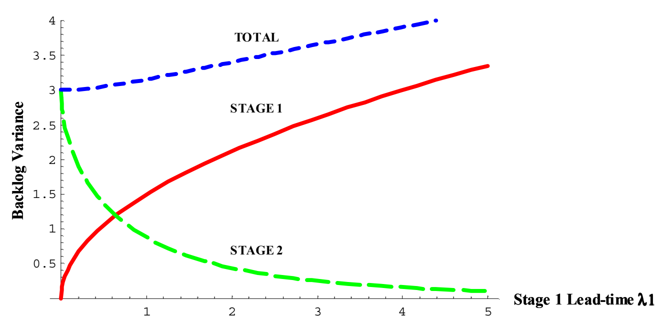

Figure 3: Backlog Variances vs. λ_1 for $\tau_1 = \tau_2 = 1$ and $\lambda_2 = 4$

In general, the minima of backlog variation for stage 1 and stage 2 occur at different values of the stage 1 target lead-time. Hence, incentive alignment is a crucial and complex issue for service-oriented supply chain management. It is evident from Figure 3 that the shape of the total backlog variance curve is a combination of the backlog variances at the first and second stages. It is also important to note that, due to correlation between the backlogs, the total variance curve is not simply the sum of the other two variances. Target lead-time reduction decreases the total backlog variation but at a decreasing rate due to trade-off between first and second stage backlog variances. Hence, although it can be argued that target lead-time reduction will always improve the backlog variance for the supply chain as a whole, the benefit may be greatly mitigated by the increase in backlog variance at the higher echelon. This is stated formally in the next proposition.

Proposition 4.2: Stage 1 backlog variance, V_{B1}, increases more quickly in λ_1 and τ_1 than the total backlog variance V_{B1+B2}.

4.2 Second-Stage Sensitivity to Second-Stage Parameters

Proposition 4.3: If the conditions assumed in Theorem 3.2 hold, then the stage 2 target lead-time λ_2 and the capacity adjustment time τ_2 have the following effects:

a. Stage 2 backlog variance V_{B2} and total backlog variance V_{B1+B2} *increase* in λ_2 and τ_2.

b. Stage 2 capacity variance V_{C2} *increases* in λ_2 and τ_2 for $\lambda_2\tau_2^2 < a\lambda_1\tau_1^2$ (where $a \approx 0.1106$) to a maximum value $V_{C2}^{\max}$. Furthermore, $V_{C2}^{\max}$ exceeds V_{C2} by no more than 0.3 percent for $\lambda_2\tau_2^2 < a\lambda_1\tau_1^2$. However, V_{C2} *decreases* in λ_2 and τ_2 asymptotically to zero for $\lambda_2\tau_2^2 > a\lambda_1\tau_1^2$.

c. The stage 2 net-hiring variance V_{W2} *decreases* in λ_2 and τ_2.

For fixed first-stage parameters, the sensitivity of the second stage capacity and backlog variances to the second-stage parameters is similar to the sensitivity of the first-stage variances to the first-stage parameters. There is, however, one minor exception: V_{C2} does not monotonically decrease in the second-stage target lead-time λ_2 and capacity adjustment time τ_2. In fact, V_{C2} behaves in λ_2 as it did in λ_1 (and in τ_2 as it did in τ_1) because it is symmetric in these two parameters (see (4.2)). However, for all practical purposes, V_{C2} essentially decreases in λ_2 and τ_2 because it never increases by more than 0.3 percent above its value at $\lambda_2\tau_2 = 0$.

4.3 Managing the Second Stage Without First Stage Demand and Backlog Information

The results in the previous sections are based on the assumption that stage 2 knows stage 1's demand and backlog (we will refer to this as the "full information case"). In an environment of improved information technology and ever increasing collaboration between supply chain trading partners, these are not unrealistic assumptions. However, for various reasons supply chains still exist in which the supply chain members do not share state information (we will refer to this as the "no information case"). Because the stage 1 capacity is equal to stage 1 output and consequently stage 2 demand, stage 1 capacity can be easily inferred. However, without end-customer demand information, this is not necessarily so easy with stage 1 backlog. Under these circumstances, a logical approach for the second stage is to assume that the stage 1 backlog equals its desired target because, according to the optimal control rule in (2.26), a rational player at stage 1 tries to match backlog to a desired target level. As a result, the control policy in (2.26) reduces to

$$W_2(t) = Q_{C1}\left[C_1(t) - \bar{r}\right] + Q_{C2}\left[C_2(t) - C_1(t)\right] + Q_{B2}\left[B_2(t) - \lambda_2 \bar{r}\right] \tag{4.5}$$

If stage 2 uses the control rule in (4.5), then its backlog, capacity, and net hiring variances are given by the following theorem.

Theorem 4.2: If stage 2 places no weight on the information from stage 1, then

a. The variance of backlog at stage 2 reduces to:

$$V_{B2} = \sigma^2 \frac{\theta_1^3\left(3\theta_1^6 + 8\theta_1^5\theta_2 + 15\theta_1^4\theta_2^2 + 18\theta_1^3\theta_2^3 + 15\theta_1^2\theta_2^4 + 8\theta_1\theta_2^5 + 3\theta_2^6\right)}{2\sqrt{2}\theta_2\left(\theta_1+\theta_2\right)^3\left(\theta_1^2+\theta_2^2\right)^3}. \tag{4.6}$$

b. The variance of capacity at stage 2 equals:

$$V_{C2} = \sigma^2\theta_1\theta_2 \frac{\theta_1^8 + 3\theta_1^7\theta_2 + 6\theta_1^6\theta_2^2 + 9\theta_1^5\theta_2^3 + 10\theta_1^4\theta_2^4 + 9\theta_1^3\theta_2^5 + 8\theta_1^2\theta_2^6 + 3\theta_1\theta_2^7 + \theta_2^8}{2\sqrt{2}\theta_2\left(\theta_1+\theta_2\right)^3\left(\theta_1^2+\theta_2^2\right)^3}. \tag{4.7}$$

c. The variance of the net hiring at stage 2 equals:

$$V_{W2} = \frac{\sigma^2\theta_1^3\theta_2^3\left(\theta_1^6 + 2\theta_1^5\theta_2 + 3\theta_1^4\theta_2^2 + 4\theta_1^3\theta_2^3 + 3\theta_1^2\theta_2^4 + 4\theta_1\theta_2^5 + \theta_2^6\right)}{2\sqrt{2}\left(\theta_1+\theta_2\right)^3\left(\theta_1^2+\theta_2^2\right)^3}. \tag{4.8}$$

d. The variance for the total backlog is:

$$V_{B1+B2}=\sigma^2\frac{3\theta_1^{10}+13\theta_1^9\theta_2+30\theta_1^8\theta_2^2+50\theta_1^7\theta_2^3+67\theta_1^6\theta_2^4+70\theta_1^5\theta_2^5+61\theta_1^4\theta_2^6+44\theta_1^3\theta_2^7+24\theta_1^2\theta_2^8+9\theta_1\theta_2^9+3\theta_2^{10}}{2\sqrt{2}\theta_1\theta_2(\theta_1+\theta_2)^3(\theta_1^2+\theta_2^2)^3}. \quad (4.9)$$

To quantify the impact of not sharing information, consider the average value over all possible initial state vectors of the objective function in (2.13) at stage 2 (where again $\tau_2 \triangleq (2\phi_2)^{-\frac{1}{2}}$).

$$F_2 = E_{X(0)}\left[E_{r(t)|X(0)}\left[\int_0^\infty \exp(-\rho_2 t)\left\{\frac{1}{4\tau_2^4\lambda_2^2}\left[B_2(t)-\lambda_2\bar{r}\right]^2+\left[W_2(t)\right]^2\right\}dt\,|X(0)\right]\right] \quad (4.10)$$

Under the assumptions of *$S_1[0,X(0)]$* and *$S_2[0,X(0)]$* and a control law for *$W_2(t)$* that is linear in the state variables with constant coefficients, then the average objective function performance is time stationary and ergodic. Using these facts and the conditional expectation formula (Taylor and Karlin 1984, p. 47), and then moving expectation operator inside the integral, (4.10) reduces to

$$F_2 = E_{r(t)}\left[\int_0^\infty \exp(-\rho_2 t)\left\{\frac{1}{4\tau_2^4\lambda_2^2}\left[B_2(t)-\lambda_2\bar{r}\right]^2+\left[W_2(t)\right]^2\right\}dt\right]=\frac{1}{\rho_2}\left(\frac{1}{4\tau_2^4\lambda_2^2}V_{B2}+V_{W2}\right). \quad (4.11)$$

Let F_2^i = the expected penalty for the full information case, and

F_2^{no} = the expected penalty for the no information case.

Proposition 4.4: Stage 2's expected penalty function increases (worsens) when stage 1 does not share information. Moreover, the degradation in performance—that is the ratio F_2^{no}/F_2^i— increases in the product of stage 1's target lead-time λ_1 and the square of its capacity adjustment time τ_1 from a minimum of unity at $\lambda_1\tau_1^2/(\lambda_2\tau_2^2)=0$ to a maximum of 4 as $\lambda_1\tau_1^2/(\lambda_2\tau_2^2)\to\infty$.

Expected performance, not surprisingly, improves when stages share information. Of interest, however, is that significant sharing benefits occur only when stage 1's target lead- and capacity adjustment times are at least comparable to stage 2's. Otherwise, stage 1's backlog will be changing too rapidly (in a relative sense) for stage 2 to react in a timely manner to the shared information.

5 Managerial Insights and Future Research

Perhaps the most important managerial insight resulting from our current work is that the optimal control rule is both intuitive and relatively easy to implement in a service-oriented supply chain. In fact, it is a weighted sum of information that is known and can be shared in most supply chains. Additionally, we have shown that the bullwhip effect occurs in service-oriented supply chains just as it does in inventory supply chains. However, unlike the numerous win-win situations present in inventory supply chains resulting from policies such as shipping delay reductions and information sharing, this paper shows that service-oriented supply chains behave differently and are characterized by numerous trade-offs as summarized in the table below.

Table 1: Summary of Trade-offs in a Service-Oriented Supply Chain

To increase local service quality, one can shorten target lead-time, *but* this increases local personnel costs, *and* decreases service quality and increases personnel costs at higher stages.
Target-lead time and capacity adjustment time at all supply chain stages must be coordinated because these quantities reinforce each other's impact on backlog and capacity variances.
As in inventory supply chains, information sharing between stages generally improves the overall performance of higher echelon stages and the supply chain as a whole. *However,* it will only be of material benefit if stages closer to the customer cannot react significantly more quickly than those further away.

As seen in this paper, synchronizing service-oriented supply chains is complex and non-intuitive, indicating the potential for specialized information systems to monitor and control job progress in service as well as in manufacturing settings. We suggest three research paths to facilitate designing these systems. One is to continue the experimentation begun in Anderson and Morrice (2000) to uncover the cognitive biases of managers responsible for service-oriented supply chains. Another is to determine optimal and robust strategies for service-oriented supply chains under real-world conditions, such as forecasting, inaccurate data, shared resources, yield loss, re-entrant flows, and parallel processing. After these strategies are discovered and the cognitive biases understood, the third path of research is to find incentive structures that will align independent companies to behave in an approximately system-optimal manner.

Appendix of Proofs

Proposition 2.1: This is a standard result that can be found in Stengel (1994), pp.451-5.

Proposition 2.2: To obtain a Minimum Mean Square Error prediction of demand at time $t+u$ forecast at time t for a linear, time-invariant system, the error in the forecast must be uncorrelated with past demand and unbiased (Shanmugan and Breipohl 1988). That is, given $r(t)$:

$$\operatorname{cov}\left[r(t+u)-\hat{C}_0(t,u), r(t-y)\right]=0 \qquad \text{for } t,u,y\geq 0 \qquad \text{(A.1)}$$

Given the distribution of $r(t)$ specified, $\bar{r}$ will clearly satisfy this criterion.

Proposition 2.3: Using the objective function (2.14), the constraints (2.15)-(2.16), and $\hat{C}_0(t,u)$ from Proposition 2.2, the following conditions are necessary for an optimal control policy for the certainty-equivalent problem $R_1[t,X(t,0)]$ (Kamien and Schwartz 1991, p. 133). As the objective function's integrand and the state constraints are all jointly convex in $X(t,u)$ and $V(t,u)$, these conditions are also sufficient for optimality. Let H_1 represent the current-value Hamiltonian.

$$\frac{\partial \mathrm{H}_1}{\partial W_1(t,u)}=0 \Rightarrow M_{C1}(u)+2W_1(t,u)=0 \qquad \text{for } t,u\geq 0 \qquad \text{(A.2)}$$

$$\frac{dM_{C1}(u)}{du}=\rho_1 M_{C1}(u)-\frac{\partial \mathrm{H}_1}{\partial C_1(t,u)}=\rho_1 M_{C1}(u)+M_{B1}(u) \quad \text{for } t,u\geq 0, \qquad \text{(A.3)}$$

$$\frac{dM_{B1}(u)}{du}=\rho_1 M_{B1}(u)-\frac{\partial \mathrm{H}_1}{\partial B_1(t,u)}=\rho_1 M_{B1}(u)-2\frac{\phi_1^2}{\lambda_1^2}\left[B_1(t,u)-\lambda_1\bar{r}\right] \quad \text{for } t,u\geq 0, \qquad \text{(A.4)}$$

$$B_1(t,u), C_1(t,u)\leq\infty . \qquad \text{(A.5)}$$

(A.2) represents the optimality condition for $X(t,u)$, in which $M_{C1}(u)$ represents the current value of the adjoint variable associated with the capacity state constraint. (A.3) is the adjoint equation for the capacity constraint, in which $M_{B1}(u)$ is the current value of the adjoint variable associated with the backlog state constraint. (A.4) is the adjoint equation for the backlog constraint. (A.5) enforces rationality constraints upon the states. Given $X(t,0)$, solving the four equations for $W_1(t,0)$—the value of $W_2(t,0)$ has no impact on this problem—yields the required certainty-equivalent control law. From Proposition 2.1, to find the required solution to the closed-loop problem, we substitute $W_1(t)$, $C_1(t)$, and $B_1(t)$ for $W_1(t,0)$, $C_1(t,0)$, and $B_1(t,0)$ respectively into the certainty-equivalent solution.

Lemma 2.1: Substituting the vector *X(t,0)* for *X(t)* and the relation (2.16) into the control law from Proposition 2.3 yields

$$\frac{d^2B_1(t,u)}{du^2}=\frac{\begin{Bmatrix}\left(\delta_{12}^2e^{-\delta_{12}u}-\delta_{11}^2e^{-\delta_{11}u}\right)\left[\bar{r}-C_1(0)\right]\\+\left(\delta_{11}\delta_{12}^2e^{-\delta_{12}u}-\delta_{11}^2\delta_{12}e^{-\delta_{11}u}\right)\left[B_1(0)-\lambda_1\bar{r}\right]\end{Bmatrix}}{\delta_{11}-\delta_{12}} \quad \text{for } t,u\geq 0. \quad \text{(A.6)}$$

Integrating this and solving for $C_1(t,u)$ using the relation (2.15) yields the required result.

Proposition 2.4: To find the MMSE forecast for a linear-quadratic system driven by white noise, we need only determine the output assuming that future values of the input disturbance are their expectation. In other words, we assume that $\hat{C}_0(t,y)=r(t-y)$ for $0\geq y\geq -t$ and $\hat{C}_0(t,y)=\bar{r}$ for $u\geq y\geq 0$. This is a standard result found in such signal-estimation texts as Shanmugan and Breipohl (1988). Because of the givens in Lemma 2.1 and Proposition 2.3, the resulting estimated output $\hat{C}_1(t,y)$ for $y\geq u\geq 0$ at time *t* is exactly the same as the certainty-equivalent solution to the capacity at stage 1 at time *t+u* using a policy determined at time *t*. Noting that the stage 1 capacity is the demand for stage 2 and using the result from (2.22) completes the proof.

Proposition 2.5: Using the objective function (2.14), the constraints (2.15)-(2.16), and $\hat{C}_1(t,u)$ from Proposition 2.4, the following conditions are necessary for an optimal control policy for the certainty-equivalent problem *R₂[t,X(t,0)]* (Kamien and Schwarz 1991). As the objective function's integrand and the state constraint are both jointly convex in *X(t,u)* and *V(t,u),* these conditions will also be sufficient for optimality. Let H_2 be the current-value Hamiltonian.

$$\frac{\partial \mathrm{H}_2}{\partial W_1(t,u)}=0\Rightarrow M_{C2}(u)+2W_2(t,u)=0 \qquad \text{for } t,u\geq 0, \qquad \text{(A.7)}$$

$$\frac{dM_{C2}(u)}{du} = \rho_2 M_{C2}(u) - \frac{\partial \mathrm{H}_2}{\partial C_2(t,u)} = \rho_2 M_{C2}(u) + M_{B2}(u) \quad \text{for } t,u \geq 0, \tag{A.8}$$

$$\frac{dM_{B2}(u)}{du} = \rho_2 M_{B2}(u) - \frac{\partial \mathrm{H}_2}{\partial B_2(t,u)} = \rho_2 M_{B2}(u) - 2\frac{\phi_2^2}{\lambda_2^2}\left[B_2(t,u) - \lambda_2 \bar{r}\right] \quad \text{for } t,u \geq 0, \tag{A.9}$$

$$B_2(t,u), C_2(t,u) \leq \infty. \tag{A.10}$$

(A.7) is the optimality condition for *X(t,u),* in which $M_{C2}(u)$ represents the current value of the adjoint variable associated with the capacity state constraint. (A.8) is the adjoint equation for the capacity constraint. (A.9) is the adjoint equation for the backlog constraint, in which $M_{B1}(u)$ represents the current value of the adjoint variable associated with the backlog state constraint. (A.10) enforces rationality constraints upon the states. Given *X(t,0), s*olving the four equations for $W_2(t,0)$ yields the required certainty-equivalent control law in terms of $X(t,0)$. From Proposition 2.1, to find the required solution to the closed-loop problem, we substitute $X(t)$ and $W_2(t)$ for $X(t,0)$ and $W_2(t,0)$ respectively in the certainty-equivalent solution.

Theorem 3.1: Transforming (2.9), (2.10), and (2.21) term-by-term into their Laplacian equivalents and using *s* as the Laplacian operator for differentiation yields: $sB_1(s) = r(s) - C_1(s)$ and $sC_1(s) = -\sqrt{2}\theta_1 C_1(s) + \theta_1^2 B_1(s)$. Solving for $B_1(s)$ and $C_1(s)$ in terms of $r(s)$ yields ratios of their Laplace transforms known as transfer functions:

$$\frac{B_1(s)}{r(s)} = \frac{\left(s + \sqrt{2}\theta_1\right)}{s^2 + \sqrt{2}s\theta_1 + \theta_1^2} \tag{A.11}$$

$$\frac{C_1(s)}{r(s)} = \frac{\theta_1^2}{s^2 + \sqrt{2}s\theta_1 + \theta_1^2}. \tag{A.12}$$

Multiplying these transfer functions respectively by λ_1 $(\mu_f - \mu_i)/s$ and $(\mu_f - \mu_i)/s$ yields the Laplace transforms of the system responses to the desired step change in end-customer demand at *t=0* (Oppenheim et al. 1983). Transforming the results back into the time domain and adding $\lambda_1 \mu_i$ and μ_i respectively gives the required results for parts a and c. For part d, setting $dC_1(t)/dt = 0$ yields the following positive roots $t=(4j-1)\pi\left(2\sqrt{2}\theta_1\right)^{-1}$ for $j=0,1,\ldots$. The root at $t=3\pi\left(2\sqrt{2}\theta_1\right)^{-1}$ is a maximum as the second derivative with respect to *t* evaluated at this root is negative, and it is a global maximum as subsequent maximums are damped by the negative exponential term in (3.4). Evaluated at $t=3\pi/\left(2\sqrt{2}\theta_1\right)$, (3.4) equals $\theta_1^{-1}\left[\sqrt{2}+Exp(-3\pi/4)\right]$. As $t\to\infty$, (3.4) converges to $\theta_1^{-1}\sqrt{2}$. Taking the ratio of the maximum to the limit yields the final result. Part b's proof is identical to part d's except for the specific algebraic manipulations involved.

Theorem 3.2: From Parseval's theorem, we can find a linear system's amplification of a white noise process' variance directly from its transfer function (Shanmugan and Breipohl 1988, p. 449]. This is done by evaluating the coefficient at $s=2\pi f\imath$ (where *f* represents the frequency and $\imath=\sqrt{-1}$), squaring the result's absolute value, and integrating it with respect to *f*. Performing this operation with the transfer functions of $C_1(s)$ and $B_1(s)$ from (A.11)-(A.12) yields $\operatorname{var}\left[C_1(t)\right]=\sigma^2\int_{-\infty}^{\infty}\left|\frac{\theta_1^2}{s^2+\sqrt{2}s\theta_1+\theta_1^2}\right|_{s=2\pi\imath f}^2 df$ and $\operatorname{var}\left[B_1(t)\right]=\sigma^2\int_{-\infty}^{\infty}\left|\frac{\left(s+\sqrt{2}\theta_1\right)}{s^2+\sqrt{2}s\theta_1+\theta_1^2}\right|_{s=2\pi\imath f}^2 df$.

Since $W_1(t)$ is the time derivative of $B_1(t)$, $\operatorname{var}\left[W_1(t)\right]=\sigma^2\int_{-\infty}^{\infty}\left|\frac{sC_1(s)}{r(s)}\right|_{s=2\pi\imath f}^2 df=\sigma^2\int_{-\infty}^{\infty}\left|\frac{s\left(s+\sqrt{2}\theta_1\right)}{s^2+\sqrt{2}s\theta_1+\theta_1^2}\right|_{s=2\pi\imath f}^2 df$.

Evaluating these expressions yields the required results.

Theorem 4.1: Obtaining the transfer functions for stage 2's capacity and backlog as a function of end-customer demand in the same manner as in Theorem 3.1 yields,

$$\frac{B_2(s)}{r(s)}=\frac{\theta_1^4\left[\sqrt{2}\theta_1\theta_2+s\left(\theta_1+\theta_2\right)\right]}{\left(s^2+\sqrt{2}s\theta_1+\theta_1^2\right)\left(\theta_1+\theta_2\right)\left(\theta_1^2+\theta_2^2\right)\left(s^2+\sqrt{2}s\theta_2+\theta_2^2\right)} \tag{A.13}$$

$$\frac{C_2(s)}{r(s)}=\frac{\theta_1^2\theta_2^2\left(s^2\left(\theta_1+\theta_2\right)+\left(\theta_1+\theta_2\right)\left(\theta_1^2+\theta_2^2\right)+\sqrt{2}s\left(\theta_1^2+\theta_1\theta_2+\theta_2^2\right)\right)}{\left(s^2+\sqrt{2}s\theta_1+\theta_1^2\right)\left(\theta_1+\theta_2\right)\left(\theta_1^2+\theta_2^2\right)\left(s^2+\sqrt{2}s\theta_2+\theta_2^2\right)}$$

$$\frac{W_2(s)}{r(s)}=\frac{s\theta_1^2\theta_2^2\left(s^2\left(\theta_1+\theta_2\right)+\left(\theta_1+\theta_2\right)\left(\theta_1^2+\theta_2^2\right)+\sqrt{2}s\left(\theta_1^2+\theta_1\theta_2+\theta_2^2\right)\right)}{\left(s^2+\sqrt{2}s\theta_1+\theta_1^2\right)\left(\theta_1+\theta_2\right)\left(\theta_1^2+\theta_2^2\right)\left(s^2+\sqrt{2}s\theta_2+\theta_2^2\right)}$$

Because the Laplace transform is a linear operator, the following transfer function results from (A.11) and (A.13): $\frac{B_1(s)+B_2(s)}{r(s)} = \frac{\left(s+\sqrt{2}\theta_1\right)\left(\theta_1+\theta_2\right)\left(\theta_1^2+\theta_2^2\right)\left(s^2+\sqrt{2}s\theta_2+\theta_2^2\right)+\theta_1^4\left[\sqrt{2}\theta_1\theta_2+s\left(\theta_1+\theta_2\right)\right]}{\left(s^2+\sqrt{2}s\theta_1+\theta_1^2\right)\left(\theta_1+\theta_2\right)\left(\theta_1^2+\theta_2^2\right)\left(s^2+\sqrt{2}s\theta_2+\theta_2^2\right)}$. Transforming the transfer functions in the same manner as in Theorem 3.2's proof yields the required results.

Proposition 4.1: To simplify this proof (and several other proofs) we will make converse arguments in $\theta_1 \triangleq \left(\tau_1\sqrt{2\lambda_1}\right)^{-1}$. For part a, it is easy to show that $\partial V_{B2}/\partial\theta_1 > 0$ and $\partial V_{W2}/\partial\theta_1 > 0$ since $\theta_1 > 0$. Similarly, for part c, $\partial V_{B1+B2}/\partial\theta_1 < 0$. For part b, $\partial V_{C2}/\partial\theta_1$ has one positive real root at $\theta_1 = \theta_2/\sqrt{a}$ where $a \approx 0.1106$, whose second derivative is negative at $\theta_1 = \theta_2/\sqrt{a}$. Hence, V_{C2} achieves a maximum of approximately $0.355\theta_2\sigma_2$ at $\theta_1 = \theta_2/\sqrt{a}$. Since $\partial V_{C2}/\partial\theta_1 > 0$ at $\theta_1 = 0$, V_{C2} increases until $\theta_1 = \theta_2/\sqrt{a}$. Since $\lim_{\theta_1\to\infty}\left(\partial V_{C2}/\partial\theta_1\right) = 0$ then V_{C2} decreases asymptotically to a constant value of $\theta_2\sigma^2/\left(2\sqrt{2}\right)$. The ratio of V_{C2}'s maximum divided by its limit as $\theta_1 \to \infty$ shows that V_{C2} decreases by less than 0.3 percent. By inspection, $V_{C2} = 0$ at $\theta_1 = 0$.

Proposition 4.2: Clearly, $\left(\partial V_{B1+B2}/\partial\theta_1\right) - \left(\partial V_{B2}/\partial\theta_1\right) > 0$. Hence, the required result.

Proposition 4.3: For part a, $\partial V_{B2}/\partial\theta_2 < 0$ and $\partial V_{B1+B2}/\partial\theta_2 < 0$ since $\theta_2 > 0$. Similarly for part c, $\partial V_{W2}/\partial\theta_2 > 0$. The proof for part b is identical to the proof for part b of Proposition 4.1 since (4.2) is completely symmetric in θ_1 and θ_2.

Theorem 4.2: Since the term associated with first stage backlog in (2.24) is eliminated to get (4.5) , we can assume that $Q_{B1} = 0$. Under this assumption (4.1)-(4.4) reduces to (4.6)-(4.9).

Proposition 4.4: Substituting into the variances of $B_2(t)$ and $W_2(t)$ from, respectively, Theorems 3 and 4 yields expressions for F_2^i and F_2^{no}. Their ratio is $\frac{F_2^{no}}{F_2^i} = 4 - \frac{6\theta_1^2\left(2\theta_1+\theta_2\right)}{4\theta_1^3+3\theta_1^2\theta_2+2\theta_1\theta_2^2+\theta_2^3}$.

Clearly, the ratio's maximum value—4—occurs when $\theta_1 \to 0$ or $\theta_2 \to \infty$. The ratio's minimum—unity—occurs when $\theta_1 \to \infty$ or $\theta_2 \to 0$. Furthermore, from our assumptions the ratio's derivative with respect to θ_1 is always negative and with respect to θ_2 is always positive.

Bibliography

Anderson, E.G. 2001. The Nonstationary Staff Management Problem with Business Cycle and Learning Effects. *Man. Sci.* **47** 817-832.

Anderson, E.G. Forthcoming. Managing the Impact of High Market Growth and Learning on Knowledge Worker Productivity and Service Quality. *Eur. J. of Oper. Res.*

Anderson, E.G. and D.J. Morrice. 2000. A Simulation Game for Service-Oriented Supply Chain Management *J. of Production Oper. Man.* **9** 40-55.

Baganha, M. P. and M. A. Cohen. 1998. The Stabilizing Effect of Inventory in Supply Chains. *Oper. Res.* **46** S72-S83

Barro, R. J. 1990. *Macroeconomics*. New York: John Wiley & Sons.

Brooks, F. P. 1982. *The Mythical Man-Month: Essays on Software Engineering.* Reading, MA: Addison-Wesley.

Chen, F.Y., J.K. Ryan, and D. Simchi-Levi. 1997. The Impact of Exponential Smoothing Forecasts on the Bullwhip Effect. Working Paper, Northwestern University.

Chen, F., Z. Drezner, J.K. Ryan, and D. Simchi-Levi. 2000. Quantifying the Bullwhip Effect: The Impact of Forecasting, Leadtime, and Information. *Man. Sci.* **46** 436-443.

Chen, H., P. Yang, D.D. Yao. 1994. Control and Scheduling in a Two-station Queueing Network: Optimal Policies and Heuristics. *Queueing Sys.* **18** 301-332.

Cohen, M.A., and S. Whang. 1997. Competing in Product and Service: A Product Life-Cycle Model. *Man. Sci.* **43** 535-545.

Dell, M., and J. Magretta. 1998. The Power of Virtual Integration. *Harvard Bus. Rev.* **76** 72-84.

Fitzsimmons, J.A., and M.J. Fitzsimmons 1998. *Service Management: Operations, Strategy, and Information Technology*. Boston: Irwin/McGraw-Hill.

Forrester, J.W. 1958. Industrial Dynamics: A Major Breakthrough for Decision Makers. *Harvard Business Rev.* **36** 37-66.

Gong, L. and H. Matsuo. 1997. Control Policy for a Manufacturing System with Random Yield and Rework. *J. of Opt. Theory and Applications* **95** 149-175.

Graves, S.C. 1986. A Tactical Planning Model for a Job Shop. *Oper. Res.* **34** 522-533.

Graves, S.C. 1999. A Single-Item Inventory Model for a Non-Stationary Demand Process. *Manufactur. & Service Oper. Man.* **1** 50-61.

Heskett, J.L., W.E. Sasser, and L.A. Schlesinger. 1997. *The Service-Profit Chain*. New York: Free Press.

Holt, C.C., F. Modigliani, J.F. Muth, and H.A. Simon. 1960. *Planning Production, Inventories, and Work Force*. Englewood Cliffs, NJ: Prentice-Hall.

Kamien, M. I. And N. L. Schwartz. 1991. *Dynamic Optimization: The Calculus of Variations and Optimal Control in Economics and Management.* Amsterdam: Elsevier Science.

Sterman, J. D., N. Repenning, and F. Kofman. 1997. Unanticipated Side Effects of Successful Quality Programs: Exploring a Paradox of Organizational Improvement. *Man. Sci.* **4** 503-521.

Lee, H.L., P. Padmanabhan, and S. Whang. 1997. Information Distortion in a Supply Chain: The Bullwhip Effect. *Man. Sci.* **43** 516-558.

Lee, H.L., K.C. So, and C.S. Tang. 2000. The Value of Information Sharing in a Two-Level Supply Chain. *Man. Sci*. **46** 626-643.

Lee, H.L. and S. Whang. 1999. Decentralized Multi-Echelon Supply Chains: Incentives and Information. *Man. Sci.* **45** 633-640.

Lilien, G.L., P.K., and K.S. Moorthy. 1992. *Marketing Models*. Englewood Cliffs, NJ: Prentice Hall.

Oliva, R. and J. D. Sterman. 2001. Cutting Corners and Working Overtime Quality Erosion in the Service Industry. *Man. Sci.* **48** 894-914.

Oppenheim, A.V., A.S. Willsky, and I.T. Young. 1983. *Signals and Systems*. Englewood Cliffs, NJ: Prentice-Hall.

Parsuraman, A., V. A. Zeithaml, and L. L. Berry. 1988. SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality. *J. of Retailing* **64** 12-40.

Papoulis, A. 1991. *Probability, Random Variables, and Stochastic Processes*. New York: McGraw Hill.

Rosberg, Z., P.P. Varaiya, J.C. Walrand. 1982. Optimal Control of Service in Tandem Queues. *IEEE Trans. on Auto. Control* **27** 600-610.

Shanmugan, K.S., and A.M. Breipohl 1988. *Random Signals: Detection, Estimation, and Data Analysis*. New York: Wiley.

Sethi, S. P. and G. L. Thompson. 2000. *Optimal Control Theory: Applications to Management Science and Economics.* Boston: Kluwer Academic Publishers.

Stengel, R.F. 1994. *Optimal Control and Estimation*. New York: Dover.

Sterman, J. D. 1989. Modeling Managerial Behavior: Misperceptions of Feedback in a Dynamic Decision Making Experiment. *Man. Sci.* **35** 321-339.

Sterman, J.D. 2000. *Business Dynamics: Systems Thinking and Modeling for a Complex World*. Boston: Irwin/McGraw-Hill.

Tang, C.S. 1990. The Impact of Uncertainty on a Production Line. *Man. Sci.* **36** 1518-1531.

Taylor, H. and S. Karlin. 1984. *An Introduction to Stochastic Modeling.* New York: Academic Press, Inc.