A TIME-SERIES FRAMEWORK FOR SUPPLY-CHAIN INVENTORY MANAGEMENT

YOSSI AVIV

Olin School of Business, Washington University, St. Louis, Missouri 63130, aviv@olin.wustl.edu

We consider a supply chain in which the underlying demand process can be described in a linear state space form. Inventory is managed at various points of the chain (members), based on local information that each member observes and continuously updates. The key feature of our model is that it takes into account the ability of the members to observe subsets of the underlying state vector, and adopt their forecasting and replenishment policies accordingly. This enables us to model situations in which the members are privy to certain explanatory variables of the demand, with the latter possibly evolving according to a vector autoregressive time series. For each member, we identify an associated demand evolution model, for which we propose an adaptive inventory replenishment policy that utilizes the Kalman filter technique. We then provide a simple methodology for assessing the benefits of various types of information-sharing agreements between members of the supply chain. We also discuss inventory positioning and cost performance assessment in the supply chain. Our performance metrics and inventory target levels are usually presented in matrix forms, allowing them to accommodate a relatively large spectrum of linear demand models, and making them simple to implement. Several illustrations for possible applications of our models are provided.

Received September 2000; revision received May 2002; accepted May 2002.
Subject classifications: Inventory/production: policies, uncertainty, applications. Forecasting: time series.
Area of review: Manufacturing, Services, and Supply-Chain Operations.

1. INTRODUCTION

In recent years, companies in various industries have been able to significantly improve their inventory management processes through integration of information technology into their forecasting and replenishment systems, and by sharing demand-related information with their supply-chain partners. As expected, the magnitude of the benefits resulting from implementation of the above practices often varies. While the management science literature has studied forecasting-integrated inventory policies, and evidenced a significant growth of research in the area of information sharing and supply-chain coordination, only a few papers address the combination of the two themes; namely, forecasting-integrated inventory policies and information sharing (see, e.g., Chen et al. 1999; Lee et al 2000; Aviv 2001, 2002).

The main objective of this paper is to propose a *methodology* for studying inventory systems in which demands evolve according to a relatively general class of statistical time-series patterns, and to propose building blocks for the construction of adaptive inventory policies for such settings. In addition, this paper provides a framework for assessing the impact of various information-sharing mechanisms on supply-chain performance. Our work is hence motivated by the following requirements faced by many supply-chain inventory managers in various industrial contexts. First, inventory planning tools should be able to incorporate a sufficiently *general* family of demand models and, specifically, demand processes that evolve over time as correlated time-series. This is because demands over consecutive time periods are rarely statistically independent (Kahn 1987, Lee et al. 2000), and because inventory systems need to manage a wide array of stock-keeping units (SKUs), sometimes numbering in the thousands. Second, demand models should be able to capture dependencies between demand realizations and other types of explanatory variables that can also evolve as correlated time-series. This need is particularly important in systems where data such as advanced customer orders, promotion plans, and expected weather conditions are collected and revised frequently. Finally, practitioners and researchers are faced today with the need to design and analyze information-sharing mechanisms in supply chains. As a consequence, a third requirement of analysis and planning tools is that they can accommodate cases in which information may not be commonly known across the supply-chain. In addition, they need to provide the ability to assess the potential benefits that can be gained by the supply-chain members if they share part or all of the data available to them with each other.

We provide below a brief review of the related literature. In terms of forecasting/inventory models, several approaches have been used. One such approach considers a case in which certain parameters of the demand distributions are unknown and uses Bayesian updating mechanisms to learn about future demand from past history (see, e.g., Scarf 1959, 1960; Iglehart 1964; Azoury 1985; Lovejoy 1990). A second approach has been to model the demand as a "classical" time-series (e.g., an autoregressive moving-average process) in which correlation exists between consecutive demand realizations; see Veinott (1965) and Johnson and Thompson (1975). A third approach has been to model the demand process as Markov-modulated; see, e.g., Song and Zipkin (1993). One can think of the latter as a case in which at the beginning of each period, the decision maker obtains information about the current "market conditions" (i.e., the exact state of the Markov chain), thus improving its forecast of future demand. While most of the abovementioned papers study the optimality of replenishment policies, we restrict our attention up-front to a particular class of forecast-adaptive policies. Specifically, our policies employ the *Kalman filter technique* to calculate minimum mean square error (MMSE) forecasts of future demands at each location of the supply chain. The periodically updated forecasts are used to determine inventory target levels and the order sizes needed in order to reach these levels. The advantage of our policies is that they are intuitive and easily implementable but, not less importantly, they can be tailored for use in information-rich supply chains, for which the characterization of optimal policies is virtually impossible.

A different type of approach to study forecasting issues in inventory management is that taken, e.g., by Chen et al. (1999). In their models they assume that the decision makers are not aware of the exact characteristics of the demand process, and hence they resort to popular forecasting mechanisms such as the moving-average technique when making replenishment decisions. Chen et al. also propose heuristic policies that are very similar in their nature to those proposed in this paper.

A large number of research papers have been written on the subject of information sharing in the management science literature, examining various types of benefits that can be achieved through different types of information-sharing mechanisms. Lee et al. (1997a, 1997b) discuss the so-called *bullwhip effect* in supply chains, a phenomenon in which the sequence of order quantities tend to have higher variability as one moves upstream in the supply chain (i.e., back to the supply side of the channel). One of the mechanisms they offer to mitigate the level of this phenomenon is by sharing point-of-sale data with the upstream members of the supply chain, hence enabling them to better anticipate future orders. Indeed, recent papers have proposed models to quantify the value of information sharing. Cachon and Fisher (2000) study a supply chain consisting of a single supplier and several retailers. They consider two impacts of information technology: speeding up the material flow in the supply chain, and expanding the flow of information. We refer the reader to their paper for a detailed survey of other related literature. Other examples of papers discussing information sharing are Bourland et al. (1996), Gavirneni et al. (1999), Lee et al. (2000), and more recently Aviv (2001). Aviv's paper is of particular relevance because it serves as an example of a supply-chain model in which not only demand realizations, but also the values of other explanatory variables, can be shared in the channel. Specifically, Aviv treats a model in which forecasting processes evolve at two stages of the supply chain. He studies the benefits of combining the decentralized forecasting processes into a centralized, co-managed forecasting process. His study sheds light on the potential gains that can be achieved by *Collaborative Forecasting Planning and Replenishment (CPFR)*, an industry-led initiative that has received a great deal of attention in recent years.

Lee et al. (2000) and Graves (1999) are the most closely related papers to our line of work, and in fact have inspired it. The approach we take is very similar to theirs, but two major differences should be pointed out. First, these papers focus on specific cases of time-series—the AR(1) model (see Example 1 in §2), and the ARIMA(0,1,1) model (see Example 2 in §2.1), respectively. Our model provides a structural time-series framework that is much more general. Second, we provide a broader framework for modeling information in the supply chain, and for studying the benefits of information sharing. This is thanks to our ability to embed explanatory variables into the state vector underlying the demand process. A recent article by Lee and Whang (2001) reports on the benefits that a demand-management solution provider was able to bring to a major U.S. drug chain by optimizing their demand-management process. One of the key value-added activities this third-party provider offers is the analysis of rich and timely data to generate demand forecasts and replenishment orders. We hence believe that the extension to linear state spaces provides the means for exploring a wide variety of timely issues in supply-chain inventory management. Aviv (2002) is an example for a study that uses the framework of this paper.

This paper constructs a *unified* framework that addresses the three requirements we specified earlier, simultaneously. Our approach is as follows: We describe the demand realization during each period as a linear function of a *state vector* that evolves as a vector autoregressive (VAR) time-series. Then, we propose a forecasting-adaptive inventory policy for each level of a two-stage supply chain. Extensions to multiproduct distribution channels follow. As far as information is concerned, our model allows one to describe situations in which the supply-chain members are able to observe different but possibly correlated parts of the underlying system state vector and integrate these observations into their forecasting and replenishment policies. We describe a simple method for investigating the impact of information sharing on supply-chain performance. Finally, we augment our forecasting and replenishment policies with a development of tools for cost assessment, and we discuss the possible use of these tools in addressing the problem of strategic inventory positioning along the supply chain. Our cost-assessment method provides an additional contribution to the literature—we show that costs in coordinated supply chains depend not only on the individual forecasting performance of the members of the supply chain, but also on the correlation between the so-called *lagged forecast errors*; see §6.1. The mathematical expressions for forecasts and inventory replenishment volumes are provided in matrix forms, and so they can be very easily translated into computer codes. Also, our cost-assessment techniques are simple, and can be implemented on spreadsheet applications.

The rest of this paper is organized as follows. Section 2 describes the notion of linear state space forms and briefly summarizes the Kalman filter technique used in our models. In §3, we begin with a treatment of the forecasting and inventory control problem for a single location system, specifying an adaptive order-up-to policy that will also serve as the basis for our more general supply-chain structures. In §4 we extend the discussion to a two-stage supply chain. We analyze the demand faced by the upper stage of the supply chain, and show that it follows the *same* type of evolution pattern as that of the demand faced by the lower stage. In §5 we describe possible extensions and applications of our framework, both for academic research and for practical analysis purposes. Particularly, we suggest extensions of our framework to multi-item, multiechelon supply chains, and we propose a method for assessing the potential benefits of information-sharing practices. Section 6 provides the means for assessing the inventory cost performance in coordinated supply chains, as well as decoupled supply chains. There we propose simulation-based algorithms for cost evaluations, as well as a simple bound on average systemwide costs. Section 7 concludes the paper.

2. PRELIMINARIES: LINEAR STATE SPACE FORMS AND THE KALMAN FILTER

In this section we briefly describe a linear state space model that serves as a building block for our inventory management policies. We then present a well-known forecasting technique associated with this model—namely, the Kalman filter. Let $\{X_t\}$ be a finite, n-dimensional vector process called the *state of the system*. In the context of inventory management, this vector may consist of early indicators of future demand in the channel, actual demand realizations at various points of the channel, and so forth. Suppose that the state vector evolves according to

$$\text{state space dynamics:} \quad X_t = FX_{t-1} + V_t, \tag{1}$$

with F being a known, time invariant, $n \times n$. The vectors $\{V_t \in \mathbf{R}^n\colon t \geqslant 1\}$ represent a white noise process, with each random vector V_t having a multivariate normal distribution with mean $\mathbf{0}_{n\times 1}$ and covariance matrix Σ_V. Next, assume that the decision maker partially or fully observes the state vector X_t. In other words, during period t, only the vector

$$\text{the observation equation:} \quad \Psi_t = HX_t \in \mathbf{R}^m, \tag{2}$$

is observed where H is a known $m \times n$ matrix. Hence, one can think of Ψ_t as the information "collected" by the decision maker during period t. In addition, let D_t represent the demand realized during period t, and suppose that the demand is a deterministic, linear function of the observed state vector,

$$\text{the demand equation:} \quad D_t = \mu + R\Psi_t, \tag{3}$$

where μ is a known scalar, and R is a $1 \times m$ vector of known parameters. To shorten the notation, we define the $1 \times n$ matrix $G \doteq RH$, so that $D_t = \mu + GX_t$. The following is an elementary example of such a model.

Example 1 (The AR(1) Model). Consider the autoregressive process of order 1, described by $D_t = \rho D_{t-1} + c + \epsilon_t$ (with $|\rho| < 1$), for all $t \geqslant 1$. An equivalent representation (1)–(3) for this case is: (i) $X_t = \rho X_{t-1} + \epsilon_t$, and (ii) $D_t = X_t + c/(1-\rho)$. Using our notation, we have: $F = \rho$, $R = H = 1$, $\mu = c/(1-\rho)$, and $\Sigma_V = \sigma^2$. (The choice of R and H is made arbitrarily from all combinations of parameters that satisfy $RH = 1$.)

We refer to the AR(1) model as an elementary example, because in this case one can think of the process $\{X_t\}$ as the actual demand realizations. While we refer to this model frequently, to relate some of our more general results to those existing in the literature (e.g., Lee et al. 2000), we see the contribution of our framework as particularly valuable in cases where X_t includes more information than the demand history. The need for such extension arises in various cases, such as: (i) when customers provide early information about their prospective orders; (ii) when planned promotions are known to one or more of the decision makers in the supply chain; or (iii) when a detailed point-of-sale data (e.g., at a store level) are provided directly to suppliers.

Because our main interest is in long-run characteristics of the system, we assume throughout this paper that the process extends back to periods $t = -1, -2, \ldots,$ unless otherwise specified. If the matrix F is *stable*, then in the long run, Equation (1) would have a unique stationary solution, given by $X_t = \sum_{j=0}^{\infty} F^j V_{t-j}$, and the corresponding demand process would also be stationary and given by $D_t = \mu + G\sum_{j=0}^{\infty} F^j V_{t-j}$ (see Chapter 12 in Brockwell and Davis 1996). (A matrix F is stable if F has all of its eigenvalues in the interior of the unit circle. If F is stable, the State Equation (1) is also said to be stable, or *causal*.)

Remark 1. The demand model (1)–(3) extends easily to cases in which the demand equation is of the form $D_t = \mu + GX_t + W_t$, with $\{W_t\}$ being a white noise process with $W_t \sim \mathrm{N}(\mu_W, \sigma_W)$ for all $t \geqslant 1$, as long as the processes $\{V_t\}$ and $\{W_t\}$ are uncorrelated at all nonzero lags (i.e., $\mathsf{E}[W_s \cdot V_t] = \mathbf{0}_{n\times 1}$ for all $s \neq t$). To see this, simply set: $\tilde{X}_t = (X_t,\ W_t)'$, $\tilde{\mu} = \mu + \mu_W$, $\tilde{G} = (G,\ 1)$, $\tilde{V}_t = (V_t,\ W_t)'$,

$$\tilde{F} = \begin{pmatrix} F & \mathbf{0}_{n\times 1} \\ \mathbf{0}_{1\times n} & 0 \end{pmatrix},$$

and

$$\tilde{\Sigma}_V = \begin{pmatrix} \Sigma_V & \mathsf{E}[W_t \cdot V_t] \\ \mathsf{E}[W_t \cdot V_t'] & \sigma_W^2 \end{pmatrix},$$

and note that the demand and system state evolution can be represented by $\tilde{X}_t = \tilde{F}\tilde{X}_{t-1} + \tilde{V}_t$, and $D_t = \tilde{\mu} + \tilde{G}\tilde{X}_t$, which maintains the structure (1)–(3).

2.1. The Kalman Filter

A fundamental element of the supply-chain's inventory control process is the estimation of future demand, based on which inventory and safety stock requirements are set. Let us focus on a single decision maker in the supply chain and assume that he faces a demand that follows the evolution model (1)–(3). To forecast future demand, it is useful for the decision maker to estimate the actual state of the system. To this end, suppose that we would like to compute the minimum mean square errors (MMSE) estimate of the state X_t, given the history of observations $\{\Psi_{t-1}, \Psi_{t-2}, \ldots, \Psi_1\}$, and denote this estimate by $\hat{X}_t = \mathsf{E}[X_t | \Psi_{t-1}, \Psi_{t-2}, \ldots, \Psi_1]$. Also, let $\Omega_t^X \doteq \mathsf{E}[(X_t - \hat{X}_t)(X_t - \hat{X}_t)']$ be the error covariance matrix associated with the estimate $\hat{X}_t$. A well-known method for computing the values of $\hat{X}_t$ and Ω_t^X for linear state space models of the type (1)–(2) is the *Kalman filter*, and in this section we provide a brief summary of this tool.

Generally speaking, the Kalman filter is a set of recursive equations used to regenerate estimates of the system state after every transition the system makes (i.e., at the beginning of each period, in our terms). Assume an initial state vector estimate $\hat{X}_1$ and an initial error covariance matrix Ω_1^X. Then, at the beginning of every period t, the estimate $\hat{X}_t$ is produced by

$$\hat{X}_t = F\hat{X}_{t-1} + FK_{t-1}(\Psi_{t-1} - H\hat{X}_{t-1}), \tag{4}$$

and its associated error covariance matrix is computed by

$$\Omega_t^X = F\Omega_{t-1}^X F' - FK_{t-1}H\Omega_{t-1}^X F' + \Sigma_V, \tag{5}$$

where $K_{t-1} = \Omega_{t-1}^X H'(H\Omega_{t-1}^X H')^{-1}$ is an $n \times m$ matrix, called the *Kalman gain* matrix. The matrix $(H\Omega_{t-1}^X H')^{-1}$ represents any *generalized inverse* of $H\Omega_{t-1}^X H'$ (a generalized inverse of an $m \times n$ matrix G always exists, and it is an $n \times m$ matrix G^{-1}, such that $GG^{-1}G = G$). Recall that the vector Ψ_t is observed by the decision maker *during* period t, *after* revising his forecasts and making his replenishment decisions at the beginning of that period, and just prior to the beginning of period $t+1$. This is the reason for including Ψ_{t-1}, and not Ψ_t in (4). The first term in (4) is the best forecast of X_t that can be generated using only the previous state estimate $\hat{X}_{t-1}$. The second term in (4) represents the adjustment to the estimate of X_t, based on the most recent observation, Ψ_{t-1}. Particularly, this term stands for the conditional expectation $\mathsf{E}\{X_t | \Psi_{t-1} - \mathsf{E}[\Psi_{t-1} | \Psi_{t-2}, \ldots, \Psi_1]\}$.

As mentioned in the introduction, we confine ourselves to cases in which the level of uncertainty surrounding future demand volumes is constant across time, or at least becomes constant when t is large. This property is clearly satisfied if the mean square error matrices Ω_t^X converge to a constant matrix (say) Ω^X, as t grows to infinity. When the Kalman filter applied to the model (1)–(3) has (or converges to) a time-invariant error covariance matrix Ω^X, we say that it is in (or converges to) a *steady state* (for related convergence and steady state conditions, see Anderson and Moore 1979 and Chan et al. 1984). For Kalman filters in steady state, we shall also use the time-invariant notation K, instead of K_{t-1}. If the matrix F is stable (i.e., has all eigenvalues inside the unit circle), this property is satisfied as long as the matrix Ω_1^X is *positive semidefinite (p.s.d.)*. But one can assume, for example, that $X_0 = \mathbf{0}_{n\times 1}$ and $\Omega_1^X = \Sigma_V$ are used, with the latter clearly being a p.s.d. matrix. When F is *not* stable, Chan et al. (1984) show that $\lim_{t\to\infty} \Omega_t^X = \Omega^X$ if the system is *observable*, and if $\Omega_1^X - \Omega^X$ is positive definite or $\Omega_1^X = \Omega^X$. The system (1)–(3) is observable if the state space X_t can be determined exactly given the observations $\Psi_t, \ldots, \Psi_{t+n-1}$, where n is the dimension of X_t. When F is unstable, it is necessary to characterize the distribution of the state vector X at a particular reference point (without loss of generality we refer to such a point as point zero). This can be done, for example, by providing a mean vector $\mathsf{E}[X_0]$ and a covariance matrix Σ_0 for the initial state X_0. Then, a recursive application of (1) yields $X_t = F^t X_0 + \sum_{j=0}^{t-1} F^j V_{t-j}$, and hence $D_t = \mu + G\sum_{j=0}^{t-1} F^j V_{t-j} + GF^t X_0$. Consider, for instance, the ARIMA(0,1,1) model:

Example 2 (The Arima(0,1,1) Model). Consider the autoregressive, integrated moving-average process: $D_t = D_{t-1} - (1-\alpha)\epsilon_{t-1} + \epsilon_t$ for all $t \geqslant 2$. In addition, we assume that $D_1 = \mu + \epsilon_1$ (or, alternatively, $D_0 = 0$ and $\epsilon_0 = 0$). An equivalent representation (1)–(3) for this case is

$$X_t = \begin{pmatrix} 1 & \alpha \\ 0 & 0 \end{pmatrix} X_{t-1} + \begin{pmatrix} 0 \\ \epsilon_t \end{pmatrix}, \quad D_t = \begin{pmatrix} 1 & 1 \end{pmatrix} X_t + \mu, \tag{6}$$

with the interpretation $X_t = (\alpha \sum_{i=1}^{t-1} \epsilon_{t-i},\ \epsilon_t)'$. Here, $H = (1,\ 1)$, $R = 1$, and Σ_V is a 2×2 matrix with $\Sigma_{V(2,2)} = \sigma^2$, and all of its other components are zeros.

In this example, the matrix F is unstable (eigenvalues $= \{0, 1\}$), but $\Omega_1^X = \Omega^X = \Sigma_V$. Thus, the Kalman filter for this model converges to a steady state, and so the framework in this paper can still be used to devise and characterize appropriate classes of *stationary* inventory policies for the ARIMA(0,1,1) case. In fact, a detailed treatment of this model is provided in Graves (1999).

The construction and applications of Kalman filters are extensively documented in the literature, and we refer the reader to Harvey (1990), Brockwell and Davis (1996, Ch. 12), and Hamilton (1994, Ch. 13) for a thorough description of this tool. Furthermore, the Kalman filter can be applied to state space forms that are more general than the one we assume in this paper, and it is widely used in control theory applications.

Remark. Because of the use of the normal distribution, negative demand realizations are not prohibited. This is particularly important in certain settings, in which the demand process can drift to the negative real line (for example, when the demand follows an AR(1) process with a large serial correlation). However, it should be emphasized that important insights into practical settings can still be gained, even if the stylized model permits negative demand volumes (see, e.g., Graves 1999), and it is therefore important that the normality assumption would be judged in a context-specific manner.

3. FORECASTING AND INVENTORY REPLENISHMENT AT THE FRONT END OF THE SUPPLY CHAIN

To construct our framework for inventory management in the supply chain, we first consider the inventory management problem at the consumption points, i.e., at the locations where customer demands are realized. We call these locations the *front-end points* of the supply chain, or simply the *end points*. An end point will usually refer to a combination of place (geographic) and product type (stock-keeping unit). Suppose that each end point faces consumer demand that evolves according to the linear state space model (1)–(3), where the parameters of the model and their interpretation are specific to the end point. The system works in a periodic review fashion, so that at the beginning of each period, replenishment orders are placed and later in the period demand is realized. Shortages are fully backlogged. The *nominal* supply lead times, excluding potential delays, is L periods for delivery of products from the immediate supplier of the particular end point; in other words, an order placed at the beginning of period t will ideally arrive at the *end* of period $t+L$. As mentioned earlier, we shall assume that by time t, the system has reached its steady state.

3.1. The Front End's Decision Problem

Consider a specific end point of the supply chain, and suppose that it does not face any supply delays for replenishment orders it places (we shall relax this assumption later). Let

$z_{t-L}, \dots, z_{t-1}$ = deliveries in transit (pipeline inventory), where z_{t-i} is the order placed in period $t-i$.

q_t = the inventory on hand at the beginning of period t, just after receiving the delivery of z_{t-L-1} units, placed in period $t-L-1$.

$C_n(q_t, z_{t-L}, \dots, z_{t-1}, \Psi_{t-1}, \Psi_{t-2}, \dots)$ = the minimum expected cost over a planning horizon of the n periods $t, \dots, t+n-1$. We set $C_0 \equiv 0$.

$c(q)$ = the single-period cost charged when the end-of-period net inventory on hand (actual units on-hand minus backlogs) is equal to q units. We assume that the function c is convex in q.

Proposition 1. *Consider the Kalman filter mechanism (4)–(5) in steady state, and suppose that the matrix $F-FKH$ has all its eigenvalues inside the unit circle. Then, the functions C_n satisfy the dynamic-programming scheme*

$$
\begin{aligned}
&C_n(q_t, z_{t-L}, \dots, z_{t-1}, \Psi_{t-1}, \Psi_{t-2}, \dots) \\
&= \min_{z_t \geqslant 0} \Bigg\{ \mathsf{E}_{D_t}\Bigg[c(q_t - D_t) \\
&\quad + \mathsf{E}_{\xi_t}\Bigg[C_{n-1}\Bigg(q_t - D_t + z_{t-L}, z_{t-L+1}, \dots, z_t, \xi_t \\
&\quad + H\sum_{i=1}^{\infty}(F-FKH)^i FK\Psi_{t-i}, \Psi_{t-1}, \Psi_{t-2}, \dots \Bigg)\Bigg] \Bigg| \\
&\qquad \Psi_{t-1}, \Psi_{t-2}, \dots \Bigg]\Bigg\},
\end{aligned}
$$

where $\xi_t \in \mathbf{R}^p$ has a multivariate normal distribution with mean zero and covariance matrix $H\Omega^X H'$, and it is independent of $\Psi_{t-1}, \Psi_{t-2}, \dots$.

Proof. Note that $\Psi_t = HX_t = H(X_t - \hat{X}_t) + H\hat{X}_t$. The term $\xi_t \doteq H(X_t - \hat{X}_t)$ has the distribution described in the proposition, because the error covariance matrix of the term $X_t - \hat{X}_t$ is given by Ω^X. Furthermore, ξ_t is independent of the information history $\{\Psi_{t-1}, \Psi_{t-2}, \dots\}$ because the estimates $\hat{X}_t$ are optimal. The term $H\hat{X}_t$ can be recursively developed using (4): $H\hat{X}_t = HFK\Psi_{t-1} + H(F-FKH)\hat{X}_{t-1} = \cdots = H\sum_{i=1}^{\infty}(F-FKH)^i FK\Psi_{t-i}$. □

Of course, it is virtually impossible to solve the dynamic program in Proposition 1, because its state space has a high dimension. We therefore continue with a simplification of the state of the dynamic program, which will then allow us to propose a heuristic inventory replenishment policy of a convenient structure. Essentially, as is common in the inventory management literature (e.g., Veinott 1965), we redefine the one-period costs c so that they are charged lead-time periods in advance. This is fine, because in the long run this phased accounting scheme does not have any effect on average cost calculations. We define $\tilde{q}_t \doteq q_t + z_{t-L} + \cdots + z_{t-1}$ as the end point's inventory *position* at the beginning of period t, before replenishment order is determined.

Proposition 2. *Let $\tilde{c}(\tilde{q}_t, \hat{X}_t, \Psi_{t-1}, \Psi_{t-2} \cdots) \doteq \mathsf{E}[c(\tilde{q}_t - \sum_{l=0}^{L} D_{t+l}) | \hat{X}_t, \Psi_{t-1}, \Psi_{t-2} \cdots]$. The function $\tilde{c}$ depends on the history $\{\Psi_{t-1}, \Psi_{t-2} \cdots\}$ through the most recent estimate $\hat{X}_t$ only, and hence can be written as $\tilde{c}(\tilde{q}_t, \hat{X}_t)$. The decision-making problem can be defined as follows:*

$$
\begin{aligned}
\tilde{C}_n(\tilde{q}_t, \hat{X}_t) = \min_{z_t \geqslant 0} \big\{ &\tilde{c}(\tilde{q}_t + z_t, \hat{X}_t) + \mathsf{E}_{\tilde{\xi}_t}\big[\tilde{C}_{n-1}(\tilde{q}_t + z_t \\
&- RH\tilde{\xi}_t - RH\hat{X}_t - \mu, F\hat{X}_t + FKH\tilde{\xi}_t)\big]\big\},
\end{aligned}
$$

where $\tilde{\xi}_t \doteq X_t - \hat{X}_t$. The vectors $\{\tilde{\xi}_t\}$ are i.i.d., each having a multivariate normal distribution with mean zero and covariance matrix Ω^X.

Proof. Recall that $D_t = RHX_t + \mu$. Hence, it follows from (1) that for every $l \geqslant 1$, the demand D_{t+l} is given by $D_{t+l} = \mu + RHF^l X_t + RH\sum_{j=1}^{l} F^{l-j} V_{t+j}$, and therefore the lead-time demand satisfies

$$
\begin{aligned}
\sum_{l=0}^{L} D_{t+l} = {}&(L+1)\mu + RH\left(\sum_{l=0}^{L} F^l\right)\hat{X}_t \\
&+ RH\left(\sum_{l=0}^{L} F^l\right)\tilde{\xi}_t + RH\sum_{l=1}^{L}\left(\sum_{j=0}^{L-l} F^j\right)V_{t+l}. \qquad (7)
\end{aligned}
$$

Next, note that the random vectors $\{V_{t+l}\colon l \geqslant 1\}$ and $\tilde{\xi}_t$ are independent of each other and of the history $\{\Psi_{t-1}, \Psi_{t-2}, \dots\}$. The dynamics of the inventory position is given by $\tilde{q}_{t+1} = \tilde{q}_t + z_t - D_t$, with $D_t = RH(\tilde{\xi}_t + \hat{X}_t) + \mu$, and the dynamics of $\hat{X}_t$ is given by the Kalman filter Equation (4). This justifies the dynamic program scheme for $\tilde{C}_n$. □

The dynamic program in Proposition 2 represents the decision maker's problem as an attempt to find the best inventory position level, given a multidimensional information state $\hat{X}_t$. For this type of problem, it is easy to verify the following.

PROPOSITION 3. *Suppose that* $\lim_{|\tilde{q}_t|\to\infty} \tilde{c}(\tilde{q}_t, \cdot) = \infty$, *and assume that* $\tilde{C}_0(\cdot,\cdot) \equiv 0$. *Then, a state-dependent order-up-to policy is optimal for the problem* $\tilde{C}_n$. *In other words, there exists a sequence of target-level functions* $\{\beta_n^*(\hat{X}_t)\colon n = 1, 2, \ldots\}$, *such that if* $\tilde{q}_t < \beta_n^*(\hat{X}_t)$, *an order for the difference* $\beta_n^*(\hat{X}_t) - \tilde{q}_t$ *is placed. Otherwise, no order is placed.*

The optimality of order-up-to policies is straightforward from the convexity of $\tilde{c}$, the linearity of the Kalman filter ($\hat{X}_{t+1} = F\hat{X}_t + FKH\tilde{\xi}_t$), and from the fact that all moments of the lead-time demands $\sum_{l=0}^{L} D_{t+l}$ and error terms $\tilde{\xi}_t$ are finite (due to the properties of the normal distribution). A formal proof can be easily written, using an induction on n. Furthermore, the extension to the case $n \to \infty$ only requires some standard adaptation from the inventory management literature; see, e.g., Karlin (1960). In the long-run problem, when $n \to \infty$, the target level is given by a function $\beta^*(\hat{X}_t)$. This implies that a good inventory policy needs to prescribe order quantities on the basis of the estimates $\hat{X}_t$, and that it suffices for this purpose to keep track of the most recent estimate only. With this in mind, we move on to our proposed heuristic inventory replenishment policy.

3.2. An Adaptive Replenishment Policy

We propose for each end point a replenishment policy that belongs to the following class of *installation-based* order-up-to policies: The end point sets an inventory position target level that is equal to its updated estimate of the aggregate demand during the immediate lead time, plus a *fixed* safety stock. The estimates of the lead-time demand $\sum_{l=0}^{L} D_{t+l}$ are given by

$$\hat{D}_t^{(L)} \doteq \mathsf{E}\left[\sum_{l=0}^{L} D_{t+l} | \Psi_{t-1}, \Psi_{t-1}, \ldots\right]$$
$$= (L+1)\mu + RH\left(\sum_{l=0}^{L} F^l\right)\hat{X}_t. \qquad (8)$$

In period t, the following target level is used:

$$\beta_t^1 = \hat{D}_t^{(L)} + \gamma^1, \qquad (9)$$

where γ^1 is a fixed safety-stock level set for the end point (we use the superscript 1 to denote an end point). Thus, although $\beta^*(\hat{X}_t)$ of Proposition 3 can take any functional form, we restrict ourselves to a target level β^1, which is a very specific linear function of the vector estimate $\hat{X}_t$. The reason we use a fixed safety stock in (9) is that the level of uncertainty surrounding the lead-time demands is constant over time. Specifically, in the long run, the variance $\Omega_t^D \doteq \mathsf{E}[(\hat{D}_t^{(L)} - \sum_{l=0}^{L} D_{t+l})^2]$ is given by

PROPOSITION 4. *Let* $G \doteq RH$. *In the long run, the mean square error of the estimates* $\hat{D}_t^{(L)}$ *is equal to*

$$\Omega^D \doteq \sum_{l=1}^{L}\left[G\left(\sum_{j=0}^{L-l} F^j\right)\Sigma_V\left(\sum_{j=0}^{L-l} F^j\right)' G'\right]$$
$$+ G\left(\sum_{j=0}^{L} F^j\right)\Omega^X\left(\sum_{j=0}^{L} F^j\right)' G' \qquad (10)$$

PROOF. Consider the expression for $\sum_{l=0}^{L} D_{t+l}$ provided in (7). The proof then follows because the random vectors $V_{t+1}, \ldots, V_{t+L}$, and $X_t - \hat{X}_t$ are independent. □

The restriction to the above class of policies is obviously convenient from a practical point of view: The Kalman filter technique is readily available to generate the estimates $\hat{X}_t$, and (8) is used to compute the corresponding values of $\hat{D}_t^{(L)}$. In fact, Lee et al. (2000) and Graves (1999) use the above class of inventory replenishment policies in their analysis of the AR(1) and ARIMA(0,1,1) models, respectively. Both of these papers use the same MMSE forecasts that can be obtained by applying the Kalman filter tool, although they do not refer to this technique explicitly. Another example is the Martingale Model for Forecast Evolution (MMFE) developed by Graves et al. (1986) and Heath and Jackson (1994). Here, as well, the evolution of forecasts is equivalent to what would be generated if the Kalman filter technique is used. Other examples for the use of this policy class can be found in Lee et al. (1997a) and Chen et al. (1999), to mention but a few. In fact, in many papers the policy (9) is specifically confined to *myopic* order-up-to policies in which the safety-stock parameter γ^1 is set in a way that minimizes the expected cost at the end of a lead time. Because myopic policies are technically and practically appealing, we discuss them further in §3.2.1 below. After that, in §3.2.2, we provide additional remarks about the complete role of the parameter γ^1. We explain under what circumstances it may be beneficial to set the safety stock γ^1 to a level that is different than that suggested by a myopic policy.

3.2.1. Myopic Policies. Suppose, as is common in the literature on inventory management, that $\tilde{c}(\tilde{q}_t, \hat{X}_t)$ can be written as a convex function $\tilde{c}'(\tilde{q}_t - \hat{D}_t^{(L)})$, and let e^* be the value of e that minimizes $\tilde{c}'(e)$. For the sake of the discussion, let us assume that e^* is unique. Using a similar argument as in Veinott (1965), we can easily show that if $\tilde{q}_t - \hat{D}_t^{(L)}$ is smaller than e^* for every t, then it is optimal to order in each period t the amount

$$z_t = \hat{D}_t^{(L)} + e^* - \tilde{q}_t. \qquad (11)$$

This is a special case of Policy (9) and it is a myopic policy because replenishment decisions are based on the best way to handle the single-step cost function $\tilde{c}'$. Johnson and Thompson (1975) prove the optimality of a *myopic* order-up-to policy for a system that faces a demand that follows an autoregressive moving-average (ARMA) process.

To do so, they truncate the normal distributions that characterize the components of the ARMA process by imposing a set of conditions on the parameters of the underlying demand process (see §3 there). Also consider the inventory model with zero set-up costs discussed in §5 of Gallego and Özer (2001). In Theorem 5 of their paper, they prove that a myopic order-up-to policy is optimal for an inventory system that faces a demand with advanced orders. This is shown by observing that the "myopic" target level can be reached in each period (see specifically the observation "$x < y_{n+1}^m$" in the proof of Theorem 5 of their paper). We note that in Gallego and Özer's paper, orders for specific periods are placed up to N periods in advance; if these orders can be characterized as normally distributed, their model in §5 would turn out to be a special case of our linear state space model. As another illustration, in the context of Markov-modulated demand, take for example the zero order cost model discussed in §4.2 of Song and Zipkin (1993). They show (Theorem 9) that when the demand rate does not decrease over time, the myopic order-up-to level is always attainable, and thus it is optimal.

Nevertheless, as they demonstrate later in Song and Zipkin (1996), a myopic policy can be substantially suboptimal when demand rate can drop over time, as is often experienced when products become obsolete. We refer the reader to Veinott (1965), Azoury (1985), and to the later paper of Lovejoy (1990) for further detailed discussion on the optimality (or near optimality) of myopic policies for inventory models with parameter-adaptive demand process. Specifically, the models treated by Lovejoy consider situations in which a parameter of the demand process is unknown and estimates of it are regenerated progressively over time as demands are realized.

A slightly different approach to the justification of myopic order-up-to policies was taken by Lee et al. (2000) and most recently by Dong and Lee (2001): If one can show that β_t^1 can be met *most* of the time, instead of at the beginning of each and every period, then a myopic order-up-to policy will be close to optimal. Note that in this case, the replenishment order A_t set by the end point at the beginning of period t, can be approximately given by

$$\begin{aligned} A_t &= \beta_t^1 - \tilde{q}_t = (\hat{D}_t^{(L)} + \gamma^1) - (\hat{D}_{t-1}^{(L)} + \gamma^1 - D_{t-1}) \\ &= \hat{D}_t^{(L)} - \hat{D}_{t-1}^{(L)} + D_{t-1}, \end{aligned} \tag{12}$$

or, in other words, $A_t = \beta_t^1 - \tilde{q}_t$, instead of $A_t = \max(\beta_t^1 - \tilde{q}_t, 0)$. But how appropriate is this assumption? Clearly, Equation (12) may be viewed as a good approximation for policies that do *not* permit negative orders, as long as $\Pr\{\hat{D}_t^{(L)} - \hat{D}_{t-1}^{(L)} + D_{t-1} \geqslant 0\} \approx 1$. When the matrix F is stable, the latter probability is given by $1 - \Phi(-\mathsf{E}[A_t]/\sqrt{\mathsf{Var}(A_t)})$. We provide later, in Proposition 7, a formula for calculating the long-run variance of the orders $\{A_t\}$. For example, in the AR(1) model of Example 1, Proposition 7 yields

$$\begin{aligned} &\text{AR(1): } \Pr\{\hat{D}_t^{(L)} - \hat{D}_{t-1}^{(L)} + D_{t-1} \geqslant 0\} \\ &= 1 - \Phi\left(-\frac{c}{\sigma} \cdot \frac{1}{\sqrt{\rho^{2L+4} \cdot (1-\rho)/(1+\rho) + (1-\rho^{L+2})^2}} \right). \end{aligned}$$

When $\rho = 0$, the last equation is reduced to $1 - \Phi(-c/\sigma)$, which is close to one when σ/c is relatively small. Furthermore, it can be easily seen that the probability above becomes even closer to one as ρ increases in the range $[0, 1)$. This specific result is also shown in Lemma 1 of Lee et al. 2000.

In our analysis we assume that (12) *always* holds, or alternatively, we permit negative values of A_t. This assumption is essential when we expand our discussion to more complex information-rich settings. Clearly, as implied by the discussion above, a myopic order-up-to policy may fail to be optimal if the order-up-to level cannot be reached in almost every period. In such cases, myopic policies may perform poorly, and it would be reasonable to assume that their performance vis-a-vis optimal policies would worsen as $\Pr\{\hat{D}_t^{(L)} - \hat{D}_{t-1}^{(L)} + D_{t-1} < 0\}$ grows. As explained in the following section, that is one of the reasons for which we suggest an adjustment to the parameter γ^1 in the target level (9).

3.2.2. The Safety-Stock Parameter γ^1. We start with a special case of the single-location (end-point) system, by investigating what happens if A_t can be negative during some periods.

Proposition 5. *Recall Proposition* 2. *If $RHF^{L+1}\hat{X}_t = 0$ for all $\hat{X}_t \in \mathbf{R}^n$, and $\tilde{c}(\tilde{q}_t, \hat{X}_t)$ can be written as a convex function $\tilde{c}'(\tilde{q}_t - \hat{D}_t^{(L)})$, then the policy structure* (9) *is optimal.*

Proof. We provide a sketch of a simple proof. Let $e_t = \tilde{q}_t - \hat{D}_t^{(L)}$ be the "effective" inventory position, and prove by induction that $\tilde{C}_n$ can be written as a function $\tilde{C}_n'$ of e_t. Suppose this is true for n. Then,

$$\begin{aligned} \tilde{C}_{n+1}(\tilde{q}_t, \hat{X}_t) &= \min_{z_t \geqslant 0} \Big\{ \tilde{c}'(e_t + z_t) + \mathsf{E}_{\tilde{\xi}_t}[\tilde{C}_n(\tilde{q}_t + z_t - RH\tilde{\xi}_t \\ &\qquad - RH\hat{X}_t - \mu, F\hat{X}_t + FKH\tilde{\xi}_t)] \Big\} \\ &= \min_{z_t \geqslant 0} \Big\{ \tilde{c}'(e_t + z_t) + \mathsf{E}_{\tilde{\xi}_t'}[\tilde{C}_n'(e_t + z_t - \tilde{\xi}_t')] \Big\} \\ &= \tilde{C}_{n+1}'(e_t), \end{aligned}$$

where $\tilde{\xi}_t' = \mu + RH(I + \sum_{l=1}^{L} F^l KH)\tilde{\xi}_t$. The second equality follows from $-\mu - RH\tilde{\xi}_t - RH\hat{X}_t - RH(\sum_{l=0}^{L} F^l)(F\hat{X}_t + FKH\tilde{\xi}_t) = -RH(\sum_{l=0}^{L} F^l)\hat{X}_t - \tilde{\xi}_t'$, in view of (8) and $RHF^{L+1}\hat{X}_t = 0$. Clearly, for the simple one-dimensional dynamic program above, a static order-up-to level policy is optimal. This implies that the policy structure (9) is optimal. □

Despite the fact that Proposition 5 only considers a special case, it makes an important point: On one hand, the myopic policy (i.e., $\gamma^1 = e^*$) may fail to be optimal if $\hat{D}_t^{(L)} - \hat{D}_{t-1}^{(L)} + D_{t-1}$ is sometimes negative, or in other words, if $\tilde{\xi}_t'$ is a random variable that can often take negative values. On the other hand, the structure of the optimal policy does not change—Policy (9) can still be used, albeit with γ^1 being set to the target-level solution of the dynamic program $\tilde{C}_{n+1}'(e_t) = \min_{z_t \geqslant 0}\{\tilde{c}'(e_t + z_t) + \mathsf{E}_{\tilde{\xi}_t'}[\tilde{C}_n'(e_t + z_t - \tilde{\xi}_t')]\}$.

The second possible reason for why a myopic order-up-to policy may not work well is because of potential supply delays, beyond the nominal lead time L, that may occur due to shortages at the front end's supplier. To design our inventory policies to cope with potential supply delays, we propose an upward adjustment in the parameter γ^1, and so (9) is perhaps no longer a myopic policy. Numerical analyses in Aviv (2001, 2002) suggest that substantial benefits can be gained by such an adjustment. We elaborate on the technical details associated with the determination of γ^1 in §6.

4. A TWO-STAGE SUPPLY CHAIN

We now extend our focus to a simple supply chain that consists of *two* members: A retailer (end point), and its supplier. To develop inventory policies for the supply-chain members, one needs to keep in mind the interaction between these members. On one hand, the retailer needs to take into account the delivery performance of the supplier, or in other words, possible delays of shipments of goods beyond the nominal lead time L. On the other hand, the supplier needs to forecast the orders that will be placed by the retailer, using all available information the supplier has. As we shall see, this is not a simple task, and in fact, the main purpose of this section is to discuss the supplier's forecasting process. First, for the retailer, we still propose the same type of forecasting and inventory policy described in the previous section. The way the retailer deals with possible supply shortages is by setting the safety-stock level γ^1 appropriately (see §6). As for the supplier's policy, we proceed as follows: In §4.1 we study the behavior of the process $\{A_t\}$—the sequence of orders placed by the retailer. We show that the retailer's orders evolve according to a linear state space. A myopic inventory policy similar in structure to that of the retailer then follows. In §4.2, we extend the discussion to consider cases in which the supplier, and not only the retailer, observes part of the system state vector X_t. While the analysis of this model extension becomes more complex, we are still able to show that the supplier's demand and information processes evolve according to a linear state space model of Type (1)–(3). This is quite pleasing, because we can adopt for the supplier the same type of forecasting and replenishment policy described in §3.

The system still works in a periodic review fashion, and the order of events is such that the retailer places his order for period t *before* the supplier needs to determine his replenishment order (dealing with the opposite order of events requires some straightforward alterations to our models). Therefore, the supplier can make use of the information A_t when making his replenishment decision in period t. The supplier faces a fixed delivery lead time of τ periods. We assume again that the retailer's inventory and forecasting process is in steady state.

4.1. The Orders Generated by the Retailer

Understanding the way by which the order quantities A_t evolve becomes useful when we develop an inventory policy for the supplier. We assume initially that the only source of demand information for the supplier is the history of orders placed by the retailer. We begin with the following observation.

Theorem 1. *Suppose that the order quantities $\{A_t\colon t \geqslant 1\}$ are determined according to* (12), *and that the Kalman filter* (4)–(5) *is in steady state. Then*

$$A_t = RHF^{L+1}\hat{X}_{t-1} + \Theta(\Psi_{t-1} - H\hat{X}_{t-1}) + \mu, \tag{13}$$

where $\Theta \doteq R[H(\sum_{l=1}^{L+1} F^l)K + I_{m\times m}]$ is a $1 \times m$ vector.

Proof. By (8) and (12), we obtain $A_t = RH(\sum_{l=0}^{L} F^l)(\hat{X}_t - \hat{X}_{t-1}) + D_{t-1}$. Next, use (4) to verify that $\hat{X}_t - \hat{X}_{t-1} = (F - I_n)\hat{X}_{t-1} + FK(\Psi_{t-1} - H\hat{X}_{t-1})$. Now note that $RH(\sum_{l=0}^{L} F^l)(F - I_n) = RHF^{L+1} - RH$. Therefore, $A_t = RHF^{L+1}\hat{X}_{t-1} - RH\hat{X}_{t-1} + RH(\sum_{l=1}^{L+1} F^l)K(\Psi_{t-1} - H\hat{X}_{t-1}) + R\Psi_{t-1} + \mu$. □

Essentially, the representation of A_t in the last proposition is equivalent to the following expression:

$$\begin{aligned}
A_t &= \mathsf{E}\left[\sum_{l=0}^{L} D_{t+l} \mid \Psi_{t-1}, \Psi_{t-2}, \ldots\right] \\
&\quad - \mathsf{E}\left[\sum_{l=0}^{L} D_{t-1+l} \mid \Psi_{t-2}, \ldots\right] + D_{t-1} \\
&= \mathsf{E}\left[\sum_{l=0}^{L} D_{t+l} - \sum_{l=0}^{L} D_{t+l-1} + D_{t-1} \mid \Psi_{t-2}, \ldots\right] \\
&\quad + \mathsf{E}\left[\sum_{l=0}^{L} D_{t+l} + D_{t-1} \mid \Psi_{t-1} - H\hat{X}_{t-1}\right] \\
&= \mathsf{E}\left[D_{t+L} \mid \Psi_{t-2}, \Psi_{t-3}, \ldots\right] \\
&\quad + \mathsf{E}\left[\sum_{l=0}^{L+1} D_{t-1+l} \mid \Psi_{t-1} - H\hat{X}_{t-1}\right]
\end{aligned}$$

The first component of the last expression is the estimate of A_t based on the information available at the retailer at the beginning of period t, and is given by $RHF^{L+1}\hat{X}_{t-1} + \mu$. The second term is based on the *forecast error* $\Psi_{t-1} - H\hat{X}_{t-1}$, observed by the retailer during period $t-1$, and it corresponds to the second term in (13). The interpretation of Equation (13) is quite interesting. Suppose that at the beginning of period t, the retailer wishes to forecast the next demand volume D_t, and that the supplier wants to forecast his next demand volume A_{t+1}. Furthermore, suppose that the supplier can retrieve the information $\{\Psi_{t-1}, \Psi_{t-2}, \ldots\}$ from the history of orders $\{A_t, A_{t-1}, \ldots\}$. Then, the standard errors of these forecasts are given by

$$\mathsf{Std}(A_{t+1} \mid \Psi_{t-1}, \Psi_{t-2}, \ldots) = \sqrt{\Theta H \Omega^X H' \Theta'}$$

and

$$\mathsf{Std}(D_t|\Psi_{t-1}, \Psi_{t-2}, \dots) = \sqrt{RH\Omega^X H'R'}.$$

Thus, we observe that even if both supply-chain members observe the data Ψ, the forecast errors of the demands at both levels may be different. Consider, for example, the ARIMA(0,1,1) model of Graves (1999), and recall that in his setting the supplier fully observes the information $\{\Psi_{t-1}, \Psi_{t-2}, \dots\}$ through the history of orders $\{A_t, A_{t-1}, \dots\}$. It may be easily shown that for the ARIMA model, $\Theta = 1 + (L+1)\alpha$ and that $H\Omega^X H' = \sigma^2$. Thus, $\mathsf{Std}(A_{t+1}|\Psi_{t-1}, \Psi_{t-2}, \dots)/\mathsf{Std}(D_t|\Psi_{t-1}, \Psi_{t-2}, \dots) = 1 + (L+1)\alpha$. The last term describes the amplification of demand uncertainty as one goes upstream the supply chain, moving from the retailer to its supplier. Similar result can be shown with respect to the AR(1) model, described in Example 1. In this case, $\Theta = (1-\rho^{L+2})/(1-\rho)$, and $H\Omega^X H' = \sigma^2$. Therefore, the amplification of demand uncertainty in this case is given by $(1-\rho^{L+2})/(1-\rho)$; see also Observation 2 in Raghunathan (2001).

In general, we observe that the retailer's orders (or, alternatively, the supplier's demands) follow the evolution.

Proposition 6. *Let* $\xi_t \doteq \Psi_t - H\hat{X}_t$, *and define the* $(n+m)\times 1$ *state vector* $Y_t \doteq (\hat{X}_t' \quad \xi_t')'$. *The following state space form represents the evolution of the process* $\{A_t\}$: $Y_t = F_A Y_{t-1} + \Xi_t$, *and* $A_{t+1} = G_A Y_t + \mu$, *where*

$$F_A = \begin{pmatrix} F & FK \\ \mathbf{0}_{m\times n} & \mathbf{0}_{m\times m} \end{pmatrix}, \quad \Xi_t = \begin{pmatrix} \mathbf{0}_{n\times 1} \\ \xi_t \end{pmatrix} \quad \textit{and}$$

$$G_A = \begin{pmatrix} RHF^{L+1} & \Theta \end{pmatrix}.$$

The matrix F_A *is stable if and only if* F *is stable.*

Proof. Verifying that $Y_t = F_A Y_{t-1} + \Xi_t$ is straightforward from (4). The expression for A_{t+1} follows directly from Theorem 1. For the second part, note that the spectrum of eigenvalues of F_A is equal to that of F, plus the eigenvalue 0, if not already an eigenvalue of F. Therefore, F_A is stable if and only if F is stable. □

Like the process $\{V_t\}$, the process $\{\Xi_t\}$ is a white noise, with vector-mean zero, and covariance matrix

$$\Sigma_\Xi = \begin{pmatrix} \mathbf{0}_{n\times n} & \mathbf{0}_{n\times m} \\ \mathbf{0}_{m\times n} & H\Omega^X H' \end{pmatrix}.$$

The significance of Proposition 6 is that it demonstrates that the orders placed with the supplier evolve according to the same type of dynamics that governs the underlying demand process. From the supplier's point of view, the system state vector consists of the most recent retailer's estimate of the underlying system state X_t and the retailer's *observed* forecast error $\Psi_t - H\hat{X}_t$. Hence, in an analogy to the retailer's forecasting and inventory process, we offer the following policy for the supplier: At the beginning of period t, update the estimate of Y_t (denoted by $\hat{Y}_t$) using the Kalman filter Equations (4)–(5), with $F_A, K_A, A_t, G_A, \Omega^Y$, and Σ_Ξ, instead of $F, K, \Psi_{t-1}, H, \Omega^X$, and Σ_V. Next, the forecast of the supplier's lead-time demand $\sum_{l=0}^{\tau-1} A_{t+1+l}$, denoted by $\hat{A}_t^{(\tau)} \doteq \mathsf{E}[\sum_{l=0}^{\tau-1} A_{t+1+l}|A_t, \dots, A_1]$, is computed via $\hat{A}_t^{(\tau)} = \tau\mu + G_A(\sum_{l=0}^{\tau-1} F_A^l)\hat{Y}_t$; see (8). Finally, a replenishment order is placed for a number of units that would bring the supplier's inventory position as close as possible to the target level

$$\beta_t^2 = \hat{A}_t^{(\tau)} + \gamma^2 \tag{14}$$

for some fixed safety stock γ^2. We describe in §6 possible ways to determine the value of γ^2.

The analysis above also allows us to provide a simple equation for calculating the long-run variance of the orders $\{A_t\}$, denoted below by $\mathsf{Var}(A_t)$. This measure is useful, for example, when one wants to assess the long-run portion of periods in which negative orders are prescribed; see our discussion in §3.2.1. For the measure $\mathsf{Var}(A_t)$ to exist, we require that the matrix F is stable.

Proposition 7. *Suppose that the matrix* F *of* (1) *is stable. Then*

$$\mathsf{Var}(A_t) = RHF^{L+1}\left[\sum_{j=1}^{\infty} F^j KH\Omega^X H'K'(F^j)'\right](RHF^{L+1})' + \Theta H\Omega^X H'\Theta'. \tag{15}$$

Proof. For a demand D_t that follows the linear state space form (1)–(3), it is easy to verify that $\mathsf{Var}(D_t) = G[\sum_{j=0}^{\infty} F^j \Sigma_V (F^j)']G'$. We apply this to the linear state representation of Proposition 6. Note first that for all $j \geqslant 1$, the upper left $(n\times m)$-sub matrix of $F_A^j \Sigma_\Xi (F_A^j)'$ is given by $F^j KH\Omega^X (F^j KH)'$, and all other entries of the matrix are zero. Thus,

$$\sum_{j=0}^{\infty} F_A^j \Sigma_\Xi (F_A^j)' = \begin{pmatrix} \sum_{j=1}^{\infty} F^j KH\Omega^X (F^j KH)' & \mathbf{0}_{n\times m} \\ \mathbf{0}_{m\times n} & H\Omega^X H' \end{pmatrix}.$$

The rest follows by simple matrix algebra. □

The variance of the orders $\{A_t\}$ is also of significant interest when the supplier faces capacity constraints (not discussed in this paper). Thus, the term in (7) has a potential use in developing replenishment strategies for the supplier in such applications.

4.1.1. Imperfect Knowledge of the System Parameters. In general, we assume that all members of the supply chain have a complete knowledge of the underlying system's *static* parameters: demand evolution parameters, the type of information observed by each member, and the types of inventory and forecasting policies used by each member. While this is often a common assumption in the literature, it may not be valid in some practical settings. For example, suppose that the retailer is able to forecast the demand for the product based on knowledge of planned promotions for the product, but the supplier is not aware of such promotions and their exact correlation with future demand for the product. Thus, the policy we devised for the supplier in §4.1 cannot be implemented. We argue, however, that the analysis in §4.1 is still valuable. Using the evolution of $\{A_t\}$ described in Proposition 6, we are able to show that this process evolves according to an AR(∞) process. In other words, the value of the demand A_{t+1} satisfies $A_{t+1} = \sum_{i=0}^{\infty} \kappa_i A_{t-i} + \epsilon_{A,t+1}$, for a white-noise process $\{\epsilon_{A,t}\}$.

PROPOSITION 8. *Suppose that the supplier has a full knowledge of the static system parameters, and that he uses a the steady state Kalman filter described in §4.1. Then, the process A_t follows the evolution*

$$A_{t+1} = \epsilon_{A,t+1} + G_A \sum_{i=1}^{\infty} (F_A - F_A K_A G_A)^i F_A K_A A_{t+1-i},$$

where $\epsilon_{A,t+1} \doteq A_{t+1} - \hat{A}_{t+1}$. The random variables $\{\epsilon_{A,t}\}$ are i.i.d., each having a normal distribution with mean zero and variance $G_A \Omega^Y G'_A$.

(The proof is similar to that of Proposition 1.) The meaning of Proposition 8 is that even without a full knowledge of all of the system parameters, the supplier can still attempt to produce the *same* MMSE forecasts of future retailer's orders. Nevertheless, two differences arise: (i) the supplier may need to keep track of a large portion of the order history $\{A_t, A_{t-1}, \ldots\}$; and (ii) the supplier would need to *estimate* the parameters of the above AR(∞) model: $\mathsf{Var}(\epsilon_{A,t})$, and $\kappa_i = G_A(F_A - F_A K_A G_A)^i F_A K_A$ for all $i \geqslant 1$.

While one may feel that the assumption of complete knowledge of system parameters is unrealistic, we argue that this may not, or should not, necessarily be true. We assert that if the supply-chain members are intelligent enough to use adaptive inventory replenishment policies, if they express their interest vis-a-vis the value of information-sharing practices, and if they operate in a business environment that advocates cooperation and coordination—they should be able and willing to discuss and jointly build an underlying demand forecast model for their products. This, however, should not be mistakenly considered as dynamic information sharing. For the latter to exist, the companies often need to make significant investments in information technology and in software packages that enable collaboration.

4.1.2. Illustration: Demand Processes with Early Signals. We consider here an example suggested (without treatment) by Raghunathan (2001): Suppose that the demand process evolves according to the dynamics

$$D_t - \mu = \rho(D_{t-1} - \mu) + \kappa U_{t-1} + \epsilon_t, \quad \epsilon_t \sim \mathrm{N}(0, \sigma), \tag{16}$$

where U_t is a retailer's observation during period t (e.g., promotion, price reduction, etc.), that directly affects the demand during the next period. Each observation U_t is seen by the retailer only, and has a normal distribution with mean $\mu_u = 0$ and standard deviation σ_u (dealing with $\mu_u \neq 0$ does not complicate the treatment). Also, suppose for simplicity that the process $\{U_t\}$ is independent of the random noise process $\{\epsilon_t\}$. A linear state space model that represents this case is given as follows: let $X'_t \doteq (D_t - \mu, U_t)$. Then

$$X_t = \begin{pmatrix} \rho & \kappa \\ 0 & 0 \end{pmatrix} X_{t-1} + \begin{pmatrix} \epsilon_t \\ U_t \end{pmatrix}, \quad \Psi_t = X_t,$$

$$D_t = \mu + \begin{pmatrix} 1 & 0 \end{pmatrix} X_t.$$

In this case, $K = H = I_2$, $\Omega^X = \Sigma_V$, $R = (1, 0)$, and

$$\Theta = R \sum_{l=0}^{L+1} F^l = \frac{1}{1-\rho} \left(1 - \rho^{L+2}, \kappa(1 - \rho^{L+1})\right).$$

Even under the relatively simple type of model presented above, the process $\{A_t\}$ does *not* follow an ARMA(1,1) evolution as under the case in which the demand evolves according to the elementary autoregressive process of Example 1. In fact, under (16), the orders $\{A_t\}$ maintain the following recursive scheme:

$$A_t - \mu = \rho(A_{t-1} - \mu) + \Theta \xi_{t-1} + \Theta_1 \xi_{t-2}, \tag{17}$$

where $\Theta_1 = \rho/(1-\rho)(1 - \rho^{L+1}, \kappa(1 - \rho^L))$, and the *vector* process $\{\xi_t\}$ consists of i.i.d. components, each having a binormal distribution. We continue with this illustration in §4.2.1.

4.2. Enriched Information Structures

In §4.1 we analyzed the evolution of the orders placed by the retailer, and we suggested an inventory replenishment policy for the supplier, when the only information available to the latter is the history of orders placed by the retailer (this is in addition to the knowledge of the static system parameters). Nonetheless, in many realistic settings the supplier may have other streams of information through which he can learn about future retailer's orders. For example, it might be the case that the supplier itself plans a promotion for his product, which will in turn cause future retailer orders to be larger (statistically). In this section, we extend our model to accommodate the supplier's desire to use all information in his possession when forecasting demand and placing replenishment orders. Our approach is as follows: We keep the demand evolution structure (1) and (3) the same. However, we replace the single observation Equation (2) by two linear observation equations. The retailer observes $\Psi_{r,t} = H_r X_t$, and for him we offer the same forecasting process and policy class described in §3. The supplier's observation during period t is given by the linear filtration $H_s X_t$, where H_s is a known $q \times n$ matrix. The supplier's policy remains of the same type discussed in §4.1, but we modify his forecasting process to

$$\hat{A}_t^{(\tau)} \doteq \mathsf{E}\left[\sum_{l=0}^{\tau-1} A_{t+1+l} \,\middle|\, \{H_s X_{t-1}, H_s X_{t-2}, \ldots\}, \{A_t, A_{t-1} \ldots\} \right].$$

Luckily, it turns out that the supplier's forecasting problem needs to deal with producing estimates of future retailer's orders based on a partially observed state vector that evolves according to the linear model of the type discussed above. The following proposition shows a sufficient representation of such a linear model.

Proposition 9. *Let $\hat{X}_t, K$, and Θ be the values associated with the retailer's steady state forecasting process and define the extended state vector*: $Z_t \doteq (\hat{X}_t' \quad X_t')'$.*This state vector follows the evolution*

$$\underbrace{\begin{pmatrix}\hat{X}_t\\ X_t\end{pmatrix}}_{Z_t} = \underbrace{\begin{pmatrix}F-FKH_r & FKH_r\\ \mathbf{0}_{n\times n} & F\end{pmatrix}}_{\tilde{F}} \cdot \underbrace{\begin{pmatrix}\hat{X}_{t-1}\\ X_{t-1}\end{pmatrix}}_{Z_{t-1}} + \underbrace{\begin{pmatrix}\mathbf{0}_{n\times 1}\\ V_t\end{pmatrix}}_{\tilde{V}_t}.$$

The complete supplier's observation structure is given by

$$\Psi_{s,t} = \begin{pmatrix}A_{t+1}\\ H_s X_t\end{pmatrix} = \underbrace{\begin{pmatrix}RH_rF^{L+1}-\Theta H_r & \Theta H_r\\ \mathbf{0}_{q\times n} & H_s\end{pmatrix}}_{\tilde{H}} \cdot Z_t.$$

Finally, $A_{t+1} = \tilde{R}\Psi_{s,t} = \tilde{R}\tilde{H}Z_t$, *where* $\tilde{R} = (1, 0, \dots, 0)$.

Proof. The state equation $Z_t = \tilde{F}Z_{t-1} + \tilde{V}_t$ is equivalent to the pair of equations (1) and (4). The observation equation $\Psi_{s,t} = \tilde{H}Z_t$ follows from Theorem 1. The last equation is straightforward. □

The representation in Proposition 9 is sufficient in the sense that a reduction in its dimension may be possible. But more importantly, the proposition enables us to suggest again the Kalman filter algorithm for estimating future order quantities, based on their history and on the supplier's observations of the underlying state vector process $\{X_t\}$. The estimates $\hat{Z}_t$ and their associated error covariance matrices Ω_t^Z can be generated through the Kalman filter Algorithm (4)–(5), replacing F, H, Ψ, and Σ_V, with $\tilde{F}, \tilde{H}, \Psi_s$, and $\Sigma_{\tilde{V}}$, respectively. Hence,

$$\hat{A}_t^{(\tau)} = \tau\mu + \tilde{G}\left(\sum_{l=0}^{\tau-1}\tilde{F}^l\right)\hat{Z}_t \tag{18}$$

and, analogously to Proposition 4, we obtain

Corollary 1. *The mean square error of the supplier's estimates $\hat{A}_t^{(\tau)}$ satisfies*

$$\Omega_t^A = \sum_{l=1}^{\tau-1}\left[\tilde{G}\left(\sum_{j=0}^{\tau-1-l}\tilde{F}^j\right)\Sigma_{\tilde{V}}\left(\sum_{j=0}^{\tau-1-l}\tilde{F}^j\right)'\tilde{G}'\right] + \tilde{G}\left(\sum_{j=0}^{\tau-1}\tilde{F}^j\right)\Omega_t^Z\left(\sum_{j=0}^{\tau-1}\tilde{F}^j\right)'\tilde{G}'. \tag{19}$$

Again, since our interest is in steady state behavior, we shall suppress the subscript t in the value of Ω^A.

4.2.1. Illustration: Demand Process with Early Signals (Continued). Consider the following extension to the demand model (16) of §4.1.2:

$$D_t - \mu = \rho(D_{t-1}-\mu) + \kappa_r U_{r,t-1} + \kappa_s U_{s,t-1} + \epsilon_t,$$
$$\epsilon_t \sim \mathrm{N}(0,\sigma), \tag{20}$$

where $U_{r,t-1}$ and $U_{s,t-1}$ are "market signals" observed by the retailer and the supplier, respectively. We assume that the bivariate process $\{(U_{r,t}, U_{s,t})\colon t \geqslant 1\}$ consists of i.i.d. components, each having a binormal distribution with mean $\mu_u = \mathbf{0}_{2\times 1}$ and covariance matrix Σ_u. The motivation for studying such types of demand processes was provided recently by Lee et al. (2000), Raghunathan (2001), and Aviv (2002). In fact, Aviv (2002) used the framework developed in this work to devise inventory control policies for various types of two-stage supply-chain structures that face demand processes with early market signals. The demand process considered by Aviv (2002) is more complex than (20), allowing both the retailer and the supplier to observe signals (i.e., U-values) about future demands, starting from several periods in advance.

Under the demand structure (20), the model representation is: let $X_t' \doteq (D_t-\mu, U_{r,t}, U_{s,t})$. Thus,

$$X_t = \begin{pmatrix}\rho & \kappa_r & \kappa_s\\ 0 & 0 & 0\\ 0 & 0 & 0\end{pmatrix} X_{t-1} + \begin{pmatrix}\epsilon_t\\ U_{r,t}\\ U_{s,t}\end{pmatrix},$$

$$\Sigma_V = \begin{pmatrix}\sigma^2 & \mathbf{0}_{1\times 2}\\ \mathbf{0}_{2\times 1} & \Sigma_u\end{pmatrix}, \quad H_r = \begin{pmatrix}I_2 & \mathbf{0}_{2\times 1}\end{pmatrix},$$

$$H_s = \begin{pmatrix}0 & 0 & 1\end{pmatrix}, \quad \text{and} \quad R = \begin{pmatrix}1 & 0\end{pmatrix}.$$

Hence, our framework is readily available to construct forecasting and inventory replenishment policies for the retailer and the supplier in the chain. In addition to specifying operating policies for the supply chain, one may find our models useful in studying important issues in supply-chain management, such as the assessment of the potential benefits of information sharing between the retailer and the supplier. It has been shown that in the elementary AR(1) and ARIMA(0,1,1) cases, there is no value for the retailer sharing information about the actual demand realizations with the supplier. This is because the supplier can observe the values of $\{D_{t-1}, D_{t-2}, \dots\}$ from the history of retailer's orders $\{A_t, A_{t-1}, \dots\}$ (see Raghunathan 2001 and Graves 1999, respectively). When the demand depends on "signals," we would expect that information sharing will enhance the ability of *both* the retailer and the supplier to more accurately forecast demand at their levels. As can be expected, when the information structure in the supply chain is complex, the derivation of exact analytical formulae for forecasting and inventory target levels is tedious. Therefore, in addition to the elegance of matrix representation, our models offer an easily implementable method for analysis and operation of information-rich inventory systems.

5. EXTENSIONS, FURTHER EXAMPLES, AND POTENTIAL APPLICATIONS

The main purpose of this section is to demonstrate the potential use of our model in a broad range of studies in the area of supply-chain inventory management. In §5.1 we propose a simple methodology for studying the value of information sharing in supply chains. In §5.2 we propose two important extensions to the two-stage supply-chain model we discussed above: The case of a serial supply chain, and the case of multiple products and retail locations. Extensions to *distribution networks*, in which each inventory-stocking location has a unique supplier, follow easily. In §5.3 we provide an illustration of a stylized model that can be used to describe and analyze the forecasting advantage that suppliers often have in industry, as a result of the fact that they observe aggregate demands. In §5.4 we conclude with some brief remarks about state space reduction.

5.1. Benefits of Information Sharing in the Supply Chain

The topic of information sharing in supply chains has received a great deal of attention in the academic literature as well as in the business press. This is particularly due to the fact that today, supply chains are usually very rich in the amount and types of information they possess. In general, it is often hard for managers to estimate the specific marginal value that each piece of data can bring to their firms when taken into account in their inventory replenishment processes. In addition, it may be even more challenging to identify the types of data that companies may want to share with their trading partners to bring the most significant added value to their system. The purpose of this section is to demonstrate that in addition to the development of forecasting and control policies for supply-chain inventory systems, our framework provides the means for assessing the benefits of information usage and information-sharing practices. First, it is quite conceivable that inventory managers may need to contemplate between the uses of different sets of data streams when determining replenishment orders. This need arises, for example, when the cost of collecting a specific type of data is high, or when a particular type of information exists in a different database than that used by the inventory planner, and so an investment in hardware and software is required to make this data available. For example, consider the demand model (16), and suppose that the retailer considers whether or not to use the data $\{U_t\}$. In such a case, one can examine the performance of the retailer's inventory policy under the case $H_r = I_{2\times 2}$ (i.e., the retailer observes and uses $\{U_t\}$), and under the case $H_r = (1, 0)$, $R = 1$, which represents the situation in which the retailer does not use the values of $\{U_t\}$. A potential application of such analysis is in supply channels for spare parts. One way by which inventory managers can reduce their uncertainty about future demand for spare parts is to collect usage data from a representative subset of their customers. While this can certainly improve forecasting performance, the main question is whether the *magnitude* of such benefits will surpass the investment in data collection and analysis.

With regards to the assessment of the value of information sharing, our methodology follows quite intuitively: Since both H_r and H_s are given as linear functions of the state vector X_t, information sharing can be described by the exact same model, but with an appropriate modification of the matrices H_r, H_s, and R. Consider, for example, the demand model (20), and suppose that the retailer and the supplier wish to gauge the benefits of "complete" information sharing. In that case, the original system (i.e., no information sharing) can be compared with $H_r = H_s = I_{3\times 3}$, and $R = (1, 0, 0)$. Clearly, various levels of information sharing can be examined and compared with each other to identify the most promising (cost beneficial) data-sharing arrangement.

5.2. Distribution Networks

Consider first a supply chain that consists of J facilities in series. The first facility (indexed $j = 1$) is a retail outlet that faces consumer demand for a single product. The rest of the facilities represent the suppliers of the product, so that facility j is the immediate supplier of facility $j-1$, for all $j = 2, \ldots, J$. We assume that the demand evolves as a linear function of an underlying state space vector X_t that maintains the form $X_t = FX_{t-1} + V_t$. Further, assume that facility j partially observes the state vector X_t through the filtration $\Psi_{j,t} = H_j X_t$; this is, of course, in addition to observing the orders placed by its immediate customer (for $j = 2, \ldots, J$, the immediate customer is facility $j-1$). The consumer demand is equal to $D_t = R\Psi_{1,t} + \mu$. It follows from Proposition 9 that

COROLLARY 2. *Suppose that the matrix F is stable. Then, each facility $j = 1, \ldots, J$ faces a demand process that evolves according to a linear state space model of type* (1)–(3).

(A proof of this corollary can be done by an induction on j.) This result means that forecasting and replenishment policies in each location of the chain are as simple as those devised for a single-facility system (§3).

Next, consider a distribution channel consisting of J retail "end points," with the understanding that an end point may represent a specific stock-keeping unit at a particular retail location. For convenience, we shall refer to each end point j, as retailer j. All retailers order their product (or components) for the same single supplier. Suppose that forecasting and inventory control at each location is done using local information, and according to the policy type described above. A key assumption that underlies our treatment of this case is that all retailers are "trustable," or in other words, they will always use their MMSE estimates of lead-time demand to set their inventory target levels. This is important, because when the retailers expect supply shortages, they may decide to inflate their orders to ensure that they will receive the number of units they actually want in the event that the supplier runs into shortage and needs to ration the goods among them (see Lee et al. 1997a).

Given our assumption about the retailers' trustable behavior, we propose for them the exact same forecasting and inventory control policy structure presented in §3. To devise a control policy for the supplier, we modify the results of Proposition 9: Let $\hat{X}_{j,t}$ be a retailer's j estimate of the state space in period t, and let K_j, H_j, R_j, and Θ_j be the matrices associated with this retailer's forecasting process. In addition, we denote retailer j's order in period t by $A_{j,t}$. We define the state vector $Z_t = (\hat{X}_{1,t}, \ldots, \hat{X}_{J,t}, X_t)$, and note that

$$Z_t = \begin{pmatrix} F-FK_1H_1 & \mathbf{0}_{n\times n} & \cdots & \mathbf{0}_{n\times n} & FK_1H_1 \\ \mathbf{0}_{n\times n} & F-FK_2H_2 & \mathbf{0}_{n\times n} & \cdots & FK_2H_2 \\ \vdots & & \ddots & & \vdots \\ \mathbf{0}_{n\times n} & \cdots & \mathbf{0}_{n\times n} & F-FK_JH_J & FK_JH_J \\ \mathbf{0}_{n\times n} & \cdots & \cdots & \mathbf{0}_{n\times n} & F \end{pmatrix} \cdot Z_{t-1} + \begin{pmatrix} \mathbf{0}_{n\times 1} \\ \vdots \\ \mathbf{0}_{n\times 1} \\ V_t \end{pmatrix}.$$

The complete supplier's observation structure is now given by

$$\Psi_{s,t} = \begin{pmatrix} A_{1,t} \\ \vdots \\ A_{J,t} \\ H_sX_t \end{pmatrix} = \begin{pmatrix} R_1H_1F^{L+1}-\Theta_1H_1 & \mathbf{0}_{1\times n} & \cdots & \Theta_1H_1 \\ \vdots & \ddots & & \vdots \\ \mathbf{0}_{1\times n} & \cdots & R_JH_JF^{L+1}-\Theta_JH_J & \Theta_JH_J \\ \mathbf{0}_{q\times n} & \cdots & \mathbf{0}_{q\times n} & H_s \end{pmatrix} Z_t.$$

Finally, the aggregate demand faced by the supplier in period $t+1$ is given by $A_{s,t+1} = \tilde{R}\Psi_{s,t}$, where $\tilde{R} = (1, \ldots, 1, 0, \ldots, 0)$. A forecasting and inventory policy for the supplier follows.

5.3. Illustration: Benefits of Observing Aggregate Demand Volumes

In this section, we briefly describe a possible application of our model in gauging the value of information the supplier has in a distribution network. Consider a supplier that serves several retailers and suppose that the aggregate demand in the market evolves according to a known dynamic pattern. Then it can be argued that since the supplier observes all retailers' orders, he might be in a better position than each individual retailer to anticipate future demand. We propose below a simple stylized model that can be useful in analyzing such situations. Again, assume a single supplier and J retailers. Let $D_{j,t}$ be the demand for retailer j in period t, and assume that the aggregate demand in the market (denoted by $D_t = \sum_{j=0}^{J} D_{j,t}$) follows an AR(1) process: $D_t - \mu = \rho(D_{t-1} - \mu) + \epsilon_t$. Assuming that the aggregate demand during every period t is split among the J retailer in a way that allows for correlation between $D_{1,t}, \ldots, D_{J,t}$, one can show that for a vector white-noise process $\{u_t\}$, each demand $D_{j,t}$ can be represented by

$$D_{j,t} = H_j \begin{pmatrix} u_{1,t} \\ \cdots \\ u_{k,t} \\ D_t \end{pmatrix} + \mu_j$$

for appropriately specified vectors $H_1, \ldots, H_J$, and an integer k. Let the vector $X_t = (u_{1,t}, \ldots, u_{J,t}, D_t)'$ be the state of the system, and note that

$$X_t = \begin{pmatrix} \mathbf{0}_{k\times(k+1)} \\ \mathbf{0}_{1\times k}\ \rho \end{pmatrix} X_{t-1} + \begin{pmatrix} u_{1,t} \\ \cdots \\ u_{k,t} \\ \epsilon_t \end{pmatrix}.$$

After developing forecasting and inventory replenishment policies for the retailers and the supplier (see §5.2), one can use this model to explore, for example, the value of the supplier sharing information about aggregate demand volumes with the retailers, and the value of the retailers sharing the values of their demand realizations with the supplier.

5.4. Remarks on State Space Reduction

In some of the model representations in this paper and the illustrations above (e.g., §5.2), we have not made an attempt to identify the most reduced forms of the Kalman filters associated with the supply-chain members' forecasting processes. For example, one can easily see from Proposition 9 that the dimension of the system vector underlying the evolution of demand at location j is exponentially increasing with j. Thus, one may wonder under what circumstances it may be possible to collapse the state vector to one with a substantially smaller dimension. In the interest of not getting the discussion carried beyond the scope of this work we shall not treat this issue in detail. Rather, we refer the reader to Anderson and Moore (1979), Harvey (1990), and references therein for more specific information on this topic. Clearly, as long as the dimensionality of the state space is not considerably large, the computational effort involved with the use of the Kalman filter tool should not be significant, given the speed of PCs today. Furthermore, some software packages (e.g., MathWorks) offer special functions for model reduction that can be utilized by practitioners. An illustration for a simple state space reduction in an inventory/forecasting model can also be found in Aviv (2002).

6. INVENTORY POSITIONING AND COST ANALYSIS

In this section we propose a method for assessing the costs associated with the inventory policies we developed for the two-stage supply chain, under the case of linear inventory and shortage penalty costs. The analysis here is different than that in the existing literature primarily due to the intricate relationship between the retailer's and the supplier's forecasting processes. We suggest a simple simulation-based procedure to assess the average long-run inventory costs for any given pair of safety-stock levels γ^1 and γ^2. In §6.1 we describe a method for search for the best values of γ^1 and γ^2 that minimize the average cost of inventory in the supply chain. In §6.2, we discuss a supply chain that is *decoupled* with inventory at the supplier's location—in other words, a setting in which the supplier holds a sufficiently large safety stock of goods at this location, so that the retailer almost never faces delays due to supplier shortages. We shall use the location index $i = 1, 2$ to denote the retailer and the supplier, respectively.

Let y_t^i be the net inventory of location i at the end of period t. Then, the following costs are charged in period t: $c^i(y_t^i) = h^i \cdot \max(0, y_t^i) + p^i \cdot \max(0, -y_t^i)$, where q^i is the cost function for location i. The nonnegative parameters h^i and p^i represent the per-unit inventory holding cost and shortage penalty cost, respectively, at location i. We focus on the long-run average cost functions

$$g^i \doteq \lim_{N\to\infty} \sum_{t=1}^{N} c^i(y_t^i), \quad i = 1, 2.$$

The long-run average total supply-chain cost is given by $g \doteq g^1 + g^2$. The next proposition provides an expression for the long-run average costs as functions of the safety-stock parameters γ^1 and γ^2.

PROPOSITION 10. *Let* $E_t^1 = \sum_{l=0}^{L} D_{t+l} - \hat{D}_t^{(L)}$, *and* $E_{t-\tau}^2 = \sum_{l=1}^{\tau} A_{t-\tau+l} - \hat{A}_{t-\tau}^{(\tau)}$. *Then*

$$g(\gamma^1, \gamma^2) = \lim_{N\to\infty} \frac{1}{N} \sum_{t=\tau+1}^{N+\tau} \Big\{ c^1(\gamma^1 - E_t^1 - \max(E_{t-\tau}^2 - \gamma^2, 0)) + c^2(\gamma^2 - E_{t-\tau}^2) \Big\}. \quad (21)$$

PROOF. It is easy to see that the supplier's net inventory at the end of period t is given by his inventory *position* τ periods in advance, minus the total retailer's orders from period $t-\tau+1$ to period t, inclusive. Thus, $y_t^2 = \hat{A}_{\tau-t}^{(\tau)} + \gamma^2 - \sum_{l=1}^{\tau} A_{t-\tau+l} = \gamma^2 - E_{t-\tau}^2$. For the retailer's y_{t+L}^1, we need to consider his inventory position in period t, but *excluding* those units that will not arrive by the end of period $t+L$. This is given by $\beta_t^1 + \min(y_t^2, 0)$. Therefore, the retailer's net inventory at the end of period $t+L$ is given by $\beta_t^1 + \min(y_t^2, 0) - \sum_{l=0}^{L} D_{t+l} = \gamma^1 - E_t^1 - \max(E_{t-\tau}^2 - \gamma^2, 0)$. To conclude, recall that we are interested in long-run average cost estimates; thus, $g(\gamma^1, \gamma^2) = \lim_{N\to\infty} \frac{1}{N} \sum_{t=\tau+1+L}^{N+\tau+L} c^1(y_t^1) + \lim_{N\to\infty} \frac{1}{N} \sum_{t=\tau+1}^{N+\tau} c^2(y_t^2)$. □

Keeping in mind the forecasting processes discussed in §§3–4, it is easy to see that the terms E_t^1 and $E_{t-\tau}^2$ stand for the forecast errors of the retailer's and the supplier's estimates of their lead-time demand. We are particularly interested in the *lagged* forecast errors; i.e., those made by the supplier during a particular period, and those made by the retailer τ periods later. Proposition 10 tells us that the cost performance of the supply chain is uniquely determined by the safety-stock parameters γ^1 and γ^2 and by the characteristics of the joint distribution of the lagged forecasts E_t^1 and $E_{t-\tau}^2$, which we identify in the next theorem.

THEOREM 2. *Consider the error terms* E_t^1 *and* $E_{t-\tau}^2$ *of Proposition* 10. *Under the assumption of Kalman filters in steady state, their joint distribution is bivariate normal*:

$$\begin{pmatrix} E_t^1 \\ E_{t-\tau}^2 \end{pmatrix} \sim N\left(\mathbf{0}_{2\times 1}, \quad \begin{pmatrix} \Omega^D & C \\ C & \Omega^A \end{pmatrix}\right), \quad (22)$$

where Ω^D *and* Ω^A *are the mean square errors associated with the retailer's and the supplier's estimates of their lead-time demands,* (10) *and* (19). *In addition,*

$$\begin{aligned} C &\doteq \mathrm{Cov}(E_t^1, E_{t-\tau}^2) \\ &= \tilde{G}\left(\sum_{j=0}^{\tau-1} \tilde{F}^j\right)\left\{\sum_{l=\tau}^{\infty} (\tilde{F} - \tilde{F}\tilde{K}\tilde{H})^{l-\tau} \mathrm{E}[\tilde{V}_t V_t']([F - FKH]^l)'\right\} \\ &\quad \times \left(\sum_{j=0}^{L} F^j\right)' G' \\ &\quad + \tilde{G}\left\{\sum_{l=1}^{\tau-1}\left(\sum_{j=0}^{l-1} \tilde{F}^j\right) \mathrm{E}[\tilde{V}_t V_t']([F - FKH]^l)'\right\}\left(\sum_{j=0}^{L} F^j\right)' G'. \end{aligned}$$

PROOF. By (7) and (8), $E_t^1 = G(\sum_{l=0}^{L} F^l)(X_t - \hat{X}_t) + G\sum_{l=1}^{L}(\sum_{j=0}^{L-l} F^j) V_{t+l}$. Applying this equation to the supplier's forecasting process (Proposition 9), we obtain: $E_{t-\tau}^2 = \tilde{G}(\sum_{l=0}^{\tau-1} \tilde{F}^l)(Z_{t-\tau} - \hat{Z}_{t-\tau}) + \tilde{G}\sum_{l=1}^{\tau-1} (\sum_{j=0}^{\tau-1-l} \tilde{F}^j) \tilde{V}_{t-\tau+l}$. Next, a recursive expansion of the term $X_t - \hat{X}_t$ yields: $X_t - \hat{X}_t = (F - FKH)(X_{t-1} - \hat{X}_{t-1}) + V_t = \cdots = \sum_{j=0}^{\infty}(F - FKH)^j V_{t-j}$, and similarly, $Z_t - \hat{Z}_t = \sum_{j=0}^{\infty}(\tilde{F} - \tilde{F}\tilde{K}\tilde{H})^j \tilde{V}_{t-j}$, where the matrix $\tilde{K}$ is defined as K, but now with respect to the estimation process $\hat{Z}_t$. Therefore, $E_t^1 = G(\sum_{j=0}^{L} F^j)\sum_{l=0}^{\infty}(F - FKH)^l V_{t-l} + G\sum_{l=1}^{L}(\sum_{j=0}^{L-l} F^j) V_{t+l}$, and $E_{t-\tau}^2 = \tilde{G}(\sum_{j=0}^{\tau-1} \tilde{F}^j)\sum_{l=\tau}^{\infty} (\tilde{F} - \tilde{F}\tilde{K}\tilde{H})^{l-\tau}\tilde{V}_{t-l} + \tilde{G}\sum_{l=1}^{\tau-1}(\sum_{j=0}^{l-1}\tilde{F}^j)\tilde{V}_{t-l}$. The proposition now follows. □

It follows from Proposition 10 and Theorem 2 that for a generic pair of random variables R^1 and R^2 that maintain the time-invariant distribution described in (22), the cost functions g^1 and g^2 are given by $g^1(\gamma^1, \gamma^2) = \mathrm{E}[c^1(\gamma^1 - R^1 - \max(R^2 - \gamma^2, 0))]$, and $g^2(\gamma^2) = \mathrm{E}[c^2(\gamma^2 - R^2)]$. This result gives rise to a straightforward simulation-based method for estimating the supply chain's long-run average inventory costs for any choice of γ^1 and γ^2; see, e.g., Aviv (2002) for a description of such a method.

In the following sections we treat the supplier's cost parameter p^2 *not* as a net cost charged to the supply chain, but rather as an internal payment mechanism that is set to provide the incentives for the supplier to hold a satisfactory amount of inventory.

6.1. Coordinated Two-Level Inventory Systems: Theoretical Results and Approximations

In this section we treat the problem of finding a pair of safety-stock parameters γ^{*1} and γ^{*2} that minimize the long-run average total supply-chain cost. We refer to a supply chain in which these optimal safety-stocks are used as a *coordinated two-level inventory system*. We assume that $h^2 < h^1$ and set $p^2 = 0$. We start with the following observation.

Theorem 3. *Suppose that $\Omega^A, \Omega^D > 0$ and consider the cost function $g(\gamma^1, \gamma^2) = g^1(\gamma^1, \gamma^2) + g^2(\gamma^2) = \mathsf{E}_{(R^1,R^2)}[c^1(\gamma^1 - R^1 - \max(R^2 - \gamma^2, 0)) + c^2(\gamma^2 - R^2)]$. Also, let $f(\gamma^2) \doteq \min_{\gamma^1} g(\gamma^1, \gamma^2)$. Then*

(i) *The functional solution $\gamma^{*1}(\gamma^2)$ to $f(\gamma^2)$ is unique, and is given by*

$$\gamma^{*1}(\gamma^2) = F_{\gamma^2}^{-1}\left(\frac{p^1}{h^1 + p^1}\right), \tag{23}$$

where F_{γ^2} is the continuous, cumulative distribution function of $R^1 + \max(R^2 - \gamma^2, 0)$, with support on the entire real line.

(ii) *For all $\gamma^2 \in \mathbf{R}$: $-1 < d\gamma^{*1}(\gamma^2)/d\gamma^2 < 0$.*

(iii) *The first-order condition $df(\gamma^2)/d\gamma^2 = 0$ is satisfied for any point γ^2, that solves the equation*

$$\mathsf{Pr}\{R^1 \leqslant \gamma^{*1}(\gamma^{*2}) | R^2 \leqslant \gamma^{*2}\} = \frac{h^2 + p^1}{h^1 + p^1}. \tag{24}$$

Proof. The proof of part (i) is straightforward, since $\gamma^{*1}(\gamma^2)$ is a solution to a newsvendor problem, with $R^1 + \max(R^2 - \gamma^2, 0)$ being the unknown demand volume. Because $\Omega^A, \Omega^D > 0$, the distribution F_{γ^2} is continuous with a full support on $\mathbf{R}$, and hence it follows that (23) has a unique solution. To prove part (ii), note that $\mathsf{Pr}\{R^1 + \max(R^2 - \gamma^2, 0) \leqslant \gamma^{*1}(\gamma^2)\} = p^1/(h^1 + p^1)$. Hence, for every positive perturbation of γ^2 (denoted below by $\delta > 0$), we obtain: $\mathsf{Pr}\{R^1 + \max(R^2 - \gamma^2, 0) \leqslant \gamma^{*1}(\gamma^2 + \delta)\} < \mathsf{Pr}\{R^1 + \max(R^2 - \gamma^2 - \delta, 0) \leqslant \gamma^{*1}(\gamma^2 + \delta)\} = p^1/(h^1 + p^1) < \mathsf{Pr}\{R^1 + \max(R^2 - \gamma^2, 0) \leqslant \gamma^{*1}(\gamma^2 + \delta) + \delta\}$. Therefore, $\gamma^{*1}(\gamma^2 + \delta) < \gamma^{*1}(\gamma^2)$ and $\gamma^{*1}(\gamma^2 + \delta) + \delta > \gamma^{*1}(\gamma^2)$ for all $\delta > 0$, from which part (ii) follows. For part (iii), consider the derivative $df(\gamma^2)/d\gamma^2$:

$$\frac{df(\gamma^2)}{d\gamma^2} = h^2 \cdot \mathsf{Pr}\{R^2 \leqslant \gamma^2\} + \left.\frac{\partial g^1(\gamma^1, \gamma^2)}{\partial \gamma^2}\right|_{\gamma^1 = \gamma^{*1}(\gamma^2)} + \underbrace{\left.\frac{\partial g^1(\gamma^1, \gamma^2)}{\partial \gamma^1}\right|_{\gamma^1 = \gamma^{*1}(\gamma^2)}}_{=0} \cdot \frac{d\gamma^{*1}(\gamma^2)}{d\gamma^2}.$$

The derivative $\partial g^1(\gamma^1, \gamma^2)/\partial \gamma^2$ is given by: $-p^1 \cdot \mathsf{Pr}\{R^1 + R^2 > \gamma^1 + \gamma^2, R^2 \geqslant \gamma^2\} + h^1 \cdot \mathsf{Pr}\{R^1 + R^2 \leqslant \gamma^1 + \gamma^2, R^2 \geqslant \gamma^2\} = -p^1 + p^1 \cdot \mathsf{Pr}\{R^2 \leqslant \gamma^2\} + (h^1 + p^1) \cdot \mathsf{Pr}\{R^1 + R^2 \leqslant \gamma^1 + \gamma^2, R^2 \geqslant \gamma^2\}$. Therefore, we obtain: $df(\gamma^2)/d\gamma^2 = -p^1 + (h^2 + p^1) \cdot \mathsf{Pr}\{R^2 \leqslant \gamma^2\} + (h^1 + p^1) \cdot \mathsf{Pr}\{R^1 + R^2 \leqslant \gamma^{*1}(\gamma^2) + \gamma^2, R^2 \geqslant \gamma^2\}$. But observe that in view of part (i) of this proposition: $\mathsf{Pr}\{R^1 + R^2 \leqslant \gamma^{*1}(\gamma^2) + \gamma^2, R^2 \geqslant \gamma^2\} = p^1/(h^1 + p^1) - \mathsf{Pr}\{R^1 \leqslant \gamma^{*1}(\gamma^2), R^2 \leqslant \gamma^2\}$. Hence, $df(\gamma^2)/d\gamma^2 = (h^2 + p^1) \cdot \mathsf{Pr}\{R^2 \leqslant \gamma^2\} - (h^1 + p^1) \cdot \mathsf{Pr}\{R^1 \leqslant \gamma^{*1}(\gamma^2), R^2 \leqslant \gamma^2\}$. It is easy to see now that the derivative is equal to zero at a point γ^{*2} that satisfies (24). □

The result of Theorem 3 is quite interesting. From (24), we see that when γ^2 is set to its best value γ^{*2}, the retailer sets his own policy parameter γ^1 by adjusting his cost parameters accordingly ($\tilde{h} = h^1 - h^2$ and $\tilde{p} = p^1 + h^2$), and plans only for the cases in which there are no supply delays (i.e., $E_{t-\tau}^2 \leqslant \gamma^{*2}$). This is done by setting γ^1 so that the probability of a stockout occurring at the retailer's location after a lead time (i.e., $E_t^1 \leqslant \gamma^{*1}$) is given by the cost ratio $\tilde{p}/(\tilde{p} + \tilde{h}) = (h^2 + p^1)/(h^1 + p^1)$. This result is reminiscent of that of Clark and Scarf (1960): In the case of $\mathsf{Cov}(R^1, R^2) = 0$, (24) reduces to $\mathsf{Pr}\{R^1 \leqslant \gamma^{*1}\} = \tilde{p}/(\tilde{p} + \tilde{h})$, or $\gamma^{*1} = \sqrt{\Omega^D} \cdot \Phi^{-1}(\tilde{p}/(\tilde{p} + \tilde{h}))$. Part (ii) of the theorem shows that regardless of the correlation between the lagged forecast error terms, an increase in the supplier's inventory will enable the retailer to decrease his safety stock, when the latter is concerned with local cost performance (i.e., the function $f(\gamma^2)$). Part (ii) also supports the intuition that to achieve a certain reduction in the retailer's safety-stock inventory, at least the same amount would need to be added to the supplier's safety stock.

We use the result in the theorem to propose a simple optimization method for the supply chain. Using the result of part (i) of the theorem, one can write a simple routine to calculate the optimal value of $\gamma^{*1}(\gamma^2)$ for any given value of γ^2. Then, a simple and efficient line search procedure can be written to search for the best value of γ^{*2}. In fact, when $\mathsf{Cov}(R^1, R^2) \geqslant 0$, it is easy to show that γ^{*2} is unique, and that the function $f(\gamma^2)$ is quasi-convex. It is possible that this result extends to $\mathsf{Cov}(R^1, R^2) < 0$ as well.

Next, we consider a special case in which the retailer fully observes the value of X_t during period t. This case is of particular interest because it is satisfied in a large part of the literature on information sharing in the context of supply-chain inventory management (e.g., Lee et al. 2000, Graves 1999). In this case, we show that the lagged forecast error terms are statistically independent of each other.

Proposition 11. *Consider the individual forecasting processes in steady state and suppose that $\hat{X}_t = FX_{t-1}$. Then, $\mathsf{Cov}(E_t^1, E_{t-\tau}^2) = 0$.*

Proof. Note that $X_t - \hat{X}_t = V_t$, and so the term for E_t^1 includes the random vectors $V_t, \ldots, V_{t+L}$ only. Next, observe that the expression for $E_{t-\tau}^2$ includes the nonoverlapping vectors $V_{t-1}, V_{t-2}, \ldots$. □

Of course, the assumption $\hat{X}_t = FX_{t-1}$ is satisfied when the retailer observes the state X_t during period t. We argue that since the correlation between the lagged forecast errors E_t^1 and $E_{t-\tau}^2$ is a determining factor in the cost performance of the supply chain, our framework provides both the control policy and cost estimation procedures that may add important new insights into practical settings. Indeed, Aviv (2002) has been able, through the use of this framework, to provide a better understanding of the relationship between a supplier's and a retailer's forecasting correlation pattern and the benefits they can gain from collaborative forecasting programs.

We conclude this section by proposing a simple upper bound for the long-run average inventory cost for the coordinated supply chain.

PROPOSITION 12. *The following is an upper bound on the cost* $g(\gamma^{*1}, \gamma^{*2})$:

$$g(\gamma^{*1}, \gamma^{*2}) \leqslant (h^1+p^1)\cdot\phi\Big[\Phi^{-1}\Big(\frac{p^1}{h^1+p^1}\Big)\Big] \cdot\sqrt{\Omega^D + 2\cdot\mathsf{Cov}(E_t^1, E_{t-\tau}^2)+\Omega^A}, \qquad (25)$$

where ϕ *and* Φ *are the density function and cumulative distribution function of the standard normal distribution, respectively.*

PROOF. We show that the right-hand-side term of (25) is equal to $\lim_{\gamma^2\to-\infty} g(\gamma^{*1}(\gamma^2), \gamma^2)$, and hence it can serve as an upper bound. Note first that $\lim_{\gamma^2\to-\infty} g^2(\gamma^2)=0$. In addition, when $\gamma^2\to-\infty$, the term $\max(R^2-\gamma^2, 0)$ can be replaced by $R^2-\gamma^2$. Hence, $\lim_{\gamma^2\to-\infty} g(\gamma^{*1}(\gamma^2), \gamma^2) = \min_\delta E[c^1(\delta - R^1 - R^2)]$. The latter is a simple newsvendor problem for which the solution is known to be $(h^1+p^1)\cdot \mathsf{Std}(R^1+R^2)\phi[\Phi^{-1}(p^1/(h^1+p^1))]$. $\square$

We have conducted an experiment to test how appropriate it is to use the bound (25) as an approximation for the average systemwide cost. Across 300 instances studied in Aviv (2002), we have achieved an almost perfect correlation (0.986) between the optimal costs and the cost approximations computed via (25). It has to be mentioned, however, that in Aviv's instances the correlation $\mathsf{Cov}(R^1, R^2)$ was negative, so it is not entirely clear whether the bound would also be tight for positive values of the covariance term. We conjecture that when the correlation between the lagged forecast errors is low, it may be appropriate for the supplier not to hold a large safety stock in his position. Of course, this suggestion deserves further study. It is important to note that the upper bound in (25) does *not* mimic the cost under a setting in which there is no supplier and the retailer faces an "extended" lead time of $L+\tau$ periods. This is because the supplier's forecast error (Ω^A) is used in the approximation. In the case in which the retailer is responsible for forecasting the demand during a lead time of $L+\tau$ periods, and the supplier's only role is to pass on the retailer's replenishment orders, we obtain the cost estimate

$$(h^1+p^1)\cdot\phi\Big[\Phi^{-1}\Big(\frac{p^1}{h^1+p^1}\Big)\Big] \cdot\sqrt{\mathsf{E}\Big[\Big(\hat{D}_t^{(L+\tau)} - \sum_{l=0}^{L+\tau} D_{t+l}\Big)^2\Big]}.$$

It might be of interest to check which of the upper bounds is smaller. If the latter term is smaller, then it may be appropriate to use the supplier as a cross-docking point and let the retailer take responsibility of the forecasting process.

6.2. Decoupled Two-Level Inventory Systems

We now turn our attention to a so-called *decoupled inventory system*, in which the supplier maintains a high level of product availability to the retailer. Consequently, the retailer will rarely face supply delays, and hence he does not have to be concerned about shortages when setting safety stocks at his level. Specifically, the retailer will set

$$\gamma_D^1 = \Phi^{-1}\left(\frac{p^1}{h^1+p^1}\right)\cdot\sqrt{\Omega^D},$$

and his long-run average costs will be (approximately)

$$(h^1+p^1)\cdot\phi\left[\Phi^{-1}\left(\frac{p^1}{h^1+p^1}\right)\right]\cdot\sqrt{\Omega^D}.$$

To achieve decoupling, we impose on the supplier a high penalty charge (p^2) per unit of shortage per period. We assume that these penalty charges are paid in full to the retailer, so that they can be ignored when calculating the average systemwide costs. In other words, since p^2 is charged to the supplier, the supplier will set

$$\gamma_D^2(p^2) = \Phi^{-1}\left(\frac{p^2}{h^2+p^2}\right)\cdot\sqrt{\Omega^A},$$

but we only consider the inventory *holding* costs at his location:

$$h^2\left[\frac{p^2}{p^2+h^2}\cdot\Phi^{-1}\left(\frac{p^2}{p^2+h^2}\right)+\phi\left(\Phi^{-1}\frac{p^2}{p^2+h^2}\right)\right]\cdot\sqrt{\Omega^A}.$$

We use the subscript "D" in the safety-stock parameters because they relate to a decoupled system. We believe that in some business settings, it is interesting and relevant to examine the inventory performance of a decoupled supply chain (see, e.g., Graves 1999). However, we argue that one should generally refrain from analyzing (or recommending) a decoupled inventory system, only because it simplifies the complexity of the mathematical analysis. In Aviv (2002) we observed that from the overall supply-chain perspective, a decoupled chain can perform very poorly compared to a coordinated one. In Proposition 13 below, we provide a simple formula that specifies a range of values of p^2 for which the associated decoupled system performs more poorly than a nondecoupled setting in which $\gamma^2 = -\infty$.

Proposition 13. *Consider the supply chain's long-run average cost* $g(\gamma_D^1, \gamma_D^2(p^2))$. *Let*

$$u_t \doteq \frac{h^1+p^1}{h^2}\cdot\phi\left[\Phi^{-1}\left(\frac{p^1}{h^1+p^1}\right)\right] \cdot\left[\sqrt{\frac{\Omega^D+2\cdot\mathrm{Cov}(E_t^1,E_{t-\tau}^2)+\Omega^A}{\Omega^A}}-\sqrt{\frac{\Omega^D}{\Omega^A}}\right],$$

and let u^* *be the solution to the equation* $u\Phi(u)+\phi(u)=u_t$. *Then, for all* $p^2 \geqslant h^2(\Phi(u^*)/(1-\Phi(u^*)))$, $g(\gamma_D^1, \gamma_D^2(p^2)) \geqslant g(\gamma^{*1}(\gamma^2=-\infty), \gamma^2=-\infty)$.

Proof. For any p^2,

$$g(\gamma_D^1, \gamma_D^2(p^2)) \geqslant (h^1+p^1)\cdot\phi\left[\Phi^{-1}\left(\frac{p^1}{h^1+p^1}\right)\right]\cdot\sqrt{\Omega^D} + h^2\left[\frac{p^2}{p^2+h^2}\cdot\Phi^{-1}\left(\frac{p^2}{p^2+h^2}\right) + \phi\left(\Phi^{-1}\frac{p^2}{p^2+h^2}\right)\right]\cdot\sqrt{\Omega^A}.$$

The value of $g(\gamma^{*1}(\gamma^2=-\infty), \gamma^2=-\infty)$ is given in the right-hand side of (25). We next substitute $u=\Phi^{-1}(p^2/(h^2+p^2))$, and the proposition follows by simple algebra. □

The upper bound on p^2 in Proposition 13 is crude, and we thus suggest the numerical simulation method developed in §6.1 to find the best operating policy (within our policy class) for the supply chain. Nevertheless, Proposition 13 can be used to avoid a bad choice of p^2. For example, consider a case in which $\Omega^A=\Omega^D=1$, $\mathrm{Cov}(E_t^1,E_{t-\tau}^2)=0$, $p^1=19$, $h^1=1$, $h^2=0.5$. Then, $u_t=40\cdot\phi[\Phi^{-1}(0.95)]\cdot[\sqrt{2}-1]=1.7088$, and so $u^*=1.69$. Therefore, p^2 should be lower than $h^2\cdot\Phi(1.69)/[1-\Phi(1.69)]=10.487$. This means that the service level provided by the supplier should be at most $10.487/(10.487+0.5)=95.45\%$. Thus, a rule of thumb that calls for the supplier to hold a safety stock which equals, e.g., three standard deviations of his forecast error (equivalent to $p^2=370$) would not only be bad, but even worse than the case in which γ^2 is set to a large negative value (so that the supplier does not carry *any* inventory, whatsoever). Proposition 13 suggests, for example, that when the covariance $C=\mathrm{Cov}(R^1,R^2)$ is large, one may see more benefits in a decoupling strategy than in the case when it is low.

7. CONCLUSIONS

The main contribution of this paper is that it proposes a unified time-series framework for forecasting and inventory control in supply chains that are rich with demand information. We contend that such a framework is timely and important for several reasons. First, it provides forecasting and replenishment procedures that can cope with a high variety of product demand characteristics, and with demand models that may need to be revised and modified over time. The matrix form-based procedures and expressions can be coded in software modules with great ease and can be used without the need to redevelop complex analytical formulae for forecasting each particular demand process. Second, the models in this paper enable the study of time-series models that can capture enriched information spaces, and not only the history of demand. Third, we provide a simple methodology for investigating the benefits of various types of information-sharing options (e.g., sharing subsets of the demand-related information, sharing information in one direction of the channel, etc.). The latter topic has engendered a great deal of attention recently not only in the academic community, but also from practitioners in the business world. We believe that the above three reasons are particularly relevant today, when companies collect a vast amount of demand-related data, given the high proliferation of product variety, and given the fact that companies are more open today to the idea of sharing information with their trading partners. Therefore, one should not view this paper as a collection of models that enable one to retrieve the results in the existing literature, as special cases of our underlying linear state space model. Rather, we encourage the reader to think about new potential applications of our models, examples of which we have provided throughout this paper.

We also devote a part of this paper to the development of cost estimation procedures. The cost expressions allow us to observe that system performance, and particularly that of coordinated supply chains, depends on the individual members' forecasting performance as well as on the *correlation* between them. Our cost assessment techniques may be valuable to supply-chain managers when making strategic inventory positioning decisions. We have demonstrated that extensions of our models to larger distribution networks are not hard, conceptually. The main challenge in the treatment of multi-item, multiechelon systems is to identify minimal state representation. We have not dealt with this task in this research, and more research in this direction may be valuable. We also identify the study of information-rich supply-chain systems under incomplete knowledge of the system parameters as important. Our discussion of this topic was relatively brief: We have argued as to why such an assumption may not be bad in practice, and we also provided a short description of how a supplier can generate forecasts of retailer's orders, even without a specific knowledge of the system parameters (§4.1.1). Finally, we have not considered time-variant demand models (i.e., ones in which the parameter matrices/scalars are time dependent) in our work. Nevertheless, it should be noted that Kalman prediction methods are readily available for such cases, and so extensions in this direction are perhaps possible.

ACKNOWLEDGMENTS

The author thanks the anonymous associate editor and referees for their valuable feedback on the previous versions of this work. The author also thanks Lingxiu Dong for valuable discussions and comments.

REFERENCES

Anderson, B. D. O., J. B. Moore. 1979. *Optimal Filtering*. Prentice-Hall, Englewood Cliffs, NJ.

Aviv, Y. 2001. The effect of collaborative forecasting on supply chain performance. *Management Sci.* **47**(10) 1326–1343.

——. 2002. Gaining benefits from joint forecasting and replenishment processes: The case of auto-correlated demand. *Manufacturing Service Oper. Management* **4**(1) 55–74.

Azoury, K. S. 1985. Bayes solution to dynamic inventory models under unknown demand distribution. *Management Sci.* **31** 1150–1160.

Bourland, K., S. Powell, D. Pyke. 1996. Exploring timely demand information to reduce inventories. *Eur. J. Oper. Res.* **92** 239–253.

Brockwell, P. J., R. A. Davis. 1996. *Time Series: Theory and Methods*, 2nd ed. Springer-Verlag, New York.

Cachon, G. P., M. Fisher. 2000. Supply chain inventory management and the value of shared information. *Management Sci.* **46**(8) 1032–1048.

Chan, S. W., G. C. Goodwin, K. S. Sin. 1984. Convergence properties of the Riccati difference equation in optimal filtering of nonstabilizable systems. *IEEE Trans. Automatic Control* **AC-29** 10–18.

Chen F., Z. Drezner, J. K. Ryan, D. Simchi-Levi. 1999. The bullwhip effect: Managerial insights on the impact of forecasting on variability in a supply chain. Chapter 14 in S. Tayur, R. Ganeshan, M. Magazine, eds. *Quantitative Models for Supply Chain Management*. Kluwer Academic Publishers Group, Norwell, MA.

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

Dong, L., H. L. Lee. 2001. Optimal policies and approximations for a serial multi-echelon inventory system with time correlated demand. Working paper, Washington University, St. Louis, MO.

Gallego, G., Ö. Özer. 2001. Integrating replenishment decisions with advanced demand information. *Management Sci.* **47**(10) 1344–1360.

Gavirneni, S., R. Kapuscinski, S. Tayur. 1999. Value of information in capacitated supply chains. *Management Sci.* **45**(1) 16–24.

Graves, S. C. 1999. A single-item inventory model for a nonstationary demand process. *Manufacturing Service Oper. Management* **1**(1) 50–61.

——, H. C. Meal, S. Dasu, Y. Qiu. 1986. Two-stage production planning in a dynamic environment. S. Axsäter, C. Schneeweiss, E. Silver, eds. *Multi-Stage Production Planning and Control. Lecture Notes in Economics and Mathematical Systems*. Springer-Verlag, Berlin, Germany.

Hamilton, J. D. 1994. *Time Series Analysis*. Princeton University Press, Princeton, NJ.

Harvey, S. C. 1990. *Forecasting, Structural Time Series Models and The Kalman Filter*. Cambridge University Press, Cambridge, U.K.

Heath, D. P., P. L. Jackson. 1994. Modeling the evolution of demand forecasts with application to safety stock analysis in production/distribution systems. *IIE Trans.* **26** 17–30.

Iglehart, D. 1964. The dynamic inventory model with unknown demand distribution. *Management Sci.* **10** 429–440.

Johnson, G., H. Thompson. 1975. Optimality of myopic inventory policies for certain dependent demand processes. *Management Sci.* **21** 1303–1307.

Kahn, J. 1987. Inventories and the volatility of production. *Amer. Econom. Rev.* **77**(4) 667–679.

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

Lee H. L., S. Whang. 2001. Demand chain excellence: A tale of two retailers. *Supply Chain Management Rev.* **5** 40–46.

——, P. Padmanabhan, S. Whang. 1997a. Information distortion in a supply chain: The bullwhip effect. *Management Sci.* **43** 546–558.

——, ——, ——. 1997b. Bullwhip effect in a supply chain. *Sloan Management Rev.* **38**(Spring) 93–102.

——, K. C. So, C. S. Tang. 2000. The value of information sharing in a two-level supply chain. *Management Sci.* **46** 626–643.

Lovejoy, W. S. 1990. Myopic policies for some inventory models with uncertain demand distributions. *Management Sci.* **36**(6) 724–738.

Raghunathan, S. 2001. Information sharing in a supply chain: A note on its value when demand is nonstationary. *Management Sci.* **47** 605–610.

Scarf, H. 1959. Bayes solutions of the statistical inventory problem. *Ann. Math. Statist.* **30** 490–508.

——. 1960. Some remarks on Bayes solutions to the inventory problem. *Naval Logist. Res. Quart.* **7** 591–596.

Song, J. S., P. Zipkin. 1993. Inventory control in a fluctuating demand environment. *Oper. Res.* **41** 351–370.

——, ——. 1996. Managing inventory with the prospect of obsolescence. *Oper. Res.* **44** 215–222.

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